

IRMA Lectures in Mathematics and Theoretical Physics 13

Edited by Christian Kassel and Vladimir G. Turaev

Institut de Recherche Mathématique Avancée
CNRS et Université de Strasbourg
7 rue René-Descartes
67084 Strasbourg Cedex
France

IRMA Lectures in Mathematics and Theoretical Physics

Edited by Christian Kassel and Vladimir G. Turaev

This series is devoted to the publication of research monographs, lecture notes, and other material arising from programs of the Institut de Recherche Mathématique Avancée (Strasbourg, France). The goal is to promote recent advances in mathematics and theoretical physics and to make them accessible to wide circles of mathematicians, physicists, and students of these disciplines.

Previously published in this series:

- 1 Deformation Quantization, *Gilles Halbout* (Ed.)
- 2 Locally Compact Quantum Groups and Groupoids, *Leonid Vainerman* (Ed.)
- 3 From Combinatorics to Dynamical Systems, *Frédéric Fauvet and Claude Mitschi* (Eds.)
- 4 Three courses on Partial Differential Equations, *Eric Sonnendrücker* (Ed.)
- 5 Infinite Dimensional Groups and Manifolds, *Tilman Wurzbacher* (Ed.)
- 6 *Athanase Papadopoulos*, Metric Spaces, Convexity and Nonpositive Curvature
- 7 Numerical Methods for Hyperbolic and Kinetic Problems, *Stéphane Cordier, Thierry Goudon, Michaël Gutnic and Eric Sonnendrücker* (Eds.)
- 8 AdS/CFT Correspondence: Einstein Metrics and Their Conformal Boundaries, *Oliver Biquard* (Ed.)
- 9 Differential Equations and Quantum Groups, *D. Bertrand, B. Enriquez, C. Mitschi, C. Sabbah and R. Schäfke* (Eds.)
- 10 Physics and Number Theory, *Louise Nyssen* (Ed.)
- 11 Handbook of Teichmüller Theory, Volume I, *Athanase Papadopoulos* (Ed.)
- 12 Quantum Groups, *Benjamin Enriquez* (Ed.)

Volumes 1–5 are available from Walter de Gruyter (www.degruyter.de)

Handbook of Teichmüller Theory Volume II

Athanase Papadopoulos
Editor



European Mathematical Society

Editor:

Athanase Papadopoulos
Institut de Recherche Mathématique Avancée
CNRS et Université de Strasbourg
7 Rue René Descartes
67084 Strasbourg Cedex
France

2000 Mathematics Subject Classification: Primary 30-00, 32-00, 32G15, 30F60; secondary 14A20, 14C17, 14D20, 14H15, 14H60, 14L24, 14L35, 16R30, 20F36, 20F38, 20F65, 22F10, 20H10, 30F10, 30F15, 30F20, 30F25, 30C62, 51N15, 53A20, 53A35, 53B35, 53C24, 53C25, 53C50, 57M07, 57M60, 57R20.

ISBN 978-3-03719-055-5

The Swiss National Library lists this publication in The Swiss Book, the Swiss national bibliography, and the detailed bibliographic data are available on the Internet at <http://www.helvetica.ch>.

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. For any kind of use permission of the copyright owner must be obtained.

© 2009 European Mathematical Society

Contact address:

European Mathematical Society Publishing House
Seminar for Applied Mathematics
ETH-Zentrum FLI C4
CH-8092 Zürich
Switzerland

Phone: +41 (0)44 632 34 36
Email: info@ems-ph.org
Homepage: www.ems-ph.org

Typeset using the author's $\text{T}_\text{E}\text{X}$ files: I. Zimmermann, Freiburg
Printed in Germany

9 8 7 6 5 4 3 2 1

Foreword

Classical Teichmüller theory concerns moduli spaces of conformal structures on surfaces. By the uniformization theorem, any conformal structure on a surface can be represented by a unique complete Riemannian metric of constant curvature -1 , 0 or 1 . From this point of view, Teichmüller theory can also be considered as the study of moduli spaces of metrics of constant curvature -1 , 0 or 1 on surfaces. In most cases (more precisely, when the Euler characteristic of the surface is negative), the curvature is negative, and Teichmüller theory can be viewed as the theory of moduli spaces of hyperbolic structures, that is, metrics of constant curvature -1 on surfaces.

In this multi-volume Handbook, the expression “Teichmüller theory” is used in a broader sense, namely, as the study of moduli of general geometric structures on surfaces, with methods inspired or adapted from those of classical Teichmüller theory. Such a theory has ramifications in group theory, in representation theory, in dynamical systems, in symplectic geometry, in three- and four-manifolds topology, and in other domains of mathematics.

The present volume of the Handbook contains four parts, namely:

Part A : The metric and the analytic theory, 2

Part B: The group theory, 2

Part C: Representation spaces and geometric structures, 1

Part D: The Grothendieck–Teichmüller theory

Parts A and B are sequels to parts with the same names in Volume I of the Handbook.

We hope that the various volumes of this Handbook will give the interested reader an overview of the old and of the recent work on Teichmüller theory and its applications, that they will open new perspectives and that they will contribute to further research in that field. In relation to future developments, it is worth mentioning that several chapters of the present volume contain a discussion of open problems. These include the chapters by Kojima, Korkmaz & Stipsicz, Möller, Šarić, Fletcher & Markovic, and Fujiwara.

Finally, let me mention that some of the contributions that were announced to appear in this volume will appear in later volumes. (At the time where these contributions were planned, only two volumes of the Handbook were expected.)

I would like to thank again Vladimir Turaev for his encouragement in this Handbook project, and Irene Zimmermann from the EMS publishing House for the seriousness of her work. Of course, I thank all of the 24 authors who contributed to this volume for their pleasant and fruitful collaboration.

Strasbourg, February 2009

Athanase Papadopoulos

Contents

Foreword.....	v
Introduction to Teichmüller theory, old and new, II <i>by Athanase Papadopoulos</i>	1
Part A. The metric and the analytic theory, 2	
Chapter 1. The Weil-Petersson metric geometry <i>by Scott A. Wolpert</i>	47
Chapter 2. Infinite dimensional Teichmüller spaces <i>by Alastair Fletcher and Vladimir Markovic</i>	65
Chapter 3. A construction of holomorphic families of Riemann surfaces over the punctured disk with given monodromy <i>by Yoichi Imayoshi</i>	93
Chapter 4. The uniformization problem <i>by Robert Silhol</i>	131
Chapter 5. Riemann surfaces, ribbon graphs and combinatorial classes <i>by Gabriele Mondello</i>	151
Chapter 6. Canonical 2-forms on the moduli space of Riemann surfaces <i>by Nariya Kawazumi</i>	217
Part B. The group theory, 2	
Chapter 7. Quasi-homomorphisms on mapping class groups <i>by Koji Fujiwara</i>	241
Chapter 8. Lefschetz fibrations on 4-manifolds <i>by Mustafa Korkmaz and András I. Stipsicz</i>	271

Chapter 9. Introduction to measurable rigidity of mapping class groups <i>by Yoshikata Kida</i>	297
Chapter 10. Affine groups of flat surfaces <i>by Martin Möller</i>	369
Chapter 11. Braid groups and Artin groups <i>by Luis Paris</i>	389
Part C. Representation spaces and geometric structures, 1	
Chapter 12. Complex projective structures <i>by David Dumas</i>	455
Chapter 13. Circle packing and Teichmüller space <i>by Sadayoshi Kojima</i>	509
Chapter 14. $(2 + 1)$ Einstein spacetimes of finite type <i>by Riccardo Benedetti and Francesco Bonsante</i>	533
Chapter 15. Trace coordinates on Fricke spaces of some simple hyperbolic surfaces <i>by William M. Goldman</i>	611
Chapter 16. Spin networks and $SL(2, \mathbb{C})$ -character varieties <i>by Sean Lawton and Elisha Peterson</i>	685
Part D. The Grothendieck–Teichmüller theory	
Chapter 17. Grothendieck’s reconstruction principle and 2-dimensional topology and geometry <i>by Feng Luo</i>	733
Chapter 18. Dessins d’enfants and origami curves <i>by Frank Herrlich and Gabriela Schmithüsen</i>	767

Chapter 19. The Teichmüller theory of the solenoid <i>by Dragomir Šarić</i>	811
List of Contributors	857
Index	859

Introduction to Teichmüller theory, old and new, II

Athanase Papadopoulos

Contents

1	The metric and the analytic theory	2
1.1	Weil–Petersson geometry	2
1.2	The quasiconformal theory	4
1.3	Holomorphic families	6
1.4	Uniformization	9
1.5	Combinatorial classes	9
1.6	Differential forms	12
2	The group theory	13
2.1	Quasi-homomorphisms	13
2.2	Lefschetz fibrations	15
2.3	Measure-equivalence	17
2.4	Affine groups	20
2.5	Braid groups	22
3	Representation spaces and geometric structures	25
3.1	Complex projective structures	26
3.2	Circle packings	29
3.3	Lorentzian geometry	31
3.4	Fricke–Klein coordinates	34
3.5	Diagrammatic approach	34
4	The Grothendieck–Teichmüller theory	35
4.1	The reconstruction principle	38
4.2	Dessins d’enfants	39
4.3	The solenoid	41

This introduction can be considered as a sequel to the introduction that I wrote for Volume I of the Handbook, and I shall limit myself here to a general presentation of the material covered in the present volume. The exposition will follow the four-parts division of the volume, and for each part, its division in chapters.

Beyond the information given on the content of this volume, I hope that the reader of this introduction will get (if he does not have it yet) an idea of the richness of the subject of Teichmüller theory.

All the surfaces considered in this introduction are orientable, unless otherwise stated. I have tried to give some necessary definitions to make the introduction as much self-contained as possible.

1 The metric and the analytic theory

Part A of this volume, on the metric and analytic theory of Teichmüller space, contains chapters on Weil–Petersson geometry, on biholomorphic maps between finite or infinite-dimensional Teichmüller spaces, on the theory of holomorphic families of Riemann surfaces, on uniformization of algebraic surfaces, on combinatorial classes in moduli space and on canonical differential forms on that space representing cohomology classes.

1.1 Weil–Petersson geometry

Chapter 1 by Scott Wolpert is a review of some recent work on the Weil–Petersson metric of $\mathcal{T}_{g,n}$, the Teichmüller space of a surface of genus $g \geq 0$ with $n \geq 0$ punctures, with negative Euler characteristic. Let us start by recalling some basic facts about this metric.

It is well known that the cotangent space to $\mathcal{T}_{g,n}$ at a point represented by a Riemann surface S can be identified with the space $Q(S)$ of holomorphic quadratic differentials on S that have at most simple poles at the punctures. The Weil–Petersson cometric on that cotangent space is given by the Hermitian product $\int_S \phi(z)\bar{\psi}(z)\rho^{-2}(z)|dz|^2$, for ϕ and ψ in $Q(S)$, where ρ is the density form of the length element $\rho(z)|dz|$ of the unique complete hyperbolic metric that uniformizes the Riemann surface S .¹

The Weil–Petersson metric on $\mathcal{T}_{g,n}$ is Kähler, geodesically convex and with negative and unbounded sectional curvature (its supremum is zero, and its infimum is $-\infty$). Its Ricci curvature is bounded from above by a negative constant. This metric is not complete, and a geodesic of bounded length can be obtained by making the hyperbolic length of a closed geodesic on the surface tend to zero. The last fact explains intuitively why the completion of the Weil–Petersson metric gives rise to the *augmented* Teichmüller space $\bar{\mathcal{T}}_{g,n}$, whose elements are equivalence classes of marked stable Riemann surfaces, that is, marked Riemann surfaces with nodes, with the property that each connected component of the complement of the nodes is a surface with cusps which has negative Euler characteristic. The space $\bar{\mathcal{T}}_{g,n}$ is a stratified space which is not locally compact and which is a partial compactification of $\mathcal{T}_{g,n}$. The action of the mapping class group on $\mathcal{T}_{g,n}$ extends to an action on $\bar{\mathcal{T}}_{g,n}$, and the quotient of $\bar{\mathcal{T}}_{g,n}$ by

¹The name Weil–Petersson has been given to this metric because it was André Weil who first noticed that this product, called the Petersson product and originally introduced by Hans Petersson on the space of modular forms, gives a metric on Teichmüller space.

this action is a compact orbifold, known as the Deligne–Mumford stable curve compactification of moduli space. In 1976, H. Masur obtained a beautiful result stating that the Weil–Petersson metric on Teichmüller space extends to a complete metric on the augmented Teichmüller space $\overline{\mathcal{T}}_{g,n}$. This result is one of the starting points for a topological approach to the Weil–Petersson metric.

Our knowledge of the Weil–Petersson geometry underwent a profound transformation at the beginning of 1980s, thanks to the work of Scott Wolpert, who obtained a series of particularly elegant results on the Weil–Petersson metric and on its associated symplectic form. New important results on the subject, from various points of view, were obtained in the last few years by several authors, including Wolpert, Yamada, Huang, Liu, Sun, Yau, McMullen, Mirzakhani, Brock, Margalit, Daskalopoulos and Wentworth (there are others). The recent work on Weil–Petersson geometry includes the study of the CAT(0) geometry of augmented Teichmüller space, that is, the study of its nonpositive curvature geometry in the sense of Cartan–Alexandrov–Toponogov (following a terminology introduced by Gromov). We recall that the definition of a CAT(0) metric space is based on the comparison of distances between points on the edges of arbitrary triangles in that metric space with distances between corresponding points on “comparison triangles” in the Euclidean plane. It is known that augmented Teichmüller space, equipped with the extension of the Weil–Petersson metric, is a complete CAT(0) metric space (a result due to Yamada). The Weil–Petersson isometry group action extends continuously to an action on augmented space. The Weil–Petersson isometry group coincides with the extended mapping class group of the surface except for some special surfaces (a result of Masur & Wolf, completed to some left-out special cases by Brock & Margalit, which parallels a famous result by Royden for the Teichmüller metric, completed by Earle & Kra). An analysis of the action of the mapping class group in the spirit of Thurston’s classification of mapping classes, showing in particular the existence of invariant Weil–Petersson geodesics for pseudo-Anosov mapping classes, has been carried out by Daskalopoulos & Wentworth. Brock established that (augmented) Teichmüller space equipped with the Weil–Petersson metric is quasi-isometric to the pants graph of the surface.

In Chapter 1 of this volume, Wolpert makes a review of the recent results on the metric aspect (as opposed to the analytical aspect) of the Weil–Petersson metric. He reports on a parametrization of augmented Teichmüller space using Fenchel–Nielsen coordinates and on a comparison between the Weil–Petersson metric and the Teichmüller metric in the thin part of Teichmüller space, using these coordinates. He gives formulae for the Hessian and for the gradient of the hyperbolic geodesic length functions and for the behaviour of these functions near degenerate hyperbolic surfaces. He also gives formulae for the Weil–Petersson symplectic form in terms of geodesic length functions. Weil–Petersson convexity and curvature are also reviewed. The chapter also contains a section on Alexandrov angles, in relation with Alexandrov tangent cones at points of the augmented Teichmüller space. Wolpert gives estimates on the exponential map, with applications to the first variation formula for the distance and to the length-minimizing paths connecting two given points and intersecting a

prescribed stratum. He displays a table comparing the known metric properties of the Teichmüller space of a surface of negative Euler characteristic with corresponding properties of the hyperbolic plane, which, as is well known, is the Teichmüller space of the torus.

1.2 The quasiconformal theory

In Chapter 2, Alastair Fletcher and Vladimir Markovic study analytic properties of finite-dimensional as well as infinite-dimensional Teichmüller spaces. They review some classical properties and they present some recent results, in particular concerning biholomorphic maps between Teichmüller spaces.

We recall that a Riemann surface is said to be of *finite topological type* if its fundamental group is finitely generated. It is said of *finite analytical type* if it is obtained (as a complex space) from a closed Riemann surface by removing a finite set of points. The Teichmüller space $\mathcal{T}(S)$ of a Riemann surface S is a Banach manifold which is finite-dimensional if and only if S is of finite analytical type. (Note that $\mathcal{T}(S)$ can be infinite-dimensional even if S has finite topological type.) A surface with border has an *ideal boundary*, which is the union of its *ideal boundary curves*, and the Teichmüller space of a surface with nonempty border is infinite-dimensional. The most important surface with border is certainly the unit disk $\mathbb{D} \subset \mathbb{C}$, and its Teichmüller space is called *universal Teichmüller space*. This space contains all Teichmüller spaces of Riemann surfaces, as we shall recall below.

In this chapter, S is a surface of finite or infinite type.

The Teichmüller space $\mathcal{T}(S)$ of a Riemann surface S is defined as a space of equivalence classes of marked Riemann surfaces (S', f) , with the marking f being a quasiconformal homeomorphism between the base surface S and a Riemann surface S' . We recall that for infinite-dimensional Teichmüller spaces, the choice of a base Riemann surface is an essential part of the definition, since homeomorphic Riemann surfaces are not necessarily quasiconformally equivalent. Teichmüller space can also be defined as a space of equivalence classes of Beltrami differentials on a given base Riemann surface. The relation between the two definitions stems from the fact that a quasiconformal mapping from a Riemann surface S to another Riemann surface is the solution of an equation of the form $f_{\bar{z}} = \mu f_z$ (called a Beltrami equation), with μ a Beltrami differential on S .

Fletcher and Markovic also deal with universal Teichmüller space. This is a space of equivalence classes of normalized quasiconformal homeomorphisms of the unit disk \mathbb{D} . It is well known that quasiconformal maps of \mathbb{D} extend to the boundary $\partial\mathbb{D}$ of \mathbb{D} . Such quasiconformal maps are normalized so that their extension to the boundary fixes the points 1 , -1 and i , and two quasiconformal self-maps of the disk are considered to be equivalent if they induce the same map on $\partial\mathbb{D}$. Like the other Teichmüller spaces, universal Teichmüller space can also be defined as a space of equivalence classes of Beltrami differentials. By lifting quasiconformal homeomorphisms or Beltrami differentials from a surface to the universal cover, the Teichmüller space of any surface

of hyperbolic type embeds in the universal Teichmüller space, and it is in this sense that the universal Teichmüller space is called “universal”.

The complex Banach structure of each Teichmüller space $\mathcal{T}(S)$ can be obtained from the so-called Bers embedding of $\mathcal{T}(S)$ into the Banach space $Q(S)$ of holomorphic quadratic differentials on the base surface S . In the case where S is of finite analytical type, this embedding provides a natural identification between the cotangent space at a point of $\mathcal{T}(S)$ and a Banach space Q of integrable holomorphic quadratic differentials, and the two spaces are finite-dimensional. In the general case, the spaces considered are not necessarily finite-dimensional, and the cotangent space at a point of $\mathcal{T}(S)$ is the *predual* of the Banach space Q , that is, a space whose dual is Q . The predual of Q is called the *Bergman space* of S . This distinction, which is pointed out by Fletcher and Markovic, is an important feature of the theory of infinite-dimensional Teichmüller spaces.

It is well known that the complex-analytic theory of finite-dimensional Teichmüller spaces can be developed using more elementary methods than those that involve the Bers embedding. For instance, for surfaces of finite analytical type, Ahlfors defined the complex structure of Teichmüller space using period matrices obtained by integrating systems of independent holomorphic one-forms over a basis of the homology of the surface. The complex analytic structure on Teichmüller space is then the one that makes the period matrices vary holomorphically. The description of the complex structure in the infinite-dimensional case requires more elaborate techniques.

Along the same line, we note some phenomena that occur in infinite-dimensional Teichmüller theory and not in the finite-dimensional one. There is a “mapping class group action” on infinite-dimensional Teichmüller spaces, but, unlike the finite dimensional case, this action is not always discrete. (Here, discreteness means that the orbit of any point under the group action is discrete.) Katsuhiko Matsuzaki studied limit sets and domains of discontinuity for such actions, in the infinite-dimensional case. From the metric-theoretic point of view, Zhong Li and Harumi Tanigawa proved that in each infinite-dimensional Teichmüller space, there are pairs of points that can be connected by infinitely many distinct geodesic segments (for the Teichmüller metric). This contrasts with the finite-dimensional case where the geodesic segment connecting two given points is unique. Li proved non-uniqueness of geodesic segments connecting two points in the universal Teichmüller space, and he showed that there are closed geodesics in any infinite-dimensional Teichmüller space. He also proved that the Teichmüller distance function, in the infinite-dimensional case, is not differentiable at some pairs of points in the complement of the diagonal, in contrast with the finite-dimensional case where, by a result of Earle, the Teichmüller distance function is continuously differentiable outside the diagonal.²

The mention of these differences between the finite- and infinite-dimensional cases will certainly give more importance to the results on isometries and biholomorphic

²The study of the differentiability of the Teichmüller distance function was initiated by Royden, and it was continued by Earle. More precise results on the differentiability of this function were obtained recently by Mary Rees.

maps between infinite-dimensional Teichmüller spaces that are reported on here by Fletcher and Markovic, since these are results that hold in both the finite- and in the infinite-dimensional cases. Fletcher and Markovic study biholomorphic maps between Teichmüller spaces by examining their induced actions on cotangent spaces (and Bergman spaces). In the finite-dimensional case, the idea of studying the action on cotangent space is already contained in the early work of Royden. The action of a biholomorphic map induces a \mathbb{C} -linear isometry between Bergman spaces. Fletcher and Markovic report on a rigidity result, whose most general form is due to Markovic, and with special cases previously obtained by Earle & Kra, Lakic and Matsuzaki. The result says that any surjective \mathbb{C} -linear isometry between the Bergman spaces $A^1(M)$ and $A^1(N)$ of two surfaces M and N is geometric, except in the case of some elementary surfaces. Roughly speaking, the word “geometric” means here that the isometry is a composition of two naturally defined isometries between such spaces, viz. multiplication by a complex number of norm one, and an isometry induced by the action of a conformal map between the surfaces. A corollary of this result is that the biholomorphic automorphism group of the Teichmüller space of a surface of non-exceptional (finite or infinite) type can be naturally identified with the mapping class group of that surface.

As in the finite-dimensional case, this result reduces the study of biholomorphic homeomorphisms between Teichmüller spaces to the study of linear isometries between some Banach spaces. In the course of proving this result, a proof is given of the fact that the Kobayashi and the Teichmüller metrics on (finite- or infinite-dimensional) Teichmüller space agree, again generalizing a result obtained by Royden and completed by Earle & Kra for finite type Riemann surfaces.

Chapter 2 of this volume also contains the proof of a local rigidity result due to Fletcher, saying that the Bergman spaces of any two surfaces whose Teichmüller spaces are infinite-dimensional are always isomorphic, and that any two infinite-dimensional Teichmüller spaces are locally bi-Lipschitz equivalent. More precisely, Fletcher proved that the Teichmüller metric on every Teichmüller space of an infinite-type Riemann surface is locally bi-Lipschitz equivalent to the Banach space l^∞ of bounded sequences with the supremum norm.

1.3 Holomorphic families

A *holomorphic family of Riemann surfaces of type (g, n)* is a triple (M, π, B) defined as follows:

- M is a 2-dimensional complex manifold (topologically, a 4-manifold);
- B is a Riemann surface;
- $\pi : M \rightarrow B$ is a holomorphic map;
- for all $t \in B$, the fiber $S_t = \pi^{-1}(t)$ is a Riemann surface of genus g with n punctures;
- the complex structure on S_t depends holomorphically on the parameter t .

Chapter 3, by Yoichi Imayoshi, concerns holomorphic families of Riemann surfaces. In all this chapter, it is assumed that $2g - 2 + n > 0$.

Why do we study holomorphic families of Riemann surfaces? One reason is that one way of investigating the complex analytic structure of Teichmüller space involves the study of holomorphic families. Another reason is that the study of degeneration of holomorphic families is related to the study of the stable curve compactification of moduli space.

To be more precise, we use the following notation: as before, $\mathcal{T}_{g,n}$ is the Teichmüller space of a surface of type (g, n) , that is, of genus g with n punctures and $\mathcal{M}_{g,n}$ is the corresponding moduli space. A holomorphic family (M, π, B) of type (g, n) gives rise to a holomorphic map $\Phi: \tilde{B} \rightarrow \mathcal{T}_{g,n}$, where \tilde{B} is the universal cover of B , and to a quotient holomorphic map $\tilde{B} \rightarrow \mathcal{M}_{g,n}$ called the *moduli map* of the family. A basic combinatorial tool in the study of the holomorphic family (M, π, B) is its *topological monodromy*, which is a homomorphism from the fundamental group of the base surface B to the mapping class group $\Gamma_{g,n}$ of a chosen Riemann surface $S_{g,n}$ of type (g, n) . In Chapter 3, this homomorphism is denoted by Φ_* , because its definition makes use of the map Φ . It is defined through the action of the mapping class group $\Gamma_{g,n}$ on the Teichmüller space $\mathcal{T}_{g,n}$.

Imayoshi reports on an important rigidity theorem stating that if (M_1, π_1, B) and (M_2, π_2, B) are locally non-trivial holomorphic families of Riemann surfaces of type (g, n) over the same base B , and if $(\Phi_1)_* = (\Phi_2)_*$, then $\Phi_1 = \Phi_2$ and (M_1, π_1, B) is biholomorphically equivalent to (M_2, π_2, B) .

Imayoshi mentions an application of this rigidity theorem to the proof of the geometric Shafarevich conjecture, which states that there are only finitely many locally non-trivial and non-isomorphic holomorphic families of Riemann surfaces of fixed finite type over a Riemann surface B of finite type. This conjecture was proved by Parshin in the case where B is compact, and by Arakelov in the general case. Imayoshi and Shiga gave a variant of the proof, using the rigidity theorem stated above. Imayoshi notes that the same rigidity theorem can be used to give a proof of the geometric Mordell conjecture, which concerns the existence of holomorphic sections for holomorphic families.

A large part of the study made in Chapter 3 concerns the case where the base surface B is the unit disk in \mathbb{C} punctured at the origin. We denote by Δ^* this punctured disk. In many ways, taking as base surface the punctured disk is sufficient for the study of the degeneration theory of holomorphic families. It may also be useful to recall here that the Deligne–Mumford stable reduction theorem for the moduli space of curves reduces the study of the stable (Deligne–Mumford) compactification of moduli space to that of holomorphic families over the punctured disk which degenerate by producing surfaces with nodes above the puncture.

In the 1960s, Kodaira began a study of holomorphic families over the punctured disk, in the special case where the fibers are surfaces of type $(1, 0)$. He studied in particular the behaviour of *singular fibers* of such families, that is, fibers obtained by

extending the family at the puncture. After this work, Kodaira and others considered singular fibers of more general families. This is also reported on in Chapter 3 of this Handbook.

In the case where the base surface is the punctured disk Δ^* , the topological monodromy is a cyclic group, and it gives rise to an element of the mapping class group of a fiber, called the *topological monodromy around the origin*. This element is defined after a choice of a basepoint s in Δ^* and after the identification of the fiber $\pi^{-1}(s)$ above that point with a fixed marked topological surface S . The topological monodromy is then the element of the mapping class group of S that performs the gluing as one traverses the circle in Δ^* centered at the origin and passing through s . The topological monodromy of the family is well defined up to conjugacy (the ambiguity being due to the choice of a surface among the fibers, and of its identification with a fixed marked surface).

In 1981, Imayoshi studied monodromies of holomorphic families (M, π, Δ^*) in connection with the deformation theory of Riemann surfaces with nodes. In particular, he proved that the topological monodromy of a family (M, π, Δ^*) is *pseudo-periodic*³, which means that this mapping class contains an orientation-preserving homeomorphism f that preserves a (possibly empty) collection $\{C_1, \dots, C_k\}$ of disjoint homotopically nontrivial and pairwise non-homotopic simple closed curves on the surface, such that for each $i = 1, \dots, k$, there exists an integer n_i such that a certain power of f is the composition of n_i -th powers of Dehn twists along the C_i 's.

Imayoshi studied a map from the punctured disk to the moduli space $\mathcal{M}_{g,n}$ of S which is canonically associated to the family (M, π, Δ^*) , and he showed that this map extends holomorphically to a map from the unit disk Δ to the Deligne–Mumford compactification $\bar{\mathcal{M}}_{g,n}$ of $\mathcal{M}_{g,n}$. He showed that algebraic properties of the topological monodromy (e.g. the fact that it is of finite or infinite order) depend on whether the image of 0 by the holomorphic map $\Delta \rightarrow \bar{\mathcal{M}}_{g,n}$ lies in $\mathcal{M}_{g,n}$ or in $\bar{\mathcal{M}}_{g,n} - \mathcal{M}_{g,n}$. He also showed that the topological monodromy is of *negative type*, meaning that it can be represented by a homeomorphism f of the fiber which is either periodic, or, using the above notation, such that the Dehn twists around the C_i 's are negative Dehn twists. Chapter 3 of this volume contains a new proof of Imayoshi's 1981 result.

Y. Matsumoto & J. M. Montesinos-Amilibia and (independently) S. Takamura proved recently a converse to Imayoshi's result. More precisely, starting with any pseudo-periodic self-map of negative type of a Riemann surface $S_{g,n}$ satisfying $2g - 2 + n > 0$, they constructed a holomorphic family of Riemann surfaces over the punctured disk whose monodromy is the given map up to conjugacy. Matsumoto and Montesinos-Amilibia showed that the ambient topological type of the singular fiber is determined by the monodromy.

³Such a mapping class is of elliptic type or of parabolic type in the Bers terminology of the Thurston classification of mapping classes.

1.4 Uniformization

Chapter 4, by Robert Silhol, concerns the problem of uniformization of Riemann surfaces defined by algebraic equations.

By the classical Poincaré–Koebe uniformization theorem, one can associate to any compact Riemann surface S of negative Euler characteristic a Fuchsian group, that is, a discrete subgroup Γ of $\mathrm{PSL}(2, \mathbb{R})$ acting on the hyperbolic plane \mathbb{H}^2 , such that S is conformally equivalent to the hyperbolic manifold \mathbb{H}^2/Γ . All the known proofs of the uniformization theorem are rather involved, and it is not an easy matter to explicitly exhibit the hyperbolic structure \mathbb{H}^2/Γ that uniformizes a given Riemann surface S . Silhol discusses this problem for the case where the Riemann surface S is given explicitly as an algebraic curve over \mathbb{C} , that is, as the zero set of a two-variable polynomial with coefficients in \mathbb{C} . We recall that by a result of Riemann, any compact Riemann surface can be defined as an algebraic curve. We note in passing that the question of what is the “best” field of coefficients for a polynomial defining a given Riemann surface can be dealt with in the setting of Grothendieck’s theory of dessins d’enfants, which is treated in another chapter of this volume. It is also worth noting that defining Riemann surfaces by algebraic equations does not necessarily reveal all the aspects of the complex structure of that surface. For instance, the problem of finding the holomorphic automorphism group of a Riemann surface given by means of an algebraic equation is not tractable in general.

Silhol presents classical and recent methods that are used in the study of the following two problems, which he calls the *uniformization problem* and the *inverse uniformization problem* respectively:

- given a Riemann surface S defined as an algebraic curve over \mathbb{C} , find its associated hyperbolic structure;
- given a discrete subgroup Γ of $\mathrm{PSL}(2, \mathbb{R})$ acting on \mathbb{H}^2 and satisfying certain conditions, find an algebraic curve representing the Riemann surface $S = \mathbb{H}^2/\Gamma$.

The methods that are used in the study of these problems involve the Schwarzian differential equation, theta functions, Poincaré series and other automorphic forms. The chapter also contains the discussion of explicit examples. The author also reports on recent work on the uniformization problem, by himself and S. Lelièvre, based on methods that were introduced by Fricke and Klein. This work concerns the uniformization of certain families of complex algebraic curves by hyperbolic surfaces obtained by gluing hyperbolic triangles or quadrilaterals along their boundaries.

Other questions related to uniformization are addressed in Chapter 18 of this volume, by Herrlich and Schmithüsen.

1.5 Combinatorial classes

In Chapter 5, Gabriele Mondello gives a detailed survey of the use of ribbon graphs in Teichmüller theory, in particular in the investigation of combinatorial classes in moduli space.

In this chapter, S is a compact oriented surface of genus $g \geq 0$ equipped with a nonempty finite subset of points X of cardinality n satisfying $2g - 2 + n > 0$, called the marked points. As before, $\mathcal{T}_{g,n}$ and $\mathcal{M}_{g,n}$ denote respectively the Teichmüller and the moduli space of the pair (S, X) . A *ribbon graph* (also called a *fatgraph*) associated to (S, X) is a finite graph G embedded in $S - X$ such that the inclusion $G \hookrightarrow (S - X)$ is a homotopy equivalence.

Mondello describes the two main methods that have been used so far for defining ribbon graphs in the context of Teichmüller space. One definition uses complex analysis, namely, Jenkins–Strebel quadratic differentials, and the other definition uses hyperbolic geometry, more precisely, Penner’s decoration theory.

We recall that a Jenkins–Strebel differential on a Riemann surface with marked points is a meromorphic quadratic differential with at worst double poles at the marked points, whose horizontal foliation has all of its regular leaves compact. A Jenkins–Strebel differential defines a flat metric on the surface, with isolated cone singularities. The surface, as a metric space, is obtained by gluing a finite collection of Euclidean cylinders along their boundaries. The combinatorics of this cylinder decomposition of the surface is encoded by a ribbon graph.

Ribbon graphs, as they are used in Chapter 5 of this Handbook, are equipped with weights, and are called *metric ribbon graphs*. The weights, in the case just described, come from the restriction of the singular flat metric to the cylinders.

In the hyperbolic geometry approach, one considers complete finite area hyperbolic metrics on the punctured surface $S - X$. Neighborhoods of punctures are cusps and, around each cusp, there is a cylinder foliated by closed horocycles, that is, closed leaves whose lifts to the universal cover of S are pieces of horocycles of \mathbb{H}^2 . A *decoration* on a hyperbolic punctured surface of finite area is the choice of a horocycle around each puncture. Again, these data are encoded by a metric ribbon graph.

There is a natural combinatorial structure on the space of ribbon graphs, which encodes the combinatorics of these graphs (valencies, etc.). This structure provides, via any one of the two constructions that we mentioned above, a cellularization of the space $\mathcal{T}_{g,n} \times \Delta^{n-1}$, where Δ^{n-1} is the standard simplex in \mathbb{R}^n . This cellularization is invariant under the action of the mapping class group $\Gamma_{g,n}$, and it gives a quotient cellularization of $\mathcal{M}_{g,n} \times \Delta^{n-1}$ (in the orbifold category). The last cellularization is one of the main tools that have been used in the study of the cohomology of moduli space and of its intersection theory. The basic work on this cellularization has been done by Harer–Mumford–Thurston, by Penner and by Bowditch & Epstein.

There is a dual object to a ribbon graph, namely, an *arc system* on the surface S . This is a collection of disjoint essential arcs with endpoints in X , which are pairwise non-homotopic with endpoints fixed.

Arc systems on the pair (S, X) naturally form a flag simplicial complex, where for each $k \geq 0$, a k -simplex is an arc system with $k + 1$ components. $\mathfrak{A}^\circ(S, X)$ denotes the *interior* of the complex $\mathfrak{A}(S, X)$. This is the subset of $\mathfrak{A}(S, X)$ consisting of arc systems on $S - X$ that cut this surface into disks or pointed disks. $\mathfrak{A}^\infty(S, X) = \mathfrak{A}(S, X) - \mathfrak{A}^\circ(S, X)$ is called the *boundary* of $\mathfrak{A}(S, X)$.

Penner and Bowditch & Epstein, using decorations on hyperbolic surfaces with cusps, and Harer–Mumford–Thurston, using flat structures arising from meromorphic quadratic differentials of Jenkins–Strebel type, proved that there is a $\Gamma_{g,n}$ -equivariant homeomorphism from the geometric realization $|\mathfrak{A}^\circ(S, X)|$ to the product space $\mathcal{T}_{g,n} \times \Delta^{n-1}$. In particular, there is a $\Gamma_{g,n}$ -equivariant homotopy equivalence $|\mathfrak{A}^\circ(S, X)| \simeq \mathcal{T}_{g,n}$. Via the homeomorphism $|\mathfrak{A}^\circ(S, X)| \rightarrow \mathcal{T}_{g,n} \times \Delta^{n-1}$, the cellular structure of $|\mathfrak{A}^\circ(S, X)|$ is transported to $\mathcal{T}_{g,n} \times \Delta^{n-1}$ and the homeomorphism $|\mathfrak{A}^\circ(S, X)| \rightarrow \mathcal{T}_{g,n} \times \Delta^{n-1}$ induces a homeomorphism $|\mathfrak{A}^\circ(S, X)|/\Gamma_{g,n} \rightarrow \mathcal{M}_{g,n} \times \Delta^{n-1}$.

Remarkable applications of this cellularization include the following results, which are reported on by Mondello in Chapter 5 of this volume:

- Harer used this cellularization to compute the virtual cohomological dimension of the mapping class group.

- Harer & Zagier and (independently) Penner used this cellularization to compute the orbifold Euler characteristic of moduli space.⁴

- Kontsevich used the homeomorphism $|\mathfrak{A}^\circ(S, X)|/\Gamma_{g,n} \rightarrow \mathcal{M}_{g,n} \times \Delta^{n-1}$ in his proof of Witten’s conjecture. Roughly speaking, the conjecture states that a certain formal power series whose coefficients are the intersection numbers of certain tautological classes on moduli space satisfies the classical KdV hierarchy of equations, that is, the generating series is a zero of certain differential operators that generate a truncated Virasoro algebra that appears in string theory.⁵

- Using the homeomorphism $|\mathfrak{A}^\circ(S, X)|/\Gamma_{g,n} \rightarrow \mathcal{M}_{g,n} \times \Delta^{n-1}$, Kontsevich, Penner and Arbarello & Cornalba studied a sequence of combinatorially defined cycles in moduli space. These cycles, called *Witten cycles*, are obtained by taking the cells that correspond to ribbon graphs with vertices of specified valencies. For instance, maximal cells correspond to trivalent ribbon graphs. Using Poincaré duality, Witten cycles define cohomology classes in $H^{2*}(\mathcal{M}_{g,n}; \mathbb{Q})$. Kontsevich and Penner (in different works) defined orientations on the Witten subcomplexes, Kontsevich used matrix integral techniques to express the volumes of these cycles, and Arbarello & Cornalba exploited Kontsevich’s techniques to analyze the integrals of the tautological classes over the combinatorial cycles.

- Chapter 5 also contains a sketch of a proof, obtained by Mondello and Igusa independently, of the Witten–Kontsevich conjecture (sharpened later by Arbarello & Cornalba) stating that the Witten cycles are Poincaré duals to some tautological classes defined in an algebro-geometric way on moduli space.

- Mondello introduced *generalized Witten cycles*, obtained by allowing some zero weights on the ribbon graphs that define the Witten cycles. He proved that generalized Witten cycles and tautological classes generate the same subring of $H^*(\mathcal{M}_{g,n}; \mathbb{Q})$. (This result was also obtained by Igusa.) Mondello also showed that there are explicit formulae that express Witten classes as polynomials in the tautological classes and

⁴The enumeration methods of ribbon graphs used in their works were first developed by theoretical physicists, using asymptotic expansions of Gaussian integrals over spaces of matrices.

⁵A new approach to Witten’s conjecture, which is closer in spirit to the hyperbolic geometry of surfaces, has been recently developed by Maryam Mirzakhani.

vice-versa. Mondello’s proof of the Witten–Kontsevich conjecture, claiming that these cycles are polynomials in the tautological classes, provides a recursive way to find these polynomials.

The chapter also contains a discussion about the Weil–Petersson form and how the spine construction for hyperbolic surfaces with geodesic boundary interpolates between the two cellularizations of $\mathcal{T}_{g,n} \times \Delta^{n-1}$.

Finally, Mondello recalls Harer’s result on the stability of the cohomology groups $H^k(\mathcal{M}_{g,n})$ for $g > 3k$ and fixed n . Without using the result by Igusa/Mondello stated above, he exhibits a direct proof of the fact that the Witten cycles are stable. It is not clear whether similar arguments can be used for A_∞ -classes, that is, cohomology classes of $\mathcal{M}_{g,n}$ related to certain A_∞ -algebras (Witten classes correspond to certain 1-dimensional algebras) first defined by Kontsevich, and whether these classes are tautological.

1.6 Differential forms

In Chapter 6, Nariya Kawazumi considers the problem of constructing “canonical” forms representing cohomology classes on moduli space. The theory is illustrated by several interesting examples, and the chapter provides an overview of various constructions of canonical two-forms.

To explain what this theory is about, Kawazumi recalls the following classical situation. Harer’s result, saying that the second homology group of the moduli space \mathcal{M}_g of a closed orientable surface S of genus $g \geq 3$ is of rank one, implies that there exists a de Rham cohomology class which is unique up to a constant. Kawazumi’s question in that case is to find a “canonical” two-form that represents such a class. It turns out there are several such “canonical” two-forms. One non-trivial 2-cocycle for \mathcal{M}_g is the Meyer cocycle. This cocycle is related to the signature of the total space of a family of compact Riemann surfaces.

The Morita–Mumford classes are other interesting related cohomology classes. We recall that for $n \geq 1$, the n -th Morita–Mumford class e_n (also called tautological class) is an element of the cohomology group $H^{2n}(\mathcal{M}_g)$. These classes play a prominent role in the stable cohomology of the mapping class group. In 2002, I. Madsen and M. Weiss proved a conjecture that was made by Mumford, stating that the rational stable cohomology algebra of the mapping class group is generated by the Morita–Mumford classes. Kawazumi with co-authors, in a series of papers, made a deep study of the Morita–Mumford classes and their generalizations. Wolpert showed that the Weil–Petersson Kähler form ω_{WP} represents the first Morita–Mumford class e_1 . This form is an example of a “canonical” representative of e_1 .

The ideas developed in Chapter 6 of this Handbook use the period map from Teichmüller space to the Siegel upper half-space. We recall that the Siegel upper half-space of genus $g \geq 2$, denoted by \mathfrak{H}_g , is the set of symmetric square $g \times g$ matrices with complex coefficients whose imaginary part is positive definite. The space \mathfrak{H}_g plays an important role in number theory, being the domain of some automorphic forms (Siegel

modular forms). The period map $\text{Jac}: \mathcal{T}_g \rightarrow \mathfrak{H}_g$ is a canonical map from Teichmüller space into the Siegel upper half-space, and the first Morita–Mumford class e_1 is the pull-back of a canonical two-form on \mathfrak{H}_g by the period map. More-generally, the odd Morita–Mumford classes are represented by pull-backs of $\text{Sp}(2g, \mathbb{R})$ -invariant differential forms on \mathfrak{H}_g , arising from Chern classes of holomorphic vector bundles. But the even ones are not. Kawazumi describes a higher analogue of the period map which he calls the *harmonic Magnus expansion*, which produces other canonical differential forms on moduli space representing the Morita–Mumford classes e_n . Some of the forms that are obtained in this way are related to Arakelov geometry.

2 The group theory

The group theory that is reported on in Part B of this volume concerns primarily the mapping class group of a surface. This group is studied from the point of view of quasi-homomorphisms, of measure-equivalence, and in relation to Lefschetz fibrations. Other related groups are also studied, namely, braid groups, Artin groups, and affine groups of singular flat surfaces. The study of singular flat surfaces is a subject of investigation which is part of Teichmüller theory, with ramifications in several areas in mathematics, such as dynamical systems theory, and in physics. Of particular interest in dynamical systems theory is the so-called *Teichmüller geodesic flow*, defined on the moduli space of flat surfaces.

2.1 Quasi-homomorphisms

Chapter 7, by Koji Fujiwara, concerns the theory of quasi-homomorphisms on mapping class groups. We recall that a quasi-homomorphism on a group G is a map $f: G \rightarrow \mathbb{R}$ satisfying

$$\sup_{x, y \in G} |f(xy) - f(x) - f(y)| < \infty.$$

Quasi-homomorphisms on a given group form a vector space. Examples of quasi-homomorphisms are homomorphisms and bounded maps. These two classes form vector subspaces of the vector space of quasi-homomorphisms, and their intersection is reduced to the zero element.

An example of a quasi-homomorphism on $G = \mathbb{R}$ is the integral part function, which assigns to a real number x the smallest integer $\leq x$.

The study of quasi-homomorphisms in relation with mapping class groups was initiated in joint work by Endo & Kotschick.⁶

In Chapter 7, quasi-homomorphisms on mapping class groups are studied in parallel with quasi-homomorphisms on Gromov hyperbolic groups. Although mapping

⁶We note however that the case of $\text{PSL}(2, \mathbb{Z})$, which is the mapping class group of the torus, had already been studied by several authors.

class groups are not word-hyperbolic, since they contain subgroups isomorphic to \mathbb{Z}^2 (except in some elementary cases), it is always good to find analogies between the two categories of groups. There is a well-known situation in which mapping class groups behave like generalized hyperbolic groups. This is through the action of mapping class groups on curve complexes which, by a result of Masur and Minsky, are Gromov hyperbolic. This action is co-compact but of course not properly discontinuous. Occasionally in this chapter, parallels are also made with quasi-homomorphisms on lattices in Lie groups. In the setting studied here, the techniques of proofs of corresponding results for mapping class group, hyperbolic groups and lattices present many similarities.

Using Fujiwara's notation, we let $\widetilde{\text{QH}}(G)$ be the quotient space of the vector space of quasi-homomorphisms $G \rightarrow \mathbb{R}$ by the subspace generated by bounded maps and by homomorphisms. The space $\widetilde{\text{QH}}(G)$ carries a Banach space structure. One of the primary objects of the theory is to compute the vector space $\widetilde{\text{QH}}(G)$ for a given group G , and, first of all, to find conditions under which $\widetilde{\text{QH}}(G)$ is nonempty. It turns out that the computation of the group $\widetilde{\text{QH}}(G)$ uses the theory of bounded cohomology. Indeed the group $\widetilde{\text{QH}}(G)$ is the kernel of the homomorphism $H_b^2(G; \mathbb{R}) \rightarrow H^2(G; \mathbb{R})$, where $H_b^2(G; \mathbb{R})$ is the second bounded cohomology group of G .

In many known cases, $\widetilde{\text{QH}}(G)$ is either zero- or infinite-dimensional. One of the first interesting examples of the latter occurrence is due to R. Brooks, who proved in the late 1970s that in the case where G is a free group of rank ≥ 2 , $\widetilde{\text{QH}}(G)$ is infinite-dimensional.

The vector space $\widetilde{\text{QH}}(G)$ is an interesting object associated to a hyperbolic group despite the fact that it is not a quasi-isometry invariant. Epstein & Fujiwara proved in 1997 that if G is any non-elementary word hyperbolic group, then $\widetilde{\text{QH}}(G)$ is infinite-dimensional. Since free groups of rank ≥ 2 are hyperbolic, this result generalizes Brooks' result mentioned above. In 2002, Bestvina & Fujiwara extended the result of Epstein & Fujiwara to groups acting isometrically on δ -hyperbolic spaces (with no assumption that the action is properly discontinuous). Using the action of mapping class groups on curve complexes, Bestvina & Fujiwara proved that if G is any subgroup of the mapping class group of a compact orientable surface which is not virtually abelian, then $\widetilde{\text{QH}}(G)$ is infinite-dimensional.

Chapter 7 contains a review of these results as well as a short introduction to the theory of bounded cohomology for discrete groups. The author also surveys some recent results by Bestvina & Fujiwara on the group $\widetilde{\text{QH}}(G)$ in the case where G is the fundamental group of a complete Riemannian manifold of non-positive sectional curvature. He describes some rank-one properties of mapping class groups related to quasi-homomorphisms, to some superrigidity phenomena and to the bounded generation property. We recall that a group G is said to be *boundedly generated* if there exists a finite subset $\{g_1, \dots, g_k\}$ of G such that every element of this group can be written as $g_1^{n_1} \dots g_k^{n_k}$ with n_1, \dots, n_k in \mathbb{Z} . Bounded generation is related to the existence of quasi-homomorphisms. Mapping class groups are not boundedly generated (Farb–

Lubotzky–Minsky). Non-elementary subgroups of word-hyperbolic groups are not boundedly generated (Fujiwara). A discrete subgroup of a rank-1 simple Lie group that does not contain a nilpotent subgroup of finite index is not boundedly generated (Fujiwara).

Chapter 7 also contains a survey of the theory of separation by quasi-homomorphisms in groups, with applications to mapping class groups, to hyperbolic groups and to lattices. One of the motivating results in this direction is a result by Polterovich & Rudnick (2001) saying that if two elements in $\mathrm{SL}(2, \mathbb{Z})$ are not conjugate to their inverses, then they can be separated by quasi-homomorphisms. Recent results on this subject, by Endo & Kotschick for mapping class groups and by Calegari & Fujiwara for hyperbolic groups, are presented in this chapter.

2.2 Lefschetz fibrations

Chapter 8, by Mustafa Korkmaz and András Stiepicz, concerns the theory of Lefschetz pencils and Lefschetz fibrations, a theory which is at the intersection of 4-manifold theory, algebraic geometry and symplectic topology. Mapping class groups of surfaces play an essential role in this theory, and it is for this reason that such a chapter is included in this Handbook.

Lefschetz fibrations are 4-dimensional manifolds that are simple enough to handle, but with a rich enough structure to make them interesting. One may consider a Lefschetz fibration as a natural generalization of a 4-manifold which is a surface fibration, a surface fibration being itself a generalization of a Cartesian product of two surfaces. Lefschetz pencils are slightly more general than Lefschetz fibrations; a Lefschetz pencil gives rise to a Lefschetz fibration by a “blowing-up” operation.

Lefschetz fibrations and Lefschetz pencils first appeared in algebraic geometry in the early years of the twentieth century, when Solomon Lefschetz studied such structures on complex algebraic surfaces, that is, 4-dimensional manifolds defined as zeroes of a homogeneous polynomial systems with complex coefficients. Lefschetz constructed a Lefschetz pencil structure on every algebraic surface.

Towards the end of the 1990s, Lefschetz fibrations and Lefschetz pencils played an important role in the work of Simon Donaldson, who showed that any symplectic 4-manifold has a Lefschetz pencil structure with base the two-sphere. Robert Gompf showed that conversely, any 4-manifold admitting a Lefschetz pencil structure carries a symplectic structure.⁷ In this way, Lefschetz pencils play the role of a topological analogue of symplectic 4-manifolds.

Let us say things more precisely. A *Lefschetz fibration* is a compact oriented 4-dimensional manifold X equipped with a projection $\pi : X \rightarrow S$, where S is a closed oriented surface, and where π is a fibration if we restrict it to the inverse image of some finite set of points in S , called the critical values. Furthermore, it is required that

⁷Gompf’s proof is an extension to the class of Lefschetz pencils of Thurston’s proof of the fact that any oriented surface bundle over a surface carries a symplectic structure, provided that the homology class of the fiber is nontrivial in the second homology group of the 4-manifold.

above a critical value, the local topological model of π is the map $(z_1, z_2) \mapsto z_1^2 + z_2^2$ from \mathbb{C}^2 to \mathbb{C} , in the neighborhood of the origin. (In this picture, the critical value is the origin.) The fibers of π above the critical values are singular surfaces, and a singular point on such a surface is called a *nodal point*. A nonsingular fiber is a closed orientable surface called a *generic fiber*. The *genus* of a Lefschetz fibration is, by definition, the genus of a regular fiber. (Recall that restricted to the complement of the critical values, a Lefschetz fibration is a genuine fibration, and therefore all the generic fibers are homeomorphic.) In some sense, a nodal point is a singularity of the simplest type in the dimension considered; it is the singularity that appears at a generic intersection of two surfaces. Such a singularity naturally appears in complex analysis. In a Lefschetz fibration, a singular fiber is obtained from a nearby fiber by collapsing to a point a simple closed curve, called a *vanishing cycle*. The vanishing cycle, when it is collapsed, becomes the nodal point of the corresponding singular fiber.

A natural way of studying the topology of a Lefschetz fibration $\pi : X \rightarrow S$ is to try to figure out how the fibers $\pi^{-1}(s)$ are glued together in X when the point s moves on the surface S , and in particular, near the critical values, since the complication comes from there. This leads to a combinatorial problem which in general is non-trivial, and the mapping class group of a generic fiber is an essential ingredient in this story. It is here that the study of Lefschetz fibrations gives rise to interesting problems on mapping class groups. For instance, Lefschetz fibrations were the motivation of recent work by Endo & Kotschick and by Korkmaz on commutator lengths of elements in mapping class groups. Lefschetz fibrations also motivated the study of questions related to “factorizations of the identity element” of a mapping class group, that is, an expansion of this identity as a product of positive Dehn twists.

I would like to say a few words on monodromies and on factorizations, and this needs some notation.

Let $P \subset S$ be the set of critical values of a Lefschetz fibration $\pi : X \rightarrow S$. We choose a basepoint s_0 for the surface S , in the complement of the set P . The fiber $\pi^{-1}(s_0)$ is then called the *base fiber* and we identify it with an abstract surface F . There is a natural homomorphism ψ , called the *monodromy representation* from $(\pi_1(S - P), s_0)$ to the mapping class group of F . This homomorphism is the main algebraic object that captures the combinatorics of the Lefschetz fibration. It is defined by considering, for each loop $\gamma : [0, 1] \rightarrow S$ based at s_0 , the fibration induced on the interval $[0, 1]$ (which is a trivial fibration), and then taking the isotopy class of the surface homeomorphism that corresponds to the gluing between the fibers of π above the points $\gamma(0)$ and $\gamma(1)$. The resulting *monodromy representation* is a homomorphism ψ from $(\pi_1(S - P), s_0)$ to the mapping class group of F , and it is well defined up to conjugacy. Two Lefschetz fibrations are isomorphic if and only if they have the same monodromy representation (up to an isomorphism between the images induced by inner automorphisms of the mapping class groups of the fibers, and up to isomorphisms of the fundamental groups of the bases of the fibrations). The detailed construction of the monodromy representation is recalled in Chapter 8 of this volume.

The monodromy representation homomorphism in this theory can be compared to the monodromy which appears in the study of holomorphic families of Riemann surfaces, as it is presented in Chapter 3 of this volume.

Now a few words about factorizations. The monodromy around a critical value is the class of the positive Dehn twist along the vanishing cycle on a regular fiber near the singular fiber. Modulo some standard choices and identifications, the monodromy associated to a loop that surrounds exactly one time each critical value produces an element of the mapping class group of the base fiber, which is equal to the identity word decomposed as a product of positive Dehn twists. Conversely, one can construct a Lefschetz fibration of genus g from each factorization of the identity element of the mapping class group of an oriented closed surface of genus g . There is an action of the braid group on the set of such factorizations, and the induced equivalence relation is called *Hurwitz equivalence*. The notion of factorization in this setting leads to a discussion of commutator length and of torsion length in the mapping class group. More precisely, it leads to the question of the minimal number of factors needed to express an element of the mapping class group as a product of commutators and of torsion elements respectively.

This chapter by Korkmaz and Stiepicz gives a quick overview on Lefschetz fibrations, with their relation to the works of Gompf and Donaldson on symplectic topology, and to the works of Endo & Kotschick and of Korkmaz on commutator lengths of Dehn twists in mapping class groups. The authors also mention generalizations of Lefschetz fibrations involving Stein manifolds and contact structures. They propose a list of open problems on the subject.

2.3 Measure-equivalence

Chapter 9, by Yoshikata Kida, considers mapping class groups in analogy with lattices, that is, discrete subgroups of cofinite volume of Lie groups, in the special setting of group actions on measure spaces.

Lattice examples are appealing for people studying mapping class groups, because it is a natural question to search for properties of mapping class groups that are shared by lattices, and for properties of mapping class groups that distinguish them from lattices. We already mentioned these facts in connection with Fujiwara's work in Chapter 7, and we recall in this respect that $\mathrm{PSL}(2, \mathbb{Z})$, which is the mapping class group of the torus, is a lattice in $\mathrm{PSL}(2, \mathbb{R})$.

At the same time, Chapter 9 gives a review of measure-equivalence theory applied to the study of mapping class groups.

Let us first recall a few definitions. Two discrete groups Γ and Λ are said to be *measure-equivalent* if there exists a standard Borel space (Σ, m) (that is, a Borel space equipped with a σ -finite positive measure which is isomorphic to a Borel subset of the unit interval) equipped with a measure-preserving action of the direct product $\Gamma \times \Lambda$, such that the actions of Γ and Λ obtained by restricting the $\Gamma \times \Lambda$ -action to $\Gamma \times \{e\}$ and $\{e\} \times \Lambda$ satisfy the following two properties:

- these actions are essentially free, that is, stabilizers of almost all points are trivial;
- these actions have finite-measure fundamental domains.

Measure-equivalence is an equivalence relation on the class of discrete groups. It was introduced by Gromov in his paper *Asymptotic invariants*, as a measure-theoretic analogue of quasi-isometry, the latter being defined on the class of finitely generated groups. Gromov raised the question of classifying discrete groups up to measure-equivalence.

From the definitions, it follows easily that isomorphic groups modulo finite kernels and co-kernels are measure-equivalent. In particular, any two finite groups are measure-equivalent. A group that is measure-equivalent to a finite group is finite. In any locally compact second countable Lie group, two lattices are measure-equivalent.

Two discrete groups Γ and Λ acting on two standard measure spaces (X, μ) and (Y, ν) are said to be *orbit-equivalent* if there exists a measure-preserving isomorphism $f: (X, \mu) \rightarrow (Y, \nu)$ such that $f(\Gamma x) = \Lambda f(x)$ for almost every x in X . Orbit-equivalence is an equivalence relation which is weaker than conjugacy, and it is intimately related to measure-equivalence. The study of orbit-equivalence was started a few decades ago by D. S. Ornstein and B. Weiss. These authors showed that an infinite discrete group is measure-equivalent to \mathbb{Z} if and only if it is amenable. Their result was stated in terms of orbit-equivalence. Orbit-equivalence is also related to the study of von Neumann algebras, and it was studied as such by S. Popa.

In a series of recent papers, Y. Kida made a detailed study of measure-equivalence in relation to mapping class groups. In particular, he obtained the following results, reported on in Chapter 9 of this volume.

Let $S = S_{g,p}$ be a compact surface of genus g with p boundary components satisfying $3g - 4 + p > 0$ and let $C(S)$ be the curve complex of S . If a discrete group Λ is measure-equivalent to the mapping class group of S , then there exists a homomorphism $\rho: \Lambda \rightarrow \text{Aut}(C(S))$ whose kernel and cokernel are both finite. Using the famous result by Ivanov (completed by Korkmaz and Luo) stating that (with a small number of exceptional surfaces) the automorphism group of the curve complex of a surface is the extended mapping class group of that surface, Kida's result gives a characterization of discrete groups that are measure-equivalent to mapping class groups. This result is an analogue of a result by A. Furman which gives a characterization of discrete groups that are measure-equivalent to higher rank lattices.

Kida also studied the relation of measure-equivalence between surface mapping class groups, proving that if two pairs of nonnegative integers (p, g) and (p', g') satisfy $3g - 4 + p \geq 0$ and $3g' - 4 + p' \geq 0$, and if the mapping class groups $\Gamma(S_{g,p})$ and $\Gamma(S_{g',p'})$ are measure-equivalent, then either the surfaces $S_{g,p}$ and $S_{g',p'}$ are homeomorphic or $\{(g, p), (g', p')\}$ is equal to $\{(0, 5), (1, 2)\}$ or to $\{(0, 6), (2, 0)\}$. He also settled the question of the classification of subgroups of mapping class groups from the viewpoint of measure-equivalence. An analogous result was known for lattices in the Lie groups $\text{SL}(n, R)$ and $\text{SO}(n, 1)$.

Kida showed that there exist no interesting embedding of the mapping class group as a lattice in a locally compact second countable group. V. Kaimanovich and H. Masur

had already proved that under the condition $3g - 4 + p \geq 0$, any sufficiently large subgroup of the mapping class group of $S_{g,p}$ (and in particular, the mapping class group itself) is not isomorphic to a lattice in a semisimple Lie group with real rank at least two.

Inspired by a definition made by R. Zimmer in the setting of lattices, Kida defined a notion of measure-amenability for actions on the curve complex of a surface. He proved the following: Let $S = S_{g,p}$ be a surface satisfying $3g - 4 + p \geq 0$, let $C(S)$ be the curve complex of S , let $\partial C(S)$ be its Gromov boundary and let μ be a probability measure on $\partial C(S)$ such that the action of the extended mapping class group of S on that measure space is non-singular. Then this action is measure-amenable.

Chapter 9 also contains interesting measure-theoretic descriptions of mapping class group actions, e.g., a classification of infinite subgroups of the mapping class group in terms of the fixed points of their actions on the space of probability measures on Thurston's space of projective measured foliations.

It is interesting to see that Y. Kida succeeded in replacing by measure-theoretic arguments the topological arguments that were used by various authors in the proofs of their rigidity results on mapping class group actions on several spaces (e.g., the actions on the curve complex and on other complexes, the actions on spaces of foliations, algebraic actions of the extended mapping class group on itself by conjugation, and so on). To give an example that highlights the analogy, we recall a result by N. Ivanov stating that, with the exception of some special surfaces, any isomorphism $\phi: \Gamma_1 \rightarrow \Gamma_2$ between finite index subgroups Γ_1 and Γ_2 of the extended mapping class group is a conjugation by an element of the extended mapping class group, and in particular, any automorphism of the extended mapping class group is an inner automorphism. An important step in Ivanov's proof of this result is the proof that any automorphism between Γ_1 and Γ_2 sends a sufficiently high power of a Dehn twist to a power of a Dehn twist. From this, and since Dehn twists are associated to homotopy classes of simple closed curves which are vertices of the curve complex, Ivanov obtains an automorphism of the curve complex induced by the isomorphism ϕ . He then appeals to the fact that the automorphism group of the curve complex is the natural image in that group of the extended mapping class group. To prove that ϕ sends powers of Dehn twists to powers of Dehn twists, Ivanov uses an algebraic characterization of Dehn twists. Moreover, he proves that ϕ preserves some geometric relations between Dehn twists; for instance, it sends pairs of commuting Dehn twists to pairs of commuting Dehn twists. Now the measure-theoretic setting. Kida's rigidity result is formulated in the general setting of isomorphisms of discrete measured groupoids. To say it in few words, Kida needs to show that any isomorphism of discrete measured groupoids arising from measure-preserving actions of the mapping class group preserves subgroupoids generated by Dehn twists. The proof of this fact uses a characterization of such groupoids in terms of discrete measured groupoid invariants. This is done by using the measure-amenability of non-singular actions of the extended mapping class group on the boundary of the curve complex mentioned above, and a subtle characterization of subgroupoids generated by Dehn twists in terms of measure-amenability.

More precisely, a subgroupoid generated by a Dehn twist is characterized by the fact that it is an amenable normal subgroupoid of infinite type of some maximal reducible subgroupoid. Kida concludes using the fact that measure-amenability is an invariant of isomorphism between groupoid actions.

Kida also obtained a measurable rigidity result for direct products of mapping class groups, using a technique introduced by N. Monod and Y. Shalom in a study they made of measurable rigidity of direct products of discrete groups.

Recently, D. Gaboriau showed that the sequence of ℓ^2 -Betti numbers introduced by Cheeger and Gromov is invariant under measure-equivalence, up to a multiplicative constant. Using this and results of McMullen and of Gromov, Kida gave formulae for these Betti numbers.

2.4 Affine groups

In Chapter 10, a *flat surface* is defined as a pair (S, ω) consisting of a closed Riemann surface S equipped with a nonzero holomorphic one-form ω (which we shall also call here an abelian differential). Such a surface S is naturally equipped with a flat (i.e. Euclidean) structure in the complement of the zeroes of ω . The flat structure is defined, using the holomorphic local coordinates, by parameters of the form $\phi(z) = \int_{z_0}^z \omega$, after a choice of a basepoint z_0 in the holomorphic chart. In fact, the surface S is equipped, in the complement of the zeroes of ω , with an atlas whose transition functions are better than Euclidean transformations of the plane, since they are translations. For this reason, a flat surfaces in the sense used here is also called a “translation surface”. The flat metric in the complement of the zeroes of ω extends at any zero point of order n to a singular flat metric whose singularity at such a point is locally a Euclidean cone point with total angle $2\pi(n + 1)$. We note that there are other ways of defining flat surfaces that do not use the word “holomorphic”. For instance, a flat surface can be obtained by gluing rational-angled Euclidean polygons along their boundaries by Euclidean translations.

There is a strong relation between flat surfaces and billiards. In 1975, Zelmyakov & Katok associated to each rational-angled polygon a uniquely defined flat surface, such that the billiard flow of the polygon is equivalent to the geodesic flow of the flat surface.

There is a natural action of the group $\mathrm{SL}(2, \mathbb{R})$ on the space of flat surfaces, and this action preserves the space \mathcal{A} of unit norm abelian differentials (the norm of a flat surface (S, ω) being defined by $(\int_S |\omega|^2)^{1/2}$). We also recall that the *Teichmüller geodesic flow* is the action of the diagonal subgroup of $\mathrm{SL}(2, \mathbb{R})$ on the space \mathcal{A} .

Flat surfaces appear in many ways in Teichmüller theory. One obvious reason is that a flat surface has an underlying Riemann surface structure, and it is therefore natural to study parametrizations of Teichmüller space by flat surfaces. Flat surfaces also arise from holomorphic quadratic differentials. We recall that a holomorphic quadratic differential being locally the square of a holomorphic one-form, also gives rise to a singular Euclidean metric on its underlying Riemann surface. Holomorphic

quadratic differentials play a prominent role in Teichmüller theory since the work of Teichmüller himself, in particular because there is a natural identification between the vector space of quadratic differentials and the cotangent space to Teichmüller space at each point.

To a flat surface (S, ω) is associated a subgroup of $SL(2, \mathbb{R})$ called its *affine group*, and denoted by $SL(S, \omega)$. To define this group, one first considers the group $\text{Aff}^+(S, \omega)$ of orientation-preserving diffeomorphisms of S that act affinely in the Euclidean charts associated to ω , in the complement of the zeroes of ω . (Such a diffeomorphism is allowed to permute the zeroes.) An affine map, in a chart, has a matrix form $X \mapsto AX + B$, with A being a constant nonsingular matrix which can be considered as the derivative of the affine map. Since the coordinate changes of the Euclidean atlas associated to a flat surface are translations, the matrix A is independent of the choice of the chart, and thus is canonically associated to the affine map. Composing two affine diffeomorphisms of S gives rise to matrix multiplication at the level of the linear parts. This gives a homomorphism $D: \text{Aff}^+(S, \omega) \rightarrow GL(2, \mathbb{R})$ which associates to each affine diffeomorphism its derivative. The image of D lies in the subgroup $SL(2, \mathbb{R})$ of $GL(2, \mathbb{R})$, as a consequence of the fact that the surface has finite area. The image of the diffeomorphism D in $SL(2, \mathbb{R})$ is, by definition, the *affine group* $SL(S, \omega)$ of the flat surface (S, ω) . W. Veech observed that the affine group $SL(S, \omega)$ is always a discrete subgroup of $SL(2, \mathbb{R})$. The affine group $SL(S, \omega)$ is sometimes called the *Veech group* of the flat surface.

There is a nice description of Thurston's classification of isotopy classes of affine diffeomorphism. An affine homeomorphism $f: S \rightarrow S$ is parabolic, elliptic or hyperbolic if $|\text{Tr}(Df)| = 2, < 2, \text{ or } > 2$ respectively. The hyperbolic affine homeomorphisms are the pseudo-Anosov affine diffeomorphisms. Beyond their use in this classification, we shall see below that the set of traces of affine homeomorphisms of a flat surface play a special role in this theory.

The notion of an affine group of a flat surface first appeared in Thurston's construction of a family of pseudo-Anosov homeomorphisms of a surface which are affine with respect to some flat structure. Indeed, in his paper *On the geometry and dynamics of homeomorphisms of surfaces*, Thurston constructed such a family, the flat structure being obtained by "thickening" a filling pair of transverse systems of simple closed curves on the surface.

In Chapter 10 of this Handbook, Martin Möller addresses the following natural problems:

- Which subgroups of $SL(2, \mathbb{R})$ arise as affine groups of flat surfaces?
- What does the affine group of a generic flat surface look like?

Several partial results on these problems have been obtained by various authors. For instance, Veech constructed flat surfaces whose affine groups are non-arithmetic lattices. Special types of flat surfaces, called origamis, or square-tiled surfaces, arise naturally in these kinds of questions. These surfaces are obtained by gluing Euclidean squares along their boundaries using Euclidean translations. E. Gutkin & C. Judge showed that the affine group of an origami is a subgroup of finite index in $SL(2, \mathbb{Z})$.

P. Hubert & S. Lelièvre showed that in any genus $g \geq 2$ there are origamis whose affine groups are non-congruence subgroups of $SL(2, \mathbb{R})$. We note that origamis were already considered in Volume I of this Handbook, namely in Chapter 6 by Herrlich and Schmithüsen, where these surfaces are studied in connection with Teichmüller disks in moduli space. They are also thoroughly studied in relation with the theory of dessins d'enfants in Chapter 18 of the present volume. Schmithüsen proved that all congruence subgroups of $SL(2, \mathbb{Z})$ with possibly five exceptions occur as affine groups of origamis. Möller, in Chapter 10 of this volume, asks the question of whether there is a subgroup of $SL(2, \mathbb{Z})$ that is not the affine group of an origami.

Another interesting class of flat surfaces is the class of *Veech surfaces*. These are the flat surfaces whose affine groups are lattices in $SL(2, \mathbb{R})$. A recent result of I. Bouw and M. Müller says that all triangle group (m, n, ∞) with $1/m + 1/n < 1$ and $m, n \leq \infty$ occur as affine groups of Veech surfaces.

C. McMullen, and then P. Hubert & T. Schmidt produced flat surfaces whose affine groups are infinitely generated.

Möller proved that provided the genus of S is ≥ 2 , the affine group of a generic flat surface (S, ω) is either $\mathbb{Z}/2$ or trivial, and that this depends on whether (S, ω) is in a hyperelliptic component or not, with respect to the natural stratification of the total space of the vector bundle of holomorphic one-forms minus the zero-section. (A hyperelliptic component is a component of a stratum that consists exclusively of hyperelliptic curves.) He also proved that in every stratum there exist flat surfaces whose affine groups are cyclic groups generated by parabolic elements. He raises the question of whether there exists a flat surface whose affine group is cyclic generated by a hyperbolic element.

Müller also discusses the relation between affine groups and closures of $SL(2, \mathbb{R})$ -orbits of the corresponding flat surfaces in moduli space.

Given an arbitrary subgroup Γ of $SL(2, \mathbb{R})$, one can define its *trace field* as the subfield K of \mathbb{R} generated by the set $\{\text{Tr}(A) : A \in \Gamma\}$. Thus, associated to a flat surface (S, ω) is the trace field of its affine group $SL(S, \omega)$. It turns out that the trace field of the affine group of a flat surface is an interesting object of study. R. Kenyon & J. Smillie proved that the trace field of the affine group $SL(S, \omega)$ has at most degree g over \mathbb{Q} . P. Hubert & E. Laneeau showed that if (S, ω) is given by Thurston's construction, then the trace field of $SL(S, \omega)$ is totally real. They also showed that there exist flat surfaces supporting pseudo-Anosov diffeomorphisms whose trace fields are not totally real. C. McMullen showed that all real quadratic fields arise as trace fields of lattice affine groups.

2.5 Braid groups

Chapter 11 by Luis Paris is a survey on braid groups and on some of their generalizations, and on the relations between these groups and mapping class groups.

Braid groups are related to mapping class groups in several ways. A well-known instance of such a relation is that the braid group on n strands is isomorphic to the

mapping class group of the surface $S_{0,n}$, that is, the disk with n punctures. In fact, this isomorphism can be considered as a first step for a general theory of representations of braid groups in mapping class groups, which is one of the main subjects reported on in Chapter 11.

Although braiding techniques have certainly been known since the dawn of humanity (hair braiding, rope braiding, etc.), braid groups as mathematical objects were formally introduced in 1925, by Emil Artin, and questions about representations of braid groups immediately showed up. One of the first important results in this representation theory is due to Artin himself, who proved that the braid group on n strands admits a faithful representation (now called the *Artin representation*) in the automorphism group of the free group on n generators.⁸ Artin's result can be seen as an analogue of the result by Dehn, Nielsen and Baer stating that the extended mapping class group of a closed surface of genus ≥ 1 admits a faithful representation in the automorphism group of the fundamental group of that surface (and in that case, the representation is an isomorphism). From Artin's result one deduces immediately that braid groups are residually finite and Hopfian. (Recall that a group is said to be Hopfian if it is not isomorphic to any of its subgroups.)

Historically, results on braid groups were obtained in general before the corresponding results on mapping class groups. This is due to the fact that braid groups have very simple presentations, with nothing comparable in the case of mapping class groups. Another possible reason is that homeomorphisms of the punctured disk are much easier to visualize compared to homeomorphisms of arbitrary surfaces, and therefore, it is in principle easier to have a geometric intuition on braid groups than on general surface mapping class groups. It is also safe to say that results on braid groups have inspired research on mapping class groups. Indeed, several results on mapping class groups were conjectured in analogy with results that were already obtained for braid groups. Let us mention a few examples:

- Presentations of braid groups have been known since the introduction of these groups. (In fact, right at the beginning, braid groups were defined by generators and relators.) But in the case of the mapping class groups, it took several decades after the question was addressed, to find explicit presentations.
- Automorphism groups of braid groups were computed long before analogous results were obtained for mapping class groups.
- Several algorithmic problems (conjugacy and word problems, etc.) were solved for braid groups before results of the same type were obtained for mapping class groups.
- The existence of a faithful linear representation for braid groups has been obtained in the year 2000 (by Bigelow and Krammer, independently), settling a question that had been open for many years. The corresponding question for mapping class groups is still one of the main open questions in the field.

⁸B. Perron and J. P. Vannier recently obtained results on the representation of a braid group on n strands in the automorphism group of the free group on $n - 1$ generators.

In Chapter 11 of this volume, the theory of braids is included in a very wide setting that encompasses mapping class groups, but also other combinatorially defined finitely presented groups, namely Garside groups, Artin groups and Coxeter groups. To make things more precise, we take a finite set S of cardinality n and we recall that a *Coxeter matrix* over S is an $n \times n$ matrix whose coefficients m_{st} ($s, t \in S$) belong to the set $\{1, 2, \dots, \infty\}$, with $m_{st} = 1$ if and only if $s = t$. The *Coxeter graph* Γ associated to a Coxeter matrix $M = m_{s,t}$ is a labeled graph whose vertex set is S and where two distinct vertices s and t are joined by an edge whenever $m_{s,t} \geq 3$. If $m_{st} \geq 4$, then the edge is labeled by $m_{s,t}$. Coxeter graphs are also called Dynkin diagrams. The *Coxeter group of type* Γ is the finitely presented group with generating set S and relations $s^2 = 1$ for s in S , and $(st)^{m_{st}} = 1$ for $s \neq t$ in S . Here, a relation with $m_{st} = \infty$ means that the relation does not exist.

The *Artin group* associated to a Coxeter matrix $M = m_{s,t}$ is a group defined by generators and relations, where the generators are the elements of S , ordered as a sequence $\{a_1, \dots, a_n\}$ and where the relations are defined by the equalities $\langle a_1, a_2 \rangle^{m_{1,2}} = \langle a_2, a_1 \rangle^{m_{2,1}}, \dots, \langle a_{n-1}, a_n \rangle^{m_{n-1,n}} = \langle a_n, a_{n-1} \rangle^{m_{n,n-1}}$ for all $m_{i,j} \in \{2, 3, \dots, \infty\}$, where $\langle a_i, a_j \rangle$ denotes the alternating product of a_i and a_j taken $m_{i,j}$ times, starting with a_i . (For example, $\langle a_1, a_2 \rangle^5 = a_1 a_2 a_1 a_2 a_1$.) Artin groups are also used in other domains of mathematics, for instance in the theory of random walks.

Coxeter groups were introduced by J. Tits in relation with his study of Artin groups. Garside groups were introduced by P. Dehornoy and L. Paris, as a generalization of Artin groups. There are several relations between Artin groups, Coxeter groups and Garside groups. One important aspect of Garside groups is that these groups are well-suited to the study of algorithmic problems for braid groups. An Artin group has a quotient Coxeter group.

There is a geometric interpretation of Artin groups which extends the interpretation of braid groups in terms of fundamental groups of hyperplane arrangements in \mathbb{C}^n . It is unknown whether mapping class groups are Artin groups and whether they are Garside groups. Some Artin groups, called Artin groups of spherical type, are Garside groups, and it is known that Artin groups of spherical type are generalizations of braid groups.

Chapter 11 contains algebraic results, algorithmic results, and results on the representation theory of these classes of groups.

From an algebraic point of view, Paris gives an account of known results on the cohomology of braid groups and of Artin groups of spherical type. He introduces Salvetti complexes of hyperplane arrangements. These complexes are simplicial complexes that arise naturally in the study of hyperplane arrangements; they have natural geometric realizations, and they have been successfully used as a tool in computing the cohomology of Artin groups.

From the algorithmic point of view, the author reports on Tits' solution of the word problem for Coxeter groups, on Garside's solution of the conjugacy problem for braid groups, and on recent progress made by Dehornoy and Paris on the extension of this result to Garside groups.

Paris also reports on recent progress on linear representations of Artin groups, extending the work by Bigelow and Krammer on linear representations of braid groups and the subsequent work on linear representations of certain Artin groups, which was done by Digne and by Cohen & Wales. The author also presents an algebraic and a topological approach that he recently developed for the question of linear representations.

Besides the study of linear representations, Chapter 11 contains a recent study of *geometric* representations of Artin groups, that is, representations into mapping class groups. (Recall the better-than-faithful representation of the braid group on n strands in the mapping class group of the disk with n punctures.) The chapter contains the description of a nice construction of geometric representations of Artin groups, obtained by sending generators to Dehn twists along some curves that realize the combinatorics of the associated Coxeter graph.

3 Representation spaces and geometric structures

Representation theory makes interesting relations between algebra and geometry. From our point of view, the subject may be described as the study of geometric structures by representing them by matrices and algebraic operations on these matrices.

As already mentioned, the geometric structures considered in Part C of this Handbook are more general than the structures that are dealt with in the classical Teichmüller theory (namely, conformal structures and hyperbolic structures). These general structures include complex projective structures, whose recent study involves techniques that have been introduced by Thurston in the 1990s. We recall that Thurston introduced parameters for (equivalence classes of) complex projective structures on a surface in which the space of measured laminations plays an essential role. In this setting, complex projective structures are obtained by *grafting* Euclidean annuli on hyperbolic surfaces along simple closed curves and, more generally, along measured geodesic laminations. As it is the case for hyperbolic structures, deformations of complex projective structures can be studied either directly on the surface, or within a space of representations of the fundamental group of the surface in an appropriate Lie group. The direct study can be done by considering complex projective structures defined on some “elementary” surfaces with boundary and then gluing together such surfaces so as to obtain complex projective structures on larger surfaces. For instance, one can study complex projective structures on pairs of pants in a way parallel to what is done classically in the study of hyperbolic structures, and then investigate the gluing between pairs of pants. Complex projective structures can also be studied in the context of representations of fundamental groups of surfaces in the Lie group $SL(2, \mathbb{C})$. It is also well known that the space of $SL(2, \mathbb{C})$ -representations (more precisely, the orbit space under the action of $SL(2, \mathbb{C})$ by conjugation) can also be studied for itself, as a generalized Teichmüller space.

In this part of the Handbook, Bill Goldman gives an exposition of what is usually referred to as the *Fricke–Klein trace parameters*, that is, parameters for the representations of Teichmüller spaces in the character variety of $\mathrm{SL}(2, \mathbb{C})$ -representations. He treats the cases of surfaces with two and three-generator fundamental groups in full detail. The parameters that are given are explicit. Although the subject is very classical, such a complete study is done for the first time. Also in this part, Sean Lawton and Elisha Peterson develop a diagrammatic approach to the study of the structure of the $\mathrm{SL}(2, \mathbb{C})$ -representations character variety for the free group on two generators using graphs that are called spin networks. Their work sits in the framework of geometric invariant theory, a theory that develops the idea (first started by Vogt and Fricke–Klein) of characterizing polynomial functions on $\mathrm{SL}(2, \mathbb{C})$ that are invariant under inner automorphisms, and that are expressible in terms of traces.

Another generalized Teichmüller space that is considered in this volume is the space of Lorentzian 3-manifolds of constant curvature which are products of surfaces with the real line, in which Thurston’s hyperbolic geometry techniques (laminations, earthquakes, grafting and so on) were brought in in the 1990s by Geoffrey Mess.

3.1 Complex projective structures

As already said, the study of moduli of complex projective structure is intimately related to that of Teichmüller space. Thus it is natural to include in the Handbook a chapter on complex projective structures.

Complex projective structures on surfaces already appear in a substantial manner in the work of Poincaré. The relation between complex projective structures on surfaces and Teichmüller theory was developed by Bers and his collaborators in the 1960s. For instance, the Bers embedding of Teichmüller space can be described in terms of complex projective structures.

The model space for complex projective geometry on surfaces is the complex projective line \mathbb{CP}^1 , that is, the space of 1-dimensional complex vector subspaces of \mathbb{C}^2 , with transformation group induced from the linear transformations of \mathbb{C}^2 . Equivalently, we can consider the model space of complex projective geometry on surfaces as the Riemann sphere $\mathbb{S}^2 = \mathbb{C} \cup \{\infty\}$ equipped with the group of transformations of the form $z \mapsto \frac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. Such transformations are called *fractional linear transformations*, or *Möbius transformations*, or *projective transformations*. A complex projective structure on a surface is then an atlas with charts in \mathbb{CP}^1 whose coordinate changes are restrictions of projective transformations. Markings of complex projective surfaces are defined as it is usually done in Teichmüller theory, that is, a marking is a homeomorphism from a fixed base surface to a surface equipped with a projective structure. There is a natural equivalence relation on the set of marked projective structures, defined, again as in Teichmüller theory, by the existence of a projective transformation in the correct homotopy class. If S is a closed surface of genus ≥ 2 , we shall denote by $\mathcal{P}(S)$ the space of equivalence classes of marked complex projective structures on S .

In what follows, “projective” means “complex projective”.

Since the projective transformations of the sphere are holomorphic, a projective structure on a surface has an underlying conformal structure. In other words, there is a *forgetful map* $\pi : \mathcal{P}(S) \rightarrow \mathcal{T}(S)$ from the space of marked projective structures on S to the Teichmüller space $\mathcal{T}(S)$ of S . This map makes the space $\mathcal{P}(S)$ a fiber bundle over Teichmüller space.

As it is the case for hyperbolic structures, projective structures, through their holonomy representation, can be studied in the context of the representation theory of the fundamental group of the surface S in the group $\mathrm{PSL}(2, \mathbb{C})$. There is a complex structure on the space $\mathcal{P}(S)$, and from works of Hejhal, Earle and Hubbard, it follows that the holonomy map from the space $\mathcal{P}(S)$ to the character variety of representations of $\pi_1(S)$ in $\mathrm{PSL}(2, \mathbb{C})$ is a local biholomorphism. Chapter 12 contains a review of basic properties of holonomy maps of projective structure, as well as a discussion of other issues of representation theory (discreteness, degeneration, etc.) that have been studied in depth by various authors, in particular by D. Dumas.

There are several ways of parameterizing projective structures on surfaces, and one classical way uses Schwarzian derivatives. The Schwarzian derivative is a differential operator which is invariant under Möbius transformations. It was already studied in the nineteenth century, in relation with the *Schwarzian differential equation* $w''(z) + \frac{1}{2}q(z)w(z) = 0$, where z varies in a domain of the Riemann sphere and where q is a holomorphic function.

We recall that the Schwarzian derivative of a Möbius transformation is zero, and that, in some sense, the Schwarzian derivative of a conformal map is a measure of how far this map is from being a Möbius transformation. The Schwarzian derivative can also be considered as a measure of the difference between two projective structures.

There is an intimate relationship between Schwarzian derivatives and quadratic differentials, the latter being certainly more familiar to Teichmüller theorists. The Schwarzian derivative Sf of a holomorphic function f of one complex variable is defined by the formula

$$(Sf)(z) = \frac{f'''(z)}{f'(z)} - \frac{3}{2} \left(\frac{f''(z)}{f'(z)} \right)^2$$

(the formula is not important for what follows). A quadratic differential appears from a projective structure on a surface by taking the Schwarzian derivative of a developing map of that structure. Using this fact, Schwarzian derivatives establish a correspondence between projective structures on a given surface with the space of holomorphic quadratic differentials on that surface. In this correspondence, each fiber $\pi^{-1}(X)$ of the map $\pi : \mathcal{P}(S) \rightarrow \mathcal{T}(S)$ over a point X in $\mathcal{T}(S)$ is identified with the vector space of holomorphic quadratic differentials on a Riemann surface representing X . Since the vector space of holomorphic quadratic differentials over a surface is also the cotangent space to Teichmüller space at the corresponding point, the theory of the Schwarzian derivative makes an identification between the space $\mathcal{P}(S)$ and the cotangent bundle $T^*\mathcal{T}(S)$ of Teichmüller space.

By the Riemann–Roch theorem, the space of quadratic differentials on a closed Riemann surface of genus $g \geq 2$ is a complex vector space of dimension $3g - 3$. As a consequence, the space of projective structures is a fiber bundle over Teichmüller space, with fiber a complex vector space of dimension $3g - 3$. This directly shows that $\mathcal{P}(S)$ is homeomorphic to a cell of complex dimension $6g - 6$.

The parametrization of $\mathcal{P}(S)$ by the fiber bundle of quadratic differentials obtained via the Schwarzian derivative is called the *Schwarzian parametrization* of $\mathcal{P}(S)$.

Thurston produced another parametrization for the space $\mathcal{P}(S)$, whose definition uses the techniques of hyperbolic geometry and of measured laminations. This is a parametrization by $\mathcal{ML}(S) \times \mathcal{T}(S)$, where the Teichmüller space $\mathcal{T}(S)$ is seen as a space of (equivalence classes) of hyperbolic structures and where $\mathcal{ML}(S)$ is the space of measured laminations on S .⁹ The homeomorphism $\mathcal{ML}(S) \times \mathcal{T}(S) \simeq \mathcal{P}(S)$ uses Thurston's general *grafting* operation, which is one of the main tools in the geometric study of complex projective structures. This operation produces from a measured lamination on a hyperbolic surface, considered as a projective structure, a new projective structure. Grafting is first defined when the measured lamination is a weighted simple closed geodesic. In this case, one cuts the surface along that closed geodesic, and introduces between the two boundary components thus obtained a Euclidean annulus whose circumference is equal to the common length of the geodesic boundary components and whose width is determined by the transverse measure of the closed geodesic (seen as an element of \mathcal{MF}) that we started with. Grafting a hyperbolic structure over an arbitrary measured lamination μ is then defined by taking a sequence of weighted simple closed geodesics converging to μ and showing that there is a limiting complex projective structure, which is well defined independently of the approximating sequence. The resulting homeomorphism $\text{Gr}: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{P}(S)$ is called the *grafting homeomorphism*. Continuity, smoothness, properness and other properties of various maps that are associated to the grafting construction were studied by Tanigawa, Scannell, Wolf, Dumas and others, and they are discussed by David Dumas in Chapter 12 of this volume.

Thurston defined a conformal Kobayashi-like distance on each projective surface, which we call the Thurston distance. There is an infinitesimal version of the Thurston distance, in which the norm of a tangent vector v is the infimum of the norm of all vectors v' in the Poincaré disk, such that there exists a complex projective immersion of this disk into the surface, sending v' to v . This definition is analogous to the definition of the infinitesimal Kobayashi distance on a complex space, where one also takes the infimum over all holomorphic immersions of the Poincaré disk. When the projective surface is obtained by a simple grafting operation (that is, the operation of inserting a Euclidean annulus in a hyperbolic surface), the Thurston metric is the one induced by the length structure associated to the constant-curvature structures on the parts.

Chapter 12 of this volume contains a detailed exposition of the Schwarzian and of the grafting parametrizations of the space $\mathcal{P}(S)$ of equivalence classes of projective

⁹Thurston's work on that subject is essentially unpublished, and one proof of the isomorphism $\mathcal{ML}(S) \times \mathcal{T}(S) \simeq \mathcal{P}(S)$ was written by Y. Kamishima and S. P. Tan.

structures, and a study of the various relations between these two parametrizations. This involves an analysis of the relation between quadratic differentials and grafting as well as a study of the asymptotic aspects of $\mathcal{P}(S)$ related to the two parametrizations. The chapter also contains a report on fundamental relations between holonomy homomorphisms of projective structures and the grafting construction. It also contains a description of the holonomy representation of a projective structure in terms of convex hulls, pleating loci and the bending deformation in 3-dimensional hyperbolic space. These constructions are at the basis of the beautiful relations between complex projective geometry on surfaces and 3-manifold topology, whose study was started by Sullivan and Thurston and which later on was developed by Epstein & Marden.

Chapter 12 also contains an exposition of results by Dumas on a grafting map compactification of the space $\mathcal{P}(S)$, and a description of a fiber of the map $\mathcal{P}(S) \rightarrow \mathcal{T}(S)$ with respect to this compactification. For a given point X in Teichmüller space, this description involves a beautiful map $i_X : \mathbb{P}\mathcal{ML}(S) \rightarrow \mathbb{P}\mathcal{ML}(S)$ called the *antipodal involution*, obtained by transporting the involution $\phi \mapsto -\phi$ defined on the space $Q(X)$ of quadratic differentials using the Hubbard–Masur parametrization of $Q(X)$ by the space of measured foliations $\mathcal{ML}(S)$ on the surface S . Dumas also studied another compactification of fibers of $\mathcal{P}(S) \rightarrow \mathcal{T}$, which he calls the *Schwarzian compactification*. It is obtained by attaching a copy of the projective space of quadratic differentials of a Riemann surface representing the given point in $\mathcal{T}(S)$ by taking limits of Schwarzian derivatives. He presents a result that compares the two compactifications of the fibers.

3.2 Circle packings

In Chapter 13 of this Handbook, Sadayoshi Kojima reports on rigidity and on flexibility properties of circle packings on complex projective surfaces, and on the relation of circle packings with Teichmüller space.

A *circle* in the complex projective line $\mathbb{CP}^1 = \mathbb{S}^2$ can be viewed as either a geometric circle for the canonical metric on the sphere \mathbb{S}^2 , or, using the stereographic projection that identifies \mathbb{S}^2 with $\mathbb{C} \cup \{\infty\}$, as a Euclidean circle or a straight line in \mathbb{C} . Circles are invariant by complex projective transformations. As a matter of fact, complex transformations are characterized by the fact that they send circles to circles. This shows that we have a natural local notion of a “circle” on a complex projective surface. In this sense, on a hyperbolic surface, geodesics, horocycles, hypercycles and geometric circles are all circles with respect to the underlying projective structure. This can be clearly seen by taking one of the usual models of hyperbolic space.

In Chapter 13, Kojima studies circles and circle packings on projective Riemann surfaces. Here, the definition of a circle is more restrictive, and one calls *circle* a homotopically trivial simple closed curve that is locally contained in a circle of \mathbb{S}^2 , the term “locally” referring to the image of the curve under the local charts of the projective structure. A *circle packing* is a collection of circles meeting tangentially, with the property that all the complementary regions are curvilinear triangles.

Although interesting problems on circle packings were already noticed by Koebe in the 1930s (and probably before), it is certainly William Thurston who made this into a subject in itself; first in his 1976 Princeton Notes, and then in 1985, when he made the conjecture that certain maps between circle packings converge to conformal maps. This conjecture was proved in 1987 by Burton Rodin and Dennis Sullivan, and it can be considered now as being at the heart of the theory of discrete conformal maps.

Chapter 13 first gives a report on Thurston's reconstruction and generalization of Andreev's theorem on circle packings, following Chapter 13 of Thurston's Princeton 1976 Notes. Andreev's theorem, as revisited by Thurston, is an existence and uniqueness result. The existence part says that a given graph on a Riemann surface determines a constant curvature surface equipped with a circle packing whose combinatorics is encoded by the graph. The uniqueness part says that two such structures encoded by the same graph are related by a global projective map. The only requirement on the graph is that its lift to the universal cover of the surface is a genuine triangulation.

The question of the realization of circle packings on Riemann surfaces was already studied by Koebe in the 1930s. Andreev's work on the subject was published in 1970. The results by Koebe and Andreev concern the case of the closed surface of genus zero (that is, the sphere). Thurston worked out the case of arbitrary genus. In the case of genus ≥ 2 , Thurston's result states that there is a unique hyperbolic structure equipped with a circle packing, realizing the given combinatorics.

Kojima then reports on flexibility results whose starting point is a work by R. Brooks who studied, instead of circle packings, more general circle patterns of circles, where complementary regions are allowed to be either triangles or quadrilaterals. Kojima reports on a method due to Brooks of parametrizing these generalized circle patterns by continued fractions, in the case where one of the complementary components is a quadrilateral. The idea is natural, and it consists in trying to fill in the quadrilateral region by successively inserting circles tangent to the rest of the configuration. Adding a new circle creates in general a new quadrilateral, but there are exceptional cases where the added circle is tangent to all the boundary sides of the quadrilateral. In this case the result is a genuine circle pattern, which, as was said before, is a rigid object, and the process ends there. Brooks continued fraction parameter is a projective invariant.

Kojima also outlines recent work on the moduli spaces of pairs (S, P) where S is a projective surface and P a circle packing whose combinatorics is fixed. He describes a projective invariant for such pairs, based on the cross ratio, which was worked out in joint work by Kojima, Mizushima & Tan. The deformation space has a natural structure of a semi-algebraic space. In the last part of Chapter 13, Kojima formulates and motivates a conjecture that states a precise relation between this parameter space and Teichmüller space.

3.3 Lorentzian geometry

A few words about Lorentzian geometry are in order.

A Lorentzian n -manifold M is a smooth n -dimensional manifold equipped with a nondegenerate bilinear symmetric form of signature $(-, +, \dots, +)$ at the tangent space at each point of M . A Lorentzian n -manifold is a pseudo-Riemannian manifold of signature $(1, n - 1)$. Denoting the bilinear form by $\langle \cdot, \cdot \rangle$, if v is a tangent vector, then the real number $\langle v, v \rangle$ (and not its square root) is called the *norm* of x .

Lorentzian manifolds are the most important pseudo-Riemannian manifolds after the Riemannian ones. This is due in part to the use of Lorentzian manifolds in physics. Indeed, 4-dimensional Lorentzian geometry is the setting of general relativity. As a consequence, the language of Lorentzian geometry is often borrowed from the language of physics. For instance, the local parameters in a Lorentzian 4-manifold are seen as three *spatial* parameters and one *temporal* parameter.

From the mathematical point of view, the basic problems of general relativity can be stated in terms of finding Lorentzian metrics on some given manifold that satisfy some partial differential equation (namely, Einstein's equations) involving the Ricci and the scalar curvature tensors.

As in Riemannian geometry, there is a notion of norm-preserving parallel vector transport in Lorentzian geometry. A Lorentzian manifold has a unique affine torsion-free connection which preserves the Lorentzian metric, which is also called the Levi-Civita connection. There are associated notions of curvature, of geodesics and of exponential map. However, the intuition that we have in Riemannian geometry may be misleading in Lorentzian geometry, partly because norms of vectors in a Lorentzian manifold can be negative. One consequence is that in general, geodesics are not distance-minimizing.

We need to recall some more terminology. A tangent vector to a Lorentzian manifold is said to be *time-like* (respectively *space-like*) if its norm is negative (respectively positive). A nonzero vector of zero norm is said to be a *light vector*. A *causal* vector is either a time-like vector or a light vector. A C^1 curve in a Lorentzian manifold is *time-like* (respectively, *space-like*, etc.) if all of its tangent vectors are time-like (respectively, space-like, etc.). A hypersurface in a Lorentzian manifold is *space-like* if the restriction of the Lorentzian metric tensor to the tangent space at each point of that hypersurface is Riemannian. A *flat spacetime* is an oriented Lorentzian manifold together with an orientation for every causal curve. A *Cauchy surface* in a flat spacetime is a codimension-one isometrically immersed Riemannian submanifold which intersects in exactly one point every maximally extended causal curve. A flat spacetime is said to be *globally hyperbolic* if it admits a Cauchy surface. The concept of Cauchy surface was introduced by physicists working in general relativity, and it turned out to be a fundamental concept in Lorentzian geometry, as we shall see below. From the physics point of view, the existence of a Cauchy surface has to do with the so-called "causality condition", which says that there are no time-like closed curves, as it is expected in reality.

Unlike the case of Riemannian manifolds, it is not true that any smooth manifold admits a Lorentzian structure. On the other hand, an important feature of Lorentzian geometry which parallels the Riemannian case is that two Lorentzian manifolds of the same dimension and having the same constant curvature are locally isometric.

Minkowski n -space, that is, the vector space \mathbb{R}^n equipped with a nondegenerate symmetric bilinear form of signature $(-, +, \dots, +)$, is a linear model for Lorentzian n -manifolds. The Minkowski model of $(n - 1)$ -dimensional hyperbolic geometry sits inside Minkowski n -space as one sheet of a hyperboloid with two sheets. This is a hypersurface that consists of future-directed time-like vectors. In fact, Minkowski Lorentzian n -space is foliated by $(n - 1)$ -Riemannian manifolds of constant negative curvature. This should be a hint for a strong relationship between Lorentzian geometry and hyperbolic geometry.

In each dimension n and for every real number κ , there is a “model Lorentzian manifold” X , that is, a unique simply connected Lorentzian manifold of dimension n and of constant curvature κ . Furthermore, such a space X has the “analytic continuation property”, that is, every isometry between two open sets of X extends to a global isometry of X . Using this fact, a Lorentzian manifold of constant curvature can be considered as a homogeneous geometric structure, that is, as a (G, X) manifold in the sense of Ehresmann. Thus, a Lorentzian manifold of constant curvature can be defined by an atlas whose charts take their values in the model manifold X and whose coordinate change functions are restrictions of isometries of the model manifold. Again, as in the Riemannian case, there is a notion of developing map and of holonomy representation. Restricting to $\kappa \in \{0, -1, 1\}$, the model spaces for 3-dimensional Lorentzian manifolds are called the 3-dimensional Minkowski spacetime ($\kappa = 0$), de Sitter spacetime ($\kappa = 1$), and anti de Sitter spacetime ($\kappa = -1$). De Sitter space can be thought of as the space of planes in hyperbolic space.

We now restrict the discussion to 3-dimensional (more commonly called $(2 + 1)$ -dimensional) Lorentzian manifolds.

In 1990, Geoffrey Mess wrote a fundamental paper, called *Lorentz spacetimes of constant curvature*.¹⁰ The paper realized a major breakthrough in the field; in particular because it brought into Lorentzian geometry the techniques that had been introduced a few years before by Thurston in hyperbolic geometry and in complex projective geometry (measured laminations, group actions on trees, earthquakes, grafting, bending, and so on).

To say it in very few words, Mess obtained a classification of the space of Lorentzian metrics of constant curvature on manifolds which are of the form $S \times \mathbb{R}$, where S is a closed orientable surface S of genus ≥ 2 . In other words, Mess gave a geometric parametrization of the moduli space of $(2+1)$ maximal globally hyperbolic spacetimes of constant curvature κ , for $\kappa \in \{-1, 0, 1\}$, that contain a compact Cauchy surface. (The case $\kappa = 0$ was completed by Kevin Scannell in 1999). The problem that Mess solved was explicitly posed by Edward Witten in 1989. As it is the case in Teichmüller theory, there is a natural equivalence relation on the space of Lorentzian metrics of

¹⁰For 14 years, this paper was circulated as a preprint; it is now published in *Geometriae Dedicata*.

constant curvature, and two metrics on $S \times \mathbb{R}$ are equivalent if they are isotopic (that is, if they differ by a diffeomorphism of $S \times \mathbb{R}$ which is isotopic to the identity). Mess showed that the space of equivalence classes of metrics satisfying the above properties is a generalized Teichmüller space parametrized by $\mathcal{T} \times \mathcal{ML}$, where \mathcal{T} is the (usual) Teichmüller space of S and \mathcal{ML} its space of measured laminations. It is useful to recall here the following two facts:

- There is a well-known homeomorphism between the space $\mathcal{T} \times \mathcal{ML}$ and the cotangent bundle of the Teichmüller space of S , obtained as a consequence of the result by Hubbard and Masur stating that the space of holomorphic quadratic differentials on a Riemann surface (which can be naturally identified with the cotangent space to Teichmüller space at the point represented by that surface) can be identified with the space of measured laminations on that surface.
- The space $\mathcal{T} \times \mathcal{ML}$ is also reminiscent of the parametrization of the space of equivalence classes of complex projective structures on S , obtained through Thurston's grafting operation. This is not a pure coincidence, and grafting plays an essential role in this work of Mess.

It is also interesting to note that Mess obtained a new proof of Thurston's earthquake theorem for the case of compact surfaces, using his classification of spacetimes.

Generalizing Mess's work to the case where the surface S is not compact requires more than the grafting operation. The *canonical Wick rotation*, which has been introduced in this context by Benedetti and Bonsante, is another basic tool for understanding the space of Lorentzian metrics of constant curvature on the product $S \times \mathbb{R}$, and explaining the parametrization by $\mathcal{T} \times \mathcal{ML}$. The Wick rotation¹¹ is a transformation, acting as a $\pi/2$ -rotation, that relates Lorentzian geometry and Riemannian geometry. Roughly speaking, the idea is to consider the parameter t in the formula $ds^2 = -dt^2 + dx_1^2 + dx_2^2 + \cdots + dx_{n-1}^2$ defining a Riemannian metric, and the formula $ds^2 = dt^2 + dx_1^2 + dx_2^2 + \cdots + dx_{n-1}^2$, defining a Lorentzian metric, as restrictions of one complex parameter to the imaginary axis and to the real axis respectively. The Wick rotation was already successfully used in physics. In particular, it established a relation between the Schrödinger equation of quantum mechanics and the heat equation of thermodynamics.¹²

Given a manifold M equipped with a Riemannian metric and a non-vanishing vector field X , the Wick rotation produces a Lorentzian metric on M for which X is a timelike vector field. The Lorentzian metric also depends on the choice of two

¹¹Named after the Italian theoretical physicist Gian-Carlo Wick (1909–1992).

¹²There are several well-known occurrences in geometry where the fact of complexifying a real parameter turns out to be very fruitful. To stay close to our subject matter, we can just mention here the complexification of earthquake coordinates which establishes relations between Weil–Petersson geometry, projective structures, pleated surfaces and quasifuchsian groups (see e.g. the work of McMullen on the extension of earthquake paths to proper holomorphic maps from disks into Teichmüller space), the complexification of Thurston's shear coordinates for measured laminations which also gives a parametrization of the space of hyperbolic 3-manifolds with fundamental group equal to a surface fundamental group (work of Bonahon), or the complex measures that define quake-bend maps that appear in the work of Epstein and Marden, where real measures correspond to earthquakes and imaginary measures correspond to bending.

positive functions α and β , called *rescaling functions*. The notion of rescaling is another fundamental object in the theory that is developed by Riccardo Benedetti and Francesco Bonsante in Chapter 14 of this volume. Another important tool in this theory is the notion of *cosmological time*, introduced in this context by Benedetti and Guadagnini.

Finally, we mention that the Wick rotation-rescaling theory also provides geometric relations between spacetimes of different curvatures, and between such spacetimes and complex projective structures. The theory transforms the various spacetimes into hyperbolic 3-manifolds that carry at infinity the same projective structure.

3.4 Fricke–Klein coordinates

As is well known, the Teichmüller space of a surface S can be described as a subspace of a space of conjugacy classes of representations of the fundamental group of S in Lie groups, in particular the Lie group $SL(2, \mathbb{C})$. This point of view was already used by R. Fricke and F. Klein in the nineteenth century. It is a well-known fact that the trace of a 2×2 matrix is a conjugacy invariant, and Fricke and Klein studied the question of parametrizing spaces of conjugacy classes of representations of the fundamental group of a surface in $SL(2, \mathbb{C})$ by a finite number of traces, viz. traces of images of base elements of the group and of some of their combinations. The space of conjugacy classes of representations is referred to here as the *character variety*. Trace coordinates are often called *Fricke–Klein coordinates*. In the case where the fundamental group of the surface is a free group of rank two, a result of Vogt, Fricke and Klein, which is quoted in several chapters of this volume, gives a characterization of two-variable functions that are invariant under the action of $SL(2, \mathbb{C})$ on itself by conjugation. This characterization leads to a description of the character variety by a set of polynomial equations, involving the traces of the images of three elements of the fundamental group.

Chapter 15 of this volume, written by Bill Goldman, considers Fricke–Klein coordinates in detail. Goldman presents the complete results with explicit formulae in the case of two- and three-generator surface groups. Non-orientable surfaces are also considered. The chapter also contains an exposition of the background material in invariant theory and in hyperbolic geometry that is needed in order to obtain the formulae. Goldman also gives formulae relating the trace coordinates to the Fenchel–Nielsen coordinates in the case of a particular two-generator surface, namely the one-holed torus.

3.5 Diagrammatic approach

Chapter 16 by Sean Lawton and Elisha Peterson concerns the character variety of $SL(2, \mathbb{C})$ -representations of the free group F_2 on two generators. One obvious relation with surface geometry stems from the fact that F_2 is the fundamental group of the pair

of pants and of the torus with one hole. The term *character variety* refers here to the orbit space of the subset of completely reducible representations under the action of $SL(2, \mathbb{C})$ by conjugation. As is well known, this character variety is an algebraic set. It contains the Teichmüller space of the surface as a subspace, and it also contains moduli spaces of other geometric structures.

The main object of this chapter is to develop a diagrammatic approach to the study of the character variety. The diagrams that appear here are graphs called *spin networks*. These graphs are used as a diagrammatic tool in the description of a natural additive basis for the coordinate ring of the character variety. The elements of this basis are the *central functions*, and the authors make a detailed study of the properties of this basis. Diagrammatic calculus is used to make explicit the symmetries of this basis. The authors also give a new constructive proof of results by Vogt and Fricke–Klein that are considered from a different viewpoint in Chapter 15 by Goldman.

Diagrammatic calculus has been thoroughly used by physicists, the most notable examples being certainly the diagrams that appear in the works of Richard Feynman and of Roger Penrose. In mathematics, it is known that diagrammatic calculus considerably simplifies certain proofs and algebraic computations.

Spin networks, as a diagrammatic tool, have been previously used in the description of quantum angular momentum by Penrose. They also appear, together with central function bases, in the work of John Baez (1996) in relation to gauge theory. More recent related work was done by Adam Sikora (2001), who considered graphs similar to spin networks, and who used the graphical calculus in the deformation theory of the SL_3 -character variety of the fundamental group of a 3-manifold, with a view on applications to quantum invariants of 3-manifolds. We finally mention that Florentino, Mourão and Nunes (2004) used similar tools in a work that is related to the geometric quantization of the moduli space of flat connections on a Riemann surface.

4 The Grothendieck–Teichmüller theory

The *Grothendieck–Teichmüller theory* is an expression that was coined after Alexandre Grothendieck wrote his *Esquisse d'un programme* (1983), a detailed research program which was part of an application for a researcher position at CNRS. The theory that is referred to in this expression has several facets, and the Grothendieck–Teichmüller theory that is reported on in this volume includes the subjects of *dessins d'enfants*, the *reconstruction principle*, and the theory of the *solenoid*.

Let me start by saying a few words on some of the objects that play important roles in this theory, namely, *dessins d'enfants*, the absolute Galois group, towers, profinite groups and the Grothendieck–Teichmüller modular group.

A *dessin d'enfant* is a finite graph embedded in an oriented connected surface, which has the following two properties:

- the complement of the graph is a union of cells;

- the vertices of the graph are colored black or white in such a way that the endpoints of any vertex do not have the same color.

One may wonder how such a simple definition leads to important developments, but in some sense this is often the case in mathematics.

It is good to recall that important ideas in Grothendieck–Teichmüller theory originate in algebraic geometry.

Grothendieck introduced *dessins d'enfants* in 1984 as a tool for the study of the absolute Galois group of the field of rational numbers, and in relation with some holomorphic branched covers of surfaces called *Belyi functions*. After that, the use of *dessins d'enfants* in Riemann surface theory and in low-dimensional topology has been highlighted by many authors. It turned out that *dessins d'enfants* make connections between several fields of mathematics, e.g. the Galois theory of algebraic numbers, Riemann surfaces, combinatorial group theory and hyperbolic geometry.

Let us note that *dessins d'enfants* were already used in two chapters of Volume I of this Handbook, namely, those written by Harvey and by Herrlich & Schmithüsen, in relation with Teichmüller disks. It is sometimes useful to have different points of view on an important topic, written by different authors. In Chapter 18 of the present volume, *dessins d'enfants* are considered in more detail, and from a point of view closer to that of Grothendieck's original. This point of view heavily uses the language and techniques of algebraic geometry. *Dessins d'enfants* are also considered, in the same chapter, in relation with *origamis*, which are special classes of Riemann surfaces on which significant progress has been made recently.

As already stated, the Grothendieck–Teichmüller theory studies actions of the absolute Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ of the field of rational numbers. Here, $\overline{\mathbb{Q}}$ is the field of algebraic numbers, that is, the algebraic closure of the field \mathbb{Q} of rational numbers, and $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is the topological automorphism group of the Galois extension $\overline{\mathbb{Q}}/\mathbb{Q}$. We note that there is no explicit description of the absolute Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, and partial understanding of this group is obtained by studying its actions on various spaces.

It is also worth noting that the representation theory of the absolute Galois group plays an important role in Wiles' proof of Fermat's Last Theorem.

One relation of Grothendieck's work with Teichmüller theory stems from the fact that one of Grothendieck's approaches to the analysis of the group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is via the action of that group on the “system” of all moduli spaces $\mathcal{M}_{g,n}$ (for varying g and n). Grothendieck calls this system the *Teichmüller tower*. In practice, a tower in this context is an object obtained either as the inverse limit of spaces, or as a profinite completion of groups. The word “tower” occurs at several places in the Grothendieck–Teichmüller theory. For instance, one has “towers of surfaces”, “towers of Teichmüller spaces”, “towers of fundamental groups”, “towers of mapping class groups” and so on. The Grothendieck–Teichmüller theory studies automorphisms of these objects, and makes relations between these objects and actions of the Galois groups on various associated spaces.

As already mentioned, the notion of profinite group is an important object in this theory. We recall that a profinite group is a Hausdorff, compact and totally disconnected

topological group which is isomorphic to a projective limit of an inverse system of finite groups. In some sense, a profinite group is obtained by assembling finite groups, and hence, profinite groups may be understood by studying their finite quotients. The absolute Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is itself an example of a profinite group. Indeed, $\overline{\mathbb{Q}}$ is the union of all the Galois finite normal extensions of \mathbb{Q} in \mathbb{C} , and $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is a projective limit of the finite Galois groups of these extensions. Algebraic fundamental groups of schemes, that appear in algebraic geometry, are other examples of profinite groups. (But fundamental groups in the sense of algebraic topology are not.) Any group G has a profinite completion \hat{G} , defined as the projective limit of the groups G/N , where N varies over the finite-index normal subgroups of G . There is a natural homomorphism $G \rightarrow \hat{G}$, which satisfies a natural universal property, and the image of G under this homomorphism is dense in \hat{G} .

The *Grothendieck–Teichmüller modular group* has been defined by Drinfel'd in 1991, as an extension of the absolute Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. This result by Drinfel'd somehow gave a natural setting for the action of the Galois group on the Teichmüller tower that was alluded to by Grothendieck in his *Esquisse d'un programme*. The Grothendieck–Teichmüller modular group is also the automorphism group of a tower of fundamental groupoids of a stack of moduli spaces equipped with *tangential base-points*. This group was studied by L. Schneps, P. Lochak, H. Nakamura, H. Tsunogai, H. Voelklein and T. Shaska and others. L. Schneps identified the Grothendieck–Teichmüller modular group with the automorphism group a tower of profinite completions of Artin braid groups. Let us also mention that the Grothendieck–Teichmüller theory has also applications in conformal field theory, and that there is a work in this direction done by B. Bakalov and A. Kirillov (related to previous work of Moore and Seiberg). More recently, P. Hu & I. Kriz worked out new relations between the Grothendieck–Teichmüller theory and conformal field theory. They described actions of the Galois group of a number field on the category of modular functors. We shall see in Chapter 18 that the Galois group of \mathbb{Q} also acts on origamis, which are closely related to dessins d'enfants.

The reconstruction principle is another important aspect of Teichmüller theory that was formulated by Grothendieck, inspired from ideas that originate in algebraic geometry. Chapter 17 of the present volume contains a detailed overview on that theory, written by Feng Luo, with an exposition of several important applications of that principle in low-dimensional topology. The reconstruction principle is related to the study of the Teichmüller tower and it gives rise to new kind of geometric structures, namely, $(\mathbb{Q}P^1, \text{SL}(2, \mathbb{Z}))$ structures, also called *modular structures*.

I have included the chapter on the Teichmüller space of the solenoid in the part of this volume dedicated to the Grothendieck–Teichmüller theory, because the study of the solenoid involves the Teichmüller tower, the mapping class group tower and other similar objects whose study is inherent in Grothendieck's program, without the language of algebraic geometry. This chapter could also have been included in Part A on the metric and the analytic theory, but I have the feeling that the fact of including it in the part on the Grothendieck–Teichmüller theory opens up a nice perspective.

Now let us review in more detail the three chapters that constitute Part D of this volume.

4.1 The reconstruction principle

Let S be a compact surface of negative Euler characteristic. An *essential subsurface* S' of S is a surface with boundary and with negative Euler characteristic embedded in S , such that no boundary component of S' bounds a disk in S or is isotopic to a boundary component of S . There is a hierarchy on the set of essential subsurfaces of S , in which the *level* of a surface S' is the maximal number of disjoint simple closed curves that cuts it into pairs of pants. In particular, level-zero surfaces are the pairs of pants, level-one surfaces are the four-punctured spheres and the one-holed tori, and level-two surfaces are the two-holed tori and the five-holed spheres.

Grothendieck's reconstruction principle says that some of the most important geometric, algebraic and topological objects that are associated to a surface S (e.g. the Teichmüller space, the mapping class group, the space of measured foliations, and spaces of representations in $SL(2, K)$ for a given field K) can be reconstructed from the corresponding spaces associated to the (generally infinite) set of level-zero, level-one and level-two essential subsurfaces of S . The geometric structures on the level-zero spaces are the building blocks of the general structures, and the structures on the level-one and the level-two spaces are the objects that encode the gluing. Paraphrasing Grothendieck from his *Esquisse d'un programme*, "the Teichmüller tower can be reconstructed from level zero to level two, and in this reconstruction, level-one gives a complete set of generators and level-two gives a complete set of relations".

Grothendieck's ideas were inspired by analogous situations in algebraic geometry, in particular by ideas originating in reductive group theory, where the semi-simple rank of a reductive group plays the role of "level".

In a series of extremely interesting and original papers, Feng Luo developed Grothendieck's intuition and made it precise. Chapter 17 of this volume, written by Luo, constitutes a detailed survey of various results in this theory.

A fundamental new object that appears in this theory is the notion of *modular structure*, a $(\mathbb{Q}P^1, SL(2, \mathbb{Z}))$ structure in the usual sense of a geometric structure defined by an atlas. Here, $\mathbb{Q}P^1 = \mathbb{Q} \cup \{\infty\}$ is seen as the set of rational points on the unit circle. Luo shows that the set of isotopy classes of essential simple closed curves on an oriented surface of level at least one is equipped with a modular structure which is invariant under the action of the mapping class group of the surface. The atlas for such a structure is obtained through some coordinate charts associated to level-one essential subsurfaces of the original surface. For these level-one surfaces, coordinate charts are homeomorphisms onto $\mathbb{Q}P^1$. We note that the idea of a modular structure for the set of isotopy classes of essential simple closed curves on the four-punctured sphere is already inherent in the work of Max Dehn done in the 1930s. It is easy to see that there is also a modular structure on the space of essential curves on the torus, and that this structure is natural with respect to the action of the mapping class group

of the torus on the space of curves. Luo describes in Chapter 17 a modular structure on the set of isotopy classes of pair of pants decompositions of a surface.

Another application of the reconstruction principle presented in Chapter 17 concerns characters of $\mathrm{SL}(2, K)$ -representations, where K is an arbitrary field. Let us review the definition.

Let $\mathcal{J}(S)$ be the set of isotopy classes of essential simple closed curves on S . An $\mathrm{SL}(2, K)$ -character on $\mathcal{J}(S)$ is defined here as the map induced by the trace function of a representation of $\pi_1(S)$ in $\mathrm{SL}(2, K)$. Luo calls an $\mathrm{SL}(2, K)$ -trace function on $\mathcal{J}(S)$ a function $\mathcal{J}(S) \rightarrow K$ whose restriction to every subset $\mathcal{J}(S')$ of $\mathcal{J}(S)$ is an $\mathrm{SL}(2, K)$ -character on $\mathcal{J}(S')$, where $S' \subset S$ is an essential level-one surface.

From the work of Fricke and Klein to which we already referred at several occasions, it follows that the trace function defined on the fundamental group $\pi_1(S)$ of S , with respect to an $\mathrm{SL}(2, K)$ -representation of $\pi_1(S)$, is determined by the restriction of this function to the elements of $\pi_1(S)$ that are represented by simple curves.

Luo proves that any $\mathrm{SL}(2, K)$ -trace function on $\mathcal{J}(S)$ is the $\mathrm{SL}(2, K)$ -character on that set, except for a finite number of cases which he enumerates. The result was conjectured by Grothendieck. To prove this fact, Luo produces a complete set of equations that express the fact that a function $\mathcal{J}(S) \rightarrow K$ is an $\mathrm{SL}(2, K)$ -character, and he proves that these equations are supported on the essential level-two subsurfaces of S . The consequence is that the character functions satisfy Grothendieck's reconstruction principle, except for a finite number of functions supported on surfaces of genus 0 with $n \geq 5$ punctures.

Another application of Grothendieck's reconstruction principle, which is also due to Luo, concerns geometric intersection functions defined on $\mathcal{J}(S)$. Luo calls a function $f: \mathcal{J}(S) \rightarrow \mathbb{R}$ a *geometric intersection function* if there exists a measured lamination μ on S such that f is the intersection function with μ , that is, $f(\alpha) = i(\alpha, \mu)$ for every α in \mathcal{J} . Luo proves that a function $\mathcal{J}(S) \rightarrow \mathbb{R}$ is a geometric intersection function if for every essential level-one subsurface S' of S , the restriction of f to $\mathcal{J}(S')$ is a geometric intersection function.

A related result, again due to Luo, is that geometric intersection functions on the set of isotopy classes of essential curves of a level-one surface are characterized by two homogeneous equations in the $(\mathbb{Q}P^1, \mathrm{PSL}(2, \mathbb{Z}))$ -structure on these subsurfaces.

Applications of the reconstruction principles in the study of Teichmüller spaces, measured foliation spaces, and mapping class groups are also discussed in the same chapter.

4.2 Dessins d'enfants

In Chapter 18, Frank Herrlich and Gabriela Schmithüsen give an overview of the theory of dessins d'enfants, and of another class of combinatorial objects, namely origamis, and they develop the relation between the two classes.

We already recalled the classical result of Riemann stating that any compact Riemann surface can be defined as an algebraic curve, that is, as the zero set of a two-

variable polynomial. In this setting, the most useful polynomials are probably those whose coefficients are in the field $\overline{\mathbb{Q}}$ of algebraic numbers. A celebrated result due to G. Belyĭ asserts that any compact Riemann surface represented by an algebraic curve with coefficients in $\overline{\mathbb{Q}}$ is a ramified meromorphic covering of the Riemann sphere, in which the ramification occurs over at most three points. This leads to the introduction of the following important notion: a *Belyĭ morphism* $X \rightarrow \mathbb{P}^1(\mathbb{C})$ is a ramified covering from a Riemann surface X to the complex projective line $\mathbb{P}^1(\mathbb{C})$, which is ramified over at most three points. Using this notion, Belyĭ's theorem states that the Riemann surface X can be defined as an algebraic curve over the field $\overline{\mathbb{Q}}$ if and only if there exists a Belyĭ morphism $X \rightarrow \mathbb{P}^1(\mathbb{C})$.

Chapter 18 contains an exposition of the fact that the following categories of objects are in natural one-to-one correspondence:

- equivalence classes of Belyĭ morphisms;
- equivalence classes of dessins d'enfants;
- equivalence classes of bipartite ribbon graphs;
- conjugacy classes of finite index subgroups of F_2 , the free group on two generators;
- transitive actions of F_2 on a symmetric group S_d of permutations of d objects up to conjugacy in S_d .

From the correspondence between the first two items in this list, it follows that the absolute Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the class of equivalence classes of dessins d'enfants. There is still no explicit description of this action, but the correspondence leads to important results, such as the embedding of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ into the Grothendieck–Teichmüller group \widehat{GT} . We also note that L. Schneps described a faithful action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on a class of equivalence classes of trees.

Herrlich and Schmithüsen provide a proof of the fact that the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on dessins d'enfants is faithful. They address the question of finding invariants of the actions mentioned above. In other words, the question is to find properties of equivalence classes of dessins d'enfants (and of the other related objects) that remain invariant under the action of the Galois group. There is no complete list of such invariants, but Herrlich and Schmithüsen study a few invariants such as the *genus* and the *valency lists* of a dessin d'enfants. They explain how the Galois action on dessins induces an injective group homomorphism of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ in the automorphism group $\text{Aut}(\hat{F}_2)$ of the profinite completion \hat{F}_2 of F_2 . This is then explained in the general context of actions of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on algebraic fundamental groups of schemes.

The embedding $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{Aut}(\hat{F}_2)$ leads to the introduction of the Grothendieck–Teichmüller group \widehat{GT} , introduced by Drinfel'd, which is a subgroup of $\text{Aut}(\hat{F}_2)$ which contains the image of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

The second part of Chapter 18 deals with origamis. These are surfaces obtained by taking a finite number of isometric squares in the Euclidean plane and gluing them along their boundaries by using Euclidean translations. Origamis already appeared in Volume I of this Handbook, and they are also mentioned in Chapter 10 of the present

volume, in particular regarding their affine groups. There are several questions on origamis that are still unsolved, regarding their arithmetic theory, their occurrence as Teichmüller disks in moduli space, and so on.

In Chapter 18, origamis are studied in parallel with dessins d'enfants. One can see the relation between these two classes of objects in the following manner: whereas a dessin d'enfant is associated to a finite unramified covering of the sphere with three points deleted, an origami is associated to an unramified finite covering of the torus with one point deleted. It may be useful to note here that the sphere with three punctures and the torus with one puncture are exactly the surfaces whose fundamental group is a free group on two generators.

Herrlich and Schmithüsen give a list of classes of objects that are equivalent to origami curves. This list is analogous to the list that we mentioned above, concerning dessins d'enfants. Then the authors report on the relation between origami curves and dessins d'enfants. More precisely, they show that an origami curve can be interpreted as a dessin d'enfants, and they show by examples how to produce a dessin associated to an origami curve. Dessins d'enfants can also be associated to a *cusp* of an origami curve, that is, a boundary point of the closure of the image of the origami curve in the Deligne–Mumford compactification of moduli space. Herrlich and Schmithüsen also study the action of the absolute Galois group on the set of origamis.

4.3 The solenoid

Taking a covering of a Riemann surface leads to a natural operation at the level of Teichmüller spaces. In fact, there is a contravariant functor from the category of oriented closed surfaces, with finite-degree orientation-preserving covers between them as morphisms, to the category of finite-dimensional complex manifolds with holomorphic embeddings as morphisms. This functor associates to each Riemann surface its Teichmüller space and to each orientation-preserving covering $X \rightarrow Y$, the naturally induced holomorphic map $\mathcal{T}(Y) \rightarrow \mathcal{T}(X)$ between the corresponding Teichmüller spaces obtained by lifting conformal structures on Y to conformal structures on X . In some sense, the solenoid can be considered as a universal object arising from this theory of taking covers of surfaces.

The solenoid was introduced by Dennis Sullivan in the early 1990s, as the inverse limit of a tower of finite sheeted pointed covers of a pointed closed oriented surface of genus $g \geq 2$. In this setting, “pointed” means equipped with a basepoint, all covers are unbranched, and the order relation between pointed covers is defined by the existence of a factorizing cover. We note that the fact of specifying basepoints make factorizations unique whenever they exist.

More precisely, the family \mathcal{C} of pointed finite-order covers of a pointed base surface (S_0, x_0) , equipped with the partial order \preceq defined by factorizations of covers, is inverse directed, and the *compact solenoid* (also called the *universal hyperbolic solenoid*) \mathcal{S} is the inverse limit of this family. Thus, a point in the compact solenoid \mathcal{S} is a point y_0 on the base surface S_0 together with a point y_i on each finite covering

surface $\pi_i: S_i \rightarrow S_0$ such that $\pi_i(y_i) = y_0$, with the property that if two covers $\pi_i: S_i \rightarrow S_0$ and $\pi_j: S_j \rightarrow S_0$ satisfy $\pi_i \leq \pi_j$ and if $\pi_{i,j}$ is the factorizing covering, then $\pi_{i,j}(y_j) = y_i$.

The compact solenoid \mathcal{C} does not depend on the choice of the base surface (S_0, x_0) . This is a consequence of the fact that any two finite covers have a common finite cover.

The compact solenoid \mathcal{S} is equipped with the subspace topology induced from the product topology on the infinite product of all pointed closed surfaces that finitely cover the base surface. With this topology, \mathcal{S} is compact, and its local structure is that of a surface times a Cantor set. Thus, the compact solenoid has the structure of a *foliated space*, or a *lamination*. (These are spaces more general than the familiar foliated manifolds and laminations on manifolds.) The direction of the Cantor set is called the *transversal direction*. Using the language of foliation theory, the path-connected components of \mathcal{S} are called the *leaves*. In the solenoid, each leaf is homeomorphic to a disk and is dense in \mathcal{S} .

Sullivan introduced the compact solenoid as a sort of “universal dynamical system”. Independently of Sullivan’s original motivation, the compact solenoid turned out to be an interesting object that can be studied for itself. Such a study has been carried out by Sullivan, Biswas and Nag, and, more recently, by Šarić, Markovic, Penner and others.

Using the correspondence between unbranched covers of a surface and subgroups of its fundamental group, there is an equivalent definition of the solenoid that uses the directed set of subgroups of the fundamental group of the base surface, equipped with the inclusion order relation.

The compact solenoid can also be described as a principal G -bundle over the base surface, with G being the profinite completion of the fundamental group of the surface and with fibers homeomorphic to a Cantor set. In this respect, recall that the universal cover of a pointed surface (S_0, x_0) is a principal $\pi_1(S_0, x_0)$ -bundle over that surface, and that the compact solenoid appears as the principal G -bundle obtained by extending the structure group of this bundle from the fundamental group to its completion. (We recall that any group is naturally included in its profinite completion.) From this description, the compact solenoid can be thought of as a “universal closed surface”. The compact solenoid \mathcal{S} , as a lamination, has an invariant transverse measure which is induced by the Haar measure on the fiber group. This transverse measure on the solenoid is important. For instance, it can be used for obtaining a measure on the solenoid by taking the product of this transverse measure with the area form obtained from a hyperbolic structure on the leaves. It can also be used for integrating objects like quadratic differentials which are holomorphic on the leaves, and so on.

The compact solenoid is equipped with a rich variety of natural structures, that parallel analogous structures associated to compact surfaces. The examples of such structures that are of main interest for us here are complex structures and hyperbolic structures, and there is a uniformization theorem that connects them. A complex structure on \mathcal{S} is defined by an atlas whose transition maps are holomorphic when restricted to the local leaves, and are continuous in the transverse directions. The

solenoid, equipped with a complex structure, becomes a *Riemann surface lamination*. There is a notion of a quasiconformal map between Riemann surface laminations. Markovic and Šarić proved that any two homotopic quasiconformal maps between complex solenoids are isotopic by a uniformly quasiconformal isotopy. There is a space of Beltrami differentials on the compact solenoid, and a corresponding Teichmüller space $\mathcal{T}(\mathcal{S})$. The latter can be defined, as in the case of the Teichmüller space of a surface, either as a space of equivalence classes of Beltrami differentials, or as a space of equivalence classes of marked solenoids equipped with complex structures. The space $\mathcal{T}(\mathcal{S})$ is infinite-dimensional and separable (in contrast with infinite-dimensional Teichmüller spaces of surfaces, which are all non-separable). The space $\mathcal{T}(\mathcal{S})$ can also be naturally embedded as a complex submanifold of the universal Teichmüller space.

Let us mention that there is another object which has the same flavour as the Teichmüller space of the compact solenoid, and which was studied by Biswas, Nag & Sullivan. It is also related to the functor that we mentioned above, between the categories {closed oriented surfaces, finite covers} and {complex spaces, holomorphic maps}. This functor leads to a directed system of Teichmüller spaces, with order relation stemming from existence of holomorphic maps induced from coverings. The direct limit of this system is called the *universal commensurability Teichmüller space*, and it is denoted by \mathcal{T}_∞ . Like the solenoid itself, the space \mathcal{T}_∞ does not depend on the choice of the base surface, and it is equipped with a Teichmüller metric, induced from the Teichmüller metrics of the Teichmüller spaces of the surfaces that were used to define it. The space \mathcal{T}_∞ is also equipped with a Weil–Petersson metric. By a result of Biswas, Nag & Sullivan, the Teichmüller space of the compact solenoid, $\mathcal{T}(\mathcal{S})$, is the completion of the universal commensurability Teichmüller space \mathcal{T}_∞ , with respect to the Teichmüller metric.

We also mention a relation with algebraic geometry. Biswas, Nag & Sullivan used their work on the universal commensurability Teichmüller space to obtain a genus-independent version of determinant line bundles and of connecting Mumford isomorphisms. This theory provides a natural Mumford isomorphism between genus-independent line bundles, which is defined over the universal commensurability Teichmüller space \mathcal{T}_∞ , made out of the Mumford isomorphisms between determinant line bundles defined at the finite-dimensional levels.

Now back to the compact solenoid.

There is a natural notion of an automorphism group of the Teichmüller space of the solenoid that was also introduced by Biswas, Nag & Sullivan. These authors proved in 1996 that this group is isomorphic to the virtual automorphism group of the fundamental group of the base surface. We recall that the virtual automorphism group of a group G is the set of isomorphisms between finite index subgroups of G up to the equivalence relation that identifies two such isomorphisms if they agree on a finite index subgroup. The virtual automorphism group of G is also called the *abstract commensurator group* of G . For instance, the virtual automorphism group of \mathbb{Z} is the multiplicative group \mathbb{Q}^* . The relation with the solenoid stems from the fact

that there is a natural correspondence between homotopy classes of homeomorphisms between finite covers of a surface and elements of the virtual automorphism group of the fundamental group of that surface. A related natural object of study is the *baseleaf preserving mapping class group* of the compact solenoid \mathcal{S} , defined (modulo some technicalities) after the choice of a baseleaf, as the group of isotopy classes of baseleaf preserving self-homeomorphisms of this space \mathcal{S} . C. Odden proved in 2004 that the baseleaf preserving mapping class group of \mathcal{S} is naturally isomorphic to the virtual automorphism group of the fundamental group of the base surface. This result is considered as an analogue of the Dehn–Nielsen–Baer Theorem that describes the mapping class group of a closed surface of genus ≥ 1 as the outer automorphism group of its fundamental group. Markovic & Šarić proved that the baseleaf preserving mapping class group of the solenoid does not act discretely on $\mathcal{T}(\mathcal{S})$, a result which should be compared to the fact that in general, the mapping class group of surfaces of infinite type does not act discretely on the corresponding Teichmüller space.

The *non-compact solenoid*, also called the *punctured solenoid*, and denoted by \mathcal{S}_{nc} , is defined in analogy with the compact solenoid, as the inverse limit of the system of all pointed finite sheeted coverings of a base surface S_0 of negative Euler characteristic, except that here, S_0 is a punctured surface. A study of the noncompact solenoid was done by Penner & Šarić, who equipped that space with the various kinds of structures that exist on the compact solenoid, namely, complex structures, quasiconformal maps between them, a Teichmüller space, and a mapping class group which is isomorphic to a subgroup of the commensurator group of the base surface preserving the peripheral structure (in analogy with the case of the mapping class group of a punctured surface).

Chapter 19 of this Handbook, written by Dragomir Šarić, contains a review of the theory of the compact solenoid and of recent work on the noncompact solenoid \mathcal{S}_{nc} by Penner & Šarić, as well as work by Bonnot, Penner and Šarić on a cellular action of the mapping class group of \mathcal{S}_{nc} . In analogy with the corresponding situation for punctured surfaces, there is a decorated Teichmüller space of the noncompact solenoid, with associated λ -length coordinates, and a convex hull construction of fundamental domains which gives an interesting combinatorial structure for this Teichmüller space, generalizing an analogous structure that was developed by Penner for the Teichmüller space of a punctured surface. An explicit set of generators for the mapping class group of the noncompact solenoid is also discussed. Note that no such explicit set of generators for the compact solenoid is known. It is conjectured that the mapping class groups of the compact and of the noncompact solenoids are not finitely generated.

Chapter 19 ends with a discussion of open problems on the Teichmüller space and on the mapping class group of the compact and the noncompact solenoids.

Part A

The metric and the analytic theory, 2

Chapter 1

The Weil–Petersson metric geometry

Scott A. Wolpert

Contents

1	Introduction	47
2	Basics of Teichmüller theory	48
3	The WP CAT(0) geometry	52
4	Geodesic-length functions	54
5	WP convexity and curvature	54
6	Approaching degenerate hyperbolic structures	56
7	Metrics and Fenchel–Nielsen coordinates	57
8	WP Alexandrov tangent cone	58
	References	61

1 Introduction

Our goal is to present a summary introduction for the current understanding of a Weil–Petersson metric space. With apologies to colleagues, our goal is to present an exposition following a development of concepts, rather than an exposition following the order of results discovered. Selected readings and general attributions are provided at the end of each section.

There are overlapping themes for the current research on finite dimensional Weil–Petersson metrics. Beginning with the work of Brock and in collaborations, the large-scale coarse geometry is under extensive investigation [5], [9], [10], [11], [12]. Brock also initiated the consideration of the comparison to the geometry of quasi-Fuchsian groups. Beginning with the work of Yamada, the CAT(0) geometry is also under examination [13], [32], [56], [57], [59]. Beginning with the work of Mirzakhani, the relationship to the Witten–Kontsevich conjecture and symplectic reduction are under continuing investigation [38], [39], [46], [47]. The work of Mirzakhani combines explicit integrals and elements of Thurston’s geometry to find the asymptotic count of lengths of simple closed geodesics on a hyperbolic surface [37], [36]. Following the work of Bridgeman and Taylor [7], McMullen has shown that the metric can be reconstructed from dynamical quantities, such as measures on the unit circle and limit sets on the sphere [34]. Beginning with the considerations of Weng [50], [51], the application to an arithmetic Riemann–Roch is being investigated [15]. McMullen [33],

the collaboration of Liu, Sun and Yau [27], [28], as well as Yeung [61], [60] have examined comparisons between the metric and the classical metrics for a domain and have also considered applications. Huang continues a detailed examination of the curvature of the metric [20], [21], [22], [23].

We present a summary introduction of the metric space geometry. Teichmüller space and its augmentation are described in terms of the Fenchel–Nielsen coordinates. The Weil–Petersson metric space basic properties are described. Geodesic-length functions are introduced and formulas for their gradients and Hessians are presented. Beginning applications are considered. A description of the Weil–Petersson metric in Fenchel–Nielsen coordinates is presented. A model metric is discussed. Properties of Weil–Petersson metric and geodesics, as well as properties of geodesic-length functions are combined to describe the Alexandrov tangent cone at points of the augmentation. A comparison dictionary is presented between the geometry of the space of flat tori and Teichmüller space with the Weil–Petersson metric.

2 Basics of Teichmüller theory

Let \mathcal{T} be the Teichmüller space for homotopy marked genus g , n -punctured Riemann surfaces R of negative Euler characteristic. A point of \mathcal{T} is the equivalence class of a pair (R, f) , with f a homeomorphism from a reference topological surface F to R . By uniformization a conformal structure determines a unique complete compatible hyperbolic metric ds^2 for R . The Teichmüller space is a complex manifold of dimension $3g - 3 + n$ with the cotangent space at R represented by $Q(R)$, the space of holomorphic quadratic differentials on R with at most simple poles at the punctures. Weil introduced the Hermitian cometric.

Definition 1. The Weil–Petersson cometric is $\langle \varphi, \psi \rangle = \int_R \varphi \bar{\psi} (ds^2)^{-1}$.

In the 1940s Teichmüller introduced the Finsler metric with conorm given as $\|\varphi\|_{\mathcal{T}} = \int_R |\varphi|$. The Weil–Petersson (WP) dual metric is invariant under the action of the *mapping class group*, MCG, the group of orientation preserving homeomorphisms modulo the subgroup of homeomorphisms homotopic to the identity relative to punctures. The WP metric projects to the quotient $\mathcal{M} = \mathcal{T}/\text{MCG}$, the moduli space of Riemann surfaces. First properties are that the metric is Kähler, non-complete with negative sectional curvature κ with $\sup_{\mathcal{T}} \kappa = 0$ (except for $\dim \mathcal{T} = 1$ where $\sup_{\mathcal{T}} \kappa < 0$) and $\inf_{\mathcal{T}} \kappa = -\infty$. The metric continues to be the primary metric for understanding the Kähler geometry of Teichmüller space [33], [27], [28], [54], [60]. The current exposition focuses on the metric space geometry. *In practice and experience the WP geometry of \mathcal{T} corresponds to the hyperbolic geometry of surfaces.*

A hyperbolic surface has a *thick–thin* decomposition with *thin* the region of injectivity radius below a threshold value. The thin components of a hyperbolic surface are neighborhoods of cusps or are collars (fixed area tubular neighborhoods of short

geodesics). Mumford first observed that the set of hyperbolic surfaces with lengths of closed geodesics bounded below by a constant $c > 0$ forms a compact subset \mathcal{M}_c of the moduli space \mathcal{M} . In general the totality of all thick regions of a given topological type forms a compact set of metric spaces in the Gromov–Hausdorff topology. The Teichmüller and WP metrics are comparable on \mathcal{M}_c . The Teichmüller and WP geometries of the ends of $\mathcal{M} - \mathcal{M}_c$ are examined in the references [35], [29], [56], [57], as well as discussed below.

We consider the following for the reference topological surface F .

Definition 2. A k -simplex of the complex of curves $C(F)$ is a set of $k + 1$ distinct free homotopy classes of non-trivial, non-peripheral, mutually disjoint simple closed curves of F . The pants graph $P(F)$ has vertices the maximal simplices of $C(F)$. Vertices of $P(F)$ are connected by an edge provided the corresponding sets of free homotopy classes differ by replacing a single curve by a curve intersecting the original curve one or two times.

The vertices of $C(F)$ are the free homotopy classes of non-trivial, non-peripheral simple closed curves. The simplices of $C(F)$ are the convex sums of vertices. The complex of curves $C(F)$ is a lattice, partially ordered by inclusion of simplices with maximal simplices, called *pants decompositions*, having dimension $3g - 4 + n$. A pants decomposition decomposes a surface into a union of $2g - 2 + n$ three-holed spheres. The pants graph $P(F)$ becomes a metric space by specifying the edges to have unit-length. Corresponding to a pants decomposition are global coordinates, Fenchel–Nielsen (FN) coordinates for \mathcal{T} given as gluing-parameters for constructing surfaces from right hyperbolic hexagons. The construction begins with right hexagons which can be doubled across alternating edges to obtain a pair of pants, a genus zero hyperbolic surface with three geodesic boundaries with lengths free-parameters in $\mathbb{R}_{>0}$. Boundaries of pants of a common length can be abutted to construct a larger surface.

A pants decomposition for F provides a combinatorial scheme to abut boundaries of pants to obtain a hyperbolic surface of genus g with n punctures. In abutting boundaries there is the free-parameter of the relative displacement of one boundary with respect to the other. Overall for each abutting there are two free-parameters. The first parameter is the common boundary geodesic-length ℓ valued in $\mathbb{R}_{>0}$. The second parameter is the relative displacement τ valued in \mathbb{R} measured in hyperbolic distance (τ is initially measured between appropriate footpoints and then analytically continued).

Theorem 3. *The FN coordinates $\prod_{j \in \mathcal{P}} (\ell_j, \tau_j) \rightarrow (\mathbb{R}_{>0} \times \mathbb{R})^{3g-3+n}$ for a pants decomposition \mathcal{P} provide a real analytic equivalence for \mathcal{T} . The WP Kähler form is $\omega_{\text{WP}} = \frac{1}{2} \sum_j d\ell_j \wedge d\tau_j$.*

The FN coordinate expression for ω_{WP} is independent of the particular choice of pants decomposition.

The *augmented Teichmüller space* $\overline{\mathcal{T}}$ is a partial compactification in the style of Bailey–Borel. The space $\overline{\mathcal{T}}$ is important to the understanding of the Deligne–Mumford compactification of the moduli space \mathcal{M} and for the WP geometry. Frontier spaces are adjoined to \mathcal{T} corresponding to allowing geodesic-lengths ℓ_j to assume the value zero with the FN angles $\theta_j = 2\pi\tau_j/\ell_j$ undefined (in polar coordinates the angle is undefined at the origin). The vanishing $\ell_j = 0$ describes a degenerate hexagon with a side-length vanishing and the adjacent sides meeting at a common point on the circle at infinity for the hyperbolic plane. The vanishing $\ell_j = 0$ corresponds to a degenerate hyperbolic surface with a pair of cusps in place of a simple closed curve (γ_j in the pants decomposition is now represented by the horocycles about the cusps). In general for a simplex $\sigma \subset C(F)$, the σ -null stratum is the space of structures $\mathcal{S}(\sigma) = \{R \text{ degenerate} \mid \ell_\alpha(R) = 0 \text{ iff } \alpha \in \sigma\}$. The frontier spaces $\mathcal{F}_\mathcal{P} = \bigcup_{\sigma \subset \mathcal{P}} \mathcal{S}(\sigma)$ subordinate to a pants decomposition \mathcal{P} are adjoined to \mathcal{T} with a neighborhood basis for $\mathcal{T} \cup \mathcal{F}_\mathcal{P}$ prescribed by the specification that

$$((\ell_\beta, \theta_\beta), \ell_\alpha): \mathcal{T} \cup \mathcal{S}(\sigma) \rightarrow \prod_{\beta \notin \sigma} (\mathbb{R}_{>0} \times \mathbb{R}) \times \prod_{\alpha \in \sigma} (\mathbb{R}_{\geq 0})$$

is a continuous map. For a simplex $\sigma \subset \mathcal{P}$, \mathcal{P}' , contained in two pants decompositions, the neighborhood systems are equivalent.

A structural property is that the deformation spaces $\mathcal{S}(\sigma)$ are products of lower dimensional Teichmüller spaces and the limit of the tangential component of the WP metric of \mathcal{T} is simply the WP metric of the lower dimensional stratum $\mathcal{S}(\sigma)$. A sequence of marked hyperbolic surfaces $\{R_n\}$ converges in $\overline{\mathcal{T}}$ provided there is a simplex σ contained in a pants decomposition \mathcal{P} with $\ell_\alpha(R_n)$, $\alpha \in \sigma$ limiting to zero (no convergence condition is placed on θ_α , $\alpha \in \sigma$) and FN parameters $(\ell_\beta, \theta_\beta)$, $\beta \in \mathcal{P} - \sigma$ converging.

Theorem 4. *The augmented Teichmüller space $\overline{\mathcal{T}} = \mathcal{T} \bigcup_{\sigma \in C(F)} \mathcal{S}(\sigma)$ is a non locally compact stratified space. The augmented Teichmüller space is the WP completion of Teichmüller space.*

A point of $\overline{\mathcal{T}} - \mathcal{T}$ represents a marked degenerate hyperbolic structure for which a simplex of $C(F)$ has each element represented by a pair of cusps. The augmentation is also described as the Chabauty topology closure of the faithful cofinite representations of $\pi_1(R)$ into $\mathrm{PSL}(2; \mathbb{R})$ modulo conjugacies by $\mathrm{PSL}(2; \mathbb{R})$. The quotient $\overline{\mathcal{T}}/\mathrm{MCG}$ is topologically the Deligne–Mumford stable curve compactification of the moduli space of curves.

We make the comparison between the upper half plane \mathbb{H} as the space of marked flat tori and \mathcal{T} as the space of marked hyperbolic structures. The comparison is explored in the following sections. A point $z \in \mathbb{H}$ determines the lattice in \mathbb{C} with basis vectors $\{1, z\}$. A lattice change of basis is given by the action of the elliptic modular group $\mathrm{PSL}(2; \mathbb{Z})$.

The Farey graph \mathcal{G} is realized in $\mathbb{H} \cup \mathbb{Q}$ by connecting vertices (rational numbers) p/q and r/s with a geodesic in \mathbb{H} provided $|ps - qr| = 1$. Brock and Margalit

examine geometric comparisons between the Farey graph and penultimate subsets of $C(F)$ [11] (the sets of decompositions containing a given $3g - 5 + n$ simplex of $C(F)$).

<i>flat structures</i>	<i>hyperbolic structures</i>
\mathbb{H} the space of homotopy marked tori	\mathcal{T} the space of homotopy marked hyperbolic metrics
$\mathrm{PSL}(2; \mathbb{Z})$	mapping class group MCG
$\mathbb{H}/\mathrm{PSL}(2; \mathbb{Z})$ moduli space of flat tori	\mathcal{M} moduli space of Riemann surfaces
$\mathbb{H} \cup \mathbb{Q}$ with horoball topology	$\overline{\mathcal{T}}$ augmented Teichmüller space
$\mathbb{H} \cup \mathbb{Q}/\mathrm{PSL}(2; \mathbb{Z})$ moduli space of stable elliptic curves	$\overline{\mathcal{T}}/\mathrm{MCG}$ Deligne–Mumford moduli space of stable curves
horoballs $\{\Im A(z) \geq c \mid c < 1/2\}$, $A \in \mathrm{PSL}(2; \mathbb{Z})$	Bers regions
Farey graph \mathcal{F}	complex of curves $C(F)$ and pants graph $P(F)$
hyperbolic metric	WP metric (Kähler with negative curvature)
initial tangents to geodesics ending at \mathbb{Q} are dense	initial tangents to geodesics ending at maximally degenerate structures are dense
for $A \in \mathrm{PSL}(2; \mathbb{Z})$, the function $-\log(\Im A(z))$	for a closed geodesic α , the root geodesic-length $\ell_\alpha^{1/2}$
gradient $\mu_A = \mathrm{grad} \log(\Im A(z))$ with $\langle \mu_A, \mu_A \rangle = 1$	gradient $\lambda_\alpha = \mathrm{grad} \ell_\alpha^{1/2}$ with $\langle \lambda_\alpha, \lambda_\alpha \rangle = 1/2\pi + O(\ell_\alpha^3)$
for $A \in \mathrm{PSL}(2; \mathbb{Z})$, hyperbolic metric given as $(d \log(\Im A(z)))^2 + (d \log(\Im A(z)) \circ J)^2$	for a closed geodesic α , WP metric given as $(d \ell_\alpha^{1/2})^2 + (d \ell_\alpha^{1/2} \circ J)^2 + O(\ell_\alpha^3)$
$D_U(-\mu_A) = \langle J \mu_A, U \rangle J \mu_A$	$D_U \lambda_\alpha = 3 \ell_\alpha^{1/2} \langle J \lambda_\alpha, U \rangle J \lambda_\alpha + O(\ell_\alpha^{3/2})$
$\mathrm{Hess}(-\log(\Im A(z))) \geq 0$	$\mathrm{Hess} \ell_\alpha > 0$, $\mathrm{Hess} \ell_\alpha^{1/2} > 0$

The original reference for the WP metric is [3]. A reference for FN coordinates is [2] and for the symplectic form in FN coordinates is [55]. A counterpart approach to coordinates and the symplectic form for surfaces with cusps is extensively investigated in the works of Penner [42], [43] and also in [40]. The complex of curves is introduced in [18] with a current introduction presented in the first volume of this Handbook [17] and its metric space geometry is investigated in the foundational work of Masur and Minsky [30], [31]. The original references for the augmented Teichmüller space are [1], [18] with the description in terms of the Chabauty closure of discrete faithful

representations in [19]. The original reference for analytic consideration of $\bar{\mathcal{M}}$ is [6] with [29] presenting the first expansion for the metric. A brief survey of WP geometry current to the year 2002 is presented in the introduction of [56]. Results from [56] are also presented in the following sections. Current understanding of WP curvature is presented in [20], [22], [23].

3 The WP CAT(0) geometry

The augmented Teichmüller space with the WP metric is a CAT(0) metric space (a complete, simply connected, generalized non-positively curved space). In particular $(\bar{\mathcal{T}}, d_{\text{WP}})$ is a length space, a metric space with unique distance-realizing paths (geodesics) between pairs of points. Furthermore the WP metric has the Euclidean triangle comparison property: for a triangle in $\bar{\mathcal{T}}$ and a corresponding triangle in \mathbb{E}^2 with corresponding edge lengths, chords of the former (located by endpoints on sides) have lengths bounded by lengths of corresponding chords of the latter. The conditions CAT(0) and Gromov hyperbolicity are independent, even though the latter is a generalized negative curvature condition. The strata $\mathfrak{S}(\sigma) \subset \bar{\mathcal{T}}$, $\sigma \in C(F)$ are intrinsic to the WP metric geometry and $\bar{\mathcal{T}}$ is an infinite polyhedron as follows.

Theorem 5. *$\bar{\mathcal{T}}$ is a stratified metric space with each open stratum characterized as the union of all geodesics containing a given point as an interior point. The interior of a geodesic in $\bar{\mathcal{T}}$ is contained in a single stratum (geodesics do not refract at strata). $\bar{\mathcal{T}}$ itself is characterized as the closed convex hull of the maximally degenerate hyperbolic structures (the unions of thrice punctured spheres).*

The structure of strata provides that the extended MCG (both orientation preserving and reversing classes) is the full group of isometries of $\bar{\mathcal{T}}$ as follows [32], [56]. (The property does not follow the comparison between flat structures and hyperbolic structures, since the group of orientation preserving isometries of \mathbb{H} is the Lie group $\text{PSL}(2; \mathbb{R})$). A WP isometry extends to the completion $\bar{\mathcal{T}}$ and preserves the intrinsic strata structure, as well as the partial ordering of inclusion of simplices of $C(F)$. Ivanov established that order preserving bijections of $C(F)$ are induced by elements of the extended MCG [24]. In particular for an orientation preserving isometry there is a corresponding element of MCG and the two maps agree on the maximally degenerate structures in $\bar{\mathcal{T}}$ (a maximally degenerate structure is uniquely determined by its simplex). The two maps agree on $\bar{\mathcal{T}}$, the closed convex hull of the maximally degenerate structures. The isometry coincides with the element of MCG.

There is a classification for the action of elements of MCG. A mapping class ι acting on $\bar{\mathcal{T}}$ has fixed-points or positive translation length realized on a closed convex set \mathcal{A}_ι , isometric to a metric space product $\mathbb{R} \times Y$. In the latter case the isometry acts on $\mathbb{R} \times Y$ as the product of a translation and id_Y . Following Thurston, a mapping class is *irreducible*

(pseudo Anosov) provided no power fixes the free homotopy class of a simple closed curve. Infinite order non-irreducible mapping classes are reducible, and are first analyzed in terms of mappings of proper subsurfaces. Each irreducible mapping class has a unique invariant WP axis and non-commuting irreducible mappings have divergent axes [13], [56].

The pants graph $P(F)$ provides a quasi isometric model for \mathcal{T} and $\overline{\mathcal{T}}$ as follows. Bers observed for constants c' depending on the topological type (g, n) that the sublevel sets $\mathcal{B}_{\mathcal{P}} = \{\ell_{\gamma} \leq c' \text{ for all } \gamma \in \mathcal{P}\}$ called *Bers regions* for \mathcal{P} a pants decomposition, cover Teichmüller space. The bounded valence multivalued relation $R \in \mathcal{T} \leftrightarrow \{\mathcal{P} \mid R \in \mathcal{B}_{\mathcal{P}}\}$ is the basis for (non-unique) embeddings $h: \mathcal{T} \rightarrow P(F)$ and $k: P(F) \rightarrow \mathcal{T}$. Brock established that the embeddings are quasi isometries (satisfy $d(x, y)/c' - c'' \leq d'(f(x), f(y)) \leq c'd(x, y) + c''$ for positive constants). On considering the edges of the pants graph $P(F)$ to have unit-length then Brock and in joint work with Margalit established the following [9], [11].

Theorem 6. *\mathcal{T} and $\overline{\mathcal{T}}$ are quasi isometric to $P(F)$. For the topological types $(1, 2)$ and $(0, 4)$ the spaces \mathcal{T} and $\overline{\mathcal{T}}$ are quasi isometric to the Farey graph.*

Degenerate hyperbolic surfaces given as a union of surfaces of penultimate and ultimate types $(1, 2)$, $(0, 4)$ and $(0, 3)$ are of special interest. For such a simplex $\sigma \in C(F)$ with a total of m subsurfaces of type $(1, 2)$ or $(0, 4)$ the corresponding stratum $\mathfrak{S}(\sigma) \subset \overline{\mathcal{T}}$ is quasi isometric to an m -fold product of Farey graphs and contains a WP isometric image of \mathbb{R}^m (a flat) as a product of geodesics from each of the cited factors. More generally a quasi flat is a quasi isometric embedding of \mathbb{R}^p . Quasi flats are important for understanding a geometry and are a tool for understanding quasi isometric rigidity in the setting of higher-rank symmetric spaces. Behrstock and Minsky settled [5] the open question on rank of \mathcal{T} showing that the maximal dimension of a quasi flat in \mathcal{T} is $\lfloor \frac{3g+n-2}{2} \rfloor$ (the maximal possible count of $(1, 2)$ and $(0, 4)$ subsurfaces). For $\dim \mathcal{T} = 3$ Brock and Masur have shown that any quasi flat is within a bounded distance of a stratum quasi isometric to a product of Farey graphs [11].

In collaborations Behrstock–Minsky [5] and Behrstock–Kleiner–Minsky–Mosher [4] have been investigating the asymptotic cones \mathcal{AC} of MCG (the Gromov–Hausdorff limits of rescalings of the group word-metric). The main result of the first work is that the maximal dimension of a locally-compact subset of \mathcal{AC} equals the maximal dimension of an Abelian subgroup of MCG. In the second work rescaling limits in \mathcal{AC} of flats and quasi flats of MCG, as well as the action on \mathcal{AC} of quasi isometries are examined. Leuzinger shows that the asymptotic cone of the moduli space \mathcal{M} with the Teichmüller metric is bi Lipschitz to the quotient complex $C(F)/\text{MCG}$, [26]. Hamenstädt is also investigating the geometry of MCG [16].

Select readings for the section include the investigations of Brock [9], [10], Daskalopoulos–Wentworth [13], [14] and the author [56].

4 Geodesic-length functions

Associated to each non-trivial, non-peripheral free homotopy class α on a marked hyperbolic surface is the length ℓ_α of the unique geodesic in the free homotopy class. Geodesic-lengths are explicit with $2 \cosh \ell_\alpha/2 = \text{tr } A$ for the free homotopy class α corresponding to the conjugacy class of A in the deck group within $\text{PSL}(2; \mathbb{R})$ and geodesic-lengths can be combined to provide coordinates for $\tilde{\mathcal{T}}$. Geodesic-length functions also have a direct relationship to WP geometry. For a simple closed geodesic α on a hyperbolic surface, the surface can be *cut* along the geodesic to form two circle boundaries, which can then be identified by a relative rotation to form a new surface. A flow on \mathcal{T} is defined by considering families of surfaces $\{R_t\}$ for which at time t reference points from sides of the original geodesic are displaced by t on R_t . The infinitesimal generator, the FN vector field t_α , and the WP gradient of geodesic-length satisfy the basic duality $2t_\alpha = J \text{grad } \ell_\alpha$ for J the almost complex structure of \mathcal{T} .

The relationship between hyperbolic geometry and WP geometry is displayed in the formulas for gradients. The twist-length formula is

$$\langle \text{grad } \ell_\alpha, J \text{grad } \ell_\beta \rangle = 4 \omega_{\text{WP}}(t_\alpha, t_\beta) = -2 \sum_{p \in \alpha \cap \beta} \cos \theta_p$$

for the geodesics α, β and intersection angles θ_* on the hyperbolic surface. The length-length formula for geodesics α, β with corresponding deck transformations A, B , with corresponding axes $\tilde{\alpha}, \tilde{\beta}$ is

$$\langle \text{grad } \ell_\alpha, \text{grad } \ell_\beta \rangle = \frac{2}{\pi} \left(\ell_\alpha \delta_{\alpha\beta} + \sum'_{\langle A \rangle \backslash \Gamma / \langle B \rangle} \left(u \log \frac{u+1}{u-1} - 2 \right) \right)$$

for the Kronecker delta δ_* , where for $C \in \langle A \rangle \backslash \Gamma / \langle B \rangle$ then $u = u(\tilde{\alpha}, C(\tilde{\beta}))$ is the cosine of the intersection angle if the lifts $\tilde{\alpha}$ and $C(\tilde{\beta})$ intersect and is otherwise $\cosh d(\tilde{\alpha}, C(\tilde{\beta}))$; for $\alpha = \beta$ the double-coset of the identity is omitted from the sum. In the latter case the values $d(\tilde{\alpha}, C(\tilde{\beta}))$ are the lengths of the shortest representatives on the surface for the homotopy classes connecting α to β with homotopy relative to α and relative to β . The above formulas are the result of explicit integration of the Weil–Petersson product.

Select readings for the section are the investigations [39], [45], [52], [53].

5 WP convexity and curvature

The geodesic-length functions give rise to WP convex functions (functions strictly convex along geodesics). The length of a geodesic, the square root of length, the total length of a measured geodesic lamination, as well as for a surface with cusps the distance between unit-length horocycles are all WP strictly convex functions. In

particular the geodesic-length sublevel sets, as well as the strata of $\overline{\mathcal{T}}$ are convex sets. Although a simple counterpart to the above gradient formulas is not yet available, there are bounds and expansions for the Riemannian Hessian (the intrinsic second derivative) of geodesic-length.

Expansions for small geodesic-lengths for quantities on \mathcal{T} provide important information. The primary interest are quantities given as integrals on Riemann surfaces. The WP metric, WP curvature, as well as the gradients and Hessians of geodesic-length are examples. The approach for an expansion is based on understanding the integrand on the thick and thin regions of the surface. Commonly the leading term of the expansion is the contribution of the collar zeroth rotational mode of the integrand with all other contributions higher order. The expansion for the gradient and Hessian for small geodesic-lengths are examples.

Theorem 7. *The variations of geodesic-length satisfy*

$$\langle \text{grad } \ell_\alpha^{1/2}, \text{grad } \ell_\beta^{1/2} \rangle - \delta_{\alpha\beta}/2\pi$$

is positive and bounded by $O(\ell_\alpha^{3/2} \ell_\beta^{3/2})$ and

$$2\ell_\alpha \text{Hess } \ell_\alpha[U, U] - \dot{\ell}_\alpha^2[U] - 3\dot{\ell}_\alpha^2[JU]$$

is positive and bounded by $O(\ell_\alpha^3 \|U\|_{\text{WP}}^2)$ for a tangent vector U where for c_0 positive the remainder term constants are uniform for $\ell_\alpha, \ell_\beta \leq c_0$.

The Hessian is directly related to covariant differentiation by $\text{Hess } h[U, V] = \langle D_U \text{grad } h, V \rangle$ for a smooth function h and vector fields U, V .

Corollary 8. *The WP connection D is described for bounded geodesic-length, root gradient $\lambda_\alpha = \text{grad } \ell_\alpha^{1/2}$, and a tangent vector U by*

$$D_U \lambda_\alpha = 3\ell_\alpha^{-1/2} \langle J\lambda_\alpha, U \rangle J\lambda_\alpha + O(\ell_\alpha^{3/2} \|U\|_{\text{WP}}).$$

A property of small geodesic-lengths and WP geodesics follows. For a geodesic $\gamma(t)$ with tangent field $\frac{d}{dt}$ the quantity $f(t) = \langle \lambda_\alpha, \frac{d}{dt} \rangle^2 + \langle J\lambda_\alpha, \frac{d}{dt} \rangle^2$ has vanishing principal term for its first derivative. The quantity $f(t)$ is Lipschitz along $\gamma(t)$ with constant $O(\ell_\alpha^{3/2})$.

The estimates for WP sectional curvatures are also examples of small geodesic-length expansions. The WP curvature of the span $\{\text{grad } \ell_\alpha, J \text{grad } \ell_\alpha\}$ is $O(-\ell_\alpha^{-1})$. Similarly for a pair of deformations approximately supported on different components of thick the corresponding curvature is $O(-\ell_{\text{sys}})$ for ℓ_{sys} the smallest geodesic-length. For a pair of deformations approximately supported in the same component of thick the corresponding curvature is approximately the curvature for the limiting 2-plane tangent to a stratum of $\overline{\mathcal{T}}$.

The section is based on the work [57] on behavior of geodesic-length and the works [20], [22], [23] on the curvature of the metric.

6 Approaching degenerate hyperbolic structures

A refined description of geodesic-length functions near degenerate hyperbolic structures provides further understanding of the WP metric. We consider hyperbolic structures near a proper stratum $\mathfrak{S}(\sigma)$, $\sigma \in C(F)$. The closure of $\mathfrak{S}(\sigma)$ in $\overline{\mathcal{T}}$ is a union of strata $\bigcup_{\tau, \sigma \subseteq \tau} \mathfrak{S}(\tau)$. A closed convex subset $\overline{\mathfrak{S}(\sigma)}$ of a CAT(0) space is the base of an orthogonal projection $\Pi_{\overline{\mathfrak{S}(\sigma)}}$ and we also consider the distance $d_{\overline{\mathfrak{S}(\sigma)}}$ to the stratum. The projection is distance non-increasing with fibers fibered by geodesics. For the k -simplex $\sigma = \{\alpha_1, \dots, \alpha_{k+1}\}$ there is the overall bound on the distance of $R \in \overline{\mathcal{T}}$ to $\overline{\mathfrak{S}(\sigma)}$

$$d_{\overline{\mathfrak{S}(\sigma)}}(R) \leq (2\pi(\ell_{\alpha_1}(R) + \dots + \ell_{\alpha_{k+1}}(R)))^{1/2}$$

a consequence of the root geodesic-length convexity and the gradient pairing expansion in Theorem 7. The bound displays the incompleteness of the metric. The inequality compares to the formal equation for the hyperbolic plane $d(z, \infty) = -\log \Im z$ with the difference between the logarithm and the square root demonstrating the difference between the complete hyperbolic metric and the incomplete WP metric. With the covering property of Bers regions it follows that for a constant depending only on topological type, each point of \mathcal{T} is within a fixed distance of a maximally degenerate structure.

There is an approximation of a long WP geodesic segment \widehat{pq} as follows. At the ending point q introduce a geodesic to a maximally degenerate structure r and introduce a third geodesic from the beginning point p to the maximally degenerate structure r . The triangle Δpqr has two long sides and a bounded-length side. The comparison Euclidean triangle has a small angle at its beginning point. A consequence of CAT(0) is that the corresponding angle $\angle qpr$ in \mathcal{T} is likewise bounded, the desired approximation. In particular for sufficiently long geodesics the corresponding angles are sufficiently small. An immediate consequence is that geodesics ending at maximally degenerate structures are dense in the space of geodesics.

A description of geodesics ending at a point of $\overline{\mathcal{T}} - \mathcal{T}$ is available. A geodesic is *projecting* to $\overline{\mathfrak{S}(\sigma)}$ provided its projection is a point or equivalently it is length-minimizing to $\overline{\mathfrak{S}(\sigma)}$. Projecting geodesics are almost described as integral curves of a constant sum of root gradient length functions. In particular for a unit-speed projecting geodesic ζ to $\mathfrak{S}(\sigma)$ and the root gradients $\lambda_j = \text{grad } \ell_{\alpha_j}$ there are constants a_j such that the tangent field to the geodesic ζ satisfies

$$\frac{d}{dt} = (2\pi)^{-1} \sum_{j=1}^{k+1} a_j \lambda_j + O(t^4)$$

with $(2\pi)^{1/2} \|a_j\|_{\text{Euclid}} = 1$ and the distance satisfies

$$d_{\overline{\mathfrak{S}(\sigma)}} = \left(2\pi \sum_{j=1}^{k+1} \ell_{\alpha_j} \right)^{1/2} + O\left(\sum_{j=1}^{k+1} \ell_{\alpha_j}^{5/2} \right).$$

The distance formula prefigures the approximation that the WP metric in a neighborhood of $\overline{\mathcal{B}(\sigma)}$ compares to the product of the (lower dimensional) WP metric on $\overline{\mathcal{B}(\sigma)}$ and a universal metric for the normal bundle. The approximation is discussed in the next section.

Selected readings for the section are [10], [56], [57]. The basic reference for CAT(0) geometry is [8].

7 Metrics and Fenchel–Nielsen coordinates

Fenchel–Nielsen coordinates provide a straightforward description of hyperbolic surfaces and a parameterization for Teichmüller space. The WP metric does not provide the structure of a symmetric space (the full isometry group is discrete); the metric is not expected to have an elementary closed-form expression in FN coordinates. Expansions and comparisons for the metric provide an alternative to an elementary expression.

The FN twist-length coordinates $(\ell_j, \tau_j)_{j=1}^{3g-3+n}$ for assembling hyperbolic pants provide global coordinates for \mathcal{T} with the WP Kähler form

$$\omega_{\text{WP}} = \frac{1}{2} \sum d\ell_j \wedge d\tau_j$$

and on the Bers region $\{\ell_j < c'\}$ the metric comparisons

$$\langle \cdot, \cdot \rangle \asymp \sum (d\ell_j^{1/2})^2 + (d\ell_j^{1/2} \circ J)^2 \asymp \sum \text{Hess } \ell_j$$

for J the almost complex structure with uniform comparability (given $c' > 0$ there exist constants c_1, c_2 such that the metric is bounded above and below on the Bers region in terms of the constants and the given expressions). In a neighborhood of the maximally degenerate structure $\{\ell_j = 0 \mid j = 1, \dots, 3g - 3\}$ the WP metric has the expansions

$$\begin{aligned} \langle \cdot, \cdot \rangle &= 2\pi \sum (d\ell_j^{1/2})^2 + (d\ell_j^{1/2} \circ J)^2 + o\left(\sum \ell_j^3 \langle \cdot, \cdot \rangle\right) \\ &= \frac{\pi}{6} \sum \frac{\text{Hess } \ell_j^2}{\ell_j} + o\left(\sum \ell_j^2 \langle \cdot, \cdot \rangle\right). \end{aligned}$$

There are corresponding expansions for the neighborhood of a general stratum. Metric incompleteness is immediate.

There is a comparison between the Teichmüller and WP geometry for the ends of the moduli space. On a Bers region the Teichmüller norm is comparable as follows

$$\| \cdot \|_T^2 \asymp \sum ((d\ell_j)^2 + (d\ell_j \circ J)^2) \ell_j^{-2}.$$

Each metric is comparable to a product of model metrics for the tangent planes $\bigcap_{k \neq j} \ker d\ell_k \cap \ker d\ell_k \circ J$. The model for the Teichmüller metric is the hyperbolic metric itself. We now examine the model for the WP metric.

There is a relation between the 1-form $d\ell_j \circ J$ and the FN angle θ_j . The definition of the angle $\theta_j = 2\pi \tau_j / \ell_j$ requires a pants decomposition. An alternative definition of an angle is given by starting with the FN vector field $t_j = 1/2 J \text{grad } \ell_j$, considering the WP dual, and defining the FN gauge 1-form $\rho_j = 2\pi(\ell_j^{3/2} \langle \lambda_j, \lambda_j \rangle)^{-1} \langle \cdot, J\lambda \rangle$ for $\lambda_j = \text{grad } \ell_j^{1/2}$. The FN gauge is determined without the choice of a pants decomposition and satisfies the essential property $\rho_j(t_j) = d\theta_j(t_j) = 2\pi/\ell_j$. Gauges and angles agree on the level sets of pants length $(\ell_j)_{j=1}^{3g-3}$.

The WP metric for geodesic-lengths $\ell_j = 2\pi^2 r_j^2$ and FN gauges ρ_j is comparable as follows

$$\langle \cdot, \cdot \rangle \asymp \sum 4dr_j^2 + r_j^6 \varrho_j^2$$

with the expansion

$$\langle \cdot, \cdot \rangle = \pi^3 \sum 4dr_j^2 + r_j^6 \varrho_j^2 + O\left(\sum \ell_j^3 \langle \cdot, \cdot \rangle\right)$$

at the maximally degenerate structure. There are corresponding expansions for the neighborhood of a general stratum. The general expansion is in terms of a product of model metrics and the WP metric of the stratum.

The model metric $4dr^2 + r^6 d\vartheta^2$ has Kähler form $2r^3 dr d\theta$, Riemannian connection D characterized by

$$D_{\frac{\partial}{\partial r}} \frac{\partial}{\partial r} = 0, \quad D_{\frac{\partial}{\partial \vartheta}} \frac{\partial}{\partial r} = D_{\frac{\partial}{\partial r}} \frac{\partial}{\partial \vartheta} = \frac{3}{r} \frac{\partial}{\partial \vartheta} \quad \text{and} \quad D_{\frac{\partial}{\partial \vartheta}} \frac{\partial}{\partial \vartheta} = \frac{-3}{4} r^5 \frac{\partial}{\partial r}$$

and Riemannian curvature $-3/2r^2$. The correspondence between the model and WP metrics is for $\frac{\partial}{\partial r_j}$ corresponding to $2^{3/2}\pi^2 \lambda_j$ and $\frac{\partial}{\partial \vartheta_j}$ corresponding to the FN angle variation $(2\pi)^{-1} \ell_j t_j$. The WP metric, Kähler form and connection have leading terms exactly corresponding to the expressions for the model metric. The WP curvature is comparable to the corresponding model curvature. In effect the correspondence of metrics is a C^1 approximation with a bounded C^2 comparison.

Readings for the section are [57], [58] with the comparability of the Teichmüller metric presented in [33] and a refined large scale comparison presented in [35].

8 WP Alexandrov tangent cone

In a CAT(0) metric space there is a well-defined angle between a pair of geodesics from a common initial point. The angle enters in the definition of the tangent cone and in the first variation formula for distance. At a point of the Teichmüller space \mathcal{T} the WP Alexandrov angle is given in terms of the Riemannian metric. At a point of a stratum $\mathcal{T}(\sigma) \subset \overline{\mathcal{T}} - \mathcal{T}$, $\sigma \in C(F)$, the WP Alexandrov tangent cone AC is isometric

to a product of a Euclidean orthant and the tangent space $\mathbf{T}\mathcal{T}(\sigma)$ with the WP metric. The dimension of the Euclidean orthant is the count $|\sigma|$ of geodesic-lengths trivial on $\mathcal{T}(\sigma)$.

A triple of points (p, q, r) in $\overline{\mathcal{T}}$ has Euclidean comparison triangle with angle $\angle(p, q, r)$ valued in the interval $[0, \pi]$ determined by the Law of Cosines

$$2d(p, q)d(p, r)\cos\angle(p, q, r) = d(p, q)^2 + d(p, r)^2 - d(q, r)^2.$$

The Alexandrov angle $(p, q, r) \mapsto \angle(p, q, r)$ is upper semi continuous. For constant speed geodesics $\gamma_0(t), \gamma_1(t)$ with common initial point (from the CAT(0) inequality) the comparison angle for $(\gamma_0(0), \gamma_0(t), \gamma_1(t'))$ is a non-decreasing function of t and t' . The Alexandrov angle is defined by the limit

$$\cos\angle(\gamma_0, \gamma_1) = \lim_{t \rightarrow 0} \frac{d(p, \gamma_0(t))^2 + d(p, \gamma_1(t))^2 - d(\gamma_0(t), \gamma_1(t))^2}{2d(p, \gamma_0(t))d(p, \gamma_1(t))}.$$

Geodesics at zero angle are said to define the *same direction*. At zero angle provides an equivalence relation on the geodesics beginning at a point p with the Alexandrov angle providing a metric on the space of directions. The Alexandrov tangent cone \mathbf{AC}_p is the set of constant speed geodesics beginning at p modulo the equivalence relation of same speed and at zero angle.

A relative length basis for a point p of $\mathcal{T}(\sigma)$ is a collection τ of vertices of $C(F)$ disjoint from the elements of σ such that at p the gradients $\{\text{grad } \ell_\beta\}_{\beta \in \tau}$ span the tangent space $\mathbf{T}\mathcal{T}(\sigma)$. A relative length basis can be given as the union of a partition and a dual partition for $R - \sigma$.

We describe for the augmentation point p an isometry between the Alexandrov tangent cone \mathbf{AC}_p and the product $\mathbb{R}_{\geq 0}^{|\sigma|} \times \mathbf{T}_p\mathcal{T}(\sigma)$ with the first factor the Euclidean orthant and the second factor the stratum tangent space with WP metric. The mapping for a geodesic $\gamma(t)$ terminating at p is given by associating for the lengths $\mathcal{L}(\gamma(t)) = (\ell_\alpha^{1/2}, \ell_\beta^{1/2})_{\alpha \in \sigma, \beta \in \tau}(\gamma(t))$ the initial one-sided derivatives

$$\Lambda: \gamma \mapsto (2\pi)^{1/2} \frac{d\mathcal{L}(\gamma)}{dt}(0)$$

(convexity provides for existence of the initial derivatives). By hypothesis the tuple $(\ell_\beta^{1/2})_{\beta \in \tau}$ provides local coordinates at p for the stratum $\mathcal{T}(\sigma)$ and therefore $(2\pi)^{1/2} \left(\frac{d\ell_\beta^{1/2}(\gamma)}{dt}(0) \right)_{\beta \in \tau}$ defines a vector in the tangent space $\mathbf{T}_p\mathcal{T}(\sigma)$ with WP inner product. The positive orthant $\mathbb{R}_{\geq 0}^{|\sigma|} \subset \mathbb{R}^{|\sigma|}$ is considered with the Euclidean inner product. The Alexandrov tangent cone is given the structure of a cone in an inner product space through the formal relation $\langle \gamma_0, \gamma_1 \rangle = \|\gamma'_0\| \|\gamma'_1\| \cos\angle(\gamma_0, \gamma_1)$.

Theorem 9. *The mapping Λ from the WP Alexandrov tangent cone \mathbf{AC}_p to $\mathbb{R}_{\geq 0}^{|\sigma|} \times \mathbf{T}_p\mathcal{T}(\sigma)$ is an isometry of cones with restrictions of inner products. A WP terminating geodesic γ with a root geodesic-length function initial derivative $\frac{d\ell_\alpha^{1/2}(\gamma)}{dt}(0)$*

vanishing is contained entirely in the stratum $\{\ell_\alpha = 0\}$. Geodesics γ_0 and γ_1 at zero Alexandrov angle have comparison angles $\angle(p, \gamma_0(t), \gamma_1(t))$ bounded as $O(t)$.

A property for non-positively curved Riemannian manifolds is that the exponential map is distance non-decreasing. An inverse exponential map $\exp_p^{-1}: \overline{\mathcal{T}} \rightarrow \text{AC}_p$ is defined by associating to $q \in \overline{\mathcal{T}}$ the unique geodesic connecting p to q with speed $d(p, q)$. The map is not an injection since geodesics at zero angle with common speed are mapped to a common element of AC_p . The map is distance non-increasing as follows. From the CAT(0) inequality and definition of the Alexandrov angle, points on geodesics beginning at p have distance satisfying

$$d(\gamma_0(t), \gamma_1(t))^2 \geq d(p, \gamma_0(t))^2 + d(p, \gamma_1(t))^2 - 2d(p, \gamma_0(t))d(p, \gamma_1(t))\cos\angle(\gamma_0, \gamma_1).$$

For equality for a single value the Flat Triangle Lemma provides that the geodesics are contained in a flat subspace of $\overline{\mathcal{T}}$. The flat subspaces are classified.

A second application of the Alexandrov angle is the first variation formula for distance. For the unit-speed geodesic $\gamma(t)$ the distance $d(\gamma(t), q)$ to a point not on the geodesic is convex with initial one-sided derivative satisfying

$$\frac{d}{dt}d(\gamma(t), q)(0) = -\cos\angle(\gamma, \gamma_{pq})$$

for γ_{pq} the geodesic connecting p to q . *Non refraction* of geodesics on $\overline{\mathcal{T}}$ is a consequence: a WP length minimizing path at most changes strata at its endpoints. Consider a pair of unit-speed geodesics $\gamma_0(t), \gamma_1(t)$ with initial point p such that the reverse path along γ_0 followed by γ_1 is length minimizing. The Alexandrov angle between the tangents at p is π . The distance $d(\gamma_0(t), \gamma_1(t))$ is at least that of the path from $\gamma_0(t)$ to p to $\gamma_1(t)$ and thus $\lim_{t \rightarrow 0} d(\gamma_0(t), \gamma_1(t))/2t = 1$ and the angle is π . Elements of AC_p at angle π necessarily lie in the subspace $\mathbf{T}_p\overline{\mathcal{T}}(\sigma)$ and from the theorem are segments of a single geodesic contained in $\overline{\mathcal{T}}(\sigma)$.

A third application is for length-minimizing paths connecting an initial and terminal point and intersecting a prescribed stratum. Consider a pair of geodesics γ_0 and γ_1 each with initial point p on $\overline{\mathcal{T}}(\sigma)$, γ_0 with endpoint q and γ_1 with endpoint r . Consider that the concatenation $\gamma_0 + \gamma_1$ is a length-minimizing path connecting q and r to a point of $\overline{\mathcal{T}}(\sigma)$. A geodesic κ beginning at p contained in $\overline{\mathcal{T}}(\sigma)$ provides a variation of the configuration. The initial derivative of the distance $d(q, p) + d(r, p)$ along κ is $-\cos\angle(\gamma_0, \kappa) - \cos\angle(\gamma_1, \kappa)$. The geodesics beginning at p contained in $\overline{\mathcal{T}}(\sigma)$ fill out the Alexandrov tangent cone $\text{AC}_p(\overline{\mathcal{T}}(\sigma))$. It follows that the sum in AC_p of the initial tangents of γ_0 and γ_1 has vanishing projection onto the subcone $\text{AC}_p(\overline{\mathcal{T}}(\sigma))$, the desired property.

A further application is for combinatorial harmonic maps. Certain groups acting on Euclidean buildings and group extensions acting on Cayley graphs satisfying a Poincaré type inequality for links of points will have a global fixed point for an action on $\overline{\mathcal{T}}$.

The main reading for the section is [57]. Additional readings are [13], [14], [56]. The basic reference for Alexandrov angles is [8]. The readings for combinatorial harmonic maps are [25], [48], [49].

References

- [1] W. Abikoff, Degenerating families of Riemann surfaces. *Ann. of Math.* (2) 105 (1) (1977), 29–44. [51](#)
- [2] W. Abikoff, *The real analytic theory of Teichmüller space*. Springer-Verlag, Berlin 1980. [51](#)
- [3] L. V. Ahlfors, Some remarks on Teichmüller’s space of Riemann surfaces. *Ann. of Math.* (2) 74 (1961), 171–191. [51](#)
- [4] J. A. Behrstock, B. Kleiner, Y. Minsky, and L. Mosher, Geometry and rigidity of mapping class groups. Preprint, 2008. [53](#)
- [5] J. Behrstock and Y. Minsky, Dimension and rank for mapping class groups. *Ann. of Math.* (2) 167 (3) (2008), 1055–1077. [47](#), [53](#)
- [6] L. Bers, Spaces of degenerating Riemann surfaces. In *Discontinuous groups and Riemann surfaces* (Proc. Conf., Univ. Maryland, College Park, Md., 1973), Ann. of Math. Stud. 79, Princeton University Press, Princeton, N.J., 1974, 43–55. [52](#)
- [7] M. Bridgeman and E. C. Taylor, An extension of the Weil-Petersson metric to the quasi-Fuchsian space. *Math. Ann.* 341 (4) (2008), 927–943. [47](#)
- [8] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*. Springer-Verlag, Berlin 1999. [57](#), [61](#)
- [9] J. F. Brock, The Weil-Petersson metric and volumes of 3-dimensional hyperbolic convex cores. *J. Amer. Math. Soc.* 16 (3) (2003), 495–535. [47](#), [53](#)
- [10] J. F. Brock, The Weil-Petersson visual sphere. *Geom. Dedicata* 115 (2005), 1–18. [47](#), [53](#), [57](#)
- [11] J. Brock and D. Margalit, Weil-Petersson isometries via the pants complex. *Proc. Amer. Math. Soc.* 135 (3) (2007), 795–803. [47](#), [51](#), [53](#)
- [12] J. Brock and H. Masur, Coarse and synthetic Weil-Petersson geometry: quasi-flats, geodesics, and relative hyperbolicity. Preprint, 2007. [47](#)
- [13] G. Daskalopoulos and R. Wentworth, Classification of Weil-Petersson isometries. *Amer. J. Math.* 125 (4) (2003), 941–975. [47](#), [53](#), [61](#)
- [14] G. Daskalopoulos and R. Wentworth, Harmonic maps and Teichmüller theory. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 33–109. [53](#), [61](#)
- [15] G. Freixas i Montplet, An arithmetic Riemann-Roch theorem for pointed stable curves. Preprint, 2007. [47](#)
- [16] U. Hamenstädt, Geometry of the mapping class groups III: Geometric rank. Preprint, 2006. [53](#)
- [17] U. Hamenstädt, Geometry of the complex of curves and of Teichmüller space. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich

- 2007, 447–467. [51](#)
- [18] W. J. Harvey, Boundary structure of the modular group. In *Riemann surfaces and related topics* (Proceedings of the 1978 Stony Brook Conference), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 245–251. [51](#)
- [19] W. Harvey, Chabauty spaces of discrete groups. In *Discontinuous groups and Riemann surfaces* (Proc. Conf., Univ. Maryland, College Park, Md., 1973), Ann. of Math. Stud. 79, Princeton University Press, Princeton, N.J., 1974, 239–246. [52](#)
- [20] Z. Huang, Asymptotic flatness of the Weil-Petersson metric on Teichmüller space. *Geom. Dedicata* 110 (2005), 81–102. [48](#), [52](#), [55](#)
- [21] Z. Huang, Average curvatures of Weil-Petersson geodesics in Teichmüller space. Preprint, 2007. [48](#)
- [22] Z. Huang, The Weil-Petersson geometry on the thick part of the moduli space of Riemann surfaces. *Proc. Amer. Math. Soc.* 135 (10) (2007), 3309–3316. [48](#), [52](#), [55](#)
- [23] Z. Huang, On asymptotic Weil-Petersson geometry of Teichmüller Space of Riemann surfaces. *Asian J. Math.* 11 (3) (2007), 459–484. [48](#), [52](#), [55](#)
- [24] N. V. Ivanov, Automorphisms of complexes of curves and of Teichmüller spaces. In *Progress in knot theory and related topics*, Travaux en Cours 56, Hermann, Paris 1997, 113–120. [52](#)
- [25] H. Izeki and S. Nayatani, Combinatorial harmonic maps and discrete-group actions on Hadamard spaces. *Geom. Dedicata* 114 (2005), 147–188. [61](#)
- [26] E. Leuzinger, Reduction theory for mapping class groups and applications to moduli spaces. Preprint, 2008. [53](#)
- [27] K. Liu, X. Sun, and S.-T. Yau, Canonical metrics on the moduli space of Riemann surfaces. I. *J. Differential Geom.* 68 (3) (2004), 571–637. [48](#)
- [28] K. Liu, X. Sun, and S.-T. Yau, Canonical metrics on the moduli space of Riemann surfaces. II. *J. Differential Geom.* 69 (1) (2005), 163–216. [48](#)
- [29] H. Masur, Extension of the Weil-Petersson metric to the boundary of Teichmüller space. *Duke Math. J.* 43 (3) (1976), 623–635. [49](#), [52](#)
- [30] H. A. Masur and Y. N. Minsky, Geometry of the complex of curves. I. Hyperbolicity. *Invent. Math.* 138 (1) (1999), 103–149. [51](#)
- [31] H. A. Masur and Y. N. Minsky, Geometry of the complex of curves. II. Hierarchical structure. *Geom. Funct. Anal.* 10 (4) (2000), 902–974. [51](#)
- [32] H. Masur and M. Wolf, The Weil-Petersson isometry group. *Geom. Dedicata* 93 (2002), 177–190. [47](#), [52](#)
- [33] C. T. McMullen, The moduli space of Riemann surfaces is Kähler hyperbolic. *Ann. of Math.* (2) 151 (1) (2000), 327–357. [47](#), [48](#), [58](#)
- [34] C. T. McMullen, Thermodynamics, dimension and the Weil-Petersson metric. *Invent. Math.* 173 (2008), 365–425. [47](#)
- [35] Y. N. Minsky, Extremal length estimates and product regions in Teichmüller space. *Duke Math. J.* 83 (2) (1996), 249–286. [49](#), [58](#)
- [36] M. Mirzakhani, Growth of the number of simple closed geodesics on hyperbolic surfaces. *Ann. of Math.* (2) 168 (1) (2008), 97–125. [47](#)

- [37] M. Mirzakhani, Simple geodesics and Weil-Petersson volumes of moduli spaces of bordered Riemann surfaces. *Invent. Math.* 167 (1) (2007), 179–222. [47](#)
- [38] M. Mirzakhani, Weil-Petersson volumes and intersection theory on the moduli space of curves. *J. Amer. Math. Soc.* 20 (1) (2007), 1–23. [47](#)
- [39] G. Mondello, Triangulated Riemann surfaces with boundary and the Weil-Petersson Poisson structure. *J. Differential Geom.* 81 (2009), 391–436. [47](#), [54](#)
- [40] G. Mondello, Riemann surfaces, ribbon graphs and combinatorial classes. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 151–215. [51](#)
- [41] K. Obitsu and S. A. Wolpert, Grafting hyperbolic metrics and Eisenstein series. *Math. Ann.* 341 (3) (2008), 685–706.
- [42] R. C. Penner, The decorated Teichmüller space of punctured surfaces. *Comm. Math. Phys.* 113 (2) (1987), 299–339. [51](#)
- [43] R. C. Penner, Cell decomposition and compactification of Riemann’s moduli space in decorated Teichmüller theory. In *Woods Hole mathematics*, Ser. Knots Everything 34, World Sci. Publ., Hackensack, N.J., 2004, 263–301. [51](#)
- [44] M. Pollicott, H. Weiss, and S. A. Wolpert, Topological dynamics of the Weil-Petersson geodesic flow. Preprint, 2007; arXiv:math/0711.3221.
- [45] G. Riera, A formula for the Weil-Petersson product of quadratic differentials. *J. Anal. Math.* 95 (2005), 105–120. [54](#)
- [46] B. Safnuk, Integration on moduli spaces of stable curves through localization. Preprint, 2007. [47](#)
- [47] B. Safnuk and M. Mulase, Mirzakhani’s recursion relations, Virasoro constraints and the KdV hierarchy. *Indian J. Math.* 50 (1) (2008), 189–218. [47](#)
- [48] M.-T. Wang, A fixed point theorem of discrete group actions on Riemannian manifolds. *J. Differential Geom.* 50 (2) (1998), 249–267. [61](#)
- [49] M.-T. Wang, Generalized harmonic maps and representations of discrete groups. *Comm. Anal. Geom.* 8 (3) (2000), 545–563. [61](#)
- [50] L. Weng, ω -admissible theory. *Proc. London Math. Soc.* (3) 79 (3) (1999), 481–510. [47](#)
- [51] L. Weng, Ω -admissible theory. II. Deligne pairings over moduli spaces of punctured Riemann surfaces. *Math. Ann.* 320 (2) (2001), 239–283. [47](#)
- [52] S. A. Wolpert, The Fenchel-Nielsen deformation. *Ann. of Math.* (2) 115 (3) (1982), 501–528. [54](#)
- [53] S. A. Wolpert, On the symplectic geometry of deformations of a hyperbolic surface. *Ann. of Math.* (2) 117 (2) (1983), 207–234. [54](#)
- [54] S. A. Wolpert, On obtaining a positive line bundle from the Weil-Petersson class. *Amer. J. Math.*, 107 (6) (1985), 1485–1507. [48](#)
- [55] S. A. Wolpert, On the Weil-Petersson geometry of the moduli space of curves. *Amer. J. Math.* 107 (4) (1985), 969–997. [51](#)
- [56] S. A. Wolpert, Geometry of the Weil-Petersson completion of Teichmüller space. In *Surveys in Differential Geometry VIII: Papers in Honor of Calabi, Lawson, Siu and Uhlenbeck*, Intl. Press, Cambridge, MA, 2003, 357–393. [47](#), [49](#), [52](#), [53](#), [57](#), [61](#)

- [57] S. A. Wolpert, Behavior of geodesic-length functions on Teichmüller space. *J. Differential Geom.* 79 (2) (2008), 277–334. [47](#), [49](#), [55](#), [57](#), [58](#), [61](#)
- [58] S. A. Wolpert, Extension of the Weil-Petersson connection. Preprint, 2007; arXiv: math/0709.2513. [58](#)
- [59] S. Yamada, On the geometry of Weil-Petersson completion of Teichmüller spaces. *Math. Res. Lett.* 11 (2–3) (2004), 327–344. [47](#)
- [60] S.-K. Yeung, Bounded smooth strictly plurisubharmonic exhaustion functions on Teichmüller spaces. *Math. Res. Lett.* 10 (2–3) (2003), 391–400. [48](#)
- [61] S.-K. Yeung, Quasi-isometry of metrics on Teichmüller spaces. *Int. Math. Res. Not.* (4) (2005), 239–255. [48](#)

Chapter 2

Infinite dimensional Teichmüller spaces

Alastair Fletcher and Vladimir Markovic

Contents

1	Introduction	65
2	Quasiconformal mappings and Teichmüller spaces	66
	2.1 Quasiconformality	66
	2.2 Teichmüller space	68
	2.3 Teichmüller metric	69
	2.4 Schwarzian derivatives and quadratic differentials	69
	2.5 Bers embedding and complex structure on Teichmüller space	72
3	Biholomorphic maps between Teichmüller spaces	74
	3.1 Kobayashi metric	74
	3.2 The infinitesimal Teichmüller metric	77
	3.3 Isometries of Bergman spaces	77
4	Local rigidity of Teichmüller spaces	81
	4.1 Projections on Banach spaces	82
	4.2 Bergman kernels and projections on $\mathcal{L}^1(M)$	84
	4.3 Isomorphisms of Bergman spaces	86
	4.4 Local bi-Lipschitz equivalence of Teichmüller spaces	88
5	Open problems	89
	References	90

1 Introduction

In this chapter, we consider some analytic properties of Teichmüller spaces, in particular those of infinite dimension. The Bers embedding maps the Teichmüller space $\mathcal{T}(M)$ of a hyperbolic Riemann surface M biholomorphically onto a subset of a Banach space $\mathcal{Q}(M)$ of holomorphic quadratic differentials. If M is of non-exceptional analytic type, then the dimension of $\mathcal{T}(M)$ is finite if and only if the dimension of $\mathcal{Q}(M)$ is finite if and only if M is of finite analytic type, that is, M is a compact Riemann surface of genus g with n punctures, where $2g + n \geq 5$. In this case $\mathcal{Q}(M)$ is a finite dimensional vector space over \mathbb{C} and is therefore reflexive.

Via the Bers embedding, it can be shown that the cotangent space of $\mathcal{T}(M)$ at the base-point $[0]$ can be identified with the Bergman space $\mathcal{A}^1(M)$, the pre-dual of $\mathcal{Q}(M)$.

In the finite dimensional case, $\mathcal{A}^1(M) = \mathcal{Q}(M)$. However, when M is of infinite analytic type, it is no longer true that $\mathcal{A}^1(M)$ is reflexive, and so $\mathcal{A}^1(M) \neq \mathcal{Q}(M)$.

The problem of classifying biholomorphic maps between Teichmüller spaces can be reduced, via consideration of the Kobayashi and Teichmüller metrics, to a problem of classifying isometries between the cotangent spaces of the corresponding Riemann surfaces. The infinite dimensional case requires more machinery because the Bergman spaces of Riemann surfaces of infinite analytic type are not reflexive.

We will show that if there is a surjective \mathbb{C} -linear isometry between $\mathcal{A}^1(M)$ and $\mathcal{A}^1(N)$, then the Riemann surfaces M and N , assumed to be of non-exceptional type, are conformally equivalent. Note that we do not assume that the Riemann surfaces M and N are even homeomorphic. This result implies that every biholomorphic map between Teichmüller spaces $\mathcal{T}(M)$ and $\mathcal{T}(N)$ is induced by a quasiconformal mapping between M and N , and therefore the automorphism group of $\mathcal{T}(M)$ is equal to the mapping class group of M .

We also prove a counterpoint to the above result on isometries of Bergman spaces. Namely, if M and N are any two Riemann surfaces of infinite analytic type, then the corresponding Bergman spaces are isomorphic. This then implies that the Teichmüller spaces of any two Riemann surfaces of infinite analytic type are locally bi-Lipschitz equivalent.

The chapter ends with some open problems that have arisen as a result of work in this area.

2 Quasiconformal mappings and Teichmüller spaces

2.1 Quasiconformality

A map $g : (a, b) \rightarrow \mathbb{C}$ is absolutely continuous on the interval (a, b) if

$$g(x) = \int_a^x h(t) dt + g(a)$$

for $x \in (a, b)$ and $h \in \mathcal{L}_{\text{loc}}^1(a, b)$, the space of locally integrable functions. If g is absolutely continuous, then it is differentiable almost everywhere and $g' = h$ almost everywhere.

Let Ω be a plane domain, $f : \Omega \rightarrow f(\Omega) \subset \mathbb{C}$ and let a rectangle $R \subset \Omega$ have sides parallel to the x and y axes. We say that f is absolutely continuous on lines (ACL) on R if f is absolutely continuous on almost every horizontal and vertical line in R . The map f is ACL on Ω if f is ACL on every rectangle $R \subset \Omega$.

Definition 2.1. A homeomorphism $f : \Omega \rightarrow f(\Omega)$ is K -quasiconformal if and only if the following holds:

- (i) f is ACL on Ω ,
- (ii) $|f_{\bar{z}}| \leq k|f_z|$ almost everywhere in Ω , where $k = (K - 1)/(K + 1)$.

By property (ii), $J_f = |f_z|^2 - |f_{\bar{z}}|^2 \geq (1 - k)|f_z|^2 > 0$ and so a quasiconformal map is orientation preserving. In particular, f is conformal if and only if f is 1-quasiconformal. There are equivalent definitions of quasiconformality, see [7].

Example. The ACL condition is certainly necessary, as we will show here. Let C be the Cantor set on $(0, 1)$ so that every $x \in C$ can be written as

$$x = \sum_{i=1}^{\infty} 2 \cdot 3^{-ni}$$

for some subset $\{n_i\}$ of \mathbb{Z}^+ . The Cantor function $F: (0, 1) \rightarrow (0, 1)$ is defined by setting

$$F(x) = \sum_{i=1}^{\infty} 2^{-ni}$$

for $x \in C$, and extended to the whole of $(0, 1)$ by requiring that F be monotonically increasing. Since F is constant on connected components of the complement of C , F is differentiable almost everywhere, with derivative 0, but is not differentiable at points of C . Now define the function $f: (0, 1) \times (-\infty, \infty) \rightarrow \mathbb{C}$ given by

$$f(x + iy) = F(x) + x + iy.$$

Now, $F(x) + x$ is a homeomorphism of $(0, 1)$ onto $(0, 2)$, and so f is a homeomorphism, which is differentiable almost everywhere, and in fact $f_{\bar{z}} = 0$ almost everywhere. However f cannot be conformal on $(0, 1) \times (-\infty, \infty)$ because it is not differentiable at any point of $C \times (-\infty, \infty)$. Moreover, f cannot be quasiconformal, because if it was, then it would have to be 1-quasiconformal and hence conformal. In conclusion, if $f: \Omega \rightarrow f(\Omega)$ is a homeomorphism, differentiable almost everywhere and $f_{\bar{z}} = 0$ almost everywhere, then this does not imply that f is conformal or quasiconformal.

Quasiconformality can also be defined for maps between Riemann surfaces, and note that in this chapter we are assuming that all our Riemann surfaces are hyperbolic, that is, they have the unit disk \mathbb{D} as the universal cover. The map $f: M \rightarrow N$ is said to be K -quasiconformal at $p \in M$ if there are coordinate charts (U_p, π_p) on M and $(U_{f(p)}, \pi_{f(p)})$ on N such that $\pi_{f(p)} \circ f \circ \pi_p^{-1}$ is a K -quasiconformal mapping whose domain is the plane domain $\pi_p(U_p)$. The mapping f is then said to be K -quasiconformal if it is K -quasiconformal at all $p \in M$. Note that this definition is independent of the choice of coordinate chart, since the transition maps are conformal.

Let μ be a measurable $(-1, 1)$ -form on a Riemann surface M with

$$|\mu(p)| \leq k < 1$$

for almost all $p \in M$, so $\mu \in B(M)$, the open unit ball of $\mathcal{L}^\infty(M)$. Such a μ is called a Beltrami differential. The Beltrami differential equation is

$$f_{\bar{z}} = \mu f_z.$$

The solution f (sometimes denoted f^μ) of the Beltrami differential equation is a quasiconformal mapping, and all quasiconformal mappings arise in this way, giving a correspondence between quasiconformal mappings and Beltrami differentials. The solution f^μ can be lifted to give a quasiconformal self-mapping of \mathbb{D} . If the lifted solution is normalized to fix three points of $\partial\mathbb{D}$, then the correspondence between quasiconformal mappings and Beltrami differentials is one-to-one. See [7] for the proof of these statements.

2.2 Teichmüller space

Let \mathcal{F} be the family of all quasiconformal self-mappings of \mathbb{D} which are normalised so that their extensions to $\partial\mathbb{D}$ fix $1, -1, i$. From the solution of the Beltrami differential equation, there is a one-to-one correspondence between \mathcal{F} and the open unit ball B of $\mathcal{L}^\infty(\mathbb{D})$. Now, we can put an equivalence relation $\sim_{\mathcal{F}}$ on elements of \mathcal{F} by declaring that $f_1 \sim_{\mathcal{F}} f_2$ if and only if the extensions of f_1 and f_2 to $\partial\mathbb{D}$ agree on $\partial\mathbb{D}$. Equivalently, two elements μ_1 and μ_2 of B are related by $\sim_{\mathcal{F}}$ if and only if the extensions of the normal solutions f^{μ_1} and f^{μ_2} to $\partial\mathbb{D}$ agree on $\partial\mathbb{D}$.

Definition 2.2. The set of equivalence classes of \mathcal{F} under $\sim_{\mathcal{F}}$ is called the universal Teichmüller space $\mathcal{T}(\mathbb{D})$.

The deformation space $\text{Def}(M)$ of a Riemann surface M is the set of pairs (N, f) where N is a Riemann surface, and $f: M \rightarrow N$ is a quasiconformal map. For any hyperbolic plane domain Ω with its appropriate Riemann map f , we have $(\Omega, f) \in \text{Def}(\mathbb{D})$. An equivalence relation \sim can be defined on $\text{Def}(M)$ by requiring that $(N_1, f_1) \sim (N_2, f_2)$ if and only if

$$f_1 \circ f_2^{-1}: N_2 \rightarrow N_1$$

is homotopic to a conformal map $g: N_2 \rightarrow N_1$. Two maps f and g between hyperbolic Riemann surfaces are homotopic if they can be lifted to mappings of \mathbb{D} which agree on $\partial\mathbb{D}$.

Definition 2.3. The Teichmüller space $\mathcal{T}(M)$ of a Riemann surface M is given by

$$\text{Def}(M) / \sim .$$

The base-point of $\mathcal{T}(M)$ is the Teichmüller class of the identity mapping. This definition agrees with the definition of universal Teichmüller space. To see this, first note that since all the quasiconformal images of \mathbb{D} are conformally equivalent, only the normalized quasiconformal self-mappings f^μ of \mathbb{D} need be considered. Then $f^{\mu_2} \circ (f^{\mu_1})^{-1}$ is homotopic to a conformal map if and only if $f^{\mu_2} \circ (f^{\mu_1})^{-1}$ is the identity mapping on $\partial\mathbb{D}$. Therefore f^{μ_1} agrees with f^{μ_2} on $\partial\mathbb{D}$, which is precisely the definition for f^{μ_1} and f^{μ_2} to determine the same point of universal Teichmüller space.

As in the considerations for the universal Teichmüller space, $\mathcal{T}(M)$ can be considered as a space of Beltrami differentials under the corresponding equivalence relation. The base-point of $\mathcal{T}(M)$ is the Teichmüller class of $0 \in B(M)$. Given a Beltrami differential μ on a Riemann surface M , μ lifts to a Beltrami differential $\tilde{\mu}$ on \mathbb{D} which satisfies

$$\tilde{\mu} = (\tilde{\mu} \circ g) \frac{\bar{g}'}{g'}$$

for every g in the covering group of M . We write f_μ for the quasiconformal mapping of the plane which has the complex dilatation $\tilde{\mu}$ on \mathbb{D} , and 0 on \mathbb{D}^* , where $\mathbb{D}^* = \mathbb{C} \setminus \mathbb{D}$. Note that \mathbb{D}^* is the image of \mathbb{D} under the reflection in $\partial\mathbb{D}$ given by $z \mapsto 1/\bar{z}$ for $z \in \mathbb{D}$.

Theorem 2.4. *The following are equivalent:*

- (i) *The Beltrami differentials μ and ν on the Riemann surface M are equivalent under \sim .*
- (ii) $f^\mu|_{\partial\mathbb{D}} = f^\nu|_{\partial\mathbb{D}}$.
- (iii) $f_\mu|_{\mathbb{D}^*} = f_\nu|_{\mathbb{D}^*}$.

2.3 Teichmüller metric

For $[f_0], [g_0] \in \mathcal{T}(M)$, we can define the Teichmüller distance

$$d_{\mathcal{T}}([f_0], [g_0]) = \frac{1}{2} \inf \log K_{g \circ f^{-1}} \quad (2.1)$$

where the infimum is taken over all maps f and g in the Teichmüller classes of f_0 and g_0 respectively, and where K_f is the maximal dilatation of f . We write $d_{\mathcal{T}}$ instead of $d_{\mathcal{T}(M)}$ for brevity where it is clear which Teichmüller space $[f]$ and $[g]$ are in. We can replace \inf by \min in (2.1) and $d_{\mathcal{T}}$ is in fact a metric. For the details, we refer to [7].

The Teichmüller metric on $B(M)$ is given by

$$d_B(\mu, \nu) = \frac{1}{2} \log \frac{1 + \|(\mu - \nu)/(1 - \bar{\mu}\nu)\|_\infty}{1 - \|(\mu - \nu)/(1 - \bar{\mu}\nu)\|_\infty},$$

for $\mu, \nu \in B(M)$. The Teichmüller metric $d_{\mathcal{T}}$ can be expressed as the quotient of d_B ,

$$d_{\mathcal{T}}([\mu_0], [\nu_0]) = \frac{1}{2} \inf_{\mu \in [\mu_0], \nu \in [\nu_0]} \log \frac{1 + \|(\mu - \nu)/(1 - \bar{\mu}\nu)\|_\infty}{1 - \|(\mu - \nu)/(1 - \bar{\mu}\nu)\|_\infty},$$

for $[\mu_0], [\nu_0] \in \mathcal{T}(M)$.

2.4 Schwarzian derivatives and quadratic differentials

If f is holomorphic in a domain Ω and $f'(z) \neq 0$ in Ω , then the Schwarzian derivative of f is

$$S_f = \left(\frac{f''}{f'} \right)' - \frac{1}{2} \left(\frac{f''}{f'} \right)^2. \quad (2.2)$$

If also $f(z) \neq 0$ in Ω , then a direct computation shows

$$S_f(z) = S_{1/f}(z)$$

and this formula shows how to define the Schwarzian derivative of a meromorphic function at simple poles. Thus the Schwarzian derivative can be defined for locally injective meromorphic functions, and S_f is itself holomorphic. Let

$$A(z) = \frac{az + b}{cz + d}$$

be a Möbius transformation, then differentiating gives

$$\frac{A''(z)}{A'(z)} = -\frac{2c}{cz + d}$$

and

$$\left(\frac{A''(z)}{A'(z)}\right)' = \frac{2c^2}{(cz + d)^2}$$

from which we see that

$$S_A = 0.$$

Conversely, starting with the equation $S_f = 0$ and setting $g = f''/f'$, then (2.2) gives $g' = g^2/2$. Solving this differential equation shows that every solution of $S_f = 0$ is a Möbius transformation. Schwarzian derivatives satisfy the composition rule

$$S_{f \circ g} = (S_f \circ g)g'^2 + S_g \quad (2.3)$$

and so if g is a Möbius transformation and $S_g = 0$ then

$$S_{f \circ g} = (S_f \circ g)g'^2. \quad (2.4)$$

On the other hand, if f is a Möbius transformation, then

$$S_{f \circ g} = S_g. \quad (2.5)$$

To define the Schwarzian derivative at ∞ , assume that f is locally injective and meromorphic in a neighbourhood of ∞ , then $h(z) = f(1/z)$ is defined in a neighbourhood of 0. Using (2.4),

$$z^4 S_h(z) = S_f(1/z)$$

and so we can define

$$S_f(\infty) = \lim_{z \rightarrow 0} z^4 S_h(z)$$

and S_f is holomorphic at ∞ . Thus the Schwarzian derivative can be defined for a locally injective meromorphic function f on any domain Ω . The following theorem shows that the Schwarzian derivative can be prescribed.

Theorem 2.5. *Let g be a holomorphic function in a simply connected domain Ω . Then there is a meromorphic function f in Ω such that*

$$S_f = g$$

which is unique up to an arbitrary Möbius transformation.

See [7] for the proof. Let $\mathcal{Q}(M)$ be the space of holomorphic quadratic differentials on a Riemann surface M equipped with the norm

$$\|\tilde{\varphi}\|_{\mathcal{Q}(M)} = \sup_{p \in M} \rho_M^{-2}(p) |\tilde{\varphi}(p)|,$$

where ρ_M is the hyperbolic density on M , and noting that $\rho_M^{-2}|\tilde{\varphi}|$ is a function on M , whereas $|\tilde{\varphi}|$ is usually not. We will call $\mathcal{Q}(M)$ the Bers space, and $\|\cdot\|_{\mathcal{Q}(M)}$ the Bers norm. It is straightforward to see that $\mathcal{Q}(M)$ is a Banach space.

Now let f_1 and f_2 be meromorphic functions on a domain Ω and also let $h: \Omega' \rightarrow \Omega$ be conformal. Using the invariance property (2.3),

$$S_{f_1 \circ h} - S_{f_2 \circ h} = (S_{f_1} \circ h - S_{f_2} \circ h)h^2.$$

Since the hyperbolic density is conformally invariant, $\rho_{\Omega'} = (\rho_{\Omega} \circ h)|h'|$, then by writing $w = h(z)$, we have the invariance formula

$$\frac{|S_{f_1}(w) - S_{f_2}(w)|}{\rho_{\Omega}^2(w)} = \frac{|S_{f_1 \circ h}(z) - S_{f_2 \circ h}(z)|}{\rho_{\Omega'}^2(z)}.$$

In terms of the Bers norm, this is

$$\|S_{f_1} - S_{f_2}\|_{\mathcal{Q}(\Omega)} = \|S_{f_1 \circ h} - S_{f_2 \circ h}\|_{\mathcal{Q}(\Omega')}.$$

If $f_2 = h^{-1}$ is a conformal mapping of Ω , then

$$\|S_{f_1} - S_{f_2}\|_{\mathcal{Q}(\Omega)} = \|S_{f_1 \circ f_2^{-1}}\|_{\mathcal{Q}(f_2(\Omega))}.$$

In the special case of f_1 being the identity,

$$\|S_{f_2}\|_{\mathcal{Q}(\Omega)} = \|S_{f_2^{-1}}\|_{\mathcal{Q}(f_2(\Omega))}.$$

Lastly, if $f_2 = h^{-1}$ is a Möbius transformation, then

$$\|S_{f_1}\|_{\mathcal{Q}(\Omega)} = \|S_{f_1 \circ f_2^{-1}}\|_{\mathcal{Q}(f_2(\Omega))},$$

which shows that $\|S_f\|_{\mathcal{Q}(\Omega)}$ is completely invariant with respect to Möbius transformations.

Recalling the quasiconformal mappings f^μ and f_μ , we have that $f_\mu|_{\mathbb{D}^*}$ is a conformal map, and the set of such maps characterizes $\mathcal{T}(M)$. If $M \simeq \mathbb{D}/G$, then for every $g \in G$, $f_\mu \circ g \circ f_\mu^{-1}$ is a Möbius transformation. Therefore, using the transformation rules (2.4) and (2.5), we have

$$S_{f_\mu|_{\mathbb{D}^*}} = S_{(f_\mu \circ g \circ f_\mu^{-1}) \circ f_\mu|_{\mathbb{D}^*}} = S_{f_\mu \circ g|_{\mathbb{D}^*}} = (S_{f_\mu|_{\mathbb{D}^*}} \circ g)g^2.$$

This shows that the Schwarzian derivative is a quadratic differential for the group G acting on \mathbb{D}^* , and its projection is a holomorphic quadratic differential on \mathbb{D}^*/G , which is the mirror image of the Riemann surface M , denoted by M^* .

2.5 Bers embedding and complex structure on Teichmüller space

We have the mapping

$$\mu \mapsto S_{f_\mu|_{\mathbb{D}^*}}$$

which maps the open unit ball $B(M)$ of $\mathcal{L}^\infty(M)$ into the space of quadratic differentials $\mathcal{Q}(M^*)$. This induces a mapping

$$\lambda_M: \mathcal{T}(M) \rightarrow \mathcal{Q}(M^*), \quad (2.6)$$

which is called the Bers embedding. The first thing to note is that if two holomorphic functions f_1 and f_2 on \mathbb{D}^* have the same Schwarzian derivative, then by Theorem 2.5, one is equal to the other post-composed by a Möbius transformation. Now since both functions are normalized at 3 points, then f_1 and f_2 must be identical on \mathbb{D}^* , and therefore determine the same Teichmüller class. This shows that the mapping λ_M is one-to-one onto its image. Further, the image of $\mathcal{T}(M)$ under λ_M is open in $\mathcal{Q}(M^*)$, the proof of which can be found in [7].

In fact, $\mu \mapsto S_{f_\mu|_{\mathbb{D}^*}}$ is a holomorphic map from $B(M)$ into $\mathcal{Q}(M^*)$. For a given $\mu \in B(M)$, write

$$G^\mu = \{f^\mu \circ g \circ (f^\mu)^{-1} : g \in G\},$$

where $M \simeq \mathbb{D}/G$, and $M^\mu \simeq \mathbb{D}/G^\mu$ is a Riemann surface which is quasiconformally equivalent to M . Let $\tilde{\alpha}_\mu: B(M) \rightarrow B(M^\mu)$ be the mapping given by

$$f^{\tilde{\alpha}_\mu(v)} = f^v \circ (f^\mu)^{-1}$$

or, by writing out in full,

$$\tilde{\alpha}_\mu(v) = \left(\frac{v - \mu}{1 - \bar{\mu}v} \left(\frac{f_z^\mu}{|f_z^\mu|} \right)^2 \right) \circ (f^\mu)^{-1}.$$

The function $\tilde{\alpha}_\mu$ maps $B(M)$ bijectively onto $B(M^\mu)$, and it follows that $\tilde{\alpha}_\mu$ is holomorphic. Note that the induced mapping α_μ , given by $\alpha_\mu([v]) = [\tilde{\alpha}_\mu(v)]$ is a bijective isometry of $\mathcal{T}(M)$ onto $\mathcal{T}(M^\mu)$. For $\mu \in B(M)$ and $v \in B(M^\mu)$ we write

$$\Lambda_\mu(v) = S_{f_v|_{\mathbb{D}^*}}. \quad (2.7)$$

This mapping of $B(M^\mu)$ into $\mathcal{Q}((M^\mu)^*)$ is holomorphic. Now, in the ball

$$B_\mu(0, 1/2) = \{\varphi \in \mathcal{Q}((M^\mu)^*) : \|\varphi\|_{\mathcal{Q}} < 2\},$$

the mapping (2.7) has a section $\sigma_\mu: B_\mu(0, 2) \rightarrow B(M^\mu)$. We see that σ_μ is holomorphic. Let π be the canonical projection of $B(M)$ onto $\mathcal{T}(M)$, and let λ and λ_μ be the Bers embeddings of $\mathcal{T}(M)$ and $\mathcal{T}(M^\mu)$, respectively, into $\mathcal{Q}(M^*)$ and $\mathcal{Q}((M^\mu)^*)$.

The collection

$$\{V_\mu = (\pi \circ (\tilde{\alpha}_\mu)^{-1} \circ \sigma_\mu)(B_\mu(0, 1/2)) : \mu \in B(M)\}$$

is an open covering of $\mathcal{T}(M)$. Indeed, V_μ is the pre-image of $B_\mu(0, 1/2)$ under the homeomorphism $h_\mu = \lambda_\mu \circ \alpha_\mu$ of $\mathcal{T}(M)$ onto $\mathcal{Q}((M^\mu)^*)$.

Theorem 2.6. *The atlas*

$$\{(V_\mu, h_\mu) : \mu \in B(M)\} \quad (2.8)$$

defines a complex structure on Teichmüller space $\mathcal{T}(M)$. The Bers embedding $[\mu] \rightarrow S_{f_\mu|_{\mathbb{D}^*}}$ of $\mathcal{T}(M)$ into $\mathcal{Q}(M^*)$ is holomorphic with respect to this structure.

Proof. Choose two elements $\mu_1, \mu_2 \in B(M)$ such that $V_{\mu_1} \cap V_{\mu_2}$ is non-empty. In $h_{\mu_1}(V_{\mu_1} \cap V_{\mu_2})$, we have

$$h_{\mu_2} \circ h_{\mu_1}^{-1} = \Lambda_{\mu_2} \circ \tilde{\alpha}_{\mu_2} \circ (\tilde{\alpha}_{\mu_1})^{-1} \circ \sigma_{\mu_1}.$$

We know that all the mappings on the right hand side of this equation are holomorphic, and so $h_{\mu_2} \circ h_{\mu_1}^{-1}$ is holomorphic. Switching μ_1 and μ_2 in this calculation shows that $h_{\mu_2} \circ h_{\mu_1}^{-1}$ is biholomorphic, and so (2.8) defines a complex structure for $\mathcal{T}(M)$. To show that the Bers embedding is holomorphic, we have to show that $\lambda \circ h_\mu^{-1}$ is holomorphic in $B_\mu(0, 1/2)$. Now,

$$\lambda \circ h_\mu^{-1} = \Lambda \circ (\tilde{\alpha}_\mu)^{-1} \circ \sigma_\mu$$

and since all the mappings on the right hand side are holomorphic, then $\lambda \circ h_\mu^{-1}$ must also be holomorphic. \square

A Riemann surface M is said to be of finite analytic type if it is a compact Riemann surface of genus g with a finite number n of punctures. It has exceptional type if $2g + n < 5$. All non-hyperbolic Riemann surfaces have exceptional type.

If M is of finite analytic type, then $\mathcal{Q}(M)$ can be identified with its pre-dual space $\mathcal{A}^1(M)$, the subset of $\mathcal{L}^1(M)$ consisting of holomorphic quadratic differentials on M with finite norm

$$\|\varphi\|_1 = \int_M |\varphi|,$$

for $\varphi \in \mathcal{A}^1(M)$. The Banach space $\mathcal{A}^1(M)$ is called the Bergman space. In fact, every linear functional on $\mathcal{A}^1(M)$, $L : \mathcal{A}^1(M) \rightarrow \mathbb{C}$ has the form

$$L(\varphi) = \int_M \rho_M^{-2} \bar{\psi} \varphi$$

for some $\psi \in \mathcal{Q}(M)$, and $L \equiv 0$ if and only if $\psi \equiv 0$. Let \tilde{M} be a compact Riemann surface of genus $g \geq 0$ and let E be a finite, possibly empty, subset of \tilde{M} which contains exactly $n \geq 0$ points. We assume that $2g + n \geq 5$, so that the Riemann surface $M = \tilde{M} \setminus E$ has non-exceptional finite type. Each $\varphi \in \mathcal{A}^1(M)$ can be

regarded as a quadratic differential on \tilde{M} which is holomorphic except for isolated singularities at the points of M . The integrability of φ implies that the singularities of φ are either removable or simple poles, so $\mathcal{A}^1(M)$ is the space of meromorphic quadratic differentials on \tilde{M} whose poles, if any, are simple and belong to E .

Proposition 2.7. *Let $M = \tilde{M} \setminus E$ be as above. Then the dimension of $\mathcal{A}^1(M)$ is $3g - 3 + n \geq 2$. If $x \in E$, then some $\varphi \in \mathcal{A}^1(M)$ has a pole at x .*

The Riemann surface M is of infinite analytic type if M has infinite genus or an infinite number of punctures. If M is of infinite analytic type, then the dimension of $\mathcal{A}^1(M)$ is infinite. For example, if M has an infinite number of punctures at the points (z_n) for $n = 1, 2, \dots$, then there exist functions f_n , each of which have a simple pole at z_n , and which are linearly independent. This points us in the direction of the following result, proved in [9].

Theorem 2.8. *A Riemann surface M is of finite analytic type if and only if the dimension of $\mathcal{A}^1(M)$ is finite.*

If the dimension of $\mathcal{A}^1(M)$ is infinite, then $\mathcal{A}^1(M)$ is a proper subset of $\mathcal{Q}(M)$. Therefore, via the Bers embedding, the dimension of the Teichmüller space $\mathcal{T}(M)$ is infinite if and only if M is not of finite analytic type.

3 Biholomorphic maps between Teichmüller spaces

In this section, we will classify biholomorphic maps between Teichmüller spaces by reducing the problem to the cotangent space. That is, all surjective linear isometries between Bergman spaces of Riemann surfaces of non-exceptional type are geometric, which in particular implies that the two Riemann surfaces are conformally related.

3.1 Kobayashi metric

Let X be any connected complex Banach manifold, and let $H(\mathbb{D}, X)$ be the set of all holomorphic maps from \mathbb{D} into X . The Kobayashi function $\delta_X : X \times X \rightarrow [0, +\infty]$ is

$$\delta_X(x, y) = \inf \{ \rho_{\mathbb{D}}(0, t) : f(0) = x, f(t) = y \text{ for some } f \in H(\mathbb{D}, X) \},$$

provided the set of such maps is non-empty, and $+\infty$ otherwise. If X and Y are connected complex Banach manifolds and $f : X \rightarrow Y$ is holomorphic, then

$$\delta_Y(f(x_1), f(x_2)) \leq \delta_X(x_1, x_2),$$

for all $x_1, x_2 \in X$, and with equality if f is biholomorphic.

Definition 3.1. The Kobayashi pseudo-metric σ_X on X is the largest pseudo-metric on X such that

$$\sigma_X(x, y) \leq \delta_X(x, y), \quad (3.1)$$

for all $x, y \in X$. If δ_X is a metric, as it will be in our cases, then σ_X is a metric and σ_X and δ_X are equal.

Note that when we say σ_X is the largest pseudo-metric, we mean that if d_X is any other pseudo-metric which satisfies (3.1), then

$$d_X(x, y) \leq \sigma_X(x, y)$$

for all $x, y \in X$. The Kobayashi pseudo-metric is the largest metric under which holomorphic mappings are distance decreasing.

Example 1. If X is \mathbb{C} or $\hat{\mathbb{C}}$, then the linear maps $f_n(z) = nz$, together with their inverses, from \mathbb{D} into X show that $\delta_X \equiv 0$. Hence $\sigma_X \equiv 0$. Furthermore, if X is a torus or the punctured plane $\mathbb{C} \setminus \{0\}$, then there is a covering map π from \mathbb{C} onto X . Since π is a contraction in the respective pseudo-metrics, $\sigma_X \equiv 0$.

Example 2. Let X be a hyperbolic Riemann surface, that is, it has \mathbb{D} as its universal cover. If π is the covering map from \mathbb{D} onto X , then $\pi \in H(\mathbb{D}, X)$ and so is a contraction in the corresponding Kobayashi pseudo-metrics. However, every $f \in H(\mathbb{D}, X)$ lifts to a map $\tilde{f} \in H(\mathbb{D}, \mathbb{D})$ such that $f = \pi \circ \tilde{f}$. Thus δ_X is equal to the quotient pseudo-metric on X with respect to the covering map π and the hyperbolic metric on \mathbb{D} . That is, δ_X coincides with the hyperbolic metric on X . Since δ_X is a pseudo-metric, $\sigma_X = \delta_X$ and so the Kobayashi pseudo-metric is equal to the hyperbolic metric on X .

Example 3. Let B be the unit ball in a complex Banach manifold X and pick $x \in B$. The linear function $f(t) = tx/\|x\|$ maps the unit disk \mathbb{D} into the unit ball B and maps $\|x\|$ to x and 0 to 0. Therefore

$$\sigma_B(0, x) \leq \sigma_{\mathbb{D}}(0, \|x\|).$$

However, via the Hahn–Banach Theorem, there exists a continuous linear functional L on X such that $L(x) = \|x\|$ and $\|L\| = 1$. Thus L maps B into the unit disk \mathbb{D} and so

$$\sigma_{\mathbb{D}}(0, \|x\|) \leq \sigma_B(0, x).$$

The definition of the hyperbolic metric on \mathbb{D} leads to

$$\sigma_B(0, x) = \frac{1}{2} \tanh^{-1} \|x\|.$$

Proposition 3.2. For all $\mu, \nu \in B(M)$, we have

$$\delta_{B(M)}(\mu, \nu) = d_{B(M)}(\mu, \nu).$$

Proof. We can assume that $\mu \neq \nu$. First we also assume that $\mu = 0$. Suppose that $f \in H(\mathbb{D}, B(M))$, $f(0) = 0$ and $f(t) = \nu$ for some $t \in \mathbb{D}$. By the Schwarz Lemma,

$$|t| \geq \|f(t)\|_\infty = \|\nu\|_\infty.$$

Taking the infimum over all such f ,

$$\delta_{B(M)}(0, \nu) \geq \rho_{\mathbb{D}}(0, \|\nu\|_\infty) = d_{B(M)}(0, \nu).$$

Choose the function $f(t) = t\nu/\|\nu\|_\infty$ and observe that

$$\delta_{B(M)}(0, \nu) \leq \rho_{\mathbb{D}}(0, \|\nu\|_\infty) = d_{B(M)}(0, \nu)$$

since $f(\|\nu\|_\infty) = \nu$. Now if $\mu \neq 0$, observe that the function $f: B(M) \rightarrow \mathcal{L}^\infty(M)$ defined by

$$f(\lambda) = \frac{\mu - \lambda}{1 - \bar{\mu}\lambda}$$

is a holomorphic map of $B(M)$ onto itself. Therefore

$$\delta_{B(M)}(\mu, \nu) = \delta_{B(M)}(f(\mu), f(\nu)) = d_{B(M)}(0, f(\nu)) = d_{B(M)}(\mu, \nu). \quad \square$$

As we have seen, the Teichmüller metric on $B(M)$ induces a quotient metric on $\mathcal{T}(M)$:

$$d_{\mathcal{T}}([\mu], [\nu]) = \inf \{d_{B(M)}(\mu_0, \nu_0) : \mu_0, \nu_0 \in B(M), \mu_0 \in [\mu], \nu_0 \in [\nu]\},$$

for all $[\mu], [\nu] \in \mathcal{T}(M)$. We can now prove Royden's theorem on the equality of Teichmüller and Kobayashi metrics on $\mathcal{T}(M)$, proved in [16].

Theorem 3.3. *The Teichmüller and Kobayashi metrics on $\mathcal{T}(M)$ coincide, that is,*

$$d_{\mathcal{T}}([\mu], [\nu]) = \delta_{\mathcal{T}(M)}([\mu], [\nu]).$$

Proof. Fix $[\mu], [\nu] \in \mathcal{T}(M)$. We have

$$\delta_{\mathcal{T}(M)}([\mu], [\nu]) \leq \inf \{d_{B(M)}(\mu_0, \nu_0) : \mu_0 \in [\mu], \nu_0 \in [\nu]\}.$$

Hence by Proposition 3.2,

$$\delta_{\mathcal{T}(M)}([\mu], [\nu]) \leq d_{\mathcal{T}}([\mu], [\nu]).$$

For the opposite inequality, choose $f \in H(\mathbb{D}, \mathcal{T}(M))$ such that $f(0) = [\mu]$ and $f(t) = [\nu]$ for some $t \in \mathbb{D}$. Using a theorem of Ślodkowski (see [4]), we can write $\pi \circ g = f$ with

$$g \in H(\mathbb{D}, B(M)).$$

Using Proposition 3.2,

$$\rho_{\mathbb{D}}(0, t) \geq d_{B(M)}(g(0), g(t)) \geq d_{\mathcal{T}}(\pi(g(0)), \pi(g(t))) = d_{\mathcal{T}}([\mu], [\nu]).$$

Taking the infimum over all such f , we obtain

$$\delta_{\mathcal{T}(M)}([\mu], [\nu]) \geq d_{\mathcal{T}}([\mu], [\nu]). \quad \square$$

Corollary 3.4. *Let M and N be two hyperbolic Riemann surfaces. Then every biholomorphic map between $\mathcal{T}(M)$ and $\mathcal{T}(N)$ preserves Teichmüller distances.*

3.2 The infinitesimal Teichmüller metric

Due to Corollary 3.4, investigating biholomorphic self-mappings of $\mathcal{T}(M)$ reduces to the study of biholomorphic Teichmüller isometries. Now, biholomorphic Teichmüller isometries preserve the infinitesimal Teichmüller metric, as we will see below.

Recall the Bers embedding (2.6) of $\mathcal{T}(M)$ into $\mathcal{Q}(M^*)$, where M^* is the mirror image of M . For the rest of this chapter, we will for brevity (and to avoid confusion with our notation for Banach duals) write $\mathcal{Q}(M)$ instead of $\mathcal{Q}(M^*)$ and $\mathcal{A}^1(M)$ for $\mathcal{A}^1(M^*)$, but bear in mind that the Bers embedding maps $\mathcal{T}(M)$ onto a subset of the Bers space of the mirror image of M . Via the Bers embedding, we can regard $\mathcal{Q}(M)$ as the tangent space to $\mathcal{T}(M)$ at its base-point $[0]$ (where $0 \in B(M)$). Further, we can define an isomorphism θ of $\mathcal{Q}(M)$ onto the Banach dual $(\mathcal{A}^1(M))^*$ of $\mathcal{A}^1(M)$ via

$$\theta(\varphi)(f) = \int_M \rho_M^{-2} \varphi(\bar{z}) f(z) dx dy,$$

for $\varphi \in \mathcal{Q}(M)$ and $f \in \mathcal{A}^1(M)$ and where ρ_M is the hyperbolic density on M . This was proved by Bers in [1]. We can therefore identify $(\mathcal{A}^1(M))^*$ with the tangent space to $\mathcal{T}(M)$ at its base-point. This identifies the cotangent space with $\mathcal{A}^1(M)$ in the finite dimensional case, since then $\mathcal{A}^1(M)$ is reflexive. Further, the standard norm

$$\|L\| = \sup\{|L(f)| : f \in \mathcal{A}^1(M), \|f\|_1 \leq 1\},$$

for $L \in (\mathcal{A}^1(M))^*$, on $(\mathcal{A}^1(M))^*$ is exactly the infinitesimal Teichmüller metric for tangent vectors at the base-point of $\mathcal{T}(M)$.

3.3 Isometries of Bergman spaces

Let M and N be two hyperbolic Riemann surfaces and let $f: \mathcal{T}(M) \rightarrow \mathcal{T}(N)$ be a biholomorphic map that sends base-point to base-point. We have seen that the derivative of f at the base-point of $\mathcal{T}(M)$ is a \mathbb{C} -linear isometry of $(\mathcal{A}^1(M))^*$ onto $(\mathcal{A}^1(N))^*$. In the finite dimensional case, we immediately have that the adjoint of that derivative is a \mathbb{C} -linear isometry of $\mathcal{A}^1(N)$ onto $\mathcal{A}^1(M)$. However, in the infinite dimensional case, we need to use a theorem of Earle and Gardiner, see [2], which says that if there is an invertible \mathbb{C} -linear isometry $F: (\mathcal{A}^1(M))^* \rightarrow (\mathcal{A}^1(N))^*$, then there is always an invertible \mathbb{C} -linear isometry $L: \mathcal{A}^1(N) \rightarrow \mathcal{A}^1(M)$ which is the adjoint of F . In this way, we pass from biholomorphic maps between Teichmüller spaces to linear isometries between Bergman spaces.

There are two obvious types of isometries between $\mathcal{A}^1(N)$ and $\mathcal{A}^1(M)$. The map $\varphi \mapsto \theta\varphi$ is an isometry of $\mathcal{A}^1(M)$ onto itself whenever $\theta \in \mathbb{C}$ has $|\theta| = 1$. Also, if α

is a conformal map of M onto N , each $\varphi \in \mathcal{A}^1(N)$ can be pulled back to a quadratic differential $\alpha^*(\varphi) = (\varphi \circ \alpha)(\alpha')^2$ on $\mathcal{A}^1(M)$, and the map $\varphi \mapsto \alpha^*(\varphi)$ is an isometry.

Definition 3.5. If M and N are Riemann surfaces, then a surjective linear isometry $T: \mathcal{A}^1(M) \rightarrow \mathcal{A}^1(N)$ is called geometric if there exists a conformal map $\alpha: M \rightarrow N$ and a complex number θ with $|\theta| = 1$ such that

$$T^{-1}(\varphi) = \theta(\varphi \circ \alpha)\alpha'^2,$$

for every $\varphi \in \mathcal{A}^1(N)$.

Theorem 3.6. *Suppose that M and N are Riemann surfaces which are of non-exceptional finite type and that $T: \mathcal{A}^1(M) \rightarrow \mathcal{A}^1(N)$ is a surjective complex-linear isometry. Then T is geometric.*

Royden proved Theorem 3.6 in [16] in the case where M and N are compact and hyperbolic, and his method was extended to Riemann surfaces of non-exceptional finite type, even though M and N are not assumed to be homeomorphic, by Earle and Kra in [3] and Lakic in [12]. Some further special cases of Theorem 3.6 were proved by Matsuzaki in [13]. Markovic proved 3.6 in full generality, that is, for the infinite analytic type case, in [14]. As in [5], we will use the methods of [14] to prove Theorem 3.6 in the finite analytic case, which gives a good indication of the methods used, without going into the technical detail required for the general case.

Let \tilde{M} be a compact Riemann surface of genus $g \geq 0$ and let E be a finite, possibly empty, subset of \tilde{M} which contains exactly $n \geq 0$ points. We assume that $2g + n \geq 5$, so that the Riemann surface $M = \tilde{M} \setminus E$ has non-exceptional finite type.

We will consider projective embeddings of \tilde{M} associated with $\mathcal{A}^1(M)$. Let k be a positive integer, and let \mathbb{P}^k be the k -dimensional complex projective space. Each point $(z_0, \dots, z_k) \in \mathbb{C}^{k+1} \setminus \{0\}$ determines a point $[(z_0, \dots, z_k)] \in \mathbb{P}^k$. The formula

$$\pi_0(z_1, \dots, z_k) = [(1, z_1, \dots, z_k)]$$

defines a holomorphic map of \mathbb{C}^k onto a dense open subset of \mathbb{P}^k .

Let $\mathcal{M}(\tilde{M})$ be the field of meromorphic functions on \tilde{M} . For any divisor D on \tilde{M} , we define $\mathcal{O}_D(\tilde{M})$ to be the complex vector space of all functions in $\mathcal{M}(\tilde{M})$, including the zero function, that are multiples of the divisor $-D$, that is

$$\mathcal{O}_D(\tilde{M}) = \{f \in \mathcal{M}(\tilde{M}) : \text{ord}_x(f) \geq -D(x) \text{ for all } x \in \tilde{M}\}.$$

Proposition 3.7. *Let $M = \tilde{M} \setminus E$ as above, and let $\varphi_0, \dots, \varphi_k$ be a basis for $\mathcal{A}^1(M)$. Set $f_j = \varphi_j/\varphi_0$, for $j = 1, \dots, k$ and set*

$$M_0 = \tilde{M} \setminus \{x \in \tilde{M} : \text{some } f_j \text{ has a pole at } x\}.$$

Let $F: M_0 \rightarrow \mathbb{C}^k$ be the holomorphic map $F = (f_1, \dots, f_k)$. There is a unique holomorphic embedding $\Phi: \tilde{M} \rightarrow \mathbb{P}^k$ such that

$$\Phi(x) = \pi_0(F(x)) \quad \text{for all } x \in M_0.$$

Proof. Consider the divisor $D = (\varphi_0) + \chi_E$ on \tilde{M} . Clearly, $\mathcal{A}^1(M)$ is the set of meromorphic quadratic differentials $\varphi = f\varphi_0$ such that $f \in \mathcal{O}_D(\tilde{M})$.

Again, $\deg(D) \geq 2g + 1$. Since the functions $1, f_1, \dots, f_k$ are a basis for $\mathcal{O}_D(\tilde{M})$, we can conclude the map

$$x \mapsto [(1, f_1(x), \dots, f_k(x))],$$

for $x \in \tilde{M}$, when interpreted appropriately at the poles of the f_j , defines a holomorphic embedding of \tilde{M} into \mathbb{P}^k . \square

Corollary 3.8. *The map F defined above is a homeomorphism of M_0 onto a closed subset of \mathbb{C}^k .*

Proof. Since F is holomorphic on M_0 , it is continuous. Since $\Phi^{-1} \circ \pi_0 \circ F$ is the identity map of M_0 to itself, F is a homeomorphism. To see that $F(M_0)$ is a closed set, consider a sequence (x_n) in M_0 such that $F(x_n)$ converges to $z = (z_1, \dots, z_k)$ in \mathbb{C}^k . We may assume that x_n converges to some point $\bar{x} \in M$. Then $f_j(\bar{x}) = z_j \neq \infty$ for $j = 1, \dots, k$ and so $\bar{x} \in M_0$ and $z = F(\bar{x})$. \square

Definition 3.9. Suppose that f_1, \dots, f_n are μ -measurable functions on X and that g_1, \dots, g_n are ν -measurable functions on Y . Writing $F = (f_1, \dots, f_n)$ and $G = (g_1, \dots, g_n)$, which we consider as \mathbb{C}^n valued functions, then F and G are equimeasurable if

$$\mu(F^{-1}(E)) = \nu(G^{-1}(E))$$

for every Borel set $E \subset \mathbb{C}$.

The following theorem on a condition for equimeasurability and the previous lemmas are due to Rudin in [17].

Theorem 3.10. *Let $0 < p < \infty$, $p \neq 2, 4, 6, \dots$, $n \in \mathbb{N}$, and let μ and ν be measures on measurable spaces X and Y respectively. If for $1 \leq i \leq n$, $f_i \in \mathcal{L}^p(\mu)$ and $g_i \in \mathcal{L}^p(\nu)$, and*

$$\int_X |1 + z_1 f_1 + \dots + z_n f_n|^p d\mu = \int_Y |1 + z_1 g_1 + \dots + z_n g_n|^p d\nu$$

for all $(z_1, \dots, z_n) \in \mathbb{C}^n$, then (f_1, \dots, f_n) and (g_1, \dots, g_n) are equimeasurable.

We are now in a position to prove Theorem 3.6 for the finite analytic case.

Proof of Theorem 3.6. We will prove this theorem in the case where the given Riemann surfaces M and N are the complements of finite sets in compact Riemann surfaces \tilde{M} and \tilde{N} . Again, for the proof in full generality, see [14]. Note that we do not assume \tilde{M} and \tilde{N} have the same genus.

Let $\varphi_0, \dots, \varphi_k$ be a basis for $\mathcal{A}^1(M)$, and define M_0 and the map

$$F = (f_1, \dots, f_k)$$

as in the previous proposition. Set $\psi_j = T(\varphi_j)$ for $j = 0, \dots, k$. Since

$$T: \mathcal{A}^1(M) \rightarrow \mathcal{A}^1(N)$$

is a surjective \mathbb{C} -linear isometry, ψ_0, \dots, ψ_k is a basis for $\mathcal{A}^1(N)$. Set $g_j = \psi_j/\psi_0$, for $j = 1, \dots, k$, and set

$$N_0 = \tilde{N} \setminus \{y \in \tilde{N} : \text{some } g_j \text{ has a pole at } y\}.$$

Let $G: N_0 \rightarrow \mathbb{C}^k$ be the holomorphic embedding map $G = (g_1, \dots, g_k)$. By Proposition 3.10, there is a holomorphic embedding $\Psi: \tilde{N} \rightarrow \mathbb{P}^k$ such that $\Psi = \pi_0 \circ G$ on N_0 .

Let μ and ν be the finite positive Borel measures on M_0 and N_0 defined by

$$\mu(A) = \int_A |\varphi_0|$$

for all Borel sets $A \subset M_0$, and

$$\nu(B) = \int_B |\psi_0|$$

for all Borel sets $B \subset N_0$. Since T is a \mathbb{C} -linear isometry, we have

$$\begin{aligned} \int_{M_0} \left| 1 + \sum_{j=1}^k \lambda_j f_j \right| d\mu &= \int_M \left| \varphi_0 + \sum_{j=1}^k \lambda_j \varphi_j \right| \\ &= \int_N \left| \psi_0 + \sum_{j=1}^k \lambda_j \psi_j \right| = \int_{N_0} \left| 1 + \sum_{j=1}^k \lambda_j g_j \right| d\nu, \end{aligned}$$

for all $(\lambda_1, \dots, \lambda_k) \in \mathbb{C}^k$. Therefore, the maps F and G , and the measures μ and ν satisfy Rudin's equimeasurability condition, Theorem 3.10. Applying this to the closed subset $F(M_0)$ of \mathbb{C}^k , we obtain

$$\begin{aligned} \|\varphi_0\| &= \int_{M_0} |\varphi_0| = \mu(M_0) = \nu(G^{-1}(F(M_0))) \\ &= \int_{G^{-1}(F(M_0))} |\psi_0| \leq \int_{N_0} |\psi_0| = \|\psi_0\|. \end{aligned}$$

Since $\|\psi_0\| = \|\varphi_0\|$, the weak inequality here is actually an equality, and then $G^{-1}(F(M_0))$ has full measure in N_0 . Since it is a closed subset of N_0 , $G^{-1}(F(M_0))$ equals N_0 , and $G(N_0)$ is contained in $F(M_0)$.

Similarly, applying the equimeasurability condition of Theorem 3.7 to the set $G(N_0)$, we find that $F(M_0)$ is a subset of $G(N_0)$. Therefore the sets $F(M_0)$ and $G(N_0)$ are equal, and so are their images under the map π_0 from \mathbb{C}^k to \mathbb{P}^k . Now $\pi_0(F(M_0)) = \Phi(M_0)$ is dense in the compact set $\Phi(\tilde{M})$, and $\pi_0(G(N_0))$ is dense in the compact set $\Psi(\tilde{N})$, and so the sets $\Phi(\tilde{M})$ and $\Psi(\tilde{N})$ are equal.

Let $h: \tilde{N} \rightarrow \tilde{M}$ be the bijective holomorphic map $\Phi^{-1} \circ \Psi$. The restriction of h to N_0 satisfies $F \circ h = G$ and $h(N_0) = M_0$. From the definitions of F and G , we obtain

$$\frac{T(\varphi_j)}{T(\varphi_0)} = \frac{\psi_j}{\psi_0} = g_j = f_j \circ h = \frac{\varphi_j}{\varphi_0} \circ h = \frac{h^*(\varphi_j)}{h^*(\varphi_0)}$$

for $j = 1, \dots, k$, and so

$$\frac{T(\varphi)}{T(\varphi_0)} = \frac{h^*(\varphi)}{h^*(\varphi_0)}$$

for all $\varphi \in \mathcal{A}^1(M)$, where we write $h^*(\varphi)$ for the pullback of φ by h . Let K be any compact set in N_0 . Applying the equimeasurability condition to the compact set $G(K)$ in \mathbb{C}^k , we obtain

$$\int_K |T(\varphi_0)| = \int_K |\psi_0| = \nu(K) = \mu(h(K)) = \int_{h(K)} |\varphi_0| = \int_K |h^*(\varphi_0)|.$$

Since K is arbitrary, we must have $|T(\varphi_0)| = |h^*(\varphi_0)|$ in N_0 , and hence in all of \tilde{N} . Therefore $T(\varphi_0) = e^{it} h^*(\varphi_0)$ for some $t \in \mathbb{R}$, and we see that

$$T(\varphi) = e^{it} h^*(\varphi), \tag{3.2}$$

for all $\varphi \in \mathcal{A}^1(M)$. To complete the proof, we need to show that $h(N) = M$. By Proposition 3.7, N is the set of points in \tilde{N} where every $T(\varphi)$ is finite, and $h^{-1}(M)$ is the set of points in \tilde{N} where every $h^*(\varphi)$ is finite. These sets coincide by (3.2). \square

We have that every biholomorphic map between two Teichmüller spaces $\mathcal{T}(M)$ and $\mathcal{T}(N)$ is induced by a quasiconformal map between M and N , unless one of them has exceptional type. The automorphism group of $\mathcal{T}(M)$, denoted $\text{Aut}(\mathcal{T}(M))$ is the group of all biholomorphic self-mappings of $\mathcal{T}(M)$.

Every quasiconformal mapping $g: M \rightarrow N$ induces a mapping $\rho_g: \mathcal{T}(M) \rightarrow \mathcal{T}(N)$ given by

$$\rho_g([f]) = [f \circ g^{-1}].$$

The mapping class group $\text{MC}(M)$ is the group of all Teichmüller classes of quasiconformal maps from the Riemann surface M onto itself. Further, every $g \in \text{MC}(M)$ induces an automorphism ρ_g of $\mathcal{T}(M)$. Theorem 3.6 immediately gives us the following result.

Theorem 3.11. *If M is a Riemann surface of non-exceptional type, then*

$$\text{Aut}(\mathcal{T}(M)) = \text{MC}(M).$$

4 Local rigidity of Teichmüller spaces

In the previous section, we saw that a surjective linear isometry between the Bergman spaces of two Riemann surfaces implies that the two Riemann surfaces are conformally

equivalent. In this section, we will use a classical Banach space result of Pelczynski [15] to prove a result of Fletcher [6] which shows that when two Riemann surfaces M and N are of infinite analytic type, their Bergman spaces will always be isomorphic. This then implies, via the Bers embedding, that their Teichmüller spaces will be locally bi-Lipschitz equivalent.

4.1 Projections on Banach spaces

Let \mathbb{A} be two-dimensional Lebesgue measure on the domain $\Omega \subseteq \mathbb{C}$. Then $\mathcal{L}^1(\Omega)$ is the Banach space of measurable functions on Ω which have finite \mathcal{L}^1 -norm given by

$$\|f\|_1 = \int_{\Omega} |f(z)| d\mathbb{A}(z) < \infty,$$

for $f \in \mathcal{L}^1(\Omega)$. The Banach space $\mathcal{A}^1(\Omega)$ is the subset of $\mathcal{L}^1(\Omega)$ consisting of holomorphic functions.

Lemma 4.1. *Let Ω be a simply connected precompact subset of a Riemann surface M . Then given $\varepsilon > 0$, there exists a projection $P: \mathcal{L}^1(\Omega) \rightarrow \mathcal{L}^1(\Omega)$ such that $\|P\| < 1$,*

$$\|P(f) - f\| < \varepsilon,$$

for all $f \in \mathcal{A}^1(\Omega)$ satisfying $\|f\|_1 \leq 1$, and $P(\mathcal{L}^1(\Omega))$ is isometric to $(l^1)_n$, where $(l^1)_n$ is the n -dimensional subspace of l^1 with all terms except possibly the first n being equal to 0.

Proof. We can for simplicity assume that Ω is a bounded simply connected plane domain. Subdivide Ω into a finite number of subsets, $\Omega_1, \dots, \Omega_n$. For a given $f \in \mathcal{L}^1(\Omega)$, define λ_i to be $\int_{\Omega_i} f$. We have

$$\sum_{i=1}^n |\lambda_i| = \sum_{i=1}^n \left| \int_{\Omega_i} f \right| \leq \int_{\Omega} |f| < \infty.$$

Define the map $P: \mathcal{L}^1(\Omega) \rightarrow \mathcal{L}^1(\Omega)$ by

$$P(f) = \sum_{i=1}^n \frac{\lambda_i}{m(\Omega_i)} \mathbf{1}_{\Omega_i},$$

where $\mathbf{1}_{\Omega_i}$ denotes the indicator function of Ω_i , and m is the usual two-dimensional Lebesgue measure of Ω_i . The map P is clearly linear and bounded ($\|P\| \leq 1$ in fact), and also a projection, since $P^2 = P$.

We can define a map $\mu: P(\mathcal{L}^1(\Omega)) \rightarrow (l^1)_n$ given by

$$\mu(P(f)) = (\lambda_1, \dots, \lambda_n, 0, \dots).$$

Now, $\|\mu(P(f))\|_{l^1} = \sum_{i=1}^n |\lambda_i|$. Also,

$$\|P(f)\|_1 = \int_{\Omega} |P(f)| = \int_{\Omega} \left| \sum_{i=1}^n \frac{\lambda_i}{m(\Omega_i)} \mathbf{1}_{\Omega_i} \right| = \sum_{i=1}^n \int_{\Omega_i} \left| \frac{\lambda_i}{m(\Omega_i)} \right| = \sum_{i=1}^n |\lambda_i|$$

since the supports of $\mathbf{1}_{\Omega_i}$ are disjoint. Hence μ is isometric, and so $P(\mathcal{L}^1(\Omega))$ is isometric to $(l^1)_n$.

We now have to show that we can find a fine enough subdivision of Ω so that for the corresponding projection P , $\|P(f) - f\| < \varepsilon$ for $f \in \mathcal{A}^1(\Omega)$ with $\|f\|_1 \leq 1$. Since Ω is precompact in M ,

$$\sup\{|f(z)|\}$$

is bounded, where the supremum is taken over all $f \in \mathcal{A}^1(M)$ with $\|f\|_1 \leq 1$ and over all $z \in \Omega$. This means that

$$\Theta = \{f|_{\Omega} : f \in \mathcal{A}^1(M), \|f\|_1 \leq 1\}$$

is a normal family, and hence is equicontinuous, ie. for all $f \in \Theta$ and for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if $|z - z_0| < \delta$, for $z, z_0 \in \Omega$, then $|f(z) - f(z_0)| < \varepsilon$.

If $B(z_i, \delta)$ is a ball centred at z_i of Euclidean radius δ , then for any holomorphic function f ,

$$\frac{1}{m(B(z_i, \delta))} \int_{B(z_i, \delta)} f = f(z_i).$$

If now Ω is subdivided into $\Omega_1, \dots, \Omega_n$, with each $\Omega_i \subset B(z_i, \delta)$ for some z_i , and P is the projection corresponding to this subdivision, then

$$\int_{\Omega_i} |f - P(f)| \leq \int_{B(z_i, \delta)} |f(z) - f(z_i)| < \varepsilon m(B(z_i, \delta))$$

recalling that $m(B(z_i, \delta))$ is the area of $B(z_i, \delta)$, and noting that the last inequality follows from the equicontinuity of Θ . Hence

$$\int_{\Omega} |f - P(f)| < \varepsilon m(\Omega)$$

and since we are assuming that $m(\Omega)$ is finite, and ε can be made as small as wished, then we have the desired conclusion that $\|P - I\|$ can be as small as desired for P corresponding to a suitably fine subdivision of Ω . \square

Lemma 4.2. *Let Y be a complemented subspace of a Banach space X , and let $T : Y \rightarrow X$ be a linear operator satisfying*

$$\|T - I|_Y\| < \varepsilon. \tag{4.1}$$

Then if ε is sufficiently small, $T(Y)$ is closed and complemented in X .

Proof. Let $S = I|_Y - T$. Then $\|S\| < \varepsilon$. Let also $P_Y: X \rightarrow Y$ be a projection, which is guaranteed to exist since Y is a complemented subspace of X . Then we have

$$X = \text{Im}(P_Y) \oplus \ker(P_Y). \quad (4.2)$$

Define $\tilde{S}: X \rightarrow X$ by

$$\tilde{S} = S \circ P_Y.$$

Then \tilde{S} is an extension of S and

$$\|\tilde{S}\| \leq \|P_Y\| \cdot \|S\|.$$

Thus if $\varepsilon < \|P_Y\|^{-1}$, we have $\|\tilde{S}\| < 1$. Now set

$$\tilde{T} = I - \tilde{S}: X \rightarrow X,$$

so that

$$\tilde{T}(x) = x - P_Y(x) + T(P_Y(x)) \quad (4.3)$$

for $x \in X$. Then \tilde{T} is an extension of T , and $\|\tilde{S}\| < 1$ implies that \tilde{T} is invertible and therefore a homeomorphism. Since \tilde{T} is a homeomorphism, the image of \tilde{T} is closed and, furthermore, since Y is a closed subspace of X , $T(Y) = \tilde{T}(Y)$ is closed. The fact that $T(Y) = \tilde{T}(Y)$ follows from (4.3). From (4.2) and (4.3), it follows that $T(Y)$ is complementary to $\tilde{T}(\ker(P_Y))$ in $\tilde{T}(X)$. Furthermore, we can rewrite (4.3) as

$$\tilde{T}(x) = (I - P_Y)(x - T(P_Y(x))) + P_Y(T(P_Y(x))) \quad (4.4)$$

for $x \in X$, where the first term on the right hand side of (4.4) is an element of $\ker(P_Y)$ and the second term on the right hand side of (4.4) is an element of Y . From the hypothesis (4.1) and the fact that T is invertible, it follows that $T: Y \rightarrow T(Y)$ is invertible and $P_Y: T(Y) \rightarrow Y$ is invertible so that the image of $P_Y \circ T$ is the whole of Y . It then follows from (4.4) that the image of \tilde{T} is the whole of X and therefore $T(Y)$ is complemented in X . \square

4.2 Bergman kernels and projections on $\mathcal{L}^1(M)$

The Bergman kernel on $\mathbb{D} \times \mathbb{D}$ is given by

$$K(z, \zeta) = \frac{1}{(1 - z\bar{\zeta})^4}.$$

Every hyperbolic Riemann surface M has the disk \mathbb{D} as its universal cover, that is, there is a Fuchsian covering group G such that $M \simeq \mathbb{D}/G$. Let $\pi: \omega \rightarrow M$ be the covering map from a fundamental region ω of \mathbb{D}/G to M , chosen so that π is injective. Now, given such a covering group G , form the Poincaré theta series given by

$$F(z, \zeta) = \sum_{\gamma \in G} K(\gamma(z), \zeta) \gamma'(z)^2.$$

Definition 4.3. Let M be a hyperbolic Riemann surface with covering group G of \mathbb{D} over M . The Bergman kernel function for $M \times M$ is given by the projection of F to M . That is,

$$K_M(\pi(z), \pi(\zeta))\pi'(z)^2\overline{\pi'(\zeta)^2} = F(z, \zeta).$$

Lemma 4.4. *The kernel function $K_M: M \times M \rightarrow \mathbb{C}$ defined above is holomorphic in the first argument, antiholomorphic in the second argument, and satisfies the following properties, where $p, q \in M$,*

- (i) $K_M(p, q) = \overline{K_M(q, p)}$;
- (ii) for every conformal $f: M \rightarrow M$, $K_M(f(p), f(q))f'(p)^2\overline{f'(q)^2} = K_M(p, q)$;
- (iii) $\int_M |K_M(p, q)| d\mathbb{A}(p) \leq \frac{\pi}{4}\rho_M^2(q)$, where $\mathbb{A}(p)$ is the area measure on M in the p -coordinate;
- (iv) for every $\varphi \in \mathcal{A}^1(M)$,

$$\varphi(p) = \frac{12}{\pi} \int_M \rho_M^{-2}(q) K_M(p, q) \varphi(q) d\mathbb{A}(q),$$

where $\mathbb{A}(q)$ is the area measure on M in the q -coordinate;

- (v) for each fixed $q \in M$,

$$\sup_{p \in M} |K_M(p, q)| \rho_M^{-2}(p) < \infty.$$

Proof. We will just prove the third property since it will be used shortly. See [6], [8] for more details. Let $p = \pi(z)$ and $q = \pi(\zeta)$ for $p, q \in M$ and $z, \zeta \in \mathbb{D}$. We have

$$\int_M |K_M(p, q)| d\mathbb{A}(p) = \int_\omega |F(z, \zeta)| d\mathbb{A}(z)$$

where ω is a fundamental region for M in \mathbb{D} ,

$$\begin{aligned} &= \int_\omega \left| \sum_{\gamma \in G} K_{\mathbb{D}}(\gamma(z), \zeta) \gamma'(z)^2 \right| d\mathbb{A}(z) \leq \sum_{\gamma \in G} \int_{\gamma(\omega)} |K_{\mathbb{D}}(z, \zeta)| d\mathbb{A}(z) \\ &= \int_{\mathbb{D}} |K_{\mathbb{D}}(z, \zeta)| d\mathbb{A}(z) \leq \frac{\pi}{4} \rho(z)^2 = \frac{\pi}{4} \rho_M(p)^2, \end{aligned}$$

which completes the proof. □

Define the linear map $P: \mathcal{L}^1(M) \rightarrow \mathcal{A}^1(M)$ by

$$(P(\varphi))(\mu) = \frac{12}{\pi} \int_M \rho_M^{-2}(q) K_M(p, q) \varphi(q) d\mathbb{A}(q) \quad (4.5)$$

for $p, q \in M$. For any $\varphi \in \mathcal{L}^1(M)$, it is clear that the integral formula for $P(\varphi)$ means that $P(\varphi)$ will be holomorphic, so the image of P is indeed $\mathcal{A}^1(M)$.

Theorem 4.5. *There exists a bounded linear projection $\theta: \mathcal{L}^1(M) \rightarrow \mathcal{A}^1(M)$, given by $\theta: \varphi \mapsto P(\varphi)$ for $\varphi \in \mathcal{L}^1(M)$.*

Proof. The map θ is clearly linear, and bounded, since

$$\begin{aligned} \|P(\varphi)\|_1 &= \int_M |P(\varphi(p))| d\mathbb{A}(p) \\ &= \frac{12}{\pi} \int_M \left| \int_M \rho_M^{-2}(q) K_M(p, q) \varphi(q) d\mathbb{A}(q) \right| d\mathbb{A}(p) \\ &\leq \frac{12}{\pi} \int_M \left(\int_M |K_M(p, q)| d\mathbb{A}(p) \right) \rho_M^{-2}(q) |\varphi(q)| d\mathbb{A}(q) \end{aligned}$$

by Fubini's theorem, which we can apply by the fifth property in Lemma 4.4, and then using the third property of Lemma 4.4 gives

$$\|P(\varphi)\|_1 \leq 3 \int_M |\varphi(q)| d\mathbb{A}(q)$$

Hence $\|\theta\| \leq 3$. The integral reproducing formula given in (4.5) shows that $\theta|_{\mathcal{A}^1(M)}$ is the identity, $\theta^2 = \theta$, and so θ is a projection. \square

4.3 Isomorphisms of Bergman spaces

Let X_1, X_2, \dots be Banach spaces with norms $\|x_i\|_i$ for $i = 1, 2, \dots$ and $x_i \in X_i$. Also let $p > 0$. Then the Banach space $(X_1 \oplus X_2 \oplus \dots)_p$ has elements of the form (x_1, x_2, \dots) , for $x_i \in X_i$, and norm given by

$$\|(x_1, x_2, \dots)\|_p = \left(\sum_{i=1}^{\infty} \|x_i\|_i^p \right)^{1/p}. \quad (4.6)$$

Theorem 4.6. *If M is a hyperbolic Riemann surface of infinite analytic type, then $\mathcal{A}^1(M)$ is isomorphic to the sequence space l^1 .*

Proof. As discussed previously, we know that this theorem applies to all Riemann surfaces where the dimension of $\mathcal{A}^1(M)$ is infinite, for example, the plane punctured at the integer lattice points, or an infinite genus surface.

We first subdivide M in an appropriate way. For every $p \in M$, there exists an open subset $U_p \subset M$ containing p , and a chart π_p such that $\pi_p(U_p)$ is a disk in \mathbb{C} and $\pi_p(p) = 0$. Let V_p be an open simply connected set in M whose closure is contained in U_p , so that in particular $\pi_p(V_p)$ is a precompact subset of $\pi_p(U_p)$.

As p varies through M , $(V_p)_{p \in M}$ forms an open cover of M , and it is possible to find a countable subset p_1, p_2, \dots such that

$$M = \bigcup_{i=1}^{\infty} V_{p_i}$$

Now modify the subsets V_{p_i} to give a disjoint partition of M in the following way: define $M_1 = V_{p_1}$, and then inductively,

$$M_n = V_{p_n} \setminus \left(\bigcup_{i=1}^{n-1} V_{p_i} \right).$$

$$\begin{array}{ccccccc} \mathcal{L}^1(M) & \xrightarrow{R=\bigoplus_i R_i} & \bigoplus_i \mathcal{L}^1(M_i) & \xrightarrow{P=\bigoplus_i P_i} & \bigoplus_i P_i(\mathcal{L}^1(M_i)) = \Lambda & \xrightarrow{\text{isometry}} & l^1 \\ \theta \downarrow & & \tilde{\theta} \downarrow & & \hat{\theta} \downarrow & & \\ \mathcal{A}^1(M) & \xrightarrow{R} & R(\mathcal{A}^1(M)) & \xrightarrow{T} & \bigoplus_i P_i(R_i(\mathcal{A}^1(M))) & & \end{array}$$

Refer to the diagram above for the following definitions. Let $R_i: \mathcal{L}^1(M) \rightarrow \mathcal{L}^1(M_i)$ be the restriction map given by $R_i(f) = f|_{M_i}$, for $f \in \mathcal{L}^1(M)$. Define the operator $R: \mathcal{L}^1(M) \rightarrow (\mathcal{L}^1(M_1) \oplus \mathcal{L}^1(M_2) \oplus \dots)_1$ by

$$R(f) = (R_1(f), R_2(f), \dots),$$

for $f \in \mathcal{L}^1(M)$. The operator R is isometric, since

$$\|R(f)\|_1 = \sum_{i=1}^{\infty} \|R_i(f)\|_1 = \sum_{i=1}^{\infty} \int_{M_i} |f| = \int_M |f| = \|f\|_1,$$

using (4.6), and R is also clearly surjective. Now, given $\varepsilon_i > 0$, by Lemma 4.1, we can find a projection P_i of $\mathcal{L}^1(M_i)$ into itself such that $\|P_i\| \leq 1$, $P_i(\mathcal{L}^1(M_i))$ is isometric to $(l^1)_{\alpha_i}$ for some $\alpha_i \in \mathbb{Z}^+$, and $\|P_i(R_i(f)) - R_i(f)\|_1 \leq \varepsilon_i$ for all $f \in \mathcal{A}^1(M)$ with $\|f\| < 1$.

Let

$$\Lambda = (P_1(\mathcal{L}^1(M_1)) \oplus P_2(\mathcal{L}^1(M_2)) \oplus \dots)_1,$$

a subspace of $(\mathcal{L}^1(M_1) \oplus \mathcal{L}^1(M_2) \oplus \dots)_1$. Since each $P_i(\mathcal{L}^1(M_i))$ is isometric to $(l^1)_{\alpha_i}$ for some $\alpha_i \in \mathbb{Z}^+$, Λ is isometric to l^1 . Now we define the operator $T: R(\mathcal{A}^1(M)) \rightarrow \Lambda$ by

$$T(R_1(f), R_2(f), \dots) = (P_1(R_1(f)), P_2(R_2(f)), \dots).$$

Since the dimension of $\mathcal{A}^1(M)$ is infinite, $R(\mathcal{A}^1(M))$ must also be infinite dimensional. We also have

$$\|T(\xi) - \xi\|_1 \leq \left(\sum_{i=1}^{\infty} \varepsilon_i \right) \|\xi\|_1$$

for $\xi \in R(\mathcal{A}^1(M))$, and so given $\varepsilon > 0$, it is possible to choose the $(\varepsilon_i)_i$ so that $\|T(\xi) - \xi\|_1 < \varepsilon \|\xi\|_1$, for $\xi \in R(\mathcal{A}^1(M))$.

There exists a bounded linear projection $\theta: \mathcal{L}^1(M) \rightarrow \mathcal{A}^1(M)$ by Theorem 4.5. Therefore, there is a bounded linear projection $\tilde{\theta}: R(\mathcal{L}^1(M)) \rightarrow R(\mathcal{A}^1(M))$, given by

$$\tilde{\theta}(R_1(f), R_2(f), \dots) = (R_1(\theta(f)), R_2(\theta(f)), \dots)$$

which is clearly linear, bounded and satisfies $\tilde{\theta}^2 = \tilde{\theta}$. Therefore $R(\mathcal{A}^1(M))$ is complemented in $R(\mathcal{L}^1(M))$. Thus, by Lemma 4.2, if ε is small enough, $T(R(\mathcal{A}^1(M)))$ is complemented in $R(\mathcal{L}^1(M))$ and, in particular, Λ . This follows since if $W \subset Y$ is complemented in X , then there exists a projection $S: X \rightarrow W$, $(\text{Im}(S)) \cap Y$ is complemented in Y and so W is complemented in Y . The projection from Λ onto $T(R(\mathcal{A}^1(M)))$ is denoted in the diagram above by $\hat{\theta}$.

If $\varepsilon < 1$, then $\|T - I\| < 1$, and Lemma 4.2 gives that T is thus invertible and an isomorphism. By a classical result due to Pelczynski [15], every infinite dimensional complemented subspace of l^1 is isomorphic to l^1 , and so $\mathcal{A}^1(M)$ is isomorphic to l^1 . \square

By taking the Banach duals of the Banach spaces in the statement of Theorem 4.6, we immediately get the following results.

Corollary 4.7. *If M is a hyperbolic Riemann surface of infinite analytic type, then $\mathcal{Q}(M)$ is isomorphic to the sequence space l^∞ , and we will denote this isomorphism by α_M^* .*

Corollary 4.8. *If M and N are two hyperbolic Riemann surfaces of infinite analytic type, then $\mathcal{A}^1(M)$ and $\mathcal{A}^1(N)$ are isomorphic, and $\mathcal{Q}(M)$ and $\mathcal{Q}(N)$ are isomorphic.*

4.4 Local bi-Lipschitz equivalence of Teichmüller spaces

We have the following situation,

$$\lambda_M: \mathcal{T}(M) \hookrightarrow \mathcal{Q}(M), \quad \alpha_M^*: \mathcal{Q}(M) \rightarrow l^\infty$$

where the image of the Bers embedding λ_M is contained in $\mathcal{Q}(M)$. Since λ_M is a locally bi-Lipschitz mapping, there exists a neighbourhood, X_M , of the identity class in $\mathcal{T}(M)$ such that $\lambda_M|_{X_M}$ is bi-Lipschitz. Since α_M^* is an isomorphism, X_M is mapped onto a neighbourhood of the origin of l^∞ by $\alpha_M^* \circ \lambda_M$. If

$$Y_M = (\alpha_M^* \circ \lambda_M)(X_M),$$

then X_M and Y_M are bi-Lipschitz equivalent.

Lemma 4.9. *If M and N are two hyperbolic Riemann surfaces with infinite dimensional Bergman spaces, then a neighbourhood of the identity class in $\mathcal{T}(M)$ is bi-Lipschitz equivalent to a neighbourhood of the identity class in $\mathcal{T}(N)$.*

Proof. Consider the neighbourhoods of the identity class in the respective Teichmüller spaces given by X_M and X_N , and consider their images in l^∞ under the respective maps $\alpha_M^* \circ \lambda_M$ and $\alpha_N^* \circ \lambda_N$, given by Y_M and Y_N :

$$\mathcal{T}(M) \xrightarrow{\lambda_M} \mathcal{Q}(M) \xrightarrow{\alpha_M^*} l^\infty \xleftarrow{\alpha_N^*} \mathcal{Q}(N) \xleftarrow{\lambda_N} \mathcal{T}(N).$$

The sets Y_M and Y_N are both open neighbourhoods of the origin in l^∞ , and so $Y := Y_M \cap Y_N$ is also an open neighbourhood of the origin. Since $\alpha_M^* \circ \lambda_M$ is a bi-Lipschitz mapping of X_M , it has an inverse on Y , and

$$((\alpha_M^* \circ \lambda_M)^{-1})(Y) \subseteq X_M$$

is an open neighbourhood of the origin in $\mathcal{T}(M)$.

Thus $(\alpha_N^* \circ \lambda_N) \circ (\alpha_M^* \circ \lambda_M)^{-1}$ is a bi-Lipschitz mapping from a neighbourhood of the identity class in $\mathcal{T}(M)$, namely $((\alpha_M^* \circ \lambda_M)^{-1})(Y)$, to a neighbourhood of the identity class in $\mathcal{T}(N)$, namely $(\alpha_N^* \circ \lambda_N)(Y)$. \square

Theorem 4.10. *If M and N are two hyperbolic Riemann surfaces with infinite dimensional Bergman spaces, then their Teichmüller spaces are locally bi-Lipschitz equivalent.*

Proof. Recall from Section 2.5 that α_μ is a bijective isometry from $\mathcal{T}(M)$ onto $\mathcal{T}(M^\mu)$, and we can therefore consider a chart for the neighbourhood of $[0] \in \mathcal{T}(M^\mu)$ to be a chart for $[\mu] \in \mathcal{T}(M)$. Thus charts for $[\mu] \in \mathcal{T}(M)$ and $[\nu] \in \mathcal{T}(N)$ correspond to charts for the respective identity classes in $\mathcal{T}(M^\mu)$ and $\mathcal{T}(N^\nu)$. Lemma 4.9 gives a bi-Lipschitz mapping between neighbourhoods of $[0] \in \mathcal{T}(M^\mu)$ and $[0] \in \mathcal{T}(N^\nu)$. Intersecting the neighbourhoods obtained from the complex structure of the respective Teichmüller spaces with those obtained from Lemma 4.9, we obtain a bi-Lipschitz mapping between neighbourhoods of $[\mu] \in \mathcal{T}(M)$ and $[\nu] \in \mathcal{T}(N)$. \square

5 Open problems

If X and Y are connected complex Banach manifolds, then the Kobayashi metrics on the respective spaces are the largest metrics for which holomorphic maps between X and Y are distance decreasing. Conversely, the smallest metric under which holomorphic mappings are distance decreasing is called the Carathéodory metric. The Carathéodory distance on a connected complex Banach manifold X is

$$C(x, y) = \sup_{f \in H(\mathbb{D}, X)} \{\rho_{\mathbb{D}}(0, t) : f(0) = x, f(t) = y\},$$

for $x, y \in X$ and where $\rho_{\mathbb{D}}$ is the hyperbolic metric on \mathbb{D} .

Problem 5.1. We can define the Carathéodory metric on Teichmüller space just as we did for the Kobayashi metric. The problem is, is the Carathéodory metric equal to the Teichmüller metric (or, equivalently, the Kobayashi metric) on Teichmüller space? Results in this direction can be found in [10], [11], where it is shown that the Carathéodory and Teichmüller metrics coincide on abelian Teichmüller disks.

Problem 5.2. Markovic's proof of Theorem 3.6 in the general case, see [14], involves how $\mathcal{A}^1(M)$ separates points. That is, we say $\mathcal{A}^1(M)$ separates $p, q \in M$ if there

exists $\varphi \in \mathcal{A}^1(M)$ such that $\varphi(p) = 0$ and $\varphi(q) \neq 0$. If M is of non-exceptional type, Markovic proves that the set of points E of M which are not separated by $\mathcal{A}^1(M)$ is discrete, which is enough to prove Theorem 3.6. The problem is, can E be shown to be empty?

Problem 5.3. Let $I: \mathbb{D} \rightarrow \mathcal{T}(M)$ be an isometry. If I is holomorphic, then the image of I is a Teichmüller disk. The problem is, do all isometries from \mathbb{D} into $\mathcal{T}(M)$, which are not necessarily holomorphic, have a Teichmüller disk as their image?

Problem 5.4. The conclusion of Theorem 4.6 is that $\mathcal{A}^1(M)$ is isomorphic to l^1 , i.e. there exists a map $\alpha_M: \mathcal{A}^1(M) \rightarrow l^1$ and a constant C_M depending only on M such that

$$\frac{\|\alpha_M(\varphi)\|_{l^1}}{C_M} \leq \|\varphi\|_1 \leq C_M \|\alpha_M(\varphi)\|_{l^1}, \quad (5.1)$$

for all $\varphi \in \mathcal{A}^1(M)$. Is there a universal constant C such that (5.1) holds, with C replacing C_M , independently of M ?

References

- [1] L. Bers, Automorphic forms and Poincaré series for finitely generated Fuchsian groups. *Amer. Math. J.* 87 (1965), 196–214. [77](#)
- [2] C. Earle and F. Gardiner, Geometric isomorphisms between infinite dimensional Teichmüller spaces. *Trans. Amer. Math. Soc.* 348 (3) (1996), 1163–1190. [77](#)
- [3] C. Earle and I. Kra, On holomorphic mappings between Teichmüller spaces. In *Contributions to Analysis*, Academic Press, New York 1974, 107–124. [78](#)
- [4] C. Earle, I. Kra and S. Krushkal, Holomorphic motions and Teichmüller spaces. *Trans. Amer. Math. Soc.* 343 (2) (1994), 927–948. [76](#)
- [5] C. Earle and V. Markovic, Isometries between the spaces of L^1 holomorphic quadratic differentials on Riemann surfaces of finite type. *Duke Math. J.* 120 (2) (2003), 433–440. [78](#)
- [6] A. Fletcher, Local Rigidity of Infinite Dimensional Teichmüller Spaces. *J. London Math. Soc.* 74 (1) (2006), 26–40. [82](#), [85](#)
- [7] A. Fletcher and V. Markovic, *Quasiconformal mappings and Teichmüller theory*. Oxford University Press, Oxford 2007. [67](#), [68](#), [69](#), [71](#), [72](#)
- [8] F. Gardiner, *Teichmüller Theory and Quadratic Differentials*. Pure Appl. Math. (N. Y.), John Wiley and Sons, New York, 1987. [85](#)
- [9] I. Kra, *Automorphic forms and Kleinian groups*. W. A. Benjamin, Reading, Mass., 1972. [74](#)
- [10] I. Kra, The Carathéodory metric on abelian Teichmüller disks. *J. Analyse Math.* 40 (1981), 129–143 [89](#)
- [11] S. Krushkal, Grunsky coefficient inequalities, Carathéodory metric and extremal quasiconformal mappings. *Comment. Math. Helv.* 64 (4) (1989), 650–660. [89](#)

- [12] N. Lakic, An isometry theorem for quadratic differentials on Riemann surfaces of finite genus. *Trans. Amer. Math. Soc.* 349 (7) (1997), 2951-2967. [78](#)
- [13] K. Matsuzaki, Inclusion relations between the Bers embeddings of Teichmüller spaces. *Israel J. Math.* 140 (2004), 113–123. [78](#)
- [14] V. Markovic, Biholomorphic maps between Teichmüller spaces. *Duke Math. J.* 120 (2) (2003), 405–431. [78](#), [79](#), [89](#)
- [15] A. Pelczynski, Projections in certain Banach spaces. *Studia. Math.* 19 (1960), 209–228. [82](#), [88](#)
- [16] H. Royden, Automorphisms and isometries of Teichmüller space. In *Advances of the Theory of Riemann Surfaces*, Ann. Math. Stud. 66, Princeton University Press, Princeton, N.J., 1971, 369–384. [76](#), [78](#)
- [17] W. Rudin, L^p -isometries and equimeasurability. *Indiana Univ. Math. J.* 25 (3) (1976) 215–228. [79](#)

Chapter 3

A construction of holomorphic families of Riemann surfaces over the punctured disk with given monodromy

Yoichi Imayoshi

Dedicated to Professor Yukio Matsumoto on his sixtieth birthday

Contents

1	Introduction	94
2	Holomorphic families of Riemann surfaces	96
	2.1 Definition of holomorphic families of Riemann surfaces	96
	2.2 Examples of holomorphic families of Riemann surfaces	97
3	Teichmüller spaces and Teichmüller modular groups	98
	3.1 Teichmüller spaces	98
	3.2 Teichmüller modular groups	100
	3.3 Teichmüller distance	100
	3.4 Bers' classification of Teichmüller modular transformations	102
	3.5 Thurston's classification of Teichmüller modular transformations	102
4	Holomorphic representations of holomorphic families of Riemann surfaces into Teichmüller spaces and topological monodromies	103
	4.1 Holomorphic representations of holomorphic families of Riemann surfaces into Teichmüller spaces	103
	4.2 Topological monodromies of holomorphic families of Riemann surfaces	104
	4.3 Rigidity theorem	104
	4.4 Examples of topological monodromies of holomorphic families of Riemann surfaces	105
5	Pseudo-periodic maps, screw numbers and valencies	110
	5.1 Pseudo-periodic maps	110
	5.2 Screw numbers	110
	5.3 Valencies	111
	5.4 Standard form	115
6	Deformation spaces of Riemann surfaces with nodes	116
	6.1 Riemann surfaces with nodes	116

6.2	Deformation spaces of Riemann surfaces with nodes	117
6.3	Parametrization near $\langle S_0, \text{id}, S_0 \rangle$ in $D(S_0)$	118
7	Proof of Theorem A	119
7.1	Type of the topological monodromy	119
7.2	Holomorphic extension of the moduli map	120
7.3	Negativity of the topological monodromy	122
7.4	Completion of (M, π, Δ^*)	122
8	Proof of Theorem B	122
8.1	Case 1: f is a periodic map	123
8.2	Case 2: f is a product of negative Dehn twists	123
8.3	Case 3: f is a generic pseudo-periodic map of negative type	124
	References	127

1 Introduction

A *holomorphic family* (M, π, B) of Riemann surfaces of type (g, n) over B is a triplet of a two-dimensional complex manifold M , a Riemann surface B and a holomorphic map $\pi: M \rightarrow B$ such that for every $t \in B$ the fiber $S_t = \pi^{-1}(t)$ is a Riemann surface of fixed finite type (g, n) and the complex structure of S_t depends holomorphically on the parameter t . Here g is the genus of S_t and n is the number of punctures of S_t . Unless otherwise stated we assume $2g - 2 + n > 0$.

The *topological monodromy* \mathcal{M}_0 of a holomorphic family (M, π, Δ^*) of Riemann surfaces over the punctured disk $\Delta^* = \{t \in \mathbb{C} \mid 0 < |t| < 1\}$ is constructed as follows: Take a circle C_0 given by $C_0(\theta) = r_0 \exp(2\pi i\theta)$, $0 \leq \theta \leq 1$, with $0 < r_0 < 1$. Fix a marking Σ_{r_0} on the Riemann surface $S_{r_0} = \pi^{-1}(r_0)$, i.e., Σ_{r_0} is a canonical system of generators of S_{r_0} . As the parameter t moves from r_0 to $t = C_0(\theta)$, we deform the markings continuously from Σ_{r_0} on S_{r_0} to a marking Σ_t on S_t . When $t = C_0(\theta)$ comes back to the initial point r_0 , we obtain a new marking Σ'_{r_0} on S_{r_0} . By Nielsen's theorem there exists an orientation-preserving homeomorphism $f_0: S_{r_0} \rightarrow S_{r_0}$ such that $(f_0)_*(\Sigma_{r_0}) = \Sigma'_{r_0}$. Then the topological monodromy \mathcal{M}_0 is an element $[f_0]$ of the mapping class group of S_{r_0} .

We say that an orientation-preserving homeomorphism $f_0: S_{r_0} \rightarrow S_{r_0}$ is a *pseudo-periodic map* if f_0 is isotopic to a homeomorphism $f'_0: S_{r_0} \rightarrow S_{r_0}$ such that there exists an admissible system (which might be empty) of simple closed curves C_1, \dots, C_k on S_{r_0} (and we shall say that $k = 0$ if the system is empty)) satisfying two conditions:

$$(1) \quad f'_0(\{C_1, \dots, C_k\}) = \{C_1, \dots, C_k\},$$

(2) for some positive integer n_0 , the restriction map of $(f'_0)^{n_0}$ to an annular neighborhood A_i of C_i is a power of the Dehn twist about C_i , $i = 1, \dots, k$, and $(f'_0)^{n_0} = \text{id}$ on $S_{r_0} \setminus (\bigcup_{i=1}^k A_i)$.

A pseudo-periodic map f_0 is said to be of *negative type* if either $k = 0$, i.e., f_0 is isotopic to a periodic map, or each $(f_0^j)^{n_i}$ is a power of the negative Dehn twist about C_i .

For a holomorphic family over the punctured disk Δ^* we have the following theorem.

Theorem A. *Let (M, π, Δ^*) be a holomorphic family of Riemann surfaces of type (g, n) over the punctured disk $\Delta^* = \{t \in \mathbb{C} \mid 0 < |t| < 1\}$. Then the topological monodromy \mathcal{M}_0 of (M, π, Δ^*) around the origin is a pseudo-periodic map and the following hold:*

- (1) *The holomorphic map $J: \Delta^* \rightarrow M_{g,n}$ defined by $J(t) = [S_t]$ has a holomorphic extension $\hat{J}: \Delta \rightarrow \hat{M}_{g,n}$, where Δ is the unit disk $\{t \in \mathbb{C} \mid |t| < 1\}$ and $\hat{M}_{g,n}$ is the Deligne–Mumford compactification of the moduli space $M_{g,n}$ of Riemann surfaces of type (g, n) .*
- (2) *\mathcal{M}_0 is of finite order if and only if $\hat{J}(0) \in M_{g,n}$.*
- (3) *\mathcal{M}_0 is of infinite order if and only if $\hat{J}(0) \in \partial M_{g,n} (= \hat{M}_{g,n} \setminus M_{g,n})$.*
- (4) *\mathcal{M}_0 is of negative type.*
- (5) *(M, π, Δ^*) has a completion $(\hat{M}, \hat{\pi}, \Delta)$, that is, \hat{M} is a two-dimensional normal complex analytic space, $\hat{\pi}: \hat{M} \rightarrow \Delta$ is a holomorphic map, and (M, π, Δ^*) is holomorphically equivalent to $(\hat{M} \setminus S_0, \hat{\pi}|_{\hat{M} \setminus S_0}, \Delta^*)$, where $S_0 = \hat{\pi}^{-1}(0)$.*
- (6) *\mathcal{M}_0 is trivial if and only if S_0 is a non-singular fiber.*

Theorem A except for the assertion \mathcal{M}_0 is of negative type was proved in [19]. Earle and Sipe [14], and Shiga and Tanigawa [45] showed that \mathcal{M}_0 is of negative type. See also Asada, Matsumoto and Oda [1], and Clemens [11].

The main purpose of this chapter is to give a proof of the following theorem due to Matsumoto and Montesinos [35], [36], which is the converse to Theorem A. This was announced in [20]. Takamura [47] also proved Theorem B.

Theorem B. *Let $f: \Sigma \rightarrow \Sigma$ be a pseudo-periodic map of negative type of an oriented topological surface Σ of type (g, n) onto itself. Then there exists a holomorphic family (M, π, Δ^*) of Riemann surfaces of type (g, n) over the punctured disk Δ^* whose topological monodromy \mathcal{M}_0 is conjugate to the isotopy class $[f]$ in the mapping class group of Σ .*

Matsumoto and Montesinos also showed that the conjugacy class of \mathcal{M}_0 determines topologically the structure of (M, π, Δ^*) , that is, if \mathcal{M}_0 is conjugate to the topological monodromy \mathcal{M}'_0 of a holomorphic family (M', π', Δ^*) of Riemann surfaces of type (g, n) , then there exist orientation-preserving homeomorphisms $F: M \rightarrow M'$ and $\varphi: \Delta^* \rightarrow \Delta^*$ such that $\pi' \circ F = \varphi \circ \pi$.

Matsumoto and Montesinos [36] studied the topology of degeneration of Riemann surfaces and proved that the ambient topological type of the singular fiber over $t = 0$

of (M, π, Δ^*) is completely determined by the topological monodromy \mathcal{M}_0 . The study of singular fibers of holomorphic families (M, π, Δ^*) of Riemann surfaces of type (g, n) originated from Kodaira [31] for Riemann surfaces of type $(1, 0)$, i.e., elliptic curves. Namikawa and Ueno [39], [40] classified singular fibers for type $(2, 0)$, and Uematsu [50] gave a numerical classification of singular fibers for type $(3, 0)$. An application of Matsumoto and Montesinos [36] was obtained by Ashikaga and Ishizaka [2] to give a complete list of singular fibers for type $(3, 0)$. Takamura [46], [47], [48] study extensively singular fibers for general type $(g, 0)$.

This chapter is organized as follows: In Section 2 we recall holomorphic families of Riemann surfaces and give some typical examples. In Section 3 we review Teichmüller spaces and the classification of Teichmüller modular transformations. In Section 4 we construct a holomorphic representation of a holomorphic family of Riemann surfaces into a Teichmüller space and its monodromy map. In Section 5 we review pseudo-periodic maps and give definitions of screw numbers and valencies of a pseudo-periodic map. In Section 6 we recall deformation spaces of Riemann surfaces with nodes, which are used to prove Theorem A and Theorem B. In Section 7 we give a proof of Theorem A, which is an alternative to a proof in [19]. Finally in Section 8 we prove Theorem B by using deformation spaces of Riemann surfaces with nodes. Families of Riemann surfaces over the punctured disk are also considered, from another point of view, in Chapter 5 of this volume, by G. Mondello [38].

The author thanks Professor Yukio Matsumoto and Tadashi Ashikaga for their valuable comments and encouragements. He is grateful to Professor Athanase Papadopoulos who recommended that he writes this chapter. He also gratefully acknowledges valuable comments of the referee. This work was partially supported by JSPS, Grant-in-Aid for Scientific Research, No.15340049 and Ministry of Education of Japan, Grant-in-Aid for Scientific Research, No.18654030.

2 Holomorphic families of Riemann surfaces

We review briefly holomorphic families of Riemann surfaces and give some typical examples.

2.1 Definition of holomorphic families of Riemann surfaces

A holomorphic family (M, π, B) of Riemann surfaces of type (g, n) over a Riemann surface B is defined as follows. Let \hat{M} be a two-dimensional complex manifold, C a one-dimensional analytic subset of \hat{M} or an empty set, and B a Riemann surface. Assume that a proper holomorphic map $\hat{\pi} : \hat{M} \rightarrow B$ satisfies two conditions:

- (1) by setting $M = \hat{M} \setminus C$ and $\pi = \hat{\pi}|_M$, the holomorphic map π is of maximal rank at every point of M , and
- (2) the fiber $S_t = \pi^{-1}(t)$ over each $t \in B$ is a Riemann surface of fixed analytically finite type (g, n) , where g is the genus of S_t and n is the number of punctures

of S_t , i.e., it is obtained by removing n distinct points from a compact Riemann surface of genus g .

We call such a triplet (M, π, B) a *holomorphic family of Riemann surfaces of type (g, n)* over B . Throughout this chapter we assume that $2g - 2 + n > 0$ unless otherwise stated.

A holomorphic family (M_1, π_1, B_1) of Riemann surface is *holomorphically equivalent* or *isomorphic* to a holomorphic family (M_2, π_2, B_2) if there exist biholomorphic maps $F: M_1 \rightarrow M_2$ and $f: B_1 \rightarrow B_2$ such that $f \circ \pi_1 = \pi_2 \circ F$.

$$\begin{array}{ccc} M_1 & \xrightarrow{F} & M_2 \\ \pi_1 \downarrow & & \downarrow \pi_2 \\ B_1 & \xrightarrow{f} & B_2 \end{array}$$

2.2 Examples of holomorphic families of Riemann surfaces

We give some simple examples of holomorphic families of Riemann surfaces.

Example 2.1. Take two Riemann surfaces B_0 and S_0 of analytically finite type. Let $M_0 = B_0 \times S_0$ and $\pi_0: M_0 = B_0 \times S_0 \rightarrow B_0$ be the canonical projection. Then (M_0, π_0, B_0) is a holomorphic family of Riemann surfaces of type (g_0, n_0) , where (g_0, n_0) is the type of S_0 . Such a family is called a *trivial family* of Riemann surfaces.

A holomorphic family (M, π, B) is said to be *globally trivial* if it is holomorphically equivalent to a trivial family (M_0, π_0, B_0) .

Example 2.2. Set $B = \mathbb{C} \setminus \{0\}$ and

$$M = \{(x, y, t) \in \mathbb{C}^2 \times B \mid y^2 = x^3 - t\}.$$

Let $\pi: M \rightarrow B$ be the canonical projection. Then (M, π, B) is a holomorphic family of Riemann surfaces of type $(1, 1)$. Note that (M, π, B) is analytically a locally trivial fiber bundle, because for any points t, t' of B , the fibers S_t and $S_{t'}$ are biholomorphically equivalent.

A holomorphic family is said to be *locally trivial* if it is analytically a locally trivial fiber bundle.

Example 2.3. Set $B = \mathbb{C} \setminus \{0, 1\}$ and

$$M = \{(x, y, t) \in \mathbb{C}^2 \times B \mid y^2 = x(x-1)(x-t)\}.$$

Let $\pi: M \rightarrow B$ be the canonical projection. Then (M, π, B) is a holomorphic family of Riemann surfaces of type $(1, 1)$, which is not locally trivial.

Example 2.4. Set $B = \mathbb{C} \setminus \{0, 1\}$ and

$$M = \{([z_0, z_1, z_2], t) \in P_2(\mathbb{C}) \times B \mid z_2^4 = z_1(z_1 - z_0)(z_1 - tz_0)z_0\},$$

where $P_2(\mathbb{C})$ is the complex projective plane. Let $\pi : M \rightarrow B$ be the canonical projection. Then (M, π, B) is a holomorphic family of Riemann surfaces of type $(3, 0)$, which is not locally trivial.

Example 2.5. Take a Riemann surface B of analytically finite type (g, n) . Let

$$M = \{(p, q) \in B \times B \mid p \neq q\}$$

and let $\pi : M \rightarrow B$ be the canonical projection. Then (M, π, B) is a locally non-trivial holomorphic family of Riemann surfaces of type $(g, n + 1)$.

Example 2.6. Kodaira [32] constructed a locally non-trivial holomorphic family (M, π, B) of Riemann surfaces of type $(g, 0)$ over a compact Riemann surface B . See also Atiyah [3], Barth, Peters and Van de Ven [4], Kas [27], and Riera [43]. We call such a complex surface M a *Kodaira surface*.

Example 2.7. For a two-dimensional, irreducible, smooth quasi-projective algebraic surface \hat{M} over the complex number field and for every point $p \in \hat{M}$, there exists a Zariski neighborhood M of p such that M has a holomorphic fibration (M, π, B) of Riemann surfaces of type (g, n) over a Riemann surface B of analytically finite type (see Griffiths [17]).

Example 2.8. Let Δ be the unit disk $\{z \in \mathbb{C} \mid |z| < 1\}$ in the complex plane, and Δ^* be the punctured unit disk $\{z \in \mathbb{C} \mid 0 < |z| < 1\}$. For any positive integer k we consider a two-dimensional complex manifold defined by

$$M = \{(x, y, t) \in \Delta^2 \times \Delta^* \mid xy = t^k\}.$$

Let $\pi : M \rightarrow \Delta^*$ be the canonical projection. For every $t \in \Delta^*$ the fiber $S_t = \pi^{-1}(t)$ is biholomorphic to an annulus $A_t^x = \{x \in \mathbb{C} \mid |t|^k < |x| < 1\}$ in the complex x -plane. Note that S_t is also biholomorphic to an annulus $A_t^y = \{y \in \mathbb{C} \mid |t|^k < |y| < 1\}$ in the complex y -plane, and note that $x \in A_t^x$ and $y \in A_t^y$ determine the same point $(x, y) \in S_t$ if and only if $xy = t^k$.

Since any fiber S_t of (M, π, Δ^*) is not of analytically finite type, this (M, π, Δ^*) does not satisfy our definition of holomorphic families of Riemann surfaces, but it plays an essential role in this chapter.

3 Teichmüller spaces and Teichmüller modular groups

In this section we review Teichmüller theory. For details we refer to [25].

3.1 Teichmüller spaces

Let R be a fixed Riemann surface of analytically finite type (g, n) with $2g - 2 + n > 0$. A *marked Riemann surface* (R, f, S) is a Riemann surface S of analytically finite type

(g, n) with a quasiconformal map $f: R \rightarrow S$. Two marked surfaces (R, f_1, S_1) and (R, f_2, S_2) are *equivalent* if there exists a conformal map $h: S_1 \rightarrow S_2$ such that the self-map $f_2^{-1} \circ h \circ f_1: R \rightarrow R$ is homotopic to the identity. We denote by $[R, f, S]$ the equivalence class of a representative (R, f, S) . The *Teichmüller space* $\mathcal{T}(R)$ of a Riemann surface R is the set of all these equivalence classes $[R, f, S]$.

We now introduce three different realizations of Teichmüller space.

Let G be a finitely generated Fuchsian group of the first kind with no elliptic elements acting on the upper half-plane \mathbb{H} such that the quotient space $R \cong \mathbb{H}/G$ is of type (g, n) . Let $\mathcal{Q}_{\text{norm}}(G)$ be the set of all quasiconformal automorphisms w of \mathbb{H} leaving $0, 1, \infty$ fixed and satisfying $wGw^{-1} \subset \text{PSL}(2, \mathbb{R})$, where $\text{PSL}(2, \mathbb{R})$ is the set of all real Möbius transformations. Two elements w_1 and w_2 of $\mathcal{Q}_{\text{norm}}(G)$ are *equivalent* if $w_1 = w_2$ on the real axis \mathbb{R} . The *Teichmüller space* $\mathcal{T}(G)$ of G is the set of all equivalence classes $[w]$ obtained by classifying $\mathcal{Q}_{\text{norm}}(G)$ by the above equivalence relation. We see that $\mathcal{T}(R)$ is canonically identified with $\mathcal{T}(G)$ and it is homeomorphic to $\mathbb{R}^{6g-6+2n}$.

Let $L^\infty(\mathbb{H}, G)_1$ be the complex Banach space of (equivalence classes of) bounded complex-valued measurable functions μ on \mathbb{H} satisfying

$$(\mu \circ g) \frac{\bar{g}'}{g'} = \mu, \quad g \in G, \quad \|\mu\|_\infty < 1.$$

For an element $\mu \in L^\infty(\mathbb{H}, G)_1$ denote by w_μ the element in $\mathcal{Q}_{\text{norm}}(G)$ with Beltrami coefficient μ . Let W^μ be the quasiconformal automorphism of the Riemann sphere $\hat{\mathbb{C}}$ such that W^μ has the Beltrami coefficient μ on the upper half-plane \mathbb{H} , and conformal on the lower half-plane \mathbb{H}^* , and

$$W^\mu(z) = \frac{1}{z+i} + O(|z+i|) \quad (3.1)$$

as $z \rightarrow -i$. This map W^μ is uniquely determined by $[w_\mu]$ up to the equivalence relation, i.e., $w_\mu = w_\nu$ on \mathbb{R} if and only if $W^\mu = W^\nu$ on \mathbb{H}^* . We set $\mathcal{T}_\beta(G) = \{[W^\mu] \mid \mu \in L^\infty(\mathbb{H}, G)_1\}$, which is called the *Bers Teichmüller space* of G . It is proved that $\mathcal{T}(G)$ is canonically identified with $\mathcal{T}_\beta(G)$.

Let φ_μ be the Schwarzian derivative of W^μ on \mathbb{H}^* . Then φ_μ is an element of the space $B_2(\mathbb{H}^*, G)$ of bounded holomorphic quadratic differentials φ for G on \mathbb{H}^* with hyperbolic L^∞ -norm

$$\|\varphi\|_\infty = \sup_{z \in \mathbb{H}^*} (\text{Im } z)^2 |\varphi(z)|.$$

The space $B_2(\mathbb{H}^*, G)$ is a $(3g - 3 + n)$ -dimensional complex vector space. Bers proved that the map sending $[W^\mu]$ into φ_μ is a biholomorphic map of $\mathcal{T}_\beta(G)$ onto a holomorphically convex bounded domain of $B_2(\mathbb{H}^*, G)$, which is denoted by $\mathcal{T}_B(G)$. By a lemma due to Nehari and Kraus, and a theorem due to Ahlfors and Weill we have

$$B(0, 1/2) \subset \mathcal{T}_B(G) \subset B(0, 3/2), \quad (3.2)$$

where $B(0, r)$ is the open ball in $B_2(\mathbb{H}^*, G)$ with center 0 and radius r for the hyperbolic L^∞ -norm.

Using the complex structure of $B_2(\mathbb{H}^*, G) \cong \mathbb{C}^{3g-3+n}$, we can introduce complex structures of $\mathcal{T}_B(G)$, $\mathcal{T}_\beta(G)$, $\mathcal{T}(G)$ and $\mathcal{T}(R)$.

3.2 Teichmüller modular groups

Let $\text{Mod}(R)$ be the *Teichmüller modular group* of R , i.e., the set of all homotopy classes $[f_0]$ on the space of quasiconformal self-maps $f_0: R \rightarrow R$.

Every element $[f_0]$ acts biholomorphically on $\mathcal{T}(R)$ by

$$[f_0]_*([R, f, S]) = [R, f \circ f_0^{-1}, S].$$

We call $[f_0]_*$ a *Teichmüller modular transformation*. It is shown that $\text{Mod}(R)$ acts properly discontinuously on $\mathcal{T}(R)$, and the quotient space $\mathcal{T}(R)/\text{Mod}(R)$ is a normal complex analytic space, which is canonically identified with the moduli space $M_{g,n}$ of Riemann surfaces of analytically finite type (g, n) .

Denote by $N(G)$ the set of all quasiconformal automorphisms ω of \mathbb{H} with $\omega G \omega^{-1} = G$. Two elements $\omega_1, \omega_2 \in N(G)$ are *equivalent* if $\omega_1 = \omega_2 \circ g_0$ on the real axis \mathbb{R} for some $g_0 \in G$. Denote by $[\omega]$ the equivalence class of a representative ω . Let $\text{Mod}(G)$ be the set of all equivalence classes $[\omega]$ in $N(G)$. Then $\text{Mod}(G)$ is canonically identified with $\text{Mod}(R)$, where $R \cong \mathbb{H}/G$. We call $\text{Mod}(G)$ the *Teichmüller modular group* of G . Every element $[\omega]$ acts on $\mathcal{T}(G)$ by

$$[\omega]_*([w]) = [\lambda \circ w \circ \omega^{-1}],$$

where $[w] \in \mathcal{T}(G)$ and $\lambda \in \text{PSL}(2, \mathbb{R})$ with $\lambda \circ w \circ \omega^{-1} \in \mathcal{Q}_{\text{norm}}(G)$.

3.3 Teichmüller distance

For any points $p_1 = [R, f_1, S_1]$ and $p_2 = [R, f_2, S_2]$ of $\mathcal{T}(R)$, the *Teichmüller distance* $d_\tau(p_1, p_2)$ between p_1, p_2 is defined by

$$d_\tau(p_1, p_2) = \inf_{g \in \mathcal{F}_{f_1, f_2}} \log \frac{1 + \|\mu_g\|_\infty}{1 - \|\mu_g\|_\infty},$$

where \mathcal{F}_{f_1, f_2} is the set of all quasiconformal maps g of S_1 to S_2 which are homotopic to $f_2 \circ f_1^{-1}$, and $\|\mu_g\|_\infty$ is the L^∞ -norm of the Beltrami differential $\mu_g = (g_{\bar{z}}/g_z)d\bar{z}/dz$ of g . It is proved that $\mathcal{T}(R)$ is complete with respect to the Teichmüller distance d_τ .

Let us recall Kobayashi distances and Carathéodory distances (see Kobayashi [29] and [30]). The *Kobayashi pseudo-distance* k_X on a complex manifold X is the largest pseudo-distance for which every holomorphic map f of X into another complex manifold Y is distance decreasing, i.e., $k_Y(f(p), f(q)) \leq k_X(p, q)$ for all points p, q of X . On the other hand, the *Carathéodory pseudo-distance* c_X on a complex manifold X is the smallest pseudo-distance for which every holomorphic map f of X into another

complex manifold Y is distance decreasing, i.e., $c_Y(f(p), f(q)) \leq c_X(p, q)$ for all points p, q of X .

In the cases that we consider below, the Kobayashi and the Carathéodory pseudo-distances are genuine distances (that is, they separate points).

(1) For the unit disk Δ

$$c_\Delta(z, w) = k_\Delta(z, w) = 2 \tanh^{-1} \left| \frac{z - w}{1 - z\bar{w}} \right|.$$

(2) For the punctured unit disk Δ^*

$$c_{\Delta^*}(z, w) = c_\Delta(z, w), \quad (z, w) \in \Delta^* \times \Delta^*.$$

(3) Let $p: \mathbb{C}^n \rightarrow \mathbb{R}^+$ be a norm(not necessarily the Euclidean norm), and $B = \{z \in \mathbb{C}^n \mid p(z) < 1\}$ the unit ball for this norm. Then

$$c_B(0, z) = k_B(0, z) = k_\Delta(0, p(z)) = 2 \tanh^{-1} p(z) \tag{3.3}$$

for any $z \in B$ (see Kobayashi [30], Example (3.1.24)).

Denote by k_τ the Kobayashi distance on the Teichmüller space $\mathcal{T}(R)$. Royden [44] proved that $k_\tau = d_\tau$.

Let c_τ be the Carathéodory distance on $\mathcal{T}(R)$. Note that $c_\tau \leq k_\tau$.

Lemma 3.1. *For any point $p \in \mathcal{T}(R)$ and positive number δ set*

$$\begin{aligned} B_k(p, \delta) &= \{q \in \mathcal{T}(R) \mid k_\tau(p, q) < \delta\}, \\ B_c(p, \delta) &= \{q \in \mathcal{T}(R) \mid c_\tau(p, q) < \delta\}. \end{aligned}$$

Then

$$B_k(p, \delta) \subset B_c(p, \delta) \Subset \mathcal{T}(R)$$

for any δ satisfying $\delta < 2 \tanh^{-1}(1/3)$.

Proof. Take a point $q \in B_k(p, \delta)$. Since $c_\tau(p, q) \leq k_\tau(p, q) < \delta$, we have $q \in B_c(p, \delta)$, and so $B_k(p, \delta) \subset B_c(p, \delta)$.

For $p = [R, f_0, R_0]$ the biholomorphic map $[f_0]_*: \mathcal{T}(R) \rightarrow \mathcal{T}(R_0)$ defined by $[f_0]_*([R, f, S]) = [R_0, f \circ f_0^{-1}, S]$ is an isometry for Carathéodory distances on $\mathcal{T}(R)$ and $\mathcal{T}(R_0)$. Note that $[f_0]_*([R, f_0, R_0]) = [R_0, \text{id}, R_0]$.

Take a Fuchsian group G_0 such that $\mathbb{H}/G_0 \cong R_0$. For the Carathéodory distance c_τ on $\mathcal{T}_B(G_0)$ we set

$$B_0(\delta) = \{\varphi \in \mathcal{T}_B(G_0) \mid c_\tau(0, \varphi) < \delta\}.$$

By (3.2) and (3.3) we get

$$c_\tau(0, \varphi) \geq c_{B(0, 3/2)}(0, \varphi) = 2 \tanh^{-1} \left(\frac{2}{3} \|\varphi\|_\infty \right).$$

Hence if $\delta < 2 \tanh^{-1}(1/3)$, then

$$\|\varphi\|_\infty \leq \frac{3}{2} \tanh\left(\frac{c_\tau(0, \varphi)}{2}\right) < \frac{3}{2} \tanh(\delta/2) < 1/2.$$

Therefore from (3.2) we conclude that

$$\overline{B_0(\delta)} \subset B(0, 1/2) \subset \mathcal{T}_B(G_0). \quad \square$$

3.4 Bers' classification of Teichmüller modular transformations

We shall classify Teichmüller modular transformations as follows (see Bers [6]).

Let d_τ be the Teichmüller distance on $\mathcal{T}(R)$. For any element $\chi (\neq \text{id})$ of $\text{Mod}(R)$ we set

$$a(\chi) = \inf_{p \in \mathcal{T}(R)} d_\tau(p, \chi(p)).$$

Then the Teichmüller modular transformation χ is classified as follows:

- (1) χ is *elliptic* if $a(\chi) = 0$, and $a(\chi) = d_\tau(p_0, \chi(p_0))$ for some p_0 , i.e., χ has a fixed point p_0 in $\mathcal{T}(R)$.
- (2) χ is *parabolic* if $a(\chi) = 0$, and $a(\chi) < d_\tau(p, \chi(p))$ for all p .
- (3) χ is *hyperbolic* if $a(\chi) > 0$, and $a(\chi) = d_\tau(p_0, \chi(p_0))$ for some p_0 .
- (4) χ is *pseudo-hyperbolic* if $a(\chi) > 0$, and $a(\chi) < d_\tau(p, \chi(p))$ for all p .

Remark 3.2. It is known that

- (1) χ is elliptic if and only if it is of finite order,
- (2) χ is parabolic if it is induced by a power of a Dehn twist.

3.5 Thurston's classification of Teichmüller modular transformations

A finite non-empty set of disjoint simple closed curves $\{C_1, \dots, C_k\}$ on R is said to be *admissible* if no C_i can be deformed continuously into either a point, a puncture of R , or into a C_j with $i \neq j$. We say that an orientation preserving homeomorphism $\omega: R \rightarrow R$ is *reduced* by $\{C_1, \dots, C_k\}$ if $\{C_1, \dots, C_k\}$ is admissible and if $\omega(C_1 \cup \dots \cup C_k) = C_1 \cup \dots \cup C_k$.

A self-map ω of R is called *reducible* if it is not isotopic to the identity map and is isotopic to a reduced map. A self-map of R is called *irreducible* if it is not reducible. This is a classification for self-maps ω , which was introduced by Thurston (cf. Thurston [49]). Theorem 4 of Bers [6] says that an element $[\omega] \in \text{Mod}(R)$ of infinite order is hyperbolic if and only if ω is irreducible.

If $\omega: R \rightarrow R$ is reduced by $\{C_1, \dots, C_k\}$, then we denote by R_1, \dots, R_m the components of $R \setminus (C_1 \cup \dots \cup C_k)$, and call them *parts* of R . Each surface R_j is of finite type (g_j, n_j) with $2g_j - 2 + n_j > 0$, and ω permutes the parts R_j . Let α_j be the

smallest positive integer so that f^{α_j} fixes R_j . We say that ω is *completely reduced* by $\{C_1, \dots, C_k\}$ if $f^{\alpha_j}|_{R_j}$ is irreducible for each j . Lemma 5 of Bers [6] shows that every reducible map is isotopic to a completely reduced map. If ω is completely reduced, then the maps $f^{\alpha_j}|_{R_j}$ are called the components maps of ω . A parabolic or pseudo-hyperbolic element $\chi \in \text{Mod}(R)$ can always be induced by a completely reduced map ω . The component maps of ω induce elements of Teichmüller modular groups of parts of R , which is called the *restrictions* of χ . The element χ is parabolic if all the restrictions are periodic or trivial, and pseudo-hyperbolic if at least one restriction is hyperbolic (see Theorem 7 of Bers [6] and its proof).

4 Holomorphic representations of holomorphic families of Riemann surfaces into Teichmüller spaces and topological monodromies

In order to study a holomorphic family (M, π, B) of Riemann surfaces by using the theory of Teichmüller space, we construct a representation of (M, π, B) into a Teichmüller space.

4.1 Holomorphic representations of holomorphic families of Riemann surfaces into Teichmüller spaces

Let (M, π, B) be a holomorphic family of Riemann surfaces of type (g, n) . Take a Teichmüller space $\mathcal{T}(R)$ of a Riemann surface R of type (g, n) .

Let $\pi_1(B, t_0)$ be the fundamental group of the base surface B with base point t_0 . Denote by \tilde{B} the universal covering surface of B , which consists of equivalence classes $[t, C]$, where $t \in B$ and C is a path from t_0 to t in B . Let $\rho: \tilde{B} \rightarrow B$ be the canonical projection defined by $\rho([t, C]) = t$.

A *representation* $\Phi: \tilde{B} \rightarrow \mathcal{T}(R)$ of (M, π, B) into $\mathcal{T}(R)$ is constructed as follows: Fix a quasiconformal map $f_0: R \rightarrow S_{t_0}$. For any point $[t, C] \in \tilde{B}$, take a quasiconformal map $a_C: S_{t_0} \rightarrow S_t$ by deforming $\text{id}: S_{t_0} \rightarrow S_{t_0}$ continuously when t moves from t_0 to t along the curve C . Then we set

$$\Phi([t, C]) = [R, a_C \circ f_0, S_t].$$

It is proved that $\Phi: \tilde{B} \rightarrow \mathcal{T}(R)$ is a well-defined holomorphic map satisfying the commutative diagram:

$$\begin{array}{ccc}
 \tilde{B} & \xrightarrow{\Phi} & \mathcal{T}(R) \\
 \rho \downarrow \pi_1(B, t_0) & & \downarrow \Pi \text{ Mod}(R) \\
 B & \xrightarrow{J} & M_{g,n}.
 \end{array}$$

Here $M_{g,n}$ is the moduli space of Riemann surfaces of analytically finite type (g, n) , which is identified with the quotient space $\mathcal{T}(R)/\text{Mod}(R)$, the map $\Pi: \mathcal{T}(R) \rightarrow M_{g,n} = \mathcal{T}(R)/\text{Mod}(R)$ is the canonical projection, and $J: B \rightarrow M_{g,n}$ is a holomorphic map given by $J(t) = [S_t]$. We call J the *moduli map* of (M, π, B) .

4.2 Topological monodromies of holomorphic families of Riemann surfaces

For any element $[C_0] \in \pi_1(B, t_0)$, we have

$$\begin{aligned} \Phi \circ [C_0]_*([t, C]) &= \Phi[t, C_0 \cdot C] \\ &= [R, a_C \circ a_{C_0} \circ f_0, S_t] \\ &= [R, a_C \circ f_0 \circ (f_0^{-1} \circ a_{C_0}^{-1} \circ f_0)^{-1}, S_t] \\ &= [f_0^{-1} \circ a_{C_0}^{-1} \circ f_0]_*([R, a_C, S_t]) \\ &= [f_0^{-1} \circ a_{C_0}^{-1} \circ f_0]_* \circ \Phi([t, C]), \end{aligned}$$

where $[f_0^{-1} \circ a_{C_0}^{-1} \circ f_0]$ is an element of the mapping class group $\text{Mod}(R)$ of R . Then we have a group homomorphism

$$\Phi_*: \pi_1(B, t_0) \rightarrow \text{Mod}(R)$$

given by $\Phi_*([C_0]) = [f_0^{-1} \circ a_{C_0}^{-1} \circ f_0]$, which satisfies the relation

$$\Phi \circ [C_0]_* = (\Phi_*([C_0]))_* \circ \Phi.$$

In this chapter, taking account of the action of Teichmüller modular transformations on $\mathcal{T}(R)$, we call $\Phi_*([C_0])^{-1} = [f_0^{-1} \circ a_{C_0} \circ f_0] \in \text{Mod}(R)$ the *topological monodromy* or *homotopical monodromy* of the representation Φ for $[C_0]$. The homomorphism $\Phi_*: \pi_1(B, t_0) \rightarrow \text{Mod}(R)$ is called the *topological monodromy* or *homotopical monodromy* of (M, π, B) for Φ .

4.3 Rigidity theorem

We remark that the notion of topological monodromy Φ_* plays an essential role in the proof of the Shafarevich conjecture and the Mordell conjecture in function fields (see [24], Jost and Yau [26], and McMullen [37]).

The topological monodromy Φ_* of a holomorphic family (M, π, B) of Riemann surfaces is a topological object. However, if (M, π, B) is locally non-trivial and the base surface B is of analytically finite type, then the monodromy Φ_* determines completely the complex analytic structure of (M, π, B) as follows:

Theorem 4.1 (Rigidity theorem). *Let (M_1, π_1, B) and (M_2, π_2, B) be locally non-trivial holomorphic families of Riemann surfaces of type (g, n) over a Riemann surface B of analytically finite type. If holomorphic representations Φ_1 and Φ_2 into*

$\mathcal{T}(R)$ induce the same topological monodromy $(\Phi_1)_* = (\Phi_2)_*$, then $\Phi_1 = \Phi_2$, and (M_1, π_1, B) is biholomorphically equivalent to (M_2, π_2, B) .

This rigidity theorem implies the following:

Theorem 4.2 (Shafarevich conjecture). *Let B be a Riemann surface of analytically finite type. Then, there are only finitely many locally non-trivial and non-isomorphic holomorphic families of Riemann surfaces of fixed finite type (g, n) over B .*

A holomorphic section of (M, π, B) means a holomorphic map $s : B \rightarrow M$ such that $\pi \circ s = \text{id}$ on B . It is easy to see that the Shafarevich conjecture implies the following Mordell conjecture.

Theorem 4.3 (Mordell conjecture). *Let B be a Riemann surface of analytically finite type. If a holomorphic family (M, π, B) of Riemann surfaces is locally non-trivial, then it has only finitely many holomorphic sections. If it is locally trivial, then it has only finitely many non-constant holomorphic sections.*

4.4 Examples of topological monodromies of holomorphic families of Riemann surfaces

We will give some examples of topological monodromies of holomorphic families of Riemann surfaces.

Example 4.4. If a holomorphic family of Riemann surfaces is globally trivial, it is obvious that $\Phi_*([C_0]) = [\text{id}]$ for any $[C_0] \in \pi_1(B, t_0)$.

Kodaira [31] classified topological monodromies of holomorphic families of type $(1, 0)$ over the punctured unit disk Δ^* . We give the following two examples of topological monodromies of holomorphic families of type $(1, 1)$.

Example 4.5. Set $B = \mathbb{C} \setminus \{0, 1\}$ and

$$M = \{(x, y, t) \in \mathbb{C}^2 \times B \mid y^2 = (x - t^2)(x - 1)\}.$$

Let $\pi : M \rightarrow B$ be the canonical projection. Then (M, π, B) is a holomorphic family of Riemann surfaces of type $(1, 1)$, which is not locally trivial.

In order to study topologically the monodromy of (M, π, B) , we consider the fundamental groups of its fibers (see Nielsen's Theorem in Harvey [18], Chapter 1).

For a point $t \in B$ we take two replicas of the x -plane \mathbb{C} cut along segments from $-\sqrt{t}$ to \sqrt{t} and from 1 to ∞ , respectively, and call them sheet I and sheet II. The cut on each sheet has two edges, labeled + edge and - edge. To construct a Riemann surface $S_t = \pi^{-1}(t) = \{(x, y) \in \mathbb{C}^2 \mid y^2 = (x - t^2)(x - 1)\}$, we attach the + edge on sheet I and the - edge on sheet II, and then attach the + edge on sheet II and the

– edge on sheet I. Then we obtain a Riemann surface S_t of analytically finite type $(1, 1)$, and a two-sheeted covering $\rho_t: S_t \rightarrow \mathbb{C}$ which is branched over $-\sqrt{t}$, \sqrt{t} and 1 with branch order 2.

Let C_0 be a circle in B which is defined by $C_0(\theta) = e^{i\theta}/2$, $(0 \leq \theta \leq 2\pi)$. Take generators $\{\alpha, \beta\}$ of the fundamental group of $S_{1/2}$. In Figure 1 we illustrate α and β on the x -plane. Note that the solid curves are located in sheet I and the dotted curve is located in sheet II.

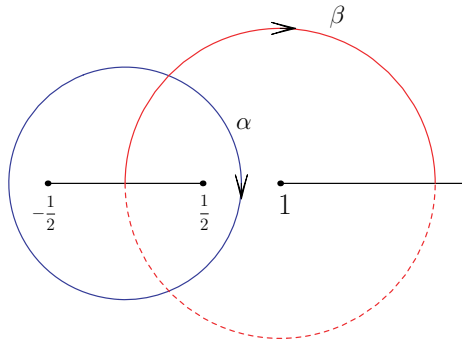


Figure 1

When θ moves from 0 to 2π , we deform generators of the fundamental group of $S_{C(\theta)}$ continuously, and we have new generators $\{\alpha', \beta'\}$ of the fundamental group of $S_{1/2}$ (see Figure 2).

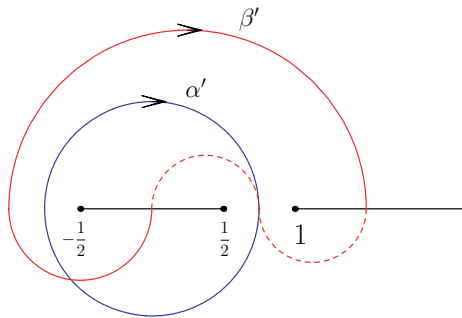


Figure 2

This means that

$$(\alpha', \beta') = (\alpha, \beta) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and the topological monodromy of (M, π, B) along $[C_0]$ is given by a Dehn twist about A , which is of parabolic type.

Example 4.6. As in Example 2.2 let us consider $B = \mathbb{C} \setminus \{0\}$,

$$M = \{(x, y, t) \in \mathbb{C}^2 \times B \mid y^2 = x^3 - t\},$$

and the canonical projection $\pi : M \rightarrow B$.

Take a circle C_0 in B defined by $C_0(\theta) = e^{i\theta}$, $(0 \leq \theta \leq 2\pi)$ and consider the generators $\{\alpha, \beta\}$ of the fundamental group of S_1 in Figure 3.

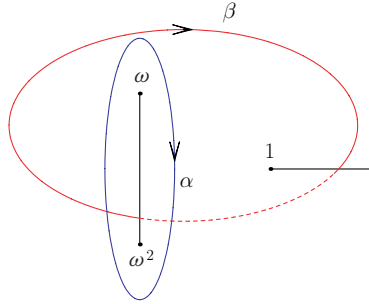


Figure 3

When θ moves from 0 to 2π , we obtain new generators $\{\alpha', \beta'\}$ of the fundamental group of S_1 in Figure 4.

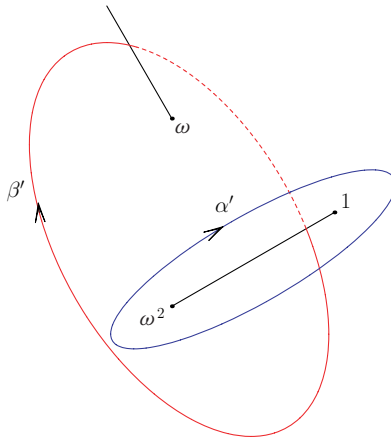


Figure 4

Hence we have

$$(\alpha', \beta') = (\alpha, \beta) \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix},$$

and the topological monodromy of (M, π, B) along $[C_0]$ is induced by a periodic map of order 6, which is of elliptic type.

Example 4.7. For the family in Example 2.4 we consider a circle C_0 in B defined by $C_0(\theta) = e^{i\theta}/2$, ($0 \leq \theta \leq 2\pi$) and generators $\{\alpha, \beta\}$ of the fundamental group of $S_{1/2}$ in Figure 5.

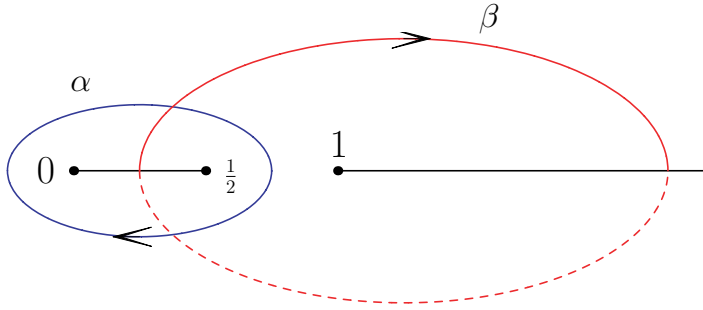


Figure 5

When θ moves from 0 to 2π , we get new generators $\{\alpha', \beta'\}$ of the fundamental group of $S_{1/2}$ in Figure 6.

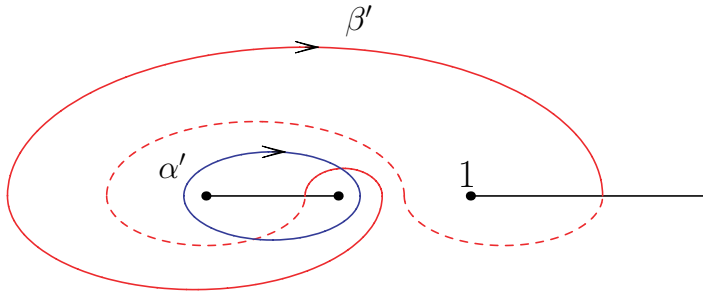


Figure 6

This implies that

$$(\alpha', \beta') = (\alpha, \beta) \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix},$$

and the topological monodromy of (M, π, B) along $[C_0]$ is induced by the twice product of a Dehn twist about A , which is of parabolic type.

Example 4.8. Kra [33] determined completely Bers' types of topological monodromies of Example 2.5 (cf. [22], [23]). Take a Riemann surface B of analytically finite type (g, n) with $2g - 2 + n > 0$. Let

$$M = \{(p, q) \in B \times B \mid p \neq q\}$$

and let $\pi : M \rightarrow B$ be the canonical projection. Then (M, π, B) is a locally non-trivial holomorphic family of Riemann surfaces of type $(g, n + 1)$. For any element γ of the

fundamental group Γ of B we denote by ω_γ the topological monodromy of (M, π, R) with respect to γ . Assume that $(g, n) \neq (0, 3)$. Then we have the following results (see Kra [33], Theorem 2).

- (1) ω_γ is not an elliptic element for any $\gamma \in \Gamma \setminus \{\text{id}\}$.
- (2) ω_γ is a parabolic element if and only if γ is either a parabolic or a simple hyperbolic element.
- (3) ω_γ is a hyperbolic element if and only if γ is an essential hyperbolic element.
- (4) ω_γ is a pseudo-hyperbolic element if and only if γ is a non-simple non-essential hyperbolic element.

Example 4.9. We may study concretely topological monodromies of Kodaira surfaces in Example 2.6. In particular, [21] classified completely Bers' types of topological monodromies of a special Kodaira surface due to Riera.

Example 4.10. Finally we consider the family of Example 2.8. For every $t \in \Delta^*$ the fiber $S_t = \pi^{-1}(t)$ is biholomorphic to an annulus $A_t^x = \{x \in \mathbb{C} \mid |t|^k < |x| < 1\}$ in the complex x -plane. Note that S_t is also biholomorphic to an annulus $A_t^y = \{y \in \mathbb{C} \mid |t|^k < |y| < 1\}$ in the complex y -plane, and note that $x \in A_t^x$ and $y \in A_t^y$ determine the same point $(x, y) \in S_t$ if and only if $xy = t^k$. For any $t, t' \in \Delta^*$, we have quasiconformal maps $f_{t,t'}$ and $g_{t,t'}$ given by

$$f_{t,t'}: A_t^x \rightarrow A_{t'}^x, \quad f_{t,t'}(x) = x |x|^{\frac{\log(t'/t)^k}{\log |t|^k}},$$

$$g_{t,t'}: A_t^y \rightarrow A_{t'}^y, \quad g_{t,t'}(y) = y |y|^{\frac{\log(t'/t)^k}{\log |t|^k}}.$$

Note that $f_{t,t'} = \text{id}$ on the unit circle $\{|x| = 1\}$ and $g_{t,t'} = \text{id}$ on $\{|y| = 1\}$.

Take a circle C_0 in Δ^* defined by $C_0(\theta) = e^{i\theta}/2$, $(0 \leq \theta \leq 2\pi)$. Setting $t = 1/2$ and $t' = C_0(\theta)$, we have

$$f_{t,t'}(x) = x |x|^{-\frac{ik\theta}{\log 2}},$$

$$g_{t,t'}(y) = y |y|^{-\frac{ik\theta}{\log 2}}.$$

Hence, when θ moves from 0 to 2π , we see that $f_{1/2,1/2} = \text{id}$ and $g_{1/2,1/2} = \text{id}$ are deformed into f and g , respectively, where

$$f(x) = x |x|^{-\frac{2\pi ki}{\log 2}},$$

$$g(y) = y |y|^{-\frac{2\pi ki}{\log 2}}.$$

We illustrate the map f for $k = 1$ in Figure 7. Such a map f is called a *negative Dehn twist* about the core curve $\{x \in \mathbb{C} \mid |x| = 3/4\}$.

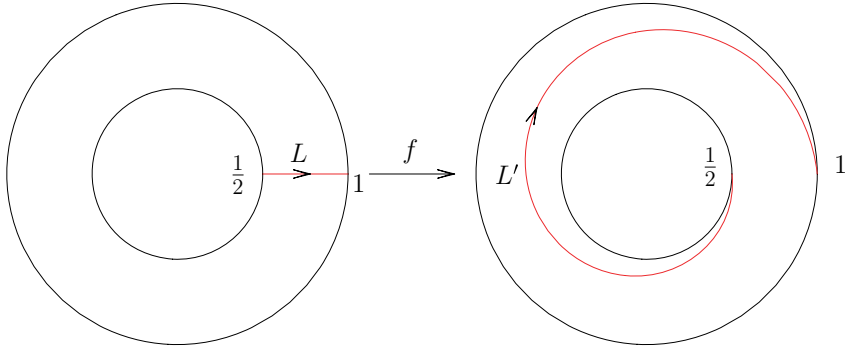


Figure 7

5 Pseudo-periodic maps, screw numbers and valencies

In this section, we assume that all surfaces are oriented, and all homeomorphisms between them are orientation-preserving. We study some topological properties of pseudo-periodic maps (see Nielsen [41], and Matsumoto and Montesinos [36]).

5.1 Pseudo-periodic maps

Let Σ be a surface of type (g, n) with $2g - 2 + n > 0$.

An orientation-preserving homeomorphism $f: \Sigma \rightarrow \Sigma$ is said to be a *pseudo-periodic map* if f is isotopic to a homeomorphism $f_0: \Sigma \rightarrow \Sigma$ which satisfies the following conditions:

- (1) f_0 is reduced by an admissible system $\mathcal{C} = \{C_j\}_{j=1}^r$ of curves on Σ (\mathcal{C} might be empty).
- (2) $f_0|_{\mathcal{B}}: \mathcal{B} \rightarrow \mathcal{B}$ is isotopic to a periodic map, where $\mathcal{B} = \Sigma \setminus (C_1 \cup \dots \cup C_r)$.

We call \mathcal{C} an *admissible system of cut curves* for f . A pseudo-periodic map is said to be *periodic* if \mathcal{C} is empty. Note that a non-periodic pseudo-periodic map is a *surface transformation of algebraically finite type* in Nielsen [42], *reducible with all component maps being of finite order* in Thurston [49], and of *parabolic type* in Bers [6].

5.2 Screw numbers

Assume that a pseudo-periodic map $f: \Sigma \rightarrow \Sigma$ is reduced by \mathcal{C} . For any $C_i \in \mathcal{C}$, there exists a minimal positive integer m_i such that $f^{m_i}(\vec{C}_i) = \vec{C}_i$. Here \vec{C}_i means an oriented curve. There also exists a minimal positive integer n_i such that $f^{n_i}|_{C_i}$ is a Dehn twist of e_i times ($e_i \in \mathbb{Z}$), where $e_i < 0$ if $f^{n_i}|_{C_i}$ is a power of a negative

(i.e., clockwise) Dehn twist in Example 4.10. We set $s(C_i) = (e_i m_i)/n_i$ and call it the *screw number* of f at C_i . We say that f is of *negative type* if $s(C_i) < 0$ for all i , ($1 \leq i \leq r$). The curve C_i is said to be *amphidrome* if m_i is even and $f^{m_i/2}(\vec{C}_i) = -\vec{C}_i$, and *non-amphidrome* otherwise.

An admissible system of cut curves $\{C_i\}_{i=1}^r$ is said to be *precise* if $s(C_i) \neq 0$ for each C_i . Note that for any pseudo-periodic map it is possible to find a precise admissible system of cut curves.

5.3 Valencies

Let Σ_0 be a surface with or without boundary. Let $f_0: \Sigma_0 \rightarrow \Sigma_0$ be a homeomorphism and \vec{C}_0 be an oriented simple closed curve on Σ_0 . Assume that there exists a positive integer n_0 such that $f_0^{n_0} = \text{id}$ on C_0 . Then the *valency* $(m, \lambda, \sigma) = (m(\vec{C}_0), \lambda(\vec{C}_0), \sigma(\vec{C}_0))$ of \vec{C}_0 with respect to f_0 is defined as follows:

- (1) m is the smallest positive integer such that $f_0^m(\vec{C}_0) = \vec{C}_0$.
- (2) λ is the order of the periodic map $f_0^m|_{\vec{C}_0}: \vec{C}_0 \rightarrow \vec{C}_0$.
- (3) For a point $q \in \vec{C}_0$, suppose that the images of q under the iteration of f^m are ordered as $\{q, f_0^{m\sigma}(q), f_0^{2m\sigma}(q), \dots, f_0^{(\lambda-1)m\sigma}(q)\}$ viewed in the orientation of \vec{C}_0 . Here, by convention, the integer σ satisfies $0 \leq \sigma < \lambda$ and $\text{gcd}(\sigma, \lambda) = 1$, so $\sigma = 0$ if and only if $\lambda = 1$.

Let $\delta = \delta(\vec{C}_0)$ be the integer determined by $\sigma\delta \equiv 1 \pmod{\lambda}$, $0 \leq \delta < \lambda$. Note that $\delta = 0$ if and only if $\lambda = 1$. The action f_0^m on \vec{C}_0 is topologically equivalent to the rotation of angle $2\pi\delta/\lambda$ in a circle.

Now consider a pseudo-periodic map $f: \Sigma \rightarrow \Sigma$ which is reduced by a precise admissible system $\mathcal{C} = \{C_i\}_{i=1}^r$ of cut curves. Take a disjoint union \mathcal{A} of annular neighborhoods $\{A_i\}_{i=1}^r$ of $\mathcal{C} = \{C_i\}_{i=1}^r$ with $f(\mathcal{A}) = \mathcal{A}$. Denote by \mathcal{B} the closure of $\Sigma \setminus \mathcal{A}$. Each annulus A_i has two boundary curves C'_i and C''_i , whose orientations are induced by the orientation of the surface \mathcal{B} . Assume that f is periodic on \mathcal{B} .

Let $(m'_i, \lambda'_i, \sigma'_i) = (m(C'_i), \lambda(C'_i), \sigma(C'_i))$ and $(m''_i, \lambda''_i, \sigma''_i) = (m(C''_i), \lambda(C''_i), \sigma(C''_i))$, i.e., they are the valencies of C'_i, C''_i with respect to f . Set $\delta'_i = \delta(C'_i)$ and $\delta''_i = \delta(C''_i)$. Then we have the following proposition (cf. Matsumoto and Montesinos [36]).

Proposition 5.1. *Let $f: \Sigma \rightarrow \Sigma$ be a pseudo-periodic map which is reduced by a precise admissible system $\mathcal{C} = \{C_i\}_{i=1}^r$ of cut curves. Then the following hold:*

- (1) *If C_i is non-amphidrome, then $m'_i = m''_i$ and*

$$K_i = -s(C_i) - \frac{\delta'_i}{\lambda'_i} - \frac{\delta''_i}{\lambda''_i} \tag{5.1}$$

is an integer.

(2) If C_i is amphidrome, then $m'_j = m''_j$ (even), $(\lambda'_j, \sigma'_j) = (\lambda''_j, \sigma''_j)$, and

$$K_j = \frac{1}{2} \left(-s(C_j) - \frac{\delta'_j}{\lambda'_j} - \frac{\delta''_j}{\lambda''_j} \right) \quad (5.2)$$

is an integer.

Proof. Let f be in the standard form in the next subsection (see Definition 5.5).

First we assume that A_i is non-amphidrome. Then $f^m(C'_i) = C'_i$ if and only if $f^m(A_i) = A_i$. Similarly, $f^m(C''_i) = C''_i$ if and only if $f^m(A_i) = A_i$. Then by the definition of m'_i and m''_i we have $m'_i = m''_i$. We set $m = m'_i = m''_i$. The lift $\tilde{f}^m|_{A_i}$ of $f^m|_{A_i}: A_i \rightarrow A_i$ to $[0, 1] \times \mathbb{R}$ is represented by

$$\tilde{f}^m|_{A_i}(t, x) = (t, x + at + b), \quad (t, x) \in [0, 1] \times \mathbb{R}$$

for some $a, b \in \mathbb{Q}$. By the geometric meaning of δ'_i/λ'_i and δ''_i/λ''_i we have

$$\frac{\delta'_i}{\lambda'_i} = -a \quad \text{and} \quad \frac{\delta''_i}{\lambda''_i} = c,$$

where $a + b = c + n$, $0 \leq c < 1$ and $n \in \mathbb{Z}$. By the definition of screw number, we have

$$s(C_i) = -b.$$

Hence, we obtain

$$\begin{aligned} -s(C_i) - \frac{\delta'_i}{\lambda'_i} - \frac{\delta''_i}{\lambda''_i} &= b + a - c \\ &= b + a - (a + b - n) \\ &= n. \end{aligned}$$

Next we assume that A_i is amphidrome. Let k be the smallest positive integer such that $f^k(A_i) = A_i$. Since A_i is amphidrome, f^k interchanges the boundary components of A_i . Thus $2k$ is the smallest positive integer such that $f^{2k}(A_i) = A_i$ and f^{2k} does not interchange the boundary components. This implies that $m'_i = m''_i$.

It is obvious that $f^k|_{C'_i}: C'_i \rightarrow C'_i$ is equivalent with respect to the actions of $f^{2k}|_{C'_i}: C'_i \rightarrow C'_i$ and $f^{2k}|_{C''_i}: C''_i \rightarrow C''_i$. This implies $(\lambda'_i, \sigma'_i) = (\lambda''_i, \sigma''_i)$.

We set $m = m'_i = m''_i$. Denote by $\tilde{f}^m|_{A_i}$ the lift of $f^m|_{A_i}: A_i \rightarrow A_i$ to $[0, 1] \times \mathbb{R}$. Then we have

$$\begin{aligned} \tilde{f}^m|_{A_i}(0, x) &= (0, x - 2a), \\ \tilde{f}^m|_{A_i}(1, x) &= (1, x + 2a) \end{aligned}$$

for some $a \in \mathbb{Q}$.

By the geometric meaning of δ'_i/λ'_i and δ''_i/λ''_i we have

$$\frac{\delta'_i}{\lambda'_i} = c \quad \text{and} \quad \frac{\delta''_i}{\lambda''_i} = c,$$

where $2a = c + n$, $0 \leq c < 1$ and $n \in \mathbb{Z}$. By the definition of screw number, we have

$$s(C_i) = -4a.$$

Hence, we obtain

$$\begin{aligned} -\frac{s(C_i)}{2} - \frac{\delta'_i}{2\lambda'_i} - \frac{\delta''_i}{2\lambda''_i} &= 2a - \frac{c}{2} - \frac{c}{2} \\ &= 2a - c \\ &= n. \end{aligned}$$

□

Example 5.2. Here we give a typical and non-trivial pseudo-periodic map. Let Σ be the closed surface of genus 3 in Figure 8. We assume that there exists a periodic homeomorphism $\varphi_0: \Sigma \rightarrow \Sigma$ of order two such that $\varphi_0(C_1) = C_2$, $\varphi_0(C'_1) = C'_2$, $\varphi_0(C''_1) = C''_2$, $\varphi_0(C_3) = C_4$, $\varphi_0(C'_3) = C'_4$, and $\varphi_0(C''_3) = C''_4$, where C_1, \dots, C''_4 are simple closed curves on Σ as illustrated in Figure 8.

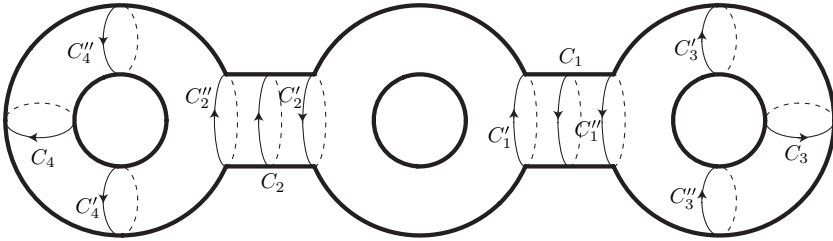


Figure 8

We consider the following subdomains of Σ .

- \mathcal{A}_1 : the ring domain bounded by C'_1, C''_1 ,
- \mathcal{D}_2 : the domain bounded by C''_1, C'_3, C''_3 ,
- \mathcal{A}_3 : the ring domain bounded by C'_3, C''_3 ,
- \mathcal{A}_4 : the ring domain bounded by C'_4, C''_4 .

Take a homeomorphism g_1 of \mathcal{A}_1 to the annulus $A_1 = \{z \in \mathbb{C} \mid 1/2 < |z| < 1\}$. As in Example 4.1, let $h_1: A_1 \rightarrow A_1$ be the negative half-Dehn twist defined by

$$h_1(z) = z|z|^{-\pi i / \log 2}.$$

Then $\varphi_1 = g_1^{-1} \circ h_1 \circ g_1: \mathcal{A}_1 \rightarrow \mathcal{A}_1$ is a negative half-Dehn twist about C_1 .

Next we take a homeomorphism g_3 of \mathcal{A}_3 to the annulus $A_3 = \{z \in \mathbb{C} \mid 1/2 < |z| < 2\}$. Setting

$$h_3(z) = \frac{1}{z},$$

we have a periodic homeomorphism $\varphi_3 = g_3^{-1} \circ h_3 \circ g_3: \mathcal{A}_3 \rightarrow \mathcal{A}_3$ of order two.

It is possible to find a periodic homeomorphism $\varphi_2: \mathcal{D}_2 \rightarrow \mathcal{D}_2$ of order two such that $\varphi_2 = \varphi_1$ on C_1'' , and $\varphi_2 = \varphi_3$ on C_3', C_3'' .

Take a homeomorphism g_4 of \mathcal{A}_4 to the annulus $A_4 = \{z \in \mathbb{C} \mid 1/2 < |z| < 1\}$. Let $h_4: A_4 \rightarrow A_4$ be the ℓ -th power of the negative Dehn twist in Example 4.10, which is defined by

$$h_4(z) = z|z|^{-2\ell\pi i/\log 2}.$$

Then $\varphi_4 = g_4^{-1} \circ h_4 \circ g_4: \mathcal{A}_4 \rightarrow \mathcal{A}_4$ is the ℓ -th power of the negative Dehn twist about C_4 .

Using these maps we have two homeomorphisms $f_1, f_2: \Sigma \rightarrow \Sigma$ given by

$$f_1(p) = \begin{cases} \varphi_1(p), & p \in \mathcal{A}_1, \\ \varphi_2(p), & p \in \mathcal{D}_2, \\ \varphi_3(p), & p \in \mathcal{A}_3, \\ p, & p \in \Sigma \setminus (\mathcal{A}_1 \cup \mathcal{D}_2 \cup \mathcal{A}_3), \end{cases}$$

$$f_2(p) = \begin{cases} \varphi_4 \circ \varphi_0(p), & p \in \mathcal{A}_3, \\ \varphi_0(p), & p \in \Sigma \setminus \mathcal{A}_3. \end{cases}$$

Now we consider a periodic homeomorphism $f_0 = f_2 \circ f_1: \Sigma \rightarrow \Sigma$ of order 4, which is a pseudo-periodic map of Σ . Note that f_0 is completely reduced by $\{C_1, C_2, C_3, C_4\}$, and that C_1, C_2 are non-amphidrome, but C_3, C_4 are amphidrome with respect to f_0 .

Let us find its screw numbers. For C_1 , we have $m_1 = 2, n_1 = 4$, and $e_1 = -1$. Thus we obtain

$$s(C_1) = \frac{e_1 m_1}{n_1} = -\frac{1}{2}.$$

For C_3 , since $m_3 = 4, n_3 = 4, e_3 = -2\ell$, we have

$$s(C_3) = \frac{e_3 m_3}{n_3} = -2\ell.$$

It is also shown that

$$(m(C_1'), \lambda(C_1'), \sigma(C_1')) = (2, 1, 0),$$

$$(m(C_1''), \lambda(C_1''), \sigma(C_1'')) = (4, 2, 1).$$

Since $\delta(C_1') = 0$ and $\delta(C_1'') = 1/2$, we obtain

$$K_1 = -s(C_1) - \frac{\delta(C_1')}{\lambda(C_1')} - \frac{\delta(C_1'')}{\lambda(C_1'')} \\ = -\left(-\frac{1}{2}\right) - 0 - \frac{1}{2} \\ = 0.$$

For C_3 , we have

$$\begin{aligned}(m(C'_3), \lambda(C'_3), \sigma(C'_3)) &= (4, 1, 0), \\ (m(C''_3), \lambda(C''_3), \sigma(C''_3)) &= (4, 1, 0).\end{aligned}$$

Hence we get $\delta(C'_3) = \delta(C''_3) = 0$, and

$$\begin{aligned}K_3 &= \frac{1}{2} \left(-s(C_3) - \frac{\delta(C'_3)}{\lambda(C'_3)} - \frac{\delta(C''_3)}{\lambda(C''_3)} \right) \\ &= \frac{1}{2}(2\ell - 0 - 0) \\ &= \ell.\end{aligned}$$

5.4 Standard form

For a pseudo-periodic map $f: \Sigma \rightarrow \Sigma$, Matsumoto and Montesinos [36] constructed a special homeomorphism which is homotopic to f and is said to be in standard form (cf. Nielsen [42], §14). In order to define standard form, we need some terminology.

Definition 5.3. Let A be an annulus, and let $\varphi: [0, 1] \times S^1 \rightarrow A$ be a parametrization (i.e., homeomorphism), where $S^1 = \mathbb{R}/\mathbb{Z}$. A homeomorphism $f: A \rightarrow A$ which does not interchange the boundary components of A is called a *linear twist* with respect to φ , if

$$f \circ \varphi(t, x) = \varphi(t, x + at + b), \quad (t, x) \in [0, 1] \times \mathbb{R}$$

for some $a, b \in \mathbb{Q}$. We say simply that $f: A \rightarrow A$ is a *linear twist* if f is a linear twist with respect to a certain parametrization $\varphi: [0, 1] \times S^1 \rightarrow A$.

Definition 5.4. Let A be an annulus, and let $\varphi: [0, 1] \times S^1 \rightarrow A$ be a parametrization, where $S^1 = \mathbb{R}/\mathbb{Z}$. A homeomorphism $f: A \rightarrow A$ which interchanges the boundary components of A is called a *special (piecewise-linear) twist* with respect to φ , if

$$f \circ \varphi(t, x) = \begin{cases} \varphi(1-t, -x - 3a(t - \frac{1}{3})), & (t, x) \in [0, \frac{1}{3}] \times \mathbb{R}, \\ \varphi(1-t, -x), & (t, x) \in [\frac{1}{3}, \frac{2}{3}] \times \mathbb{R}, \\ \varphi(1-t, -x - 3a(t - \frac{2}{3})), & (t, x) \in [\frac{2}{3}, 1] \times \mathbb{R} \end{cases}$$

for some $a \in \mathbb{Q}$. We say simply that $f: A \rightarrow A$ is a *special twist* if f is a special twist with respect to a certain parametrization $\varphi: [0, 1] \times S^1 \rightarrow A$.

Definition 5.5. A pseudo-periodic map $f: \Sigma \rightarrow \Sigma$ is said to be in *standard form* if the following conditions are satisfied:

- (1) There exists a system of disjoint annular neighborhoods $\{A_i\}_{i=1}^r$ of the precise system of cut curves for f , such that $f(\mathcal{A}) = \mathcal{A}$, where $\mathcal{A} = \bigcup_{i=1}^r A_i$.
- (2) $f|_{\mathcal{B}}: \mathcal{B} \rightarrow \mathcal{B}$ is a periodic map, where \mathcal{B} is the closure of $\Sigma \setminus \mathcal{A}$.

- (3) Let k_i be the smallest positive integer such that $f^{k_i}(A_i) = A_i$, $i = 1, 2, \dots, r$.
- (3)-(a) If $f^{k_i}|_{A_i}: A_i \rightarrow A_i$ does not interchange the boundary components of A_i , then $f^{k_i}|_{A_i}$ is a linear twist.
- (3)-(b) If $f^{k_i}|_{A_i}: A_i \rightarrow A_i$ interchanges the boundary components of A_i , then $f^{k_i}|_{A_i}$ is a special twist.

Now we state the following theorem (see Matsumoto and Montesinos [36], and Gilman [16], Theorem 13.3).

Theorem 5.6. *Any pseudo-periodic map $f: \Sigma \rightarrow \Sigma$ is isotopic to a pseudo-periodic map in standard form. If two pseudo-periodic maps in standard form $f, f': \Sigma \rightarrow \Sigma$ are mutually homotopic, then there is a homeomorphism $h: \Sigma \rightarrow \Sigma$ isotopic to the identity such that $f = h^{-1} \circ f' \circ h$.*

6 Deformation spaces of Riemann surfaces with nodes

In this section we review a kind of generalization of Teichmüller spaces (see Bers [7], [8], [9], [10], Fay [15], Kra [34], and Wolpert [51]).

6.1 Riemann surfaces with nodes

A *Riemann surface with nodes* is a connected one-dimensional complex analytic space S_0 such that every point $p \in S_0$ has a fundamental system of neighborhoods each of which is isomorphic either to the disk $\{|z| < 1\}$ in \mathbb{C} or to the set $\{z_1 z_2 = 0, |z_1| < 1, |z_2| < 1\}$ in \mathbb{C}^2 ; in the second case p is called a *node* of S_0 . Every component of the complement of the set of nodes of S_0 is called a *part* of S_0 .

By a Riemann surface S_0 of analytically finite type (g, n) with nodes we mean a Riemann surface with nodes satisfying the following conditions:

- (1) Either $n = 0$ and S_0 is compact, or $n > 0$ and S_0 is compact except for n punctures. (A puncture can never be at a node.)
- (2) S_0 has finitely many parts $\Sigma_1, \dots, \Sigma_r$, each part Σ_j is a Riemann surface of analytically finite type (g_j, n_j) with $2g_j - 2 + n_j > 0$, and

$$\sum_{j=1}^r n_j = 2k + n,$$

$$\sum_{j=1}^r (2g_j - 2 + n_j) = 2g - 2 + n,$$

where k is the number of nodes.

Note that if one “thickens” each node so as to obtain a smooth surface \check{S}_0 , then \check{S}_0 is homeomorphic to a Riemann surface of analytically finite type (g, n) .

6.2 Deformation spaces of Riemann surfaces with nodes

Let S_0 and S'_0 be two Riemann surfaces of analytically finite type (g, n) with nodes. A continuous surjection $f: S'_0 \rightarrow S_0$ is called a *deformation* if

- (1) the inverse image of every node of S_0 is either a node of S'_0 or a Jordan curve on a part of S'_0 ,
- (2) for every part Σ of S_0 , the restriction $f^{-1}|_{\Sigma}$ is an orientation preserving homeomorphism onto $f^{-1}(\Sigma)$, and
- (3) every puncture of S'_0 corresponds, under f , to a puncture of S_0 .

A holomorphic deformation is called an *isomorphism*. The *moduli space* $\hat{M}_{g,n}$ of Riemann surfaces with nodes of analytically finite type (g, n) is the set of all isomorphism classes $[S_0]$ of Riemann surfaces with nodes of analytically finite type (g, n) .

The equivalence class $\langle S, f, S_0 \rangle$ of a deformation $f: S \rightarrow S_0$ consists of all deformations $S' \rightarrow S_0$ of the form $\psi \circ f \circ \varphi^{-1}$, where $\varphi: S \rightarrow S'$ is a deformation isotopic to an isomorphism and $\psi: S_0 \rightarrow S_0$ is a deformation isotopic to the identity. The *deformation space* $D(S_0)$ consists of all equivalence classes $\langle S, f, S_0 \rangle$ of deformations to S_0 .

Every deformation $g: S_1 \rightarrow S_0$ induces an *allowable map* $\langle g \rangle_*: D(S_1) \rightarrow D(S_0)$ which sends $\langle S, f, S_1 \rangle \in D(S_1)$ into $\langle S, g \circ f, S_0 \rangle \in D(S_0)$. Let $\text{Mod}(S_0)$ be the group of allowable self-maps of $D(S_0)$ induced by all topological orientation preserving self-maps of S_0 , and let $\text{Mod}_0(S_0)$ be the subgroup induced by the isomorphisms of S_0 .

The following statements are proved in the references given at the beginning of this section.

Proposition 6.1. *The deformation space $D(S_0)$ of a Riemann surface S_0 with nodes of analytically finite type (g, n) is a complex manifold and biholomorphically equivalent to a bounded domain in \mathbb{C}^{3g-3+n} .*

Proposition 6.2. *An allowable map $\langle g \rangle_*: D(S_1) \rightarrow D(S_0)$ is holomorphic. If S_1 and S_0 have the same number of nodes, then $\langle g \rangle_*$ is biholomorphic.*

Proposition 6.3. *The group $\text{Mod}(S_0)$ is discrete, the subgroup $\text{Mod}_0(S_0)$ is finite and it is the stabilizer of $\langle S_0, \text{id}, S_0 \rangle$ in $\text{Mod}(S_0)$.*

Proposition 6.4. *The canonical projection $D(S_0) \rightarrow \hat{M}_{g,n}$ is holomorphic. Furthermore, $\langle S_0, \text{id}, S_0 \rangle \in D(S_0)$ has a $\text{Mod}_0(S_0)$ -invariant neighborhood U_0 such that $U_0/\text{Mod}_0(S_0)$ is isomorphic to a neighborhood of $[S_0]$ in $\hat{M}_{g,n}$.*

Proposition 6.5. *The moduli space $\hat{M}_{g,n}$ of Riemann surfaces with nodes of analytically finite type (g, n) is a compact normal complex analytic space and a V -manifold, i.e., a complex analytic space whose singularities are quotient singularities of a complex Euclidean space by a finite linear group.*

6.3 Parametrization near $\langle S_0, \text{id}, S_0 \rangle$ in $D(S_0)$

We parametrize Riemann surfaces with or without nodes near $\langle S_0, \text{id}, S_0 \rangle$ in the deformation space $D(S_0)$ of S_0 as follows (see Wolpert [51], §4).

At the node p_i , $(i = 1, \dots, r)$, the punctures a_i and b_i are paired. Choose disjoint neighborhoods D_i^1, D_i^2 , $i = 1, \dots, r$, of the punctures a_i and b_i and let $z_i : D_i^1 \rightarrow \Delta$, $w_i : D_i^2 \rightarrow \Delta$ be local coordinates with $z_i(a_i) = 0$ and $w_i(b_i) = 0$. Fixing an open set U disjoint from D_i^1, D_i^2 , we take Beltrami differentials μ_j with support in U spanning the Teichmüller space of $S_0 \setminus \{p_1, \dots, p_r\}$ (the dimension is $N = 3g - 3 + n - r$). Given $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{C}^N$ in a neighborhood of the origin, the sum $\mu(\alpha) = \sum_j \alpha_j \mu_j$ satisfies $\|\mu\|_\infty < 1$ and thus a μ -conformal solution $f^{\mu(\alpha)}$ of the Beltrami equation exists. The Riemann surface $f^{\mu(\alpha)}(S_0) = S_\alpha$ is a quasiconformal deformation of S_0 .

Now we parametrize the opening up of the nodes. The map $f^{\mu(\alpha)}$ is conformal on D_i^1 and D_i^2 ; therefore z_i and w_i serve as coordinates for $f^{\mu(\alpha)}(D_i^1), f^{\mu(\alpha)}(D_i^2) \subset S_\alpha$. Given $\tau = (\tau_1, \dots, \tau_r) \in \Delta^r$, we construct a surface $S_{\alpha, \tau}$ as follows. Remove the disks $\{z_i \mid |z_i| \leq |\tau_i|\}$ and $\{w_i \mid |w_i| \leq |\tau_i|\}$ from S_α . Attach $\{z_i \mid |\tau_i| < |z_i| \leq 1\}$ and $\{w_i \mid |\tau_i| < |w_i| \leq 1\}$ by identifying z_i and τ_i/w_i to obtain $S_{\alpha, \tau}$. The couple (α, τ) gives holomorphic coordinates at $\langle S_0, \text{id}, S_0 \rangle$ in the deformation space $D(S_0)$ of S_0 .

Next we write explicitly the action $\langle g_0 \rangle_*$ of an element $g_0 \in \text{Aut}(S_0)$ on the deformation space $D(S_0)$ of S_0 by using the above coordinates (α, τ) . Select the charts z_i and w_i at the punctures a_i and b_i of $S_0 \setminus \{p_1, \dots, p_r\}$ such that the hyperbolic metric is given locally as $|dz_i|/(|z_i| \log(1/|z_i|))$ and $|dw_i|/(|w_i| \log(1/|w_i|))$; such coordinates are unique modulo rotation. Since the isomorphism g_0 is an isometry in the hyperbolic metric, we have

$$\begin{aligned} z_j &= g_0(z_i) = \xi_i z_i, \\ w_j &= g_0(w_i) = \chi_i w_i \end{aligned}$$

for constants $\xi_i, \chi_i, |\xi_i| = |\chi_i| = 1$. Note that

$$\begin{aligned} \xi_i &= dz_j/dz_i, \\ \chi_i &= dw_j/dw_i. \end{aligned}$$

We may assume that the finite dimensional vector space spanned by the Beltrami differentials $\{\mu_j\}$ is invariant under $\text{Aut}(S_0)$. Then for a representation of the action $\langle g_0 \rangle_*$ of g_0 on the deformation space, writing $\langle g_0 \rangle_*(\alpha, \tau) = (\beta, \sigma)$ in the above local

coordinates, we have

$$\mu(\beta) = \mu(\alpha) \circ g_0^{-1} \frac{\overline{(g_0^{-1})'}}{(g_0^{-1})'}, \quad (6.1)$$

$$\sigma_j = \xi_i \chi_i \tau_i = \frac{dz_j}{dz_i} \frac{dw_j}{dw_i} \tau_i, \quad (6.2)$$

where g_0 maps the i -th node to j -th node.

7 Proof of Theorem A

Theorem A. *Let (M, π, Δ^*) be a holomorphic family of Riemann surfaces of type (g, n) over the punctured disk $\Delta^* = \{t \in \mathbb{C} \mid 0 < |t| < 1\}$. Then the topological monodromy \mathcal{M}_0 of (M, π, Δ^*) around the origin is a pseudo-periodic map and the following hold:*

- (1) *The holomorphic map $J: \Delta^* \rightarrow M_{g,n}$ defined by $J(t) = [S_t]$ has a holomorphic extension $\hat{J}: \Delta \rightarrow \hat{M}_{g,n}$, where Δ is the unit disk $\{t \in \mathbb{C} \mid |t| < 1\}$ and $\hat{M}_{g,n}$ is the Deligne–Mumford compactification of the moduli space $M_{g,n}$ of Riemann surfaces of type (g, n) .*
- (2) *\mathcal{M}_0 is of finite order if and only if $\hat{J}(0) \in M_{g,n}$.*
- (3) *\mathcal{M}_0 is of infinite order if and only if $\hat{J}(0) \in \partial M_{g,n} (= \hat{M}_{g,n} \setminus M_{g,n})$.*
- (4) *\mathcal{M}_0 is of negative type.*
- (5) *(M, π, Δ^*) has a completion $(\hat{M}, \hat{\pi}, \Delta)$, that is, \hat{M} is a two-dimensional normal complex analytic space, $\hat{\pi}: \hat{M} \rightarrow \Delta$ is a holomorphic map, and (M, π, Δ^*) is holomorphically equivalent to $(\hat{M} \setminus S_0, \hat{\pi}|_{\hat{M} \setminus S_0}, \Delta^*)$, where $S_0 = \hat{\pi}^{-1}(0)$.*
- (6) *\mathcal{M}_0 is trivial if and only if S_0 is a non-singular fiber.*

A proof of Theorem A was given in [19]. Here we will give an alternative proof of this theorem. It is divided into four parts.

7.1 Type of the topological monodromy

Proposition 7.1. *The topological monodromy \mathcal{M}_0 of (M, π, Δ^*) is represented by a pseudo-periodic map, i.e., it is of elliptic or parabolic type in the sense of Bers.*

Proof. This is proved by the distance decreasing property of holomorphic maps with respect to Kobayashi distances (see McMullen [37]).

Denote by \mathbb{H} the upper half-plane $\{\tau \in \mathbb{C} \mid \text{Im } \tau > 0\}$. Setting $\rho_0(\tau) = \exp(2\pi i \tau)$, we have a universal covering $\rho_0: \mathbb{H} \rightarrow \Delta^*$ with covering transformation group $\Gamma_0 = \langle \gamma_0 \rangle$, where $\gamma_0 \in \text{Aut}(\mathbb{H})$ is defined by $\gamma_0(\tau) = \tau + 1$.

Let $\Phi: \mathbb{H} \rightarrow \mathcal{T}(R)$ be a holomorphic representation of (M, π, Δ^*) into the Teichmüller space $\mathcal{T}(R)$ of a Riemann surface R of type (g, n) . By the definition of Φ_* we have

$$\Phi \circ \gamma_0 = \Phi_*(\gamma_0) \circ \Phi.$$

Since the Teichmüller distance d_τ on $\mathcal{T}(R)$ coincides with the Kobayashi distance, we have

$$d_\tau(\Phi(\tau), \Phi(\tau')) \leq d_{\mathbb{H}}(\tau, \tau')$$

for any $\tau, \tau' \in \mathbb{H}$, where $d_{\mathbb{H}}$ is the Poincaré distance on \mathbb{H} . Thus for any positive integer n we obtain

$$\begin{aligned} a(\Phi_*(\gamma_0)) &\leq d_\tau(\Phi(in), \Phi_*(\gamma_0) \circ \Phi(in)) \\ &= d_\tau(\Phi(in), \Phi(in+1)) \\ &\leq d_{\mathbb{H}}(in, in+1). \end{aligned}$$

It is easy to see that $\lim_{n \rightarrow \infty} d_{\mathbb{H}}(in, in+1) = 0$, and $a(\Phi_*(\gamma_0)) = 0$. This shows that $\Phi_*(\gamma_0)$ is of elliptic or parabolic type. As stated in §3.4 and 3.5 $\Phi_*(\gamma_0)$ is induced by a pseudo-periodic map. Therefore $\mathcal{M}_0 = \Phi_*(\gamma_0)^{-1}$ is also represented by a pseudo-periodic map. \square

7.2 Holomorphic extension of the moduli map

Proposition 7.2. *The moduli map $J: \Delta^* \rightarrow M_{g,n}$ defined by $J(t) = [S_t]$ has a holomorphic extension $\hat{J}: \Delta \rightarrow \hat{M}_{g,n}$. Here $\hat{M}_{g,n}$ is the Deligne–Mumford compactification of the moduli space $M_{g,n}$ of Riemann surfaces of analytically finite type (g, n) (see Deligne and Mumford [12], and Bers [9]). Moreover*

- (1) \mathcal{M}_0 is of finite order if and only if $\hat{J}(0) \in M_{g,n}$,
- (2) \mathcal{M}_0 is of infinite order if and only if $\hat{J}(0) \in \partial M_{g,n} (= \hat{M}_{g,n} \setminus M_{g,n})$.

Proof. Let ε be a sufficiently small positive constant satisfying $\delta = 2 \tanh^{-1} \varepsilon < \tanh^{-1}(1/3)$. Denote by $\Delta(\varepsilon)$ the disk $\{t \in \mathbb{C} \mid |t| < \varepsilon\}$, and by $\Delta(\varepsilon)^*$ the punctured disk $\Delta(\varepsilon) \setminus \{0\}$.

First assume that \mathcal{M}_0 is trivial. Then we may assume that the representation Φ of (M, π, Δ^*) is a holomorphic map of Δ^* into $\mathcal{T}(R)$.

Fix a point t_0 with $0 < |t_0| < \varepsilon$. The distance decreasing property of holomorphic maps with respect to Carathéodory distances implies

$$\begin{aligned} c_\tau(\Phi(t_0), \Phi(t)) &\leq c_{\Delta^*}(t_0, t) \\ &\leq c_{\Delta^*}(0, t_0) + c_{\Delta^*}(0, t) \\ &\leq \delta (< 2 \tanh^{-1}(1/3)) \end{aligned}$$

for any t with $0 < |t| < \varepsilon$. From Lemma 3.1, Φ maps $\Delta(\varepsilon)^*$ into $B_c(\Phi(t_0), \delta) \Subset \mathcal{T}(R)$. Since $\mathcal{T}(R)$ is biholomorphically equivalent to a bounded domain in \mathbb{C}^{3g-3+n}

we see that $\Phi: \Delta(\delta)^* \rightarrow B_c(\Phi(t_0), \delta) \subseteq \mathcal{T}(R)$ has a holomorphic extension $\hat{\Phi}: \Delta(\varepsilon) \rightarrow \mathcal{T}(R)$ (cf. Earle [13]). Hence $J = \Pi \circ \Phi: \Delta^* \rightarrow \mathcal{T}(R)/\text{Mod}(R) \cong M_{g,n}$ has a holomorphic extension $\hat{J} = \Pi \circ \hat{\Phi}: \Delta \rightarrow M_{g,n}$.

Next assume that \mathcal{M}_0 is of finite order m . Denote by G_m a finite subgroup of $\text{Aut}(\Delta^*)$ generated by $g_m(t) = t \exp(2\pi i/m)$. Setting $\rho_m(t) = t^m$, we have a covering map $\rho_m: \Delta^* \rightarrow \Delta^*$ with covering transformation group G_m . Then we see that $\Phi_m = \Phi \circ \rho_m: \Delta^* \rightarrow \mathcal{T}(R)$ is a holomorphic map and $\Phi_m \circ g_m = \mathcal{M}_0^{-1} \circ \Phi_m$. By the same argument as the case $\mathcal{M}_0 = \text{id}$ we conclude that Φ_m has a holomorphic extension $\hat{\Phi}_m: \Delta \rightarrow \mathcal{T}(R)$. Hence J has a holomorphic extension $\hat{J}: \Delta \rightarrow M_{g,n}$. Note that $\hat{\Phi}_m(0)$ is a fixed point of \mathcal{M}_0 .

Last we suppose that \mathcal{M}_0 is of infinite order. Since $\hat{M}_{g,n}$ is compact, we can find a sequence $\{t_j\}$ in $\Delta^* \cap \mathbb{R}$ such that $J(t_j) = [S_{t_j}] \in M_{g,n}$ converges to $[S_0] \in \hat{M}_{g,n}$ as $j \rightarrow \infty$.

Case 1. Suppose that $[S_0] \in M_{g,n}$. Put $p_0 = [R, f_0, S_0] \in \mathcal{T}(R)$. Take a positive constant δ so that two points $[R, f_1, S_1], [R, f_2, S_2] \in B_c(p_0, \delta)$ are equivalent under $\text{Mod}(R)$, i.e., $[R, f_2, S_2] = [\omega]_*([R, f_1, S_1])$ for some $[\omega] \in \text{Mod}(R)$, if and only if $[\omega] = [f_0^{-1} \circ h \circ f_0]$ for some $h \in \text{Aut}(S_0)$. Set $\tilde{t}_j = (\log t_j)/(2\pi i)$. We may assume that $\Phi(\tilde{t}_1) \in B_c(p_0, \delta/2)$ and $c_{\mathbb{H}}(\tilde{t}_1, \gamma_0(\tilde{t}_1)) < \delta/2$. Then

$$\begin{aligned} c_{\tau}(p_0, \Phi_*(\gamma_0) \circ \Phi(\tilde{t}_1)) &= c_{\tau}(p_0, \Phi \circ \gamma_0(\tilde{t}_1)) \\ &\leq c_{\tau}(p_0, \Phi(\tilde{t}_1)) + c_{\tau}(\Phi(\tilde{t}_1), \Phi \circ \gamma_0(\tilde{t}_1)) \\ &\leq c_{\tau}(p_0, \Phi(\tilde{t}_1)) + c_{\mathbb{H}}(\tilde{t}_1, \gamma_0(\tilde{t}_1)) \\ &< \delta. \end{aligned}$$

Hence $\Phi_*(\gamma_0) = [f_0^{-1} \circ h \circ f_0]$ for some $h \in \text{Aut}(S_0)$, which is a contradiction, for h is of finite order and $\Phi_*(\gamma_0)$ is of infinite order.

Case 2. Suppose now that $[S_0] \in \partial M_{g,n} = \hat{M}_{g,n} \setminus M_{g,n}$. Fix a deformation $\sigma_0: R \rightarrow S_0$. Let $[\sigma_0]_*: \mathcal{T}(R) \rightarrow D(S_0)$ be the allowable map induced by σ_0 , i.e., $[\sigma_0]_*([R, f, S]) = \langle S, \sigma_0 \circ f^{-1}, S_0 \rangle$. Then we have a holomorphic map $\Psi: \mathbb{H} \rightarrow D(S_0)$.

We take a positive constant δ such that two points $\langle S_1, \sigma_1, S_0 \rangle, \langle S_2, \sigma_2, S_0 \rangle \in B_{D(S_0)}(p_0, \delta)$ are equivalent under $\text{Mod}(S_0)$, i.e., $\langle S_2, \sigma_2, S_0 \rangle = \langle \alpha \rangle_* (\langle S_1, \sigma_1, S_0 \rangle)$ for some $\langle \alpha \rangle \in \text{Mod}(S_0)$, if and only if $\langle \alpha \rangle \in \text{Mod}_0(S_0)$, where $B_{D(S_0)}(p_0, \delta) = \{p \in D(S_0) \mid c_{D(S_0)}(p_0, p) < \delta\}$. The same argument as Case 1 implies that we can find $\langle \alpha_0 \rangle \in \text{Mod}_0(S_0)$ so that

$$\Psi \circ \gamma_0 = \langle \alpha_0 \rangle_* \circ \Psi.$$

Let m be the order of $\langle \alpha_0 \rangle_*$. Then we have a holomorphic map $\Psi_m = \Psi \circ \rho_m: \Delta^* \rightarrow D(S_0)$. By a similar argument to the case where \mathcal{M}_0 is trivial, it is proved that Ψ_m has a holomorphic extension $\hat{\Psi}_m: \Delta \rightarrow D(S_0)$, which shows $J: \Delta^* \rightarrow M_{g,n}$ has a holomorphic extension $\hat{J}: \Delta \rightarrow \hat{M}_{g,n}$ with $\hat{J}(0) = [S_0] \in \partial M_{g,n}$. □

Remark 7.3. In the Teichmüller space $\mathcal{T}_B(R)$ this proposition means the following (see [19]). Let $\mathcal{D} = \{t \in \Delta \mid 0 \leq \theta_1 < \arg t < \theta_2 < 2\pi\}$. Then the holomorphic representation $\Phi(t)$ converges to $\varphi_0 \in \overline{\mathcal{T}_B(R)}$ uniformly, as t tends to zero through \mathcal{D} . Furthermore,

- (1) \mathcal{M}_0 is of finite order if and only if $\varphi_0 \in \mathcal{T}_B(R)$. In this case, φ_0 is a fixed point of \mathcal{M}_0 .
- (2) \mathcal{M}_0 is of infinite order if and only if $\varphi_0 \in \partial\mathcal{T}_B(R)$. In this case, φ_0 corresponds to a regular b-group, i.e., its quotient space is a Riemann surface of type (g, n) with $k \geq 1$ nodes, and φ_0 is a fixed point of \mathcal{M}_0 in the augment space of $\mathcal{T}_B(R)$.

7.3 Negativity of the topological monodromy

Proposition 7.4. \mathcal{M}_0 is a pseudo-periodic map of negative type.

Proof. Assume that \mathcal{M}_0 is of infinite order. Denote by $\hat{\Psi}_m: \Delta \rightarrow D(S_0)$ the holomorphic map with $\hat{\Psi}_m(0) = \langle S_0, \text{id}, S_0 \rangle$ which is constructed in the proof of Proposition 7.2. Let p_1, \dots, p_r be nodes of S_0 , and let $\sigma_0: R \rightarrow S_0$ be a deformation.

Recall the parametrization (α, τ) on a neighborhood of $\langle S_0, \text{id}, S_0 \rangle$ in $D(S_0)$ (see §6.3). Then we may assume that the τ_j component of the map $\hat{\Psi}_m$ is given by $t \mapsto t^{k_j}$ for some positive integer k_j . Let ω_j be a k_j -th power of the Dehn twist of negative type about a Jordan curve $\sigma_0^{-1}(p_j)$. Then Example 4.10 implies that \mathcal{M}_0^m is induced by $\omega_1 \circ \dots \circ \omega_r$. Therefore, \mathcal{M}_0 is of negative type. \square

7.4 Completion of (M, π, Δ^*)

We state the following two propositions. Their proofs were given in [19], where fiber spaces of Teichmüller spaces and fiber spaces of deformation spaces of Riemann surfaces with nodes were used. For these fiber spaces we refer to Bers [5], [7], [8], [9], and [10].

Proposition 7.5. (M, π, Δ^*) has a completion $(\hat{M}, \hat{\pi}, \Delta)$, that is, \hat{M} is a two-dimensional normal complex analytic space, $\hat{\pi}: \hat{M} \rightarrow \Delta$ is a holomorphic map, and (M, π, Δ^*) is holomorphically equivalent to $(\hat{M} \setminus S_0, \hat{\pi}|_{\hat{M} \setminus S_0}, \Delta^*)$, where $S_0 = \hat{\pi}^{-1}(0)$.

Proposition 7.6. \mathcal{M}_0 is trivial if and only if S_0 is a non-singular fiber.

8 Proof of Theorem B

Theorem B. Let $f: \Sigma \rightarrow \Sigma$ be a pseudo-periodic map of negative type of an oriented topological surface Σ of type (g, n) onto itself. Then there exists a holomorphic family

(M, π, Δ^*) of Riemann surfaces of type (g, n) over the punctured disk Δ^* whose topological monodromy \mathcal{M}_0 is conjugate to the isotopy class $[f]$ in the mapping class group of Σ .

We will give a proof of Theorem B. Let $f: \Sigma \rightarrow \Sigma$ be a pseudo-periodic map of negative type of a topological surface Σ of type (g, n) with $2g - 2 + n > 0$. We may assume that f is periodic, or that there exists an admissible system of simple closed curves $\mathcal{C} = \{C_1, \dots, C_r\}$ such that for each i , (1) $f(C_i) = C_j$ for some j , and (2) for some positive integer n_i , $f^{n_i}|_{C_i}$ is the ℓ_i -th power of the negative Dehn twist about C_i .

8.1 Case 1: f is a periodic map

We may assume that Σ is a Riemann surface of analytically finite type (g, n) such that f is a biholomorphic map f of order m . This is an immediate consequence of Nielsen’s realization problem (cf. Kerckhoff [28], Wolpert [52]).

We give a proof of Theorem B in this case which is due to Bers [6], Theorem 1. We may assume that Σ is a surface which is obtained by removing n distinct points from a smooth closed surface of genus g . Then take a Riemannian metric ds^2 on Σ so that near every puncture of Σ this metric ds^2 is represented in a form

$$ds^2 = \left(\frac{1}{\log |z|} \left| \frac{dz}{z} \right| \right)^2.$$

Let

$$ds_0^2 = \sum_{j=1}^m (f^*)^j (ds^2)$$

and let R_0 be a Riemann surface of analytically finite type (g, n) induced by isothermal coordinates for the metric ds_0^2 . Then $f: R_0 \rightarrow R_0$ is conformal with respect to the metric ds_0^2 , and so f is biholomorphic.

Now we put $M_0 = \Delta^* \times R_0$ and define the biholomorphic maps

$$\begin{aligned} g_m: \Delta^* &\rightarrow \Delta^*, & g_m(t) &= (\exp(2\pi i/m))t, \\ F_0: M_0 &\rightarrow M_0, & F_0(t, p) &= (g_m(t), f(p)). \end{aligned}$$

We consider the quotient space $M = M_0/\langle F_0 \rangle$ and the projection $\pi: M \rightarrow \Delta^*/\langle g_m \rangle \cong \Delta^*$ given by $\pi([t, p]) = [t]$. Then (M, π, Δ^*) is a holomorphic family of Riemann surfaces of type (g, n) with given monodromy f .

8.2 Case 2: f is a product of negative Dehn twists

We assume that f is a product of ℓ_i -th powers of negative Dehn twists about C_i , $i = 1, \dots, r$. Construct a Riemann surface S_0 with nodes p_1, \dots, p_r by shrink-

ing C_1, \dots, C_r into points. Let $(\alpha, \tau) = (\alpha_1, \dots, \alpha_N, \tau_1, \dots, \tau_r)$ be coordinates at (S_0, id, S_0) in the deformation space $D(S_0)$ of S_0 (see §6.3). For any $t \in \Delta^*$, denote by S_t the Riemann surface represented by the coordinates $(\alpha, \tau) = (0, \dots, 0, t^{\ell_1}, \dots, t^{\ell_r})$. We construct $M = \coprod_{t \in \Delta^*} \{t\} \times S_t$ and the projection $\pi: M \rightarrow \Delta^*$ defined by $\pi(t, p) = t$. Then it is shown that (M, π, Δ^*) is a holomorphic family of Riemann surfaces of type (g, n) with given monodromy f (cf. Earle and Sipe [14]).

8.3 Case 3: f is a generic pseudo-periodic map of negative type

We may assume that f is completely reduced by $\mathcal{C} = \{C_1, \dots, C_r\}$. The connected components of $\Sigma \setminus \mathcal{C}$ can be denoted by $\Sigma_{\alpha\beta}$, $\alpha = 1, \dots, \alpha_0$, $\beta = 1, \dots, N_\alpha$ in such a way that

$$f(\Sigma_{\alpha\beta}) = \Sigma_{\alpha, \beta+1},$$

where we agree, once and for all, that

$$\Sigma_{\alpha, N_\alpha+1} = \Sigma_{\alpha 1}.$$

Note that $\Sigma_{\alpha\beta}$ is of type $(g_{\alpha\beta}, n_{\alpha\beta})$ with $2g_{\alpha\beta} - 2 + n_{\alpha\beta} > 0$ and $f^{N_\alpha}|_{\Sigma_{\alpha 1}}: \Sigma_{\alpha 1} \rightarrow \Sigma_{\alpha 1}$ is periodic. As in §8.1 we may assume that for any $\alpha = 1, \dots, \alpha_0$ the topological surface $\Sigma_{\alpha 1}$ is a Riemann surface of analytically finite type $(g_{\alpha 1}, n_{\alpha 1})$ and $f^{N_\alpha}|_{S_{\alpha 1}}: S_{\alpha 1} \rightarrow S_{\alpha 1}$ is an isomorphism. Then we may also assume that for every $\beta = 1, \dots, N_\alpha$, the topological surface $\Sigma_{\alpha\beta}$ is a Riemann surface $S_{\alpha\beta}$ of analytically finite type $(g_{\alpha\beta}, n_{\alpha\beta})$ and $f|_{S_{\alpha\beta}}: S_{\alpha\beta} \rightarrow S_{\alpha, \beta+1}$ is an isomorphism. Hence we can construct a Riemann surface S_0 of type (g, n) with r nodes p_1, \dots, p_r , a continuous map $\omega_0: \Sigma \rightarrow S_0$ and an isomorphism $f_0: S_0 \rightarrow S_0$ satisfying the following conditions:

- (1) $C_i = \omega_0^{-1}(p_i)$ for all $i = 1, \dots, r$.
- (2) $\omega_0: \Sigma_{\alpha\beta} \rightarrow S_{\alpha\beta}$ is homeomorphic for all $\alpha = 1, \dots, \alpha_0$ and $\beta = 1, \dots, N_\alpha$.
- (3) $\omega_0 \circ f = f_0 \circ \omega_0$.

Let us decompose the admissible system of curves C_1, \dots, C_r into C_{ij} , $i = 1, \dots, r_0$, $j = 1, \dots, r_i$ so that

$$f(C_{ij}) = C_{i, j+1},$$

where we agree, once and for all, that

$$C_{i, r_i+1} = C_{i1}.$$

Let m_i be a minimal positive integer such that $f^{m_i}(\overrightarrow{C_{i1}}) = \overrightarrow{C_{i1}}$, and let n_i be a minimal positive integer such that $f^{n_i}|_{\mathcal{A}_{i1}}$ is the ℓ_i -th power of the negative Dehn twist about C_{i1} , where \mathcal{A}_{i1} is an annular neighborhood of C_{i1} . Then the screw number $s(C_{ij})$ of f at C_{ij} is given by

$$s(C_{ij}) = -\frac{m_i}{n_i} \ell_i$$

for any $i = 1, \dots, r_0$ and $j = 1, \dots, r_i$.

Denote by n_0 the least common multiple of n_1, \dots, n_{r_0} and set $n'_i = n_0/n_i$ for $i = 1, \dots, r_0$.

For each $i = 1, \dots, r_0$ we define a holomorphic map $\Psi_i: \Delta \rightarrow \Delta^{r_i}$ by

$$\Psi_i(t) = (A_{i1}, A_{i2}, \dots, A_{im'_i}) t^{n'_i \ell_i},$$

where $A_{i1} = 1$, and $A_{i2}, \dots, A_{im'_i}$ are complex numbers with $|A_{i2}| = \dots = |A_{im'_i}| = 1$, and

$$\begin{cases} m'_i = m_i & (C_{i1} \text{ is non-amphidrome}), \\ m'_i = \frac{m_i}{2} & (C_{i1} \text{ is amphidrome}). \end{cases}$$

Take holomorphic coordinates $(\alpha, \tau) \in \Delta_0^N \times \Delta^r$ at $\langle S_0, \text{id}, S_0 \rangle$ in the deformation space $D(S_0)$, which is constructed in §6.3. Here $N = 3g - 3 + n - r$, and Δ_0 is a sufficiently small disk $\{t \in \mathbb{C} \mid |t| < \varepsilon\}$.

Let $\Psi: \Delta \rightarrow \Delta_0^N \times \Delta^r$ be a holomorphic map given by

$$\Psi(t) = (\underbrace{0, \dots, 0}_{n \text{ times}}, \Psi_1(t), \dots, \Psi_{r_0}(t)).$$

Denote by g_{n_0} an element of $\text{Aut}(\Delta)$ sending t into $t \exp(2\pi i/n_0)$. Then we have the following

Lemma 8.1. *There exist constants A_{ij} so that*

$$\Psi \circ g_{n_0} = \langle f_0 \rangle_* \circ \Psi \quad \text{on } \Delta.$$

Proof. We recall the representation in §6.3 of the action $\langle f_0 \rangle_*$ in a neighborhood of $\langle S_0, \text{id}, S_0 \rangle$. For any $i = 1, \dots, r_0$ and $j = 1, \dots, m'_i$, put $p_{ij} = \omega_0(C_{ij})$ and take holomorphic coordinates z_{ij}, w_{ij} at p_{ij} such that

$$\begin{aligned} z_{i,j+1} &= f_0(z_{ij}) = \xi_{ij} z_{ij}, \\ w_{i,j+1} &= f_0(w_{ij}) = \chi_{ij} w_{ij}. \end{aligned}$$

First we assume C_{i1} is non-amphidrome. Then $A_{i1}, A_{i2}, \dots, A_{im'_i}$ must satisfy

$$\begin{aligned} A_{i1} e_0^{n'_i \ell_i} &= \xi_{im'_i} \chi_{im'_i} A_{im'_i}, \\ A_{i2} e_0^{n'_i \ell_i} &= \xi_{i1} \chi_{i1} A_{i1}, \\ A_{i3} e_0^{n'_i \ell_i} &= \xi_{i2} \chi_{i2} A_{i2}, \\ &\vdots \\ A_{im'_i} e_0^{n'_i \ell_i} &= \xi_{i,m'_i-1} \chi_{i,m'_i-1} A_{i,m'_i-1}. \end{aligned}$$

This implies

$$\begin{aligned}
 A_{i2} &= \xi_{i1} \chi_{i1} e_0^{-n'_i \ell_i}, \\
 A_{i3} &= \xi_{i2} \chi_{i2} e_0^{-n'_i \ell_i} A_{i2}, \\
 &\vdots \\
 A_{im_i} &= \xi_{i,m_i-1} \chi_{i,m_i-1} e_0^{-n'_i \ell_i} A_{i,m_i-1} \\
 &= (\xi_{im_i} \chi_{im_i} e_0^{-n'_i \ell_i})^{-1}.
 \end{aligned}$$

Hence we need the compatibility condition

$$(\xi_{i1} \cdots \xi_{im_i}) (\chi_{i1} \cdots \chi_{im_i}) e_0^{-m_i n'_i \ell_i} = 1. \quad (8.1)$$

By the definition of screw number and valency, we have

$$(\xi_{i1} \cdots \xi_{im_i}) (\chi_{i1} \cdots \chi_{im_i}) e_0^{-m_i n'_i \ell_i} = \exp(2\pi i K_i),$$

where

$$K_i = -s(C_{i1}) - \frac{\delta'_{i1}}{\lambda'_{i1}} - \frac{\delta''_{i1}}{\lambda''_{i1}}.$$

Since K_i is an integer from (5.1) of Proposition 5.1, we have the compatibility condition (8.1).

Next we assume C_{i1} is amphidrome. Then $A_{i1}, A_{i2}, \dots, A_{im'_i}$ must satisfy

$$\begin{aligned}
 A_{i1} e_0^{n'_i \ell_i} &= \xi_{im'_i} \chi_{im'_i} A_{im'_i}, \\
 A_{i2} e_0^{n'_i \ell_i} &= \xi_{i1} \chi_{i1} A_{i1}, \\
 A_{i3} e_0^{n'_i \ell_i} &= \xi_{i2} \chi_{i2} A_{i2}, \\
 &\vdots \\
 A_{im'_i} e_0^{n'_i \ell_i} &= \xi_{i,m'_i-1} \chi_{i,m'_i-1} A_{i,m'_i-1}.
 \end{aligned}$$

Then we have

$$\begin{aligned}
 A_{i2} &= \xi_{i1} \chi_{i1} e_0^{-n'_i \ell_i}, \\
 A_{i3} &= \xi_{i2} \chi_{i2} e_0^{-n'_i \ell_i} A_{i2}, \\
 &\vdots \\
 A_{im'_i} &= \xi_{i,m'_i-1} \chi_{i,m'_i-1} e_0^{-n'_i \ell_i} A_{i,m'_i-1} \\
 &= (\xi_{im'_i} \chi_{im'_i} e_0^{-n'_i \ell_i})^{-1}.
 \end{aligned}$$

In this case the compatibility condition is

$$(\xi_{i1} \cdots \xi_{im'_i}) (\chi_{i1} \cdots \chi_{im'_i}) e_0^{-m'_i n'_i \ell_i} = 1. \quad (8.2)$$

By the definition of screw number and valency, we have

$$(\xi_{i1} \cdots \xi_{im'_i}) (\chi_{i1} \cdots \chi_{im'_i}) e_0^{-m'_i n'_i \ell_i} = \exp(2\pi i K_i),$$

where

$$K_i = \frac{1}{2} \left(-s(C_{i1}) - \frac{\delta(C'_{i1})}{\lambda(C'_{i1})} - \frac{\delta(C''_{i1})}{\lambda(C''_{i1})} \right).$$

From (5.2) we obtain the compatibility condition (8.2). \square

Now we can construct a holomorphic family (M, π, Δ^*) of Riemann surfaces of type (g, n) with given monodromy f as follows.

For any $t \in \Delta^*$, let S_t be the Riemann surface of (g, n) which is represented by the coordinates $(\alpha, \tau) = \Psi(t)$. We set

$$M_0 = \coprod_{t \in \Delta^*} \{t\} \times S_t.$$

This becomes a two-dimensional complex manifold. For every $t \in \Delta^*$, the isomorphism $f_0: S_0 \rightarrow S_0$ induces a biholomorphic map $F_0(t, \cdot): S_t \rightarrow S_{g_{n_0}(t)}$. Then the map $\mathcal{F}_0: M_0 \rightarrow M_0$ given by $\mathcal{F}_0(t, p) = (g_{n_0}(t), F_0(t, p))$ is biholomorphic. We consider the quotient space $M = M_0 / \langle \mathcal{F}_0 \rangle$ and the projection $\pi: M \rightarrow \Delta^* / \langle g_{n_0} \rangle \cong \Delta^*$ defined by $\pi([t, p]) = [t]$. Then (M, π, Δ^*) is a desired holomorphic family of Riemann surfaces of type (g, n) which have the given monodromy f .

References

- [1] M. Asada, Y. Matsumoto and T. Oda, Local monodromy on the fundamental groups of algebraic curves along a degenerate stable curves. *J. Pure Appl. Algebra* 103 (1995), 235–283. [95](#)
- [2] T. Ashikaga and M. Ishizaka, Classification of degenerations of curves of genus three via Matsumoto-Montesinos's theory. *Tôhoku Math. J.* **54** (2002), 195–226.
- [3] M. F. Atiyah, The signature of fibre-bundles. In *Global Analysis, Papers in honor of K. Kodaira*, (D. C. Spencer and S. Iyanaga, eds.), University of Tokyo Press and Princeton University Press, Tokyo and Princeton, N.J., 1969, 73–84. [96](#)
- [4] W. Barth, C. Peters, and A. Van de Ven, *Compact Complex Surfaces*. *Ergeb. Math. Grenzgeb.* 4, Springer-Verlag, Berlin 1984. [98](#)
- [5] L. Bers, Fiber spaces over Teichmüller spaces. *Acta Math.* 130 (1973), 89–126. [98](#)
- [6] L. Bers, An extremal problem for quasiconformal mappings and a theorem by Thurston. *Acta Math.* 141 (1978), 73–98. [122](#)

- [7] L. Bers, On spaces of Riemann surfaces with nodes. *Bull. Amer. Math. Soc.* 80 (1974), 1219–1222. [102](#), [103](#), [110](#), [123](#)
- [8] L. Bers, Spaces of degenerating Riemann surfaces. In *Discontinuous Groups and Riemann Surfaces, 1973 Maryland Conference* (L. Greenberg, ed.), Ann. of Math. Stud. 79, Princeton University Press, Princeton, N.J., 1974, 43–55. [116](#), [122](#)
- [9] L. Bers, Deformations and moduli of Riemann surfaces with nodes and signatures. *Math. Scand.* 36 (1975), 12–16. [116](#), [122](#)
- [10] L. Bers, Finite dimensional Teichmüller spaces and generalizations. *Bull. Amer. Math. Soc. (N.S.)* 5 (1981), 131–172. [116](#), [120](#), [122](#)
- [11] C. H. Clemens, Picard-Lefschetz theorem for families of nonsingular algebraic varieties acquiring ordinary singularities. *Trans. Amer. Math. Soc.* 136 (1969), 93–108. [116](#), [122](#)
- [12] P. Deligne and D. Mumford, The irreducibility of the space of curves of given genus. *Inst. Hautes Études Sci. Publ. Math.* 36 (1969), 75–110. [95](#)
- [13] C. J. Earle, On the Carathéodory metric in Teichmüller spaces. In *Discontinuous groups and Riemann surfaces* (L. Greenberg, ed.), Ann. of Math. Stud. 79 (1974), 99–103. [120](#)
- [14] C. J. Earle, and P. L. Sipe, Families of Riemann surfaces over the punctured disk. *Pacific J. Math.* 150 (1991), 79–86. [121](#)
- [15] J. D. Fay, *Theta Functions on Riemann Surfaces*. Lecture Notes in Math. 352, Springer-Verlag, Berlin 1973. [95](#), [124](#)
- [16] J. Gilman, On the Nielsen Type and the Classification for the Mapping Class Group. *Adv. Math.* 40 (1981), 68–96. [116](#)
- [17] P. A. Griffiths, Complex analytic properties of certain Zariski open sets on algebraic varieties. *Ann. of Math.* 94 (1971), 21–55. [116](#)
- [18] W. J. Harvey (ed.), *Discrete Groups and Automorphic Functions*. 1975 Cambridge Institutional Conference, Academic Press, London 1977. [98](#)
- [19] Y. Imayoshi, Holomorphic families of Riemann surfaces and Teichmüller spaces. In *Riemann Surfaces and Related Topics, 1978 Stony Brook Conference* (I. Kra and B. Maskit, eds.), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 277–300. [105](#)
- [20] Y. Imayoshi, A construction of holomorphic families of Riemann surfaces over the punctured disk with given monodromy. In *Proceedings of the 12th International Conference on Finite or Infinite Dimensional Complex Analysis and Applications* (H. Kazama, M. Morimoto and C-C. Yang, eds.), Kyushu University Press, Fukuoka 2005, 97–104. [95](#), [96](#), [119](#), [122](#)
- [21] Y. Imayoshi, M. Ito and H. Yamamoto, The Nielsen-Thurston-Bers type of elements of the monodromy group of a holomorphic family of Riemann surfaces induced by a Kodaira surface I: Riera's example. In *Finite or Infinite Dimensional Complex Analysis*, ed. by Joji Kajiwara, Zhong Li, and Kwang Ho Shon, Lecture Notes in Pure Appl. Math. 214, Marcel Dekker, New York 2000, 169–177, [95](#)
[109](#)
- [22] Y. Imayoshi, M. Ito and H. Yamamoto, On the Nielsen-Thurston-Bers type of some self-maps of Riemann surfaces with two specified points. *Osaka J. Math.* 40 (2003), 27–47. [108](#)

- [23] Y. Imayoshi, M. Ito and H. Yamamoto, A reducibility problem for monodromy of some surface bundles. *J. Knot Theory Ramifications* 13 (2004) 597–616. [108](#)
- [24] Y. Imayoshi and H. Shiga, A finiteness theorem of holomorphic families of Riemann surfaces, in *Holomorphic Functions and Moduli II* (D. Drain et al., eds.), Math. Sci. Res. Inst. Publ. 11, Springer-Verlag, New York 1988, 207–219. [104](#)
- [25] Y. Imayoshi and M. Taniguchi, *An Introduction to Teichmüller Spaces*. Springer-Verlag, Tokyo 1992. [98](#)
- [26] J. Jost and S.-T. Yau, Harmonic mappings and algebraic varieties over function fields. *Amer. J. Math.* 115 (1993), 1197–1227. [104](#)
- [27] A. Kas, On deformations of a certain type of irregular algebraic surface. *Amer. J. Math.* 90 (1968), 789–804. [98](#)
- [28] S.P. Kerckhoff, The Nielsen realization problem. *Ann. Math.* 117 (1983), 235–265. [123](#)
- [29] S. Kobayashi, *Hyperbolic manifolds and holomorphic mappings*. Pure Appl. Math. 2, Marcel Dekker Inc., New York 1970. [100](#)
- [30] S. Kobayashi, *Hyperbolic complex spaces*. Grundlehren Math. Wiss. 318, Springer-Verlag, Berlin 1998. [100](#), [101](#)
- [31] K. Kodaira, On compact complex surfaces II. *Ann. of Math.* 77 (1963), 563–626. [96](#), [105](#)
- [32] K. Kodaira, A certain type of irregular algebraic surfaces. *J. d'Analyse Math.* 19 (1967), 207–215. [98](#)
- [33] I. Kra, On the Nielsen-Thurston-Bers type of some self-maps of Riemann surfaces. *Acta Math.* 146 (1981), 231–270. [108](#), [109](#)
- [34] I. Kra, Horocyclic coordinates for Riemann surfaces and moduli spaces I: Teichmüller and Riemann spaces of Kleinian groups. *J. Amer. Math. Soc.* 3 (1990), 497–578. [116](#)
- [35] Y. Matsumoto and J. M. Montesinos-Amilibia, Pseudo-periodic homomorphisms and degeneration of Riemann surfaces I, II. Preprints, Univ. of Tokyo and Univ. Complutense de Madrid, 1991/1992. [95](#)
- [36] Y. Matsumoto and J. M. Montesinos-Amilibia, Pseudo-periodic homomorphisms and degeneration of Riemann surfaces. *Bull. Amer. Math. Soc. (N.S.)* 30 (1994), 70–75. [95](#), [96](#), [110](#), [111](#), [115](#), [116](#)
- [37] C. T. McMullen, From dynamics on surfaces to rational points on curves. *Bull. Amer. Math. Soc.* 37 (2000), 119–140. [104](#), [119](#)
- [38] G. Mondello, Riemann surfaces, ribbon graphs and combinatorial classes. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich, 2009, 151–215. [96](#)
- [39] Y. Namikawa and K. Ueno, On fibers in families of curves of genus two I, singular fibers of elliptic type. In *Number Theory Algebraic Geometry and Commutative Algebra*, in honor of Y. Akizuki (Y. Kusunoki, S. Mizohata, H. Toda, M. Yamaguchi, H. Yoshizawa, eds.), Kinokuniya, Tokyo 1973, 297–371. [96](#)
- [40] Y. Namikawa and K. Ueno, The complete classification of fibers in pencils of curves of genus two. *Manuscripta Math.* 9 (1973), 143–186. [96](#)
- [41] J. Nielsen, Die Structur periodischer Transformationen von Flächen. *Mat.-Fys. Medd. Danske Vid. Selsk.*, 15 (1937); English transl. in *Collected Papers 2*, Birkhäuser, Boston, Basel, Stuttgart, 1986. [110](#)

- [42] J. Nielsen, Surface transformation classes of algebraically finite type. *Mat.-Fys. Medd. Danske Vid. Selsk.* 21 (1944), 3–89; English transl. in *Collected Papers* 2, Birkhäuser, Boston, Basel, Stuttgart, 1986. [110](#), [115](#)
- [43] G. Riera, Semi-direct products of Fuchsian groups and uniformization. *Duke Math. J.* 44 (1977), 291–304. [98](#)
- [44] H. L. Royden, Automorphisms and isometries of Teichmüller space. In *Advances in the Theory of Riemann Surfaces* (1969 Stony Brook Conference) (L. V. Ahlfors et al., eds.), Ann. Math. Stud. 66, Princeton University Press, Princeton, N. J., 1971, 369–383. [101](#)
- [45] H. Shiga and H. Tanigawa, On the Maskit coordinates of Teichmüller spaces and modular transformations. *Kodai Math. J.* 12 (1989), 437–443. [95](#)
- [46] S. Takamura, Towards the classification of atoms of degenerations, I, (Splitting criteria via configurations of singular fibers. *J. Math. Soc. Japan* 56 (1) (2004), 115–145. [96](#)
- [47] S. Takamura, Towards the classification of atoms of degenerations, II (Cyclic quotient construction of degenerations of complex curves). RIMS Preprint 1344, 2001. [95](#), [96](#)
- [48] S. Takamura, *Splitting Deformations of Degenerations of Complex Curves, Towards the Classification of Atoms of Degenerations, III*. Lecture Notes in Math. 1886, Springer-Verlag, Berlin 2006. [96](#)
- [49] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc.*, 19 (1989), 763–782. [102](#), [110](#)
- [50] K. Uematsu, Numerical classification of singular fibers of genus 3 pencils. *J. Math. Kyoto Univ.* 39 (1999), 417–431. [96](#)
- [51] S. Wolpert, On the homology of the moduli space of stable curves. *Ann. of Math.* 118 (1983), 491–523. [116](#), [118](#)
- [52] S. Wolpert, Geodesic length functions and the Nielsen problem. *J. Differential Geom.* 25 (1987), 275–296. [123](#)

Chapter 4

The uniformization problem

Robert Silhol

Contents

1	Introduction	131
2	The classical strategy for solving the uniformization problem	132
	2.1 The standard differential equation	133
	2.2 The Schwarzian derivative and accessory parameters	136
	2.3 Hyperelliptic curves with many automorphisms	137
3	Geometric methods for the uniformization problem	138
	3.1 Algebraic curves with many automorphisms	138
	3.2 Tiling by quadrangles and other polygons	139
4	Automorphic forms and the inverse uniformization problem	141
	4.1 Poincaré series	143
	4.2 Modular curves	144
	4.3 Other methods	146
	References	147

1 Introduction

As is often the case with classical and difficult problems the uniformization problem has many variants. We will essentially only deal here with one of these: the Fuchsian uniformization, and only of some aspects of this extremely vast subject. In fact the subject is so vast and with so many ramifications in other branches of mathematics that we will barely be able to scratch the surface of things (a complete treatment would need at least a complete volume). Essentially we will deal here with the more elementary aspects of the problem and indicate some references for more advanced developments. On the other hand there are many aspects we have not touched at all. One of these is Schottky uniformization for which we refer to B. Maskit [22] (see also for recent developments [30] and [10]).

The starting point is of course the famous Poincaré–Koebe Theorem [17], [25] (1907), that asserts that if S is a compact Riemann surface of genus $g > 1$ then there is a discrete faithful representation of the fundamental group of S

$$\rho: \pi_1(S) \rightarrow \Gamma \subset \mathrm{PSL}_2(\mathbb{R}) \tag{1.1}$$

and this representation is such that S , as a Riemann surface, is isomorphic to the quotient of the upper half space by the discrete group Γ

$$S \cong \mathbb{H}/\Gamma.$$

Actually we do not need S to be compact, it can have finitely many punctures $\{x_1, \dots, x_n\}$ and the result is also true for genus 0 or 1 provided the Euler characteristic of $S \setminus \{x_1, \dots, x_n\}$ is strictly negative. On the other hand we will restrict to surfaces of finite area, and Fuchsian groups of the first kind i.e. discrete subgroups of $\mathrm{PSL}_2(\mathbb{R})$ with finite co-volume.

In this general setting the Fuchsian uniformization problem can be formulated as follows:

Problem 1. Given a Riemann surface S , defined by an algebraic curve for example, compute the representation ρ .

Intimately linked with Problem 1 is the inverse problem:

Problem 2. Given a Fuchsian group of the first kind Γ compute an algebraic curve isomorphic to \mathbb{H}/Γ .

There are of course many variants and subproblems of great interest, and we will encounter a few, but if we restrict to the Fuchsian uniformization these are the basic questions.

Historically the first attempts to solve Problem 1, or at least a partial version, focused on the question of finding functions, holomorphic on some open set, satisfying algebraic relations. Not surprisingly theta relations provide a host of these and so probably the first solutions to the uniformization problem can be found in the work of Jacobi. This is also the point of view taken by Burnside [6] who found two functions x and y , expressed as rational functions in the Weierstrass functions \wp and \wp' , that satisfy the relation $y^2 = x(x^4 - 1)$, an equation for Bolza's curve. For a modern account see Rankin [27] who also computes the corresponding Fuchsian group.

More recent approaches involve essentially one of the following two methods

- (i) solving the Schwarzian differential equation;
- (ii) constructing a fundamental domain for the Fuchsian group.

We will describe these in Sections 2 and 3.

For Problem 2 the classical method is to compute the space of automorphic forms of given weight and use these to obtain a pluricanonical embedding. This works remarkably well for congruence subgroups and is an extremely active field of research.

2 The classical strategy for solving the uniformization problem

Let S be a Riemann surface of genus $g \geq 2$, which for simplicity we assume to be compact. Let Γ be its Fuchsian group and $\pi : \mathbb{H} \rightarrow \mathbb{H}/\Gamma = S$ the covering map.

Let φ be a local inverse of π defined on some simply connected open subset Ω of S . If φ_1 is another such inverse, then it will differ by an element of Γ , $\varphi_1 = \gamma \varphi$, γ a Möbius transformation in Γ . This leads to the introduction of the Schwarzian derivative

Definition 2.1. Let φ be a C^3 function of a complex variable. *The Schwarzian derivative of φ is*

$$S\varphi(z) = \left(\frac{\varphi''(z)}{\varphi'(z)} \right)' - \frac{1}{2} \left(\frac{\varphi''(z)}{\varphi'(z)} \right)^2. \quad (2.1)$$

The reason for the introduction of this differential operator lies in the next classical lemma (see for example [36] or [24])

Lemma 2.2. (i) $Sf \equiv 0$ if and only if f is a Möbius transformation;

(ii) $S(f \circ g)(z) = (g'(z))^2 Sf(g(z)) + Sg(z)$.

In particular we have that if f is a Möbius transformation, then

(iii) $S(f \circ g)(z) = Sg(z)$.

From this one can show that, although the inverse of π is multi-valued, $S\varphi$ is single valued and defines in fact a meromorphic function on S . Hence if S is an algebraic curve defined by some polynomial equation $P(x, y) = 0$, then $S\varphi = R(x, y)$ for some rational function. So the strategy is to

- (i) compute the rational function R ;
- (ii) solve the equation $S\varphi = R$.

As we will see the difficulty lies with (i) and not with (ii).

2.1 The standard differential equation

By considering an algebraic curve as a ramified cover of the Riemann sphere one can extend the arguments of the next two sections to arbitrary algebraic curves. But to simplify the exposition, and since all the difficulties already appear here, we will concentrate on the simplest case: the case of hyperelliptic curves.

Definition 2.3. A hyperelliptic curve is a ramified double cover of the complex projective plane. If C is hyperelliptic one can always define it by an affine equation of the form

$$y^2 = P(x) = \prod_{i=1}^n (x - x_i) \quad \text{where the } x_i \text{ are distinct.} \quad (2.2)$$

If C is of genus g then the polynomial P will be of degree $2g + 1$ or $2g + 2$ depending on whether the point at infinity is or not a branch point. We assume here

that the curve is projective non-singular and consider Equation (2.2) to be an affine equation of the curve.

In this context the double covering map from the curve to the plane is simply $h: (x, y) \mapsto x$, appropriately extended at infinity.

One reason why this case is more favorable for the uniformization problem is that one has an easy description of the fundamental group by lifting loops in the plane minus the ramification points (see for example [23]).

The Schwarzian differential equation (sometimes called Fuchsian equation) is the second order differential equation

$$y''(z) + \frac{1}{2}S\varphi(z)y(z) = 0. \quad (2.3)$$

Concerning this differential equation we have another classical lemma

Lemma 2.4. *If y_1 and y_2 are two independent solutions of Equation (2.3) defined in a neighborhood of a point q , then $S(y_1/y_2) = S\varphi$. Conversely if $S\psi = S\varphi$, then $\psi = y_1/y_2$ for two independent solutions. Moreover these are unique if we fix $y_1(q)$.*

See for example [36] or [24].

Now fix a point \tilde{q} in the hyperelliptic curve C that is not a Weierstrass point i.e. in the notations of (2.2) is not of the form $(x_i, 0)$ or a point at infinity. Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ be a Fuchsian group for C and let $\rho_0: \pi_1(C, \tilde{q}) \rightarrow \mathrm{PSL}_2(\mathbb{R})$ be the corresponding representation of the fundamental group. We identify C with \mathbb{H}/Γ . Let $\pi: \mathbb{H} \rightarrow \mathbb{H}/\Gamma = C$ be the covering map and $X = h \circ \pi$, where h is the hyperelliptic map extending the map $(x, y) \mapsto x$. The set A of ramification points of X is $A = \{x_1, \dots, x_n\}$ or $A = \{x_1, \dots, x_n, \infty\}$ depending on the parity of n . Finally we set $q = h(\tilde{q})$.

The relevance of Equation (2.3) is the following. Let Ω be a simply connected neighborhood of q in $\hat{C} \setminus A$ and let φ be a branch of the inverse of X defined in Ω . Then, by Lemma 2.4, we can find two independent solutions u and v of Equation (2.3), holomorphic in Ω and such that $u/v = \varphi$. An important point to recall here is that although φ is only locally defined $S\varphi$ is a global meromorphic function with double poles at the ramification points and hence is holomorphic on $\hat{C} \setminus A$.

Let γ be a closed loop in $\hat{C} \setminus A$ based at q and that lifts to a non trivial loop $\tilde{\gamma}$ in C based at \tilde{q} (for example a loop surrounding two ramification points). Using analytic continuation along γ we find two new solutions u_γ and v_γ , holomorphic in a neighborhood of q . Note that u_γ and v_γ only depend of the homotopy class of γ in $\hat{C} \setminus A$ and hence only depend of the homotopy class of $\tilde{\gamma}$ in C .

Since the space of solutions of the differential equation (2.3) is two-dimensional

$$\begin{bmatrix} u_\gamma \\ v_\gamma \end{bmatrix} = A_\gamma \begin{bmatrix} u \\ v \end{bmatrix} \quad (2.4)$$

for some matrix $A_\gamma \in \text{GL}_2(\mathbb{C})$. Moreover if a_{ij} are the coefficients of A_γ we have

$$\frac{u_\gamma}{v_\gamma} = \frac{a_{11}u/v + a_{12}}{a_{21}u/v + a_{22}}. \quad (2.5)$$

On the other hand we can also continue analytically the local inverse $\varphi = u/v$ along γ to obtain an inverse branch φ_γ defined in a neighborhood of q . We of course have

$$\rho_0(\tilde{\gamma})(\varphi) = \varphi_\gamma.$$

By the uniqueness statement of Lemma 2.4 and the uniqueness of analytic continuation we have $\varphi_\gamma = u_\gamma/v_\gamma$. Hence

Proposition 2.5. *The Möbius transformation of (2.5) associated to the matrix A_γ of Formula (2.4) is $\rho_0(\tilde{\gamma}) \in \Gamma \subset \text{PSL}_2(\mathbb{R})$.*

Starting with an arbitrary pair (u_1, v_1) of independent solutions defined in a neighborhood of q we will get with the above method a representation conjugate to ρ_0 . More precisely

Corollary 2.6. *Let (u_1, v_1) be a basis of solutions to the differential equation (2.3) defined in a neighborhood of q . Let γ be as above and let $((u_1)_\gamma, (v_1)_\gamma)$ be solutions obtained from (u_1, v_1) by analytic continuation along γ . Then there exists a Möbius transformation g , independent of γ , such that we have*

$$\begin{pmatrix} (u_1)_\gamma \\ (v_1)_\gamma \end{pmatrix} = g^{-1} \rho_0(\tilde{\gamma}) g \begin{pmatrix} u_1 \\ v_1 \end{pmatrix}$$

Proof. Let (u, v) be independent solutions such that u/v is a local inverse of the covering map X . Let $u = au_1 + bv_1$ and $v = cu_1 + dv_1$.

By the uniqueness of analytic continuation we will also have $u_\gamma = a(u_1)_\gamma + b(v_1)_\gamma$ and $v_\gamma = c(u_1)_\gamma + d(v_1)_\gamma$. Hence by Proposition 2.5 we can take g to be $z \mapsto \frac{az+b}{cz+d}$.

In practice we only need to find a representation of $\pi_1(C, \tilde{q})$ in $\text{PSL}_2(\mathbb{R})$. Technically this can be done as follows. Applying the method above to an arbitrary pair of independent solutions of Equation (2.3) yields a representation ρ of $\pi_1(C, \tilde{q})$ in $\text{PSL}_2(\mathbb{C})$. Let $\gamma_1, \dots, \gamma_n$ be a set of generators of $\pi_1(C, \tilde{q})$. Let x_1, x_2 be the fixed points of $\rho(\gamma_1)$ and let x_3 be the repelling fixed point of $\rho(\gamma_2)$. Then there is a Möbius transformation g sending x_1, x_2, x_3 to $\infty, 0, 1$. Conjugating by g will yield the desired representation in $\text{PSL}_2(\mathbb{R})$.

In conclusion finding a complete solution to the uniformization problem for hyperelliptic curves reduces to solving the differential equation (2.3). Unfortunately the problem now lies in the computation of this equation and more precisely in the computation of the Schwarzian derivative $S\varphi$. This is where other difficulties appear.

2.2 The Schwarzian derivative and accessory parameters

We limit our discussion to elementary aspects of the theory. For far deeper results see for example the papers of L. Takhtajan and P. Zograf [33] and [34]. See also [20] and [16] as well as the abundant literature on the subject.

Let $\varphi = h \circ \pi$ be as above the map from the upper-half plane to the Riemann sphere. Let P be the polynomial of (2.2) and let $x_1, \dots, x_n, n = 2g + 1$ or $2g + 2$, be the distinct roots of P . A local analysis of $S\varphi$ (see for example [24]) shows that $S\varphi$ is in fact a rational function and more precisely of the form

$$S\varphi(z) = \sum \left(\frac{3}{8} \frac{1}{(z - x_i)^2} + \frac{b_i}{z - x_i} \right) \quad (2.6)$$

where the b_i are the so-called accessory parameters (sic! accessory maybe but rather essential for our problem). In the general case the b_i satisfy two (resp. three) relations if $n = 2g + 1$ (resp. $n = 2g + 2$). These are, if $n = 2g + 1$,

- (i) $\sum_{i=1}^n b_i = 0$;
- (ii) $\sum_{i=1}^n x_i b_i + \frac{3}{8} = \frac{3}{8}$.

If $n = 2g + 2$ we must replace (ii) by (ii)' and we have the additional relation (iii):

- (ii)' $\sum_{i=1}^n x_i b_i + \frac{3}{8} = 0$;
- (iii) $\sum_{i=1}^n x_i^2 b_i + \frac{3}{4} x_i = 0$.

See [24, Chap. V] for more details (where the assertion on (ii) is false but can be easily corrected using the arguments given there). It should be noted here that if we consider the general situation of a ramified cover of the sphere one can express $S\varphi$ in terms of the ramification points and the ramification indices. This general expression is very similar to the one given in (2.6) and we again have accessory parameters that satisfy three relations.

Unfortunately these relations are not enough to compute the b_i in the general case. In fact although they have been studied for over a century, these accessory parameters are largely mysterious and to quote Nehari “the determination of the $n - 3$ independent constants [...] is an exceedingly difficult task”.

On the other hand we can note the following. If the curve has an automorphism distinct from the hyperelliptic involution, then such an automorphism is induced by a Möbius transformation fixing globally the roots x_i (and infinity if $n = 2g + 1$). This statement follows easily from the fact that the hyperelliptic involution commutes with all other automorphisms. Thus applying Lemma 2.2 (iii) we obtain additional relations, and in the favorable case when the curve has a “very large” automorphism group (see Definition 2.7 for a precise meaning of this) we will have enough relations to compute the b_i and solve the problem.

2.3 Hyperelliptic curves with many automorphisms

In [35] (1929) E. T. Whittaker conjectured that, for a hyperelliptic curve defined by (2.2)

$$S\varphi(z) = W_P(z) = \frac{3}{8} \left(\left(\frac{P'(z)}{P(z)} \right)^2 - \frac{(2g+2)P''(z)}{(2g+1)P(z)} \right). \quad (2.7)$$

In the following years many examples were found of curves satisfying this conjecture (see the bibliography in [12]) and in 1958 R. A. Rankin substantially enlarged the list of examples by proving in [26] that the conjecture was satisfied if the set of roots of P satisfied certain symmetrical properties. In practice these symmetrical properties ensure that the curve has enough automorphisms to apply the method above and find sufficiently many relations to compute the b_i .

A more conceptual approach is given by E. Girono and G. González-Diez in [12]. For this recall

Definition 2.7. An algebraic curve C is said to have a *very large automorphism group* if one of the following conditions holds

- (i) $C/\text{Aut}(C)$ is the Riemann Sphere and the covering map $C \rightarrow C/\text{Aut}(C)$ is ramified precisely over three points;
- (ii) The space of holomorphic quadratic differentials invariant under $\text{Aut}(C)$ is of dimension 0.

The two conditions in the definition are of course equivalent since the dimension of the space of quadratic differentials invariant under a group G acting by automorphisms on C is $3\tilde{g} - 3 + n$, where \tilde{g} is the genus of C/G and n is the number of points of C/G with ramification index > 1 (see [9, p. 273]).

Since the tangent space to the moduli space at the point defined by the isomorphism class of C is the space of quadratic differentials we can reformulate Definition 2.7 by saying that C has a very large automorphism group if $(C, \text{Aut}(C))$ is an isolated point in moduli space, i.e. the deformations of C will have a smaller automorphism groups.

Lemma 2.8. Let $P(z) = \prod (z - x_i)$ and if M is a Möbius transformation let $M(P)(z) = P \circ M^{-1}(z) = \prod (z - M(x_i))$. Then for W_P as in Formula (2.7) we have

$$W_P(z) = (M'(z))^2 W_{M(P)}(M(z))$$

This can be found by direct computation (see [26], p. 41).

On the other hand we can deduce from Lemma 2.2 that if M is a Möbius transformation and φ is as before then

$$S\varphi(z) = (M'(z))^2 S(\varphi \circ M^{-1})(M(z)). \quad (2.8)$$

Combining Lemma 2.8 and Formula (2.8) we get that if $Q_C = W_P - S\varphi$ and $h: (x, y) \mapsto x$ is the hyperelliptic projection, then $h^*(Q_C dz^2)$ is a quadratic differential on C , invariant under the group $\text{Aut}(C)$. A priori this is a meromorphic

quadratic differential but a local analysis at the poles of Q_C shows that it is in fact holomorphic. The equality $S\varphi = W_P$, and Whittaker's conjecture, immediately follows for curves with a very large automorphism group.

But apart from this (and some scattered exceptional cases — see [12] for a discussion) the conjecture is hopelessly false. The main reason being that, as proved by I. Kra in [20], the coefficients of the rational function $S\varphi$ only depend real analytically, and not holomorphically, on the roots x_1, \dots, x_n of the polynomial P , whereas of course W_P does depend holomorphically (see also [16]). In fact even for a curve for which the space of holomorphic quadratic forms invariant under $\text{Aut}(C)$ is of dimension one the conjecture can be proven to be generically false (see [31]).

On the positive side this does give us a method to completely solve the uniformization problem for curves with equations of the form

- (i) $y^2 = x^{2g+2} - 1$,
 - (ii) $y^2 = x^{2g+1} - 1$,
 - (iii) $y^2 = x(x^{2g} - 1)$,
- for all $g > 1$.

There are also a few other curves for which one can compute directly the accessory parameters and hence apply the methods of Section 2.1 (see [31]).

3 Geometric methods for the uniformization problem

The methods we are going to describe here originated with Fricke and Klein and the most famous example is probably the construction of a fundamental domain for Klein's quartic with homogeneous equation

$$x y^3 + y z^3 + z x^3 = 0. \tag{3.1}$$

For this construction one can note two points: the domain constructed is tiled by 56 copies of a hyperbolic triangle; the curve has a very large automorphism group (of order 168, the maximum possible in genus 3). These two facts are of course related.

3.1 Algebraic curves with many automorphisms

Recall that a triangle group is a group generated by reflections along the sides of a hyperbolic triangle. In terms of these groups one can reinterpret the condition (i) of Definition 2.7 by saying that an algebraic curve has a very large automorphism group if its Fuchsian group Γ is a normal subgroup of a triangle group. As a consequence the group Γ has a fundamental domain tiled by copies of a hyperbolic triangle.

For curves for which one knows the precise structure of the automorphism group, e.g. the hyperelliptic curves indicated at the end of Section 2, the problem now becomes a combinatorial problem, that can sometimes be solved.

For example consider the curve X_g defined by the equation $y^2 = x^{2g+2} - 1$, then it is known that the full group of automorphisms is generated by $u: (x, y) \mapsto (\xi x, y)$, where ξ is a primitive $(2g + 2)$ -root of unity, and $v: (x, y) \mapsto (1/x, i y/x^{g+1})$. This yields the presentation

$$\langle u, v; u^{2g+2}, v^4, (uv)^2, uv^2u^{-1}v^2 \rangle. \quad (3.2)$$

Let $n = g + 1$ and let \mathcal{T} be the hyperbolic triangle with angles $\pi/2n, \pi/2n$ and π/n . Label O one of the vertices with angle $\pi/2n$ and paste $4n$ copies of \mathcal{T} at the point O . We obtain in this way a hyperbolic polygon \mathcal{P} with $4n$ edges and interior angles alternately $2\pi/n$ and π/n . Number the edges from 1 to $4n$ in cyclic order and let Γ be the group identifying, for k odd, edge number k to edge number $k + 3 \pmod{4n}$. The group Γ thus constructed is a Fuchsian group for X_g (for a complete justification that this is indeed the case see [5]).

If $g \equiv 3 \pmod{4}$, then the same polygon can also be used to compute the Fuchsian group for Kulkarni's curve with equation

$$y^{2g+2} = x(x-1)^{g-1}(x+1)^{g+2}. \quad (3.3)$$

We only need to identify, for k odd, edge number k to edge number $2n + k + 3 \pmod{4n}$ (see [5] for complete details). Note that these are not hyperelliptic curves. For $g = 3$ this is actually isomorphic to Fermat's quartic with homogeneous equation $x^4 + y^4 + z^4 = 0$.

Similar constructions can be made for curves with equations

$$y^2 = x(x^{2g} - 1) \quad \text{or} \quad y^2 = x^{2g+1} - 1.$$

For these the construction is even simpler as one only needs to consider the regular hyperbolic $4g$ -gon (respectively $(4g + 2)$ -gon) with interior angles $\pi/2g$ (respectively $2\pi/(2g + 1)$) and the group that identifies opposite sides.

3.2 Tiling by quadrangles and other polygons

The limits of the method described in the preceding section come from the fact that one cannot deform hyperbolic triangles. This is just a rephrasing of the remark made earlier that curves with very large automorphism groups are isolated points in moduli space.

On the other hand one can deform hyperbolic quadrangles and also of course more general polygons.

To illustrate what we have in mind consider the family of curves defined by the equations

$$y^2 = x^{2g+2} + ax^{g+1} + 1 \quad \text{with } a \neq \pm 2. \quad (3.4)$$

It can be shown that for a generic member of this family the full automorphism group is generated by $u_1: (x, y) \mapsto (\zeta x, y)$, where ζ is this time a primitive $(g + 1)$ -root of

unity, and $v: (x, y) \mapsto (1/x, i y/x^{g+1})$. The presentation of the group is similar to the presentation (3.2).

Now consider again the tiling of the polygon \mathcal{P} by copies of the triangle \mathcal{T} constructed for the curve with equation $y^2 = x^{2g+2} - 1$. We can also view \mathcal{P} as tiled by $2n$ copies of the quadrangle \mathcal{Q} , with interior angles equal to π/n , obtained by pasting two copies of \mathcal{T} . Let \mathcal{Q}_1 be a hyperbolic quadrangle, with opposite angles equal and angle sum equal to $4\pi/n$, $n = g + 1$. Replace in \mathcal{P} the copies of \mathcal{Q} by copies of \mathcal{Q}_1 alternating the angle at the origin (see Figure 1 for a representation in the unit disk).

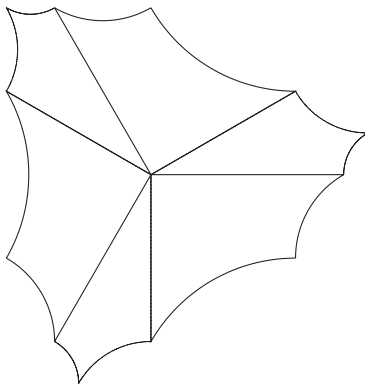


Figure 1. A fundamental domain for $y^2 = x^6 - 6x^3 + 1$.

With this construction we can again use the identification pattern i.e. for k odd, edge number k identified to edge number $k + 3 \pmod{4n}$.

Since hyperbolic quadrangles with fixed angle sum and opposite angles equal form a real two-dimensional family we obtain in this way a complex one-dimensional family of Riemann surfaces with the same automorphism group as the family (3.4). These two families are in fact the same (for a complete proof using a different method see [21]).

Establishing a more precise correspondence between the coefficient a of (3.4) and the quadrangle \mathcal{Q}_1 reduces to the problem of uniformizing a genus 1 curve with one elliptic point or equivalently the sphere with 4 conical points. A still unsolved problem but simpler and better understood than the general one (see for example [13]). There are also some cases when one can achieve the computations. For example in Figure 1 the quadrangle has angles equal to $\pi/6$ and $\pi/2$ and the hyperbolic cosines of the side lengths are 3 and 5 respectively. If we let A and B be hyperbolic transformations identifying opposite sides, then $\mathbb{H}/\langle A, B \rangle$ is an elliptic curve and the image of the axes of A and B define a symplectic basis for the homology of this elliptic curve. In this concrete case it can be shown to be the one with normalized period $\tau = \frac{1}{2} + \frac{i}{2}$. From this using the methods of [21] one can show that the coefficient a is equal to -6 . Since the genus is 2 in this case, the equation is $y^2 = x^6 - 6x^3 + 1$ (see [21] for other specific examples).

Variants of the above construction can be applied to other one-dimensional families in particular to the family defined by

$$y^2 = x(x^{2g} + ax^g + 1). \quad (3.5)$$

One can also use tilings by other polygons. For example tilings by pentagons and hexagons are considered in [1].

4 Automorphic forms and the inverse uniformization problem

Let S be a compact Riemann surface of genus g . Let ω and η be q th order differential forms (holomorphic q -differentials), then ω/η is a well defined meromorphic function on S .

Let $\Omega^q(S)$ be the space of holomorphic q -differentials. By Riemann–Roch the dimension of Ω^q is g if $q = 1$ or $k = (2q - 1)(g - 1)$ if $q \geq 2$. Let $\omega_1, \dots, \omega_k$ be a basis of Ω^q , then by the remark made above $f_q: p \mapsto (\omega_1(p), \dots, \omega_k(p))$ defines a map from S to the projective space \mathbb{P}^{k-1} , the so-called q -canonical map. It depends of course on the choice of a basis of $\Omega^q(S)$ but two such maps only differ by a projective automorphism of \mathbb{P}^{k-1} .

For non-hyperelliptic surfaces f_1 is an embedding and a two-to-one map onto a rational curve for hyperelliptic surfaces. For $q \geq 3$, f_q is always an embedding. The most interesting case is however f_2 which is always an embedding if $g > 2$ and is two-to-one onto a rational curve if $g = 2$ (see for example [28, Chap. III, §5] or [9, Chap. III, §10]).

For surfaces with cusps we have a very similar construction using the space of meromorphic q -differentials having at most poles of order less or equal to $q - 1$ at the cusps.

If S is defined as \mathbb{H}/Γ , then a q -differential ω lifts to $u(z) dz^q$ on \mathbb{H} , with

$$u(\gamma z)\gamma'(z)^q = u(z) \quad \text{for all } \gamma \in \Gamma. \quad (4.1)$$

Definition 4.1. A holomorphic function on \mathbb{H} satisfying condition (4.1) is called an *automorphic form of weight $2q$ for Γ* .

Since the transformations we are considering are Möbius transformations we can reformulate the condition (4.1):

$$u(\gamma z) = (cz + d)^{2q} u(z) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma. \quad (4.2)$$

Actually in the presence of cusps (we assume as before that Γ does not contain elliptic elements) we will need an additional condition to control the behavior at the punctures, namely

$$\sup \{y^{2q} |u(x + iy)| \mid x + iy \in \mathbb{H}\} < \infty. \quad (4.3)$$

The space $S^q(\Gamma)$ of functions satisfying Conditions (4.1) and (4.3) is the space of cusp forms of weight $2q$ for Γ .

Using the Petersson inner product one can show that condition (4.3) is equivalent to

$$\int_{\mathbb{H}/\Gamma} y^{2q-2} |u(x+iy)| dx dy < \infty$$

where integration is over a fundamental domain for Γ .

The importance of the space of cusp forms $S^q(\Gamma)$ comes from

Lemma 4.2. *Let $S = \mathbb{H}/\Gamma$ and let S^* be the smooth compactification of S . Let $\{p_1, \dots, p_n\} = S^* \setminus S$ and let D be the divisor $\sum p_i$. Finally let $\Omega^q((q-1)D)$ be the space of meromorphic q -differentials on S^* , holomorphic on S and with at worst poles of order $q-1$ at the p_i . Then*

$$S^q(\Gamma) \cong \Omega^q((q-1)D).$$

Proof. Let $\pi: \mathbb{H} \rightarrow S$ be the canonical projection. Let $p \in \mathbb{H}$ and let ζ be a local coordinate in the neighborhood of $\pi(p)$. An automorphic form φ of weight $2q$ for Γ projects to a holomorphic q -differential on S , $\Phi(\zeta)d\zeta^q$ with the rule

$$\varphi(z) = \Phi(\zeta) \left(\frac{d\zeta}{dz} \right)^q. \quad (4.4)$$

Let $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ be the extended real line. If $p \in \hat{\mathbb{R}}$ is a cusp for Γ we may assume without loss of generality that $p = \infty$ and that the stabilizer of p is generated by $z \mapsto z+1$. In this case φ has a Fourier expansion

$$\varphi(z) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi inz}. \quad (4.5)$$

Condition (4.3) then implies that

$$\lim_{y \rightarrow \infty} y^{2q} \sum a_n e^{-2\pi ny} = 0.$$

Hence $a_n = 0$ for $n \leq 0$.

A local coordinate in the neighborhood of the puncture is $\zeta = e^{2\pi iz}$. This implies that if we have $a_n = 0$ for $n < r$ and $a_r \neq 0$ in the expansion (4.5), then, by (4.4), the q -differential $\Phi(\zeta)d\zeta^q$ will have order $r-q$ and hence at worst a pole of order $q-1$ at the puncture.

Conversely if ω is a meromorphic q -differential on S^* , holomorphic on S and with at worst poles of order $q-1$ at the points of $S^* \setminus S$, then ω lifts via (4.4) to a cusp form of weight $2q$.

In this context solving the inverse uniformization problem reduces to finding a basis of the space $S^q(\Gamma)$ for some $q \geq 1$.

4.1 Poincaré series

The standard method to construct automorphic forms goes back to Poincaré who introduced what is now known as Poincaré series. Let h be a function on \mathbb{H} and

$$\Theta^q(h)(z) = \sum_{\gamma \in \Gamma} h(\gamma z) \gamma'(z)^q. \quad (4.6)$$

Obviously, provided the sum in (4.6) converges absolutely, $\Theta^q(h)$ satisfies condition (4.1). For $q \geq 2$ conditions for convergence were found by Poincaré, but we are going to restrict to a special class of functions.

Let \mathcal{R} be the space of rational functions holomorphic on $\hat{\mathbb{C}} \setminus \hat{\mathbb{R}}$ that have at most simple poles in $\hat{\mathbb{R}}$. If \mathcal{P} is a subset of $\hat{\mathbb{R}}$ we define

$$\mathcal{R}(\mathcal{P}) = \{f \in \mathcal{R} \mid f \text{ is holomorphic on } \hat{\mathbb{C}} \setminus \mathcal{P}\}.$$

Proposition 4.3. *For $R \in \mathcal{R}$ and $q \geq 2$ the series*

$$\Theta^q(R)(z) = \sum_{\gamma \in \Gamma} R(\gamma z) \gamma'(z)^q$$

converges absolutely and uniformly on compact subsets of \mathbb{H} . Moreover the Poincaré operator Θ^q maps \mathcal{R} onto the space of weight $2q$ cusp forms for Γ i.e.

$$\Theta^q(\mathcal{R}) = S^q(\Gamma).$$

For a proof see [19, Proposition 1.5]. The fact that Θ^q is onto is due to L. Bers [4]. The next result is again due to I. Kra [19, Theorem 1].

Theorem 4.4. *Let a_1, \dots, a_{2q-1} , $q > 1$, be distinct points in $\hat{\mathbb{R}}$. Let $\gamma_0 = \text{Id}$ and let $\gamma_1, \dots, \gamma_N$ be generators for Γ . Let*

$$\mathcal{P} = \{\gamma_j(a_k) \mid 1 \leq k < 2q, 0 \leq j \leq N\}.$$

Then

$$\Theta^q(\mathcal{R}(\mathcal{P})) = S^q(\Gamma).$$

In order to find a basis for $S^q(\Gamma)$ we first note that, with $\mathcal{R}(\mathcal{P})$ as in Theorem 4.4, we have on the one hand

$$\dim \mathcal{R}(\mathcal{P}) = \#\mathcal{P} + 1$$

and that on the other hand, if $\mathbb{H}/\Gamma = S$ is a surface of genus g with n punctures, then Riemann–Roch applied to the space $\Omega^q((q-1)D)$ of Lemma 4.2 yields

$$\dim S^q(\Gamma) = (2q-1)(g-1) + n(q-1).$$

The next point to note is that [19] provides an explicit algorithm to determine whether the Poincaré series vanishes i.e. $\Theta^q(R) = 0$. These facts put together indicate a clear strategy to find a basis of $S^q(\Gamma)$.

For $q = 2$, that is quadratic forms, one can use a result of S. Wolpert [37] that we proceed to describe. Let

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \quad \text{and} \quad \gamma(z) = \frac{az + b}{cz + d}.$$

The trace $\mathrm{tr}(\gamma) = a + d$ is only defined up to sign but its square $\mathrm{tr}^2(\gamma)$ is well defined and > 4 for γ hyperbolic. Also the fixed points of γ are the roots of $c z^2 + (d - a) z - b$. This noted we define

$$\omega_\gamma(z) = \sqrt{\mathrm{tr}^2(\gamma) - 4(c z^2 + (d - a) z - b)^{-2}}. \tag{4.7}$$

The importance of ω_γ lies in the fact that

$$(\omega_\gamma(\gamma(z))\gamma'(z))^2 = \omega_\gamma(z) \tag{4.8}$$

as can be checked by a direct computation. In particular ω_γ is an automorphic 2-form for the cyclic group $\langle \gamma \rangle$.

Now assume $S = \mathbb{H}/\Gamma$ is compact and let γ be a simple closed curve in S . Such a curve defines a hyperbolic element in Γ that we will also denote by γ . We consider the Petersson–Poincaré series

$$\theta_\gamma^*(z) = \sum_{h \in \langle \gamma \rangle \backslash \Gamma} \omega_\gamma(h(z))h'(z)^2. \tag{4.9}$$

The result of S. Wolpert [37, Theorem 3.7, p. 521] is

Theorem 4.5. *Let $S = \mathbb{H}/\Gamma$ be a compact surface of genus g . Let $\gamma_1, \dots, \gamma_{3g-3}$ be a maximal set of non-intersecting simple closed curves in S . Then the Petersson–Poincaré series $\theta_{\gamma_i}^*$ converge absolutely and uniformly on compact subsets and the $\{\theta_{\gamma_i}^*\}$ form a basis of $S^2(\Gamma)$.*

The conclusion is that we have a complete theoretical solution to the inverse uniformization problem. In practice however there is one difficulty left: evaluating the series (4.6) or (4.9), which is not so easy to do explicitly.

4.2 Modular curves

We will briefly indicate here a few aspects of the theory of modular curves in relation with the uniformization problem and in particular with the theory presented in the last section.

For an integer $N \geq 1$ define the principal level N congruence subgroup of the modular group $\mathrm{SL}_2(\mathbb{Z})$ to be

$$\Gamma(N) = \{A \in \mathrm{SL}_2(\mathbb{Z}) \mid A \equiv \text{Identity matrix (mod } N)\}. \tag{4.10}$$

A congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ is a subgroup containing $\Gamma(N)$ for some N . Of particular interest are

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid c \equiv 0 \pmod{N} \right\}, \quad (4.11)$$

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) \mid a \equiv d \equiv 1 \pmod{N} \right\}. \quad (4.12)$$

A congruence subgroup Γ is clearly a discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$. We can consider the quotients of the upper-half plane under such a group and this defines a Riemann surface $Y(\Gamma) = \mathbb{H}/\Gamma$. At this stage one should note that congruence subgroups have parabolic elements, hence the modular curve has cusps and is non-compact. It is the affine modular curve for Γ . One can easily compactify by considering $\mathbb{H}^* = \mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$ with an appropriate topology, on which the group obviously acts, and the quotient $X(\Gamma) = \mathbb{H}^*/\Gamma$ will be a compact Riemann surface and hence a projective algebraic curve. This is what is generally called the modular curve for Γ . On the other hand because of the cusps, Γ is a Fuchsian group for the affine curve $Y(\Gamma)$ and not the compact curve $X(\Gamma)$. If we start with $X(\Gamma)$ and compute a Fuchsian group G for this curve, then G will not contain parabolics and hence will not be conjugate to Γ .

In addition to parabolic elements a congruence subgroup may also have elliptic elements (of order 2 or 3) so even for the affine curve $Y(\Gamma)$ the situation is a little different than the one considered in the introduction.

We limit the discussion to the groups $\Gamma(N)$, $\Gamma_0(N)$ and $\Gamma_1(N)$ for which the surfaces are denoted by $Y(N)$, $Y_0(N)$, $Y_1(N)$, $X(N)$, $X_0(N)$ and $X_1(N)$ respectively. Henceforth Γ will designate one of these groups.

The first thing to note is that the curves $Y(N)$, $Y_0(N)$ and $Y_1(N)$ are moduli spaces for moduli problems involving elliptic curves and N -torsion points or subgroups of N -torsion points (see [29, §6] or [32, Theorem 13.1]). Using this one can sometimes, for small values of N , directly compute equations for these spaces.

Another approach is to find generators for the function field. For example it can be shown that the function field of $Y_0(N)$ is generated by $X = j(z)$ and $Y = j(Nz)$ (where j is the classical j -invariant for elliptic curves), hence an equation of the form $F_N(X, Y)$ for some polynomial F_N . Unfortunately this leads to equations of high degree with huge coefficients. For recent developments and a variant of this approach see [38].

The general method however is to deal with the compact surfaces $X(\Gamma)$.

We again consider the space of weight $2q$ cusp forms for Γ , that is, holomorphic functions on \mathbb{H} satisfying condition (4.2) that vanish at every cusp (note that in view of the proof of Lemma 4.2 this is equivalent to condition (4.3)).

The Hecke operators $T(p)$, p prime to N , are linear operators on the space $S^q(\Gamma)$ that satisfy

$$T(p)T(p') = T(p')T(p), \quad (4.13)$$

$$\langle T(p)f, g \rangle = \langle f, T(p)g \rangle, \quad (4.14)$$

where

$$\langle f, g \rangle = \int_{\mathbb{H}/\Gamma} f(x + iy) \overline{g(x + iy)} y^{2q-2} dx dy, \quad (4.15)$$

is the Petersson inner product.

The product (4.15) turns the finite-dimensional space $S^q(\Gamma)$ into a Hilbert space. Since the operators $T(p)$ are commuting and self-adjoint they can be simultaneously diagonalised by a unitary matrix. Hence

Theorem 4.6 (Hecke–Petersson). *The space $S^q(\Gamma)$ has a basis consisting of eigenforms for the operators $T(p)$, that is a basis $\{f_i\}_i$ with $f_i \in S^q(\Gamma)$ such that*

$$T(p)f_i = \lambda_i(p)f_i$$

for all p such that $(p, N) = 1$.

As noted above the congruence subgroups contain parabolics of the form $\begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix}$ and ∞ is always a cusp. Hence if f is a cusp form we can consider its Fourier expansion

$$f(z) = \sum_{n \geq 1} a_n e(z)^n, \quad \text{where } e(z) = e^{2\pi iz}. \quad (4.16)$$

The Hecke operators can be defined in terms of (4.16). For example if $\Gamma = \Gamma_0(N)$ we have, for p prime to N ,

$$T(p)f(z) = \sum_{n \geq 1} (a_{np} + p^{2q-1} a_{n/p}) e(z)^n \quad \text{with } a_{n/p} = 0 \text{ if } p \nmid n. \quad (4.17)$$

For a more intrinsic definition see [29, Chap. 3].

The important consequence of this is that one can recover the Fourier expansion of the forms f_i of Theorem 4.6 in terms of the eigenvalues of the Hecke operators. Moreover there are many number theoretic methods to effectively compute these eigenvalues. For recent developments see for example [11] and [3].

4.3 Other methods

We very briefly describe the approach of [7] and [8] to the inverse uniformization problem for hyperelliptic curves.

First consider a hyperelliptic curve C of genus g defined over \mathbb{R} and with $g + 1$ real components. Then C is the union of four isometric copies of a hyperbolic $2g + 2$ -gon D . Let x_1, \dots, x_{2g+2} be the vertices of D and denote by $[x_i, x_{i+1}]$ the edges. Now consider the harmonic function u_i on D satisfying the mixed boundary conditions

$$\begin{aligned} u_i(z) &= 0 && \text{on } [x_{2g+2}, x_1] \text{ and } [x_{2j}, x_{2j+1}], \text{ for } j < i, \\ u_i(z) &= 1 && \text{on } [x_{2j}, x_{2j+1}], \text{ for } i \leq j, \\ v[u_i] &= 0 && \text{on } [x_{2k-1}, x_{2k}], \end{aligned}$$

where $v[u_i]$ is the derivative of u_i with respect to the outward pointing unit vector field. The capacity of u_i on D is

$$c_i = \int_D \|\nabla u_i\|^2 dx dy = - \int_{\partial D} u_i v[u_i] d\mu. \quad (4.18)$$

The importance of these capacities is

Proposition 4.7 ([7]). *The coefficients of a normalized period matrix for C are simple and explicit linear combinations of the capacities c_i in (4.18).*

Since C is hyperelliptic one can use standard methods to recover an equation from the period matrix (see [9] or [23]). As a final remark on this method one should note that it is a fairly easy matter to obtain efficient approximations of the c_i .

The basic idea of [8] is to reduce the problem to the uniformization of elliptic curves and use the classical solution of the problem in genus 1. In genus 2 this can be done as follows. A hyperbolic genus 2 surface S can always be represented as a symmetric hyperbolic octagon with opposite sides identified. Let g_i be the Möbius transformation identifying sides $i + 4$ and side i and let p_i be the hyperbolic midpoint of side i . Let h_i be the order 2 elliptic transformation centered at p_i . The group $\tilde{G} = \langle g_1, g_3, h_2, h_4, h_6, h_8 \rangle$ has signature $(1; 2, 2, 2, 2)$. The relation between this genus 1 curve E and the genus 2 surface is that one can find an equation $y^2 = x(x-1)(x-a_1)$ such that the elliptic points have coordinates $(a_2, \pm y_2)$, $(a_3, \pm y_3)$ and the genus 2 surface has equation $y^2 = x(x-1)(x-a_1)(x-a_2)(x-a_3)$. The interest of the method is that a map from the octagon to the parallelogram defining E can easily be approximated numerically.

References

- [1] A. Aigon and R. Silhol, Hyperbolic hexagons and algebraic curves in genus 3. *J. London Math. Soc.* 66 (2002), 671–690. [141](#)
- [2] A. O. L. Atkin and J. Lehmer, Hecke Operators on $\Gamma_0(m)$. *Math. Ann.* 185 (1970), 134–160.
- [3] L. A. Borisov, P. E. Gunnels and S. Popescu, Elliptic functions and equations of modular curves. *Math. Ann.* 321 (2001), 553–568. [146](#)
- [4] L. Bers, Poincaré series for Kleinian groups. *Comm. Pure Appl. Math.* 26 (1973), 667–672; Correction to: “Poincaré series for Kleinian groups”, *ibid.* 27 (1974), 583. [143](#)
- [5] E. Bujalance, A. F. Costa, J. M. Gamboa and G. Riera, Period matrices of Accola-Maclachlan and Kulkarni surfaces. *Ann. Acad. Sci. Fenn. Math.* 25 (2000), 161–177. [139](#)
- [6] W. Burnside, Note on the equation $y^2 = x(x^4 - 1)$. *Proc. London Math. Soc.* 24 (1893), 17–20. [132](#)

- [7] P. Buser and R. Silhol, Geodesics, periods and equations of real hyperelliptic curves. *Duke Math. J.* 108 (2001), 211–250. [146](#), [147](#)
- [8] P. Buser and R. Silhol, Some remarks on the uniformizing function in genus 2. *Geom. Dedicata* 115 (2005), 121–133. [146](#), [147](#)
- [9] H. Farkas and I. Kra, *Riemann surfaces*. Grad. Texts in Math. 71, Springer-Verlag, New York 1992. [137](#), [141](#), [147](#)
- [10] J. Figueroa and R. Hidalgo, Numerical Schottky uniformizations. *Geom. Dedicata* 111 (2005), 125–157. [131](#)
- [11] S. D. Galbraith, *Equations for modular curves*. Doctoral Thesis, Oxford 1996. [146](#)
- [12] E. Gironde and G. González-Diez, On a conjecture of Whittaker concerning uniformization of hyperelliptic curves. *Trans. Amer. Math. Soc.* 356 (2004), 691–702. [137](#), [138](#)
- [13] L. Hadasz and Z. Jaskólski, Liouville theory and uniformization of four punctured spheres. *J. Math. Phys.* 47 (2006), 082304. [140](#)
- [14] D. Hejhal, Sur les paramètres accessoires pour l’uniformisation fuchsienne. *C. R. Acad. Sci. Paris A-B* 282 (1976), 403–406.
- [15] D. Hejhal, Monodromy groups and Poincaré series. *Bull. Amer. Math. Soc.* 84 (1978), 339–376.
- [16] L. Keen, H. E. Rauch and A. T. Vasquez, Moduli of punctured tori and the accessory parameters of Lamé’s equation. *Trans. Amer. Math. Soc.* 255 (1979), 201–230. [136](#), [138](#)
- [17] P. Koebe, Über die Uniformisierung beliebiger analytischer Kurven. *Göttinger Nachr.* (1907), 191–210; Zweite Mitteilung. *ibid.* (1907), 633–669. [131](#)
- [18] I. Kra, *Automorphic forms and Kleinian groups*. W. A. Benjamin, Inc., Reading, Mass., 1972.
- [19] I. Kra, On the vanishing of and spanning sets for Poincaré series for cusp forms. *Acta Math.* 153 (1984), 47–116. [143](#)
- [20] I. Kra, Accessory parameters for punctured spheres. *Trans. Amer. Math. Soc.* 313 (1989), 589–617. [136](#), [138](#)
- [21] S. Lelièvre and R. Silhol, Multi-geodesic tessellations, fractional Dehn twists and uniformization of algebraic curves. Preprint Hal-00129643, Feb. 2007. [140](#)
- [22] B. Maskit, *Kleinian Groups*. Grundlehren Math. Wiss. 287, Springer-Verlag, Berlin 1988. [131](#)
- [23] D. Mumford, *Tata lectures on Theta II*. Progr. Math. 43, Birkhäuser, Boston, Mass., 1984. [134](#), [147](#)
- [24] Z. Nehari, *Conformal Mapping*. McGraw-Hill, New York 1952. [133](#), [134](#), [136](#)
- [25] H. Poincaré, Sur l’uniformisation des fonctions analytiques. *Acta. Math.* 31 (1907), 1–64. [131](#)
- [26] R. A. Rankin, The differential equation associated with the uniformization of certain Algebraic curves. *Proc. Roy. Soc. Edinburgh Sect. A* 65 (1958), 35–62. [137](#)
- [27] R. A. Rankin, Burnside’s uniformization. *Acta Arith.* 79 (1997), 53–57. [132](#)
- [28] I. R. Shafarevich, *Basic Algebraic Geometry*. Springer Study Edition, Springer-Verlag, Berlin, New York 1977. [141](#)

- [29] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*. Princeton University Press, Princeton, N.J., 1971. [145](#), [146](#)
- [30] M. Seppälä, Myrberg's numerical uniformization of hyperelliptic curves. *Ann. Acad. Sci. Fenn. Math.* 29 (2004), 3–20. [131](#)
- [31] R. Silhol, A functional relation for accessory parameters for genus 2 algebraic curves with an order 4 automorphism. *J. London Math. Soc.* 71 (2005), 133–145. [138](#)
- [32] J. Silverman, *The arithmetic of elliptic curves*. Grad. Texts in Math. 106, Springer-Verlag, New York 1985. [145](#)
- [33] L. Takhtajan and P. Zograf, On the Liouville equation, accessory parameters and the geometry of Teichmüller space for Riemann surfaces of genus 0. *Math. USSR Sb.* 60 (1988), 143–161. [136](#)
- [34] L. Takhtajan and P. Zograf, Hyperbolic 2-spheres with conical singularities, accessory parameters and Kähler metrics on $\mathcal{M}_{0,n}$. *Trans. Amer. Math. Soc.* 355 (2003), 1857–1867. [136](#)
- [35] E. T. Whittaker, On hyperlemniscate functions. A family of automorphic functions. *J. London Math. Soc.* 4 (1929), 274–278. [137](#)
- [36] E. T. Whittaker and G. N. Watson, *A course of modern Analysis*. Cambridge University Press, Cambridge 1927. [133](#), [134](#)
- [37] S. Wolpert, The Fenchel-Nielsen deformation. *Ann. of Math.* 115 (1982), 501–528. [144](#)
- [38] Y. Yang, Defining equations of modular curves. *Adv. Math.* 204 (2006), 481–508. [145](#)

Chapter 5

Riemann surfaces, ribbon graphs and combinatorial classes

Gabriele Mondello

Contents

1	Introduction	151
1.1	Overview	151
1.2	Structure of the chapter	156
2	Systems of arcs and ribbon graphs	157
2.1	Systems of arcs	157
2.2	Ribbon graphs	162
3	Differential and algebro-geometric point of view	171
3.1	The Deligne–Mumford moduli space	171
3.2	The system of moduli spaces of curves	173
3.3	Augmented Teichmüller space	174
3.4	Tautological classes	178
3.5	Kontsevich’s compactification	180
4	Cell decompositions of the moduli space of curves	182
4.1	Harer–Mumford–Thurston construction	182
4.2	Penner–Bowditch–Epstein construction	191
4.3	Hyperbolic surfaces with boundary	192
5	Combinatorial classes	198
5.1	Witten cycles	198
5.2	Witten cycles and tautological classes	204
5.3	Stability of Witten cycles	206
	References	211

1 Introduction

1.1 Overview

1.1.1 Moduli space and Teichmüller space. Consider a compact oriented surface S of genus g together with a finite subset $X = \{x_1, \dots, x_n\}$, such that $2g - 2 + n > 0$.

The moduli space $\mathcal{M}_{g,X}$ is the set of all X -pointed Riemann surfaces of genus g up to isomorphism. Its universal cover (in the orbifold sense) can be identified with

the Teichmüller space $\mathcal{T}(S, X)$, which parametrizes complex structures on S up to isotopy (relative to X); equivalently, $\mathcal{T}(S, X)$ parametrizes isomorphism classes of (S, X) -marked Riemann surfaces. Thus, $\mathcal{M}_{g, X}$ is the quotient of $\mathcal{T}(S, X)$ under the action of the mapping class group $\Gamma(S, X) = \text{Diff}_+(S, X)/\text{Diff}_0(S, X)$.

As $\mathcal{T}(S, X)$ is contractible (Teichmüller [71]), we also have $\mathcal{M}_{g, X} \simeq B\Gamma(S, X)$. However, $\Gamma(S, X)$ acts on $\mathcal{T}(S, X)$ discontinuously but with finite stabilizers. Thus, $\mathcal{M}_{g, X}$ is naturally an orbifold and $\mathcal{M}_{g, X} \simeq B\Gamma(S, X)$ must be intended in the orbifold category.

1.1.2 Algebro-geometric point of view. As compact Riemann surfaces are complex algebraic curves, $\mathcal{M}_{g, X}$ has an algebraic structure and is in fact a Deligne–Mumford stack, which is the algebraic analogue of an orbifold. The underlying space $M_{g, X}$ (forgetting the isotropy groups) is a quasi-projective variety.

The problem of counting curves with suitable properties, a topic which is also called “enumerative geometry of curves”, has always been central in algebraic geometry. The usual set-up is to describe the loci in $\mathcal{M}_{g, X}$ of curves that satisfy the wished properties and then to compute their intersection, which naturally leads to seeking for a suitable compactification of $\mathcal{M}_{g, X}$. Deligne and Mumford [16] understood that it was sufficient to consider algebraic curves with mild singularities to compactify $\mathcal{M}_{g, X}$. In fact, their compactification $\bar{\mathcal{M}}_{g, X}$ is the moduli space of X -pointed stable (algebraic) curves of genus g , where a complex projective curve C is “stable” if its only singularities are nodes (that is, in local analytic coordinates C looks like $\{(x, y) \in \mathbb{C}^2 \mid xy = 0\}$) and every irreducible component of the smooth locus of $C \setminus X$ has negative Euler characteristic.

The main tool to prove the completeness of $\bar{\mathcal{M}}_{g, X}$ is the stable reduction theorem, which essentially says that a smooth holomorphic family $\mathcal{C}^* \rightarrow \Delta^*$ of X -pointed Riemann surfaces of genus g over the pointed disc can be completed to a flat family over Δ (after a suitable change of base $z \mapsto z^k$) using a stable curve.

The beauty of $\bar{\mathcal{M}}_{g, X}$ is that it is smooth (as an orbifold) and that its coarse space $\bar{M}_{g, X}$ is a projective variety (Mumford [59], Gieseker [22], Knudsen [40] [41], Kollár [42] and Cornalba [15] and [7]).

1.1.3 Tautological maps. The map $\bar{\mathcal{M}}_{g, X \cup \{y\}} \rightarrow \bar{\mathcal{M}}_{g, X}$ that forgets the y -marking (and then stabilizes the possibly unstable X -marked curve) can be identified with the universal family over $\bar{\mathcal{M}}_{g, X}$ and is the first example of tautological map.

Moreover, $\bar{\mathcal{M}}_{g, X}$ has a natural algebraic stratification, in which each stratum corresponds to a topological type of curve: for instance, smooth curves correspond to the open stratum $\mathcal{M}_{g, X}$. As another example: irreducible curves with one node correspond to an irreducible locally closed subvariety of (complex) codimension 1, which is the image of the (generically $2 : 1$) tautological boundary map $\mathcal{M}_{g-1, X \cup \{y_1, y_2\}} \rightarrow \bar{\mathcal{M}}_{g, X}$ that glues y_1 to y_2 . Thus, every stratum is the image of a (finite-to-one) tautological boundary map, and thus is isomorphic to a finite quotient of a product of smaller moduli spaces.

1.1.4 Augmented Teichmüller space. Teichmüller theorists are more interested in compactifying $\mathcal{T}(S, X)$ rather than $\mathcal{M}_{g,X}$. One of the most popular ways to do this is due to Thurston (see [21]): the boundary of $\mathcal{T}(S, X)$ is thus made of projective measured laminations and it is homeomorphic to a sphere.

Clearly, there cannot be any clear link between a compactification of $\mathcal{T}(S, X)$ and of $\mathcal{M}_{g,X}$, as the infinite discrete group $\Gamma(S, X)$ would not act discontinuously on a compact boundary $\partial\mathcal{T}(S, X)$.

Thus, a $\Gamma(S, X)$ -equivariant bordification of $\mathcal{T}(S, X)$ whose quotient is $\overline{\mathcal{M}}_{g,X}$ cannot be compact. A way to understand such a bordification is to endow $\mathcal{M}_{g,X}$ (and $\mathcal{T}(S, X)$) with the Weil–Petersson metric [73] and to show that its completion is exactly $\overline{\mathcal{M}}_{g,X}$ [51]. Hence, the Weil–Petersson completion $\overline{\mathcal{T}}(S, X)$ can be identified with the set of (S, X) -marked stable Riemann surfaces.

Similarly to $\overline{\mathcal{M}}_{g,X}$, $\overline{\mathcal{T}}(S, X)$ has a stratification by topological type and each stratum is a (finite quotient of a) product of smaller Teichmüller spaces.

1.1.5 Tautological classes. The moduli space $\overline{\mathcal{M}}_{g,X}$ comes equipped with natural vector bundles: for instance, \mathcal{L}_i is the holomorphic line bundle whose fiber at $[C]$ is the cotangent space T_{C,x_i}^* . Chern classes of these line bundles and their push-forward through tautological maps generate the so-called tautological classes (which can be seen in the Chow ring or in cohomology). The κ classes were defined by Mumford [60] and Morita [57] and then modified (to make them behave better under tautological maps) by Arbarello and Cornalba [5]. The ψ classes were defined by E. Miller [52] and their importance was successively rediscovered by Witten [74].

The importance of the tautological classes is due to the following facts (among others):

- Their geometric meaning appears quite clear.
- They behave very naturally under the tautological maps (see, for instance, [5]).
- They often occur in computations of enumerative geometry; that is, Poincaré duals of interesting algebraic loci are often tautological (see [60]) but not always (see [25])!
- They are defined on $\overline{\mathcal{M}}_{g,X}$ for every g and X (provided $2g - 2 + |X| > 0$), and they generate the stable cohomology ring over \mathbb{Q} due to Madsen–Weiss’s solution [49] of Mumford’s conjecture (see Section 5.3).
- There is a set of generators (ψ ’s and κ ’s) which have non-negativity properties (see [4] and [60]).
- They are strictly related to the Weil–Petersson geometry of $\overline{\mathcal{M}}_{g,X}$ (see [76], [79], [80] and [53]).

1.1.6 Simplicial complexes associated to a surface. One way to analyze the (co)homology of $\mathcal{M}_{g,X}$, and so of $\Gamma(S, X)$, is to construct a highly connected simplicial complex on which $\Gamma(S, X)$ acts. This is usually achieved by considering complexes of disjoint, pairwise non-homotopic simple closed curves on $S \setminus X$ with suitable properties (for instance, Harvey’s complex of curves [30]).

If X is nonempty (or if S has boundary), then one can construct a complex using systems of homotopically nontrivial, disjoint arcs joining two (not necessarily distinct) points in X (or in ∂S), thus obtaining the arc complex $\mathfrak{A}(S, X)$ (see [28]). It has an “interior” $\mathfrak{A}^\circ(S, X)$ made of systems of arcs that cut $S \setminus X$ in discs (or pointed discs) and a complementary “boundary” $\mathfrak{A}^\infty(S, X)$.

An important result, which has many fathers (Harer–Mumford–Thurston [28], Penner [61], Bowditch–Epstein [13]), says that the topological realization $|\mathfrak{A}^\circ(S, X)|$ of $\mathfrak{A}^\circ(S, X)$ is $\Gamma(S, X)$ -equivariantly homeomorphic to $\mathcal{T}(S, X) \times \Delta_X$ (where Δ_X is the standard simplex in \mathbb{R}^X). Thus, we can transfer the cell structure of $|\mathfrak{A}^\circ(S, X)|$ to an (orbi)cell structure on $\mathcal{M}_{g,X} \times \Delta_X$.

The homeomorphism is realized by coherently associating a weighted system of arcs to every X -marked Riemann surface, equipped with a decoration $\underline{p} \in \Delta_X$. There are two traditional ways to do this: using the flat structure arising from a Jenkins–Strebel quadratic differential (Harer–Mumford–Thurston) with prescribed residues at X or using the hyperbolic metric coming from the uniformization theorem (Penner and Bowditch–Epstein). Quite recently, several other ways have been introduced (see [46], [47], [56] and [55]).

1.1.7 Ribbon graphs. To better understand the homeomorphism between $|\mathfrak{A}^\circ(S, X)|$ and $\mathcal{T}(S, X) \times \Delta_X$, it is often convenient to adopt a dual point of view, that is to think of weighted systems of arcs as of metrized graphs \mathbb{G} , embedded in $S \setminus X$ through a homotopy equivalence.

This can be done by picking a vertex in each disc cut by the system of arcs and joining these vertices by adding an edge transverse to each arc. What we obtain is an (S, X) -marked metrized ribbon graph. Thus, points in $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X) \cong \mathcal{M}_{g,X} \times \Delta_X$ correspond to metrized X -marked ribbon graphs of genus g .

This point of view is particularly useful to understand singular surfaces (see also [13], [43], [45], [65], [82], [7] and [55]). The object dual to a system of arcs in $\mathfrak{A}^\infty(S, X)$ is a collection of data that we called an (S, X) -marked “enriched” ribbon graph. Notice that an X -marked “enriched” metrized ribbon graph does not carry all the information needed to construct a stable Riemann surface. Hence, the map $\bar{\mathcal{M}}_{g,X} \times \Delta_X \rightarrow |\mathfrak{A}(S, X)|/\Gamma(S, X)$ is not injective on the locus of singular curves, but still it is a homeomorphism on a dense open subset.

1.1.8 Topological results. The utility of the $\Gamma(S, X)$ -equivariant homotopy equivalence $\mathcal{T}(S, X) \simeq |\mathfrak{A}^\circ(S, X)|$ relies on the possibility of making topological computations on $|\mathfrak{A}^\circ(S, X)|$. For instance, Harer [28] determined the virtual cohomological dimension of $\Gamma(S, X)$ (and so of $\mathcal{M}_{g,X}$) using the high connectivity of $|\mathfrak{A}^\circ(S, X)|$ and he established that $\Gamma(S, X)$ is a virtual duality group, by showing that $|\mathfrak{A}^\circ(S, X)|$ is spherical. An analysis of the singularities of $|\mathfrak{A}(S, X)|/\Gamma(S, X)$ is in [66].

Successively, Harer–Zagier [29] and Penner [62] have computed the orbifold Euler characteristic of $\mathcal{M}_{g,X}$, where by “orbifold” we mean that a cell with stabilizer G has Euler characteristic $1/|G|$. Because of the cellularization, the problem translates into

enumerating X -marked ribbon graphs of genus g and counting them with the correct sign.

Techniques for enumerating graphs and ribbon graphs (see, for instance, [10]) have been known to physicists for long time: they use asymptotic expansions of Gaussian integrals over spaces of matrices. The combinatorics of iterated integrations by parts is responsible for the appearance of (ribbon) graphs (Wick's lemma). Thus, the problem of computing $\chi^{orb}(\mathcal{M}_{g,X})$ can be reduced to evaluating a matrix integral (a quick solution is also given by Kontsevich in Appendix D of [43]).

1.1.9 Intersection-theoretical results. As $\mathcal{M}_{g,X} \times \Delta_X$ is not just homotopy equivalent to $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$ but actually homeomorphic (through a piecewise real-analytic diffeomorphism), it is clear that one can try to rephrase integrals over $\mathcal{M}_{g,X}$ as integrals over $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$, that is as sums over maximal systems of arcs of integrals over a single simplex. This approach looked promising in order to compute Weil–Petersson volumes (see Penner [63]). Kontsevich [43] used it to compute volumes coming from a “symplectic form” $\Omega = p_1^2\psi_1 + \dots + p_n^2\psi_n$, thus solving Witten's conjecture [74] on the intersection numbers of the ψ classes.

However, in Witten's paper [74] matrix integrals entered in a different way. The idea was that, in order to integrate over the space of all conformal structures on S , one can pick a random decomposition of S into polygons, give each polygon a natural Euclidean structure and extend it to a conformal structure on S , thus obtaining a “random” point of $\mathcal{M}_{g,X}$. Refining the polygonalization of S leads to a measure on $\mathcal{M}_{g,X}$. Matrix integrals are used to enumerate these polygonalizations.

Witten also noticed that this refinement procedure may lead to different limits, depending on which polygons we allow. For instance, we can consider decompositions into A squares, or into A squares and B hexagons, and so on. Dualizing this last polygonalization, we obtain ribbon graphs embedded in S with A vertices of valence 4 and B vertices of valence 6. The corresponding locus in $|\mathfrak{A}^\circ(S, X)|$ is called a Witten subcomplex.

1.1.10 Witten classes. Kontsevich [43] and Penner [64] proved that Witten subcomplexes obtained by requiring that the ribbon graphs have m_i vertices of valence $(2m_i + 3)$ can be oriented (see also [14]) and they give cycles in $\overline{\mathcal{M}}_{g,X}^{\text{-comb}} := |\mathfrak{A}(S, X)|/\Gamma(S, X) \times \mathbb{R}_+$, which are denoted by $\overline{W}_{m_*,X}$. The Ω -volumes of these $\overline{W}_{m_*,X}$ are also computable using matrix integrals [43] (see also [17]).

In [44], Kontsevich constructed similar cycles using structure constants of finite-dimensional cyclic A_∞ -algebras with positive-definite scalar product and he also claimed that the classes $W_{m_*,X}$ (restriction of $\overline{W}_{m_*,X}$ to $\mathcal{M}_{g,X}$) are Poincaré dual to tautological classes.

This last statement (usually called Witten–Kontsevich's conjecture) was settled independently by Igusa [32] [33] and Mondello [54], whereas very little is known about the nature of the (non-homogeneous) A_∞ -classes.

1.1.11 Surfaces with boundary. The key point of all constructions of a ribbon graph out of a surface is that X must be nonempty, so that $S \setminus X$ can be retracted by deformation onto a graph. In fact, it is not difficult to see that the spine construction of Penner and Bowditch–Epstein can be performed (even in a more natural way) on hyperbolic surfaces Σ with geodesic boundary. The associated cellularization of the corresponding moduli space is due to Luo [46] (for smooth surfaces) and by Mondello [55] (also for singular surfaces, using Luo’s result).

The interesting fact (see [56] and [55]) is that gluing semi-infinite cylinders at $\partial\Sigma$ produces (conformally) punctured surfaces that “interpolate” between hyperbolic surfaces with cusps and flat surfaces arising from Jenkins–Strebel differentials.

1.2 Structure of the chapter

In Sections 2.1 and 2.2, we carefully define systems of arcs and ribbon graphs, both in the singular and in the nonsingular case, and we explain how the duality between the two works. Moreover, we recall Harer’s results on $\mathfrak{A}^\circ(S, X)$ and $\mathfrak{A}^\infty(S, X)$ and we state a simple criterion for compactness inside $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$.

In Sections 3.1 and 3.2, we describe the Deligne–Mumford moduli space of curves and the structure of its boundary, the associated stratification and boundary maps. In 3.3, we explain how the analogous bordification of the Teichmüller space $\overline{\mathcal{T}}(S, X)$ can be obtained as completion with respect to the Weil–Petersson metric.

Tautological classes and rings are introduced in 3.4 and Kontsevich’s compactification of $\mathcal{M}_{g,X}$ is described in 3.5.

In 4.1, we explain and sketch a proof of Harer–Mumford–Thurston cellularization of the moduli space and we illustrate the analogous result of Penner–Bowditch–Epstein in 4.2. In 4.3, we quickly discuss the relations between the two constructions using hyperbolic surfaces with geodesic boundary.

In 5.1, we define Witten subcomplexes and Witten cycles and we prove (after Kontsevich) that Ω orients them. We sketch the ideas involved in the proof the Witten cycles are tautological in Section 5.2.

Finally, in 5.3, we recall Harer’s stability theorem and we exhibit a combinatorial construction that shows that Witten cycles are stable. The fact (and probably also the construction) is well-known and it is also a direct consequence of Witten–Kontsevich’s conjecture and Miller’s work.

Acknowledgments. It is a pleasure to thank Shigeyuki Morita, Athanase Papadopoulos and Robert C. Penner for the stimulating workshop “Teichmüller space (Classical and Quantum)” they organized in Oberwolfach (May 28th–June 3rd, 2006) and the MFO for the hospitality.

I would like to thank Enrico Arbarello for all I learnt from him about Riemann surfaces and for his constant encouragement.

2 Systems of arcs and ribbon graphs

Let S be a compact oriented differentiable surface of genus g with $n > 0$ distinct marked points $X = \{x_1, \dots, x_n\} \subset S$. We will always assume that the Euler characteristic of the punctured surface $\dot{S} := S \setminus X$ is negative, that is $2 - 2g - n < 0$. This restriction only rules out the cases in which \dot{S} is the sphere with less than 3 punctures.

Let $\text{Diff}_+(S, X)$ be the group of orientation-preserving diffeomorphisms of S that fix X pointwise. The *mapping class group* $\Gamma(S, X)$ is the group of connected components of $\text{Diff}_+(S, X)$.

In what follows, we borrow some notation and some ideas from [45].

2.1 Systems of arcs

2.1.1 Arcs and arc complex. An oriented *arc* in S is a smooth path $\vec{\alpha} : [0, 1] \rightarrow S$ such that $\vec{\alpha}([0, 1]) \cap X = \{\vec{\alpha}(0), \vec{\alpha}(1)\}$, up to reparametrization. Let $\mathcal{A}^{\text{or}}(S, X)$ be the space of oriented arcs in S , endowed with its natural topology. Define $\sigma_1 : \mathcal{A}^{\text{or}}(S, X) \rightarrow \mathcal{A}^{\text{or}}(S, X)$ to be the orientation-reversing operator and we will write $\sigma_1(\vec{\alpha}) = \overleftarrow{\alpha}$. Call α the σ_1 -orbit of $\vec{\alpha}$ and denote by $\mathcal{A}(S, X)$ the (quotient) space of σ_1 -orbits in $\mathcal{A}^{\text{or}}(S, X)$.

A *system of $(k+1)$ -arcs* in S is a collection $\underline{\alpha} = \{\alpha_0, \dots, \alpha_k\} \subset \mathcal{A}(S, X)$ of $k+1$ unoriented arcs such that:

- if $i \neq j$, then the intersection of α_i and α_j is contained in X ,
- no arc in $\underline{\alpha}$ is homotopically trivial,
- no pair of arcs in $\underline{\alpha}$ are homotopic to each other.

We will denote by $S \setminus \underline{\alpha}$ the *complementary subsurface* of S obtained by removing $\alpha_0, \dots, \alpha_k$.

Each connected component of the space of systems of $(k+1)$ -arcs $\mathcal{AS}_k(S, X)$ is clearly contractible, with the topology induced by the inclusion $\mathcal{AS}_k(S, X) \hookrightarrow \mathcal{A}(S, X)^{k+1} / \mathfrak{S}_{k+1}$.

Let $\mathfrak{A}_k(S, X)$ be the set of homotopy classes of systems of $k+1$ arcs, that is $\mathfrak{A}_k(S, X) := \pi_0 \mathcal{AS}_k(S, X)$.

The *arc complex* is the simplicial complex $\mathfrak{A}(S, X) = \bigcup_{k \geq 0} \mathfrak{A}_k(S, X)$.

Notation. We will implicitly identify arc systems $\underline{\alpha}$ and $\underline{\alpha}'$ that are homotopic to each other. Similarly, we will identify the isotopic subsurfaces $S \setminus \underline{\alpha}$ and $S \setminus \underline{\alpha}'$.

2.1.2 Proper simplices. An arc system $\underline{\alpha} \in \mathfrak{A}(S, X)$ *fills* (resp. *quasi-fills*) a subsurface $R \subseteq S$ if $\underline{\alpha} \subset R$ and $\dot{R} \setminus \underline{\alpha}$ is a disjoint union of subsurfaces homeomorphic to discs (resp. discs and annuli isotopic to an end of \dot{R}). It is easy to check that the star of $\underline{\alpha}$ is finite if and only if $\underline{\alpha}$ quasi-fills S . In this case, we also say that $\underline{\alpha}$ is a *proper simplex* of $\mathfrak{A}(S, X)$.

Denote by $\mathfrak{A}^\infty(S, X) \subset \mathfrak{A}(S, X)$ the subcomplex of non-proper simplices and let $\mathfrak{A}^\circ(S, X) = \mathfrak{A}(S, X) \setminus \mathfrak{A}^\infty(S, X)$ be the collection of proper ones.

Notation. We denote by $|\mathfrak{A}^\infty(S, X)|$ and $|\mathfrak{A}(S, X)|$ the topological realizations of $\mathfrak{A}^\infty(S, X)$ and $\mathfrak{A}(S, X)$. We will use the symbol $|\mathfrak{A}^\circ(S, X)|$ to mean the complement of $|\mathfrak{A}^\infty(S, X)|$ inside $|\mathfrak{A}(S, X)|$.

2.1.3 Topologies on $|\mathfrak{A}(S, X)|$. The realization $|\mathfrak{A}(S, X)|$ of the arc complex can be endowed with two natural topologies (as is remarked in [13], [45] and [7]).

The former (which we call *standard*) is the finest topology that makes the inclusions $|\underline{\alpha}| \hookrightarrow |\mathfrak{A}(S, X)|$ continuous for all $\underline{\alpha} \in \mathfrak{A}(S, X)$; in other words, a subset $U \subset |\mathfrak{A}(S, X)|$ is declared to be open if and only if $U \cap |\underline{\alpha}|$ is open for every $\underline{\alpha} \in \mathfrak{A}(S, X)$. The latter topology is induced by the path *metric* d , which is the largest metric that restricts to the Euclidean one on each closed simplex.

The two topologies are the same where $|\mathfrak{A}(S, X)|$ is locally finite, but the latter is coarser elsewhere. We will always consider all realizations to be endowed with the metric topology.

2.1.4 Visible subsurfaces. For every system of arcs $\underline{\alpha} \in \mathfrak{A}(S, X)$, define $S(\underline{\alpha})_+$ to be the largest isotopy class of open subsurfaces of S such that

- every arc in $\underline{\alpha}$ is contained in $S(\underline{\alpha})_+$,
- $\underline{\alpha}$ quasi-fills $S(\underline{\alpha})_+$.

The *visible subsurface* $S(\underline{\alpha})_+$ can be constructed by taking the union of a thickening a representative of $\underline{\alpha}$ inside S and all those connected components of $S \setminus \underline{\alpha}$ which are homeomorphic to discs with at most one marked point (this construction appears already in [13]). We will always consider $S(\underline{\alpha})_+$ as an open subsurface (up to isotopy), homotopically equivalent to its closure $S(\underline{\alpha})_+$, which is an embedded surface with boundary.

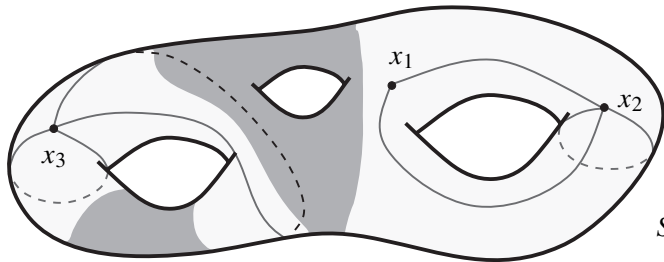


Figure 1. The invisible subsurface is the dark non-cylindrical component.

One can rephrase 2.1.2 by saying that $\underline{\alpha}$ is proper if and only if all S is $\underline{\alpha}$ -visible. We call *invisible subsurface* $S(\underline{\alpha})_-$ associated to $\underline{\alpha}$ the union of the interior of

the connected components of $S \setminus \overline{S(\underline{\alpha})}_+$ which are not unmarked cylinders. Thus, $S \setminus (S(\underline{\alpha})_+ \cup S(\underline{\alpha})_-)$ is a disjoint union of cylinders.

We also say that a marked point x_i is (in)visible for $\underline{\alpha}$ if it belongs to the $\underline{\alpha}$ -(in)visible subsurface.

2.1.5 Ideal triangulations. A maximal system of arcs $\underline{\alpha} \in \mathfrak{A}(S, X)$ is also called an *ideal triangulation* of S . In fact, it is easy to check that, in this case, each component of $S \setminus \underline{\alpha}$ bounded by three arcs and so is a “triangle”. (The term “ideal” comes from the fact that one often thinks of (S, X) as a hyperbolic surface with cusps at X and of $\underline{\alpha}$ as a collection of hyperbolic geodesics.) It is also clear that such an $\underline{\alpha}$ is proper.

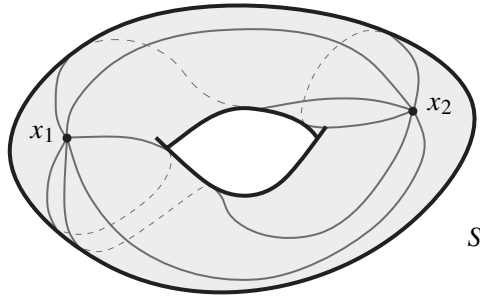


Figure 2. An example of an ideal triangulation for $(g, n) = (1, 2)$.

A simple calculation with the Euler characteristic of S shows that an ideal triangulation is made of exactly $6g - 6 + 3n$ arcs.

2.1.6 The spine of $|\mathfrak{A}^\circ(S, X)|$. Consider the barycentric subdivision $\mathfrak{A}(S, X)'$, whose k -simplices are chains $(\underline{\alpha}_0 \subsetneq \underline{\alpha}_1 \subsetneq \dots \subsetneq \underline{\alpha}_k)$. There is an obvious piecewise-affine homeomorphism $|\mathfrak{A}(S, X)'| \rightarrow |\mathfrak{A}(S, X)|$, that sends a vertex $(\underline{\alpha}_0)$ to the barycenter of $|\underline{\alpha}_0| \subset |\mathfrak{A}(S, X)|$.

Denote $\mathfrak{A}^\circ(S, X)'$ the subcomplex of $\mathfrak{A}(S, X)'$ whose simplices are chains of simplices that belong to $\mathfrak{A}^\circ(S, X)$. Clearly, $|\mathfrak{A}^\circ(S, X)'| \subset |\mathfrak{A}(S, X)'|$ is contained in $|\mathfrak{A}^\circ(S, X)| \subset |\mathfrak{A}(S, X)|$ through the above homeomorphism.

It is a general fact that there is a deformation retraction of $|\mathfrak{A}^\circ(S, X)|$ onto the spine $|\mathfrak{A}^\circ(S, X)'|$: on each simplex of $|\mathfrak{A}(S, X)'| \cap |\mathfrak{A}^\circ(S, X)|$ this is given by projecting onto the face contained in $|\mathfrak{A}^\circ(S, X)'|$. It is also clear that the retraction is $\Gamma(S, X)$ -equivariant.

In the special case of $X = \{x_1\}$, a proper system contains at least $2g$ arcs; whereas a maximal system contains exactly $6g - 3$ arcs. Thus, the (real) dimension of $|\mathfrak{A}^\circ(S, X)'|$ is $(6g - 3) - 2g = 4g - 3$.

Proposition 2.1 (Harer [28]). *If $X = \{x_1\}$, the spine $|\mathfrak{A}^\circ(S, X)'|$ has dimension $4g - 3$.*

2.1.7 Action of σ -operators. For every arc system $\underline{\alpha} = \{\alpha_0, \dots, \alpha_k\}$, denote by $E(\underline{\alpha})$ the subset $\{\vec{\alpha}_0, \overleftarrow{\alpha}_0, \dots, \vec{\alpha}_k, \overleftarrow{\alpha}_k\}$ of $\pi_0 \mathcal{A}^{\text{or}}(S, X)$. The action of σ_1 clearly restricts to $E(\underline{\alpha})$.

For each $i = 1, \dots, n$, the orientation of S induces a cyclic ordering of the oriented arcs in $E(\underline{\alpha})$ outgoing from x_i .

If $\vec{\alpha}_j$ starts at x_i , then define $\sigma_\infty(\vec{\alpha}_j)$ to be the oriented arc in $E(\underline{\alpha})$ outgoing from x_i that comes just *after* $\vec{\alpha}_j$. Moreover, σ_0 is defined by $\sigma_0 = \sigma_\infty^{-1} \sigma_1$.

If we denote by $E_t(\underline{\alpha})$ the orbits of $E(\underline{\alpha})$ under the action of σ_t , then

- $E_1(\underline{\alpha})$ can be identified with $\underline{\alpha}$,
- $E_\infty(\underline{\alpha})$ can be identified with the set of $\underline{\alpha}$ -visible marked points,
- $E_0(\underline{\alpha})$ can be identified with the set of connected components of $S(\underline{\alpha})_+ \setminus \underline{\alpha}$.

Denote by $[\vec{\alpha}_j]_t$ the σ_t -orbit of $\vec{\alpha}_j$, so that $[\vec{\alpha}_j]_1 = \alpha_j$ and $[\vec{\alpha}_j]_\infty$ is the starting point of $\vec{\alpha}_j$, whereas $[\vec{\alpha}_j]_0$ is the component of $S(\underline{\alpha})_+ \setminus \underline{\alpha}$ adjacent to α_j and which induces the orientation $\vec{\alpha}_j$ on it.

2.1.8 Action of $\Gamma(S, X)$ on (S, X) . There is a natural right action of the mapping class group

$$\begin{aligned} \mathcal{A}(S, X) \times \Gamma(S, X) &\longrightarrow \mathcal{A}(S, X), \\ (\alpha, g) &\longmapsto \alpha \circ g. \end{aligned}$$

The induced action on $\mathfrak{A}(S, X)$ preserves $\mathfrak{A}^\infty(S, X)$ and so $\mathfrak{A}^\circ(S, X)$.

It is easy to see that the stabilizer (under $\Gamma(S, X)$) of a simplex $\underline{\alpha}$ fits in the following exact sequence

$$1 \rightarrow \Gamma_{\text{cpt}}(S \setminus \underline{\alpha}, X) \rightarrow \text{stab}_\Gamma(\underline{\alpha}) \rightarrow \mathfrak{S}(\underline{\alpha})$$

where $\mathfrak{S}(\underline{\alpha})$ is the group of permutations of $\underline{\alpha}$ and $\Gamma_{\text{cpt}}(S \setminus \underline{\alpha}, X)$ is the mapping class group of orientation-preserving diffeomorphisms of $S \setminus \underline{\alpha}$ with compact support that fix X . Define the image of $\text{stab}_\Gamma(\underline{\alpha}) \rightarrow \mathfrak{S}(\underline{\alpha})$ to be the *automorphism group of $\underline{\alpha}$* .

We can immediately conclude that $\underline{\alpha}$ is proper if and only if $\text{stab}_\Gamma(\underline{\alpha})$ is finite (equivalently, if and only if $\Gamma_{\text{cpt}}(S \setminus \underline{\alpha}, X)$ is trivial).

2.1.9 Weighted arc systems. A point \bar{w} contained in $|\mathfrak{A}(S, X)|$ consists of a map $\bar{w}: \mathfrak{A}_0(S, X) \rightarrow [0, 1]$ such that

- the support of \bar{w} is a simplex $\underline{\alpha} = \{\alpha_0, \dots, \alpha_k\} \in \mathfrak{A}(S, X)$,
- $\sum_{i=0}^k \bar{w}(\alpha_i) = 1$.

We will call \bar{w} the (*projective*) *weight* of $\underline{\alpha}$. A *weight* for $\underline{\alpha}$ is a point of $w \in |\mathfrak{A}(S, X)|_{\mathbb{R}} := |\mathfrak{A}(S, X)| \times \mathbb{R}_+$, that is a map $w: \mathfrak{A}_0(S, X) \rightarrow \mathbb{R}_+$ with support on $\underline{\alpha}$. Call \bar{w} its associated projective weight.

2.1.10 Compactness in $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$. We are going to prove a simple criterion for a subset of $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$ to be compact.

Call $\mathcal{C}(S, X)$ the set of free homotopy classes of simple closed curves on $S \setminus X$, which are neither contractible nor homotopic to a puncture.

Define the “intersection product”

$$1: \mathcal{C}(S, X) \times |\mathfrak{A}(S, X)| \rightarrow \mathbb{R}_{\geq 0}$$

as $1(\gamma, \bar{w}) = \sum_{\alpha} 1(\gamma, \alpha) \bar{w}(\alpha)$, where $1(\gamma, \alpha)$ is the *geometric* intersection number. We will also refer to $1(\gamma, \bar{w})$ as to the *length* of γ at \bar{w} . Consequently, we will say that the *systole* at \bar{w} is

$$\text{sys}(\bar{w}) = \inf\{1(\gamma, \bar{w}) \mid \gamma \in \mathcal{C}(S, X)\}.$$

Clearly, the function sys descends to

$$\text{sys}: |\mathfrak{A}(S, X)|/\Gamma(S, X) \rightarrow \mathbb{R}_+.$$

Lemma 2.2. *A closed subset $K \subset |\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$ is compact if and only if there is an $\varepsilon > 0$ such that $\text{sys}([\bar{w}]) \geq \varepsilon$ for all $[\bar{w}] \in K$.*

Proof. In \mathbb{R}^N we easily have $d_2 \leq d_1 \leq \sqrt{N} \cdot d_2$, where d_r is the L^r -distance. Similarly, in $|\mathfrak{A}(S, X)|$ we have

$$d(\bar{w}, |\mathfrak{A}^\infty(S, X)|) \leq \text{sys}(\bar{w}) \leq \sqrt{N} \cdot d(\bar{w}, |\mathfrak{A}^\infty(S, X)|)$$

where $N = 6g - 7 + 3n$. The same holds in $|\mathfrak{A}(S, X)|/\Gamma(S, X)$.

Thus, if $[\underline{\alpha}] \in |\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$, then $[\underline{\alpha}] \cap \text{sys}^{-1}([\varepsilon, \infty)) \cap |\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$ is compact for every $\varepsilon > 0$. As $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$ contains finitely many cells, we conclude that $\text{sys}^{-1}([\varepsilon, \infty)) \cap |\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$ is compact.

Vice versa, if $\text{sys}: K \rightarrow \mathbb{R}_+$ is not bounded from below, then we can find a sequence $[\bar{w}_m] \subset K$ such that $\text{sys}(\bar{w}_m) \rightarrow 0$. Thus, $[\bar{w}_m]$ approaches $|\mathfrak{A}^\infty(S, X)|/\Gamma(S, X)$ and so is divergent in $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$. \square

2.1.11 Boundary weight map. Let Δ_X be the standard simplex in \mathbb{R}^X and let $\Delta_X^\circ := \Delta_X \cap \mathbb{R}_+^X$. The *boundary weight map* $\ell_\partial: |\mathfrak{A}(S, X)|_{\mathbb{R}} \rightarrow \Delta_X \times \mathbb{R}_+ \subset \mathbb{R}^X$ is the piecewise-linear map that sends $\{\alpha\}$ to $[\vec{\alpha}]_\infty + [\overleftarrow{\alpha}]_\infty$. The projective boundary weight map $\frac{1}{2}\ell_\partial: |\mathfrak{A}(S, X)| \rightarrow \Delta_X$ instead sends $\{\alpha\}$ to $\frac{1}{2}[\vec{\alpha}]_\infty + \frac{1}{2}[\overleftarrow{\alpha}]_\infty$.

2.1.12 Results on the arc complex. A few things are known about the topology of $|\mathfrak{A}(S, X)|$.

- The space of proper arc systems $|\mathfrak{A}^\circ(S, X)|$ can be naturally given the structure of a piecewise-affine topological manifold with boundary (Hubbard–Masur [31], credited to Whitney) of (real) dimension $6g - 7 + 3n$.
- The space $|\mathfrak{A}^\circ(S, X)|$ is $\Gamma(S, X)$ -equivariantly homeomorphic to $\mathcal{T}(S, X) \times \Delta_X$, where $\mathcal{T}(S, X)$ is the Teichmüller space of (S, X) (see 3.1.1 for definitions and

Section 4 for an extensive discussion on this result), and so is contractible. This result could also be probably extracted from [31], but it is more explicitly stated in Harer [28] (who attributes it to Mumford and Thurston), Penner [61] and Bowditch–Epstein [13]. As the moduli space of X -marked Riemann surfaces of genus g can be obtained as $\mathcal{M}_{g,X} \cong \mathcal{T}(S, X)/\Gamma(S, X)$ (see 3.1.2), then $\mathcal{M}_{g,X} \simeq B\Gamma(S, X)$ in the orbifold category.

- (c) The space $|\mathfrak{A}^\infty(S, X)|$ is homotopy equivalent to an infinite wedge of spheres of dimension $2g - 3 + n$ (Harer [28]).

Results (b) and (c) are the key step in the following.

Theorem 2.3 (Harer [28]). *$\Gamma(S, X)$ is a virtual duality group (that is, it has a subgroup of finite index which is a duality group) of dimension $4g - 4 + n$ for $g, n > 0$ (and $4g - 5$ for $n = 0$) and $n - 3$ if $g = 0$.*

This is immediate for $g = 0$, because $\mathcal{M}_{0,X}$ is a complex affine variety. If $g > 0$, it is sufficient to work with $X = \{x_1\}$, in which case the upper bound is given by (b) and Proposition 2.1, and the duality by (c).

2.2 Ribbon graphs

2.2.1 Graphs. A graph G is a triple (E, \sim, σ_1) , where E is a finite set, $\sigma_1: E \rightarrow E$ is a fixed-point-free involution and \sim is an equivalence relation on E .

In ordinary language

- E is the set of *oriented edges* of the graph,
- σ_1 is the orientation-reversing involution of E , so that the set of unoriented edges is $E_1 := E/\sigma_1$,
- two oriented edges are equivalent if and only if they come out from the same vertex, so that the set V of vertices is E/\sim and the valence of $v \in E/\sim$ is exactly the cardinality of the equivalence class $[v]$.

A *ribbon graph* \mathbb{G} is a triple (E, σ_0, σ_1) , where E is a (finite) set, $\sigma_1: E \rightarrow E$ is a fixed-point-free involution and $\sigma_0: E \rightarrow E$ is a permutation. Define $\sigma_\infty := \sigma_1 \circ \sigma_0^{-1}$ and denote by E_t the set of orbits of σ_t and by $[\cdot]_t: E \rightarrow E_t$ the natural projection. A disjoint union of two ribbon graphs is defined in the natural way.

Remark 2.4. Given a ribbon graph \mathbb{G} , the underlying ordinary graph $G = \mathbb{G}^{ord}$ is obtained by declaring that oriented edges in the same σ_0 -orbit are equivalent and forgetting about the precise action of σ_0 .

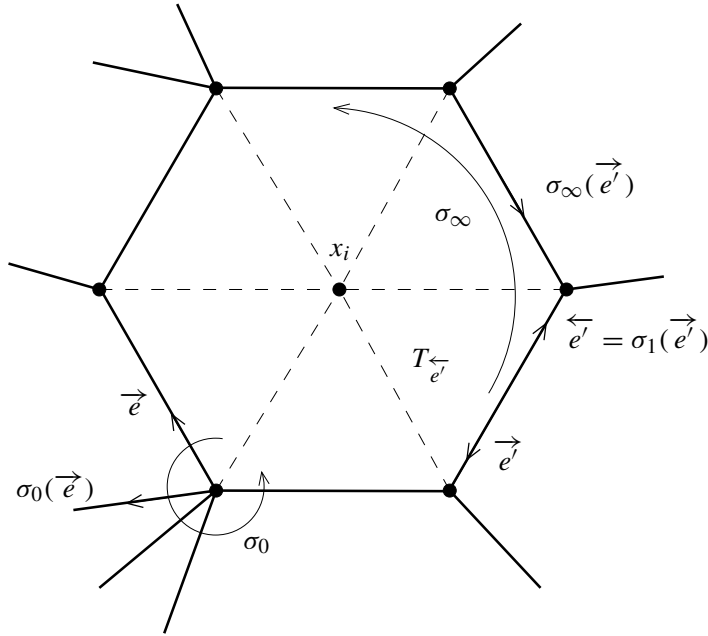


Figure 3. Geometric representation of a ribbon graph.

In ordinary language, a ribbon graph is an ordinary graph endowed with a cyclic ordering of the oriented edges outgoing from each vertex.

The σ_∞ -orbits are sometimes called *holes*. A *connected component* of \mathbb{G} is an orbit of $E(\mathbb{G})$ under the action of $\langle \sigma_0, \sigma_1 \rangle$.

The *Euler characteristic* of a ribbon graph \mathbb{G} is $\chi(\mathbb{G}) = |E_0(\mathbb{G})| - |E_1(\mathbb{G})|$ and its *genus* is $g(\mathbb{G}) = 1 + \frac{1}{2}(|E_1(\mathbb{G})| - |E_0(\mathbb{G})| - |E_\infty(\mathbb{G})|)$.

A (*ribbon*) *tree* is a connected (ribbon) graph of genus zero with one hole.

2.2.2 Subgraphs and quotients. Let $\mathbb{G} = (E, \sigma_0, \sigma_1)$ be a ribbon graph and let $Z \subsetneq E_1$ be a nonempty subset of edges.

The *subgraph* \mathbb{G}_Z is given by $(\tilde{Z}, \sigma_0^Z, \sigma_1^Z)$, where $\tilde{Z} = Z \times_{E_1} E$ and σ_0^Z, σ_1^Z are the induced operators (that is, for every $e \in \tilde{Z}$ we define $\sigma_0^Z(e) := \sigma_0^k(e)$, where $k = \min\{k > 0 \mid \sigma_0^k(e) \in \tilde{Z}\}$).

Similarly, the *quotient* \mathbb{G}/Z is $(\mathbb{G} \setminus \tilde{Z}, \sigma_0^{Z^c}, \sigma_1^{Z^c})$, where $\sigma_1^{Z^c}$ and $\sigma_\infty^{Z^c}$ are the operators induced on $E \setminus \tilde{Z}$ and $\sigma_0^{Z^c}$ is defined accordingly. A *new vertex* of \mathbb{G}/Z is a $\sigma_0^{Z^c}$ -orbit of $E \setminus \tilde{Z}$, which is not a σ_0 -orbit.

2.2.3 Bicolored graphs. A *bicolored graph* ζ is a finite connected graph with a partition $V = V_+ \cup V_-$ of its vertices. We say that ζ is *reduced* if no two vertices of

V_- are adjacent. If not differently specified, we will always understand that bicolored graphs are reduced.

If ζ contains an edge z that joins $w_1, w_2 \in V_-$, then we can obtain a new graph ζ' merging w_1 and w_2 along z into a new vertex $w' \in V'_-$ (by simply forgetting \vec{z} and \overleftarrow{z} and by declaring that vertices outgoing from w_1 are equivalent to vertices outgoing from w_2).

If ζ comes equipped with a function $g: V_- \rightarrow \mathbb{N}$, then $g': V'_- \rightarrow \mathbb{N}$ is defined so that $g'(w') = g(w_1) + g(w_2)$ if $w_1 \neq w_2$, or $g'(w') = g(w_1) + 1$ if $w_1 = w_2$.

As merging reduces the number of edges, we can iterate the process only a finite number of times. The result is independent of the choice of which edges to merge first and is a reduced graph ζ^{red} (possibly with a g^{red}).

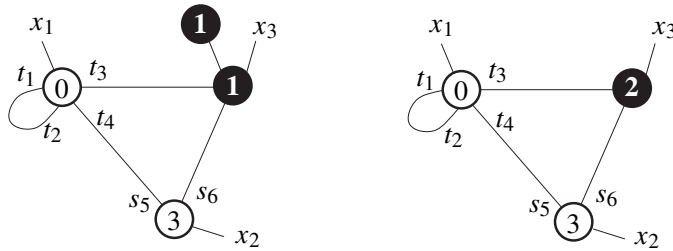


Figure 4. A non-reduced bicolored graph (on the left) and its reduction (on the right). Vertices in V_- are black. See Example 2.5.

2.2.4 Enriched ribbon graphs. An *enriched X -marked ribbon graph* \mathbb{G}^{en} is the datum of

- a connected bicolored graph (ζ, V_+) ,
- a ribbon graph \mathbb{G} plus a bijection $V_+ \rightarrow \{\text{connected components of } \mathbb{G}\}$,
- an (invisible) genus function $g: V_- \rightarrow \mathbb{N}$,
- an X -marking map $m: X \rightarrow V_- \cup E_\infty(\mathbb{G}) \cup E_0(\mathbb{G})$ such that the restriction $m^{-1}(E_\infty(\mathbb{G})) \rightarrow E_\infty(\mathbb{G})$ is bijective and the restriction $m^{-1}(E_0(\mathbb{G})) \rightarrow E_0(\mathbb{G})$ is injective (a vertex in the image of this last map is called *marked*),
- an injection $s_v: \{\text{oriented edges of } \zeta \text{ outgoing from } v\} \rightarrow E_0(\mathbb{G}_v)$ (vertices of \mathbb{G}_v in the image of s_v are called *nodal*; a vertex is called *special* if it is either marked or nodal),

that satisfy the following properties:

- for every $v \in V_+$ and $y \in E_0(\mathbb{G}_v)$ we have $|m^{-1}(y) \cup s_v^{-1}(y)| \leq 1$ (i.e. no more than one marking or one node at each vertex of \mathbb{G}_v),
- $2g(v) - 2 + |\{\text{oriented edges of } \zeta \text{ outgoing from } v\}| + |\{\text{marked points on } v\}| > 0$ for every $v \in V$ (stability condition),
- every non-special vertex of \mathbb{G}_v must be at least trivalent for all $v \in V_+$.

We say that \mathbb{G}^{en} is *reduced* if ζ is.

If the graph ζ is not reduced, then we can merge two vertices of ζ along an edge of ζ and obtain a new enriched X -marked ribbon graph. \mathbb{G}_1^{en} and \mathbb{G}_2^{en} are considered equivalent if they are related by a sequence of merging operations. It is clear that each equivalence class can be identified with its reduced representative. Unless differently specified, we will always refer to an enriched graph as the canonical reduced representative.

The total *genus* of \mathbb{G}^{en} is $g(\mathbb{G}^{\text{en}}) = 1 - \chi(\zeta) + \sum_{v \in V_+} g(\mathbb{G}_v) + \sum_{w \in V_-} g(w)$.

Example 2.5. In Figure 4, the genus of each vertex is written inside, x_1 and x_2 are marking the two holes of \mathbb{G} (sitting in different components), whereas x_3 is an invisible marked point. Moreover, t_1, t_2, t_3, t_4 (resp. s_5, s_6) are distinct (nodal) vertices of the visible component of genus 0 (resp. of genus 3): in particular, t_4 is marking the oriented edge that goes from the genus 0 component to the genus 3 component, whereas s_5 is marking the same edge with the opposite orientation. (Note that, if x_i marked a vertex s of some visible component, then we would have written “ s ” close the tail that joined v to s .) The total genus of the associated \mathbb{G}^{en} is 7.

Remark 2.6. If an edge z of ζ joins $v \in V_+$ and $w \in V_-$ and this edge is marked by the special vertex $y \in E_0(\mathbb{G}_v)$, then we will say, for brevity, that z joins w and y .

An enriched X -marked ribbon graph is *nonsingular* if ζ consists of a single visible vertex. Equivalently, an enriched nonsingular X -marked ribbon graph consists of a connected ribbon graph \mathbb{G} together with an injection $X \hookrightarrow E_\infty(\mathbb{G}) \cup E_0(\mathbb{G})$, whose image is exactly $E_\infty(\mathbb{G}) \cup \{\text{special vertices}\}$, such that non-special vertices are at least trivalent and $\chi(\mathbb{G}) - |\{\text{marked vertices}\}| < 0$.

2.2.5 Category of nonsingular ribbon graphs. A *morphism* of nonsingular X -marked ribbon graphs $\mathbb{G}_1 \rightarrow \mathbb{G}_2$ is an injective map $f: E(\mathbb{G}_2) \hookrightarrow E(\mathbb{G}_1)$ such that

- f commutes with σ_1, σ_∞ and respects the X -marking,
- $\mathbb{G}_{1,Z}$ is a disjoint union of trees, where $Z = E_1(\mathbb{G}_1) \setminus E_1(\mathbb{G}_2)$.

Notice that, as f preserves the X -markings (which are *injections* $X \hookrightarrow E_\infty(\mathbb{G}_i) \cup E_0(\mathbb{G}_i)$), then each component of Z may contain at most one special vertex.

Vice versa, if \mathbb{G} is a nonsingular X -marked ribbon graph and $\emptyset \neq Z \subsetneq E_1(\mathbb{G})$ such that \mathbb{G}_Z is a disjoint union of trees (each one containing at most a special vertex), then the inclusion $f: E_1(\mathbb{G}) \setminus Z \hookrightarrow E_1(\mathbb{G})$ induces a morphism of nonsingular ribbon graphs $\mathbb{G} \rightarrow \mathbb{G}/Z$.

Remark 2.7. A morphism is an *isomorphism* if and only if f is bijective.

$\mathfrak{RG}_{X,\text{ns}}$ is the small category whose objects are nonsingular X -marked ribbon graphs \mathbb{G} (where we assume that $E(\mathbb{G})$ is contained in a fixed countable set) with the

morphisms defined above. We use the symbol $\mathfrak{RG}_{g,X,\text{ns}}$ to denote the full subcategory of ribbon graphs of genus g .

2.2.6 Topological realization of nonsingular ribbon graphs. The *topological realization* $|G|$ of the graph $G = (E, \sim, \sigma_1)$ is the one-dimensional CW-complex obtained from $I \times E$ (where $I = [0, 1]$) by identifying

- $(t, \vec{e}) \sim (1 - t, \overleftarrow{e})$ for all $t \in I$ and $\vec{e} \in E$,
- $(0, \vec{e}) \sim (0, \vec{e}')$ whenever $e \sim e'$.

The *topological realization* $|\mathbb{G}|$ of the nonsingular X -marked ribbon graph $\mathbb{G} = (E, \sigma_0, \sigma_1)$ is the oriented surface obtained from $T \times E$ (where $T = I \times [0, \infty]/I \times \{\infty\}$) by identifying

- $(t, 0, \vec{e}) \sim (1 - t, 0, \overleftarrow{e})$ for all $t \in I$ and $\vec{e} \in E$,
- $(1, y, \vec{e}) \sim (0, y, \sigma_1 \sigma_\infty(\overleftarrow{e}))$ for all $\vec{e} \in E$ and $y \in [0, \infty]$.

If G is the ordinary graph underlying \mathbb{G} , then there is a natural embedding $|G| \hookrightarrow |\mathbb{G}|$, which we call the *spine*.

The points at infinity in $|\mathbb{G}|$ are called *centers* of the holes and can be identified with $E_\infty(\mathbb{G})$. Thus, $|\mathbb{G}|$ is naturally an X -marked surface.

Notice that a morphism of nonsingular X -marked ribbon graphs $\mathbb{G}_1 \rightarrow \mathbb{G}_2$ induces an isotopy class of orientation-preserving homeomorphisms $|\mathbb{G}_1| \rightarrow |\mathbb{G}_2|$ that respect the X -marking.

2.2.7 Nonsingular (S, X) -markings. An (S, X) -*marking* of the nonsingular X -marked ribbon graph \mathbb{G} is an isotopy class of orientation-preserving homeomorphisms $f: S \rightarrow |\mathbb{G}|$, compatible with $X \hookrightarrow E_\infty(\mathbb{G}) \cup E_0(\mathbb{G})$.

Define $\mathfrak{RG}_{\text{ns}}(S, X)$ to be the category whose objects are (S, X) -marked nonsingular ribbon graphs (\mathbb{G}, f) and whose morphisms $(\mathbb{G}_1, f_1) \rightarrow (\mathbb{G}_2, f_2)$ are morphisms $\mathbb{G}_1 \rightarrow \mathbb{G}_2$ such that $S \xrightarrow{f_1} |\mathbb{G}_1| \rightarrow |\mathbb{G}_2| \xrightarrow{f_2} S$ is isotopic to $f_2 \circ f_1^{-1}: S \rightarrow S$.

As usual, there is a right action of the mapping class group $\Gamma(S, X)$ on $\mathfrak{RG}_{\text{ns}}(S, X)$ and the quotient category $\mathfrak{RG}_{\text{ns}}(S, X)/\Gamma(S, X)$ is obtained from $\mathfrak{RG}_{\text{ns}}(S, X)$ by adding an (iso)morphism $[f: S \rightarrow \mathbb{G}] \mapsto [f \circ g: S \rightarrow \mathbb{G}]$ for each $g \in \Gamma(S, X)$ and each object $[f: S \rightarrow \mathbb{G}]$. It can be shown that the functor $\mathfrak{RG}_{\text{ns}}(S, X)/\Gamma(S, X) \rightarrow \mathfrak{RG}_{g,X,\text{ns}}$ that forgets the S -marking is an equivalence.

2.2.8 Nonsingular arcs/graph duality. Let $\underline{\alpha} = \{\alpha_0, \dots, \alpha_k\} \in \mathfrak{A}^\circ(S, X)$ be a proper arc system and let $\sigma_0, \sigma_1, \sigma_\infty$ be the corresponding operators on the set of oriented arcs $E(\underline{\alpha})$. The *ribbon graph dual to $\underline{\alpha}$* is $\mathbb{G}_{\underline{\alpha}} = (E(\underline{\alpha}), \sigma_0, \sigma_1)$, which comes naturally equipped with an X -marking (see 2.1.7).

Define the (S, X) -marking $f: S \rightarrow |\mathbb{G}_{\underline{\alpha}}|$ in the following way. Fix a point c_v in each component v of $S \setminus \underline{\alpha}$ (which must be exactly the marked point, if the component is a pointed disc) and let f send it to the corresponding vertex v of $|\mathbb{G}_{\underline{\alpha}}|$. For each arc

$\alpha_i \in \underline{\alpha}$, consider a transverse path β_i from $c_{v'}$ to $c_{v''}$ that joins the two components v' and v'' separated by α_i , intersecting α_i exactly once, in such a way that the interiors of β_i and β_j are disjoint, if $i \neq j$. Define f to be a homeomorphism of β_i onto the oriented edge in $|\mathbb{G}_{\underline{\alpha}}|$ corresponding to α_i that runs from v' to v'' .

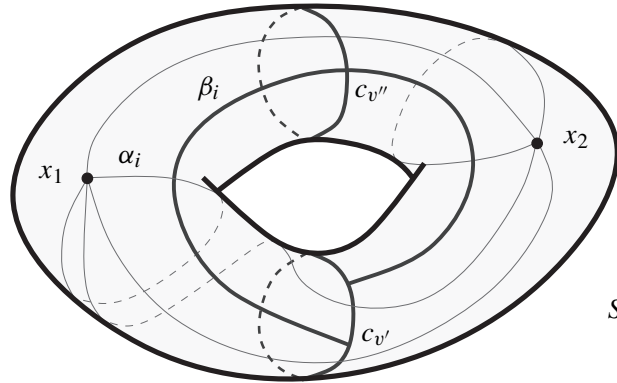


Figure 5. Thick curves represent $f^{-1}(|G_{\underline{\alpha}}|)$ and thin ones their dual arcs.

Because all components of $S \setminus \underline{\alpha}$ are discs (or pointed discs), it is easy to see that there is a unique way of extending f to a homeomorphism (up to isotopy).

Proposition 2.8. *The association above defines a $\Gamma(S, X)$ -equivariant equivalence of categories*

$$\widehat{\mathfrak{A}}^{\circ}(S, X) \longrightarrow \mathfrak{RG}_{\text{ns}}(S, X)$$

where $\widehat{\mathfrak{A}}^{\circ}(S, X)$ is the category of proper arc systems on (S, X) , whose morphisms are reversed inclusions.

In fact, an inclusion $\underline{\alpha} \hookrightarrow \underline{\beta}$ of proper systems induces a morphism $\mathbb{G}_{\underline{\beta}} \rightarrow \mathbb{G}_{\underline{\alpha}}$ of nonsingular (S, X) -marked ribbon graphs.

A pseudo-inverse is constructed as follows. Let $f: S \rightarrow |\mathbb{G}|$ be a nonsingular (S, X) -marked ribbon graph and let $|G| \hookrightarrow |\mathbb{G}|$ be the spine. The graph $f^{-1}(|G|)$ decomposes S into a disjoint union of one-pointed discs. For each edge e of $|G|$, let α_e be the simple arc joining the points in the two discs separated by e . Thus, we can associate the system of arcs $\{\alpha_e \mid e \in E_1(\mathbb{G})\}$ to (\mathbb{G}, f) and this defines a pseudo-inverse $\mathfrak{RG}_{\text{ns}}(S, X) \rightarrow \widehat{\mathfrak{A}}^{\circ}(S, X)$.

2.2.9 Metrized nonsingular ribbon graphs. A *metric* on a ribbon graph \mathbb{G} is a map $\ell: E_1(\mathbb{G}) \rightarrow \mathbb{R}_+$. Given a simple closed curve $\gamma \in \mathcal{C}(S, X)$ and an (S, X) -marked nonsingular ribbon graph $f: S \rightarrow |\mathbb{G}|$, there is a unique simple closed curve $\tilde{\gamma} = |e_{i_1}| \cup \dots \cup |e_{i_k}|$ contained inside $|G| \subset |\mathbb{G}|$ such that $f(\gamma)$ is freely homotopic to $\tilde{\gamma}$ inside $|\mathbb{G}| \setminus \{\text{centers of holes}\}$.

If \mathbb{G} is metrized, then we can define the *length* $\ell(\gamma)$ to be $\ell(\tilde{\gamma}) = \ell(e_{i_1}) + \dots + \ell(e_{i_k})$. Consequently, the *systole* is given by $\inf\{\ell(\gamma) \mid \gamma \in \mathcal{C}(S, X)\}$.

Given a proper weighted arc system $w \in |\mathcal{A}^\circ(S, X)|_{\mathbb{R}}$, supported on $\underline{\alpha} \in \mathcal{A}^\circ(S, X)$, we can endow the corresponding ribbon graph $\mathbb{G}_{\underline{\alpha}}$ with a *metric*, by simply setting $\ell(\alpha_i) := w(\alpha_i)$. Thus, one can extend the correspondence to proper weighted arc systems and metrized (S, X) -marked nonsingular ribbon graphs. Moreover, the notions of length and systole agree with those given in 2.1.10.

Notice the similarity between Lemma 2.2 and the Mumford–Mahler criterion for compactness in $\mathcal{M}_{g,n}$.

2.2.10 Category of enriched ribbon graphs. An *isomorphism* of enriched X -marked ribbon graphs $\mathbb{G}_1^{\text{en}} \rightarrow \mathbb{G}_2^{\text{en}}$ is the datum of compatible isomorphisms of their (reduced) graphs $c: \zeta_1 \rightarrow \zeta_2$ and of the ribbon graphs $\mathbb{G}_1 \rightarrow \mathbb{G}_2$, such that $c(V_{1,+}) = V_{2,+}$ and respecting the rest of the data.

Let \mathbb{G}^{en} be an enriched X -marked ribbon graph and let $e \in E_1(\mathbb{G}_v)$, where $v \in V_+$. Assume that $|V_+| > 1$ or that $|E_1(\mathbb{G}_v)| > 1$. We define \mathbb{G}^{en}/e in the following way.

- (a) If e is the only edge of \mathbb{G}_v , then we just turn v into an invisible component and we define $g(v) := g(\mathbb{G}_v)$ and $m(x_i) = v$ for all $x_i \in X$ that marked a hole or a vertex of \mathbb{G}_v . In what follows, suppose that $|E_1(\mathbb{G}_v)| > 1$.
- (b) If $[\vec{e}]_0$ and $[\overleftarrow{e}]_0$ are distinct and not both special, then we obtain \mathbb{G}^{en}/e from \mathbb{G}^{en} by simply replacing \mathbb{G}_v by \mathbb{G}_v/e .
- (c) If $[\vec{e}]_0 = [\overleftarrow{e}]_0$ is not special, then replace \mathbb{G}_v by \mathbb{G}_v/e . If $\{\vec{e}\}$ was a hole marked by x_j , then mark the new vertex of \mathbb{G}_v/e by x_j . Otherwise, add an edge to ζ that joins the two new vertices of \mathbb{G}_v/e (which may or may not split into two visible components).
- (d) In case $[\vec{e}]_0$ and $[\overleftarrow{e}]_0$ are both special vertices (whether or not they are distinct), add a new invisible component w of genus 0 to ζ , replace \mathbb{G}_v by \mathbb{G}_v/e (if \mathbb{G}_v/e is disconnected, the vertex v splits) and join w to the new vertices (one or two) of \mathbb{G}_v/e and mark w by $m^{-1}([\vec{e}]_0) \cup m^{-1}([\overleftarrow{e}]_0)$. Moreover, if $\{\vec{e}\}$ was a hole marked by x_j , then mark w by x_j .

Notice that \mathbb{G}^{en}/e can be not reduced, so we may want to consider the reduced enriched graph $\widetilde{\mathbb{G}^{\text{en}}/e}$ associated to it. We define $\mathbb{G}^{\text{en}} \rightarrow \widetilde{\mathbb{G}^{\text{en}}/e}$ to be an *elementary contraction*.

X -marked enriched ribbon graphs form a (small) category \mathfrak{RG}_X , whose morphisms are compositions of isomorphisms and elementary contractions. Denote by $\mathfrak{RG}_{g,X}$ the full subcategory of \mathfrak{RG}_X whose objects are ribbon graphs of genus g .

Remark 2.9. In reality, the automorphism group of an enriched ribbon graph must be defined as the product of the automorphism group as defined above by $\prod_{v \in V_-} \text{Aut}(v)$, where $\text{Aut}(v)$ is the group of automorphisms of the generic Riemann surface of type $(g(v), n(v))$ (where $n(v)$ is the number of oriented edges of ζ outgoing from v and of marked points on v). Fortunately, $\text{Aut}(v)$ is almost always trivial, except if $g(v) = n(v) = 1$, when $\text{Aut}(v) \cong \mathbb{Z}/2\mathbb{Z}$.

2.2.11 Topological realization of enriched ribbon graphs. The *topological realization* of the enriched X -marked ribbon graph \mathbb{G}^{en} is the nodal X -marked oriented surface $|\mathbb{G}^{\text{en}}|$ obtained as a quotient of

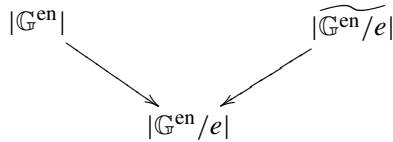
$$\left(\coprod_{v \in V_+} |\mathbb{G}_v| \right) \amalg \left(\coprod_{w \in V_-} S_w \right)$$

by a suitable equivalence relation, where S_w is a compact oriented surface of genus $g(w)$ with special points given by $m^{-1}(w)$ and by the oriented edges of ζ outgoing from w . The equivalence relation identifies couples of special points corresponding to the same edge of ζ .

As in the nonsingular case, for each $v \in V_+$ the positive component $|\mathbb{G}_v|$ naturally contains an embedded *spine* $|G_v|$. Notice that there is an obvious correspondence between edges of ζ and nodes of $|\mathbb{G}^{\text{en}}|$.

Moreover, the elementary contraction $\mathbb{G}^{\text{en}} \rightarrow \mathbb{G}^{\text{en}}/e$ to the (possibly) non-reduced \mathbb{G}^{en}/e defines a unique homotopy class of maps $|\mathbb{G}^{\text{en}}| \rightarrow |\mathbb{G}^{\text{en}}/e|$, which may shrink a circle inside a positive component of $|\mathbb{G}^{\text{en}}|$ to a singular point (only in cases (c) and (d)), and which are homeomorphisms elsewhere.

If $\widetilde{\mathbb{G}^{\text{en}}/e}$ is the reduced graph associated to \mathbb{G}^{en}/e , then we also have a map $|\widetilde{\mathbb{G}^{\text{en}}/e}| \rightarrow |\mathbb{G}^{\text{en}}/e|$ that shrinks some circles inside the invisible components to singular points and is a homeomorphism elsewhere.



2.2.12 (S, X) -markings of \mathbb{G}^{en} . An (S, X) -*marking* of an enriched X -marked ribbon graph \mathbb{G}^{en} is a map $f: S \rightarrow |\mathbb{G}^{\text{en}}|$ compatible with $X \hookrightarrow E_\infty(\mathbb{G}) \cup E_0(\mathbb{G})$ such that $f^{-1}(\{\text{nodes}\})$ is a disjoint union of circles and f is an orientation-preserving homeomorphism elsewhere, up to isotopy. The *visible subsurface* is the subsurface $S_+ := f^{-1}(|\mathbb{G}| \setminus \{\text{special points}\})$.

An *isomorphism* of (S, X) -marked (reduced) enriched ribbon graphs is an isomorphism $\mathbb{G}_1^{\text{en}} \rightarrow \mathbb{G}_2^{\text{en}}$ such that $S \xrightarrow{f_1} |\mathbb{G}_1^{\text{en}}| \rightarrow |\mathbb{G}_2^{\text{en}}|$ is homotopic to $f_2: S \rightarrow |\mathbb{G}_2^{\text{en}}|$.

Given (S, X) -markings $f: S \rightarrow |\mathbb{G}^{\text{en}}|$ and $f': S \rightarrow |\widetilde{\mathbb{G}^{\text{en}}/e}|$ such that $S \xrightarrow{f} |\mathbb{G}^{\text{en}}| \rightarrow |\mathbb{G}^{\text{en}}/e|$ is homotopic to $S \xrightarrow{f'} |\widetilde{\mathbb{G}^{\text{en}}/e}| \rightarrow |\mathbb{G}^{\text{en}}/e|$, then we define $(\mathbb{G}^{\text{en}}, f) \rightarrow (\widetilde{\mathbb{G}^{\text{en}}/e}, f')$ to be an *elementary contraction* of (S, X) -marked enriched ribbon graphs.

Define $\mathfrak{RG}(S, X)$ to be the category whose objects are (equivalence classes of) (S, X) -marked enriched ribbon graphs $(\mathbb{G}^{\text{en}}, f)$ and whose morphisms are compositions of isomorphisms and elementary contractions.

Again, the mapping class group $\Gamma(S, X)$ acts on $\mathfrak{R}\mathfrak{G}(S, X)$ and the quotient $\mathfrak{R}\mathfrak{G}(S, X)/\Gamma(S, X)$ is equivalent to $\mathfrak{R}\mathfrak{G}_{g, X}$.

2.2.13 Arcs/graph duality. Let $\underline{\alpha} = \{\alpha_0, \dots, \alpha_k\} \in \mathfrak{A}^\circ(S, X)$ be an arc system and let $\sigma_0, \sigma_1, \sigma_\infty$ the corresponding operators on the set of oriented arcs $E(\underline{\alpha})$.

Define V_+ to be the set of connected components of $S(\underline{\alpha})_+$ and V_- the set of components of $S(\underline{\alpha})_-$. Let ζ be a graph whose vertices are $V = V_+ \cup V_-$ and whose edges correspond to connected components of $S \setminus (S(\underline{\alpha})_+ \cup S(\underline{\alpha})_-)$, where an edge connects v and w (possibly $v = w$) if the associated cylinder bounds v and w .

Define $g: V_- \rightarrow \mathbb{N}$ to be the genus function associated to the connected components of $S(\underline{\alpha})_-$.

Denote by S_v the subsurface associated to $v \in V_+$ and let \hat{S}_v be the quotient of \bar{S}_v obtained by identifying each component of ∂S_v to a point. We denote by $\underline{\alpha} \cap \hat{S}_v$ the system of arcs induced on \hat{S}_v by $\underline{\alpha}$.

As $\underline{\alpha} \cap \hat{S}_v$ quasi-fills \hat{S}_v , we can construct a dual ribbon graph \mathbb{G}_v and a homeomorphism $\hat{S}_v \rightarrow |\mathbb{G}_v|$ by sending ∂S_v to nodal vertices of $|\mathbb{G}_v|$ and marked points on \hat{S}_v to centers or marked vertices of $|\mathbb{G}_v|$. These homeomorphisms glue to give a map $S \rightarrow |\mathbb{G}^{\text{en}}|$ that shrinks cylinders in $S \setminus (S(\underline{\alpha})_+ \cup S(\underline{\alpha})_-)$ to nodes and is a homeomorphism elsewhere, which is thus homotopic to a marking of $|\mathbb{G}^{\text{en}}|$.

We obtain an enriched (S, X) -marked (reduced) ribbon graph $\mathbb{G}_{\underline{\alpha}}^{\text{en}}$ dual to $\underline{\alpha}$.

Proposition 2.10. *The above construction defines a $\Gamma(S, X)$ -equivariant equivalence of categories*

$$\hat{\mathfrak{A}}(S, X) \longrightarrow \mathfrak{R}\mathfrak{G}(S, X)$$

where $\hat{\mathfrak{A}}(S, X)$ is the category of arc systems on (S, X) , whose morphisms are reversed inclusions.

As before, an inclusion $\underline{\alpha} \hookrightarrow \underline{\beta}$ of systems of arcs induces a morphism $\mathbb{G}_{\underline{\beta}}^{\text{en}} \rightarrow \mathbb{G}_{\underline{\alpha}}^{\text{en}}$ of (S, X) -marked enriched ribbon graphs.

To construct a pseudo-inverse, start with $(\mathbb{G}^{\text{en}}, f)$ and write \hat{S}_v for the surface obtained from $f^{-1}(|\mathbb{G}_v|)$ by shrinking each boundary circle to a point. By nonsingular duality, we can construct a system of arcs $\underline{\alpha}_v$ inside \hat{S}_v dual to $f_v: \hat{S}_v \rightarrow |\mathbb{G}_v|$. As the arcs miss the vertices of $f_v^{-1}(|\mathbb{G}_v|)$ by construction, $\underline{\alpha}_v$ can be lifted to S . The wanted arc system on S is $\underline{\alpha} = \bigcup_{v \in V_+} \underline{\alpha}_v$.

2.2.14 Metrized enriched ribbon graphs. A metric on \mathbb{G}^{en} is a map $\ell: E_1(\mathbb{G}) \rightarrow \mathbb{R}_+$. Given $\gamma \in \mathcal{C}(S, X)$ and an (S, X) -marking $f: S \rightarrow |\mathbb{G}^{\text{en}}|$, we can define $\gamma_+ := \gamma \cap S_+$. As in the nonsingular case, there is a unique $\tilde{\gamma}_+ = |e_{i_1}| \cup \dots \cup |e_{i_k}|$ inside $|G| \subset |\mathbb{G}|$ such that $f(\gamma_+) \simeq \tilde{\gamma}_+$.

Hence, we can define $\ell(\gamma) := \ell(\gamma_+) = \ell(e_{i_1}) + \dots + \ell(e_{i_k})$. Clearly, $\ell(\gamma) = i(\gamma, w)$, where w is the weight function supported on the arc system dual to $(\mathbb{G}^{\text{en}}, f)$.

Thus, the arc/graph duality also establishes a correspondence between weighted arc systems on (S, X) and metrized (S, X) -marked enriched ribbon graphs.

3 Differential and algebro-geometric point of view

3.1 The Deligne–Mumford moduli space

3.1.1 The Teichmüller space. Fix a compact oriented surface S of genus g and a subset $X = \{x_1, \dots, x_n\} \subset S$ such that $2g - 2 + n > 0$.

A *smooth family* of (S, X) -marked Riemann surfaces is a commutative diagram

$$\begin{array}{ccc} B \times S & \xrightarrow{f} & \mathcal{C} \\ & \searrow & \downarrow \pi \\ & & B \end{array}$$

where f is an orientation-preserving diffeomorphism (relative to π), $B \times S \rightarrow B$ is the projection onto the first factor and the fibers \mathcal{C}_b of π are Riemann surfaces, whose complex structure varies smoothly with $b \in B$.

Two families (f_1, π_1) and (f_2, π_2) over B are *isomorphic* if there exists a continuous map $h: \mathcal{C}_1 \rightarrow \mathcal{C}_2$ such that

- $h_b \circ f_{1,b}: (S, X) \rightarrow (\mathcal{C}_{2,b}, h_b \circ f_{1,b}(X))$ is homotopic to $f_{2,b}$ for every $b \in B$,
- $h_b: (\mathcal{C}_{1,b}, f_{1,b}(X)) \rightarrow (\mathcal{C}_{2,b}, f_{2,b}(X))$ is a biholomorphism for every $b \in B$.

The functor $\mathcal{T}(S, X): (\text{manifolds}) \rightarrow (\text{sets})$ defined by

$$B \mapsto \left\{ \begin{array}{l} \text{smooth families of } (S, X)\text{-marked} \\ \text{Riemann surfaces over } B \end{array} \right\} / \text{iso}$$

is represented by the *Teichmüller space* $\mathcal{T}(S, X)$.

It is a classical result that $\mathcal{T}(S, X)$ is a complex-analytic manifold of (complex) dimension $3g - 3 + n$ (Ahlfors [1], Bers [8] and Ahlfors-Bers [3]) and is diffeomorphic to a ball (Teichmüller [71]).

3.1.2 The moduli space of Riemann surfaces. A *smooth family* of X -marked Riemann surfaces of genus g is the datum of

- a submersion $\pi: \mathcal{C} \rightarrow B$,
- a smooth embedding $s: X \times B \rightarrow \mathcal{C}$

such that the fibers \mathcal{C}_b are Riemann surfaces of genus g , whose complex structure varies smoothly in $b \in B$, and $s_{x_i}: B \rightarrow \mathcal{C}$ is a section for every $x_i \in X$.

Two families (π_1, s_1) and (π_2, s_2) over B are isomorphic if there exists a diffeomorphism $h: \mathcal{C}_1 \rightarrow \mathcal{C}_2$ such that $\pi_2 \circ h = \pi_1$, the restriction of h to each fiber $h_b: \mathcal{C}_{1,b} \rightarrow \mathcal{C}_{2,b}$ is a biholomorphism and $h \circ s_1 = s_2$.

The existence of Riemann surfaces with nontrivial automorphisms (for $g \geq 1$) prevents the functor

$$\begin{aligned} (\text{manifolds}) &\longrightarrow (\text{sets}), \\ B &\longmapsto \left\{ \begin{array}{l} \text{smooth families of } X\text{-marked} \\ \text{Riemann surfaces over } B \end{array} \right\} / \text{iso} \end{aligned}$$

from being representable. However, Riemann surfaces with $2g - 2 + n > 0$ have finitely many automorphisms and so $\mathcal{M}_{g,X}$ is actually represented by an orbifold, which is in fact $\mathcal{T}(S, X) / \Gamma(S, X)$ (in the orbifold sense). In the algebraic category, we would rather say that $\mathcal{M}_{g,X}$ is a Deligne–Mumford stack with quasi-projective coarse space. In any case, we will always refer to $\mathcal{M}_{g,X}$ as the *moduli space* of X -marked Riemann surfaces of genus g .

3.1.3 Stable curves. Enumerative geometry is traditionally reduced to intersection theory on suitable moduli spaces. In our case, $\mathcal{M}_{g,X}$ is not a compact orbifold. To compactify it in an algebraically meaningful way, we need to look at how algebraic families of complex projective curves can degenerate.

In particular, given a holomorphic family $\mathcal{C}^* \rightarrow \Delta^*$ of algebraic curves over the punctured disc, we must understand how to complete the family over Δ .

Families of algebraic curves over the punctured disc are also studied, from a different point of view, in Chapter 3 of this volume, by Imayoshi (see [35]).

Example 3.1. Consider the family $\mathcal{C}^* = \{(b, [x : y : z]) \in \Delta^* \times \mathbb{C}\mathbb{P}^2 \mid y^2z = x(x - bz)(x - 2z)\}$ of curves of genus 1 with the marked point $[2 : 0 : 1] \in \mathbb{C}\mathbb{P}^2$, parametrized by $b \in \Delta^*$. Notice that the projection $\mathcal{C}_b^* \rightarrow \mathbb{C}\mathbb{P}^1$ given by $[x : y : z] \mapsto [x : z]$ (where $[0 : 1 : 0] \mapsto [1 : 0]$) is a $2 : 1$ cover, branched over $\{0, b, 2, \infty\}$. Fix a $\bar{b} \in \Delta^*$ and consider a closed curve $\gamma \subset \mathbb{C}\mathbb{P}^1$ that separates $\{\bar{b}, 2\}$ from $\{0, \infty\}$ and pick one of the two (simple closed) lifts $\tilde{\gamma} \subset \mathcal{C}_{\bar{b}}^*$.

This $\tilde{\gamma}$ determines a nontrivial element of $H_1(\mathcal{C}_{\bar{b}}^*)$. A quick analysis tells us that the endomorphism $T : H_1(\mathcal{C}_{\bar{b}}^*) \rightarrow H_1(\mathcal{C}_{\bar{b}}^*)$ induced by the monodromy around a generator of $\pi_1(\Delta^*, \bar{b})$ is nontrivial. Thus, the family $\mathcal{C}^* \rightarrow \Delta^*$ cannot be completed over Δ as a smooth family (because it would have trivial monodromy).

If we want to compactify our moduli space, we must allow our curves to acquire some singularities. Thus, it makes no longer sense to ask them to be submersions. Instead, we will require them to be *flat*.

Given an open subset $0 \in B \subset \mathbb{C}$, a flat family of connected projective curves $\mathcal{C} \rightarrow B$ may typically look like (up to shrinking B)

- $\Delta \times B \rightarrow B$ around a smooth point of \mathcal{C}_0 ,
- $\{(x, y) \in \mathbb{C}^2 \mid xy = 0\} \times B \rightarrow B$ around a node of \mathcal{C}_0 that persists on each \mathcal{C}_b ,
- $\{(b, x, y) \in B \times \mathbb{C}^2 \mid xy = b\} \rightarrow B$ around a node of \mathcal{C}_0 that does not persist on the other curves \mathcal{C}_b with $b \neq 0$

in local analytic coordinates.

Notice that, in the above cases, the (arithmetic) genus of each fiber $g_b = 1 - \frac{1}{2}[\chi(\mathcal{C}_b) - \nu_b]$ is constant in b , where ν_b is the number of nodes in \mathcal{C}_b .

To prove that allowing nodal curves is enough to compactify $\mathcal{M}_{g,X}$, one must show that it is always possible to complete any family $\mathcal{C}^* \rightarrow \Delta^*$ to a family over Δ . However, because nodal curves may have nontrivial automorphisms, we shall consider also the case in which $0 \in \Delta$ is an orbifold point. Thus, it is sufficient to be able to complete not exactly the family $\mathcal{C}^* \rightarrow \Delta^*$ but its pull-back under a suitable map $\Delta^* \rightarrow \Delta^*$ given by $z \mapsto z^k$. This is exactly the *semi-stable reduction theorem*.

One can observe that it is always possible to avoid producing genus 0 components with 1 or 2 nodes. Thus, we can consider only *stable curves*, that is nodal projective (connected) curves such that all irreducible components have finitely many automorphisms (equivalently, no irreducible component is a sphere with less than three nodes/marked points).

The *Deligne–Mumford compactification* $\bar{\mathcal{M}}_{g,X}$ of $\mathcal{M}_{g,X}$ is the moduli space of X -marked stable curves of genus g , which is a compact orbifold (algebraically, a Deligne–Mumford stack with projective coarse moduli space).

Its underlying topological space is a projective variety of complex dimension $3g - 3 + n$.

3.2 The system of moduli spaces of curves

3.2.1 Boundary maps. Many facts suggest that one should not look separately at each of the moduli spaces of X -pointed genus g curves $\bar{\mathcal{M}}_{g,X}$, but one must consider the whole system $(\bar{\mathcal{M}}_{g,X})_{g,X}$. An evidence is given by the existence of three families of maps that relate different moduli spaces.

- (1) The *forgetful map* is a projective flat morphism

$$\pi_q: \bar{\mathcal{M}}_{g,X \cup \{q\}} \longrightarrow \bar{\mathcal{M}}_{g,X}$$

that forgets the point q and stabilizes the curve (i.e. contracts a possible two-pointed sphere). This map can be identified with the universal family and so is endowed with *tautological sections*

$$\vartheta_{0,\{x_i,q\}}: \bar{\mathcal{M}}_{g,X} \rightarrow \bar{\mathcal{M}}_{g,X \cup \{q\}}$$

for all $x_i \in X$.

- (2) The *boundary map* corresponding to irreducible curves is the finite map

$$\vartheta_{\text{irr}}: \bar{\mathcal{M}}_{g-1,X \cup \{x',x''\}} \longrightarrow \bar{\mathcal{M}}_{g,X}$$

(defined for $g > 0$) that glues x' and x'' together. It is generically $2 : 1$ and its image sits in the boundary of $\bar{\mathcal{M}}_{g,X}$.

- (3) The *boundary maps* corresponding to reducible curves are the finite maps

$$\vartheta_{g',I}: \bar{\mathcal{M}}_{g',I \cup \{x'\}} \times \bar{\mathcal{M}}_{g-g',I \cup \{x''\}} \longrightarrow \bar{\mathcal{M}}_{g,X}$$

(defined for every $0 \leq g' \leq g$ and $I \subseteq X$ such that the spaces involved are nonempty) that take two curves and glue them together identifying x' and x'' . They are generically of degree 1 onto their images (except in the case $g = 2g'$ and $X = \emptyset$, when the map is generically $2 : 1$), which sit in the boundary of $\overline{\mathcal{M}}_{g,X}$ too.

Let $\delta_{0,\{x_i,q\}}$ be the Cartier divisor in $\overline{\mathcal{M}}_{g,X \cup \{q\}}$ corresponding to the image of the tautological section $\vartheta_{0,\{x_i,q\}}$ and write $D_q := \sum_i \delta_{0,\{x_i,q\}}$.

3.2.2 Stratification by topological type. We observe that $\overline{\mathcal{M}}_{g,X}$ has a natural *stratification* by topological type of the complex curve. In fact, we can attach to every stable curve Σ its *dual graph* ζ_Σ , whose vertices V correspond to irreducible components and whose edges correspond to nodes of Σ . Moreover, we can define a *genus function* $g : V \rightarrow \mathbb{N}$ such that $g(v)$ is the genus of the normalization $\tilde{\Sigma}_v$ of the irreducible component Σ_v corresponding to v and a *marking function* $m : X \rightarrow V$ (determined by requiring that x_i is marking a point on the irreducible component corresponding to $m(x_i)$). Equivalently, we will also say that the vertex $v \in V$ is *labelled* by $(g(v), X_v := m^{-1}(v))$. Call Q_v the inverse image through the normalization map $\tilde{\Sigma}_v \rightarrow \Sigma_v$ of the singular points of Σ_v .

For every such labeled graph ζ , we can construct a boundary map

$$\vartheta_\zeta : \prod_{v \in V} \overline{\mathcal{M}}_{g_v, X_v \cup Q_v} \longrightarrow \overline{\mathcal{M}}_{g,X}$$

which is a finite morphism.

3.3 Augmented Teichmüller space

3.3.1 Bordifications of $\mathcal{T}(S, X)$. Fix a compact oriented surface S of genus g and let $X = \{x_1, \dots, x_n\} \subset S$ such that $2g - 2 + n > 0$.

It is natural to look for natural *bordifications* of $\mathcal{T}(S, X)$: that is, we look for a space $\overline{\mathcal{T}}(S, X) \supset \mathcal{T}(S, X)$ that contains $\mathcal{T}(S, X)$ as a dense subspace and such that the action of the mapping class group $\Gamma(S, X)$ extends to $\overline{\mathcal{T}}(S, X)$.

A remarkable example is given by Thurston’s compactification $\overline{\mathcal{T}}^{\text{Th}}(S, X) = \mathcal{T}(S, X) \cup \mathbb{P}\mathcal{ML}(S, X)$, in which points at infinity are (isotopy classes of) projective measured laminations with compact support in $S \setminus X$. Thurston showed that $\mathbb{P}\mathcal{ML}(S, X)$ is compact and homeomorphic to a sphere. As $\Gamma(S, X)$ is infinite and discrete, this means that the quotient $\overline{\mathcal{T}}^{\text{Th}}(S, X) / \Gamma(S, X)$ cannot be too good and so this does not sound like a convenient way to compactify $\mathcal{M}_{g,X}$.

We will see in Section 4 that $\mathcal{T}(S, X)$ can be identified with $|\mathfrak{A}^\circ(S, X)|$. Thus, another remarkable example will be given by $|\mathfrak{A}(S, X)|$.

A natural question is how to define a bordification $\overline{\mathcal{T}}(S, X)$ such that

$$\overline{\mathcal{T}}(S, X) / \Gamma(S, X) \cong \overline{\mathcal{M}}_{g,X}.$$

3.3.2 Deligne–Mumford augmentation. A (continuous) family of stable (S, X) -marked curves is a commutative diagram

$$\begin{array}{ccc} B \times S & \xrightarrow{f} & \mathcal{C} \\ & \searrow & \downarrow \pi \\ & & B \end{array}$$

where $B \times S \rightarrow B$ is the projection onto the first factor and

- the family π is obtained as a pull-back of a flat stable family of X -marked curves $\mathcal{C}' \rightarrow B'$ through a continuous map $B \rightarrow B'$,
- if $N_b \subset \mathcal{C}_b$ is the subset of nodes, then $f^{-1}(v)$ is a smooth loop in $S \times \{b\}$ for every $v \in N_b$,
- for every $b \in B$ the restriction $f_b: S \setminus f^{-1}(N_b) \rightarrow \mathcal{C}_b \setminus N_b$ is an orientation-preserving homeomorphism, compatible with the X -marking.

Isomorphisms of such families are defined in the obvious way.

Example 3.2. Start with a flat family $\mathcal{C}' \rightarrow \Delta$ such that \mathcal{C}'_b are all homeomorphic for $b \neq 0$. Then consider the path $B = [0, \varepsilon) \subset \Delta$ and write $\mathcal{C} := \mathcal{C}' \times_{\Delta} B$. Over $(0, \varepsilon)$, the family \mathcal{C} is topologically trivial, whereas \mathcal{C}_0 may contain some new nodes.

Consider a marking $S \rightarrow \mathcal{C}_{\varepsilon/2}$ that pinches circles to nodes, is an oriented homeomorphism elsewhere and is compatible with X . The induced map $S \times (0, \varepsilon) \rightarrow \mathcal{C}_{\varepsilon/2} \times (0, \varepsilon)$ extends over 0 to $S \times [0, \varepsilon) \rightarrow \mathcal{C}_{\varepsilon/2} \times [0, \varepsilon) \cong \text{Bl}_{\mathcal{C}_0} \mathcal{C}$, where $\text{Bl}_{\mathcal{C}_0} \mathcal{C}$ is the real-oriented blow-up of \mathcal{C} along \mathcal{C}_0 . Composing with $\text{Bl}_{\mathcal{C}_0} \mathcal{C} \rightarrow \mathcal{C}$, we obtain our wished (S, X) -marking.

The *Deligne–Mumford augmentation* of $\mathcal{T}(S, X)$ is the topological space $\overline{\mathcal{T}}^{\text{DM}}(S, X)$ that classifies families of stable (S, X) -marked curves.

It follows easily that $\overline{\mathcal{T}}^{\text{DM}}(S, X)/\Gamma(S, X) = \overline{\mathcal{M}}_{g, X}$ as topological spaces. However, $\overline{\mathcal{T}}^{\text{DM}}(S, X) \rightarrow \overline{\mathcal{M}}_{g, X}$ has infinite ramification at $\partial^{\text{DM}}\mathcal{T}(S, X)$, due to the Dehn twists around the pinched loops.

3.3.3 Hyperbolic length functions. Let $[f: S \rightarrow \Sigma]$ be a point of $\mathcal{T}(S, X)$. As $\chi(S \setminus X) = 2 - 2g - n < 0$, the uniformization theorem provides a universal cover $\mathbb{H} \rightarrow \Sigma \setminus f(X)$, which endows $\Sigma \setminus f(X)$ with a hyperbolic metric of finite volume, with cusps at $f(X)$.

In fact, we can interpret $\mathcal{T}(S, X)$ as the classifying space of (S, X) -marked families of hyperbolic surfaces. It is clear that continuous variation of the complex structure corresponds to continuous variation of the hyperbolic metric (uniformly on the compact subsets, for instance), and so to continuity of the holonomy map $H: \pi_1(S \setminus X) \times \mathcal{T}(S, X) \rightarrow \text{PSL}_2(\mathbb{R}) \cong \text{Aut}(\mathbb{H})$.

In particular, for every $\gamma \in \pi_1(S \setminus X)$ the function $\ell_{\gamma}: \mathcal{T}(S, X) \rightarrow \mathbb{R}$ that associates to $[f: S \rightarrow \Sigma]$ the *length* of the unique geodesic in the free homotopy class

$f_*\gamma$ is continuous. As $\cosh(\ell_\gamma/2) = |\text{Tr}(H_\gamma/2)|$, one can check that H can be reconstructed from sufficiently (but finitely) many length functions. So that the continuity of these is equivalent to the continuity of the family.

3.3.4 Fenchel–Nielsen coordinates. Let $\underline{\gamma} = \{\gamma_1, \dots, \gamma_N\}$ be a maximal system of disjoint simple closed curves of $\dot{S} = S \setminus X$ (and so $N = 3g - 3 + n$) such that no γ_i is contractible in \dot{S} or homotopic to a puncture and no couple γ_i, γ_j bounds a cylinder contained in \dot{S} .

The system $\underline{\gamma}$ induces a *pair of pants decomposition* of \dot{S} , that is $\dot{S} \setminus (\gamma_1 \cup \dots \cup \gamma_N) = P_1 \cup P_2 \cup \dots \cup P_{2g-2+n}$, and each P_i is a pair of pants (i.e. a surface of genus 0 with $\chi(P_i) = -1$).

Given $[f : S \rightarrow \Sigma] \in \mathcal{T}(S, X)$, we have *lengths* $\ell_i(f) = \ell_{\gamma_i}(f)$ for $i = 1, \dots, N$, which determine the hyperbolic type of all pants P_1, \dots, P_{2g-2+n} . The information about how the pants are glued together is encoded in the *twist parameters* $\tau_i = \tau_{\gamma_i} \in \mathbb{R}$, which are well-defined up to some choices. What is important is that, whatever choices we make, the difference $\tau_i(f_1) - \tau_i(f_2)$ is the same and it is well-defined.

The *Fenchel–Nielsen coordinates* $(\ell_i, \tau_i)_{i=1}^N$ exhibit a real-analytic diffeomorphism $\mathcal{T}(S, X) \xrightarrow{\sim} (\mathbb{R}_+ \times \mathbb{R})^N$ (which clearly depends on the choice of $\underline{\gamma}$).

3.3.5 Fenchel–Nielsen coordinates around nodal curves. Points of $\partial^{\text{DM}}\mathcal{T}(S, X)$ are (S, X) -marked stable curves or, equivalently (using the uniformization theorem componentwise), (S, X) -marked hyperbolic surfaces with nodes, i.e. homotopy classes of maps $f : S \rightarrow \Sigma$, where Σ is a hyperbolic surface with nodes ν_1, \dots, ν_k , the fiber $f^{-1}(\nu_j)$ is a simple closed curve γ_j and f is an orientation-preserving diffeomorphism outside the nodes.

Complete $\{\gamma_1, \dots, \gamma_k\}$ to a maximal set $\underline{\gamma}$ of simple closed curves in (S, X) and consider the associated Fenchel–Nielsen coordinates (ℓ_j, τ_j) on $\mathcal{T}(S, X)$. As we approach the point $[f]$, the holonomies $H_{\gamma_1}, \dots, H_{\gamma_k}$ tend to parabolics and so the lengths ℓ_1, \dots, ℓ_k tend to zero. In fact, the hyperbolic metric on the surface Σ has a pair of cusps at each node ν_j .

This shows that the lengths functions ℓ_1, \dots, ℓ_k continuously extend to zero at $[f]$. On the other hand, the twist parameters $\tau_1(f), \dots, \tau_k(f)$ make no longer sense.

If we look at what happens on $\bar{\mathcal{M}}_{g,X}$, we may notice that the couples $(\ell_j, \tau_j)_{j=1}^k$ behave like polar coordinate around $[\Sigma]$, so that it seems natural to set $\vartheta_m = 2\pi \tau_m / \ell_m$ for all $m = 1, \dots, N$ and define consequently a map $F_{\underline{\gamma}} : (\mathbb{R}^2)^N \rightarrow \bar{\mathcal{M}}_{g,X}$, that associates to $(\ell_1, \vartheta_1, \dots, \ell_N, \vartheta_N)$ the surface with Fenchel–Nielsen coordinates $(\ell_m, \tau_m = \ell_m \vartheta_m / 2\pi)$, where (ℓ_i, ϑ_i) are polar coordinates on the i -th copy of \mathbb{R}^2 . Notice that the map is well-defined, because a twist along γ_j by ℓ_j is a diffeomorphism of the surface (a Dehn twist).

The map $F_{\underline{\gamma}}$ is an orbifold local chart $F_{\underline{\gamma}} : \mathbb{R}^{2N} \rightarrow F_{\underline{\gamma}}(\mathbb{R}^{2N}) \subset \bar{\mathcal{M}}_{g,X}$ and its image contains $[\Sigma]$. Varying $\underline{\gamma}$, we can cover the whole $\bar{\mathcal{M}}_{g,X}$ and thus give it a *Fenchel–Nielsen smooth structure*.

The bad news, analyzed by Wolpert [78], is that the Fenchel–Nielsen smooth structure is different (at $\partial\mathcal{M}_{g,X}$) from the Deligne–Mumford one. In fact, if a boundary divisor is locally described by $\{z_1 = 0\}$, then the length ℓ_γ of the corresponding vanishing geodesic is related to z_1 by $|z_1| \approx \exp(-1/\ell_\gamma)$, which shows that the identity map $\overline{\mathcal{M}}_{g,X}^{\text{FN}} \rightarrow \overline{\mathcal{M}}_{g,X}^{\text{DM}}$ is Lipschitz, but its inverse is not Hölder-continuous.

3.3.6 Weil–Petersson metric. Let Σ be a Riemann surface of genus g with marked points $X \hookrightarrow \Sigma$ such that $2g - 2 + n > 0$. First-order deformations of the complex structure can be rephrased in terms of $\bar{\partial}$ operator as $\bar{\partial} + \varepsilon\mu\partial + o(\varepsilon)$, where the *Beltrami differential* $\mu \in \Omega^{0,1}(T_\Sigma(-X))$ can be locally written as $\mu(z) \frac{d\bar{z}}{dz}$ with respect to some holomorphic coordinate z on Σ and $\mu(z)$ vanishes at X .

Given a smooth vector field $V = V(z) \frac{\partial}{\partial z}$ on Σ that vanishes at X , the deformations induced by μ and $\mu + \bar{\partial}V$ differ only by an isotopy of Σ generated by V (which fixes X).

Thus, the *tangent space* $T_{[\Sigma]}\mathcal{M}_{g,X}$ can be identified with $H^{0,1}(\Sigma, T_\Sigma(-X))$. As a consequence, the *cotangent space* $T_{[\Sigma]}^*\mathcal{M}_{g,X}$ identifies with the space $\mathcal{Q}(\Sigma, X)$ of *integrable holomorphic quadratic differentials* on $\dot{\Sigma} = \Sigma \setminus X$, that is, which are allowed to have a simple pole at each $x_i \in X$. The duality between $T_{[\Sigma]}\mathcal{M}_{g,X}$ and $T_{[\Sigma]}^*\mathcal{M}_{g,X}$ is given by

$$H^{0,1}(\Sigma, T_\Sigma(-X)) \times H^0(\Sigma, K_\Sigma^{\otimes 2}(X)) \longrightarrow \mathbb{C},$$

$$(\mu, \varphi) \longmapsto \int_\Sigma \mu\varphi.$$

If $\dot{\Sigma}$ is given the hyperbolic metric λ , then elements in $H^{0,1}(\Sigma, T_\Sigma(-X))$ can be identified with the space of *harmonic Beltrami differentials* $\mathcal{H}(\Sigma, X) = \{\bar{\varphi}/\lambda \mid \varphi \in \mathcal{Q}(\Sigma, X)\}$.

The *Weil–Petersson Hermitian metric* $h = g + i\omega$ (defined by Weil [73] using Petersson’s pairing of modular forms) is

$$h(\mu, \nu) := \int_\Sigma \mu\bar{\nu} \cdot \lambda$$

for $\mu, \nu \in \mathcal{H}(\Sigma, X) \cong T_\Sigma\mathcal{M}_{g,X}$.

This metric has a lot of properties: it is Kähler (Weil [73] and Ahlfors [2]) and it is mildly divergent at $\partial\mathcal{M}_{g,X}$, so that the Weil–Petersson distance extends to a non-degenerate distance on $\overline{\mathcal{M}}_{g,X}$ and all points of $\partial\mathcal{M}_{g,X}$ are at finite distance (Masur [51], Wolpert [75]). A recent review of the Weil–Petersson geometry is contained in Chapter 1 of this volume of the Handbook (see [81]).

Because $\overline{\mathcal{M}}_{g,X}$ is compact and so WP-complete, the lifting of the Weil–Petersson metric to $\overline{\mathcal{T}}(S, X)$ is also complete. Thus, $\overline{\mathcal{T}}(S, X)$ can be seen as the *Weil–Petersson completion* of $\mathcal{T}(S, X)$.

3.3.7 Weil–Petersson form. We should emphasize that the Weil–Petersson symplectic form ω_{WP} depends more directly on the hyperbolic metric on the surface than on its holomorphic structure.

In particular, Wolpert [77] has shown that

$$\omega_{WP} = \sum_i d\ell_i \wedge d\tau_i$$

on $\mathcal{T}(S, X)$, where (ℓ_i, τ_i) are Fenchel–Nielsen coordinates associated to any pair of pants decomposition of (S, X) .

On the other hand, if we identify the space $\mathcal{T}(S, X)$ with an open subset of $\text{Hom}(\pi_1(\dot{S}), \text{SL}_2(\mathbb{R}))/\text{SL}_2(\mathbb{R})$, then points of $\mathcal{T}(S, X)$ are associated \mathfrak{g} -local systems ρ on \dot{S} (with parabolic holonomies at X and hyperbolic holonomies otherwise), where $\mathfrak{g} = \mathfrak{sl}_2(\mathbb{R})$ is endowed with the symmetric bilinear form $\langle \alpha, \beta \rangle = \text{Tr}(\alpha\beta)$.

Goldman [24] has proved that, in this description, the tangent space to $\mathcal{T}(S, X)$ at ρ is naturally $H^1(S, X; \mathfrak{g})$ and that ω_{WP} is given by $\omega(\mu, \nu) = (1/8)\text{Tr}(\mu \cup \nu) \cap [S]$.

Remark 3.3. Another description of ω in terms of shear coordinates and Thurston’s symplectic form on measured laminations is given by Bonahon–Sözen [68].

One can feel that the complex structure J on $\mathcal{T}(S, X)$ inevitably shows up whenever we deal with the Weil–Petersson metric, as $g(\cdot, \cdot) = \omega(\cdot, J\cdot)$. On the other hand, the knowledge of ω is sufficient to compute volumes and characteristic classes.

3.4 Tautological classes

3.4.1 Relative dualizing sheaf. All the maps between moduli spaces we have defined are in some sense tautological as they are very naturally constructed and they reflect intrinsic relations among the various moduli spaces. It is evident that one can look at these as classifying maps to the Deligne–Mumford stack $\overline{\mathcal{M}}_{g, X}$ (which obviously descend to maps between coarse moduli spaces). Hence, we can consider all the cycles obtained by pushing forward or pulling back via these maps as being “tautologically” defined.

Moreover, there is an ingredient we have not considered yet: it is the *relative dualizing sheaf* of the universal family $\pi_q: \overline{\mathcal{M}}_{g, X \cup \{q\}} \rightarrow \overline{\mathcal{M}}_{g, X}$. One expects that it carries many information and that it can produce many classes of interest.

The relative dualizing sheaf ω_{π_q} is the sheaf on $\overline{\mathcal{M}}_{g, X \cup \{q\}}$, whose local sections are (algebraically varying) Abelian differentials that are allowed to have simple poles at the nodes, provided the two residues at each node are opposite. The local sections of $\omega_{\pi_q}(D_q)$ (the logarithmic variant of ω_{π_q}) are sections of ω_{π_q} that may have simple poles at the X -marked points.

3.4.2 MMMAC classes. The *Miller classes* are

$$\psi_{x_i} := c_1(\mathcal{L}_i) \in CH^1(\overline{\mathcal{M}}_{g, X})_{\mathbb{Q}}$$

where $\mathcal{L}_i := \vartheta_{0, \{x_i, q\}}^* \omega_{\pi_q}$ and the *Mumford–Morita classes* (suitably modified by Arbarello–Cornalba) are

$$\kappa_j := (\pi_q)_*(\psi_q^{j+1}) \in CH^j(\bar{\mathcal{M}}_{g, X})_{\mathbb{Q}}.$$

One could moreover define the l -th *Hodge bundle* as $\mathbb{E}_l := (\pi_q)_*(\omega_{\pi_q}^{\otimes l})$ and consider the Chern classes of these bundles (for example, the λ classes $\lambda_i := c_i(\mathbb{E}_1)$). However, using Grothendieck–Riemann–Roch, Mumford [60] and Bini [11] proved that $c_i(\mathbb{E}_j)$ can be expressed as a linear combination of Mumford–Morita classes up to elements in the boundary, so that they do not introduce anything really new.

When there is no risk of ambiguity, we will denote in the same way the classes ψ and κ belonging to different $\bar{\mathcal{M}}_{g, X}$'s as it is now traditional.

Remark 3.4. Wolpert has proven [76] that, on $\bar{\mathcal{M}}_g$, we have $\kappa_1 = [\omega_{WP}]/\pi^2$ and that the amplitude of $\kappa_1 \in A^1(\bar{\mathcal{M}}_g)$ (and so the projectivity of $\bar{\mathcal{M}}_g$) can be recovered from the fact that $[\omega_{WP}/\pi^2]$ is an integral Kähler class [79]. He also showed that the cohomological identity $[\omega_{WP}/\pi^2] = \kappa_1 = (\pi_q)_*\psi_q^2$ admits a clean pointwise interpretation [80].

3.4.3 Tautological rings. Because of the natural definition of the κ and ψ classes, as explained before, the subring $R^*(\mathcal{M}_{g, X})$ of $CH^*(\mathcal{M}_{g, X})_{\mathbb{Q}}$ they generate is called the *tautological ring* of $\mathcal{M}_{g, X}$. Its image $RH^*(\mathcal{M}_{g, X})$ through the cycle class map is called the cohomology tautological ring.

From an axiomatic point of view, the *system of tautological rings* ($R^*(\bar{\mathcal{M}}_{g, X})$) is the minimal system of subrings of $(CH^*(\bar{\mathcal{M}}_{g, X}))$ such that

- every $R^*(\bar{\mathcal{M}}_{g, X})$ contains the fundamental class $[\bar{\mathcal{M}}_{g, X}]$,
- the system is closed under push-forward maps π_* , $(\vartheta_{\text{irr}})_*$ and $(\vartheta_{g', l})_*$.

$R^*(\mathcal{M}_{g, X})$ is defined as the image of the restriction map $R^*(\bar{\mathcal{M}}_{g, X}) \rightarrow CH^*(\mathcal{M}_{g, X})$. The definition for the rational cohomology is analogous (where the role of $[\bar{\mathcal{M}}_{g, X}]$ is here played by its Poincaré dual $1 \in H^0(\bar{\mathcal{M}}_{g, X}; \mathbb{Q})$).

It is a simple fact to remark that all tautological rings contain ψ and κ classes and in fact that $R^*(\mathcal{M}_{g, X})$ is generated by them. Really, this was the original definition of $R^*(\mathcal{M}_{g, X})$.

3.4.4 Faber’s formula. The ψ classes interact reasonably well with the forgetful maps. In fact

$$\begin{aligned} (\pi_q)_*(\psi_{x_1}^{r_1} \cdots \psi_{x_n}^{r_n}) &= \sum_{\{i | r_i > 0\}} \psi_{x_1}^{r_1} \cdots \psi_{x_i}^{r_i-1} \cdots \psi_{x_n}^{r_n}, \\ (\pi_q)_*(\psi_{x_1}^{r_1} \cdots \psi_{x_n}^{r_n} \psi_q^{b+1}) &= \psi_{x_1}^{r_1} \cdots \psi_{x_n}^{r_n} \kappa_b, \end{aligned}$$

where the first one is the so-called *string equation* and the second one for $b = 0$ is the *dilaton equation* (see [74]). They have been generalized by Faber for maps that forget more than one point: *Faber’s formula* (which we are going to describe below) can be proven using the second equation above and the relation $\pi_q^*(\kappa_j) = \kappa_j - \psi_q^j$ (proven in [5]).

Let $Q := \{q_1, \dots, q_m\}$ and let $\pi_Q : \overline{\mathcal{M}}_{g, X \cup Q} \rightarrow \overline{\mathcal{M}}_{g, X}$ be the forgetful map. Then

$$(\pi_Q)_*(\psi_{x_1}^{r_1} \dots \psi_{x_n}^{r_n} \psi_{q_1}^{b_1+1} \dots \psi_{q_m}^{b_m+1}) = \psi_{x_1}^{r_1} \dots \psi_{x_n}^{r_n} K_{b_1 \dots b_m}$$

where $K_{b_1 \dots b_m} = \sum_{\sigma \in \mathfrak{S}_m} \kappa_{b(\sigma)}$ and $\kappa_{b(\sigma)}$ is defined in the following way. If $\gamma = (c_1, \dots, c_l)$ is a cycle, then set $b(\gamma) := \sum_{j=1}^l b_{c_j}$. If $\sigma = \gamma_1 \dots \gamma_\nu$ is the decomposition in disjoint cycles (including 1-cycles), then we let $\kappa_{b(\sigma)} := \prod_{i=1}^\nu \kappa_{b(\gamma_i)}$. We refer to [38] for more details on Faber’s formula, to [5] and [6] for more properties of tautological classes and to [20] (and [58]) for a conjectural description (which is now partially proven) of the tautological rings.

3.5 Kontsevich’s compactification

3.5.1 The line bundle \mathbb{L} . It has been observed by Witten [74] that the intersection theory of κ and ψ classes can be reduced to that of ψ classes only by using the push-pull formula with respect to the forgetful morphisms. Moreover recall that

$$\psi_{x_i} = c_1(\omega_{\pi_{x_i}}(D_{x_i}))$$

on $\overline{\mathcal{M}}_{g, X}$, where $D_{x_i} = \sum_{j \neq i} 0, \{x_i, x_j\}$ (as shown in [74]). So, in order to find a “minimal” projective compactification of $\mathcal{M}_{g, X}$ where to compute the intersection numbers of the ψ classes, it is natural to look at the maps induced by the linear system $\mathbb{L} := \bigotimes_{x_i \in X} \omega_{\pi_{x_i}}(D_{x_i})$. It is well-known that \mathbb{L} is nef and big (Arakelov [4] and Mumford [60]), so that the problem is to decide whether \mathbb{L} is semi-ample and to determine its exceptional locus $E_X(\mathbb{L}^{\otimes d})$ for $d \gg 0$.

It is easy to see that $\mathbb{L}^{\otimes d}$ pulls back to the trivial line bundle via the boundary map $\overline{\mathcal{M}}_{g', \{x'\}} \times \{C\} \rightarrow \overline{\mathcal{M}}_{g, X}$, where C is a fixed curve of genus $g - g'$ with an $X \cup \{x''\}$ -marking and the map glues x' with x'' . Hence the map induced by the linear system $\mathbb{L}^{\otimes d}$ (if base-point-free) should restrict to the projection $\overline{\mathcal{M}}_{g, \{x'\}} \times \overline{\mathcal{M}}_{g-g', X \cup \{x''\}} \rightarrow \overline{\mathcal{M}}_{g-g', X \cup \{x''\}}$ on these boundary components.

Whereas \mathbb{L} is semi-ample in characteristic $p > 0$, it is not so in characteristic 0 (Keel [39]). However, one can still topologically contract the exceptional (with respect to \mathbb{L}) curves to obtain Kontsevich’s map

$$\xi' : \overline{\mathcal{M}}_{g, X} \longrightarrow \overline{\mathcal{M}}_{g, X}^K$$

which is a proper continuous surjection of orbispaces. A consequence of Keel’s result is that the coarse $\overline{\mathcal{M}}_{g, P}^K$ cannot be given a scheme structure such that the contraction

map is a morphism. This is in some sense unexpected, because the morphism behaves as if it were algebraic: in particular, the fiber product $\overline{\mathcal{M}}_{g,X} \times_{\overline{\mathcal{M}}_{g,X}^K} \overline{\mathcal{M}}_{g,X}$ is projective.

Remark 3.5. $\overline{\mathcal{M}}_{g,X}^K$ can be given the structure of a stratified orbispace, where the stratification is again by topological type of the generic curve in the fiber of ξ' . Also, the stabilizer of a point s in $\overline{\mathcal{M}}_{g,X}^K$ will be the same as the stabilizer of the generic point in $(\xi')^{-1}(s)$.

3.5.2 Visibly equivalent curves. So now we leave the realm of algebraic geometry and proceed topologically to construct and describe this different compactification. In fact we introduce a slight modification of Kontsevich's construction (see [43]). We realize it as a quotient of $\overline{\mathcal{M}}_{g,X} \times \Delta_X$ by an equivalence relation, where Δ_X is the standard simplex in \mathbb{R}^X .

If (Σ, \underline{p}) is an element of $\overline{\mathcal{M}}_{g,X} \times \Delta_X$, then we say that an irreducible component of Σ (and so the associated vertex of the dual graph ζ_Σ) is *visible* with respect to \underline{p} if it contains a point $x_i \in X$ such that $p_i > 0$.

Next, we declare that (Σ, \underline{p}) is equivalent to $(\Sigma', \underline{p}')$ if $\underline{p} = \underline{p}'$ and there is a homeomorphism of pointed surfaces $\Sigma \xrightarrow{\sim} \Sigma'$, which is biholomorphic on the visible components of Σ . As this relation would not give back a Hausdorff space we consider its closure, which we are now going to describe.

Consider the following two moves on the dual graph ζ_Σ :

- (1) if two invisible vertices w and w' are joined by an edge e , then we can build a new graph discarding e , merging w and w' along e , thus obtaining a new vertex w'' , which we label with $(g_{w''}, X_{w''}) := (g_w + g_{w'}, X_w \cup X_{w'})$
- (2) if an invisible vertex w has a loop e , we can make a new graph discarding e and relabeling w with $(g_w + 1, X_w)$.

Applying these moves to ζ_Σ iteratively until the process ends, we end up with a *reduced dual graph* $\zeta_{\Sigma, \underline{p}}^{\text{red}}$. Denote by $V_-(\Sigma, \underline{p})$ the subset of invisible vertices and $V_+(\Sigma, \underline{p})$ the subset of visible vertices of $\zeta_{\Sigma, \underline{p}}^{\text{red}}$.

For every couple (Σ, \underline{p}) denote by $\overline{\Sigma}$ the quotient of Σ obtained collapsing every invisible component to a point.

We say that (Σ, \underline{p}) and $(\Sigma', \underline{p}')$ are *visibly equivalent* if $\underline{p} = \underline{p}'$ and there exist a homeomorphism $\overline{\Sigma} \xrightarrow{\sim} \overline{\Sigma}'$, whose restriction to each component is analytic, and a compatible isomorphism $f^{\text{red}}: \zeta_{\Sigma, \underline{p}}^{\text{red}} \xrightarrow{\sim} \zeta_{\Sigma', \underline{p}'}^{\text{red}}$ of reduced dual graphs.

Remark 3.6. In other words, $(\Sigma, \underline{p}), (\Sigma', \underline{p}')$ are visibly equivalent if and only if: $\underline{p} = \underline{p}'$ and there exists a third stable Σ'' and maps $h: \Sigma'' \rightarrow \Sigma$ and $h': \Sigma'' \rightarrow \Sigma'$ such that h, h' are biholomorphic on the visible components and are a stable marking on the invisible components of $(\Sigma'', \underline{p})$ (that is, they may shrink some disjoint simple closed curves to nodes and are homeomorphisms elsewhere).

Finally let

$$\xi: \bar{\mathcal{M}}_{g,X} \times \Delta_X \longrightarrow \bar{\mathcal{M}}_{g,X}^\Delta := \bar{\mathcal{M}}_{g,X} \times \Delta_X / \sim$$

be the quotient map and remark that $\bar{\mathcal{M}}_{g,X}^\Delta$ is compact and that ξ commutes with the projection onto Δ_X .

Similarly, one can say that two (S, X) -marked stable surfaces $([f: S \rightarrow \Sigma], \underline{p})$ and $([f': S \rightarrow \Sigma'], \underline{p}')$ are visibly equivalent if there exists a third stable (S, X) -marked surface $[f'': S \rightarrow \Sigma'']$ and maps $h: \Sigma'' \rightarrow \Sigma$ and $h': \Sigma'' \rightarrow \Sigma'$ such that $h \circ f'' \simeq f$, $h' \circ f'' \simeq f'$ and $(\Sigma, \underline{p}), (\Sigma', \underline{p}')$ are visibly equivalent through h, h' (see the remark above). Consequently, we can define $\bar{\mathcal{T}}^\Delta(S, X)$ as the quotient of $\bar{\mathcal{T}}(S, X) \times \Delta_X$ obtained by identifying visibly equivalent (S, X) -marked surfaces.

For every \underline{p} in Δ_X , we will denote by $\bar{\mathcal{M}}_{g,X}^\Delta(\underline{p})$ the subset of points of the type $[\Sigma, \underline{p}]$. Then it is clear that $\bar{\mathcal{M}}_{g,X}^\Delta(\Delta_X^\circ)$ is in fact homeomorphic to a product $\bar{\mathcal{M}}_{g,X}^\Delta(\underline{p}) \times \Delta_X^\circ$ for any given $\underline{p} \in \Delta_X^\circ$. Observe that $\bar{\mathcal{M}}_{g,X}^\Delta(\underline{p})$ is isomorphic to $\bar{\mathcal{M}}_{g,X}^K$ for all $\underline{p} \in \Delta_X^\circ$ in such a way that

$$\xi_{\underline{p}}: \bar{\mathcal{M}}_{g,X} \cong \bar{\mathcal{M}}_{g,X} \times \{\underline{p}\} \longrightarrow \bar{\mathcal{M}}_{g,X}^\Delta(\underline{p})$$

is identified with ξ' .

Notice, by the way, that the fibers of ξ are isomorphic to moduli spaces. More precisely consider a point $[\Sigma, \underline{p}]$ of $\bar{\mathcal{M}}_{g,X}^\Delta$. For every $w \in V_-(\Sigma, \underline{p})$, denote by Q_w the subset of oriented edges of $\zeta_{\Sigma, \underline{p}}^{\text{red}}$ outgoing from w . Then we have the natural isomorphism

$$\xi^{-1}([\Sigma, \underline{p}]) \cong \prod_{w \in V_-(\Sigma, \underline{p})} \bar{\mathcal{M}}_{g_w, X_w \cup Q_w}$$

according to the fact that $\bar{\mathcal{M}}_{g,X} \times_{\bar{\mathcal{M}}_{g,X}^K} \bar{\mathcal{M}}_{g,X}$ is projective.

4 Cell decompositions of the moduli space of curves

4.1 Harer–Mumford–Thurston construction

One traditional way to associate a weighted arc system to a Riemann surface endowed with weights at its marked points is to look at critical trajectories of Jenkins–Strebel quadratic differentials. Equivalently, to decompose the punctured surface into a union of semi-infinite flat cylinders with lengths assigned to their circumference.

4.1.1 Quadratic differentials. Let Σ be a compact Riemann surface and let φ be a meromorphic quadratic differential, that is $\varphi = \varphi(z)dz^2$ where z is a local holomorphic

coordinate and $\varphi(z)$ is a meromorphic function. Being a quadratic differential means that, if $w = w(z)$ is another local coordinate, then $\varphi = \varphi(w) \left(\frac{dz}{dw}\right)^2 dw^2$.

Regular points of Σ for φ are points where φ has neither a zero nor a pole; *critical points* are zeroes or poles of φ .

We can attach a metric to φ , by simply setting $|\varphi| := \sqrt{\varphi\bar{\varphi}}$. In coordinates, $|\varphi| = |\varphi(z)|dzd\bar{z}$. The metric is well-defined and flat at the regular points and it has conical singularities (with angle $\alpha = (k+2)\pi$) at simple poles ($k = -1$) and at zeroes of order k . Poles of order 2 or higher are at infinite distance.

If P is a regular point, we can pick a local holomorphic coordinate z at $P \in U \subset \Sigma$ such that $z(P) = 0$ and $\varphi = dz^2$ on U . The choice of z is unique up to sign. Thus, $\{Q \in U \mid z(Q) \in \mathbb{R}\}$ defines a real-analytic curve through P on Σ , which is called a *horizontal trajectory* of φ . Similarly, $\{Q \in U \mid z(Q) \in i\mathbb{R}\}$ defines the *vertical trajectory* of φ through P .

Horizontal (resp. vertical) trajectories τ are intrinsically defined by asking that the restriction of φ to τ is a positive-definite (resp. negative-definite) symmetric bilinear form on the tangent bundle of τ .

If φ has at worst double poles, then the local aspect of horizontal trajectories is as in Figure 6 (horizontal trajectories through q are drawn thicker).

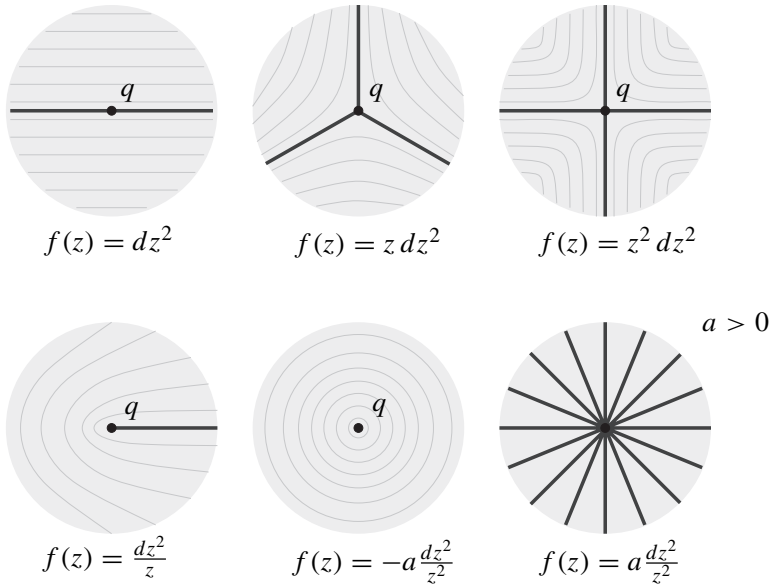


Figure 6. Local structure of horizontal trajectories.

Trajectories are called *critical* if they meet a critical point. It follows from the general classification (see [70]) that

- a trajectory is *closed* if and only if it is either periodic or it starts and ends at a critical point;

- if a horizontal trajectory τ is *periodic*, then there exists a maximal open annular domain $A \subset \Sigma$ and a number $c > 0$ such that

$$(A, \varphi|_A) \xrightarrow{\sim} \left(\{z \in \mathbb{C} \mid r < |z| < R\}, -c \frac{dz^2}{z^2} \right)$$

and, under this identification, $\tau = \{z \in \mathbb{C} \mid h = |z|\}$ for some $h \in (r, R)$;

- if all horizontal trajectories are closed of finite length, then φ has at most double poles and there it has negative quadratic residue (i.e. at a double pole, φ looks like $-a \frac{dz^2}{z^2}$, with $a > 0$).

4.1.2 Jenkins–Strebel differentials. There are many theorems about existence and uniqueness of quadratic differentials φ with specific behaviors of their trajectories and about their characterization using extremal properties of the associated metric $|\varphi|$ (see Jenkins [37]). The following result is the one we are interested in.

Theorem 4.1 (Strebel [69]). *Let Σ be a compact Riemann surface of genus g and $X = \{x_1, \dots, x_n\} \subset \Sigma$ such that $2g - 2 + n > 0$. For every $(p_1, \dots, p_n) \in \mathbb{R}_+^X$ there exists a unique quadratic differential φ such that*

- (a) φ is holomorphic on $\Sigma \setminus X$,
- (b) all horizontal trajectories of φ are closed,
- (c) it has a double pole at x_i with quadratic residue $-\left(\frac{p_i}{2\pi}\right)^2$,
- (d) the only annular domains of φ are pointed discs at the x_i 's.

Moreover, φ depends continuously on Σ and on $\underline{p} = (p_1, \dots, p_n)$.

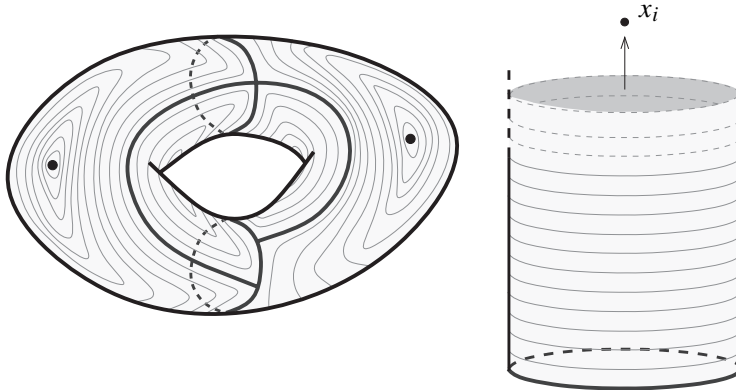


Figure 7. Example of horizontal foliation of a Jenkins–Strebel differential.

Remark 4.2. Notice that the previous result establishes the existence of a continuous map

$$\mathbb{R}_+^X \longrightarrow \{\text{continuous sections of } \mathcal{Q}(S, 2X) \rightarrow \mathcal{T}(S, X)\}$$

where $\mathcal{Q}(S, 2X)$ is the vector bundle whose fiber over $[f: S \rightarrow \Sigma]$ is the space of quadratic differentials on Σ , which can have double poles at X and are holomorphic elsewhere. Hubbard and Masur [31] proved (in a slightly different case, though) that the sections of $\mathcal{Q}(S, 2X)$ in the image of the above map are piecewise real-analytic and gave precise equations for them.

Quadratic differentials that satisfy (a) and (b) are called *Jenkins–Strebel differentials*. They are particularly easy to understand because their critical trajectories form a graph $G = G_{\Sigma, \underline{p}}$ embedded inside the surface Σ and G decomposes Σ into a union of cylinders (with respect to the flat metric $|\varphi|$), whose circumferences are horizontal trajectories.

Property (d) is telling us that $\Sigma \setminus X$ retracts by deformation onto G , flowing along the vertical trajectories out of X .

Remark 4.3. It can be easily seen that Theorem 4.1 still holds for $p_1, \dots, p_n \geq 0$ but $\underline{p} \neq 0$. Condition (d) can be rephrased by saying that every annular domain corresponds to some x_i for which $p_i > 0$, and that $x_j \in G$ if $p_j = 0$. It is still true that $\Sigma \setminus X$ retracts by deformation onto G .

We sketch the traditional existence proof of Theorem 4.1.

Definition 4.4. The *modulus* of a standard annulus $A(r, R) = \{z \in \mathbb{C} \mid r < |z| < R\}$ is $m(A(r, R)) = \frac{1}{2\pi} \log(R/r)$ and the modulus of an annulus A is defined to be that of a standard annulus biholomorphic to A . Consider a simply connected domain $0 \in U \subset \mathbb{C}$ and let z be a holomorphic coordinate at 0. The *reduced modulus* of the annulus $U^* = U \setminus \{0\}$ is $m(U^*, z) = m(U^* \cap \{|z| > \varepsilon\}) + \frac{1}{2\pi} \log(\varepsilon)$, which is independent of the choice of a sufficiently small $\varepsilon > 0$.

Notice that the *extremal length* E_γ of a circumference γ inside $A(r, R)$ is exactly $1/m(A(r, R))$.

Existence of Jenkins–Strebel differential. Fix holomorphic coordinates z_1, \dots, z_n at x_1, \dots, x_n . A *system of annuli* is a holomorphic injection $s: \Delta \times X \hookrightarrow \Sigma$ such that $s(0, x_i) = x_i$, where Δ is the unit disc in \mathbb{C} . We write $m_i(s)$ for the reduced modulus $m(s(\Delta \times \{x_i\}), z_i)$ and define the functional

$$F: \{\text{systems of annuli}\} \longrightarrow \mathbb{R},$$

$$s \longmapsto \sum_{i=1}^n p_i^2 m_i(s),$$

which is bounded above, because $\Sigma \setminus X$ is hyperbolic. A maximizing sequence s_n converges (up to extracting a subsequence) to a system of annuli s_∞ . Let $D_i = s_\infty(\Delta \times \{x_i\})$. Notice that the restriction of s_∞ to $\Delta \times \{x_i\}$ is injective if $p_i > 0$ and is constantly x_i if $p_i = 0$.

Clearly, s_∞ is maximizing for every choice of z_1, \dots, z_n and so we can assume that, whenever $p_i > 0$, z_i is the coordinate induced by s_∞ .

Define the L^1_{loc} -quadratic differential φ on $\Sigma \setminus X$ as

$$\varphi := \begin{cases} \left(-\frac{p_i^2}{4\pi^2} \frac{dz_i^2}{z_i^2} \right) & \text{on } D_i \text{ if } p_i > 0, \\ 0 & \text{elsewhere.} \end{cases}$$

Notice that $F(s_\infty) = \|\varphi\|_{\text{red}}$, where the *reduced norm* is given by

$$\|\varphi\|_{\text{red}} := \int_\Sigma \left[|\varphi|^2 - \sum_{i:p_i>0} \frac{p_i^2}{4\pi^2} \frac{dz_i d\bar{z}_i}{|z_i|^2} \chi(|z_i| < \varepsilon_i) \right] + \sum_{i=1}^n \frac{p_i^2}{2\pi} \log(\varepsilon_i)$$

which is independent of the choice of sufficiently small $\varepsilon_1, \dots, \varepsilon_n > 0$.

As s_∞ is a stationary point for F , so is for $\|\cdot\|_{\text{red}}$. Thus, for every smooth vector field $V = V(z)\partial/\partial z$ on Σ , compactly supported on $\Sigma \setminus X$, the first-order variation of

$$\|f_t^*(\varphi)\|_{\text{red}} = \|\varphi\|_{\text{red}} + 2t \int_S \text{Re}(\varphi \bar{\partial} V) + o(t)$$

must vanish, where $f_t = \exp(tV)$. Thus, φ is holomorphic on $\Sigma \setminus X$ by Weyl's lemma and it satisfies all the requirements. \square

4.1.3 The nonsingular case. Using the construction described above, we can attach to every (Σ, X, p) a graph $G_{\Sigma, p} \subset \Sigma$ (and thus an (S, X) -marked ribbon graph $\mathbb{G}_{\Sigma, p}$) which is naturally metrized by $|\varphi|$. By arc/graph duality (in the nonsingular case, see 2.2.8), we also have a weighted proper system of arcs in Σ . Notice that, because of (c), the boundary weights are exactly p_1, \dots, p_n .

If $[f : S \rightarrow \Sigma]$ is a point in $\mathcal{T}(S, X)$ and $\underline{p} \in (\mathbb{R}_{\geq 0}^X) \setminus \{0\}$, then the previous construction (which is explicitly mentioned by Harer in [28], where he attributes it to Mumford and Thurston) provides a point in $|\mathfrak{A}^\circ(S, X)| \times \mathbb{R}_+$. It is however clear that, if $a > 0$, then the Strebel differential associated to $(\Sigma, a\underline{p})$ is $a\varphi$. Thus, we can just consider $\underline{p} \in \mathbb{P}(\mathbb{R}_{\geq 0}^X) \cong \Delta_X$, so that the corresponding weighted arc system belongs to $|\mathfrak{A}^\circ(S, X)|$ (after multiplying by a factor 2).

Because of the continuous dependence of φ on Σ and \underline{p} , the map

$$\Psi_{\text{JS}} : \mathcal{T}(S, X) \times \Delta_X \longrightarrow |\mathfrak{A}^\circ(S, X)|$$

is *continuous*.

We now show that a point $\bar{w} \in |\mathfrak{A}^\circ(S, X)|$ determines exactly one (S, X) -marked surface, which proves that Ψ_{JS} is *bijective*.

By 2.2.9, we can associate a metrized (S, X) -marked nonsingular ribbon graph \mathbb{G}_α to each $w \in |\mathfrak{A}^\circ(S, X)|_{\mathbb{R}}$ supported on α . However, if we realize $|\mathbb{G}_\alpha|$ by gluing semi-infinite tiles $T_{\vec{\alpha}_i}$ of the type $[0, w(\alpha_i)]_x \times [0, \infty)_y \subset \hat{\mathbb{C}}_z$, which naturally come together with a complex structure and a quadratic differential dz^2 , then $|\mathbb{G}_\alpha|$ becomes

a Riemann surface endowed with the (unique) Jenkins–Strebel quadratic differential φ determined by Theorem 4.1. Thus, $\Psi_{\text{JS}}^{-1}(w) = ([f: S \rightarrow |\mathbb{G}_{\underline{\alpha}}|], \underline{p})$, where p_i is obtained from the quadratic residue of φ at x_i . Moreover, the length function defined on $|\mathfrak{A}^\circ(S, X)|_{\mathbb{R}}$ exactly corresponds to the $|\varphi|$ -length function on $\mathcal{T}(S, X) \times \Delta_X \times \mathbb{R}_+$.

Notice that Ψ_{JS} is $\Gamma(S, X)$ -equivariant by construction and so induces a continuous bijection $\bar{\Psi}_{\text{JS}}: \mathcal{M}_{g, X} \times \Delta_X \rightarrow |\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$ on the quotient. If we prove that $\bar{\Psi}_{\text{JS}}$ is *proper*, then $\bar{\Psi}_{\text{JS}}$ is a homeomorphism. To conclude that Ψ_{JS} is a homeomorphism too, we will use the following.

Lemma 4.5. *Let Y and Z be metric spaces acted on discontinuously by a discrete group of isometries G and let $h: Y \rightarrow Z$ be a G -equivariant continuous injection such that the induced map $\bar{h}: Y/G \rightarrow Z/G$ is a homeomorphism. Then h is a homeomorphism.*

Proof. To show that h is surjective, let $z \in Z$. Because \bar{h} is bijective, there exists a unique $[y] \in Y/G$ such that $\bar{h}([y]) = [z]$. Hence, $h(y) = z \cdot g$ for some $g \in G$ and so $h(y \cdot g^{-1}) = z$.

To prove that h^{-1} is continuous, let $(y_m) \subset Y$ be a sequence such that $h(y_m) \rightarrow h(y)$ as $m \rightarrow \infty$ for some $y \in Y$. Clearly, $[h(y_m)] \rightarrow [h(y)]$ in Z/G and so $[y_m] \rightarrow [y]$ in Y/G , because \bar{h} is a homeomorphism. Let $(v_m) \subset Y$ be a sequence such that $[v_m] = [y_m]$ and $v_m \rightarrow y$ and denote by $g_m \in G$ the element such that $y_m = v_m \cdot g_m$. By continuity of h , we have $d_Z(h(v_m), h(y)) \rightarrow 0$ and by hypothesis $d_Z(h(v_m) \cdot g_m, h(y)) \rightarrow 0$. Hence, $d_Z(h(y), h(y) \cdot g_m) \rightarrow 0$ and so $g_m \in \text{stab}(h(y)) = \text{stab}(y)$ for large m , because G acts discontinuously on Z . As a consequence, $y_m \rightarrow y$ and so h^{-1} is continuous. \square

The final step is the following.

Lemma 4.6. $\bar{\Psi}_{\text{JS}}: \mathcal{M}_{g, X} \times \Delta_X \rightarrow |\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$ is *proper*.

Proof. Let $([\Sigma_m], \underline{p}_m)$ be a diverging sequence in $\mathcal{M}_{g, X} \times \Delta_X$ and denote by \mathfrak{l}_m the hyperbolic metric on $\Sigma \setminus X$. By the Mumford–Mahler criterion, there exist simple closed hyperbolic geodesics $\gamma_m \subset \Sigma_m$ such that $\ell_{\mathfrak{l}_m}(\gamma_m) \rightarrow 0$. By Maskit’s inequalities comparing extremal and hyperbolic length [50], we conclude that the extremal length $E(\gamma_m) \rightarrow 0$.

Consider now the metric $|\varphi_m|$ induced by the Jenkins–Strebel differential φ_m uniquely determined by $(\Sigma_m, \underline{p}_m)$. Denote by $\ell_\varphi(\gamma_m)$ the length of the unique geodesic $\tilde{\gamma}_m$ with respect to the metric $|\varphi_m|$, freely homotopic to $\gamma_m \subset \Sigma_m$. Notice that $\tilde{\gamma}_m$ is a union of critical horizontal trajectories.

Because $|\varphi_m|$ has infinite area, define a modified metric g_m on Σ_m in the same conformal class as $|\varphi_m|$ as follows.

- g_m agrees with $|\varphi_m|$ on the critical horizontal trajectories of φ_m .

- Whenever $p_{i,m} > 0$, consider a coordinate z at x_i such that the annular domain of φ_m at x_i is exactly $\Delta^* = \{z \in \mathbb{C} \mid 0 < |z| < 1\}$ and $\varphi_m = -\frac{p_{i,m}^2 dz^2}{4\pi^2 z^2}$. Then define g_m to agree with $|\varphi_m|$ on $\exp(-2\pi/p_{i,m}) \leq |z| < 1$ (which becomes isometric to a cylinder of circumference $p_{i,m}$ and height 1, so with area $p_{i,m}$) and to be the metric of a flat Euclidean disc of circumference $p_{i,m}$ centered at $z = 0$ (so with area $\pi p_{i,m}^2$) on $|z| < \exp(-2\pi/p_{i,m})$.

Notice that the total area $A(g_m)$ is $\pi(p_{1,m}^2 + \dots + p_{n,m}^2) + (p_{1,m} + \dots + p_{n,m}) \leq \pi + 1$.

Call $\ell_g(\gamma_m)$ the length of the shortest g_m -geodesic $\hat{\gamma}_m$ in the class of γ_m . By definition, $\ell_g(\gamma_m)^2/A(g_m) \leq E(\gamma_m) \rightarrow 0$ and so $\ell_g(\gamma_m) \rightarrow 0$. As a g_m -geodesic is either longer than 1 or contained in the critical graph of φ , then $\hat{\gamma}_m$ coincides with $\tilde{\gamma}_m$ for $m \gg 0$.

Hence, $\ell_\varphi(\gamma_m) \rightarrow 0$ and so $\text{sys}(\bar{w}_m) \rightarrow 0$. By Lemma 2.2, we conclude that $\bar{\Psi}_{\text{JS}}(\Sigma_m, \underline{p}_m)$ diverges in $|\mathfrak{A}^\circ(S, X)|/\Gamma(S, X)$. □

Remark 4.7. Suppose that $([f_m : S \rightarrow \Sigma_m], \underline{p}_m)$ converges to $([f : S \rightarrow \Sigma], \underline{p}) \in \bar{\mathcal{T}}_{g,X} \times \Delta_X$ and let $\Sigma' \subset \Sigma$ be an invisible component, that is a component of Σ with no positively weighted marked points. Then, $S' = f^{-1}(\Sigma')$ is bounded by simple closed curves $\gamma_1, \dots, \gamma_k \subset S$ and $\ell_{f_m^* \varphi_m}(\gamma_i) \rightarrow 0$ for $i = 1, \dots, k$. Just analyzing the shape of the critical graph of φ_m , one can check that $\ell_{\varphi_m}(\gamma) \leq \sum_{i=1}^k \ell_{\varphi_m}(\gamma_i)$ for all $\gamma \subset S'$. Hence, $\ell_{f_m^* \varphi_m}(\gamma) \rightarrow 0$ uniformly in γ , and so $f_m^* \varphi_m$ tends to zero uniformly on the compact subsets of $(S')^\circ$.

4.1.4 The case of stable curves. We want to extend the map Ψ_{JS} to Deligne–Mumford’s augmentation and, by abuse of notation, we will still write

$$\Psi_{\text{JS}} : \bar{\mathcal{T}}(S, X) \times \Delta_X \rightarrow |\mathfrak{A}(S, X)|$$

for this extension.

Given $([f : S \rightarrow \Sigma], \underline{p})$, we can construct a Jenkins–Strebel differential φ on each visible component of Σ , by considering nodes as marked points with zero weight. Extend φ to zero over the invisible components. Clearly, φ is a holomorphic section of $\omega_\Sigma^{\otimes 2}(2X)$ (the square of the logarithmic dualizing sheaf on Σ): call it the *Jenkins–Strebel differential associated to (Σ, \underline{p})* . Notice that it clearly maximizes the functional F , used in the proof of Theorem 4.1.

As φ defines a metrized ribbon graph for each visible component of Σ , one can easily see that thus we have an (S, X) -marked enriched ribbon graph \mathbb{G}^{en} (see 2.2.4), where ζ is the dual graph of Σ and V_+ is the set of visible components of (Σ, \underline{p}) , m is determined by the X -marking and s by the position of the nodes.

By arc/graph duality (see 2.2.13), we obtain a system of arcs $\underline{\alpha}$ in (S, X) and the metrics provide a system of weights \bar{w} with support on $\underline{\alpha}$. This defines the set-theoretic extension of Ψ_{JS} . Clearly, it is still $\Gamma(S, X)$ -equivariant and it identifies visibly equivalent (S, X) -marked surfaces. Thus, it descends to a bijection

$\Psi_{JS}: \bar{\mathcal{T}}^\Delta(S, X) \rightarrow |\mathfrak{A}(S, X)|$ and we also have

$$\bar{\Psi}_{JS}: \bar{\mathcal{M}}_{g,X}^\Delta \longrightarrow |\mathfrak{A}(S, X)|/\Gamma(S, X)$$

where $|\mathfrak{A}(S, X)|/\Gamma(S, X)$ can be naturally given the structure of an orbispace (essentially, forgetting the Dehn twists along curves of S that are shrunk to points, so that the stabilizer of an arc system just becomes the automorphism group of the corresponding enriched X -marked ribbon graph).

The only thing left to prove is that Ψ_{JS} is continuous. In fact, $\bar{\mathcal{M}}_{g,X}^\Delta$ is compact and $|\mathfrak{A}(S, X)|/\Gamma(S, X)$ is Hausdorff: hence, $\bar{\Psi}_{JS}$ would be (continuous and) automatically proper, and so a homeomorphism. Using Lemma 4.5 again (using a metric pulled back from $\bar{\mathcal{M}}_{g,X}^\Delta$), we could conclude that Ψ_{JS} is a homeomorphism too.

Continuity of Ψ_{JS} . Consider a differentiable stable family

$$\begin{array}{ccc} S \times [0, \varepsilon] & \xrightarrow{f} & \mathcal{C} \\ & \searrow & \downarrow g \\ & & [0, \varepsilon] \end{array}$$

of (S, X) -marked curves (that is, obtained restricting to $[0, \varepsilon]$ a smooth family over the unit disc Δ), such that g is topologically trivial over $(0, \varepsilon]$ with fiber a curve with k nodes. Let also $\underline{p}: [0, \varepsilon] \rightarrow \Delta_X$ be a differentiable family of weights.

We can assume that there are disjoint simple closed curves $\gamma_1, \dots, \gamma_k, \eta_1, \dots, \eta_h \subset S$ such that $f(\gamma_i \times \{t\})$ is a node for all t , that $f(\eta_j \times \{t\})$ is a node for $t = 0$ and that \mathcal{C}_t is smooth away from these nodes.

Fix a nonempty open relatively compact subset K of $S \setminus (\gamma_1 \cup \dots \cup \gamma_k \cup \eta_1 \cup \dots \cup \eta_h)$ that intersects every connected component. Define a reduced L^1 norm of a section ψ_t of $\omega_{\mathcal{C}_t}^{\otimes 2}(2X)$ to be $\|\psi\|_{\text{red}} = \int_{f_t(K)} |\psi|$. Notice that L^1 convergence of holomorphic sections ψ_t as $t \rightarrow 0$ implies uniform convergence of $f_t^* \psi_t$ on the compact subsets of $S \setminus (\gamma_1 \cup \dots \cup \gamma_k \cup \eta_1 \cup \dots \cup \eta_h)$.

Denote by φ_t the Jenkins–Strebel differential associated to $(\mathcal{C}_t, \underline{p}_t)$ with annular domains $D_{1,t}, \dots, D_{n,t}$.

As all the components of \mathcal{C}_t are hyperbolic, $\|\varphi_t\|_{\text{red}}$ is uniformly bounded and we can assume (up to extracting a subsequence) that φ_t converges to a holomorphic section φ'_0 of $\omega_{\mathcal{C}_0}^{\otimes 2}(2X)$ in the reduced norm. Clearly, φ'_0 will have double poles at x_i with prescribed residue.

Remark 4.7 implies that φ'_0 vanishes on the invisible components of \mathcal{C}_0 , whereas it certainly does not on the visible ones.

For all those $(i, t) \in \{1, \dots, n\} \times [0, \varepsilon]$ such that $p_{i,t} > 0$, let $z_{i,t}$ be the coordinate at x_i (uniquely defined up to phase) given by $z_{i,t} = u_{i,t}^{-1}|_{D_{i,t}}$ and

$$u_{i,t}: \bar{\Delta} \longrightarrow \bar{D}_{i,t} \subset \mathcal{C}_t$$

is continuous on $\bar{\Delta}$ and biholomorphic in the interior for all $t > 0$ and $\varphi_t|_{D_{i,t}} = -\frac{p_{i,t}^2 dz_{i,t}^2}{4\pi^2 z_{i,t}^2}$ for $t \geq 0$. Whenever $p_{i,t} = 0$, choose $z_{i,t}$ such that $\varphi_t|_{D_{i,t}} = z^k dz^2$, with $k = \text{ord}_{x_i} \varphi_t$. When $p_{i,t} > 0$, we can choose the phases of $u_{i,t}$ in such a way that $u_{i,t}$ vary continuously with $t \geq 0$.

If $p_{i,0} = 0$, then set $D_{i,0} = \emptyset$. Otherwise, $p_{i,0} > 0$ and so $D_{i,0}$ cannot shrink to $\{x_i\}$ (because F_t would go to $-\infty$ as $t \rightarrow 0$). In this case, denote by $D_{i,0}$ the region $\{|z_{i,0}| < 1\} \subset \mathcal{C}_0$. Notice that φ'_0 has a double pole at x_i with residue $p_{i,0} > 0$ and clearly $\varphi'_0|_{D_{i,0}} = -\frac{p_{i,0}^2 dz_{i,0}^2}{4\pi^2 z_{i,0}^2}$.

We want to prove that the visible subsurface of \mathcal{C}_0 is covered by $\bigcup_i \bar{D}_{i,0}$ and so φ'_0 is a Jenkins–Strebel differential on each visible component of \mathcal{C}_0 . By uniqueness, it must coincide with φ_0 .

Consider a point y in the interior of $f_0^{-1}(\mathcal{C}_{0,+}) \setminus X$. For every $t > 0$ there exists a $y_t \in S$ such that $f_t(y_t)$ does not belong to the critical graph of φ_t and the $f_t^*|\varphi_t|$ -distance $d_t(y, y_t) < t$. As $\varphi_t \rightarrow \varphi_0$ in reduced norm and $y, y_t \notin X$, then $d_0(y, y_t) \rightarrow 0$ as $t \rightarrow 0$.

We can assume (up to discarding some t 's) that $f_t(y_t)$ belongs to $D_{i,t}$ for a fixed i and in particular that $f_t(y_t) = u_{i,t}(c_t)$ for some $c_t \in \Delta$. Up to discarding some t 's, we can also assume that $c_t \rightarrow c_0 \in \bar{\Delta}$. Denote by y'_t the point given by $f_0(y'_t) = u_{i,0}(c_t)$.

$$\begin{aligned} d_0(y'_t, y) &\leq d_0(y_t, y) + d_0(y'_t, y_t) \leq d_0(y_t, y) + d_0(f_0^{-1}u_{i,0}(c_t), f_t^{-1}u_{i,t}(c_t)) \\ &\leq d_0(y_t, y) + d_0(f_0^{-1}u_{i,0}(c_t), f_0^{-1}u_{i,0}(c_0)) \\ &\quad + d_0(f_0^{-1}u_{i,0}(c_0), f_t^{-1}u_{i,t}(c_0)) + d_0(f_t^{-1}u_{i,t}(c_0), f_t^{-1}u_{i,t}(c_t)) \end{aligned}$$

and all terms go to zero as $t \rightarrow 0$. Thus, every point in the smooth locus $\mathcal{C}_{0,+} \setminus X$ is at $|\varphi_0|$ -distance zero from some $D_{i,0}$. Hence, φ_0 is a Jenkins–Strebel differential on the visible components.

With a few simple considerations, one can easily conclude that

- the zeroes of φ_t move continuously as $t \in [0, \varepsilon]$;
- for every edge e_0 of the critical graph of φ_0 , and for t small enough, there are corresponding edges e_t of the critical graphs of φ_t such that $e_t \rightarrow e_0$ and $\ell_{|\varphi_t|}(e_t) \rightarrow \ell_{|\varphi_0|}(e_0)$;
- the critical graph of φ_t converges to that of φ_0 for the Gromov–Hausdorff distance.

Thus, the associated weighted arc systems $\bar{w}_t \in |\mathfrak{A}(S, X)|$ converge to \bar{w}_0 for $t \rightarrow 0$. \square

Thus, we have proved the following result, claimed by Kontsevich in [43] (see Looijenga's [45] and Zvonkine's [82]).

Proposition 4.8. *The map defined above*

$$\Psi_{\text{JS}}: \bar{\mathcal{T}}^\Delta(S, X) \longrightarrow |\mathfrak{A}(S, X)|$$

is a $\Gamma(S, X)$ -equivariant homeomorphism, which commutes with the projection onto Δ_X . Hence, $\overline{\Psi}_{JS}: \overline{\mathcal{M}}_{g,X}^\Delta \rightarrow |\mathfrak{A}(S, X)|/\Gamma(S, X)$ is a homeomorphism of orbispaces too.

A consequence of the previous proposition and of 2.2.13 is that the realization $B\mathfrak{R}\mathfrak{G}_{g,X,ns}$ is the classifying space of $\Gamma(S, X)$ and that $B\mathfrak{R}\mathfrak{G}_{g,X} \rightarrow \overline{\mathcal{M}}_{g,X}^\Delta$ is a homotopy equivalence (in the category of orbispaces).

4.2 Penner–Bowditch–Epstein construction

The other traditional way to obtain a weighted arc system out of a Riemann surface with weighted marked points is to look at the spine of the truncated surface obtained by removing horoballs of prescribed circumference. Equivalently, to decompose the surface into a union of hyperbolic cusps.

4.2.1 Spines of hyperbolic surfaces. Let $[f: S \rightarrow \Sigma]$ be an (S, X) -marked hyperbolic surface and let $\underline{p} \in \Delta_X$. Denote by $H_i \subset \Sigma$ the horoball at x_i with circumference p_i (as $p_i \leq 1$, the horoball is embedded in Σ) and let $\Sigma_{tr} = \Sigma \setminus \bigcup_i H_i$ be the *truncated surface*. The datum $(\Sigma, \partial H_1, \dots, \partial H_n)$ is also called a *decorated surface*.

For every $y \in \Sigma \setminus X$ at finite distance from $\partial \Sigma_{tr}$, let the *valence* $\text{val}(y)$ be the number of paths that realize $\text{dist}(y, \partial \Sigma_{tr})$, which is generically 1. We will call a *projection* of y a point on $\partial \Sigma_{tr}$ which is at shortest distance from y : clearly, there are $\text{val}(y)$ of them.

Let the *spine* $\text{Sp}(\Sigma, \underline{p})$ be the locus of points of Σ which are at finite distance from $\partial \Sigma_{tr}$ and such that $\text{val}(y) \geq 2$ (see Figure 8).

In particular, $\text{val}^{-1}(2)$ is a disjoint union of finitely many geodesic arcs (the *edges*) and $\text{val}^{-1}(\{3, \infty\})$ is a finite collection of points (the *vertices*). If $p_i = 0$, then we include x_i in $\text{Sp}(\Sigma, \underline{p})$ and we consider it a vertex. Its valence is defined to be the number of half-edges of the spine incident at x_i .

There is a deformation retraction of $\Sigma_{tr} \cap \Sigma_+$ (where Σ_+ is the visible subsurface) onto $\text{Sp}(\Sigma, \underline{p})$, defined on $\text{val}^{-1}(1)$ simply flowing away from $\partial \Sigma_{tr}$ along the unique geodesic that realizes the distance from $\partial \Sigma_{tr}$.

This shows that $\text{Sp}(\Sigma, \underline{p})$ defines an (S, X) -marked enriched ribbon graph \mathbb{G}_{sp}^{en} . By arc/graph duality, we also have an associated *spinal arc system* $\underline{\alpha}_{sp} \in \mathfrak{A}(S, X)$.

4.2.2 Horocyclic lengths and weights. As Σ is a hyperbolic surface, we could metrize $\text{Sp}(\Sigma, \underline{p})$ by inducing a length on each edge. However, the relation between this metric and \underline{p} would be complicated.

Instead, for every edge e of \mathbb{G}_{sp}^{en} (that is, of $\text{Sp}(\Sigma, \underline{p})$), consider one of its two projections $\text{pr}(e)$ to $\partial \Sigma_{tr}$ and define $\ell(e)$ to be the *horocyclic length* of e , that is the hyperbolic length of $\text{pr}(e)$, which clearly does not depend on the chosen projection. Thus, the boundary weights vector ℓ_∂ is exactly \underline{p} .

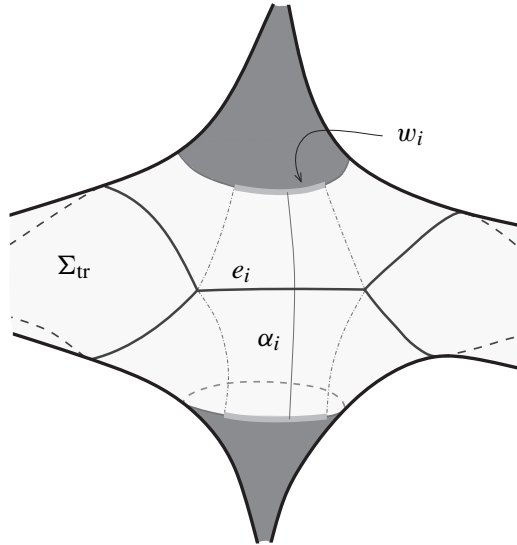


Figure 8. Weights come from lengths of horocyclic arcs.

This endows \mathbb{G}_{sp}^{en} with a metric and so $\underline{\alpha}_{sp}$ with a projective weight $\bar{w}_{sp} \in |\mathfrak{A}(S, X)|$. Notice that visibly equivalent surfaces are associated to the same point of $|\mathfrak{A}(S, X)|$.

This defines a $\Gamma(S, X)$ -equivariant map

$$\Phi_0: \bar{\mathcal{T}}^\Delta(S, X) \longrightarrow |\mathfrak{A}(S, X)|$$

that commutes with the projection onto Δ_X .

Penner [61] proved that the restriction of Φ_0 to $\mathcal{T}(S, X) \times \Delta_X^\circ$ is a homeomorphism; the statement that the whole Φ_0 is a homeomorphism appears in Bowditch–Epstein’s [13] (and a very detailed treatment will appear in [7]). We refer to these papers for a proof of this result.

4.3 Hyperbolic surfaces with boundary

The purpose of this informal subsection is to briefly illustrate the bridge between the cellular decomposition of the Teichmüller space obtained using Jenkins–Strebel differentials and that obtained using spines of decorated surfaces.

4.3.1 Teichmüller and moduli space of hyperbolic surfaces. Fix a compact oriented surface S as before and $X = \{x_1, \dots, x_n\} \subset S$ a nonempty subset.

A *stable hyperbolic surface* Σ is a nodal surface such that $\Sigma \setminus \{\text{nodes}\}$ is hyperbolic with geodesic boundary and/or cusps. Notice that, by convention, $\partial\Sigma$ includes the cusps but it does not include the possible nodes of Σ .

An X -marking of a (stable) hyperbolic surface Σ is a bijection $X \rightarrow \pi_0(\partial\Sigma)$.

An (S, X) -marking of the (stable) hyperbolic surface Σ is an isotopy class of maps $f: S \setminus X \rightarrow \Sigma$, that may shrink disjoint simple closed curves to nodes and are homeomorphisms onto $\Sigma \setminus (\partial\Sigma \cup \{\text{nodes}\})$ elsewhere.

Let $\overline{\mathcal{T}}^\partial(S, X)$ be the Teichmüller space of (S, X) -marked stable hyperbolic surfaces. There is a natural map $\ell_\partial: \overline{\mathcal{T}}^\partial(S, X) \rightarrow \mathbb{R}_{\geq 0}^X$ that associates to $[f: S \rightarrow \Sigma]$ the boundary lengths of Σ , which thus descends to $\bar{\ell}_\partial: \overline{\mathcal{M}}_{g,X}^\partial \rightarrow \mathbb{R}_{\geq 0}^X$. Write $\overline{\mathcal{T}}^\partial(S, X)(\underline{p})$ (resp. $\overline{\mathcal{M}}_{g,X}^\partial(\underline{p})$) for the leaf $\ell_\partial^{-1}(\underline{p})$ (resp. $\bar{\ell}_\partial^{-1}(\underline{p})$).

There is an obvious identification between $\overline{\mathcal{T}}^\partial(S, X)(0)$ (resp. $\overline{\mathcal{M}}_{g,X}^\partial(0)$) and $\overline{\mathcal{T}}(S, X)$ (resp. $\overline{\mathcal{M}}_{g,X}$).

Let $\hat{\mathcal{M}}_{g,X}$ be the blow-up of $\overline{\mathcal{M}}_{g,X}^\partial$ along $\overline{\mathcal{M}}_{g,X}^\partial(0)$: the exceptional locus can be naturally identified with the space of (projectively) decorated surfaces with cusps (which is homeomorphic to $\overline{\mathcal{M}}_{g,X} \times \Delta_X$). Define similarly $\hat{\mathcal{T}}(S, X)$.

4.3.2 Tangent space to the moduli space. The conformal analogue of a hyperbolic surface with geodesic boundary Σ is a Riemann surface with real boundary. In fact, the double of Σ is a hyperbolic surface with no boundary and an orientation-reversing involution, that is a Riemann surface with an anti-holomorphic involution. As a consequence, $\partial\Sigma$ is a real-analytic submanifold.

This means that first-order deformations of a smooth Σ are determined by Beltrami differentials on Σ which are real on $\partial\Sigma$, and so $T_{[\Sigma]} \overline{\mathcal{M}}_{g,X}^\partial \cong H^{0,1}(\Sigma, T_\Sigma)$, where T_Σ is the sheaf of tangent vector fields $V = V(z)\partial/\partial z$, which are real on $\partial\Sigma$.

Dually, the cotangent space $T_{[\Sigma]}^* \overline{\mathcal{M}}_{g,X}^\partial$ is given by the space $\mathcal{Q}(\Sigma)$ of holomorphic quadratic differentials that are real on $\partial\Sigma$. If we write $\mathcal{H}(\Sigma) = \{\bar{\varphi}/\lambda \mid \varphi \in \mathcal{Q}(\Sigma)\}$, where λ is the hyperbolic metric on Σ , then $H^{0,1}(\Sigma, T_\Sigma)$ is identified with the space of harmonic Beltrami differentials $\mathcal{H}(\Sigma)$.

If Σ has nodes, then the situation is more complicated. The logarithmic cotangent bundle $T_{\overline{\mathcal{M}}_{g,X}^\partial}^*$ ($\partial\overline{\mathcal{M}}_{g,X}^\partial$) can be related to quadratic differentials with double poles at the nodes (with the same quadratic residue on both branches). Details can be found in [9] and [51].

4.3.3 Weil–Petersson metric. Mimicking what is done for surfaces with cusps, we can define Hermitian pairings on $\mathcal{Q}(\Sigma)$ and $\mathcal{H}(\Sigma)$, where Σ is a smooth hyperbolic surface with boundary. In particular,

$$h(\mu, \nu) = \int_\Sigma \mu \bar{\nu} \cdot \lambda,$$

$$h^*(\varphi, \psi) = \int_\Sigma \frac{\varphi \bar{\psi}}{\lambda}$$

where $\mu, \nu \in \mathcal{H}(\Sigma)$ and $\varphi, \psi \in \mathcal{Q}(\Sigma)$.

Thus, if $h = g + i\omega$, then g is the Weil–Petersson Riemannian metric and ω is the Weil–Petersson form. Write similarly $h^* = g^* + i\omega^*$, where g^* is the cometric dual to g and ω^* is the Weil–Petersson bivector field.

Notice that ω and ω^* are degenerate. This can be easily seen, because Wolpert’s formula $\omega = \sum_i d\ell_i \wedge d\tau_i$ still holds. We can also conclude that the symplectic leaves of ω^* are exactly the fibers of the boundary length map ℓ_∂ .

4.3.4 Spines of hyperbolic surfaces with boundary. The spine construction can be carried on, even in a more natural way, on hyperbolic surfaces with geodesic boundary.

In fact, given such a Σ with boundary components x_1, \dots, x_n , we can define the distance from $\partial\Sigma$ and so the valence of a point in Σ and consequently the spine $\text{Sp}(\Sigma)$, with no need of further information.

Similarly, if Σ has also nodes (that is, some holonomy degenerates to a parabolic element), then $\text{Sp}(\Sigma)$ is embedded inside the *visible components* of Σ , i.e. those components of Σ that contain a boundary circle of positive length.

The weight of an arc $\alpha_i \in \underline{\alpha}_{\text{sp}}$ dual to the edge e_i of $\text{Sp}(\Sigma)$ is still defined as the hyperbolic length of one of the two projections of e_i to $\partial\Sigma$. Thus, the above construction gives a point $w_{\text{sp}} \in |\mathfrak{A}(S, X)| \times (0, \infty)$.

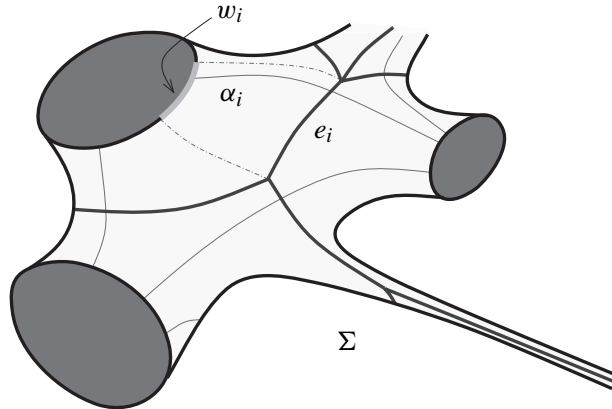


Figure 9. Weights come from lengths of geodesic boundary arcs.

It is easy to check (see [56] or [55]) that w_{sp} converges to the \bar{w}_{sp} defined before when the hyperbolic surface with boundary converges to a decorated surface with cusps in $\hat{\mathcal{T}}(S, X)$. Thus, the $\Gamma(S, X)$ -equivariant map

$$\Phi: \hat{\mathcal{T}}(S, X) \longrightarrow |\mathfrak{A}(S, X)| \times [0, \infty)$$

reduces to Φ_0 for decorated surfaces with cusps.

Theorem 4.9 (Luo [46]). *The restriction of Φ to smooth surfaces with no boundary cusps gives a homeomorphism onto its image.*

The continuity of the whole Φ is proven in [55], using Luo's result.

The key point of Luo's proof is the following. Pick a generic hyperbolic surface with geodesic boundary Σ and suppose that the spinal arc system is the ideal triangulation $\underline{\alpha}_{\text{sp}} = \{\alpha_1, \dots, \alpha_M\} \in \mathfrak{A}^\circ(\Sigma, X)$ with weight w_{sp} . We can define the length ℓ_{α_i} as the hyperbolic length of the shortest geodesic $\tilde{\alpha}_i$ in the free homotopy class of α_i .

The curves $\{\tilde{\alpha}_i\}$ cut Σ into hyperbolic hexagons, which are completely determined by $\{\ell_{\underline{j}_1}, \dots, \ell_{\underline{j}_{2M}}\}$, where the \underline{j} 's are the sides of the hexagons lying on $\partial\Sigma$. Unfortunately, going from the $\ell_{\underline{j}}$'s to w_{sp} is much easier than the converse. In fact, $w_{\alpha_1}, \dots, w_{\alpha_M}$ can be written as explicit linear combinations of the $\ell_{\underline{j}}$'s: in matrix notation, $B = (\ell_{\underline{j}})$ is a solution of the system $W = RB$, where R is a fixed $(M \times 2M)$ -matrix (that encodes the combinatorics is $\underline{\alpha}_{\text{sp}}$) and $W = (w_{\alpha_i})$. Clearly, there is a whole affine space E_W of dimension M of solutions of $W = RB$. The problem is that a random point in E_W would determine hyperbolic structures on the hexagons of $\Sigma \setminus \underline{\alpha}_{\text{sp}}$ that do not glue, because we are not requiring the two sides of each α_i to have the same length.

Starting from very natural quantities associated to hyperbolic hexagons with right angles, Luo defines a functional on the space $(b_1, \dots, b_{2M}) \in \mathbb{R}_{\geq 0}^{2M}$. For every W , the space E_W is not empty (which proves the surjectivity of Φ) and the restriction of Luo's functional to E_W is strictly concave and achieves its (unique) maximum exactly when $B = (\ell_{\underline{j}})$ (which proves the injectivity of Φ).

The geometric meaning of this functional is still not entirely clear, but it seems related to some volume of a three-dimensional hyperbolic manifold associated to Σ . Quite recently, Luo [47] (see also [26]) has introduced a modified functional F_c , which depends on a parameter $c \in \mathbb{R}$, and he has produced other realizations of the Teichmüller space as a polytope, and so different systems of "simplicial" coordinates.

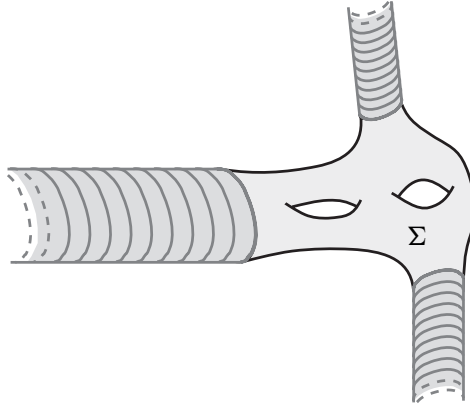
4.3.5 Surfaces with large boundary components. To close the circle, we must relate the limit of Φ for surfaces whose boundary lengths diverge to $\Psi_{\mathcal{J}_S}$. This is the topic of [55]. Here, we only sketch the main ideas. To simplify the exposition, we will only deal with smooth surfaces.

Consider an X -marked hyperbolic surface with geodesic boundary Σ . Define $\text{gr}_\infty(\Sigma)$ to be the surface obtained by gluing semi-infinite flat cylinders at $\partial\Sigma$ of lengths $(p_1, \dots, p_n) = \ell_\partial(\Sigma)$.

Thus, $\text{gr}_\infty(\Sigma)$ has a hyperbolic core and flat ends and the underlying conformal structure is that of an X -punctured Riemann surface. This *infinite grafting* procedure defines a map

$$(\text{gr}_\infty, \ell_\partial): \mathcal{T}^\partial(S, X) \longrightarrow \mathcal{T}(S, X) \times \mathbb{R}_{\geq 0}^N.$$

For more details about (finite) grafting, see [19].

Figure 10. A grafted surface $\text{gr}_\infty(\Sigma)$.

Proposition 4.10 ([55]). *The map $(\text{gr}_\infty, \ell_\partial)$ is a $\Gamma(S, X)$ -equivariant homeomorphism.*

The proof is a variation of Scannell–Wolf’s [67] that finite grafting is a self-homeomorphism of the Teichmüller space.

Thus, the composition of $(\text{gr}_\infty, \ell_\partial)^{-1}$ and Φ gives (after blowing up the locus $\{\ell_\partial = 0\}$) the homeomorphism

$$\Psi: \mathcal{T}(S, X) \times \Delta_X \times [0, \infty) \longrightarrow |\mathfrak{A}^\circ(S, X)| \times [0, \infty).$$

Proposition 4.11 ([55]). *The map Ψ extends to a $\Gamma(S, X)$ -equivariant homeomorphism*

$$\Psi: \mathcal{T}(S, X) \times \Delta_X \times [0, \infty] \longrightarrow |\mathfrak{A}^\circ(S, X)| \times [0, \infty]$$

and Ψ_∞ coincides with Harer–Mumford–Thurston’s $\Psi_{\mathcal{S}}$.

The main point is to show that a surface Σ with large boundaries and with spine $\text{Sp}(\Sigma)$ is very close in $\mathcal{T}(S, X)$ to the flat surface whose Jenkins–Strebel differential has critical graph isomorphic to $\text{Sp}(\Sigma)$ (as metrized ribbon graphs).

To understand why this is reasonable, consider a sequence of hyperbolic surfaces Σ_m whose spine has fixed isomorphism type \mathbb{G} and fixed projective metric and such that $\ell_\partial(\Sigma_m) = c_m(p_1, \dots, p_n)$, where c_m diverges as $m \rightarrow \infty$. Consider the grafted surfaces $\text{gr}_\infty(\Sigma_m)$ and rescale them so that $\sum_i p_i = 1$. The flat metric on the cylinders is naturally induced by a holomorphic quadratic differential, which has negative quadratic residue at X . Extend this differential to zero on the hyperbolic core.

Because of the rescaling, the distance between the flat cylinders and the spine goes to zero and the differential converges in L_{red}^1 to a Jenkins–Strebel differential.

Dumas [18] has shown that an analogous phenomenon occurs for closed surfaces grafted along a measured lamination $t\lambda$ as $t \rightarrow +\infty$.

4.3.6 Weil–Petersson form and Penner’s formula. Using Wolpert’s result and hyperbolic geometry, Penner [63] proved that the pull-back of the Weil–Petersson form on the space of decorated hyperbolic surfaces with cusps, which can be identified with $\mathcal{T}(S, X) \times \Delta_X$, can be neatly written in the following way. Fix a triangulation $\underline{\alpha} = \{\alpha_1, \dots, \alpha_M\} \in \mathfrak{A}^\circ(S, X)$. For every $([f: S \rightarrow \Sigma], p) \in \mathcal{T}(S, X) \times \Delta_X$, let $\tilde{\alpha}_i$ be the geodesic representative in the class of $f_*(\alpha_i)$ and write $a_i := \ell(\tilde{\alpha}_i \cap \Sigma_{\text{tr}})$, where Σ_{tr} is the truncated hyperbolic surface. Then

$$\pi^* \omega_{WP} = \sum_{t \in T} (da_{t_1} \wedge da_{t_2} + da_{t_2} \wedge da_{t_3} + da_{t_3} \wedge da_{t_1})$$

where $\pi: \mathcal{T}(S, X) \times \Delta_X \rightarrow \mathcal{T}(S, X)$ is the projection, T is the set of ideal triangles in which the $\tilde{\alpha}_i$ ’s decompose Σ , and the sides of t are $(\alpha_{t_1}, \alpha_{t_2}, \alpha_{t_3})$ in the cyclic order induced by the orientation of t (see Figure 11).

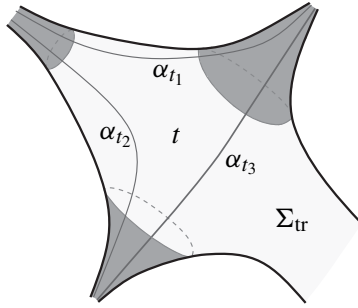


Figure 11. An ideal triangle in T .

To work on $\mathcal{M}_{g,X} \times \Delta_X$ (for instance, to compute Weil–Petersson volumes), one can restrict to the interior of the cells $\Phi_0^{-1}(|\underline{\alpha}|)$ whose associated system of arcs $\underline{\alpha}$ is a triangulation and write the pull-back of ω_{WP} with respect to $\underline{\alpha}$.

4.3.7 Weil–Petersson form for surfaces with boundary. Still using methods of Wolpert [77], one can generalize Penner’s formula to hyperbolic surfaces with boundary. The result is better expressed using the Weil–Petersson bivector field than the 2-form.

Proposition 4.12 ([56]). *Let Σ be a hyperbolic surface with boundary components C_1, \dots, C_n and let $\underline{\alpha} = \{\alpha_1, \dots, \alpha_M\}$ be a triangulation. Then the Weil–Petersson bivector field can be written as*

$$\omega^* = \frac{1}{4} \sum_{b=1}^n \sum_{\substack{y_i \in \alpha_i \cap C_b \\ y_j \in \alpha_j \cap C_b}} \frac{\sinh(p_b/2 - d_b(y_i, y_j))}{\sinh(p_b/2)} \frac{\partial}{\partial a_i} \wedge \frac{\partial}{\partial a_j}$$

where $a_i = \ell(\alpha_i)$ and $d_b(y_i, y_j)$ is the length of the geodesic arc running from y_i to y_j along C_b in the positive direction (according to the orientation induced by Σ on C_b).

The idea is to use Wolpert’s formula $\omega^* = -\sum_i \partial_{\ell_i} \wedge \partial_{\tau_i}$ on the double $d\Sigma$ of Σ with the pair of pants decomposition induced by doubling the arcs $\{\alpha_i\}$. Then one must compute the (first-order) effect on the a_i ’s of twisting $d\Sigma$ along α_j .

Though not immediate, the above formula can be shown to reduce to Penner’s, when the boundary lengths go to zero, as we approximate $\sinh(x) \approx x$ for small x . Notice that Penner’s formula shows that ω linearizes (with constant coefficients!) in the coordinates given by the a_i ’s.

More interesting is to analyze what happens for $(\Sigma, t\underline{p})$ with $\underline{p} \in \Delta_X$, as $t \rightarrow +\infty$. Assume the situation is generic and so $\Psi_{JS}(\Sigma)$ is supported on a triangulation, whose dual graph is \mathbb{G} .

Once again, the formula dramatically simplifies as we approximate $2 \sinh(x) \approx \exp(x)$ for $x \gg 0$. Under the rescalings $\tilde{\omega}^* = c^2 \omega^*$ and $\tilde{w}_i = w_i/c$ with $c = \sum_b p_b/2$, we obtain that

$$\lim_{t \rightarrow \infty} \tilde{\omega}^* = \omega_\infty^* := \frac{1}{2} \sum_{v \in E_0(\mathbb{G})} \left(\frac{\partial}{\partial \tilde{w}_{v_1}} \wedge \frac{\partial}{\partial \tilde{w}_{v_2}} + \frac{\partial}{\partial \tilde{w}_{v_2}} \wedge \frac{\partial}{\partial \tilde{w}_{v_3}} + \frac{\partial}{\partial \tilde{w}_{v_3}} \wedge \frac{\partial}{\partial \tilde{w}_{v_1}} \right)$$

where $v = \{v_1, v_2, v_3\}$ and $\sigma_0(v_j) = v_{j+1}$ (and $j \in \mathbb{Z}/3\mathbb{Z}$).

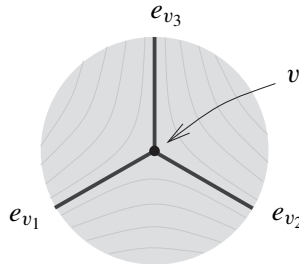


Figure 12. A trivalent vertex v of \mathbb{G} .

Thus, the Weil–Peterson symplectic structure is again linearized (and with constant coefficients!), but in the system of coordinates given by the w_j ’s, which are in some sense dual to the a_i ’s.

It would be nice to exhibit a clear geometric argument for the perfect symmetry of these two formulae.

5 Combinatorial classes

5.1 Witten cycles

Fix as usual a compact oriented surface S of genus g and a subset $X = \{x_1, \dots, x_n\} \subset S$ such that $2g - 2 + n > 0$.

We introduce some remarkable $\Gamma(S, X)$ -equivariant subcomplexes of $\mathfrak{A}(S, X)$, which define interesting cycles in the homology of $\overline{\mathcal{M}}_{g,X}^K$ as well as in the Borel–Moore homology of $\mathcal{M}_{g,X}$ and so, by Poincaré duality, in the cohomology of $\mathcal{M}_{g,X}$ (that is, of $\Gamma(S, X)$).

These subcomplexes are informally defined as the locus of points of $|\mathfrak{A}^\circ(S, X)|$, whose associated ribbon graphs have prescribed odd valences of their vertices. It can be easily shown that, if we assign even valence to some vertex, the subcomplex we obtain is not a cycle (even with $\mathbb{Z}/2\mathbb{Z}$ coefficients!).

We follow Kontsevich ([43]) for the orientation of the combinatorial cycles, but an alternative way is due to Penner [64] and Conant and Vogtmann [14].

Later, we will mention a slight generalization of the combinatorial classes by allowing some vertices to be marked.

Notice that we are going to use the cellularization of the moduli space of curves given by Ψ_{JS} , and so we will identify $\overline{\mathcal{M}}_{g,X}^\Delta$ with the orbispace $|\mathfrak{A}(S, X)|/\Gamma(S, X)$. As the arguments will be essentially combinatorial/topological, any of the decompositions described before would work.

5.1.1 Witten subcomplexes. Let $m_* = (m_0, m_1, \dots)$ be a sequence of nonnegative integers such that

$$\sum_{i \geq 0} (2i + 1)m_i = 4g - 4 + 2n$$

and define $(m_*)! := \prod_{i \geq 0} m_i!$ and $r := \sum_{i \geq 0} i m_i$.

Definition 5.1. The *combinatorial subcomplex* $\mathfrak{A}_{m_*}(S, X) \subset \mathfrak{A}(S, X)$ is the smallest simplicial subcomplex that contains all proper simplices $\underline{\alpha} \in \mathfrak{A}^\circ(S, X)$ such that $S \setminus \underline{\alpha}$ is the disjoint union of exactly m_i polygons with $2i + 3$ sides.

It is convenient to set $|\mathfrak{A}_{m_*}(S, X)|_{\mathbb{R}} := |\mathfrak{A}_{m_*}(S, X)| \times \mathbb{R}_+$. Clearly, this subcomplex is $\Gamma(S, X)$ -equivariant. Hence, if we write $\overline{\mathcal{M}}_{g,X}^{\text{comb}} := \overline{\mathcal{M}}_{g,X}^\Delta \times \mathbb{R}_+ \cong |\mathfrak{A}(S, X)|_{\mathbb{R}}/\Gamma(S, X)$, then we can define $\overline{\mathcal{M}}_{m_*,X}^{\text{comb}}$ to be the subcomplex of $\overline{\mathcal{M}}_{g,X}^{\text{comb}}$ induced by $\mathfrak{A}_{m_*}(S, X)$.

Remark 5.2. We can introduce also univalent vertices by allowing $m_{-1} > 0$. It is still possible to define the complexes $\mathfrak{A}_{m_*}(S, X)$ and $\mathfrak{A}_{m_*}^\circ(S, X)$, just allowing (finitely many) contractible loops (i.e. unmarked tails in the corresponding ribbon graph picture). However, $\mathfrak{A}_{m_*}(S, X)$ would no longer be a subcomplex of $\mathfrak{A}(S, X)$. Thus, we should construct an associated family of Riemann surfaces over $\overline{\mathcal{M}}_{m_*,X}^{\text{comb}}$ (which can be easily done) and consider the classifying map $\overline{\mathcal{M}}_{m_*,X}^{\text{comb}} \rightarrow \overline{\mathcal{M}}_{g,X}^{\text{comb}}$, whose existence is granted by the universal property of $\overline{\mathcal{M}}_{g,X}$, but which would no longer be cellular.

For every $\underline{p} \in \Delta_X \times \mathbb{R}_+$ write $\overline{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p}) := \bar{\ell}_\partial^{-1}(\underline{p}) \subset \overline{\mathcal{M}}_{g,X}^{\text{comb}}$ and define $\overline{\mathcal{M}}_{m^*,X}^{\text{comb}}(\underline{p}) := \overline{\mathcal{M}}_{m^*,X}^{\text{comb}} \cap \overline{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p})$.

Notice that the dimensions of the slices are the expected ones because in every cell they are described by n independent linear equations.

5.1.2 Combinatorial ψ classes. Define L_i^* as the space of couples (\mathbb{G}, y) , where \mathbb{G} is an X -marked metrized ribbon graph in $\overline{\mathcal{M}}_{g,X}^{\text{comb}}(\{p_i > 0\})$ and y is a ray that joins x_i to a point of $|G| \subset |\mathbb{G}|$ that bounds the x_i -th hole.

Clearly $L_i^* \rightarrow \overline{\mathcal{M}}_{g,X}^{\text{comb}}(\{p_i > 0\})$ is a topological bundle with fiber homeomorphic to S^1 . It is easy to see that, for a fixed $\underline{p} \in \Delta_X \times \mathbb{R}_+$ such that $p_i > 0$, the pull-back of L_i^* via

$$\xi_{\underline{p}}: \overline{\mathcal{M}}_{g,X} \longrightarrow \overline{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p})$$

is isomorphic (as a topological bundle) to the sphere bundle associated to \mathcal{L}_i^* .

The proof of the following lemma is very easy.

Lemma 5.3 ([43]). *Fix x_i in X and $\underline{p} \in \Delta_X \times \mathbb{R}_+$ such that $p_i > 0$. Then on every simplex $|\underline{\alpha}|(\underline{p}) \subset \overline{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p})$ define*

$$\bar{\eta}_i|_{|\underline{\alpha}|(\underline{p})} := \sum_{1 \leq s < t \leq k-1} d\tilde{e}_s \wedge d\tilde{e}_t$$

where $\tilde{e}_j = \frac{\ell(e_j)}{p_i}$ and x_i marks a hole with cyclically ordered sides (e_1, \dots, e_k) . These 2-forms glue to give a piecewise-linear 2-form $\bar{\eta}_i$ on $\overline{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p})$, that represents $-c_1(L_i^*)$. Hence, the pull-back class $\xi_{\underline{p}}^*[\bar{\eta}_i]$ is exactly $\psi_i = c_1(\mathcal{L}_i)$ in $H^2(\overline{\mathcal{M}}_{g,X})$.

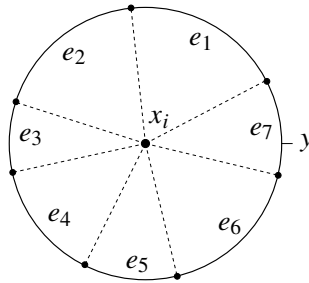


Figure 13. A fiber of the bundle L_i over a hole with 7 sides.

5.1.3 Orientation of Witten subcomplexes. The following lemma says that the η forms can be assembled in a piecewise-linear “symplectic form”, that can be used to orient maximal cells of Witten subcomplexes.

Lemma 5.4 ([43]). *For every $\underline{p} \in \Delta_X \times \mathbb{R}_+$ the restriction of*

$$\bar{\Omega} := \sum_{i=1}^n p_i^2 \bar{\eta}_i$$

to the maximal simplices of $\bar{\mathcal{M}}_{m^,X}^{\text{comb}}(\underline{p})$ is a non-degenerate symplectic form. Hence, $\bar{\Omega}^r$ defines an orientation on $\bar{\mathcal{M}}_{m^*,X}^{\text{comb}}(\underline{p})$. Also, $\bar{\Omega}^r \wedge \bar{\ell}_0^* \text{Vol}_{\mathbb{R}^X}$ is a volume form on $\bar{\mathcal{M}}_{m^*,X}^{\text{comb}}$.*

Proof. Let $|\underline{\alpha}|(\underline{p})$ be a cell of $\bar{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p})$, whose associated ribbon graph $\mathbb{G}_{\underline{\alpha}}$ has only vertices of odd valence.

On $|\underline{\alpha}|(\underline{p})$, the differentials de_i span the cotangent space. As the p_i 's are fixed, we have the relation $dp_i = 0$ for all $i = 1, \dots, n$. Hence

$$T^* \bar{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p})|_{|\underline{\alpha}|(\underline{p})} \cong |\underline{\alpha}|(\underline{p}) \times \bigoplus_{e \in E_1(\underline{\alpha})} \mathbb{R} \cdot de / \left(\sum_{[\vec{e}]_0 = x_i} de \mid i = 1, \dots, n \right).$$

On the other hand the tangent bundle is

$$T \bar{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p})|_{|\underline{\alpha}|(\underline{p})} \cong |\underline{\alpha}|(\underline{p}) \times \left\{ \sum_{e \in E_1(\underline{\alpha})} b_e \frac{\partial}{\partial e} \mid \sum_{[\vec{e}]_0 \in x_i} b_e = 0 \text{ for all } i = 1, \dots, n \right\}.$$

In order to prove that $\bar{\Omega}|_{\underline{\alpha}} : T|\underline{\alpha}|(\underline{p}) \rightarrow T^*|\underline{\alpha}|(\underline{p})$ is non-degenerate, we construct its right-inverse. Define $B : T^*|\underline{\alpha}|(\underline{p}) \rightarrow T|\underline{\alpha}|(\underline{p})$ as

$$B(de) = \sum_{i=1}^{2s} (-1)^i \frac{\partial}{\partial [\sigma_0^i(\vec{e})]_1} + \sum_{j=1}^{2t} (-1)^j \frac{\partial}{\partial [\sigma_0^j(\overleftarrow{e})]_1}$$

where \vec{e} is any orientation of e , while $2s + 1$ and $2t + 1$ are the cardinalities of $[\vec{e}]_0$ and $[\overleftarrow{e}]_0$ respectively. We want to prove that $\bar{\Omega}B(de) = 4de$ for every $e \in E_1(\underline{\alpha})$.

To shorten the notation, set $f_i := [\sigma_0^i(\vec{e})]_1$ and $h_j := [\sigma_0^j(\overleftarrow{e})]_1$ and call $F_i := [\sigma_1 \sigma_0^i(\vec{e})]_\infty$ for $i = 1, \dots, 2s - 1$ and $H_j := [\sigma_1 \sigma_0^j(\overleftarrow{e})]_\infty$ for $j = 1, \dots, 2t - 1$ the holes bordered respectively by $\{f_i, f_{i+1}\}$ and $\{h_j, h_{j+1}\}$. Finally denote by E_+ and E_- the holes adjacent to e as in Figure 14. Remark that neither the edges f and h nor the holes F and H are necessarily distinct. This however has no importance in the following computation:

$$B(de) = \sum_{i=1}^{2s} (-1)^i \frac{\partial}{\partial f_i} + \sum_{j=1}^{2t} (-1)^j \frac{\partial}{\partial h_j}.$$

It is easy to see (using that the perimeters are constant) that

$$p_{F_i}^2 \bar{\eta}_{F_i} \left(\frac{\partial}{\partial f_i} - \frac{\partial}{\partial f_{i+1}} \right) = df_i + df_{i+1}$$

and analogously for the h 's. Moreover

$$p_{E_+}^2 \bar{\eta}_{E_+} \left(\frac{\partial}{\partial h_{2s}} - \frac{\partial}{\partial f_1} \right) = dh_{2s} + df_1 + 2de$$

and similarly for E_- . Finally, we obtain $\bar{\Omega}B(de) = 4de$. □

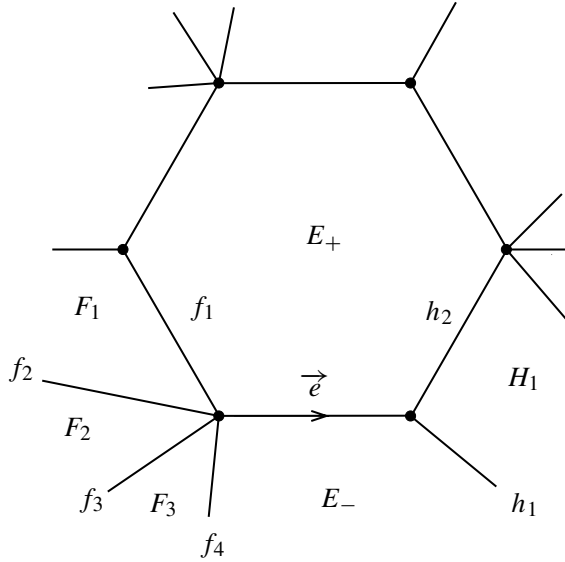


Figure 14. An example with $s = 2$ and $t = 1$.

Remark 5.5. Notice that B is the piecewise-linear extension of the restriction of the Weil–Petersson bivector field $2\tilde{\omega}_\infty^*$ to the open maximal simplices. Thus, $\bar{\Omega}$ is the piecewise-linear extension of $2\tilde{\omega}_\infty$.

Finally, we can show that the (cellular) chain obtained by adding maximal simplices of Witten subcomplexes (with the orientation determined by Ω) is in fact a cycle.

Lemma 5.6 ([43]). *With the given orientation $\bar{\mathcal{M}}_{m^*,X}^{\text{comb}}(\underline{p})$ is a cycle for all $\underline{p} \in \Delta_X \times \mathbb{R}_+$ and $\bar{\mathcal{M}}_{m^*,X}^{\text{comb}}(\mathbb{R}_+^X)$ is a cycle with non-compact support.*

Proof. Given a top-dimensional cell $|\underline{\alpha}|(\underline{p})$ in $\bar{\mathcal{M}}_{m^*,X}^{\text{comb}}(\underline{p})$, each face in the boundary $\partial|\underline{\alpha}|(\underline{p})$ is obtained by shrinking one edge of $\mathbb{G}_{\underline{\alpha}}$. This contraction may merge two vertices as in Figure 15.

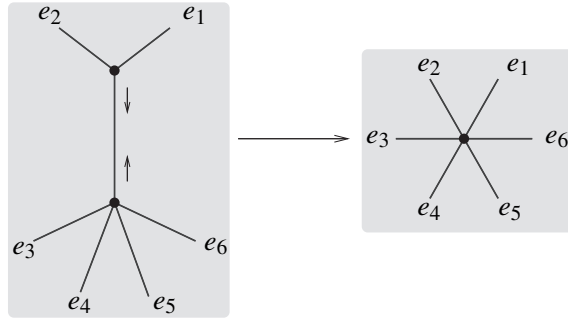


Figure 15. A contraction that merges a 3-valent and a 5-valent vertex.

Otherwise the shrinking produces a node, as in Figure 16.

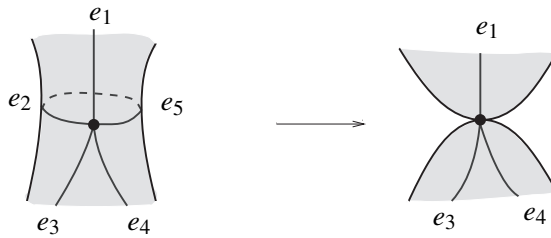


Figure 16. A contraction produces a node.

Let $|\underline{\alpha}'|(p) \subset \partial|\underline{\alpha}|(p)$ be the face of $|\underline{\alpha}|(p)$ obtained by shrinking the edge e . Then $\Lambda^{6g-7+2n-2r}T|\underline{\alpha}'|(p) = \Lambda^{6g-6+2n-2r}T|\underline{\alpha}|(p) \otimes N_{|\underline{\alpha}'|/|\underline{\alpha}|}^*$ and so the dual of the orientation form induced by $|\underline{\alpha}|(p)$ on $|\underline{\alpha}'|(p)$ is $\iota_{de}(B_{\underline{\alpha}}^{6g-6+2n-2r}) = (6g - 6 + 2n - 2r)\iota_{de}(B_{\underline{\alpha}}) \wedge B_{\underline{\alpha}}^{6g-8+2n-2r}$, where $B_{\underline{\alpha}}$ is the bivector field on $|\underline{\alpha}|(p)$ defined in Lemma 5.4.

Consider the graph $\mathbb{G}_{\underline{\alpha}'}$ that occurs in the boundary of a top-dimensional cell of $\overline{\mathcal{M}}_{m_*, X}^{\text{comb}}(p)$. Suppose it is obtained merging two vertices of valences $2t_1 + 3$ and $2t_2 + 3$ in a vertex v of valence $2(t_1 + t_2) + 4$. Then $|\underline{\alpha}'|(p)$ is in the boundary of exactly $2(t_1 + t_2) + 4$ cells of $\overline{\mathcal{M}}_{m_*, X}^{\text{comb}}(p)$. In any case, the number of cells $|\underline{\alpha}'|(p)$ is bordered by are even: we need to prove that half of them induces on $|\underline{\alpha}'|(p)$ an orientation and the other half induces the opposite one. If $\mathbb{G}_{\underline{\alpha}'}$ is obtained from some $\mathbb{G}_{\underline{\alpha}}$ contracting an edge e , then we just have to compute the vector field $\iota_{de}(B_{\underline{\alpha}})$, which turns out to be

$$\iota_{de}(B_{\underline{\alpha}}) = \pm \sum_{i=1}^{2(t_1+t_2)+4} (-1)^i \frac{\partial}{\partial f_i}$$

where $f_1, \dots, f_{2(t_1+t_2)+4}$ are the edges of $\mathbb{G}_{\underline{\alpha}'}$ outgoing from v . It is a straightforward computation to check that one obtains in half the cases a plus and in half the cases a minus.

When $\mathbb{G}_{\underline{\alpha}'}^{\text{en}}$ has a node with $2t_1 + 2$ edges on one side (which we will denote by f_1, \dots, f_{2t_1+2}) and $2t_2 + 3$ edges on the other side, the computation is similar. The cell occurs as boundary of exactly $(2t_1 + 2)(2t_2 + 3)$ top-dimensional cells and, if $\mathbb{G}_{\underline{\alpha}'}$ is obtained by $\mathbb{G}_{\underline{\alpha}}$ contracting the edge e , then

$$\iota_{de}(B_{\underline{\alpha}}) = \pm 2 \sum_{i=1}^{2t_1+2} (-1)^i \frac{\partial}{\partial f_i}.$$

A quick check ensures that the signs cancel. □

Define the *Witten classes* $\overline{W}_{m_*,X}(\underline{p}) := [\overline{\mathcal{M}}_{m_*,X}^{\text{comb}}(\underline{p})]$ and let $W_{m_*,X}(\underline{p})$ be the restriction of such a class to $\mathcal{M}_{g,X}^{\text{comb}}(\underline{p})$, which defines (by Poincaré duality) a cohomology class in $H^{2r}(\mathcal{M}_{g,X})$, independent of \underline{p} .

5.1.4 Generalized Witten cycles. It is possible to define a slight generalization of the previous classes, prescribing that some markings hit vertices with assigned valence.

These *generalized Witten classes* are related to the previous $W_{m_*,X}$ in an intuitively obvious way, because forgetting the markings of some vertices will map them onto one another. We will omit the details and refer to [54].

5.2 Witten cycles and tautological classes

In this subsection, we will sketch the proof of the following result, due to K. Igusa [32] and [33] (see also [34]) and Mondello [54] independently.

Theorem 5.7. *Witten cycles $W_{m_*,X}$ on $\mathcal{M}_{g,X}$ are Poincaré dual to polynomials in the κ classes and vice versa.*

In [54], the following results are also proven:

- Witten generalized cycles on $\mathcal{M}_{g,X}$ are Poincaré dual to polynomials in the ψ and the κ classes;
- ordinary and generalized Witten cycles on $\overline{\mathcal{M}}_{g,X}^{\text{comb}}(\underline{p})$ are push-forward of (the Poincaré dual of) tautological classes from $\overline{\mathcal{M}}_{g,X}$; an explicit recipe to produce such tautological classes is given.

5.2.1 The case with one special vertex. We want to consider a combinatorial cycle on $\mathcal{M}_{g,X}$ supported on ribbon graphs, whose vertices are generically all trivalent except

one, which is $(2r + 3)$ -valent (and $r \geq 1$). To shorten the notation, call this Witten cycle W_{2r+3} .

We also define a generalized Witten cycle on the universal curve $\mathcal{C}_{g,X} \subset \overline{\mathcal{M}}_{g,X \cup \{y\}}$ supported on the locus of ribbon graphs, which have a $(2r + 3)$ -valent vertex marked by y and all the other vertices are trivalent and unmarked. Call W_{2r+3}^y this cycle.

We would like to show that $\text{PD}(W_{2r+3}^y) = c(r)\psi_y^{r+1}$, where $c(r)$ is some constant. As a consequence, pushing the two hand-sides down through the proper map $\pi_y: \mathcal{C}_{g,X} \rightarrow \mathcal{M}_{g,X}$, we would obtain $\text{PD}(W_{2r+3}) = c(r)\kappa_r$.

Lemma 5.3 gives us the nice piecewise-linear 2-form η_y , that is pulled back to ψ_y through ξ . The only problem is that η_y is defined only for $p_y > 0$, whereas W_{2r+3}^y is exactly contained in the locus $\{p_y = 0\}$.

To compare the two, one can look at the blow-up $\text{Bl}_{p_y=0} \overline{\mathcal{M}}_{g,X \cup \{y\}}^{\text{-comb}}$ of $\overline{\mathcal{M}}_{g,X \cup \{y\}}^{\text{-comb}}$ along the locus $\{p_y = 0\}$. Points in the exceptional locus E can be identified with metrized (nonsingular) ribbon graphs \mathbb{G} , in which y marks a vertex, plus *angles* ϑ between consecutive oriented edges outgoing from y . One must think of these angles as of *infinitesimal edges*.

It is clear now that η_y extends to E by

$$\eta_y|_{\underline{\alpha}(p)} := \sum_{1 \leq s < t \leq k-1} d\tilde{e}_s \wedge d\tilde{e}_t$$

where $\tilde{e}_j = \frac{\vartheta_j}{2\pi}$, y marks a vertex with cyclically ordered outgoing edges $(\vec{e}_1, \dots, \vec{e}_k)$ and ϑ_j is the angle between \vec{e}_j and \vec{e}_{j+1} (with $j \in \mathbb{Z}/k\mathbb{Z}$).

Thus, pushing forward η_y^{r+1} through $E \rightarrow \overline{\mathcal{M}}_{g,X}^{\text{-comb}} (p_y = 0)$, we obtain $c(r)\overline{W}_{2r+3}^y$ plus other terms contained in the boundary, and the coefficient $c(r)$ is exactly the integral of η_y^{r+1} on a fiber (that is, a simplex), which turns out to be $c(r) = \frac{(r+1)!}{(2r+2)!}$. Thus, W_{2r+3}^y is Poincaré dual to $2^{r+1}(2r+1)!\psi_y^{r+1}$.

5.2.2 The case with many special vertices. To mimic what is done for one non-trivalent vertex, consider combinatorial classes with two non-trivalent vertices. Thus, we examine the class $\psi_y^{r+1}\psi_z^{s+1}$ (with $r, s \geq 1$) on $\mathcal{C}_{g,X}^2 := \mathcal{C}_{g,X} \times_{\mathcal{M}_{g,X}} \mathcal{C}_{g,X}$.

Let us look at the blow-up $\text{Bl}_{p_y=0, p_z=0} \overline{\mathcal{M}}_{g,X \cup \{y\}}^{\text{-comb}}$ of $\overline{\mathcal{M}}_{g,X \cup \{y,z\}}^{\text{-comb}}$ along the locus $\{p_y = 0\} \cup \{p_z = 0\}$ and let $E = E_y \cap E_z$, where E_y and E_z are the exceptional loci.

As before, we can identify $E \cap \{y \neq z\}$ with the set of metrized ribbon graphs \mathbb{G} , with angles at the vertices y and z . Thus, pushing $\eta_y^{r+1}\eta_z^{s+1}$ forward through the blow-up map (which forgets the angles at y and z), we obtain a multiple of the generalized combinatorial cycles given by y marking a $(2r + 3)$ -valent vertex and z marking a $(2s + 3)$ -valent (distinct) vertex. The coefficient $c(r, s)$ will just be $\frac{(r+1)!(s+1)!}{(2r+2)!(2s+2)!}$.

Points in $E \cap \{y = z\}$ can be thought of as metrized ribbon graphs \mathbb{G} with two infinitesimal holes (respectively marked by y and z) adjacent to each other. If we perform the push-forward of $\eta_y^{r+1}\eta_z^{s+1}$ forgetting first the angles at z and then the angles

at y , then we obtain some contribution only from the loci in which the infinitesimal z -hole has $(2s + 3)$ edges and the infinitesimal y -hole has $(2r + 4)$ edges (including the common one). Thus, we obtain the same contribution for each of the $b(r, s)$ configurations of two adjacent holes of valences $(2s + 3)$ and $(2r + 4)$.

Thus, we obtain a cycle supported on the locus of metrized ribbon graphs \mathbb{G} in which $y = z$ marks a $(2r + 2s + 3)$ -valent vertex, with coefficient $b(r, s)c(r, s)$.

Hence, $\psi_y^{r+1}\psi_z^{s+1}$ is Poincaré dual to a linear combination of generalized combinatorial cycles. As before, using the forgetful map, the same holds for the Witten cycles obtained by deleting the y and the z markings.

One can easily see that the transformation laws from ψ classes to combinatorial classes are invertible (because they are “upper triangular” in a suitable sense).

Clearly, in order to deal with many ψ classes (that is, with many non-trivalent marked vertices), one must compute more and more complicated combinatorial factors like $b(r, s)$.

We refer to [34] and [54] for two (complementary) methods to calculate these factors.

5.3 Stability of Witten cycles

5.3.1 Harer’s stability theorem. The (co)homologies of the mapping class groups have the remarkable property that they stabilize when the genus of the surface increases. This was proven by Harer [27], and the stability bound was then improved by Ivanov [36] (and successively again by Harer for homology with rational coefficients, in an unpublished paper). We now want to recall some of Harer’s results.

Let $S_{g,n,b}$ be a compact oriented surface of genus g with n marked points and b boundary components C_1, \dots, C_b . We denote the group of isotopy classes of diffeomorphisms of S that fix the marked points and ∂S pointwise by $\Gamma(S_{g,n,b})$.

Write $P = S_{0,0,3}$ for a fixed pair of pants and denote by B_1, B_2, B_3 its boundary components.

Consider the following two operations:

- (y) gluing $S_{g,n,b}$ and P by identifying C_b with B_1 , thus producing an oriented surface of genus g with n marked points and $b + 1$ boundary components,
- (v) identify C_{b-1} with C_b of $S_{g,n,b}$, thus producing an oriented surface of genus $g + 1$ with n marked points and $b - 2$ boundary components.

Clearly, they induce homomorphism at the level of mapping class groups

$$\mathcal{Y}: \Gamma(S_{g,n,b}) \longrightarrow \Gamma(S_{g,n,b+1})$$

when $b \geq 1$ (by extending the diffeomorphism as the identity on P) and

$$\mathcal{V}: \Gamma(S_{g,n,b}) \longrightarrow \Gamma(S_{g+1,n,b-2})$$

when $b \geq 2$.

Theorem 5.8 (Harer [27]). *The induced maps in homology*

$$\begin{aligned} \mathcal{Y}_* : H_k(\Gamma(S_{g,n,b})) &\longrightarrow H_k(\Gamma(S_{g,n,b+1})), \\ \mathcal{V}_* : H_k(\Gamma(S_{g,n,b})) &\longrightarrow H_k(\Gamma(S_{g+1,n,b-2})) \end{aligned}$$

are isomorphisms for $g \geq 3k$.

The exact bound is not important for our purposes. We only want to stress that the theorem implies that $H_k(\Gamma(S_{g,n,b}))$ stabilizes for large g . In particular, for a fixed $n \geq 0$, the rational homology of $\mathcal{M}_{g,n}$ stabilizes for large g .

Remark 5.9. We have $B\Gamma(S_{g,n,b}) \simeq \mathcal{M}_{g,X,T}$, where $\mathcal{M}_{g,X,T}$ is the moduli space of Riemann surfaces of genus g with $X \cup T$ marked points ($X = \{x_1, \dots, x_n\}$ and $T = \{t_1, \dots, t_b\}$) and a nonzero tangent vector at each point of T . If $b \geq 1$, then $\mathcal{M}_{g,X,T}$ is a smooth variety: in fact, an automorphism of a Riemann surface that fixes a point and a tangent direction at that point is the identity (this follows from uniformization and Schwarz lemma).

5.3.2 Mumford's conjecture. Write $\Gamma_{\infty,n} = \lim_{g \rightarrow \infty} \Gamma(S_{g,n,1})$, where the map $\Gamma(S_{g,n,1}) \rightarrow \Gamma(S_{g+1,n,1})$ corresponds to gluing a torus with two holes at the boundary component of $S_{g,n,1}$.

Then, $H^k(\Gamma_{\infty,n})$ coincides with $H^k(\Gamma_{g,n})$ for $g \gg k$.

Mumford conjectured that $H^*(\Gamma_{\infty}; \mathbb{Q})$ is the polynomial algebra on the κ classes. Miller [52] showed that $H^*(\Gamma_{\infty}; \mathbb{Q})$ is a Hopf algebra that contains $\mathbb{Q}[\kappa_1, \kappa_2, \dots]$.

Recently, after works of Tillmann (for instance, [72]) and Madsen–Tillmann [48], Madsen and Weiss [49] proved a much stronger statement of homotopy theory, which in particular implies Mumford's conjecture.

Thanks to a result of Bödigheimer–Tillmann [12], it follows that $H^*(\Gamma_{\infty,n}; \mathbb{Q})$ is a polynomial algebra on ψ_1, \dots, ψ_n and the κ classes.

Thus, generalized Witten classes, being polynomials in ψ and κ , are also stable. In what follows, we would like to prove this stability in a direct way.

5.3.3 Ribbon graphs with tails. One way to cellularize the moduli space of curves with marked points and tangent vectors at the marked points is to use ribbon graphs with tails (see, for instance, [23]).

Consider Σ a compact Riemann surface of genus g with marked points $X \cup T = \{x_1, \dots, x_n\} \cup \{t_1, \dots, t_b\}$ and nonzero tangent vectors v_1, \dots, v_b at t_1, \dots, t_b .

Given $p_1, \dots, p_n \geq 0$ and $q_1, \dots, q_b > 0$, we can construct the ribbon graph \mathbb{G} associated to $(\Sigma, \underline{p}, \underline{q})$, say using the Jenkins–Strebel differential φ .

For every $j = 1, \dots, b$, move from the center t_j along a vertical trajectory γ_j of φ determined by the tangent vector v_j , until we hit the critical graph. Parametrize the opposite path γ_j^* by arc-length, so that $\gamma_j^* : [0, \infty] \rightarrow \Sigma$, $\gamma_j^*(0)$ lies on the critical graph and $\gamma_j^*(\infty) = t_j$. Then, construct a new ribbon graph out of \mathbb{G} by “adding” a

new vertex (which we will denote by \tilde{v}_j) and a new edge e_{v_j} of length $|v_j|$ (a *tail*), whose realization is $\gamma_j^*([0, |v_j|])$ (see Figure 17).

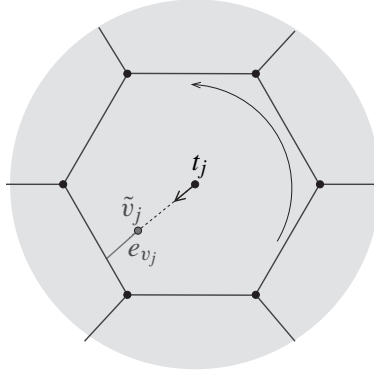


Figure 17. Correspondence between a tail and a nonzero tangent vector.

Thus, we have realized an embedding of $\mathcal{M}_{g,X,T} \times \mathbb{R}_{\geq 0}^X \times \mathbb{R}_+^T$ inside $\mathcal{M}_{g,X \cup T \cup V}^{\text{comb}}$, where $V = \{\tilde{v}_1, \dots, \tilde{v}_b\}$. If we denote its image by $\mathcal{M}_{g,X,T}^{\text{comb}}$, we have obtained the following result.

Lemma 5.10. $\mathcal{M}_{g,X,T}^{\text{comb}} \simeq B\Gamma(S_{g,n,b})$.

Notice that the embedding $\mathcal{M}_{g,X,T}^{\text{comb}} \hookrightarrow \mathcal{M}_{g,X \cup T \cup V}^{\text{comb}}$ allows us to define (generalized) Witten cycles $W_{m^*,X,T}$ on $\mathcal{M}_{g,X,T}^{\text{comb}}$ simply by restriction.

5.3.4 Gluing ribbon graphs with tails. Let \mathbb{G}' and \mathbb{G}'' be two ribbon graphs with tails \vec{e}' and \vec{e}'' , i.e. $\vec{e}' \in E(\mathbb{G}')$ and $\vec{e}'' \in E(\mathbb{G}'')$ with the property that $\sigma_0'(\vec{e}') = \vec{e}'$ and $\sigma_0''(\vec{e}'') = \vec{e}''$.

We produce a third ribbon graph \mathbb{G} by *gluing* \mathbb{G}' and \mathbb{G}'' in the following way.

We set $E(\mathbb{G}) = (E(\mathbb{G}') \cup E(\mathbb{G}'')) / \sim$, where we declare that $\vec{e}' \sim \overleftarrow{e}''$ and $\overleftarrow{e}' \sim \vec{e}''$. Thus, we have a natural σ_1 induced on $E(\mathbb{G})$. Moreover, we define σ_0 acting on $E(\mathbb{G})$ as

$$\sigma_0([\vec{e}]) = \begin{cases} [\sigma_0'(\vec{e})] & \text{if } \vec{e} \in E(\mathbb{G}') \text{ and } \vec{e} \neq \vec{e}', \\ [\sigma_0''(\vec{e})] & \text{if } \vec{e} \in E(\mathbb{G}'') \text{ and } \vec{e} \neq \vec{e}''. \end{cases}$$

If \mathbb{G}' and \mathbb{G}'' are metrized, then we induce a metric on \mathbb{G} in a canonical way, declaring the length of the new edge of \mathbb{G} to be $\ell(e') + \ell(e'')$.

Suppose that \mathbb{G}' is marked by $\{x_1, \dots, x_n, t'\}$ and e' is a tail contained in the hole t' and that \mathbb{G}'' is marked by $\{y_1, \dots, y_m, t''\}$ and if e'' is a tail contained in the hole

t'' , then \mathbb{G} is marked by $\{x_1, \dots, x_n, y_1, \dots, y_m, t\}$, where t is a new hole obtained merging the holes centered at t' and t'' .

Thus, we have constructed a *combinatorial gluing map*

$$\mathcal{M}_{g',X',T' \cup \{t'\}}^{\text{comb}} \times \mathcal{M}_{g'',X'',T'' \cup \{t''\}}^{\text{comb}} \longrightarrow \mathcal{M}_{g'+g'',X' \cup X'' \cup \{t\},T' \cup T''}^{\text{comb}}$$

5.3.5 The combinatorial stabilization maps. Consider the gluing maps in two special cases which are slightly different from what we have seen before.

Call $S_{g,X,T}$ a compact oriented surface of genus g with boundary components labelled by T and marked points labeled by X .

Fix a trivalent ribbon graph \mathbb{G}_j , with genus 1, one hole and j tails for $j = 1, 2$ (for instance, $j = 2$ in Figure 18).

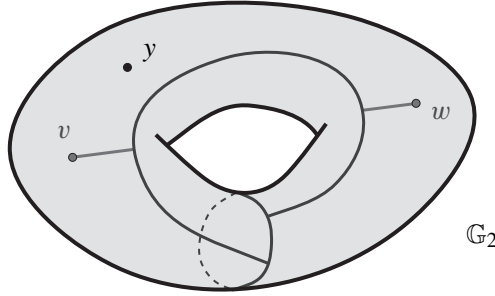


Figure 18. Example of a fixed torus.

Consider the combinatorial gluing maps

$$\begin{aligned} \mathfrak{S}_1^{\text{comb}} : \mathcal{M}_{g,X,\{t\}}^{\text{comb}} &\longrightarrow \mathcal{M}_{g+1,X \cup \{t\}}^{\text{comb}}, \\ \mathfrak{S}_2^{\text{comb}} : \mathcal{M}_{g,X,\{t\}}^{\text{comb}} &\longrightarrow \mathcal{M}_{g+1,X,\{t\}}^{\text{comb}}, \end{aligned}$$

where $\mathfrak{S}_j^{\text{comb}}$ is obtained by simply gluing a graph \mathbb{G} in $\mathcal{M}_{g,X,\{t\}}^{\text{comb}}$ with the fixed graph \mathbb{G}_j , identifying the unique tail of $\mathcal{M}_{g,X,\{t\}}^{\text{comb}}$ with the v -tail of \mathbb{G}_j and renaming the new hole by t .

It is easy to see that $\mathfrak{S}_2^{\text{comb}}$ incarnates a stabilization map (obtained by composing twice \mathcal{Y} and once \mathcal{V}).

On the other hand, consider the map $\mathfrak{S}_1 : B\Gamma(S_{g,X,\{t\}}) \rightarrow B\Gamma(S_{g+1,X \cup \{t\}})$, that glues a torus $S_{1,\{y\},\{t'\}}$ with one puncture and one boundary component to the unique boundary component of $S_{g,X,\{t\}}$, by identifying t and t' , and relabels the y -puncture by t .

The composition of \mathfrak{S}_1 followed by the map π_t that forgets the t -marking

$$B\Gamma(S_{g,X,\{t\}}) \xrightarrow{\mathfrak{S}_1} B\Gamma(S_{g+1,X \cup \{t\}}) \xrightarrow{\pi_t} B\Gamma(S_{g+1,X})$$

induces an isomorphism on H_k for $k \gg g$, because it can also be obtained composing \mathcal{Y} and \mathcal{V} .

Notice that $\pi_t: B\Gamma(S_{g+1, X \cup \{t\}}) \rightarrow B\Gamma(S_{g+1, X})$ can be realized as a *combinatorial forgetful map* $\pi_t^{\text{comb}}: \mathcal{M}_{g+1, X \cup \{t\}}^{\text{comb}}(\mathbb{R}_+^X \times \{0\}) \rightarrow \mathcal{M}_{g+1, X}^{\text{comb}}(\mathbb{R}_+^X)$ in the following way.

Let \mathbb{G} be a metrized ribbon graph in $\mathcal{M}_{g+1, X \cup \{t\}}^{\text{comb}}(\mathbb{R}_+^X \times \{0\})$. If t is marking a vertex of valence 3 or more, then just forget the t -marking. If t is marking a vertex of valence 2, then forget the t marking and merge the two edges outgoing from t in one new edge. Finally, if t is marking a univalent vertex of \mathbb{G} lying on an edge e , then replace \mathbb{G} by \mathbb{G}/e and forget the t -marking.

5.3.6 Behavior of Witten cycles. The induced homomorphism on Borel–Moore homology

$$(\pi_t^{\text{comb}})^*: H_*^{BM}(\mathcal{M}_{g+1, X}^{\text{comb}}(\mathbb{R}_+^X)) \longrightarrow H_*^{BM}(\mathcal{M}_{g+1, X \cup \{t\}}^{\text{comb}}(\mathbb{R}_+^X \times \{0\}))$$

pulls $W_{m_*, X}$ back to the combinatorial class $W_{m_* + \delta_0, X}^t$, corresponding to (the closure of the locus of) ribbon graphs with one univalent vertex marked by t and $m_i + \delta_{0, i}$ vertices of valence $(2i + 3)$ for all $i \geq 0$.

We now use the fact that, for X nonempty, there is a homotopy equivalence

$$E: \mathcal{M}_{g+1, X \cup \{t\}}^{\text{comb}}(\mathbb{R}_+^X \times \mathbb{R}_+) \xrightarrow{\sim} \mathcal{M}_{g+1, X \cup \{t\}}^{\text{comb}}(\mathbb{R}_+^X \times \{0\})$$

and that $E^*(W_{m_* + \delta_0, X}^t) = W_{m_* + 2\delta_0, X \cup \{t\}}$.

This last phenomenon can be understood by simply observing that E^{-1} corresponds to opening the (generically univalent) t -marked vertex to a small t -marked hole, thus producing an extra trivalent vertex.

Finally, $(\mathcal{G}_1^{\text{comb}})^*(W_{m_* + 2\delta_0, X \cup \{t\}}) = W_{m_* - \delta_0, X, \{t\}}$, because \mathbb{G}_1 has exactly 3 trivalent vertices.

As a consequence, we have obtained that

$$(\pi_t^{\text{comb}} \circ E \circ \mathcal{G}_1^{\text{comb}})^*: H_*^{BM}(\mathcal{M}_{g+1, X}^{\text{comb}}(\mathbb{R}_+^X)) \longrightarrow H_*^{BM}(\mathcal{M}_{g, X, \{t\}}^{\text{comb}}(\mathbb{R}_+^X \times \mathbb{R}_+))$$

is an isomorphism for $g \gg *$ and pulls $W_{m_*, X}$ back to $W_{m_* - \delta_0, X, \{t\}}$.

The other gluing map is much simpler: the induced

$$(\mathcal{G}_2^{\text{comb}})^*: H_*^{BM}(\mathcal{M}_{g+1, X, \{t\}}^{\text{comb}}(\mathbb{R}_+^X \times \mathbb{R}_+)) \longrightarrow H_*^{BM}(\mathcal{M}_{g, X, \{t\}}^{\text{comb}}(\mathbb{R}_+^X \times \mathbb{R}_+))$$

carries $W_{m_*, X, \{t\}}$ to $W_{m_* - 4\delta_0, X, \{t\}}$, because \mathbb{G}_2 has 4 trivalent vertices.

We recall that a class in $H^k(\Gamma_{\infty, X})$ (i.e. a stable class) is a sequence of classes $\{g \in H^k(\mathcal{M}_{g, X}) \mid g \geq g_0\}$, which are compatible with the stabilization maps, and that two sequences are equivalent (i.e. they represent the same stable class) if they are equal for large g .

Proposition 5.11. *Let $m_* = (m_0, m_1, \dots)$ be a sequence of nonnegative integers such that $m_N = 0$ for large N and let $|X| = n > 0$. Define*

$$c(g) = 4g - 4 + 2n - \sum_{j \geq 1} (2j + 1)m_j$$

and let $g_0 = \inf\{g \in \mathbb{N} \mid c(g) \geq 0\}$. Then, the collection

$$\{W_{m_*+c(g)\delta_0, X} \in H^{2k}(\mathcal{M}_{g, X}) \mid g \geq g_0\}$$

is a stable class, where $k = \sum_{j>0} j m_j$.

It is clear that an analogous statement can be proven for generalized Witten cycles. Notice that Proposition 5.11 implies Miller's result [52] that ψ and κ classes are stable.

References

- [1] L. V. Ahlfors, The complex analytic structure of the space of closed Riemann surfaces. In 1960 *Analytic functions*, Princeton University Press, Princeton, N.J., 45–66. [171](#)
- [2] L. V. Ahlfors, Some remarks on Teichmüller's space of Riemann surfaces. *Ann. of Math.* (2) 74 (1961), 171–191. [177](#)
- [3] L. V. Ahlfors and L. Bers, Riemann's mapping theorem for variable metrics. *Ann. of Math.* (2) 72 (1960), 385–404. [171](#)
- [4] S. J. Arakelov, Families of algebraic curves with fixed degeneracies. *Izv. Akad. Nauk SSSR Ser. Mat.* 35 (1971), 1269–1293. [153](#), [180](#)
- [5] E. Arbarello and M. Cornalba, Combinatorial and algebro-geometric cohomology classes on the moduli spaces of curves. *J. Algebraic Geom.* 5 (4) (1996), 705–749. [153](#), [180](#)
- [6] E. Arbarello and M. Cornalba, Calculating cohomology groups of moduli spaces of curves via algebraic geometry. *Inst. Hautes Études Sci. Publ. Math.* 88 (1998), 97–127. [180](#)
- [7] E. Arbarello, M. Cornalba, P. Griffiths, and J. Harris, *Geometry of Algebraic Curves II*. Book in preparation. [152](#), [154](#), [158](#), [192](#)
- [8] L. Bers, Simultaneous uniformization. *Bull. Amer. Math. Soc.* 66 (1960), 94–97. [171](#)
- [9] L. Bers, Spaces of degenerating Riemann surfaces. In *Discontinuous groups and Riemann surfaces* (Proc. Conf., Univ. Maryland, College Park, MD, 1973), Ann. of Math. Stud. 79, Princeton University Press, Princeton, NJ, 1974, 43–55. [193](#)
- [10] D. Bessis, C. Itzykson, and J. B. Zuber, Quantum field theory techniques in graphical enumeration. *Adv. in Appl. Math.* 1 (2) (1980), 109–157. [155](#)
- [11] G. Bini, A combinatorial algorithm related to the geometry of the moduli space of pointed curves. *J. Algebraic Combin.* 15 (3) (2002), 211–221. [179](#)
- [12] C.-F. Bödigheimer and U. Tillmann, Stripping and splitting decorated mapping class groups. In *Cohomological methods in homotopy theory* (Bellaterra, 1998), Progr. Math. 196, Birkhäuser, Basel 2001, 47–57. [207](#)
- [13] B. H. Bowditch and D. B. A. Epstein, Natural triangulations associated to a surface. *Topology* 27 (1) (1988), 91–117. [154](#), [158](#), [162](#), [192](#)

- [14] J. Conant and K. Vogtmann, On a theorem of Kontsevich. *Algebr. Geom. Topol.* 3 (2003), 1167–1224. [155](#), [199](#)
- [15] M. Cornalba, On the projectivity of the moduli spaces of curves. *J. Reine Angew. Math.* 443 (1993), 11–20. [152](#)
- [16] P. Deligne and D. Mumford, The irreducibility of the space of curves of given genus. *Inst. Hautes Études Sci. Publ. Math.* 36 (1969), 75–109. [152](#)
- [17] P. Di Francesco, C. Itzykson, and J.-B. Zuber, Polynomial averages in the Kontsevich model. *Comm. Math. Phys.* 151 (1) (1993), 193–219. [155](#)
- [18] D. Dumas, The Schwarzian Derivative and Measured Laminations on Riemann Surfaces. *Duke Math J.* 140 (2) (2007), 203–243. [196](#)
- [19] D. Dumas, Complex projective structures. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 455–508. [195](#)
- [20] C. Faber, A conjectural description of the tautological ring of the moduli space of curves. In *Moduli of curves and Abelian varieties*, Aspects Math. E33, Vieweg, Braunschweig 1999, 109–129. [180](#)
- [21] A. Fathi, F. Laudenbach, and V. Poénaru (eds.), *Travaux de Thurston sur les surfaces*. *Astérisque* 66–67 (1979). [153](#)
- [22] D. Gieseker, *Lectures on moduli of curves*. Tata Inst. Fund. Res. Lectures on Math. and Phys. 69, Tata Institute of Fundamental Research, Bombay 1982. [152](#)
- [23] V. Godin, The unstable integral homology of the mapping class groups of a surface with boundary. *Math. Ann.* 337 (1) (2007), 15–60. [207](#)
- [24] W. M. Goldman, The symplectic nature of fundamental groups of surfaces. *Adv. in Math.* 54 (2) (1984), 200–225. [178](#)
- [25] T. Graber and R. Pandharipande, Constructions of nontautological classes on moduli spaces of curves. *Michigan Math. J.* 51 (1) (2003), 93–109. [153](#)
- [26] R. Guo, On parametrizations of Teichmüller spaces of surfaces with boundary. Preprint, 2006; arXiv:math/0612221. [195](#)
- [27] J. L. Harer, Stability of the homology of the mapping class groups of orientable surfaces. *Ann. of Math.* (2) 121 (2) (1985), 215–249. [206](#), [207](#)
- [28] J. L. Harer, The virtual cohomological dimension of the mapping class group of an orientable surface. *Invent. Math.* 84 (1) (1986), 157–176. [154](#), [159](#), [162](#), [186](#)
- [29] J. L. Harer and D. Zagier, The Euler characteristic of the moduli space of curves. *Invent. Math.* 85 (3) (1986), 457–485. [154](#)
- [30] W. J. Harvey, Geometric structure of surface mapping class groups. In *Homological group theory* (Proc. Sympos., Durham, 1977), London Math. Soc. Lecture Note Ser. 36, Cambridge University Press, Cambridge 1979, 255–269. [153](#)
- [31] J. Hubbard and H. Masur, Quadratic differentials and foliations. *Acta Math.* 142 (3–4) (1979), 221–274. [161](#), [162](#), [185](#)
- [32] K. Igusa, Combinatorial Miller-Morita-Mumford classes and Witten cycles. *Algebr. Geom. Topol.* 4 (2004), 473–520. [155](#), [204](#)
- [33] K. Igusa, Graph cohomology and Kontsevich cycles. *Topology* 43 (6) (2004), 1469–1510. [155](#), [204](#)

- [34] K. Igusa and M. Kleber, Increasing trees and Kontsevich cycles. *Geom. Topol.* 8 (2004), 969–1012. [204](#), [206](#)
- [35] Y. Imayoshi, A construction of holomorphic families of Riemann surfaces over the punctured disk with given monodromy. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich, 2009, 93–130. [172](#)
- [36] N. V. Ivanov, On the homology stability for Teichmüller modular groups: closed surfaces and twisted coefficients. In *Mapping class groups and moduli spaces of Riemann surfaces* (Göttingen, 1991/Seattle, WA, 1991), *Contemp. Math.* 150, Amer. Math. Soc., Providence, RI, 1993, 149–194. [206](#)
- [37] J. A. Jenkins, On the existence of certain general extremal metrics. *Ann. of Math.* (2) 66 (1957), 440–453. [184](#)
- [38] R. Kaufmann, Y. Manin, and D. Zagier, Higher Weil-Petersson volumes of moduli spaces of stable n -pointed curves. *Comm. Math. Phys.* 181 (3) (1996), 763–787. [180](#)
- [39] S. Keel, Basepoint freeness for nef and big line bundles in positive characteristic. *Ann. of Math.* (2) 149 (1) (1999), 253–286. [180](#)
- [40] F. F. Knudsen, The projectivity of the moduli space of stable curves. II. The stacks $\mathcal{M}_{g,n}$. *Math. Scand.* 52 (2) (1983), 161–199. [152](#)
- [41] F. F. Knudsen, The projectivity of the moduli space of stable curves. III. The line bundles on $\mathcal{M}_{g,n}$, and a proof of the projectivity of $\overline{\mathcal{M}}_{g,n}$ in characteristic 0. *Math. Scand.* 52 (2) (1983), 200–212. [152](#)
- [42] J. Kollár, Projectivity of complete moduli. *J. Differential Geom.* 32 (1) (1990), 235–268. [152](#)
- [43] M. Kontsevich, Intersection theory on the moduli space of curves and the matrix Airy function. *Comm. Math. Phys.* 147 (1) (1992), 1–23. [154](#), [155](#), [181](#), [190](#), [199](#), [200](#), [201](#), [202](#)
- [44] M. Kontsevich, Feynman diagrams and low-dimensional topology. In *First European Congress of Mathematics* (Paris, 1992), Vol. II, *Progr. Math.* 120, Birkhäuser, Basel 1994, 97–121. [155](#)
- [45] E. Looijenga, Cellular decompositions of compactified moduli spaces of pointed curves. In *The moduli space of curves* (Texel Island, 1994), Birkhäuser Boston, Boston, Mass., 1995, 369–400. [154](#), [157](#), [158](#), [190](#)
- [46] F. Luo, On Teichmüller Space of a Surface with Boundary. *Duke Math J.* 139 (3) (2007), 463–482. [154](#), [156](#), [195](#)
- [47] F. Luo, Rigidity of Polyhedral Surfaces. Preprint, 2006; arXiv:math/0612714. [154](#), [195](#)
- [48] I. Madsen and U. Tillmann, The stable mapping class group and $Q(\mathbb{C}P_{\mp}^{\infty})$. *Invent. Math.* 145 (3) (2001), 509–544. [207](#)
- [49] I. Madsen and M. S. Weiss, The stable moduli space of Riemann surfaces: Mumford’s conjecture. *Ann. of Math.* (2) 165 (3) (2007), 843–941. [153](#), [207](#)
- [50] B. Maskit, Comparison of hyperbolic and extremal lengths. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 10 (1985), 381–386. [187](#)
- [51] H. Masur, Extension of the Weil-Petersson metric to the boundary of Teichmüller space. *Duke Math. J.* 43 (3) (1976), 623–635. [153](#), [177](#), [193](#)

- [52] E. Y. Miller, The homology of the mapping class group. *J. Differential Geom.* 24 (1) (1986), 1–14. [153](#), [207](#), [211](#)
- [53] M. Mirzakhani, Weil-Petersson volumes and intersection theory on the moduli space of curves. *J. Amer. Math. Soc.* 20 (1) (2007), 1–23. [153](#)
- [54] G. Mondello, Combinatorial classes on $\overline{\mathcal{M}}_{g,n}$ are tautological. *Internat. Math. Res. Notices* 2004 (44) (2004), 2329–2390. [155](#), [204](#), [206](#)
- [55] G. Mondello, Riemann surfaces with boundary and natural triangulations of the Teichmüller space. Submitted paper. [154](#), [156](#), [194](#), [195](#), [196](#)
- [56] G. Mondello, Triangulated Riemann surfaces with boundary and the Weil-Petersson Poisson structure. *J. Differential Geom.* 81 (2009), 391–436. [154](#), [156](#), [194](#), [197](#)
- [57] S. Morita, Characteristic classes of surface bundles. *Invent. Math.* 90 (3) (1987), 551–577. [153](#)
- [58] S. Morita, Introduction to mapping class groups of surfaces and related groups. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 353–386. [180](#)
- [59] D. Mumford, Stability of projective varieties. *Enseignement Math.* (2) 23 (1–2) (1977), 39–110. [152](#)
- [60] D. Mumford, Towards an enumerative geometry of the moduli space of curves. In *Arithmetic and geometry*, Vol. II, Birkhäuser Boston, Boston, Mass., 1983, 271–328. [153](#), [179](#), [180](#)
- [61] R. C. Penner, The decorated Teichmüller space of punctured surfaces. *Comm. Math. Phys.* 113 (2) (1987), 299–339. [154](#), [162](#), [192](#)
- [62] R. C. Penner, Perturbative series and the moduli space of Riemann surfaces. *J. Differential Geom.* 27 (1) (1988), 35–53. [154](#)
- [63] R. C. Penner, Weil-Petersson volumes. *J. Differential Geom.* 35 (3) (1992), 559–608. [155](#), [197](#)
- [64] R. C. Penner, The Poincaré dual of the Weil-Petersson Kähler two-form. *Comm. Anal. Geom.* 1 (1) (1993), 43–69. [155](#), [199](#)
- [65] R. C. Penner, Cell decomposition and compactification of Riemann’s moduli space in decorated Teichmüller theory. In *Woods Hole mathematics* (2004), Ser. Knots Everything 34, World Sci. Publ., Hackensack, N.J., 263–301. [154](#)
- [66] R. C. Penner, The structure and singularities of arc complexes. *J. Topol.* 1 (2008), 527–550. [154](#)
- [67] K. P. Scannell and M. Wolf, The grafting map of Teichmüller space. *J. Amer. Math. Soc.* 15 (4) (2002), 893–927. [196](#)
- [68] Y. Sözen and F. Bonahon, The Weil-Petersson and Thurston symplectic forms. *Duke Math. J.* 108 (3) (2001), 581–597. [178](#)
- [69] K. Strebel, On quadratic differentials with closed trajectories and second order poles. *J. Analyse Math.* 19 (1967), 373–382. [184](#)
- [70] K. Strebel, *Quadratic differentials*. *Ergeb. Math. Grenzgeb.* (3) 5, Springer-Verlag, Berlin 1984. [183](#)
- [71] O. Teichmüller, *Gesammelte Abhandlungen*. Edited by L. V. Ahlfors and F. W. Gehring, Springer-Verlag, Berlin 1982. [152](#), [171](#)

- [72] U. Tillmann, On the homotopy of the stable mapping class group. *Invent. Math.* 130 (2) (1997), 257–275. [207](#)
- [73] A. Weil, On the moduli of Riemann surfaces. 1979, In *Collected papers of André Weil* (originally unpublished, 1958), 381–389. [153](#), [177](#)
- [74] E. Witten, Two-dimensional gravity and intersection theory on moduli space. In *Surveys in differential geometry* (Cambridge, MA, 1990), Lehigh University, Bethlehem, PA, 1991, 243–310. [153](#), [155](#), [180](#)
- [75] S. A. Wolpert, The finite Weil-Petersson diameter of Riemann space. *Pacific J. Math.* 70 (1) (1977), 281–288. [177](#)
- [76] S. A. Wolpert, On the homology of the moduli space of stable curves. *Ann. of Math.* (2) 118 (3) (1983), 491–523. [153](#), [179](#)
- [77] S. A. Wolpert, On the symplectic geometry of deformations of a hyperbolic surface. *Ann. of Math.* (2) 117 (2) (1983), 207–234. [178](#), [197](#)
- [78] S. A. Wolpert, On the Weil-Petersson geometry of the moduli space of curves. *Amer. J. Math.* 107 (4) (1985), 969–997. [177](#)
- [79] S. A. Wolpert, On obtaining a positive line bundle from the Weil-Petersson class. *Amer. J. Math.* 107 (6) (1985), 1485–1507. [153](#), [179](#)
- [80] S. A. Wolpert, Chern forms and the Riemann tensor for the moduli space of curves. *Invent. Math.* 85 (1) (1986), 119–145. [153](#), [179](#)
- [81] S. A. Wolpert, The Weil-Petersson metric geometry. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich, 2009, 47–64. [177](#)
- [82] D. Zvonkine, Strebel differentials on stable curves and Kontsevich’s proof of Witten’s conjecture. Preprint, 2002; arXiv:math/0209071. [154](#), [190](#)

Chapter 6

Canonical 2-forms on the moduli space of Riemann surfaces

Nariya Kawazumi

Contents

1	Introduction	217
2	The cotangent space of the moduli space	219
3	The Weil–Petersson Kähler form	222
4	The first Chern form on the Siegel upper halfspace	224
5	Rauch’s variational formula	225
6	The Earle class and the twisted Morita–Mumford classes	229
7	A higher analogue of the period map	230
8	Secondary objects on the moduli space	232
	References	234

1 Introduction

Let $g \geq 2$ be an integer. The moduli space $\mathbb{M}_g = \mathcal{T}_g/\mathcal{M}_g$ of compact Riemann surfaces of genus g is the quotient space of Teichmüller space \mathcal{T}_g by the natural action of the mapping class group \mathcal{M}_g . Since Teichmüller space is contractible, the real cohomology of the mapping class group is isomorphic to that of the moduli space. As was shown by Harer [14], [15], the second homology of \mathbb{M}_g is of rank 1 if $g \geq 3$. This means that there exists a nontrivial second de Rham cohomology class on \mathbb{M}_g which is unique up to a constant factor. But several canonical 2-forms on the moduli space have been constructed in various geometric contexts, and they differ from each other. In this chapter we review some constructions of such canonical 2-forms in order to provide material for future research on the “secondary geometry” of the moduli space \mathbb{M}_g .

The signature of the total space of a fiber bundle is not necessarily equal to the product of the signatures of the base space and the fiber. The first example for this phenomenon was given by Kodaira [27] and Atiyah [6], who constructed a certain branched covering space of the product of two compact Riemann surfaces. The covering space has non-zero signature, while the signature of any compact Riemann surface is zero. We may regard the covering space as a family of compact Riemann surfaces parametrized by a compact Riemann surface, so that it defines a non-trivial 2-cycle

on the space \mathbb{M}_g . As was formulated by Meyer [30], [31], the signature of the total space of a family of compact Riemann surfaces defines a non-trivial 2-cocycle of the mapping class group \mathcal{M}_g and this provides a non-trivial cohomology class of degree 2 on the space \mathbb{M}_g . Nowadays this cocycle is called the Meyer cocycle and it has been playing an essential role in the topological study of fibered complex surfaces. See [4] and [5] for details.

The first and the second Betti numbers of the space \mathbb{M}_g , or, equivalently, those of the group \mathcal{M}_g , are given by

$$b_1(\mathbb{M}_g) = 0, \quad ([41], [45], [14, \text{p. 223}]), \quad (1.1)$$

$$b_2(\mathbb{M}_g) = 1, \quad \text{if } g \geq 3 \quad ([14], [15]). \quad (1.2)$$

For alternative computations of $b_2(\mathbb{M}_g)$, see [2], [28], [44]. The group $H^2(\mathbb{M}_g; \mathbb{R})$ is generated by the cohomology class of the Meyer cocycle. In the case $g = 2$ we have $b_2(\mathbb{M}_2) = 0$ because of Igusa's result $\mathbb{M}_2 = \mathbb{C}^3/(\mathbb{Z}/5) \simeq *$ (cf. [12]).

Mumford [42] and Morita [33] independently introduced a series of cohomology classes $e_n = (-1)^{n+1} \kappa_n \in H^{2n}(\mathbb{M}_g)$, $n \geq 1$, the Morita–Mumford classes or the tautological classes. They are defined as follows. Let $\pi: \mathbb{C}_g \rightarrow \mathbb{M}_g$ be the universal family of compact Riemann surfaces of genus g . The relative tangent bundle of the map π , $T_{\mathbb{C}_g/\mathbb{M}_g}$, the kernel of the differential $d\pi: T\mathbb{C}_g \rightarrow \pi^*T\mathbb{M}_g$, is a complex line V-bundle over \mathbb{C}_g . The n -th Morita–Mumford class $e_n = (-1)^{n+1} \kappa_n$, $n \geq 1$, is defined to be the integral of the $(n + 1)$ -st power of the Chern class of the bundle $T_{\mathbb{C}_g/\mathbb{M}_g}^\times$ along the fiber

$$e_n = (-1)^{n+1} \kappa_n = \int_{\text{fiber}} c_1(T_{\mathbb{C}_g/\mathbb{M}_g}^\times)^{n+1} \in H^{2n}(\mathbb{M}_g). \quad (1.3)$$

The first one $e_1 = \kappa_1$ is 3 times the cohomology class of the Meyer cocycle. As was proved by Morita [34] and Miller [32], the Morita–Mumford classes are algebraically independent in the stable range $* < \frac{2}{3}g$ [16] of the cohomology algebra $H^*(\mathbb{M}_g; \mathbb{R})$. Their proofs generalize the construction of Kodaira and Atiyah. Madsen and Weiss [29] proved that the cohomology algebra $H^*(\mathbb{M}_g; \mathbb{R})$ in the stable range is generated by the Morita–Mumford classes.

From the results (1.1) and (1.2) the simplest non-trivial cohomology classes on \mathbb{M}_g are of degree 2, and they are unique up to a constant factor. But several 2-forms on \mathbb{M}_g , or, equivalently, \mathcal{M}_g -equivariant 2-forms on Teichmüller space \mathcal{T}_g , have been canonically constructed in various geometric contexts.

From the uniformization theorem any compact Riemann surface C of genus $g \geq 2$ admits a unique hyperbolic metric. The volume form of the hyperbolic metric defines the Weil–Petersson pairing on the cotangent space $T_{[C]}^*\mathbb{M}_g$ involved with no additional information. As was shown by Wolpert [49] the Weil–Petersson–Kähler form ω_{WP} represents the first Morita–Mumford class e_1 . Thus we obtain a canonical 2-form representing e_1 .

The period map is a canonical map defined on Teichmüller space into the Siegel upper halfspace \mathfrak{H}_g . We have a canonical 2-form on \mathfrak{H}_g whose pullback represents the class e_1 on the moduli space \mathbb{M}_g .

We have another canonical metric on a compact Riemann surface. A natural Hermitian product on the space of holomorphic 1-forms defines the volume form B in 5.3 which induces a Hermitian metric on the Riemann surface. The Arakelov–Green function is derived from the volume form B . As will be stated in §7 and §8, a higher analogue of the period map is constructed and yields other canonical 2-forms representing e_1 . These forms are closely related to the volume form B .

All of them differ from each other. As to 2-forms representing non-trivial cohomology classes of degree 2 on the moduli space \mathbb{M}_g , the term ‘canonical’ does not imply ‘unique’. The difference of such forms should induce some secondary object on the moduli space \mathbb{M}_g . Assume $g \geq 3$. If we have two real $(1, 1)$ -forms ψ_1 and ψ_2 on \mathbb{M}_g representing e_1 , then there exists a real-valued function $f \in C^\infty(M; \mathbb{R})$ such that $\psi_2 - \psi_1 = \frac{\sqrt{-1}}{2\pi} \partial \bar{\partial} f$. Such a function f is unique up to a constant. See Lemma 8.1. This function captures the difference between these two forms, so that it should describe a certain relation between the two geometric contexts behind these forms.

In this chapter we review some constructions of canonical 2-forms. In §2 we give a short review on the cotangent spaces of moduli spaces. They are naturally isomorphic to some spaces of quadratic differentials. In §3 we take a quick glance at the Weil–Petersson Kähler form, which is related to the Virasoro cocycle through the Krichever construction. The most classical 2-form on \mathbb{M}_g is the pullback of the first Chern form on the Siegel upper halfspace \mathfrak{H}_g by the period map Jac , or, equivalently, the first Chern form of the Hodge bundle on \mathbb{M}_g . We explain this form in §§4 and 5. The Hodge bundle yields all the odd Morita–Mumford classes but not the even ones. We can obtain other canonical differential forms on the moduli space representing all the Morita–Mumford class e_i , $i \geq 1$, through a higher analogue of the period map, and this is described in §§6 and 7. Among them some 2-forms seem to be related to Arakelov geometry, as will be discussed in §8.

Acknowledgments. First of all the author thanks Athanase Papadopoulos, the editor, for careful reading and valuable comments on this chapter. Furthermore he thanks Leon Takhtajan for helpful comments on an earlier version. This work was partially supported by Grant-in-Aid for Scientific Research (A) (No.18204002), the Japan Society for Promotion of Sciences.

2 The cotangent space of the moduli space

Let C be a compact Riemann surface of genus $g \geq 2$, P_0 a point on C . Then we denote by $H^q(C; aK + bP_0)$, $q = 0, 1$, and $a, b \in \mathbb{Z}$, the q -th cohomology group $H^q(C; \mathcal{O}_C(T^*C^{\otimes a} \otimes [P_0]^{\otimes b}))$. Moreover we denote by $\Omega^q(C)$ the complex-valued

q -currents on C for $0 \leq q \leq 2$. The Hodge $*$ -operator $*$: $(T_{\mathbb{R}}^*C) \otimes \mathbb{C} \rightarrow (T_{\mathbb{R}}^*C) \otimes \mathbb{C}$ on the cotangent bundle of C depends only on the complex structure of C . The $-\sqrt{-1}$ -eigenspace is the holomorphic cotangent bundle T^*C , and the $\sqrt{-1}$ -eigenspace is the antiholomorphic cotangent bundle $\overline{T^*C}$. The operator $*$ decomposes the space $\Omega^1(C)$ into the $\pm\sqrt{-1}$ -eigenspaces

$$\Omega^1(C) = \Omega^{1,0}(C) \oplus \Omega^{0,1}(C),$$

where $\Omega^{1,0}(C)$ is the $-\sqrt{-1}$ -eigenspace and $\Omega^{0,1}(C)$ the $\sqrt{-1}$ -eigenspace. Throughout this chapter we denote by φ' and φ'' the $(1, 0)$ - and the $(0, 1)$ -parts of $\varphi \in \Omega^1(C)$, respectively, i.e.,

$$\varphi = \varphi' + \varphi'', \quad *\varphi = -\sqrt{-1}\varphi' + \sqrt{-1}\varphi''.$$

If φ is harmonic, then φ' is holomorphic and φ'' anti-holomorphic.

The Kodaira–Spencer map gives a natural isomorphism

$$T_{[C]}\mathbb{M}_g = H^1(C; -K). \quad (2.1)$$

To look at the isomorphism (2.1) more explicitly, consider a C^∞ family of compact Riemann surfaces C_t , $t \in \mathbb{R}$, $|t| \ll 1$, with $C_0 = C$. The family $\{C_t\}$ is trivial as a C^∞ fiber bundle over an interval near $t = 0$, so that we have a C^∞ family of C^∞ diffeomorphisms $f^t: C \rightarrow C_t$ with $f^0 = 1_C$. In general, if $\bigcirc = \bigcirc_t$ is a “function” in $t \in \mathbb{R}$, $|t| \ll 1$, then we write simply

$$\dot{\bigcirc} = \left. \frac{d}{dt} \right|_{t=0} \bigcirc_t.$$

For example, we denote

$$\dot{\mu} = \left. \frac{d}{dt} \right|_{t=0} \mu(f^t).$$

Here $\mu(f^t)$ is the complex dilatation of the diffeomorphism f^t . Let z_1 be a complex coordinate on C , and ζ_1 on C_t . The complex dilatation $\mu(f^t)$ is defined locally by

$$\mu(f^t) = \mu(f^t)(z_1) \frac{d}{dz_1} \otimes d\bar{z}_1 = \frac{(\zeta_1 \circ f^t)_{\bar{z}_1}}{(\zeta_1 \circ f^t)_{z_1}} \frac{d}{dz_1} \otimes d\bar{z}_1,$$

which does not depend on the choice of the coordinates z_1 and ζ_1 . The Dolbeault cohomology class $[\dot{\mu}] \in H^1(C; -K)$ is exactly the tangent vector $\left. \frac{d}{dt} \right|_{t=0} [C_t] \in T_{[C]}\mathbb{M}_g$.

We define a linear operator $S = S[\dot{\mu}]: \Omega^1(C) \rightarrow \Omega^1(C)$ by

$$S(\varphi) = S(\varphi') + S(\varphi'') := -2\varphi'\dot{\mu} - 2\varphi''\overline{\dot{\mu}}$$

for $\varphi = \varphi' + \varphi''$, $\varphi' \in \Omega^{1,0}(C)$, $\varphi'' \in \Omega^{0,1}(C)$. From straightforward computation we have

$$\dot{*} = *S = -S*: \Omega^1(C) \rightarrow \Omega^1(C). \quad (2.2)$$

By Serre duality we have a natural isomorphism

$$T_{[C]}^* \mathbb{M}_g = H^0(C; 2K). \quad (2.3)$$

The space $H^0(C; 2K)$ consists of the holomorphic quadratic differentials on C . For any holomorphic quadratic differential q the covariant tensor $q \dot{\mu}$ can be regarded as a $(1, 1)$ -form on C . The integral $\int_C q \dot{\mu}$ is just the value of the covector q at the tangent vector $[\dot{\mu}] = \left. \frac{d}{dt} \right|_{t=0} [C_t]$.

Let \mathbb{C}_g denote the moduli space of pointed compact Riemann surfaces (C, P_0) of genus g with $P_0 \in C$. The forgetful map $\pi: \mathbb{C}_g \rightarrow \mathbb{M}_g$, $[C, P_0] \mapsto [C]$, can be interpreted as the universal family of compact Riemann surfaces on the moduli space \mathbb{M}_g . We identify

$$T_{[C, P_0]} \mathbb{C}_g = H^1(C; -K - P_0) \quad \text{and} \quad T_{[C, P_0]}^* \mathbb{C}_g = H^0(C; 2K + P_0) \quad (2.4)$$

in a way similar to the space \mathbb{M}_g .

The relative tangent bundle of the forgetful map π with the zero section deleted

$$T_{\mathbb{C}_g/\mathbb{M}_g}^\times = T_{\mathbb{C}_g/\mathbb{M}_g} \setminus (\text{zero section})$$

can be interpreted as the moduli space of triples (C, P_0, v) of genus g . Here C is a compact Riemann surface of genus g , $P_0 \in C$, and $v \in T_{P_0} C \setminus \{0\}$. Similarly the space of quadratic differentials $H^0(C; 2K + 2P_0)$ is identified with the cotangent space of $T_{\mathbb{C}_g/\mathbb{M}_g}^\times$

$$T_{[C, P_0, v]}^* T_{\mathbb{C}_g/\mathbb{M}_g}^\times = H^0(C; 2K + 2P_0). \quad (2.5)$$

Moreover this space is closely related to Ehresmann connections on the bundle $T_{\mathbb{C}_g/\mathbb{M}_g}$. In general, let $\varpi: L \rightarrow M$ be a holomorphic line bundle over a complex manifold M , and L^\times the total space with the zero section deleted $L^\times = L \setminus (\text{zero section})$. We denote by R_a the right action of $a \in \mathbb{C}^\times := \mathbb{C} \setminus \{0\}$ on the space L^\times , and by Z the vector field on L^\times generated by the action R_a

$$Z := \left. \frac{d}{dt} \right|_{t=0} R_{e^t}.$$

An Ehresmann connection A (of type $(1, 0)$) on the bundle L is a $(1, 0)$ -form on the space L^\times with the conditions

$$A(Z) = 1,$$

and

$$R_{e^t}^* A = A, \quad \text{for all } t \in \mathbb{C}$$

([7], [26]). In other words, it is a splitting of the extension of holomorphic vector bundles over M

$$0 \rightarrow T^*M \xrightarrow{\varpi^*} (T^*L^\times)/\mathbb{C}^\times \xrightarrow{Z} \mathbb{C} \rightarrow 0.$$

Then there exists a unique $(1, 1)$ -form $c_1(A)$ on M such that $\frac{\sqrt{-1}}{2\pi}dA = \varpi^*c_1(A)$. The form $c_1(A)$ is, by definition, the Chern form of the connection A and represents the first Chern class of the line bundle L

$$[c_1(A)] = c_1(L) \in H^2(M; \mathbb{R}).$$

Now we let $M = \mathbb{C}_g$ and $L = T_{\mathbb{C}_g/\mathbb{M}_g}$. By straightforward computation we have a natural commutative diagram:

$$\begin{CD} 0 @>>> T_{[C, P_0]}^*M @>\varpi^*>> ((T^*L^\times)/\mathbb{C}^\times)_{[C, P_0]} @>Z>> \mathbb{C} @>>> 0 \\ @. @| @VVV @| @. \\ 0 @>>> H^0(C; 2K + P_0) @>>> H^0(C; 2K + 2P_0) @>2\pi\sqrt{-1}\text{Res}_{P_0}>> \mathbb{C} @>>> 0. \end{CD}$$

Here $\text{Res}_{P_0} : H^0(C; 2K + 2P_0) \rightarrow \mathbb{C}$ is the residue map of quadratic differentials at P_0 defined by

$$\text{Res}_{P_0}(q_{-2}z^{-2} + q_{-1}z^{-1} + q_0 + q_1z^1 + \dots)dz^{\otimes 2} = q_{-2},$$

where z is a complex coordinate centered at P_0 . It is easy to check that q_{-2} does not depend on the choice of the coordinate z . Consequently any C^∞ family $q = \{q(C, P_0)\}_{[C, P_0] \in \mathbb{C}_g}$, $q(C, P_0) \in H^0(C; 2K + 2P_0)$ of quadratic differentials parametrized by the space \mathbb{C}_g satisfying the condition $\text{Res}_{P_0} q(C, P_0) = \frac{1}{2\pi\sqrt{-1}}$ for any $[C, P_0] \in \mathbb{C}_g$ corresponds to an Ehresmann connection on the relative tangent bundle $T_{\mathbb{C}_g/\mathbb{M}_g}$. The $(1, 1)$ form $\frac{\sqrt{-1}}{2\pi}\bar{\partial}q$ on the space \mathbb{C}_g represents the first Chern class of the bundle $T_{\mathbb{C}_g/\mathbb{M}_g}$ (see [20]):

$$\frac{\sqrt{-1}}{2\pi}[\bar{\partial}q] = c_1(T_{\mathbb{C}_g/\mathbb{M}_g}) \in H^2(\mathbb{C}_g; \mathbb{R}). \tag{2.6}$$

3 The Weil–Petersson Kähler form

As was shown in §2 the cotangent space of the moduli space \mathbb{M}_g at $[C]$ is naturally isomorphic to the space of holomorphic quadratic differentials, $H^0(C; 2K)$. Let $d\text{vol}$ denote the hyperbolic volume form on the Riemann surface C . It is regarded as a Hermitian metric on the relative tangent bundle $T_{\mathbb{C}_g/\mathbb{M}_g}$. For any two differentials $q_1, q_2 \in H^0(C; 2K)$ the Weil–Petersson pairing $\langle q_1, q_2 \rangle_{\text{WP}}$ is defined by the integral

$$\langle q_1, q_2 \rangle_{\text{WP}} = \int_C q_1 \bar{q}_2 / d\text{vol}.$$

Here $q_1 \bar{q}_2 / d\text{vol}$ is regarded as a $(1, 1)$ -form on C . The pairing induces a Hermitian metric on the moduli space \mathcal{M}_g , the Weil–Petersson metric. Ahlfors [1] proved that it

is Kähler. See [10] for an alternative gauge-theoretic proof. Let ω_{WP} denote the Kähler form of the Weil–Petersson metric.

Now recall the original definition of the i -th Morita–Mumford class $e_i = (-1)^{i+1} \kappa_i$, $i \geq 1$ ([42], [33]). It is defined to be the integral along the fiber of the $(i + 1)$ -st power of the first Chern class of the relative tangent bundle $T_{\mathbb{C}_g/\mathbb{M}_g}$:

$$e_i = (-1)^{i+1} \kappa_i = \int_{\text{fiber}} c_1(T_{\mathbb{C}_g/\mathbb{M}_g})^{i+1} \in H^{2i}(\mathbb{M}_g). \tag{3.1}$$

It is one of the most orthodox ways to obtain differential forms representing the Morita–Mumford classes to take the integral of powers of the hyperbolic Chern form of the relative tangent bundle $T_{\mathbb{C}_g/\mathbb{M}_g}$ along the fiber. This was carried out by Wolpert [49]. He computed the Chern form $c_1^{\text{hyperbolic}}(T_{\mathbb{C}_g/\mathbb{M}_g})$ of the hyperbolic metric explicitly, and he proved that

$$\int_{\text{fiber}} c_1^{\text{hyperbolic}}(T_{\mathbb{C}_g/\mathbb{M}_g})^2 = \frac{1}{2\pi^2} \omega_{\text{WP}} \tag{3.2}$$

as differential forms on the moduli space \mathbb{M}_g . As a corollary we have

$$\frac{1}{2\pi^2} [\omega_{\text{WP}}] = e_1 \in H^2(\mathbb{M}_g; \mathbb{R}).$$

Furthermore Wolpert [50] gave a description of the Weil–Petersson Kähler form in terms of the Fenchel–Nielsen coordinates (τ_j, ℓ_j) , $1 \leq j \leq 3g - 3$, for any pants decomposition of the surface

$$\omega_{\text{WP}} = \sum d\ell_i \wedge d\tau_i. \tag{3.3}$$

Here ℓ_j denotes the geodesic length of each simple closed curve in the decomposition, and $\tau_j \in \mathbb{R}$ the hyperbolic displacement parameter. Penner [43] described explicitly the pullback of ω_{WP} to the decorated Teichmüller space. Goldman [11] generalized the Weil–Petersson geometry to the space of surface group representations in a reductive Lie group.

Now we consider the Lie algebra \mathfrak{d} of complex analytic vector fields on the punctured disk $\{z \in \mathbb{C}; 0 < |z| < \varepsilon\}$, $0 < \varepsilon \ll 1$. The 2-cochain vir on \mathfrak{d} defined by

$$\begin{aligned} \text{vir} \left(f_1(z) \frac{d}{dz}, f_2(z) \frac{d}{dz} \right) &:= \frac{1}{2\pi\sqrt{-1}} \oint_{|z|=1} \det \begin{pmatrix} f_1'(z) & f_2'(z) \\ f_1''(z) & f_2''(z) \end{pmatrix} dz \\ &= \frac{\sqrt{-1}}{2\pi} \oint_{|z|=1} \det \begin{pmatrix} f_1(z) & f_2(z) \\ f_1'''(z) & f_2'''(z) \end{pmatrix} dz \end{aligned}$$

is a cocycle and it is called the Virasoro cocycle. Its cohomology class generates the second Lie algebra cohomology group $H^2(\mathfrak{d}) = \mathbb{C}$.

Arbarello, De Concini, Kac and Procesi [3] established an isomorphism of $H^2(\mathbf{d})$ onto the second cohomology group of \mathbb{M}_g

$$v: H^2(\mathbf{d}) \xrightarrow{\cong} H^2(\mathbb{M}_g; \mathbb{C}) \tag{3.4}$$

induced by the Krichever construction.

For a local coordinate z on a Riemann surface one can define a local differential operator, or a local complex analytic Gel'fand–Fuks 1-cocycle with values in quadratic differentials by

$$\nabla_2^{d/dz}: f(z) \frac{d}{dz} \mapsto \frac{1}{6} f'''(z)(dz)^{\otimes 2}$$

(see [19, p. 666]). The cocycle $\nabla_2^{d/dz}$ is equivalent to a projective structure. In fact, if w is another coordinate, then

$$\nabla_2^{d/dw} X - \nabla_2^{d/dz} X = \mathcal{L}_X(\{w, z\}(dz)^{\otimes 2})$$

for any local complex analytic vector field X . Here $\{w, z\}$ denotes the Schwarzian derivative. In particular, the hyperbolic structure on a (hyperbolic) Riemann surface defines a global operator $\nabla_2^{\text{hyperbolic}}$.

The Krichever construction relates the 2-cocycle vir with the operator $\nabla_2^{\text{hyperbolic}}$. By straightforward computation using the Bers embedding we have

$$\bar{\partial} \nabla_2^{\text{hyperbolic}} = 8\omega_{\text{wp}} \tag{3.5}$$

as (1, 1)-forms on the moduli space \mathbb{M}_g . This result, the first variation of the hyperbolic structure coincides with ω_{wp} , was first proved by Zograf and Takhtajan [51, p. 310].

4 The first Chern form on the Siegel upper halfspace

The Hodge bundle $\Lambda_{\mathbb{M}_g}$ is defined to be the holomorphic vector bundle on \mathbb{M}_g whose fiber over $[C]$ is the space of holomorphic 1-forms on C

$$\Lambda_{\mathbb{M}_g} = \coprod_{[C] \in \mathbb{M}_g} H^0(C; K).$$

We write simply c_1 for the first Chern class of $\Lambda_{\mathbb{M}_g}$

$$c_1 = c_1(\Lambda_{\mathbb{M}_g}) \in H^2(\mathbb{M}_g; \mathbb{R}).$$

The bundle $\Lambda_{\mathbb{M}_g}$ comes from a symplectic equivariant vector bundle on the Siegel upper halfspace \mathfrak{H}_g . In fact, the space \mathfrak{H}_g can be identified with the space of almost complex structures J on the real $2g$ -dimensional symplectic vector space (\mathbb{R}^{2g}, \cdot) with

the conditions

$$\begin{aligned} Jx \cdot Jy &= x \cdot y, & \text{for all } x, y \in \mathbb{R}^{2g}, \\ x \cdot Jx &> 0, & \text{for all } x \in \mathbb{R}^{2g} \setminus \{0\}. \end{aligned}$$

We have a holomorphic vector bundle $E'_{\mathfrak{H}_g}$ on \mathfrak{H}_g whose fiber over J is the $-\sqrt{-1}$ -eigenspace of J . We have a natural isomorphism of vector bundles

$$T^*\mathfrak{H}_g = \text{Sym}^2 E'_{\mathfrak{H}_g}. \tag{4.1}$$

For each Riemann surface C the Hodge $*$ -operator on the 1-forms induces such an almost complex structure on the space of real harmonic 1-forms. This induces a holomorphic map $\text{Jac}: \mathbb{M}_g \rightarrow \mathfrak{H}_g / \text{Sp}_{2g}(\mathbb{Z})$ known as the period map in the classical context. The pullback of $E'_{\mathfrak{H}_g}$ by the map Jac is exactly the Hodge bundle $\Lambda_{\mathbb{M}_g}$.

Thus the cohomology class c_1 can be regarded as an integral cohomology class of the Siegel modular group $\text{Sp}_{2g}(\mathbb{Z})$, $c_1 \in H^2(\text{Sp}_{2g}(\mathbb{Z}); \mathbb{Z})$. Meyer [30] proved that the cohomology class of the Meyer cocycle is equal to $4c_1 \in H^2(\text{Sp}_{2g}(\mathbb{Z}); \mathbb{Z})$. From the Grothendieck–Riemann–Roch formula, or, equivalently, the Atiyah–Singer index theorem for families, it follows that

$$\frac{1}{12}e_1 = c_1 \in H^2(\mathbb{M}_g; \mathbb{R}). \tag{4.2}$$

To describe a canonical 2-form representing $c_1(E'_{\mathfrak{H}_g})$ we consider the quotient vector bundle $E''_{\mathfrak{H}_g} := (\mathfrak{H}_g \times \mathbb{C}^{2g})/E'_{\mathfrak{H}_g}$, and the family of projections $\pi = \{\pi_J\}_{J \in \mathfrak{H}_g}$ on \mathbb{C}^{2g} , $\pi_J := \frac{1}{2}(1 - \sqrt{-1}J)$, parametrized by \mathfrak{H}_g . Then $\{\pi_J \circ d\}_{J \in \mathfrak{H}_g}$ is a covariant derivative ∇ of type $(1, 0)$ on the bundle $E''_{\mathfrak{H}_g} \cong \coprod_{J \in \mathfrak{H}_g} \text{Image } \pi_J$, whose curvature form R^∇ is given by

$$R^\nabla = \pi(\partial\pi)(\bar{\partial}\pi). \tag{4.3}$$

The 2-form $c_1(\nabla)$ defined by $c_1(\nabla) = \frac{\sqrt{-1}}{2\pi} \text{trace } R^\nabla$ represents $c_1(E'_{\mathfrak{H}_g})$. Let $J_\alpha(t) \in \mathfrak{H}_g$, $|t| \ll 1$, $\alpha = 1, 2$, be C^∞ paths on \mathfrak{H}_g with $J_1(0) = J_2(0) = J$. Then, one can compute

$$c_1(\nabla)_J = \frac{1}{8\pi} \text{trace}(\dot{J}_1 J \dot{J}_2). \tag{4.4}$$

In the next section we prove Rauch’s variational formula to obtain the pullback of $c_1(\nabla)_J$ by the period map Jac explicitly.

5 Rauch’s variational formula

Rauch’s variational formula describes the differential of the period map Jac . Let C be a compact Riemann surface of genus g . We denote by H the real first homology group $H_1(C; \mathbb{R})$. Consider the map $H^* = H^1(C; \mathbb{R}) \rightarrow \Omega^1(C)$ assigning to each

cohomology class the harmonic 1-form representing it. The map can be regarded as an H -valued 1-form $\omega_{(1)} \in \Omega^1(C) \otimes H$.

Let $\{X_i, X_{g+i}\}_{i=1}^g$ be a symplectic basis of $H_{\mathbb{C}} = H_1(C; \mathbb{C})$

$$X_i \cdot X_{g+j} = \delta_{ij}, \quad X_i \cdot X_j = X_{g+i} \cdot X_{g+j} = 0, \quad 1 \leq i, j \leq g,$$

and $\{\xi_i, \xi_{g+i}\}_{i=1}^g \subset \Omega^1(C)$ the basis of the harmonic 1-forms dual to $\{X_i, X_{g+i}\}_{i=1}^g$. Then we have

$$\omega_{(1)} = \sum_{i=1}^g \xi_i X_i + \xi_{g+i} X_{g+i} \in \Omega^1(C) \otimes H_{\mathbb{C}}.$$

In particular, if $\{\psi_i\}_{i=1}^g \subset H^0(C; K)$ is an orthonormal basis

$$\frac{\sqrt{-1}}{2} \int_C \psi_i \wedge \bar{\psi}_j = \delta_{ij}, \quad 1 \leq i, j \leq g, \quad (5.1)$$

then we obtain

$$\omega_{(1)} = \sum_{i=1}^g \psi_i Y_i + \bar{\psi}_i \bar{Y}_i, \quad (5.2)$$

where $\{Y_i, Y_{g+i}\}_{i=1}^g \subset H_{\mathbb{C}}$ is the dual basis of the symplectic basis $\{[\psi_i], \frac{\sqrt{-1}}{2}[\bar{\psi}_i]\}_{i=1}^g$ of $H_{\mathbb{C}}^* = H^1(C; \mathbb{C})$. Since the complete linear system of the canonical divisor on the complex algebraic curve C has no basepoint, the 2-form

$$B = \frac{1}{2g} \omega_{(1)} \cdot \omega_{(1)} = \frac{\sqrt{-1}}{2g} \sum_{i=1}^g \psi_i \wedge \bar{\psi}_i \in \Omega^2(C) \quad (5.3)$$

is a volume form on C .

Now we recall the Hodge decomposition of the 1-forms on C . We have an exact sequence

$$0 \rightarrow \mathbb{C} \rightarrow \Omega^0(C) \xrightarrow{d^*d} \Omega^2(C) \xrightarrow{\int_C} \mathbb{C} \rightarrow 0.$$

The vector space \mathbb{C} on the left side means the constant functions. A Green operator $\Psi: \Omega^2(C) \rightarrow \Omega^0(C)$ is a linear map satisfying the property

$$d * d\Psi \Omega = \Omega$$

for any $\Omega \in \Omega^2(C)$ with $\int_C \Omega = 0$. In this chapter we use two sorts of Green operators $\hat{\Phi} = \hat{\Phi}_C$ and $\Phi = \Phi^{(C, P_0)}$. The former is characterized by the conditions

$$d * d\hat{\Phi}(\Omega) = \Omega - \left(\int_C \Omega \right) B \quad \text{and} \quad \int_C \hat{\Phi}(\Omega) B = 0 \quad (5.4)$$

for any $\Omega \in \Omega^2(C)$. Let $\delta_{P_0}: C^\infty(C) \rightarrow \mathbb{C}$, $f \mapsto f(P_0)$, be the delta current on C at the point P_0 . We define the latter Φ to be a linear map with values in $\Omega^0(C)/\mathbb{C}$ instead of $\Omega^0(C)$. Then the operator $d\Phi: \Omega^2(C) \rightarrow \Omega^1(C)$ makes sense, and the operator Φ

is defined by the condition

$$d * d\Phi\Omega = \Omega - \left(\int_C \Omega \right) \delta_{P_0}$$

for any $\Omega \in \Omega^2(C)$.

Any Green operator Ψ induces the Hodge decomposition of the 1-currents

$$\varphi = \mathcal{H}\varphi + d\Psi d * \varphi + *d\Psi d\varphi \tag{5.5}$$

for any $\varphi \in \Omega^1(C)$, where $\mathcal{H}: \Omega^1(C) \rightarrow \Omega^1(C)$ is the harmonic projection on the 1-currents on C .

In the setting of §2 the first variation of $\omega_{(1)}$ is given by

$$\dot{\omega}_{(1)} = -d\Psi d * S\omega_{(1)}. \tag{5.6}$$

In fact, differentiating $d * \omega_{(1)} = 0$, we get

$$d * \dot{\omega}_{(1)} = -d \dot{*} \omega_{(1)} = -d * S\omega_{(1)}.$$

Since $f^{t*} \omega_{(1)}$ is cohomologous to $\omega_{(1)}$, we have some function u such that $\dot{\omega}_{(1)} = du$. Hence from (5.5) we obtain

$$\dot{\omega}_{(1)} = d\Psi d * \dot{\omega}_{(1)} = -d\Psi d * \omega_{(1)},$$

as was to be shown.

Theorem 5.1 (Rauch). *The diagram*

$$\begin{array}{ccc} T_{[C]}^* \mathbb{M}_g & \xleftarrow{(d \text{ Jac})^*} & T_{[\text{Jac}(C)]}^* \mathcal{H}_g / \text{Sp}_{2g}(\mathbb{Z}) \\ \parallel & & \parallel \\ H^0(C; 2K) & \xleftarrow{2\sqrt{-1}(\text{multiplication})} & \text{Sym}^2 H^0(C; K) \end{array}$$

commutes. Here the lower horizontal arrow maps $\psi_1 \otimes \psi_2$ to the quadratic differential $2\sqrt{-1}\psi_1\psi_2$ for any 1-forms ψ_1 and $\psi_2 \in H^0(C; K)$.

Proof. The integral $\int_C * \omega_{(1)} \wedge \omega_{(1)} \in H \otimes H = H^* \otimes H = \text{Hom}(H, H)$ coincides with the almost complex structure on $H = H_1(C; \mathbb{R})$ induced by the Hodge $*$ -operator.

Since $\omega_{(1)}$ is harmonic and $\dot{\omega}_{(1)}$ is d -exact by (5.6), we have

$$\int_C * \omega_{(1)} \wedge \dot{\omega}_{(1)} = - \int_C \omega_{(1)} \wedge * \dot{\omega}_{(1)} = 0.$$

Hence

$$\begin{aligned} \left(\int_C * \omega_{(1)} \wedge \omega_{(1)} \right)' &= \int_C \dot{*} \omega_{(1)} \wedge \omega_{(1)} = \int_C (*S\omega_{(1)}) \wedge \omega_{(1)} \\ &= 2\sqrt{-1} \int_C \omega_{(1)'} \omega_{(1)'} \dot{\mu} - 2\sqrt{-1} \overline{\left(\int_C \omega_{(1)'} \omega_{(1)'} \dot{\mu} \right)}. \end{aligned}$$

This proves the theorem. \square

Substituting the theorem into the formula (4.4) we have

Corollary 5.2.

$$\text{Jac}^* c_1(\nabla) = \frac{1}{8\pi\sqrt{-1}} \sum_{i,j=1}^g \psi_i \psi_j \otimes \bar{\psi}_i \bar{\psi}_j \in T_{[C]}^* \mathbb{M}_g \otimes \overline{T_{[C]}^* \mathbb{M}_g}.$$

Here $\{\psi_i\}_{i=1}^g \subset H^0(C; K)$ is any orthonormal basis (5.1).

The elementary polynomials $\sigma_1, \dots, \sigma_g$ in indeterminates x_1, \dots, x_g are given by $\prod_{i=1}^g (t - x_i) = t^g + \sum_{k=1}^g (-1)^k \sigma_k t^{g-k}$. The equation $\sum_{i=1}^g x_i^m = s_m(\sigma_1, \dots, \sigma_g)$ defines the m -th Newton polynomial s_m . The m -th Newton class of the Hodge bundle $\Lambda = \Lambda_{\mathbb{M}_g}$ is defined by

$$s_m(\Lambda) = s_m(c_1(\Lambda), \dots, c_g(\Lambda)) \in H^{2m}(\mathbb{M}_g; \mathbb{R}),$$

where $c_k(\Lambda)$ is the k -th Chern class of the bundle Λ .

The complex conjugate $\bar{\Lambda}$ satisfies $s_m(\bar{\Lambda}) = (-1)^m s_m(\Lambda)$. Since $\Lambda \oplus \bar{\Lambda}$ is a flat vector bundle on \mathbb{M}_g whose fiber over $[C]$ is the homology group $H_1(C; \mathbb{C})$, we have

$$s_{2n}(\Lambda) = \frac{1}{2} s_{2n}(\Lambda \oplus \bar{\Lambda}) = 0.$$

From the Grothendieck–Riemann–Roch formula, or, equivalently, the Atiyah–Singer index theorem for families, it follows that

$$e_{2n-1} = (-1)^{n-1} \frac{2n}{B_{2n}} s_{2n-1}(\Lambda) \in H^{4n-2}(\mathbb{M}_g; \mathbb{R}). \quad (5.7)$$

Here B_{2n} is the n -th Bernoulli number. In the case $n = 1$ this is exactly the formula (4.2).

Hence the Hodge bundle yields all the odd Morita–Mumford classes, but not the even ones. To get all the Morita–Mumford classes we introduce a higher analogue of the period map, as will be discussed in the succeeding sections.

6 The Earle class and the twisted Morita–Mumford classes

Let Σ_g be a closed oriented C^∞ surface of genus g , $p_0 \in \Sigma_g$ a point, and $v_0 \in T_{p_0}\Sigma_g \setminus \{0\}$ a non-zero tangent vector at the point p_0 . We denote by \mathcal{M}_g , $\mathcal{M}_{g,*}$ and $\mathcal{M}_{g,1}$ the mapping class groups for the surface Σ_g , the pointed surface (Σ_g, p_0) and the triple (Σ_g, p_0, v_0) respectively. They are the orbifold fundamental groups of the spaces \mathbb{M}_g , \mathbb{C}_g and $T_{\mathbb{C}_g/\mathbb{M}_g}^\times$. The fundamental group $\pi_1(\Sigma_g, p_0)$ is naturally embedded into the group $\mathcal{M}_{g,*}$ (cf. [40]).

By abuse of notation let H denote the real first homology group of Σ_g , $H_1(\Sigma_g; \mathbb{R})$, on which the mapping class groups act in an obvious way. The module H can be interpreted as a flat vector bundle on the moduli space \mathbb{M}_g . In 1978 Earle [9] constructed an explicit 1-cocycle $\psi: \mathcal{M}_{g,*} \rightarrow H$ such that $(2 - 2g)\psi$ has values in $H_1(\Sigma_g; \mathbb{Z})$, and $\psi|_{\pi_1(\Sigma_g)}$ is equal to the abelianization map of the group $\pi_1(\Sigma_g)$. Later Morita [35] independently discovered a cohomology class $k \in H^1(\mathcal{M}_{g,*}; H_1(\Sigma_g; \mathbb{Z}))$ which is equal to $[(2 - 2g)\psi]$. Furthermore he proved that

$$H^1(\mathcal{M}_{g,*}; H_1(\Sigma_g; \mathbb{Z})) = \mathbb{Z}k \cong \mathbb{Z} \quad (6.1)$$

for $g \geq 2$. The author would like to propose the class k should be called *the Earle class*.

The square of the class k is related to the first Morita–Mumford class $e_1 = \kappa_1$ through the intersection pairing

$$m: H \otimes H = H_1(\Sigma_g; \mathbb{R}) \otimes H_1(\Sigma_g; \mathbb{R}) \rightarrow \mathbb{R}. \quad (6.2)$$

Morita [36] proved that

$$m_*(k^{\otimes 2}) = -e_1 + 2g(2 - 2g)e \in H^2(\mathcal{M}_{g,*}). \quad (6.3)$$

Here e is the first Chern class of the relative tangent bundle $c_1(T_{\mathbb{C}_g/\mathbb{M}_g}) \in H^2(\mathbb{C}_g) = H^2(\mathcal{M}_{g,*})$.

These phenomena have a higher analogue. The twisted Morita–Mumford class $m_{i,j} \in H^{2i+j-2}(\mathcal{M}_{g,1}; \Lambda^j H)$, $i, j \geq 0$, was introduced in [21]. We have $m_{1,1} = k$ and $m_{i+1,0} = e_i$, $i \geq 1$. All the cohomology classes on the mapping class groups with trivial coefficients (even in the unstable range) obtained from any products of the twisted Morita–Mumford classes by contracting the coefficients using the intersection pairing are exactly the polynomials in the Morita–Mumford classes, see [25].

This fact is closely related to the Johnson homomorphisms on the mapping class group. The fundamental group $\pi_1(\Sigma_g, p_0, v_0) = \pi_1(\Sigma_g \setminus \{p_0\}, v_0)$ with tangential basepoint v_0 is a free group of rank $2g$. Let Γ_k , $k \geq 0$, denote the lower central series of the free group $\pi_1(\Sigma_g, p_0, v_0)$. We have $\Gamma_0 = \pi_1(\Sigma_g, p_0, v_0)$ and $\Gamma_{k+1} = [\Gamma_k, \Gamma_0]$ for $k \geq 0$. The quotient Γ_1/Γ_2 is naturally isomorphic to $\bigwedge^2 H_1(\Sigma_g; \mathbb{Z}) \subset \bigwedge^2 H$. Let $\mathcal{J}_{g,1}$ be the Torelli group, that is, the kernel of the natural action of $\mathcal{M}_{g,1}$ on the homology group $H_1(\Sigma_g; \mathbb{Z})$. For any $\varphi \in \mathcal{J}_{g,1}$ and $\gamma \in \Gamma_0$, the difference $\gamma^{-1}\varphi(\gamma)$

belongs to Γ_1 from the definition of $\mathcal{J}_{g,1}$. Hence we can define a homomorphism

$$\tau_1(\varphi): H_1(\Sigma_g; \mathbb{Z}) \rightarrow \bigwedge^2 H_1(\Sigma_g; \mathbb{Z}), \quad [\gamma] \mapsto \gamma^{-1}\varphi(\gamma) \bmod \Gamma_2.$$

It is easy to check this induces a homomorphism $\tau_1: \mathcal{J}_{g,1} \rightarrow H^* \otimes \bigwedge^2 H \cong H \otimes \bigwedge^2 H$. The last isomorphism comes from Poincaré duality. Johnson [18] proved that the image $\tau_1(\mathcal{J}_{g,1})$ is included in $\bigwedge^3 H$. The homomorphism τ_1 is called the first Johnson homomorphism. Morita [38] proved there exists a unique cohomology class $\tilde{k} \in H^1(\mathcal{M}_{g,1}; \bigwedge^3 H)$ which restricts to τ_1 on the Torelli group $\mathcal{J}_{g,1}$. We call it the extended first Johnson homomorphism. See [40, §7] for more information on the Johnson homomorphisms.

The class $\frac{1}{6}m_{0,3}$ is equal to the extended first Johnson homomorphism $\tilde{k}: \mathcal{M}_{g,1} \rightarrow \bigwedge^3 H$ ([25]). Each of the Morita–Mumford classes is obtained from some power of \tilde{k} by contracting the coefficients using the intersection pairing m (see [39]). Conversely for any Sp-module V and any Sp-homomorphism $f: (\bigwedge^3 H)^{\otimes n} \rightarrow V$ induced by the intersection pairing, the cohomology class $f_*(\tilde{k}^{\otimes n})$ is a polynomial in the twisted Morita–Mumford class [25]. An extension of the second Johnson homomorphism to the whole mapping class group provides a fundamental relation among the twisted Morita–Mumford classes (see [22]). In the next section we introduce a flat connection on a vector bundle on the space $T_{\mathbb{C}_g/\mathbb{M}_g}^\times$, whose holonomy is an extension of the Johnson homomorphisms to the whole mapping class group $\mathcal{M}_{g,1}$.

7 A higher analogue of the period map

A complex-analytic counterpart of the first Johnson homomorphism is the (pointed) harmonic volume introduced by Harris [17], [46]. It is a real analytic section of a fiber bundle on the moduli space \mathbb{C}_g whose fiber over $[C, P_0]$ is $(\bigwedge^3 H_1(C; \mathbb{Z})) \otimes (\mathbb{R}/\mathbb{Z})$. The first variation of the (pointed) harmonic volumes is a twisted 1-form representing the cohomology class $[\tilde{k}]$ (see [23]).

To obtain “canonical” differential forms representing all the twisted Morita–Mumford classes and their higher relations, we construct a higher analogue of the classical period map and the harmonic volume, the harmonic Magnus expansion $\theta: \mathcal{T}_{g,1} \rightarrow \Theta_{2g}$ ([23]). The space $\mathcal{T}_{g,1} = \widetilde{T_{\mathbb{C}_g/\mathbb{M}_g}^\times}$ is the Teichmüller space of *triples* (C, P_0, v) of genus g . Here C is a compact Riemann surface of genus g , $P_0 \in C$, and v a non-zero tangent vector of C at P_0 as in §2. For any triple (C, P_0, v) one can define the fundamental group of the complement $C \setminus \{P_0\}$ with the tangential basepoint v denoted by $\pi_1(C, P_0, v)$, which is a free group of rank $2g$. The space Θ_n is the set of all Magnus expansions of the free group F_n of rank $n \geq 2$ in a wider sense stated as follows.

We denote by H the first real homology group of the group F_n , $H_1(F_n; \mathbb{R})$, H^* the first real cohomology group of F_n , $H^1(F_n; \mathbb{R})$, and $[\gamma] \in H$ the homology class of $\gamma \in F_n$. The completed tensor algebra generated by H , $\hat{T} = \hat{T}(H) = \prod_{m=0}^{\infty} H^{\otimes m}$, has a decreasing filtration of two-sided ideals $\{\hat{T}_p\}_{p \geq 1}$ defined by $\hat{T}_p = \prod_{m \geq p} H^{\otimes m}$. The subset $1 + \hat{T}_1$ is a subgroup of the multiplicative group of the algebra \hat{T} . We call a map $\theta: F_n \rightarrow 1 + \hat{T}_1$ a *Magnus expansion of the free group F_n* in a wider sense ([22]), if $\theta: F_n \rightarrow 1 + \hat{T}_1$ is a group homomorphism, and if $\theta(\gamma) \equiv 1 + [\gamma] \pmod{\hat{T}_2}$ for any $\gamma \in F_n$. One can endow the set of all Magnus expansions Θ_n with a natural structure of a (projective limit of) real analytic manifold(s). A certain (projective limit of) Lie group(s) $\text{IA}(\hat{T})$ acts on Θ_n in a free and transitive way. This induces a series of 1-forms $\eta_p \in \Omega^1(\Theta_n) \otimes H^* \otimes H^{\otimes(p+1)}$, $p \geq 1$, the Maurer–Cartan forms of the action of $\text{IA}(\hat{T})$, which are invariant under a natural action of the automorphism group of the group F_n , $\text{Aut}(F_n)$. The Maurer–Cartan formula $d\eta = \eta \wedge \eta$ allows us to regard the forms η_p as an equivariant flat connection on the vector bundle $\Theta_n \times H^* \otimes \hat{T}_2$. The holonomy of the connection is an extension of all the Johnson homomorphisms to the whole group $\text{Aut}(F_n)$. The 1-forms η_p represent the twisted Morita–Mumford classes on the group $\text{Aut}(F_n)$, cf. [22], [23].

Let (C, P_0, v) be a triple of genus g . From now on we denote by H the real first homology group $H_1(C; \mathbb{R})$. As in §5 we denote by $\delta_{P_0}: C^\infty(C) \rightarrow \mathbb{R}$, $f \mapsto f(P_0)$, the delta 2-current on C at P_0 . Then there exists a \hat{T}_1 -valued 1-current $\omega \in \Omega^1(C) \otimes \hat{T}_1$, satisfying the following 3 conditions

- (1) $d\omega = \omega \wedge \omega - I \cdot \delta_{P_0}$, where $I \in H^{\otimes 2}$ is the intersection form.
- (2) The first term of ω is equal to $\omega_{(1)} \in \Omega^1(C) \otimes H$ introduced in §5.
- (3) $\int_C (\omega - \omega_{(1)}) \wedge * \varphi = 0$ for any closed 1-form φ and each $p \geq 2$.

Using Chen’s iterated integrals [8], we can define a Magnus expansion

$$\theta = \theta^{(C, P_0, v)}: \pi_1(C, P_0, v) \rightarrow 1 + \hat{T}_1(H_1(C; \mathbb{R})), \quad [\ell] \mapsto 1 + \sum_{m=1}^{\infty} \int_{\ell} \underbrace{\omega \omega \dots \omega}_m.$$

Let a point $p_0 \in \Sigma_g$ and a non-zero tangent vector $v_0 \in T_{p_0} \Sigma_g \setminus \{0\}$ be fixed as in §6. Moreover we fix an isomorphism $\pi_1(\Sigma_g, p_0, v_0) \cong F_{2g}$. A marking α of a triple (C, P_0, v) is an orientation-preserving diffeomorphism of Σ_g onto C satisfying the conditions $\alpha(p_0) = P_0$ and $(d\alpha)_{p_0}(v_0) = v$. For any marked triple $[(C, P_0, v), \alpha]$ we define a Magnus expansion of the free group F_{2g} by

$$F_{2g} \cong \pi_1(\Sigma_g, p_0, v_0) \xrightarrow{\alpha_*} \pi_1(C, P_0, v) \xrightarrow{\theta^{(C, P_0, v)}} 1 + \hat{T}_1(H_1(C; \mathbb{R})) \xrightarrow{\alpha_*^{-1}} 1 + \hat{T}_1.$$

Consequently, the Magnus expansions $\theta^{(C, P_0, v)}$ for all the triples (C, P_0, v) define a canonical real analytic map $\theta: T_{\mathbb{C}_g/\mathbb{M}_g}^\times = \mathcal{T}_{g,1} \rightarrow \Theta_{2g}$, which we call *the harmonic Magnus expansion on the universal family of Riemann surfaces*. The pullbacks of the Maurer–Cartan forms η_p define a flat connection on a vector bundle on the space

$T_{\mathbb{C}_g/\mathbb{M}_g}^\times$, and give the canonical differential forms representing the Morita–Mumford classes and their higher relations.

Theorem 7.1 ([23]). *For any $[C, P_0, v, \alpha] \in \mathcal{T}_{g,1}$ we have*

$$(\theta^* \eta)_{[C, P_0, v, \alpha]} = 2\Re(N(\omega' \omega') - 2\omega_{(1)}' \omega_{(1)}') \in T_{[C, P_0, v, \alpha]}^* \mathcal{T}_{g,1} \otimes \hat{T}_3.$$

Here $N: \hat{T}_1 \rightarrow \hat{T}_1$ is defined by $N|_{H^{\otimes m}} = \sum_{k=0}^{m-1} \binom{1 \ 2 \ \dots \ m-1 \ m}{2 \ 3 \ \dots \ m \ 1}^k$, and the meromorphic quadratic differential $N(\omega' \omega')$ is regarded as a $(1, 0)$ -cotangent vector at $[C, P_0, v, \alpha] \in \mathcal{T}_{g,1}$ in a natural way.

The third homogeneous term $N(\omega' \omega')_{(3)} = N(\omega'_{(1)} \omega'_{(2)} + \omega'_{(2)} \omega'_{(1)})$ is the first variation of the (pointed) harmonic volumes of pointed Riemann surfaces. It represents the extended first Johnson homomorphism \bar{k} . The higher terms provide higher relations among the twisted Morita–Mumford classes. Hence all of the Morita–Mumford classes are represented by some algebraic combinations of $N(\omega' \omega')$.

The second term coincides with $2\omega_{(1)}' \omega_{(1)}'$, which is exactly the first variation of the period matrices given by Rauch’s formula in §5. Hence we may regard the harmonic Magnus expansion as a higher analogue of the classical period map Jac.

8 Secondary objects on the moduli space

The determinant of the Laplacian acting on the space of k -differentials on Riemann surfaces is a ‘secondary’ object on the moduli space. Zograf and Takhtajan [52] proved that it yields the difference on the moduli space of compact Riemann surfaces, \mathbb{M}_g , between a multiple of the Weil–Petersson form ω_{wp} and the Chern form of the Hodge line bundle for the k -differentials induced by the hyperbolic metric. Moreover, they studied analogous phenomena for punctured Riemann surfaces to introduce their Kähler metric, the Zograf–Takhtajan metric, on the moduli space of punctured Riemann surfaces [48].

In this section we discuss other secondary objects, which come from the higher analogue of the period map introduced in §7. Now we can obtain explicit 2-forms from the connection form $N(\omega' \omega')$ on $T_{\mathbb{C}_g/\mathbb{M}_g}^\times$, e^J on \mathbb{C}_g and e_1^J on \mathbb{M}_g . Consider the quadratic differential η_2' defined by

$$\eta_2' = N(\omega' \omega')_{(4)} \in H^0(C; 2K + 2P_0) \otimes H^{\otimes 4},$$

which satisfies

$$\frac{1}{2g(2g+1)} \operatorname{Res}_{P_0}((m \otimes m)(\eta_2')) = -\frac{1}{8\pi^2}.$$

Here m is the intersection pairing $m: H \otimes H \rightarrow \mathbb{R}$ as in (6.2). We define

$$e^J = \frac{-2}{2g(2g+1)} \bar{\partial}((m \otimes m)(\eta_2')) \in \Omega^{1,1}(\mathbb{C}_g).$$

From (2.6) e^J represents the first Chern class of the relative tangent bundle

$$[e^J] = e = c_1(T_{\mathbb{C}_g/\mathbb{M}_g}) \in H^2(\mathbb{C}_g; \mathbb{R}).$$

We obtain a twisted 1-form $\eta_1^H \in \Omega^1(\mathbb{C}_g; H)$ representing the Earle class k by contracting the coefficients of $\eta_1' = N(\omega'\omega')_{(3)}$. By (6.3) $m(\eta_1^H)^{\otimes 2} \in \Omega^{1,1}(\mathbb{C}_g)$ represents $-e_1 + 2g(2 - 2g)e$. So we define

$$e_1^J = -m(\eta_1^H)^{\otimes 2} + 2g(2 - 2g)e^J$$

which can be regarded as a $(1, 1)$ -form on \mathbb{M}_g , see [23, §8].

Hain and Reed [13] already constructed the same form e_1^J in a Hodge-theoretical context. They applied the following lemma to $\frac{1}{12}e_1^J - \text{Jac}^*c_1(\nabla)$ to get a function $\beta_g \in C^\infty(\mathbb{M}_g; \mathbb{R})/\mathbb{R}$, the Hain–Reed function, a secondary object on the moduli space \mathbb{M}_g .

Lemma 8.1. *Let M be a connected complex orbifold with $H^0(M; \mathcal{O}) = \mathbb{C}$ and $H^1(M; \mathbb{C}) = H^1(M; \mathcal{O}) = 0$. If a real $C^\infty(1, 1)$ -form ψ is d -exact, then there exists a real-valued function $f \in C^\infty(M; \mathbb{R})$ such that $\psi = \frac{\sqrt{-1}}{2\pi} \partial \bar{\partial} f$. Such a function f is unique up to a constant.*

Here we remark all the holomorphic functions on \mathbb{M}_g are constants provided $g \geq 3$. In fact, each of the boundary component of the Satake compactification of \mathbb{M}_g is of complex codimension ≥ 2 . The vanishing of the first cohomology follows from (1.1). See [41]. In [13], Hain and Reed also studied the asymptotic behavior of the function β_g towards the boundary of the Deligne–Mumford compactification $\overline{\mathbb{M}}_g^{\text{DM}}$.

We have another ‘secondary’ phenomenon around the 2-forms e^J and e_1^J (see [24]). Let $B = \frac{1}{2g} \omega_{(1)} \cdot \omega_{(1)}$ be the volume form in (5.3). On any pointed Riemann surface (C, P_0) there exists a function $h = h_{P_0} = -\hat{\Phi}(\delta_{P_0})$ with $d * dh = B - \delta_{P_0}$ and $\int_C hB = 0$. The function $G(P_0, P_1) := \exp(-4\pi h_{P_0}(P_1))$ is just the Arakelov–Green function. We regard G as a function on the fiber product $\mathbb{C}_g \times_{\mathbb{M}_g} \mathbb{C}_g$ and define the $(1, 1)$ -form e^A on \mathbb{C}_g by

$$e^A := \frac{1}{2\pi\sqrt{-1}} \partial \bar{\partial} \log G|_{\text{diagonal}} \in \Omega^{1,1}(\mathbb{C}_g),$$

representing the Chern class $e = c_1(T_{\mathbb{C}_g/\mathbb{M}_g})$. In fact, the normal bundle of the diagonal map $\mathbb{C}_g \rightarrow \mathbb{C}_g \times_{\mathbb{M}_g} \mathbb{C}_g$ is exactly the relative tangent bundle $T_{\mathbb{C}_g/\mathbb{M}_g}$.

Furthermore we introduce an explicit real-valued function a_g on \mathbb{M}_g by

$$a_g(C) := \int_C \omega_{(1)} \cdot \hat{\Phi}(\omega_{(1)} \wedge \omega_{(1)}) \cdot \omega_{(1)}, \tag{8.1}$$

where $\hat{\Phi}$ is the Green operator introduced in (5.4). By (5.2) we have

$$a_g(C) = - \sum_{i,j=1}^g \int_C \psi_i \wedge \bar{\psi}_j \hat{\Phi}(\bar{\psi}_i \wedge \psi_j). \quad (8.2)$$

We have $a_g(C) > 0$ if $g \geq 2$. Then comparing ∂a_g with η'_2 as explicit quadratic differentials, we obtain

$$e^A - e^J = \frac{-2\sqrt{-1}}{2g(2g+1)} \partial \bar{\partial} a_g. \quad (8.3)$$

On the other hand, the integral along the fiber

$$e_1^F := \int_{\text{fiber}} (e^J)^2 \in \Omega^{1,1}(\mathbb{M}_g)$$

also represents the first Morita–Mumford class e_1 . By straightforward computation on $\partial \bar{\partial} a_g$ we deduce

Theorem 8.2 ([24]).

$$e^A - e^J = \frac{-2\sqrt{-1}}{2g(2g+1)} \partial \bar{\partial} a_g = \frac{1}{(2-2g)^2} (e_1^F - e_1^J).$$

The function $a_g(C)$ is also a secondary object on the moduli space \mathbb{M}_g , and it defines a conformal invariant of the compact Riemann surface C , but the author does not know any of its further properties.

References

- [1] L. Ahlfors, Curvature properties of Teichmüller space. *J. d'Analyse Math.* 9 (1961), 161–176. [222](#)
- [2] E. Arbarello and M. Cornalba, Calculating cohomology groups of moduli spaces of curves via algebraic geometry. *Inst. Hautes Études Sci. Publ. Math.* 88 (1998), 97–127. [218](#)
- [3] E. Arbarello, C. De Concini, V. G. Kac, and C. Procesi, Moduli spaces of curves and representation theory. *Comm. Math. Phys.* 117 (1988), 1–36. [224](#)
- [4] T. Ashikaga and H. Endo, Various aspects of degenerate families of Riemann surfaces. *Sugaku Expositions* 19 (2006), 171–196. [218](#)
- [5] T. Ashikaga and K. Konno, Global and local properties of pencils of algebraic curves. In *Algebraic Geometry 2000, Azumino*, Adv. Stud. Pure. Math. 36, Math. Soc. Japan, Tokyo 2002, 1–49. [218](#)
- [6] M. Atiyah, The signature of fibre-bundles. In *Global Analysis*, University of Tokyo Press, Tokyo 1969, 73–84. [217](#)
- [7] M. Atiyah, Complex Analytic Connections in Fibre Bundles. *Trans. Amer. Math. Soc.* 85 (1957), 181–207. [221](#)

- [8] K. T. Chen, Iterated integrals of differential forms and loop space homology. *Ann. of Math.* 97 (1973), 217–246. [231](#)
- [9] C. J. Earle, Families of Riemann surfaces and Jacobi varieties. *Ann. of Math.* 107 (1978), 255–286. [229](#)
- [10] W. Goldman, The symplectic nature of fundamental groups of surfaces. *Adv. Math.* 54 (1984), 200–225. [223](#)
- [11] W. Goldman, Invariant functions on Lie groups and Hamiltonian flows of surface group representations. *Invent. Math.* 85 (1986), 263–302. [223](#)
- [12] J. Igusa, Arithmetic variety of moduli for genus 2. *Ann. of Math.* 72 (1960), 612–649. [218](#)
- [13] R. Hain and D. Reed, On the Arakelov geometry of moduli spaces of curves. *J. Differential Geom.* 67 (2004), 195–228. [233](#)
- [14] J. Harer, The second homology group of the mapping class group of an orientable surface. *Invent. Math.* 72 (1983), 221–239. [217](#), [218](#)
- [15] J. Harer, The cohomology of the moduli space of curves. In *Theory of moduli* (Montecatini Terme, 1985), Lecture Notes in Math. 1337, Springer-Verlag, Berlin 1988, 138–221. [217](#), [218](#)
- [16] J. Harer, Stability of the homology of the mapping class groups of orientable surfaces. *Ann. of Math.* 121 (1985), 215–249. [218](#)
- [17] B. Harris, Harmonic volumes. *Acta Math.* 150 (1983), 91–123. [230](#)
- [18] D. Johnson, An abelian quotient of the mapping class group \mathcal{I}_g . *Math. Ann.* 249 (1980), 225–242. [230](#)
- [19] N. Kawazumi, On the complex analytic Gel'fand-Fuks cohomology of Riemann surfaces. *Ann. Inst. Fourier.* 43 (1993), 655–712. [224](#)
- [20] N. Kawazumi, An infinitesimal approach to the stable cohomology of the moduli of Riemann surfaces. In *Topology and Teichmüller Spaces*, World Scientific, River Edge, NJ, 1996, 79–100. [222](#)
- [21] N. Kawazumi, A generalization of the Morita-Mumford classes to extended mapping class groups for surfaces. *Invent. Math.* 131 (1998), 137–149. [229](#)
- [22] N. Kawazumi, Cohomological aspects of Magnus expansions. Preprint, arXiv:math.GT/0505497. [230](#), [231](#)
- [23] N. Kawazumi, Harmonic Magnus Expansion on the Universal Family of Riemann Surfaces. Preprint, arXiv:math.GT/0603158. [230](#), [231](#), [232](#), [233](#)
- [24] N. Kawazumi, Johnson homomorphisms and the Arakelov-Green function. Preprint, arXiv:math.0801.4218. [233](#), [234](#)
- [25] N. Kawazumi and S. Morita, The primary approximation to the cohomology of the moduli space of curves and cocycles for the stable characteristic classes. *Math. Research Lett.* 3 (1996), 629–641. [229](#), [230](#)
- [26] S. Kobayashi and K. Nomizu, *Foundations of differential geometry, I*. Interscience, New York, London 1963. [221](#)
- [27] K. Kodaira, A certain type of irregular algebraic surfaces. *J. Anal. Math.* 19 (1967), 207–215. [217](#)

- [28] M. Korkmaz and A. I. Stipsicz, The second homology groups of mapping class groups of oriented surfaces. *Math. Proc. Cambridge Philos. Soc.* 134 (2003), 479–489. [218](#)
- [29] I. Madsen and M. Weiss, The stable moduli space of Riemann surfaces: Mumford’s conjecture. *Ann. of Math. (2)* 165 (2007), 843–941. [218](#)
- [30] W. Meyer, Die Signatur von lokalen Koeffizientensystemen und Faserbündeln. Dissertation, *Bonner Mathematische Schriften* 53 (1972). [218](#), [225](#)
- [31] W. Meyer, Die Signatur von Flächenbündeln. *Math. Ann.* 201 (1973), 239–264. [218](#)
- [32] E.Y. Miller, The homology of the mapping class group. *J. Differential Geom.* 24 (1986), 1–14. [218](#)
- [33] S. Morita, Characteristic classes of surface bundles. *Bull. Amer. Math. Soc.* 11 (1984), 386–388. [218](#), [223](#)
- [34] S. Morita, Characteristic classes of surface bundles. *Invent. Math.* 90 (1987), 551–577. [218](#)
- [35] S. Morita, Families of Jacobian manifolds and characteristic classes of surface bundles, I. *Ann. Inst. Fourier* 39 (1989), 777–810. [229](#)
- [36] S. Morita, Families of Jacobian manifolds and characteristic classes of surface bundles, II. *Math. Proc. Cambridge Philos. Soc.* 105 (1989), 79–101. [229](#)
- [37] S. Morita, Casson’s invariant for homology 3-spheres and characteristic classes of surface bundles, I. *Topology* 28 (1989), 305–323.
- [38] S. Morita, The extension of Johnson’s homomorphism from the Torelli group to the mapping class group. *Invent. Math.* 111 (1993), 197–224. [230](#)
- [39] S. Morita, A linear representation of the mapping class group of orientable surfaces and characteristic classes of surface bundles. In *Topology and Teichmüller Spaces*, World Scientific, River Edge, NJ, 1996, 159–186. [230](#)
- [40] S. Morita, Introduction to mapping class groups of surfaces and related groups. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 353–386. [229](#), [230](#)
- [41] D. Mumford, Abelian quotients of the Teichmüller modular group. *Journal d’Analyse Math.* 18 (1967), 227–244. [218](#), [233](#)
- [42] D. Mumford, Towards an enumerative geometry of the moduli space of curves. In *Arithmetic and Geometry*, Progr. Math. 36, Birkhäuser, Boston, Mass., 1983, 271–328. [218](#), [223](#)
- [43] R. C. Penner, Weil-Petersson volumes. *J. Differential Geom.* 35 (1992), 559–608. [223](#)
- [44] W. Pitsch, Un calcul élémentaire de $H^2(\mathcal{M}_{g,1}, \mathbb{Z})$ pour $g \geq 4$. *C. R. Acad. Sci. Paris Sér. I Math.* 329 (1999), 667–670. [218](#)
- [45] J. Powell, Two theorems on the mapping class group of a surface. *Proc. Amer. Math. Soc.* 68 (1978), 347–350. [218](#)
- [46] M. J. Pulte, The fundamental group of a Riemann surface: Mixed Hodge structures and algebraic cycles. *Duke Math. J.* 57 (1988), 721–760. [230](#)
- [47] H. E. Rauch, On the transcendental moduli of algebraic Riemann surfaces. *Proc. Nat. Acad. Sci. U.S.A.* 41 (1955), 42–49.

- [48] L. Takhtajan and P. Zograf, A local index theorem for families of $\bar{\partial}$ -operators on punctured Riemann surfaces and a new Kähler metric on their moduli spaces. *Comm. Math. Phys.* 137 (1991), 399–426. [232](#)
- [49] S. Wolpert, Chern forms and the Riemann tensor for the moduli space of curves. *Invent. Math.* 85 (1986), 119–145. [218](#), [223](#)
- [50] S. Wolpert, On the Weil-Petersson geometry of deformations of a hyperbolic surface. *Amer. J. Math.* 107 (1985), 969–997. [223](#)
- [51] P. Zograf and L. Takhtajan, On uniformization of Riemann surfaces and the Weil-Petersson metric on Teichmüller and Schottky spaces. *Math. USSR-Sb.* 60 (1988), 297–313. [224](#)
- [52] P. Zograf and L. Takhtajan, A local index theorem for families of $\bar{\partial}$ -operators on Riemann surfaces. *Russian Math. Surveys* 42 (6) (1987), 169–190. [232](#)

Part B

The group theory, 2

Quasi-homomorphisms on mapping class groups

Koji Fujiwara

Contents

1	Introduction	241
	1.1 Quasi-homomorphisms	242
	1.2 Stable commutator length	243
	1.3 Bounded cohomology	244
2	Brooks' counting quasi-homomorphism on free groups	245
3	Delta-hyperbolicity and quasi-homomorphism	246
	3.1 Word-hyperbolic groups	248
	3.2 Mapping class groups and curve complexes	251
	3.3 Rank-1 manifolds	254
4	Rigidity	254
5	Bounded generation	255
6	Separation by quasi-homomorphisms	257
7	Gaps in stable commutator length	259
	7.1 Word-hyperbolic groups	259
	7.2 Mapping class groups	260
8	Appendix. Bounded cohomology	261
	8.1 Riemannian geometry	262
	8.2 Group theory	265
	References	266

1 Introduction

We survey some results on quasi-homomorphism on mapping class groups from the viewpoint of hyperbolic geometry in the sense of Gromov. Most of the results in this chapter are shown both for word-hyperbolic groups and mapping class groups by the same techniques. The mapping class group, $\text{MCG}(S)$, of a compact orientable surface S is typically not word-hyperbolic, but it acts on its complex of curves $\mathcal{C}(S)$, which is δ -hyperbolic, [48]. The action is co-finite, but not proper (otherwise, the mapping class would be word-hyperbolic). Another aspect of the geometry of $\mathcal{C}(S)$ is that this space is not locally compact. Thanks to the study of $\mathcal{C}(S)$ by Masur–Minsky [48] regarding the geometry of $\mathcal{C}(S)$, we can apply the standard methods developed in the theory of word-hyperbolic groups to $\text{MCG}(S)$.

1.1 Quasi-homomorphisms

For proofs of the material in 1.1–1.3 see [3, §3] and also [15] for an updated account.

Definition 1.1 (Quasi-homomorphism). Let G be a group. A *quasi-homomorphism* is a function $f: G \rightarrow \mathbb{R}$ such that

$$D(f) = \sup_{a,b \in G} |f(a) + f(b) - f(ab)| < \infty.$$

$D(f)$ is called the *defect* of f . If a quasi-homomorphism satisfies $f(a^n) = nf(a)$ for all $a \in G$ and n , it is said *homogeneous*. We denote the vector space of all homogeneous quasi-homomorphisms on G by $\text{HQH}(G)$.

Quasi-homomorphisms are also called *quasimorphisms* (for example in [3], [17]). If f is a quasi-homomorphism on G , then one can obtain a homogeneous quasi-homomorphism \bar{f} as follows:

$$\bar{f}(a) = \lim_{n \rightarrow \infty} \frac{f(a^n)}{n}.$$

Note that the limit exists since the sequence $\{f(a^n)\}$ is additive with bounded error (cf. [59, Part One 99]). For any $a \in G$, $|f(a) - \bar{f}(a)| \leq D(f)$. Namely, a quasi-homomorphism f is (uniquely) written as the sum of a homogeneous quasi-homomorphism \bar{f} and a bounded function. The defect $D(\bar{f})$ is related to $D(f)$ by

$$D(\bar{f}) \leq 4D(f).$$

If f is a homogeneous quasi-homomorphism, then it is easy to check that for all $a, b \in G$, $f(aba^{-1}) = f(b)$, and therefore $|f([a, b])| \leq D(f)$. It turns out that there is an equality

$$\sup_{a,b \in G} |f([a, b])| = D(f).$$

The following result follows from a result on bounded cohomology (see Section 1.3).

Theorem 1.2 ([3, §3]). *Suppose that G is an amenable group. Then a homogeneous quasi-homomorphism on G is a homomorphism.*

Let $\mathcal{V}(G)$ be the vector space of all quasi-homomorphisms $G \rightarrow \mathbb{R}$. We denote by $\text{BDD}(G)$ and $\text{HOM}(G) = H^1(G; \mathbb{R})$ the subspaces of $\mathcal{V}(G)$ consisting of bounded functions and respectively homomorphisms. Note that $\text{BDD}(G) \cap \text{HOM}(G) = 0$. We will be concerned with the quotient spaces

$$\text{QH}(G) = \mathcal{V}(G)/\text{BDD}(G)$$

and

$$\widetilde{\text{QH}}(G) = \mathcal{V}(G)/(\text{BDD}(G) + \text{HOM}(G)) \cong \text{QH}(G)/H^1(G; \mathbb{R}).$$

Each element $f \in \mathcal{V}(G)$ defines $\bar{f} \in \text{HQH}(G)$. This implies that $\text{QH}(G) \cong \text{HQH}(G)$, therefore $\widetilde{\text{QH}}(G) \cong \text{HQH}(G)/H^1(G; \mathbb{R})$. Theorem 1.2 says $\widetilde{\text{QH}}(G)$ is trivial if G is amenable.

1.2 Stable commutator length

Let G be a group. Given $g \in [G, G]$, the *commutator length* of g , denoted by $\text{cl}(g)$, is the least number of commutators in G whose product is equal to g . Namely, $\min l = \text{cl}(g)$ such that $a_i, b_i \in G$ and

$$g = [a_1, b_1] \cdots [a_l, b_l].$$

The *stable commutator length*, denoted by $\text{scl}(g)$, is defined by

$$\text{scl}(g) = \lim_{n \rightarrow \infty} \frac{\text{cl}(g^n)}{n}.$$

The limit exists since the sequence $\{\text{cl}(g^n)\}$ is subadditive (cf. [59, Part One 98]). Note that cl and scl are class functions, namely, they are constant on each conjugacy class in G . The function scl is defined whenever some power of g is contained in $[G, G]$. By convention, we may extend scl to all of G by setting $\text{scl}(g) = \infty$ if no power of g is contained in $[G, G]$.

The following fact [3, §1.1] already appears in [52].

Proposition 1.3. *Let $f: G \rightarrow \mathbb{R}$ be a homogeneous quasi-homomorphism. If $f(a) = 1$ for $a \in [G, G]$ then $\frac{1}{2D(f)} \leq \text{scl}(a)$.*

Proof. Since $f(a) = 1$, f is not a homomorphism, therefore $D(f) > 0$. Denote $D(f)$ by D . For $n > 0$, put $l(n) = \text{cl}(a^n)$. The element a^n is a product of $l(n)$ commutators, c_i , in G . Since f is a quasi-homomorphism,

$$n = f(a^n) \leq |f(c_1)| + \cdots + |f(c_{l(n)})| + (l(n) - 1)D.$$

Since f is homogeneous, $|f(c_i)| \leq D$ for all i , therefore $n \leq (2l(n) - 1)D$. Thus, $\frac{1}{D} \leq \frac{2l(n)-1}{n}$ for all $n > 0$. Letting $n \rightarrow \infty$, we obtain $\frac{1}{2D} \leq \text{scl}(a)$. \square

Quasi-homomorphisms and stable commutator length are related by Bavard's Duality Theorem in a more precise way ([3, §3.6]):

Theorem 1.4 (Bavard's Duality Theorem). *Let G be a group and $a \in [G, G]$. If $\text{HQH}(G) = H^1(G; \mathbb{R})$ then $\text{scl}(a) = 0$. Otherwise, we have an equality*

$$\text{scl}(a) = \frac{1}{2} \sup_{\phi \in \text{HQH}(G) \setminus H^1(G; \mathbb{R})} \frac{|\phi(a)|}{D(\phi)}.$$

The argument is based on the Hahn–Banach theorem. In particular, the quasi-homomorphisms promised by Bavard's theorem are typically non constructive.

By Theorems 1.4 and 1.2, if G is amenable, then $\text{scl} = 0$ on $[G, G]$. On the other hand, if F is a free group of rank at least two, then for any $1 \neq g \in [F, F]$, $\text{scl}(g) \geq 1/6$ ([19, Corollary 3.3]). Recently, D. Calegari [16] proved that $\text{scl}(g) \in \mathbb{Q}$ for any $g \in [F, F]$. On the other hand, D. Zhuang [61] found a finitely presented group G such that $\text{scl}(g)$ is irrational (indeed, transcendental) for some $g \in [G, G]$.

A group G is called *perfect* if $G = [G, G]$ and *uniformly perfect* if G is perfect and cl is bounded on G , which implies that $\text{scl} = 0$. It is known that $\text{SL}_n(\mathbb{Z})$ is uniformly perfect if $n \geq 3$ (cf. [2]).

We discuss the stable commutator length in Section 7 in connection to hyperbolicity.

1.3 Bounded cohomology

To define the bounded cohomology group ([30]) of a discrete group G , let

$$C_b^k(G; \mathbb{R}) = \{f: G^k \rightarrow \mathbb{R} \mid f \text{ has bounded image}\}$$

The boundary $\delta: C_b^k(G; \mathbb{R}) \rightarrow C_b^{k+1}(G; \mathbb{R})$ is given by

$$\begin{aligned} \delta f(g_0, \dots, g_k) &= f(g_1, \dots, g_k) + \sum_{i=1}^k (-1)^i f(g_0, \dots, g_{i-1}g_i, \dots, g_k) \\ &\quad + (-1)^{k+1} f(g_0, \dots, g_{k-1}). \end{aligned}$$

The cohomology of the complex $\{C_b^k(G; \mathbb{R}), \delta\}$ is the *bounded cohomology group* of G , denoted by $H_b^*(G; \mathbb{R})$. See [30], [40], [54] as general references for the theory of bounded cohomology. $H_b^1(G; \mathbb{R})$ is trivial for any group G , and $H_b^n(G; \mathbb{R})$ is trivial for all $n \geq 1$ if G is amenable.

By definition, for each n , there is a natural homomorphism, sometimes called *comparison map*, $H_b^n(G; \mathbb{R}) \rightarrow H^n(G; \mathbb{R})$ induced by the inclusion $C_b^n(G; \mathbb{R}) \rightarrow C^n(G; \mathbb{R})$. An element $f \in \text{QH}(G)$ defines a bounded class $[\delta f] \in H_b^2(G; \mathbb{R})$. There is an exact sequence ([3])

$$0 \rightarrow H^1(G; \mathbb{R}) \rightarrow \text{QH}(G) \rightarrow H_b^2(G; \mathbb{R}) \rightarrow H^2(G; \mathbb{R}).$$

Since $\widetilde{\text{QH}}(G)$ is the quotient $\text{QH}(G)/H^1(G; \mathbb{R})$, we see that $\widetilde{\text{QH}}(G)$ can also be identified with the kernel of $H_b^2(G; \mathbb{R}) \rightarrow H^2(G; \mathbb{R})$. It follows from Theorem 1.4 that the kernel is trivial if G is uniformly perfect, [49].

If $G \rightarrow G'$ is an epimorphism then the induced maps $\text{QH}(G') \rightarrow \text{QH}(G)$ and $\widetilde{\text{QH}}(G') \rightarrow \widetilde{\text{QH}}(G)$ are injective.

Calculations of $\widetilde{\text{QH}}(G)$ have been made for many groups G . In many cases $\widetilde{\text{QH}}(G)$ is either 0 or infinite dimensional. But, a group G such that $\widetilde{\text{QH}}(G)$ is nontrivial and finite dimensional has been constructed (Appendix in [46]), using a group H such that $H_b^2(H; \mathbb{R})$ is nontrivial and finite dimensional (see [13, Remark 25] for such H).

If G is finitely generated by k elements, then $H^1(G; \mathbb{R})$ is at most k -dimensional, therefore $\widetilde{QH}(G)$ is infinite dimensional if $QH(G)$ is infinite dimensional (cf. Theorem 5.1).

As we said, if G is amenable then $H_b^2(G; \mathbb{R}) = 0$ ([30]), therefore the kernel of $H_b^2(G; \mathbb{R}) \rightarrow H^2(G; \mathbb{R})$ is trivial. In other words, $HQH(G) = \widetilde{QH}(G) = 0$. This is indeed how Theorem 1.2 is shown in [3]. $\widetilde{QH}(G)$ also vanishes when G is an irreducible lattice in a semisimple Lie group of real rank > 1 ([12], see Theorem 4.1).

2 Brooks' counting quasi-homomorphism on free groups

Our first example of a group G such that $\widetilde{QH}(G)$ is non trivial is a free group.

Theorem 2.1 ([11]). *Suppose F is a free group of rank at least two. Then $\widetilde{QH}(F)$ is an infinite dimensional vector space over \mathbb{R} .*

We explain Brooks' construction of a quasi-homomorphism f on F which is non-trivial in $\widetilde{QH}(F)$. For simplicity suppose the rank of F is two and let x, y be free generators of F . Fix a reduced word w on x, y . Any element $1 \neq a \in F$ is uniquely written as a (non-empty) reduced word on x, y , which we also denote by a . Define $|a|_w$ to be the maximal number of times that w can be seen as an (oriented) subword of a without overlapping. Define $|1|_w = 0$.

Example 2.2. $|xyxyx|_{xy} = 2; |xyxyx|_{xyx} = 1; |xxyxy|_{yx} = 1$.

Let w^{-1} be the reduced word which is the inverse of w as a group element. Define a function on F by $h_w(a) = |a|_w - |a|_{w^{-1}}$. The following says that h_w is a quasi-homomorphism.

Lemma 2.3. $D(h_w) \leq 3$.

To see this, let $a, b \in F$. We think of them as reduced words too. Let $a \cdot b$ be the word which we obtain by placing the word b after a . This word represents the group element ab , but may be not reduced. If the word is reduced, we see

$$||a \cdot b|_w - |a|_w - |b|_w| \leq 1, \quad ||a \cdot b|_{w^{-1}} - |a|_{w^{-1}} - |b|_{w^{-1}}| \leq 1.$$

Therefore $|h_w(ab) - h_w(a) - h_w(b)| \leq 2$. In general, $a \cdot b$ is not reduced, and each function $|\cdot|_w, |\cdot|_{w^{-1}}$ on F is not a quasi-homomorphism. One verifies that h_w is a quasi-homomorphism by writing $a = a' \cdot c, b = c^{-1} \cdot b'$ such that a', b', c and $a' \cdot b'$ are reduced, therefore the above inequalities apply to $a = a' \cdot c$ and $b = c^{-1} \cdot b'$ (use $|c|_w = |c^{-1}|_{w^{-1}}$). From this, one easily gets $D(h_w) \leq 6$ and indeed $D(h_w) \leq 3$ as Figure 1 shows. Suppose w is cyclically reduced, namely, $w^n (n > 0)$ is reduced. Then, $|w^n|_w = n$ for all $n > 0$. On the other hand, $|w^n|_{w^{-1}} = 0$, therefore

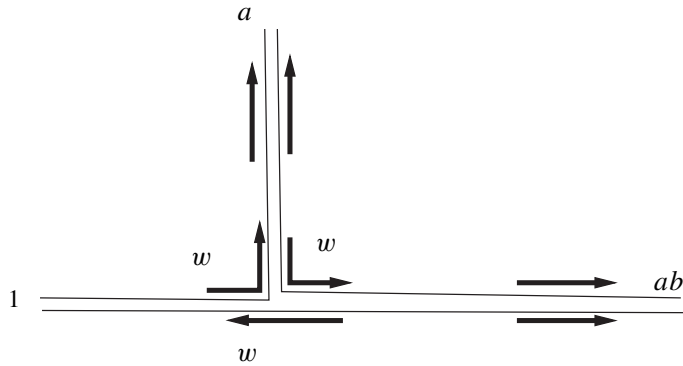


Figure 1. At most three subwords w count for $\delta h_w(a, b)$. The other pairs of subwords w cancel for δh_w (not necessarily for δc_w and $\delta c_{w^{-1}}$).

$h_w(w^n) = n$ for all $n > 0$. We find that h_w is non-trivial in $\text{QH}(F)$. If we take w to represent an element in $[F, F]$, then we obtain from h_w a non-trivial element in $\widetilde{\text{QH}}(F)$. For example, one can take $w = xyx^{-1}y^{-1}$. Moreover we can find a sequence of reduced and cyclically reduced words w_i such that h_{w_i} are linearly independent in $\widetilde{\text{QH}}(F)$. This proves Theorem 2.1.

We remark that Lemma 2.3 and Proposition 1.3 give a uniform positive lower bound of $\text{scl}(a)$ for any $1 \neq a \in [F, F]$. To see this, since scl is invariant by taking conjugates, one may assume that a , as a reduced word, is shortest among its conjugates. Then the word a is cyclically reduced, therefore $D(h_a) \leq 3$. Then $D(\bar{h}_a) \leq 12$ and $\bar{h}_a(a) = 1$. By Proposition 1.3, $\text{scl}(a) \geq \frac{1}{24}$. As we said, Culler [19] showed that $\text{scl}(a) \geq \frac{1}{6}$.

3 Delta-hyperbolicity and quasi-homomorphism

The construction of quasi-homomorphisms by Brooks has been generalized to the δ -hyperbolic setting. δ -hyperbolic geometry, or the hyperbolic geometry in the sense of Gromov, was invented by Gromov [29]. We only give a few basic definitions and facts. See for example [10].

Definition 3.1 (δ -hyperbolic space, δ -thin, word-hyperbolic group). Let X be a geodesic metric space and $\delta \geq 0$. We say that X is δ -hyperbolic if for any points a, b, c of X , and any geodesic segments $[a, b]$, $[b, c]$ and $[c, a]$, the segment $[a, b]$ is contained in the δ -neighborhood of the union of $[b, c]$ and $[c, a]$ (then the geodesic triangle $[a, b] \cup [b, c] \cup [c, a]$ is said δ -thin).

Let G be a finitely generated group with a fixed set of generators, and let Γ be its Cayley graph. We say G is word-hyperbolic if Γ is δ -hyperbolic for some δ .

Note that a geodesic between two points a, b is not unique, but we denote it by $[a, b]$. If X is δ -hyperbolic, then the Hausdorff distance of any two geodesics between a, b is at most δ .

If a geodesic space X is quasi-isometric (cf. [10]) to a geodesic space which is δ -hyperbolic, then there exists $\delta' \geq 0$ such that X is δ' -hyperbolic. As a consequence, the word-hyperbolicity of a finitely generated group, G , does not depend on the choice of a set of generators since the Cayley graphs of G for two sets of generators are quasi-isometric to each other.

Clearly, finite groups and \mathbb{Z} are word-hyperbolic. If G contains an infinite cyclic subgroup of finite index, then G is quasi-isometric to \mathbb{Z} (to be precise, the Cayley graphs of those two groups are quasi-isometric to each other), therefore, G is word-hyperbolic. A word-hyperbolic group which contains a cyclic subgroup of finite index is called an *elementary* word-hyperbolic group.

Definition 3.2 (Quasi-geodesic). Let X be a geodesic space. Let I be an interval of \mathbb{R} (bounded or unbounded). A (K, ε) -quasi-geodesic in X is a map $\alpha: I \rightarrow X$ such that for all $t, s \in I$

$$\frac{|t-s|}{K} - \varepsilon \leq d(\alpha(t), \alpha(s)) \leq K|t-s| + \varepsilon.$$

We may denote the image of α by α .

The following fact, sometimes called Morse Lemma, is important. (Cf. [10, III.H. Theorem 1.7].)

Proposition 3.3 (Stability of quasi-geodesics). *For all $\delta \geq 0, \varepsilon \geq 0, K \geq 1$ there exists $L(\delta, K, \varepsilon)$ with the following property: If X is a δ -hyperbolic space, α is a (K, ε) -quasi-geodesic in X and $[a, b]$ is a geodesic segment joining the endpoints of α , then the Hausdorff distance between $[a, b]$ and the image of α is at most L .*

Definition 3.4 (Hyperbolic isometry). Let X be a δ -hyperbolic space. An isometry a of X is called *hyperbolic* if there exist $x \in X$ and a constant $C > 0$ such that

$$d(x, a^n(x)) \geq Cn$$

for all $n \geq 1$.

Definition 3.5 (Translation length). If a is an isometry of a metric space X , the *translation length* of a , $\tau(a)$, is defined as follows. Let $x \in X$ be a point in X . Then,

$$\tau(a) = \liminf_{n \rightarrow \infty} \frac{d(x, a^n(x))}{n}.$$

The number $\tau(a)$ does not depend on the choice of x .

A finitely generated group G acts on a Cayley graph of G by isometries. It is an important fact that if G is word-hyperbolic, then each element $a \in G$ of infinite

order acts as a hyperbolic isometry, [29]. Therefore, a has infinite order if and only if $\tau(a) > 0$ on the Cayley graph.

If a is a hyperbolic isometry, then there exists a quasi-geodesic α in X with $\alpha = a(\alpha)$. The quasi-geodesic α is called a quasi-geodesic *axis* of a . It is not always true that α can be taken to be a geodesic. It is known that if G is word-hyperbolic and Γ is a Cayley graph, then there exists a constant P such that for any element $a \in G$ of infinite order, there exists a geodesic α such that $a^P(\alpha) = \alpha$. (For an argument, see for example [20]).

3.1 Word-hyperbolic groups

The following classification of subgroups in a word-hyperbolic group is a standard fact. We may regard it as a Tits alternative.

Theorem 3.6 (cf. [10]). *Let H be a subgroup of a word-hyperbolic group G . Then one of the following holds.*

- (1) H contains a free group of rank two.
- (2) H contains a cyclic group as a subgroup of finite index.

A subgroup H of the second type in Theorem 3.6 is called *elementary*. In other words, H is elementary if it is finite, or if it contains \mathbb{Z} as a subgroup of finite index. Note that a subgroup of a word-hyperbolic group is not necessarily word-hyperbolic. N. Brady constructed an example of a word-hyperbolic group which contains a finitely presented non-word-hyperbolic subgroup.

The following theorem is a generalization of Theorem 2.1 since a free group of rank at least two is a non-elementary word-hyperbolic group.

Theorem 3.7 ([20]). *Let G be a non-elementary word-hyperbolic group. Then $\widetilde{\text{QH}}(G)$ is infinite dimensional.*

Remark 3.8. The argument in [20] shows that if H is a non-elementary subgroup of a word-hyperbolic group, then $\widetilde{\text{QH}}(H)$ is infinite dimensional.

The argument for Theorem 3.7 is based on a generalization of the construction of quasi-homomorphisms, *counting functions*, by Brooks that we explain in Section 2. We outline the argument. See [20], [25] or [6] for more details.

Suppose G is a group with a fixed symmetric generating set S , and $\Gamma = \Gamma_S(G)$ is its Cayley graph. Let w be a (reduced) word in the generating set. Let α be a (directed) path in Γ , and $|\alpha|$ its length. Define $|\alpha|_w$ to be the maximal number of times that w can be seen as an (oriented) subword of α without overlapping (see Example 2.2 and Figure 2). An (oriented) path labeled by w is called a *copy* of w .

If we see α as a word on S , then it represents an element in G which we denote by $\bar{\alpha}$. We can uniquely identify α and the path in Γ from 1 to $\bar{\alpha}$ with the label by α .

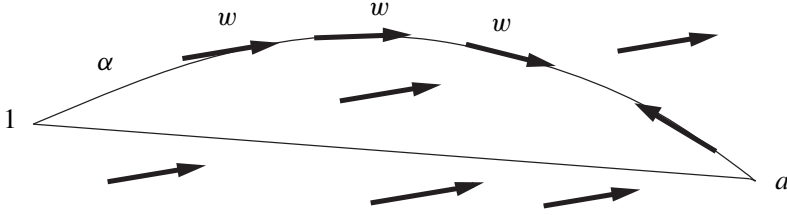


Figure 2. $|\alpha|_w = 3$.

In general, for an element $a \in G$, there is more than one geodesic, therefore reduced, path α in Γ from 1 to a . It is natural to define $|\alpha|_w = \max |\alpha|_w$ such that α runs through all geodesics with $\bar{\alpha} = a$, but indeed we need to modify the definition to have something similar to Lemma 2.3.

Let $0 < W < |w|$ be a constant. For $x, y \in \Gamma$, define

$$c_{w,W}(x, y) = d(x, y) - \inf_{\alpha} (|\alpha| - W|\alpha|_w),$$

where α ranges over all the paths from x to y . If the infimum is attained by α , we say that α is a *realizing path* for $c_{w,W}$ from x to y . If γ is a geodesic from x to y , then we may also write $c_{w,W}(x, y)$ as $c_{w,W}(\gamma)$.

Fix a point $x \in \Gamma$. (We may take $x = 1$.) Define for $a \in G$

$$c_{w,W}(a) = c_{w,W}(x, a(x)).$$

The function $c_{w,W}$ is called the *counting function* for the pair (w, W) . Let w^{-1} denote the inverse word of w . We define

$$h_{w,W} = c_{w,W} - c_{w^{-1},W}.$$

In [20], the normalization $W = 1$ is used. This is an appropriate choice of constant when $w^* := \dots w w w w \dots$ is a bi-infinite geodesic. Then w^* is a geodesic axis for w . In spirit, $h_{w,1}$ is same as h_w which is defined in Section 2 for free groups.

The following fact is not so difficult to prove. This does not require that Γ is δ -hyperbolic.

Proposition 3.9 (cf. Lemma 3.3, [25], and Proposition 3.9, [20]). *If α is a realizing path for $c_{w,W}$, then it is a (K, ε) -quasigeodesic, where*

$$K = \frac{|w|}{|w| - W}, \quad \varepsilon = \frac{2W|w|}{|w| - W}.$$

Since Γ is δ -hyperbolic, Proposition 3.3 applies to realizing paths. Let $L = L(\delta, \frac{|w|}{|w|-W}, \frac{2W|w|}{|w|-W})$. Let γ be a geodesic from x to y . From Proposition 3.9 we deduce that a realizing path from x to y must be contained in the L -neighborhood

of γ . Consequently, if the L -neighborhood of γ does not contain a copy of w , then $c_{w,W}(x, y) = 0$.

Suppose the L -neighborhood of w^n does not contain a copy of w^{-1} (see Figure 3). Here we are thinking of the L -neighborhood of w^n , for large n , like a long narrow tube whose core has a definite orientation, agreeing with the orientation on w . By “a copy of w^{-1} ”, we mean a copy of w whose orientation *disagrees* with that of the core of the tube. We will find a necessary and sufficient algebraic condition for \bar{w} to satisfy regarding this combinatorial/geometric property (see Condition 6.2, cf. Example 3.11).

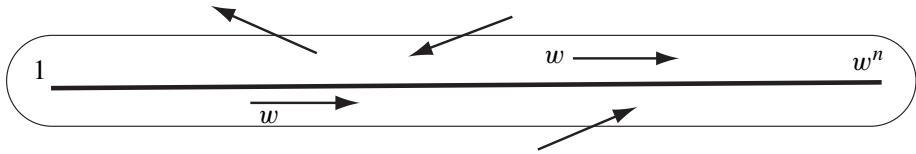


Figure 3. Copies of w with the opposite direction do not fit in the L -neighborhood of a geodesic from 1 to w^n .

It follows that $c_{w^{-1},W}(w^n) = 0$ because for a realizing path α for $c_{w^{-1},W}$ at w^n we must have $|\alpha|_{w^{-1}} = 0$. We thus obtain for all $n > 0$ an inequality $h_{w,W}(w^n) \geq nW$.

Consider a triangle of realizing paths. We have observed that it is L -close to a geodesic triangle, which is δ -thin. Therefore the triangle of realizing paths is $(\delta + 2L)$ -thin. The following inequality on the defect then follows. This is an analogue of Lemma 2.3. The argument is same in spirit.

Proposition 3.10 (cf. Proposition 3.10, [25], Proposition 2.13, [20]).

$$D(h_{w,W}) \leq 12L + 6W + 48\delta.$$

Note that the defect only depends on $|w|$, W and δ . If we take $W = 1$, then L depends only on δ if $|w| \geq 2$. In particular, the upper bound in Proposition 3.10 depends only on δ .

Although h_w is unbounded if w is cyclically reduced in Section 2, $h_{w,W}$ may be bounded.

Example 3.11. Let $G = \langle a, b \mid a^2 = b^2 = 1 \rangle \cong \mathbb{Z}_2 * \mathbb{Z}_2$. The group G is an elementary word-hyperbolic group. Since G is generated by torsion elements a, b , there is no non-trivial homomorphism. It follows that any quasi-homomorphism is bounded (use Theorem 1.2. G is amenable).

Indeed, this conclusion can be thought of as a consequence of an algebraic property. Let h be a homogeneous quasi-homomorphism. To see that $h(w) = 0$ for all w , we may assume that w is either a, b or $(ab)^n$ since w is conjugate to one of those. We have $h(a) = h(b) = 0$ since $a = a^{-1}, b = b^{-1}$. Since ab is conjugate to $ba = (ab)^{-1}$ by a , $h(ab) = 0$. What is essential in this argument is the algebraic property that $(ab)^n$

is conjugate to $(ab)^{-n}$. We will state this as an axiom in Condition 6.2. This property can be thought of as a dynamical property concerning the action of G on its Cayley graph. Namely, the points $(ab)^n$ are on a geodesic axis α for the action of ab , which is flipped by a to α with the opposite direction.

The following result (cf. [20], and [6] for WPD-actions) guarantees that there are many choices w such that $h_{w,1}$ are unbounded quasi-homomorphisms. We already know that G contains a (quasi-convex) free group F of rank two by Theorem 3.6. Proposition 3.12 says that one can take F to satisfy an additional dynamical property (no flip of an axis), which is explained in Example 3.11. This property is critical to show (2). For the counting functions $c_{w,1}, c_{w^{-1},1}$ for $1 \neq w \in F$ to make sense, we take a geodesic path/word from 1 to w , which we also denote w . For the definition of quasi-convexity, see [29], [10].

Proposition 3.12. *Let G be a non-elementary word-hyperbolic group. Then there exist a quasi-convex subgroup $F < [G, G]$ which is isomorphic to a rank-two free group and a constant D such that for each non-trivial element $w \in F$ we have the following:*

- (1) $c_{w,1}(w^n) \geq n/2$ for all $n > 0$.
- (2) $c_{w^{-1},1}(w^n) = 0$ for all $n > 0$.
- (3) $D(h_{w,1}) \leq D$, where $h_{w,1} = c_{w,1} - c_{w^{-1},1}$.

In particular, $h_{w,1}$ is an unbounded quasi-homomorphism. Moreover, one can show (see [20]) that there is a sequence of elements $w_i \in F$ such that the corresponding quasi-homomorphisms h_i are linearly independent in $\text{QH}(G)$. This proves Theorem 3.7. Since $w_i \in [G, G]$, it follows that $\bar{h}_i \in \text{HQH}(G)$ is not a homomorphism.

3.2 Mapping class groups and curve complexes

We apply the construction of quasi-homomorphisms in Section 3.1 to mapping class groups.

Let S be a compact orientable surface of genus g and p punctures. The *mapping class group* of S , $\text{MCG}(S)$, is the group of isotopy classes of orientation-preserving homeomorphisms $S \rightarrow S$. This group acts on the *curve complex* $\mathcal{C}(S)$ of S defined by Harvey [37] and successfully used in the study of mapping class groups by Harer [36], [35]. For our purposes, we will restrict to the 1-skeleton of Harvey's complex, so that $\mathcal{C}(S)$ is a graph whose vertices are isotopy classes of essential, nonparallel, nonperipheral, simple closed curves in S and two distinct vertices are joined by an edge if they can be realized simultaneously by pairwise disjoint curves. If a non-empty (finite) collection of vertices are realized simultaneously by pairwise disjoint curves, we call it a *curve system* (or *multi-curve*). (The actual curve complex of S is the flag

complex made from $\mathcal{C}(S)$, and it is quasi-isometric to $\mathcal{C}(S)$. A curve system defines a simplex in the curve complex.)

In certain *sporadic* cases $\mathcal{C}(S)$ as defined above is 0-dimensional or empty. This happens when there are no curve systems consisting of two curves, i.e. when $g = 0$, $p \leq 4$ and when $g = 1$, $p \leq 1$. One could rectify the situation by declaring that two vertices are joined by an edge if the corresponding curves can be realized with only one intersection point in the case $g = 1$, $p \leq 1$ and with two intersection points in the case $g = 0$, $p = 4$.

The mapping class group $\text{MCG}(S)$ acts on $\mathcal{C}(S)$ by $a \cdot [c] = [a(c)]$, where $a \in \text{MCG}(S)$ and $[c]$ is the isotopy class of a simple closed curve c on S . A classification of each element a in $\text{MCG}(S)$ is known (cf. [38, Section 7.1]):

- (1) a has finite order.
- (2) There exists a curve system M on S such that the simplex that M defines is invariant by a (maybe its vertices are permuted). Then a is called *reducible*.
- (3) a is not reducible and has infinite order. a is called *pseudo-Anosov*.

Two pseudo-Anosov elements a, b are called *independent* if the subgroup generated by a, b does not contain \mathbb{Z} as a subgroup of finite index.

H. Masur and Y. Minsky proved the following remarkable result.

Theorem 3.13 ([48]). *Let S be a nonsporadic surface. The curve complex $\mathcal{C}(S)$ is δ -hyperbolic. An element of $\text{MCG}(S)$ acts hyperbolically on $\mathcal{C}(S)$ if and only if it is pseudo-Anosov.*

An alternative proof of the δ -hyperbolicity of the curve complex is given in Chapter 10 of Volume I of this Handbook [34].

It follows that $a \in \text{MCG}(S)$ has positive (indeed, uniformly positive by [9]) translation length on $\mathcal{C}(S)$ (Definition 3.5) if and only if a is pseudo-Anosov.

Remark 3.14. Theorem 3.13 is generalized to a non-orientable surface [7]. When a surface S is non-orientable, we consider the group of isotopy classes of all homeomorphisms $S \rightarrow S$. This group is called the *extended mapping class group* of S . When S is orientable, the extended mapping class group contains $\text{MCG}(S)$ as a subgroup of index two.

The action of $\text{MCG}(S)$ on $\mathcal{C}(S)$ is not proper. We introduce the following notion.

Definition 3.15 (WPD). We say that the action of G on a δ -hyperbolic space X satisfies *WPD* (weak proper discontinuity) if

- G contains at least one element that acts on X as a hyperbolic isometry, and
- fix $x \in X$. For every hyperbolic element $g \in G$ and for every $C > 0$, there exists $N > 0$ such that the set

$$\{\gamma \in G \mid d(x, \gamma(x)) \leq C, d(g^N(x), \gamma g^N(x)) \leq C\}$$

is finite. (Note that this does not depend on the choice of x .)

Proposition 3.16 ([6]). *Let S be a nonsporadic surface. The action of $\text{MCG}(S)$ on the curve complex $\mathcal{C}(S)$ satisfies WPD.*

The following is a generalization of Theorem 3.7, which is the case when the action of G on X is proper (and co-compact). As we point out in Remark 3.8, that the action is co-compact is not important.

A group is said to *virtually* have some property if some subgroup of finite index in the group has this property.

Theorem 3.17 ([6]). *Let X be a δ -hyperbolic space and suppose G acts on X by isometry and WPD. If G contains a hyperbolic isometry and is not virtually \mathbb{Z} , then $\widehat{\text{QH}}(G)$ is infinite dimensional.*

The argument for Theorem 3.17 is similar to the one for Theorem 3.7. To construct counting functions on G using its action on X , we modify the definition of counting functions (Section 3.1) as follows. Let w be a path in X and call $a(w)$ for $a \in G$ a copy of w . For a path α in Γ , define $|\alpha|_w$ to be the maximal number of disjoint oriented copies of w which can be obtained as subpaths of α . All other definitions are the same as before. To find many elements w which give unbounded quasi-homomorphisms, we prove something similar to Proposition 3.12. This is where WPD is essentially used, in particular to verify (2).

By Theorem 3.13 and Proposition 3.16, we can apply Theorem 3.17 to the action of $\text{MCG}(S)$ on $\mathcal{C}(S)$. We obtain the following. This settles Morita's conjectures 6.19 and 6.21 [57] in the affirmative.

Theorem 3.18 ([6]). *Let S be a compact orientable surface. Suppose $G < \text{MCG}(S)$ is a subgroup. If G is not virtually abelian, then $\widehat{\text{QH}}(G)$ is infinite dimensional.*

In the argument for Theorem 3.18, we use the following classification of subgroups of a mapping class group (see [51]).

Theorem 3.19. *Let G be an infinite subgroup of the mapping class group of an orientable surface S . Then one of the following holds:*

- (1) *G contains two pseudo-Anosov elements which are independent (and G is called sufficiently large). Then G contains a free group of rank two.*
- (2) *G contains \mathbb{Z} as a subgroup of finite index.*
- (3) *G fixes a multi-curve on S (and G is called reducible).*

From this classification, a Tits alternative follows (cf. Theorem 3.6), namely, either G contains a free group of rank two, or else G contains a free abelian group of finite rank as a subgroup of finite index ([50], [39]).

3.3 Rank-1 manifolds

Let M be a complete Riemannian manifold of non-positive sectional curvature of finite volume, and $G = \pi_1(M)$. We briefly discuss $\widetilde{QH}(G)$ in this section. Suppose $\dim M \geq 2$. Assume that G is *irreducible*, namely, it does not contain a subgroup H of finite index such that H is a product of two infinite groups. If M is a locally symmetric space, namely the universal cover \tilde{M} is a symmetric space (cf. [10]), then $\widetilde{QH}(G)$ is trivial if the rank of M is at least two (Theorem 4.1), or $\widetilde{QH}(G)$ is infinite dimensional if the rank is one (see the proof of Theorem 5.4, cf. Theorem 3.17).

Indeed the converse of Theorem 4.1 is true. In other words, $\widetilde{QH}(\Gamma) = 0$ characterizes locally symmetric spaces of rank at least two.

Theorem 3.20 ([8]). *Let M be a complete Riemannian manifold of nonpositive curvature and finite volume. Assume that $\Gamma = \pi_1(M)$ is finitely generated and does not contain a subgroup of finite index which is cyclic or a Cartesian product of two infinite groups. Then the universal cover \tilde{M} is a symmetric space of rank at least two if and only if $\widetilde{QH}(\Gamma) = 0$. Otherwise, $\widetilde{QH}(\Gamma)$ is infinite-dimensional.*

The proof uses the celebrated Rank Rigidity Theorem ([1]), as well as a construction of quasi-homomorphisms on groups that act on CAT(0) spaces and contain rank-1 elements, which can be thought of as a generalization of Theorem 3.17. (See [10], [1] for the definitions of CAT(0) spaces and rank-1 elements.) In connection to Theorem 4.1, we remark that a symmetric space of non-compact type is CAT(0), and if it has rank at least two then any hyperbolic isometry of the space is not rank-1.

4 Rigidity

We discuss a version of superrigidity for mapping class groups. Theorem 4.2 was conjectured by N. V. Ivanov and proved by Kaimanovich and Masur [41] using random walks in the case when the image group contains independent pseudo-Anosov elements and it was extended to the general case by Farb and Masur [23] using the classification of subgroups of $MCG(S)$ (see Section 3.2). We give an argument based on the work of M. Burger and N. Monod [12] on bounded cohomology of lattices.

Theorem 4.1 ([12], [13]). *Let Γ be an irreducible lattice in a connected semi-simple Lie group G with no compact factors, with finite center, and of rank > 1 . Then the kernel of $H_b^2(\Gamma; \mathbb{R}) \rightarrow H^2(\Gamma; \mathbb{R})$ is trivial.*

They indeed show that $\widetilde{QH}(\Gamma)$ is trivial. Their approach is out of the range of this chapter. It was known that $H^1(\Gamma; \mathbb{R})$ is trivial by Matsushima and others.

Theorem 4.2. *Let Γ be an irreducible lattice in a connected semi-simple Lie group G with no compact factors, with finite center, and of rank > 1 . Then every homomorphism $\Gamma \rightarrow \text{MCG}(S)$ has finite image.*

Proof. Let $\phi: \Gamma \rightarrow \text{MCG}(S)$ be a homomorphism. By the Margulis–Kazhdan theorem [62, Theorem 8.1.2] either the image of ϕ is finite or the kernel of ϕ is contained in the center. When Γ is a nonuniform lattice, the proof is easier and was known to Ivanov before the work of Kaimanovich–Masur (see Ivanov’s comments to Problem 2.15 on Kirby’s list “Problems in low-dimensional topology”). Since the rank is ≥ 2 the lattice Γ then contains a solvable subgroup N which does not become abelian after quotienting out a finite normal subgroup. If the kernel is finite, then $\phi(N)$ is a solvable subgroup of $\text{MCG}(S)$ which is not virtually abelian, contradicting [50] (see the classification of subgroups in mapping class groups in Section 3.2).

Now assume that Γ is a uniform lattice. If the kernel $\text{Ker}(\phi)$ is finite then there is an unbounded quasi-homomorphism $q: \text{Im}(\phi) \rightarrow \mathbb{R}$ by Theorem 3.18. But then $q\phi: \Gamma \rightarrow \mathbb{R}$ is an unbounded quasi-homomorphism contradicting Theorem 4.1 that says that every quasi-homomorphism $\Gamma \rightarrow \mathbb{R}$ is bounded. \square

In connection to Theorem 4.1, we ask a question.

Question 4.3 ([12]). Let Γ be as in Theorem 4.1. Is there a constant C such that for all $a \in [\Gamma, \Gamma]$, $\text{cl}(a) \leq C$?

Note that $[\Gamma, \Gamma]$ has finite index in Γ since $H^1(\Gamma; \mathbb{R})$ is trivial. The answer is yes if Γ is $\text{SL}_n(\mathbb{Z})$ with $n \geq 3$ (see Section 1.2).

5 Bounded generation

A group G is said to be *boundedly generated* if there exist finitely many elements $g_1, \dots, g_k \in G$ such that for any $g \in G$ there exist $n_i \in \mathbb{Z}$ with

$$g = g_1^{n_1} \dots g_k^{n_k}.$$

One may say G is boundedly generated by g_1, \dots, g_k .

Kotschick related bounded generation of a group G and $\text{HQH}(G)$ as follows.

Theorem 5.1 (Proposition 5 [44]). *If G is boundedly generated by g_1, \dots, g_k then the dimension of $\text{HQH}(G)$ as a vector space is at most k .*

If G is generated by k elements, then the vector space of all homomorphisms from G to \mathbb{R} is at most k -dimensional. One may see this theorem as a generalization. Kotschick combined this result and Theorem 3.18, and gave a new proof to the following theorem by Farb–Lubotzky–Minsky.

Theorem 5.2 ([24]). *The mapping class group MCG of a closed orientable surface S of genus at least one is not boundedly generated.*

In fact, since Theorem 3.18 applies to all subgroups in $MCG(S)$, a subgroup G in $MCG(S)$ is not boundedly generated if G is not virtually abelian (cf. [27]).

It is observed in [24] that a non-elementary word-hyperbolic group G is not boundedly generated. Their argument uses the deep result by Gromov [29] saying that such a group G has an infinite quotient which is a torsion group. Clearly, a boundedly generated group cannot have an infinite torsion quotient. By Theorems 5.1 and 3.7 (and Remark 3.8), we have the following ([27]).

Theorem 5.3. *A non-elementary subgroup in a word-hyperbolic group is not boundedly generated.*

It follows that a uniform lattice G in a simple Lie group of rank one is not boundedly generated since G is a non-elementary word-hyperbolic group. Margulis and Vinberg [47] showed that many discrete subgroups in a rank-1 simple Lie group are virtually mapped by homomorphisms to non-abelian free groups, so that they are not boundedly generated. In fact we have the following.

Theorem 5.4 ([27]). *Let G be a discrete subgroup in a rank-1 simple Lie group. If G does not contain a nilpotent subgroup of finite index then it is not boundedly generated.*

Proof. G acts on a rank-1 symmetric space, which is δ -hyperbolic. The action is proper. If G is not virtually nilpotent, then G contains a hyperbolic isometry (we use a classification of discrete subgroups in a rank-1 simple Lie group). Then a theorem from [25] (the theorem applies to proper G -actions on δ -hyperbolic spaces. Or one can use Theorem 3.18) says that $\widetilde{QH}(G)$ is infinite dimensional since G is not virtually cyclic. \square

Note that Theorem 5.4 gives a classification of virtually nilpotent subgroups among discrete subgroups in terms of bounded generation since the converse is true. It is not hard to check that a finitely generated nilpotent group is boundedly generated. It then follows that a finitely generated virtually nilpotent group is boundedly generated. If G as in the theorem is virtually nilpotent, then it is finitely generated, therefore, boundedly generated.

Non-uniform lattices in a Lie group of rank at least two are known to be boundedly generated (cf. [60]). For example, $SL(n, \mathbb{Z})$, $n > 2$ and $SL(2, \mathbb{Z}[1/p])$ such that p is a prime number are boundedly generated.

There is a more direct way to show Theorems 5.2, 5.3, 5.4 using quasi-homomorphisms. We discuss it in the next section (for example see Remark 6.7).

6 Separation by quasi-homomorphisms

Definition 6.1 (Separation, [58]). Let G be a group and $a \neq b \in G$. If there exists a homogeneous quasi-homomorphism f on G such that $f(a) = 1$ and $f(b) = 0$, then we say that a is *separated* from b (by f).

Let $B \subset G$ be a set of elements such that $a \notin B$. If there exists a homogeneous quasi-homomorphism f on G such that $f(a) = 1$ and $f(b) = 0$ for all $b \in B$, then we say that a is *separated* from B (by f).

The condition that a quasi-homomorphism f is homogeneous is necessary, otherwise, one can always separate a from b (by letting $f(a) = 1$ and $f(c) = 0$ for all $c \neq a$). On the other hand, as long as $f(a) \neq 0$, one can always normalize f such that $f(a) = 1$. The normalization $f(a) = 1$ becomes important when one tries to bound the defect $D(f)$ from above. See (the second part of) Theorem 7.3 and 7.4.

Our separation property has a similar flavor to the residual finiteness of a group. A group G is said to be *residually finite* if for any non-trivial element $a \in G$, there exists a finite group F and a homomorphism $f: G \rightarrow F$ such that $f(a)$ is non-trivial. Similarly, we may try to separate two elements by a homomorphism to \mathbb{Z} . But, for example, if $G \simeq \mathrm{SL}(2, \mathbb{Z})$, then any homomorphism $G \rightarrow \mathbb{R}$ is trivial since G is generated by two torsion elements. Therefore, it is impossible to separate two elements by a homomorphism to \mathbb{Z} . On the other hand, we know that $\widetilde{\mathrm{QH}}(G)$ is infinite dimensional (G is non-elementary word-hyperbolic. Apply Theorem 3.7).

Suppose that one can separate a from b by a homogeneous quasi-homomorphism f such that $f(a) = 1$, $f(b) = 0$. Then the elements a and b must satisfy the following condition since f is a class function.

Condition 6.2. (1) For all $n \neq m$ and $c \in G$, $a^n \neq ca^m c^{-1}$.
 (2) For all $n \neq 0, m$ and $c \in G$, $a^n \neq cb^m c^{-1}$.

Note that by Condition (1), a has infinite order. It is interesting to know if Condition 6.2 is sufficient to separate a from b by a homogeneous quasi-homomorphism. An affirmative answer is found by Polterovich and Rudnick [58] for $\mathrm{SL}(2, \mathbb{Z})$.

Theorem 6.3. *Suppose $a, b \in \mathrm{SL}(2, \mathbb{Z})$ satisfy Condition 6.2. Then there is a homogeneous quasi-homomorphism f such that $f(a) = 1$, $f(b) = 0$.*

Polterovich and Rudnick asked if one can generalize the theorem to word-hyperbolic groups.

Theorem 6.4 ([17], [20]). *Let G be a word-hyperbolic group. Suppose $a, b \in G$ satisfy Condition 6.2. Then there is a homogeneous quasi-homomorphism f on G such that $f(a) = 1$, $f(b) = 0$.*

Moreover, let $B \subset G$ be a finite collection of elements such that for a and each $b \in B$ Condition 6.2 holds. Then there is a homogeneous quasi-homomorphism f on G such that $f(a) = 1$ and for all $b \in B$, $f(b) = 0$.

We also have a separation theorem for mapping class groups.

Theorem 6.5 ([17], [6]). *Let S be a compact orientable surface and let $\text{MCG}(S)$ be its mapping class group. Suppose $a, b \in \text{MCG}(S)$ satisfy Condition 6.2 and a is a pseudo-Anosov element. Then there is a homogeneous quasi-homomorphism f on G such that $f(a) = 1$, $f(b) = 0$.*

Moreover, let $B \subset \text{MCG}(S)$ be a collection of elements such that Condition 6.2 holds for a and each $b \in B$. Suppose there exists T such that the translation length of each $b \in B$ on $\mathcal{C}(S)$ is at most T . Then there is a homogeneous quasi-homomorphism f on G such that $f(a) = 1$ and for all $b \in B$, $f(b) = 0$.

In fact, Theorem 6.4, 6.5 are part of Theorem 7.3, 7.4, in which we obtain upper bounds on the defect of f .

Note that it is free to assume that the set B contains all non-pseudo-Anosov elements. This is because if $c \in \text{MCG}(S)$ is not pseudo-Anosov, then the translation length of c on $\mathcal{C}(S)$ is zero as c has a bounded orbit. It follows from the construction that a homogeneous quasi-homomorphism f obtained in Theorem 6.5 satisfies $f(c) = 0$.

To explain the connection of separation and bounded generation, we need one definition.

Definition 6.6 (Product of subgroups). Let G be a group and $H_1, \dots, H_n < G$ subgroups. Then the *product* $H_1 \dots H_n$ is a subset of G defined as follows:

$$H_1 \dots H_n = \{h_1 \dots h_n \mid h_i \in H_i\}.$$

Remark 6.7. One can show Theorem 5.3 using Theorem 6.4 as follows. Let G be a non-elementary word-hyperbolic group. Suppose that elements $b_1, \dots, b_n \in G$ are given. Then one can find an element $a \in G$ such that a and each b_i satisfy Condition 6.2 (this is not trivial). By Theorem 6.4, there exists a homogeneous quasi-homomorphism f with $f(a) = 1$ and $f(b_i) = 0$ for all i . Then $|f|$ is bounded by $(n-1)D(f)$ on the following subset in G :

$$\langle b_1 \rangle \dots \langle b_n \rangle$$

Since f is unbounded on $\langle a \rangle$, we have $G \neq \langle b_1 \rangle \dots \langle b_n \rangle$. Therefore G is not boundedly generated by b_1, \dots, b_n .

Similarly, one can show that $\text{MCG}(S)$ is not boundedly generated using Theorem 6.5.

7 Gaps in stable commutator length

We discuss the image, or the spectrum, of the function scl on $[G, G]$.

7.1 Word-hyperbolic groups

D. Calegari [14] showed the following theorem.

Theorem 7.1. *For every dimension n and any $\varepsilon > 0$, there is a constant $\delta(\varepsilon, n) > 0$ such that if M is a complete hyperbolic n -manifold and $a \in \pi_1(M)$ has stable commutator length $\leq \delta(\varepsilon, n)$, then a is represented by a closed geodesic in M with length $\leq \varepsilon$.*

Since there are only finitely many closed geodesics of length at most ε in M , this theorem says that there is a gap (at zero) in the spectrum of stable commutator length. Calegari uses pleated surfaces in M to estimate stable commutator length from below. A similar argument appears in [29], where Gromov asserts that the hyperbolicity implies the positivity of scl . The existence of a gap at zero was found by Calegari.

Via Theorem 1.4, Theorem 7.1 is related to quasi-homomorphisms on $\pi_1(M)$. In some way, the following result [17] is a generalization to word-hyperbolic groups.

Theorem 7.2 (Gap Theorem in hyperbolic groups, weak version [17]). *Let G be a word-hyperbolic group whose Cayley graph is δ -hyperbolic with respect to a symmetric generating set S with $|S|$ generators. Then there is a constant $C(\delta, |S|) > 0$ such that for every $a \in G$, either $\text{scl}(a) \geq C$ or else there is some positive integer n and some $b \in G$ such that $ba^{-n}b^{-1} = a^n$.*

Note that $\text{scl}(a) = 0$ if the condition $ba^{-n}b^{-1} = a^n$ holds for $n > 0$ (cf. Condition 6.2 (1)). This condition is called *mirror condition* in [17]). It follows from this condition that b has finite order if a has infinite order. Therefore the condition never holds in the fundamental group of a hyperbolic manifold since there is no nontrivial torsion element (cf. Theorem 7.1).

Theorem 7.2 is a consequence of the first part of the following theorem by Proposition 1.3 (cf. Theorem 1.4) with $C = \frac{1}{2D}$. The second part of the theorem can be thought of a separation theorem (see Section 6).

Theorem 7.3 (Gap Theorem in hyperbolic groups, strong version [17]). *Let G be a word-hyperbolic group whose Cayley graph is δ -hyperbolic with respect to a symmetric generating set S with $|S|$ generators. Then there exists a constant $D(\delta, |S|)$ with the following property. Let $a \in G$ be a (non-torsion) element. Assume there is no $n > 0$ and no $b \in G$ with $ba^{-n}b^{-1} = a^n$. Then there is a homogeneous quasi-homomorphism h on G such that*

- (1) $h(a) = 1$,
- (2) the defect of h is $\leq D(\delta, |S|)$.

Moreover, let $a_i \in G$ be a collection of elements for which $T = \sup_i \tau(a_i)$ is finite. Suppose that for all integers $n \neq 0, m$ and all elements $b \in G$ and indices i , there is an inequality

$$ba^n b^{-1} \neq a_i^m.$$

Then there is a homogeneous quasi-homomorphism h on G such that

- (1) $h(a) = 1$, and $h(a_i) = 0$ for all i ,
- (2) the defect of h is $\leq D'(\delta, |S|, T, \tau(a))$.

Note that the translation length τ concerns the Cayley graph of G with respect to S . The argument for Theorems 7.2, 7.3 is a refinement of the one for Theorem 3.7. We construct a quasi-homomorphisms f by counting functions, and the issue is to bound the defect of f .

7.2 Mapping class groups

We show a theorem similar to Theorem 7.3 for mapping class groups. For $a \in \text{MCG}(S)$, $\tau(a)$ denotes the translation length of a on $\mathcal{C}(S)$.

Theorem 7.4 ([17]). *Let S be a compact orientable surface of hyperbolic type and $\text{MCG}(S)$ its mapping class group. Then there is a positive integer P depending on S such that for any pseudo-Anosov element a , either there is an $0 < n \leq P$ and an element $b \in \text{MCG}(S)$ with $ba^{-n}b^{-1} = a^n$, or else there exists a homogeneous quasi-homomorphism h on $\text{MCG}(S)$ such that $h(a) = 1$ and the defect of h is $\leq D(S)$, where $D(S)$ depends only on S .*

Moreover, let $a_i \in \text{MCG}(S)$ be a collection of elements for which $T = \sup_i \tau(a_i)$ is finite. Suppose that for all integers $n \neq 0, m$ and all elements $b \in \text{MCG}(S)$ and indices i , there is an inequality

$$ba^n b^{-1} \neq a_i^m$$

Then there is a homogeneous quasi-homomorphism h on $\text{MCG}(S)$ such that

- (1) $h(a) = 1$, and $h(a_i) = 0$ for all i ,
- (2) the defect of h is $\leq D'(S, T, \tau(a))$.

The construction of a quasi-homomorphism is the same as in Theorem 3.18, but to have the desired bound on the defect, we need extra ingredients. This extra part is more difficult than for word-hyperbolic groups since the action of $\text{MCG}(S)$ on $\mathcal{C}(S)$ is not proper, and $\mathcal{C}(S)$ is not locally finite. The standard argument which has been developed in the theory of word-hyperbolic groups does not apply immediately. To deal with this difficulty, we use the notion of *tight geodesics*, which is introduced by Masur–Minsky [48]. They show a certain local finiteness property in terms of tight geodesics. Bowditch [9] obtains more refined information than [48], which we use.

Theorem 7.5 ([9]). *Let S be a compact orientable surface and $\text{MCG}(S)$ its mapping class group. For $R > 0$, there exist $D(R)$, $K(R)$, which depends on S , such that for any two vertices $x, y \in \mathcal{C}(S)$ with $d(x, y) \geq D$, the following set contains at most K elements:*

$$\{a \in \text{MCG}(S) \mid d(x, a(x)) \leq R, d(y, a(y)) \leq R\}.$$

Proposition 3.16 also follows from Theorem 7.5.

Theorem 7.6 ([9]). *Let S be a compact orientable surface and $\text{MCG}(S)$ its mapping class group. Then there exists a constant $M = M(S) > 0$ such that for any pseudo-Anosov element $a \in \text{MCG}(S)$, there exists a geodesic $\alpha \subset \mathcal{C}(S)$ with $a^M(\alpha) = \alpha$.*

A similar result is known for word-hyperbolic groups in terms their action on their Cayley graphs (for example, see [20, Theorem 5.1]).

Combining the first part of Theorem 7.4 and Proposition 1.3, we obtain the following with $C(S) = \frac{1}{2D(S)}$.

Theorem 7.7 (Gap theorem [17]). *Let S be a compact orientable surface of hyperbolic type and $\text{MCG}(S)$ its mapping class group. Then there exists $C(S) > 0$ such that for any pseudo-Anosov element $a \in \text{MCG}(S)$, either there is an $0 < n \leq P(S)$ and an element $b \in \text{MCG}(S)$ with $ba^{-n}b^{-1} = a^n$ (then $\text{scl}(a) = 0$), or else $\text{scl}(a) \geq C$.*

This theorem is complementary to the following results.

Theorem 7.8. *Let S be a closed orientable surface of genus $g \geq 2$.*

- (1) [21] (cf. [43]) *If $a \in \text{MCG}(S)$ is a Dehn-twist along a separating simple closed curve, then $\text{scl}(a) \geq \frac{1}{6(3g-1)}$.*
- (2) [22] *There exists $a \in \text{MCG}(S)$ such that for all $n > 0$ and $c \in \text{MCG}(S)$, $a^n \neq ca^{-n}c^{-1}$ and that $\text{scl}(a) = 0$.*

Note that the element a in (2) is not pseudo-Anosov by Theorem 7.7. It follows from (1) that $\text{MCG}(S)$ is not uniformly perfect, and that $H_b^2(\text{MCG}(S); \mathbb{R})$ is not trivial (and indeed infinite dimensional by Theorem 3.18).

8 Appendix. Bounded cohomology

The theory of bounded cohomology was developed in Gromov's seminal work [30]. We already mentioned in Section 1.3 that the space of quasi-homomorphisms on a group is closely related to the second bounded cohomology of the group. We review a part of the theory in this section. We recommend the survey articles [5] and [55] for interested readers. All spaces and manifolds in this section are connected.

8.1 Riemannian geometry

In [30], Gromov defined the minimal volume, $\text{MinVol}(M)$, of a compact manifold M to be the infimum of the volume of all Riemannian metric g on M such that the sectional curvature K_g satisfies $-1 \leq K_g \leq 1$. If $\dim M = 2$, then Gauss–Bonnet formula gives

$$\int_M K_g dv_g = 2\pi \chi(M),$$

where $\chi(M)$ denotes the Euler characteristic of M . It immediately follows that $\text{MinVol}(M) = 2\pi |\chi(M)|$, and if $\chi(M) < 0$, then the minimal volume is attained (only) by a metric of constant curvature -1 .

It is difficult to compute $\text{MinVol}(M)$ in general. To give a lower bound for $\text{MinVol}(M)$, Gromov defined the simplicial volume, $\|M\|$, of M , which can be used in general as a replacement of the Euler characteristic of a surface. Let $c = \sum r_i c_i$ ($r_i \in \mathbb{R}$) be a real singular chain of M . Consider the l^1 -norm defined by $\|c\|_1 = \sum |r_i|$. For a homology class $\alpha \in H_*(M; \mathbb{R})$, define a semi-norm by

$$\|\alpha\| = \inf\{\|z\|_1 \mid z \text{ is closed and } [z] = \alpha\}.$$

If M is orientable, define $\|M\| = \|[M]\|$, where $[M]$ is the fundamental n -class. If M is not orientable, then pass to the double cover M' and define $\|M\| = \frac{1}{2}\|M'\|$.

Theorem 8.1 ([30]). *If M is a compact n -dimensional manifold, then*

$$C_n \|M\| \leq \text{MinVol}(M),$$

where $C_n > 0$ is a constant which depends only on n .

Of course, if $\|M\| = 0$, then this estimate is useless. Suppose $f: M \rightarrow N$ is a continuous map such that M and N are compact orientable manifolds of the same dimension. Then it is easy to see from the definition that

$$\|M\| \geq |\deg f| \cdot \|N\|.$$

It follows that if there exists a continuous map $g: M \rightarrow M$ such that $\deg g \neq 0, \pm 1$, then $\|M\| = 0$ (if M is compact). For example, if M is a sphere or a torus, then $\|M\| = 0$.

There are examples of M with $\|M\| > 0$.

Theorem 8.2 (Gromov–Thurston [30]). *Let (M, g) be an n -dimensional compact Riemannian manifold. Suppose there exists a constant k such that $-k \leq K_g \leq -1$. Then,*

$$\text{vol}(M, g) \leq c_n \|M\|,$$

where c_n is a constant which depends only on n .

Moreover, if $K_g = -1$, then

$$\text{vol}(M, g) = T_n \|M\|,$$

where T_n is the supremum of the volume of all geodesic n -simplices in the n -dimensional real hyperbolic space, \mathbb{H}^n .

It is shown in [30] that one can take $c_n = (n - 1)^n n!$. A simplex is called *geodesic* if all of its faces are totally geodesic. The proof is by “straightening” (into a geodesic one in the case $K_g = -1$) the lift of an n -simplex contained in $[M]$ in the universal cover of M . That’s how T_n comes into the estimate. It is known by now ([32]) that T_n is equal to the volume of ideal regular n -simplices in \mathbb{H}^n . Thus one needs to consider only regular geodesic n -simplices in the definition of T_n . Note that a simplex (possibly ideal) is *regular* if any permutation of vertices is induced by an isometry of \mathbb{H}^n .

We explain the connection between simplicial volume and bounded cohomology. The definition of bounded cohomology of a topological space X differs from the one for the ordinary real singular cohomology in that one considers only the set of singular cochains each of which is bounded as a function.

Let $S_n(X)$ be the set of n -dimensional singular simplices in X . Real n -dimensional singular cochains are functions $S_n(X) \rightarrow \mathbb{R}$. They form a vector space over \mathbb{R} , which we denote $C^n(X)$. Let δ be the standard coboundary map $C^n(X) \rightarrow C^{n+1}(X)$ for each n . The real singular cohomology of X , $H^*(X; \mathbb{R})$ (sometimes we omit \mathbb{R} in this chapter), is the cohomology of this cochain complex.

Now let $B^n(X) \subset C^n(X)$ be the set of all bounded functions on $S_n(X)$. Each element in $B^n(X)$ is called a bounded n -cochain. It is easy to see that $\delta(c) \in B^{n+1}(X)$ if $c \in B^n(X)$. The cohomology of the complex $B^*(X)$ is the *bounded cohomology* of X , denoted by $H_b^*(X)$. Each element $c \in C^n(X)$ has a natural l^∞ -norm.

$$\|c\|_\infty = \sup_{\sigma \in S_n(X)} c(\sigma) \leq \infty.$$

For an element $\beta \in H^*(X)$, define

$$\|\beta\| = \|\beta\|_\infty = \inf_y \|\beta\|_\infty \leq \infty,$$

where y are all cochains such that $\delta y = 0$ and $[y] = \beta$.

The inclusion $B^n(X) \rightarrow C^n(X)$ induces a canonical map $H_b^n(X) \rightarrow H^n(X)$, the comparison map. We say $\beta \in H^n(X)$ is *bounded* if it is contained in the image of this map, in other words, $\|\beta\|_\infty < \infty$.

The following two results in [30] are fundamental. There is a detailed account of the argument in [40], where he discusses a countable CW-complex X .

Theorem 8.3. *Let X be a topological space. Then,*

$$H_b^n(K(\pi_1(X), 1); \mathbb{R}) \simeq H_b^n(X; \mathbb{R})$$

for all n .

$H_b^n(K(\pi_1(X), 1); \mathbb{R})$ can be computed as $H_b^n(\pi_1(X); \mathbb{R})$ using the definition of the bounded cohomology of a group given in Section 1.3. We obtain the following

theorem, which says that the bounded cohomology depends only on the fundamental group.

Theorem 8.4. *Let X be a topological space. Then,*

$$H_b^n(X; \mathbb{R}) \simeq H_b^n(\pi_1(X); \mathbb{R}).$$

By this theorem, if M is a closed Riemannian manifold of negative sectional curvature, then $H_b^2(M; \mathbb{R})$ is infinite dimensional, in particular, non-trivial. This is because $G = \pi_1(M)$ is a non-elementary word-hyperbolic group, therefore $\widetilde{\text{QH}}(G)$ is infinite dimensional by Theorem 3.7, so that $H_b^2(G; \mathbb{R})$ is also infinite dimensional since $\widetilde{\text{QH}}(G)$ is a subspace as a vector space over \mathbb{R} in $H_b^2(G; \mathbb{R})$ (see Section 1.3).

The simplicial volume of a manifold M is related to the bounded cohomology of M as follows.

Theorem 8.5. *Let M be an n -dimensional closed orientable manifold and $\alpha \in H^n(M; \mathbb{R})$ the fundamental class such that $\langle \alpha, [M] \rangle = 1$. Then,*

$$\|M\|^{-1} = \|\alpha\|_\infty.$$

In particular, if α is bounded, namely $\|\alpha\|_\infty < \infty$, then $\|M\| \neq 0$.

It follows that if M is simply connected, then $\|M\| = 0$. This is because $H_b^n(M; \mathbb{R})$ is trivial since $\pi_1(M)$ is trivial. Therefore, $\|\alpha\|_\infty = \infty$.

The following is also proved using straightening.

Theorem 8.6 ([30]). *Let M be a closed Riemannian manifold whose sectional curvature is negative. Then the map $H_b^n(M; \mathbb{R}) \rightarrow H^n(M; \mathbb{R})$ is surjective for all $n > 1$.*

If M is an n -dimensional closed hyperbolic manifold ($K_g = -1$), then by Theorems 8.1 and 8.2, $\frac{C_n}{T_n} \text{vol}(M) \leq \text{MinVol}(M)$. The following result was conjectured in [30].

Theorem 8.7 ([4]). *Let (M, g) be a closed Riemannian manifold such that $K_g = -1$. Then $\text{MinVol}(M) = \text{vol}(M, g)$ and a metric which attains $\text{MinVol}(M)$ is isometric to g .*

We record one more recent progress. This is an answer in the affirmative to a question in [30].

Theorem 8.8 ([45]). *Let M be a closed locally symmetric space of non-compact type. Then $\|M\| > 0$.*

In particular it follows that $\text{MinVol}(M) > 0$ for such manifolds by Theorem 8.1, which was known for most cases ([31], [18]).

8.2 Group theory

Theorem 8.6 is generalized to word-hyperbolic groups. In general, $H_b^1(G; \mathbb{R}) = 0$ since a bounded homomorphism from G to \mathbb{R} is trivial.

Theorem 8.9 ([53]). *Let G be a non-elementary word-hyperbolic group. Then the map $H_b^n(G; \mathbb{R}) \rightarrow H^n(G; \mathbb{R})$ is surjective for all $n > 1$.*

In this chapter, we have seen several examples of groups G such that $\widetilde{\text{QH}}(G, \mathbb{R})$ is infinite dimensional. Those groups have infinite dimensional $H_b^2(G, \mathbb{R})$. Here is a list of such G .

- (1) Free groups of rank at least two (Theorem 2.1).
- (2) Non-elementary subgroups of a word-hyperbolic group (see Theorem 3.7 and Remark 3.8).
- (3) Subgroups in $\text{MCG}(S)$ which are not virtually abelian (Theorem 3.18).
- (4) Discrete subgroups in a rank-1 simple Lie group which are not virtually nilpotent (see the proof of Theorem 5.4).
- (5) The fundamental group G of a complete Riemannian manifold M of dimension at least two such that $\text{vol}(M) < \infty$ and the sectional curvature is non-positive, and such that M is not locally symmetric of rank at least two and G is irreducible (Theorem 3.20).
- (6) $G = A *_C B$ such that $|C \setminus A/C| \geq 3$ and $|B/C| \geq 2$; or $G = A *_C, \phi$ such that $|A/C| \geq 2$ and $|A/\phi(C)| \geq 2$ (see [26]).

If there is a surjective homomorphism $h: \widetilde{G} \rightarrow F$, where F is a rank two free group (sometimes then G is called *large*), then $\widetilde{\text{QH}}(G)$, therefore, $H_b^2(G, \mathbb{R})$ is infinite dimensional. This is because if $f: F \rightarrow \mathbb{R}$ is a homogeneous quasi-homomorphism, then $f \circ h: G \rightarrow \mathbb{R}$ is a homogeneous quasi-homomorphism. (We do not need that G is finitely generated. $\widetilde{\text{QH}}(F)$ is indeed infinite dimensional if we restrict it to $[F, F]$ as well.) For example, this argument applies to the fundamental group of a closed orientable surface of genus at least two, which is a non-elementary word-hyperbolic group. For the same reason, if a group G has a surjective homomorphism to one of the groups in the list, then $\widetilde{\text{QH}}(G)$ is infinite dimensional.

Not much is known about $H_b^n(G; \mathbb{R})$ for $n > 2$. If M is an n -dimensional closed locally symmetric space, then $H_b^n(\pi_1(M); \mathbb{R})$ is non-trivial by Theorems 8.4, 8.6, 8.8. It is not known in general if the dimension of $H_b^n(\pi_1(M); \mathbb{R})$ is finite.

There is a new direction of study of the second bounded cohomology with non-trivial coefficient. It is revealed that it has a connection to rigidity in terms of orbit equivalence of actions.

Let Γ and Λ be countable groups and $(X, \mu), (Y, \nu)$ probability Γ - and Λ - spaces respectively. A measurable isomorphism $F: X \rightarrow Y$ is said to be an *orbit equivalence* (OE) of the actions if for a.e. $x \in X$, $F(\Gamma x) = \Lambda F(x)$. (See [55], [56] and Chapter 9 of this volume [42].)

Let C_{reg} be the class of countable groups G such that $H_b^2(G, \ell^2(G)) \neq 0$.

Theorem 8.10 ([56]). *A countable group G belongs to C_{reg} if it admits one of the following actions:*

- (1) *a non-elementary simplicial action on a simplicial tree, proper on the set of edges,*
- (2) *a non-elementary, proper isometric action on a proper CAT(−1) space,*
- (3) *a non-elementary, proper isometric action on a δ -hyperbolic graph with bounded valency.*

In particular, a countable group which is free of rank at least two, a non-trivial free product of two countable groups except for $\mathbb{Z}_2 * \mathbb{Z}_2$, and a non-elementary subgroup of a word-hyperbolic group are in C_{reg} .

Among many rigidity theorems, Monod and Shalom showed the following.

Theorem 8.11 ([56]). *Let Γ_1, Γ_2 be torsion-free groups in C_{reg} , $\Gamma = \Gamma_1 \times \Gamma_2$, and let (X, μ) be an irreducible probability Γ -space. Let (Y, ν) be any other probability Γ -space. If the Γ -actions on X and Y are orbit equivalent, then they are isomorphic with respect to an automorphism of Γ .*

Acknowledgment. I would like to thank my collaborators M. Bestvina, D. Calegari and D. B. A. Epstein. I am grateful to A. Papadopoulos for his valuable comments on the first draft.

My series of works on quasi-homomorphisms started when I was visiting David Epstein in 1993 at the University of Warwick, partially supported by the Canon Foundation. I would like to dedicate this chapter to him for his seventieth birthday.

References

- [1] W. Ballmann, *Lectures on spaces of nonpositive curvature*. DMV Seminar 25, Birkhäuser, Basel 1995. [254](#)
- [2] H. Bass, J. Milnor, J.-P. Serre, Solution of the congruence subgroup problem for SL_n ($n \geq 3$) and Sp_{2n} ($n \geq 2$). *Inst. Hautes Études Sci. Publ. Math.* 33 (1967), 59–137. [244](#)
- [3] C. Bavard, Longueur stable des commutateurs. *L'Enseign. Math.* 37 (1991), 109–150 [242](#), [243](#), [244](#), [245](#)
- [4] G. Besson, G. Courtois, S. Gallot, Entropies et rigidités des espaces localement symétriques de courbure strictement négative. *Geom. Funct. Anal.* 5 (5) (1995), 731–799. [264](#)
- [5] G. Besson, G. Courtois, S. Gallot, A real Schwarz lemma and some applications. *Rend. Mat. Appl.* (7) 18 (2) (1998), 381–410. [261](#)
- [6] M. Bestvina, K. Fujiwara, Bounded cohomology of subgroups of mapping class groups. *Geom. Topol.* 6 (2002), 69–89. [248](#), [251](#), [253](#), [258](#)
- [7] M. Bestvina, K. Fujiwara, Quasi-homomorphisms on mapping class groups. *Glasnik Matematički* 42 (1) (2007), 213–236. [252](#)

- [8] M. Bestvina, K. Fujiwara, A characterization of higher rank symmetric spaces via bounded cohomology. Preprint, arXiv:math.GR/0702274; *Geom. Funct. Anal.*, to appear. [254](#)
- [9] B. Bowditch, Tight geodesics in the curve complex. *Invent. Math.* 171 (2) (2008), 281–300. [252](#), [260](#), [261](#)
- [10] M. Bridson, A. Haefliger, *Metric spaces of nonpositive curvature*. Grundlehren Math. Wiss. 319, Springer-Verlag, Berlin 1999. [246](#), [247](#), [248](#), [251](#), [254](#)
- [11] R. Brooks, Some remarks on bounded cohomology. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference* (State Univ. New York, Stony Brook, N.Y., 1978), Princeton University Press, Princeton, N.J., 1981, 53–63. [245](#)
- [12] M. Burger, N. Monod, Bounded cohomology of lattices in higher rank Lie groups. *J. Europ. Math. Soc. (JEMS)* 1 (2) (1999), 199–235. [245](#), [254](#), [255](#)
- [13] M. Burger, N. Monod, Continuous bounded cohomology and applications to rigidity theory. *Geom. Funct. Anal.* 12 (2) (2002), 219–280. [244](#), [254](#)
- [14] D. Calegari, Length and stable length. *Geom. Funct. Anal.* 18 (1) (2008), 50–76. [259](#)
- [15] D. Calegari, *scl*. MSJ Memoirs, Math. Soc. Japan., to appear. [242](#)
- [16] D. Calegari, Stable commutator length is rational in free groups. Preprint, arXiv:0802.1352. [244](#)
- [17] D. Calegari, K. Fujiwara, Stable commutator length in word-hyperbolic groups. Preprint 2006; arXiv:math..GR/0611889 [242](#), [257](#), [258](#), [259](#), [260](#), [261](#)
- [18] C. Connell, B. Farb, The degree theorem in higher rank. *J. Differential Geom.* 65 (2003), 19–59. [264](#)
- [19] M. Culler, Using surfaces to solve equations in free groups. *Topology* 20 (2) (1981), 133–145. [244](#), [246](#)
- [20] D. B. A. Epstein, K. Fujiwara, The second bounded cohomology of word-hyperbolic groups. *Topology* 36 (6) (1997), 1275–1289. [248](#), [249](#), [250](#), [251](#), [257](#), [261](#)
- [21] H. Endo, D. Kotschick, Bounded cohomology and non-uniform perfection of mapping class groups. *Invent. Math.* 144 (1) (2001), 169–175. [261](#)
- [22] H. Endo, D. Kotschick, Failure of separation by quasi-homomorphisms in mapping class groups. *Proc. Amer. Math. Soc.* 135 (9) (2007), 2747–2750. [261](#)
- [23] B. Farb, H. Masur, Superrigidity and mapping class groups. *Topology* 37 (6) (1998), 1169–1176. [254](#)
- [24] B. Farb, A. Lubotzky, Y. Minsky, Rank-1 phenomena for mapping class groups. *Duke Math. J.* 106 (3) (2001), 581–597. [256](#)
- [25] K. Fujiwara, The second bounded cohomology of a group acting on a Gromov-hyperbolic space. *Proc. London Math. Soc.* (3) 76 (1) (1998), 70–94. [248](#), [249](#), [250](#), [256](#)
- [26] K. Fujiwara, The second bounded cohomology of an amalgamated free product of groups. *Trans. Amer. Math. Soc.* 352 (3) (2000), 1113–1129. [265](#)
- [27] K. Fujiwara, On non-bounded generation of discrete subgroups in rank-1 Lie group. In *Geometry, spectral theory, groups, and dynamics*, Contemp. Math. 387, Amer. Math. Soc., Providence, RI, 2005, 153–156. [256](#)
- [28] M. Gromov, *Asymptotic invariants of infinite groups*. London Math. Soc. Lecture Note Ser. 182, Cambridge University Press, Cambridge 1993.

- [29] M. Gromov, Hyperbolic groups. In *Essays in group theory*, Springer-Verlag, New York 1987, 75–263. [246](#), [248](#), [251](#), [256](#), [259](#)
- [30] M. Gromov, Volume and bounded cohomology. *Inst. Hautes Études Sci. Publ. Math.* 56 (1983), 5–99. 1982. [244](#), [245](#), [261](#), [262](#), [263](#), [264](#)
- [31] M. Gromov, Filling Riemannian manifolds. *J. Differential Geom.* 18 (1983) 1–147. [264](#)
- [32] U. Haagerup, H. Munkholm, Simplices of maximal volume in hyperbolic n -space. *Acta Math.* 147 (1–2) (1981), 1–11. [263](#)
- [33] U. Hamenstädt, Bounded cohomology and isometry groups of hyperbolic spaces. *J. Eur. Math. Soc.* 10 (2) (2008), 315–349.
- [34] U. Hamenstädt, Geometry of the complex of curves and of Teichmüller space. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 447–467. [252](#)
- [35] J. L. Harer, Stability of the homology of the mapping class groups of orientable surfaces. *Ann. of Math.* (2) 121 (2) (1985), 215–249. [251](#)
- [36] J. L. Harer, The virtual cohomological dimension of the mapping class group of an orientable surface. *Invent. Math.* 84 (1) (1986), 157–176. [251](#)
- [37] W. J. Harvey, Boundary structure of the modular group. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference* (State Univ. New York, Stony Brook, N.Y., 1978), Princeton University Press, Princeton, N.J., 1981, 245–251. [251](#)
- [38] N. V. Ivanov, Mapping class groups. In *Handbook of geometric topology*, North-Holland, Amsterdam 2002, 523–633. [252](#)
- [39] N. V. Ivanov, *Subgroups of Teichmüller modular groups*. Transl. Math. Monogr. 115, Amer. Math. Soc., Providence, RI, 1992. [253](#)
- [40] N. V. Ivanov, Foundations of the theory of bounded cohomology. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov* 143 (1985), 69–109. [244](#), [263](#)
- [41] V. A. Kaimanovich, H. Masur, The Poisson boundary of the mapping class group. *Invent. Math.* 125 (2), (1996), 221–264. [254](#)
- [42] Y. Kida, Introduction to measurable rigidity of mapping class groups. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 297–367. [265](#)
- [43] M. Korkmaz, Stable commutator length of a Dehn twist. *Michigan Math. J.* 52 (1) (2004), 23–31. [261](#)
- [44] D. Kotschick, Quasi-homomorphisms and stable lengths in mapping class groups. *Proc. Amer. Math. Soc.* 132 (11) (2004), 3167–3175. [255](#)
- [45] J.-F. Lafont, B. Schmidt, Simplicial volume of closed locally symmetric spaces of non-compact type. *Acta Math.* 197 (1) (2006), 129–143. [264](#)
- [46] J. F. Manning, Quasi-actions on trees and property (QFA). With an appendix by N. Monod and B. Rémy. *J. London Math. Soc.* (2) 73 (1) (2006), 84–108. [244](#)
- [47] G. A. Margulis, E. B. Vinberg, Some linear groups virtually having a free quotient. *J. Lie Theory* 10 (1) (2000), 171–180. [256](#)
- [48] H. A. Masur and Y. N. Minsky, Geometry of the complex of curves. I. Hyperbolicity. *Invent. Math.* 138 (1) (1999), 103–149. [241](#), [252](#), [260](#)

- [49] S. Matsumoto, S. Morita, Bounded cohomology of certain groups of homeomorphisms. *Proc. Amer. Math. Soc.* 94 (3) (1985), 539–544. [244](#)
- [50] J. McCarthy, A “Tits-alternative” for subgroups of surface mapping class groups. *Trans. Amer. Math. Soc.* 291 (2) (1985), 583–612. [253](#), [255](#)
- [51] J. McCarthy, A. Papadopoulos, Dynamics on Thurston’s sphere of projective measured foliations. *Comment. Math. Helv.* 64 (1) (1989), 133–166. [253](#)
- [52] J. Milnor, On the existence of a connection with curvature zero. *Comment. Math. Helv.* 32 (1958), 215–223. [243](#)
- [53] I. Mineyev, Straightening and bounded cohomology of hyperbolic groups. *Geom. Funct. Anal.* 11 (4) (2001), 807–839. [265](#)
- [54] N. Monod, *Continuous bounded cohomology of locally compact groups*. Lecture Notes in Math. 1758. Springer-Verlag, Berlin 2001. [244](#)
- [55] N. Monod, An invitation to bounded cohomology. In *Proceedings of the International Congress of Mathematicians* (Madrid, 2006), Vol. II, EMS Publishing House, Zürich 2006, 1183–1211. [261](#), [265](#)
- [56] N. Monod, Y. Shalom, Orbit equivalence rigidity and bounded cohomology. *Ann. of Math.* (2) 164 (3) (2006), 825–878. [265](#), [266](#)
- [57] S. Morita, Structure of the mapping class groups of surfaces: A survey and a prospect. In *Proceedings of the Kirbyfest* (Berkeley, 1998), Geom. Topol. Mongr. 2, Geom. Topol. Publ., Coventry 1999, 349–406. [253](#)
- [58] L. Polterovich, Z. Rudnick, Stable mixing for cat maps and quasi-morphisms of the modular group. *Ergodic Theory Dynam. Systems* 24 (2) (2004), 609–619. [257](#)
- [59] G. Pólya, G. Szegő, *Problems and theorems in analysis*. I. Classics in Mathematics, Springer-Verlag, Berlin 1998. [242](#), [243](#)
- [60] O. Tavgen, Bounded generation of Chevalley groups over rings of algebraic S -integers. *Math. USSR-Izv.* 36 (1991), 101–128. [256](#)
- [61] D. Zhuang, Irrational stable commutator length in finitely presented groups. *J. Mod. Dyn.* 2 (3) (2008), 499–507. [244](#)
- [62] R. J. Zimmer, *Ergodic theory and semisimple groups*. Birkhäuser, Basel 1984. [255](#)

Chapter 8

Lefschetz fibrations on 4-manifolds

Mustafa Korkmaz and András I. Stipsicz

Contents

1	Introduction	271
2	Lefschetz pencils and Lefschetz fibrations	272
	2.1 Lefschetz pencils on $X \subset \mathbb{C}\mathbb{P}^n$	272
	2.2 Lefschetz fibrations	273
3	Topology of Lefschetz fibrations	274
	3.1 Vanishing cycles, topological description	275
	3.2 The mapping class group and the monodromy factorization	275
	3.3 Sections of Lefschetz fibrations	277
4	Relation to symplectic topology: the work of Gompf and Donaldson	278
5	Results on Lefschetz fibrations	279
	5.1 Lefschetz fibrations of low genus	279
	5.2 Holomorphicity of Lefschetz fibrations	281
	5.3 Commutator lengths of Dehn twists	282
	5.4 Minimal number of singular fibers	284
	5.5 Fundamental groups of Lefschetz fibrations	287
	5.6 Sections of Lefschetz fibrations	288
	5.7 Surface bundles over surfaces with nonzero signature	289
	5.8 Teichmüller spaces	289
6	Variations of Lefschetz fibrations	290
	6.1 Achiral Lefschetz fibrations	290
	6.2 Lefschetz fibrations and Stein structures	290
	6.3 Achiral Lefschetz fibrations and contact structures	291
	6.4 Further generalizations	292
7	Open problems	292
	References	293

1 Introduction

Lefschetz pencils and fibrations were introduced for studying topological properties of smooth complex projective varieties. More recently, as an application of Donaldson's asymptotically holomorphic methods [13], Lefschetz pencils have been found

on all symplectic manifolds [14], [15]. Conversely, Gompf showed that a 4-manifold admitting a Lefschetz pencil/fibration carries a symplectic structure [23], [24]. Since symplectic 4-manifolds play a prominent role in modern low-dimensional topology, the Donaldson–Gompf correspondence piqued interest in the study of Lefschetz fibrations. In addition, these structures provide a connection between symplectic topology and geometric group theory. The utilization of this correspondence led, for example, to results on the commutator lengths of certain elements in the mapping class groups of oriented surfaces (of genus > 1) [17], [29]. The arguments in [17], [29] relied on Seiberg–Witten theory, and had interesting consequences on the algebraic structure of mapping class groups.

In this chapter we collect the most basic definitions, describe the fundamental results and explain some of the consequences of the correspondence mentioned above. In some cases we provide a full or partial proof of the result. Most of the time, however, we restrict ourselves to quoting the theorems, and try to put them in perspective. The only new result is stated in Corollary 5.16. We would like to point out that although most of the fundamental results hold in any (even) dimension, we restrict our attention to the case of symplectic 4-manifolds, in which case the Lefschetz fibrations have surface fibers, and therefore the theory of mapping class groups enters in an essential way.

Acknowledgement. The second author would like to acknowledge partial support from the Clay Mathematics Institute. He was also supported by OTKA T49449.

2 Lefschetz pencils and Lefschetz fibrations

In this section we define Lefschetz pencils and Lefschetz fibrations.

2.1 Lefschetz pencils on $X \subset \mathbb{C}\mathbb{P}^n$

Suppose that X is a smooth complex projective variety in $\mathbb{C}\mathbb{P}^n$. The idea Lefschetz had for studying the topology of X was to ‘slice it up’ into smaller dimensional pieces, study those slices first and then understand how they are pieced together to form X . To this end, we consider a (complex) codimension-2 hyperplane $A \cong \mathbb{C}\mathbb{P}^{n-2}$, say

$$A = \{[x_0 : \dots : x_n] \in \mathbb{C}\mathbb{P}^n \mid x_0 = x_1 = 0\}.$$

The subset A (called the *axis*) is expected to intersect X in a (complex) codimension-2 submanifold – for a generic choice of A this is exactly what happens. Now consider the family of codimension-1 hyperplanes containing the axis A . This family $\{H_t\}$ can be parametrized by $\mathbb{C}\mathbb{P}^1$; in the above example

$$H_t = \{[x_0 : \dots : x_n] \mid t_0 x_0 + t_1 x_1 = 0, t = [t_0 : t_1] \in \mathbb{C}\mathbb{P}^1\}.$$

Notice that since $\{H_t\}$ sweeps out $\mathbb{C}\mathbb{P}^n$, this construction gives a map

$$\psi: \mathbb{C}\mathbb{P}^n - A \rightarrow \mathbb{C}\mathbb{P}^1$$

by sending a point $x \in \mathbb{C}\mathbb{P}^n - A$ to $t \in \mathbb{C}\mathbb{P}^1$ if $x \in H_t$.

Assuming that the intersection $A \cap X$ is transverse, now we consider the slices $F_t = H_t \cap X$. For a generic choice of t this intersection is transverse again, and so F_t is a smooth submanifold of X . For finitely many t , however, the transversality fails to hold. Once again, for a sufficiently generic choice of A the singularity of F_t is mild enough; it is a transverse double point, hence it is modelled on a canonical example. (Notice that $A \cap X \subset F_t$ for every $t \in \mathbb{C}\mathbb{P}^1$.) Also, the restriction of the map ψ from $\mathbb{C}\mathbb{P}^n - A$ to $X - A$ provides a map

$$\psi: X - A \rightarrow \mathbb{C}\mathbb{P}^1.$$

The argument above indicates that the singularities of ψ and its behavior near $X - A$ can be given on canonical models, leading us to

Definition 2.1. Let X be a closed, connected, oriented, smooth 4-manifold. A *Lefschetz pencil* on X is a nonempty finite subset B of X , called the *base locus*, together with a smooth map $\psi: X - B \rightarrow \mathbb{C}\mathbb{P}^1$ such that each point $b \in B$ has an orientation-preserving coordinate chart in which ψ is given by the projectivization map $\mathbb{C}^2 - \{0\} \rightarrow \mathbb{C}\mathbb{P}^1$, and each critical point of ψ has an orientation-preserving chart on which $\psi(z_1, z_2) = z_1^2 + z_2^2$ relative to a suitable smooth chart on $\mathbb{C}\mathbb{P}^1$. For $t \in \mathbb{C}\mathbb{P}^1$, the *fiber* F_t is $\psi^{-1}(t) \cup B \subset X$.

2.2 Lefschetz fibrations

By appropriately blowing up the points of B in a Lefschetz pencil, ultimately we get a fibration map on the blown-up X into $\mathbb{C}\mathbb{P}^1$, where the fibers (the proper transforms of F_t) are smooth submanifolds with finitely many exceptions, and in these exceptions the total transform \tilde{F}_t has a transverse double point singularity. In real dimension four (when B is a finite set of points) this construction leads to the following more general definition – when we allow the base curve to have higher genus, and X to be a smooth 4-manifold with possibly non-empty boundary ∂X .

Definition 2.2. Let X be a compact, connected, oriented, smooth 4-manifold and let Σ be a compact, connected, oriented, smooth surface. A *Lefschetz fibration* is a smooth map $f: X \rightarrow \Sigma$ such that

- $f^{-1}(\partial \Sigma) = \partial X$,
- the set $C = \{p_1, p_2, \dots, p_k\}$ of critical points of f lies in the interior of X ,
- for each i , there are orientation-preserving complex local charts around p_i and $f(p_i)$ where f is given by $f(z_1, z_2) = z_1^2 + z_2^2$.

A regular fiber of a Lefschetz fibration is a smooth, closed, connected, oriented surface whose genus is called the *genus* of the Lefschetz fibration. There are finitely many singular fibers which are immersed surfaces with transverse positive double points. In the following we will assume that f is injective on C , that is, a singular fiber admits a unique singular point. (By a slight perturbation of the map f this property can always be achieved.) A singular fiber is called *reducible* if it becomes disconnected after removing the double point. Reducible fibers have two components. Both of these components have square -1 in homology. The topology of the neighborhood of a reducible singular fiber is determined by the genera of the two components we get after removing the singular point; the smaller of these genera is called the *type* of the reducible singular fiber. A singular fiber which is not reducible is called *irreducible*. The vector giving the number of irreducible singular fibers, together with the number of reducible singular fibers of various types is called the *combinatorial data* of the Lefschetz fibration. Notice, for example, that the Euler characteristic of the total space X is determined by the fiber and base genera and the number of singular fibers, while the signature typically depends on the actual fibration map. A fibration is *relatively minimal* if there is no (-1) -sphere contained in a fiber; in other words, the 4-manifold cannot be blown down in such a way that the fibration structure is preserved. A map $\varphi: \Sigma \rightarrow X$ is a *section* if $f \circ \varphi = \text{id}_\Sigma$, that is, $\varphi(t) \in F_t$ for every $t \in \Sigma$.

Two Lefschetz fibrations $f: X \rightarrow \Sigma$ and $f': X' \rightarrow \Sigma'$ are *equivalent* if there are diffeomorphisms $\Phi: X \rightarrow X'$ and $\phi: \Sigma \rightarrow \Sigma'$ such that $f' \circ \Phi = \phi \circ f$. In particular, equivalent Lefschetz fibrations have equal fiber genera and equal combinatorial data.

The *fiber sum* $f: X \#_f X' \rightarrow \Sigma \# \Sigma'$ of the two fibrations $f: X \rightarrow \Sigma$ and $f': X' \rightarrow \Sigma'$ of the same genus is defined as follows: remove the neighborhood of a regular fiber from each fibration and glue the resulting 4-manifolds with boundary with a fiber preserving, orientation-reversing diffeomorphism of their diffeomorphic boundaries. The actual gluing map is suppressed from the notation, although various choices give rise to drastically different results; examples where different choices give total spaces with different first integer homologies can be found in [39].

3 Topology of Lefschetz fibrations

The fibration map $X \rightarrow \Sigma$ provides a relative handlebody decomposition structure for X (built on the neighborhood of a regular fiber), a structure which ultimately can be used to describe f as a factorization of a certain element in the mapping class group of the regular fiber. We devote this section to outline the construction of the handlebody decomposition and to set up the correspondence with mapping class groups.

3.1 Vanishing cycles, topological description

Let us suppose that $f: X \rightarrow \Sigma$ is a given Lefschetz fibration. Consider two embedded disks $D_1 \subset D_2 \subset \Sigma$ in such a way that there is no critical value in the annulus $D_2 - D_1$. Then it is fairly easy to see that $f^{-1}(D_1)$ is a deformation retract of $f^{-1}(D_2)$. The topology of this inverse image will change, however, if $D_2 - D_1$ contains a critical value $f(p_i)$. Away from the critical point we still have the retraction, but near p_i we detect a 2-handle attachment.

Recall that each critical point has a local coordinate chart in which $f(z_1, z_2) = z_1 z_2$ (which is the same as $(z'_1)^2 + (z'_2)^2$ through the coordinate change $z_1 = z'_1 + i z'_2$, $z_2 = z'_1 - i z'_2$). On this chart, the unique critical value is 0, and $f^{-1}(0) = \{(z_1, z_2) \mid z_1 = 0 \text{ or } z_2 = 0\}$ is a pair of intersecting planes. Thus, each singular fiber is a smoothly immersed surface, and each critical point corresponds to a positive transverse double point. Nearby fibers $F_t = f^{-1}(t)$, $t \neq 0$, are nonsingular, and are obtained from $f^{-1}(0)$ by removing the intersection. That is, we perform a surgery on a 0-sphere in the fiber (the pair of identified points) by removing the intersecting disks and replacing them with the annulus $z_1 z_2 = t$. Equivalently, each critical point corresponds to an embedded surgery circle called a *vanishing cycle* in a nearby regular fiber, and the singular fiber is obtained by collapsing the vanishing cycle to a point to create a transverse double point. If the singular fiber is reducible, that is, the vanishing cycle separates the generic fiber, the singular fiber will be the image of an immersion of a disconnected surface.

Next we show that a Lefschetz critical point corresponds to a 2-handle, and determine its attaching map. Near the critical point we write $f(z_1, z_2) = z_1^2 + z_2^2$, so a nearby regular fiber is given by $z_1^2 + z_2^2 = t$, and after multiplying f by a unit complex number we can assume $t > 0$. If we intersect the fiber with $\mathbb{R}^2 \subset \mathbb{C}^2$, we obtain the circle $x_1^2 + x_2^2 = t$ in \mathbb{R}^2 (where $z_j = x_j + i y_j$). This circle bounds a disk $D_t \subset \mathbb{R}^2$, which was called a *thimble* by Lefschetz. As t approaches zero, the thimble D_t shrinks to a point in \mathbb{R}^2 . Thus, $\partial D_t = F_t \cap \mathbb{R}^2$ is the vanishing cycle of the critical point, and we explicitly see the singular fiber F_0 being created from F_t by the collapse of D_t . A regular neighborhood νF_0 of the singular fiber is obtained from the neighborhood νF_t by adding a regular neighborhood of D_t . The latter neighborhood is clearly a 2-handle h attached to νF_t . As an explicit computation in the local model shows, the 2-handle h is attached along $\partial D_s \subset \partial \nu F$ (the vanishing cycle) with framing -1 relative to the framing induced by the surface F_s containing ∂D_s .

3.2 The mapping class group and the monodromy factorization

Lefschetz fibrations can be described combinatorially by means of their *monodromy*. Let Γ_g denote the *mapping class group* of the closed connected orientable genus- g surface F , that is, the group of isotopy classes of orientation-preserving self-diffeomorphisms of F . For a smooth fiber bundle $f: E \rightarrow B$ with fibers diffeomorphic to the manifold F , we define the *monodromy representation* $\Psi: \pi_1(B) \rightarrow \Gamma_g$ of

f relative to a fixed identification φ of F with the fiber over the base point of B : For $I = [0, 1]$ and for each loop $\gamma: I \rightarrow B$ the bundle $f_\gamma: \gamma^*(E) \rightarrow I$ is canonically trivial, inducing a diffeomorphism $f_\gamma^{-1}(0) \rightarrow f_\gamma^{-1}(1)$ (up to isotopy). Using φ to identify $f_\gamma^{-1}(0)$ and $f_\gamma^{-1}(1)$ with F , we obtain the element $\Psi(\gamma) \in \Gamma_g$. Changing the identification φ changes Ψ by a conjugation by an element of Γ_g .

For a relatively minimal, genus- g Lefschetz fibration $f: X \rightarrow \Sigma$ with a regular fiber F , we define the *monodromy* of the Lefschetz fibration to be the monodromy of the fiber bundle $X - f^{-1}(Q) \rightarrow \Sigma - Q$, where $Q = f(C)$ is the set of critical values. For $f: X \rightarrow \Sigma$ as above with f injective on the set C of critical points, the monodromy representation $\Psi: \pi_1(\Sigma - Q) \rightarrow \Gamma_g$ determines f up to isomorphism, except in the cases of sphere and torus bundles ($C = \emptyset$) over closed surfaces. This follows from the fact that for $g \geq 2$ the space of self-diffeomorphisms of F isotopic to the identity is contractible (while for $g = 0, 1$ it has nontrivial topology, resulting in nontrivial S^2 - and T^2 -bundles over the sphere). In fact, two Lefschetz fibrations $X \rightarrow \Sigma$ and $X' \rightarrow \Sigma'$ with diffeomorphic regular fibers F and F' with genera at least two are *equivalent* if and only if their monodromy representations agree, that is, there are isomorphisms making the diagram

$$\begin{array}{ccc} \pi_1(\Sigma - Q) & \longrightarrow & \Gamma_g \\ \downarrow & & \downarrow \\ \pi_1(\Sigma' - Q') & \longrightarrow & \Gamma'_g \end{array}$$

commute, where Γ'_g is the mapping class group of F' , and the isomorphism $\pi_1(\Sigma - Q) \rightarrow \pi_1(\Sigma' - Q')$ is induced by a smooth map $\Sigma \rightarrow \Sigma'$.

As a local computation shows, the monodromy of a Lefschetz fibration $f: X \rightarrow D^2$ over the disk with a single critical point is a right-handed Dehn twist along the vanishing cycle corresponding to the singular fiber. Therefore the monodromy of a Lefschetz fibration $f: X \rightarrow \Sigma_h$ comprises a factorization of $1 \in \Gamma_g$ as

$$1 = \prod_{i=1}^n t_{v_i} \prod_{i=1}^h [a_i, b_i],$$

where v_i are the vanishing cycles of the singular fibers and t_{v_i} is the right-handed Dehn twist about v_i . This factorization of the identity is called the *monodromy factorization*. In particular, a product $\prod_{i=1}^h [a_i, b_i]$ of h commutators in Γ_g gives an F -bundle over the surface Σ_h^1 of genus h with one boundary component. The mapping classes a_i and b_i specify the monodromy along the obvious free generating system $\langle \alpha_1, \beta_1, \dots, \alpha_k, \beta_k \rangle$ of $\pi_1(\Sigma_h^1)$. If $\prod_{i=1}^h [a_i, b_i] = 1$ in Γ_g , we get an F -bundle $X \rightarrow \Sigma_h$. (The bundle is uniquely determined by the word once $g \geq 2$.)

An expression $\prod_{i=1}^k t_i \in \Gamma_g$ with t_i right-handed Dehn twists provides a genus- g Lefschetz fibration $X \rightarrow D^2$ over the disk with fiber F . If $\prod_{i=1}^k t_i = 1$ in Γ_g then the fibration closes up to a fibration over the sphere S^2 and the closed-up manifold is

uniquely determined by the word $\prod_{i=1}^k t_i$ once $g \geq 2$. By combining the above two constructions, a word

$$w = \prod_{i=1}^{k'} t_i \prod_{j=1}^h [a_j, b_j]$$

gives a Lefschetz fibration over $\Sigma_h - D^2$ and if $w = 1$ in Γ_g we get a Lefschetz fibration $X \rightarrow \Sigma_h$.

In a genus- g Lefschetz fibration over the base S^2 , a theorem of Ivan Smith shows that there must be some irreducible fibers. (A Lefschetz fibration is called *trivial* if there are no singular fibers.)

Theorem 3.1 ([3]). *There are no nontrivial Lefschetz fibrations with base S^2 whose monodromy is contained in the Torelli group, the subgroup of the mapping class group acting trivially on the first homology of the surface.* \square

A sharper lower bound for the number of irreducible singular fibers was given in [49].

Theorem 3.2 ([49]). *A nontrivial genus- g Lefschetz fibration over the sphere S^2 has at least $\frac{4g+2}{5}$ irreducible singular fibers.* \square

It is easy to see that the monodromy factorization corresponding to a fiber sum is simply the product of the factorizations defining the individual factors. The only point one should be careful about is how the typical fiber is identified with the genus- g surface Σ_g using which the mapping class group is defined: for different identifications, terms in the monodromy factorizations should be conjugated. Consequently, the dependence of the result of the fiber sum on the gluing diffeomorphism will be visible through an overall conjugation of one side of the fiber sum.

3.3 Sections of Lefschetz fibrations

Suppose that the Lefschetz fibration $f: X \rightarrow \Sigma$ admits a section. In terms of the monodromy representation this simply means that the point of the fiber $f^{-1}(t)$ where it is intersected by the section should stay fixed. Therefore a section provides a lift of the representation Ψ from $\pi_1(\Sigma - Q)$ to the mapping class group $\Gamma_{g,1}$ with one marked point. Recall that Q denotes the subset of Σ consisting of critical values. Conversely, every such representation determines a fibration with a section. Similarly, the existence of n disjoint sections is equivalent to the existence of a lift of the monodromy representation to $\Gamma_{g,n}$, the mapping class group of the genus- g surface with n marked points.

Recall that there is a canonical homomorphism $p: \Gamma_g^1 \rightarrow \Gamma_{g,1}$ where Γ_g^1 denotes the mapping class group of a genus- g surface with one boundary component consisting of the relative isotopy classes of diffeomorphisms which fix the boundary pointwise.

It is not hard to see that $\ker p$ is isomorphic to \mathbb{Z} , generated by the right-handed Dehn twist t_δ along a simple closed curve δ parallel to the boundary. If the factorization $1 = \prod_i t_i \prod_j [a_j, b_j]$ lifts from Γ_g to $\Gamma_{g,1}$ then (according to the above said) the corresponding fibration has a section. In addition, it is not hard to see that if this product (when viewed in Γ_g^1) is equal to t_δ^m then this section is of self-intersection $-m$. A similar identity holds for n disjoint sections (in which case we use the mapping class groups $\Gamma_{g,n}$ and Γ_g^n).

This observation can be used to determine Lefschetz pencils using the monodromy representation. Recall that a Lefschetz pencil can be blown up at the base points to get a Lefschetz fibration (on the blown-up 4-manifold). Therefore a Lefschetz pencil can be regarded as a Lefschetz fibration over S^2 together with a distinguished set of sections, all of self-intersection (-1) . Consequently an expression $\prod_{i=1}^n t_i = t_{\delta_1} \dots t_{\delta_k}$ in Γ_g^k (where all t_i stand for right-handed Dehn twists and t_{δ_i} are right-handed Dehn twists along circles parallel to the boundary components of the Riemann surface at hand) naturally describes a Lefschetz pencil: The relation determines a Lefschetz fibration with k (disjoint) sections, each of self-intersection (-1) , and after blowing these sections down we get a Lefschetz pencil. Conversely, by blowing up the base locus of a Lefschetz pencil we arrive to a Lefschetz fibration which can be captured (together with the exceptional divisors of the blow-ups, which are all sections now) by a relator of the above type.

4 Relation to symplectic topology: the work of Gompf and Donaldson

In 1994 Donaldson showed that in a symplectic manifold (X, ω) with symplectic form ω satisfying

$$\frac{1}{2\pi}[\omega] \in \text{Im}(H^2(X; \mathbb{Z}) \rightarrow H^2(X; \mathbb{R}))$$

the Poincaré dual of $\frac{k}{2\pi}[\omega]$ (for k sufficiently large) can be represented by a symplectic submanifold. In his proof Donaldson considered the complex line bundle $L \rightarrow X$ with $c_1(L) = \frac{1}{2\pi}[\omega]$ and showed that for k sufficiently large the line bundle $L^{\otimes k}$ admits an ‘asymptotically holomorphic section’, for which the zero-set is a symplectic submanifold. By finding two transverse such sections σ_1, σ_2 of $L^{\otimes k}$ (for an appropriate, and possibly higher k), Donaldson showed that the map

$$x \mapsto [\sigma_1(x) : \sigma_2(x)]$$

(defined on $X - \{x \in X \mid \sigma_1(x) = \sigma_2(x) = 0\}$) provides a Lefschetz pencil on X . For a symplectic 4-manifold therefore (after possibly rescaling and perturbing its symplectic structure) an appropriate blow-up admits a Lefschetz fibration over S^2 .

Conversely, under a mild assumption (which will be explained later) a Lefschetz fibration $f: X \rightarrow \Sigma$ provides a way to equip the 4-manifold X with a symplectic structure [23], [24]. Notice first that the tangents of the fibers (and an orientation of the

fiber) together with their orthogonal complements give (away from the critical points) an almost complex structure. Because of the canonical local model near the critical points, this structure obviously extends to an almost complex structure on X . Suppose now that the typical fiber F_t represents a nonzero homology class in $H_2(X; \mathbb{R})$. Using a partition of unity then a form can be constructed on X which will be closed and nondegenerate along the fiber directions. (Here again the local model near the critical points plays a crucial role.) By pulling back an appropriate multiple of a volume form on the base, the construction ultimately provides a symplectic structure on X . With a little more care one can arrange that a fixed (finite) set of disjoint sections will be symplectic submanifolds, hence if these sections are (-1) -spheres, those can be symplectically blown down. Notice that the fiber of a Lefschetz fibration is nontrivial in real homology provided either (a) the fibration admits a section, (b) the fiber genus is not equal to 1 (since in this case the first Chern class of the almost complex structure gives $\langle c_1(X, J), [F_t] \rangle = 2 - 2g \neq 0$), or (c) the fibration admits singular fibers. A fibration stemming from blowing up a pencil always admits sections (the exceptional spheres of the blow-ups).

Remark 4.1. The assumption on the nontriviality of the real homology class of the fiber is essential: taking the product of a nontrivial circle bundle over S^2 with the circle S^1 we get a torus bundle over S^2 with vanishing second real cohomology, so although the 4-manifold admits a torus fibration, it cannot be symplectic.

In conclusion we get

Theorem 4.2. *A 4-manifold X admits a symplectic structure if and only if it carries a Lefschetz pencil, equivalently if some blow-up X' of X carries a Lefschetz fibration.*

□

Notice that this construction provides the possibility of describing symplectic structures through the factorization of the unit element in some mapping class group, defined by a Lefschetz fibration associated to the symplectic structure. There is a delicate equivalence relation among those factorizations providing the same fibration, and in order to get invariants of symplectic structures in this way, one also needs to understand the relation between factorizations corresponding to different fibrations resulting in the same symplectic manifold. The first step towards the definition of such invariants was made in [6].

5 Results on Lefschetz fibrations

5.1 Lefschetz fibrations of low genus

We start with a definition. Suppose that a mapping class is represented as a product of right-handed Dehn twists. A *Hurwitz move* replaces a factor $t_a t_b$ in this representation

by $t_{a(b)}t_a = (t_a t_b t_a^{-1})t_a$ or by $t_b t_{b^{-1}(a)} = t_b(t_b^{-1}t_a t_b)$. Two such representations are called *Hurwitz equivalent* if they can be obtained from each other by a sequence of Hurwitz moves.

Let $f: X \rightarrow S^2$ be a relatively minimal genus- g Lefschetz fibration. If the fiber genus $g = 0$, then X is diffeomorphic to $S^2 \times S^2$ or $S^2 \tilde{\times} S^2$, the trivial and twisted sphere bundles over S^2 . In complex surface theory these manifolds are called (*minimal*) *ruled surfaces*.

If $g = 1$ then the 4-manifold X is either a torus bundle (in case there are no singular fibers), or the monodromy of the Lefschetz fibration is Hurwitz equivalent to $(t_a t_b)^{6n} = 1$ for some integer $n \geq 0$ [37], where a and b are two simple closed curves on the torus (a regular fiber) intersecting transversely at one point. In the mapping class group Γ_1 of the torus we have $(t_a t_b)^6 = 1$, and it can be fairly easily checked that the 4-manifold corresponding to the factorization $(t_a t_b)^6 = 1$ is $\mathbb{C}\mathbb{P}^2 \# 9\overline{\mathbb{C}\mathbb{P}^2}$ with an elliptic fibration on it. This manifold is usually called the *rational elliptic surface*, and is denoted by $E(1)$. For the general case, the factorization $(t_a t_b)^{6n} = 1$ gives rise to X_n , which is diffeomorphic to the elliptic surface $E(n)$, the n -fold fiber sum of $E(1)$. We note here that this simple classification of genus-1 Lefschetz fibrations stems from the simplicity of the mapping class group Γ_1 , which is isomorphic to $SL_2(\mathbb{Z})$ through the map

$$t_a \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad t_b \mapsto \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

In the mapping class group Γ_2 of the oriented closed surface of genus 2 (depicted in Figure 1), let $\sigma = (t_1 t_2)^6 = (t_4 t_5)^6$. Define the words

$$\begin{aligned} T &= t_1 t_2 t_3 t_4 t_5 t_5 t_4 t_3 t_2 t_1, \\ W_0 &= T^2, \\ W_1 &= (t_1 t_2 t_3 t_4 t_5)^6, \\ W_2 &= \sigma (t_3 t_4 t_5 t_2 t_3 t_4 t_1 t_2 t_3)^2 T. \end{aligned}$$

It can be shown that each of the words W_0 , W_1 and W_2 represents the identity in the mapping class group Γ_2 .

Theorem 5.1 (Auroux [4]). *Let F be any factorization of the identity as a product of positive Dehn twists in the mapping class group Γ_2 . Then there exist integers $\varepsilon \in \{0, 1\}$, $k \geq 0$, and $m \geq 0$ such that, for any large enough integer n , the factorization $F \cdot (W_0)^n$ is Hurwitz equivalent to*

$$(W_0)^{n+k} (W_1)^\varepsilon (W_2)^m.$$

Consequently, there are three genus-2 Lefschetz fibrations X_0, X_1, X_2 such that for any genus-2 Lefschetz fibration X the fiber sums $X \#_f n X_0$ and $(n+k)X_0 \#_f \varepsilon X_1 \#_f m X_2$ are equivalent Lefschetz fibrations. \square

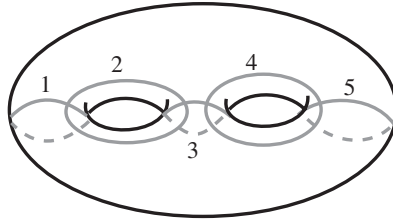


Figure 1

For Lefschetz fibrations of arbitrary genus we do not hope to get a classification result of the above type. In general, Auroux proved the following stable classification result.

Theorem 5.2 (Auroux [5]). *For any g , there exists a genus- g Lefschetz fibration f_g^0 with the following property. Let $f: X \rightarrow S^2$ and $f': X' \rightarrow S^2$ be two genus- g Lefschetz fibrations, each equipped with a distinguished section. Assume that*

- (1) *the 4-manifolds X and X' have the same Euler characteristic and signature,*
- (2) *the distinguished sections of f and f' have the same self intersection,*
- (3) *f and f' have the same number of reducible fibers of each type, i.e., the fibrations have identical combinatorial data.*

Then, for all large enough values of n , the fiber sums $f \# n f_g^0$ and $f' \# n f_g^0$ are isomorphic. □

5.2 Holomorphicity of Lefschetz fibrations

A Lefschetz fibration is called *holomorphic* if the total space X is a complex surface, and for a suitable complex structure on the base 2-manifold Σ the fibration map $f: X \rightarrow \Sigma$ is holomorphic. Fiber sums of holomorphic Lefschetz fibrations might not be holomorphic, as the examples provided by the following theorem show:

Theorem 5.3 ([39], [30]). *For each $g \geq 2$, there exist infinitely many genus- g Lefschetz fibrations with base S^2 admitting no complex structure with either orientation. These Lefschetz fibrations can be chosen to be fiber sums of holomorphic Lefschetz fibrations.* □

By definition, a Lefschetz fibration has *transitive monodromy* if its monodromy generates the mapping class group of a general fiber.

Theorem 5.4 (Siebert–Tian [43]). *Let $f: X \rightarrow S^2$ be a genus-2 Lefschetz fibration with transitive monodromy. If all singular fibers are irreducible, then f is isomorphic to a holomorphic Lefschetz fibration.* \square

As a corollary of Theorem 5.1 we get

Corollary 5.5 (Auroux [4]). *Let $f: X \rightarrow S^2$ be a genus-2 Lefschetz fibration. Then the fiber sum of f with sufficiently many copies of the rational genus-2 Lefschetz fibration with 20 singular fibers whose monodromy group is W_0 is isomorphic to a holomorphic fibration.* \square

5.3 Commutator lengths of Dehn twists

Let G be a group and let $[G, G]$ denote the commutator subgroup of G , i.e., the subgroup generated by all commutators $[a, b] = aba^{-1}b^{-1}$ for $a, b \in G$. For $x \in [G, G]$, let us define the *commutator length* $c(x)$ of x to be the minimum number of factors needed to express x as a product of commutators. Clearly, the commutator length is subadditive:

$$c(x^{n+m}) \leq c(x^n) + c(x^m)$$

for all x in $[G, G]$. It can be shown easily that the limit

$$\|x\| = \lim_{n \rightarrow \infty} \frac{c(x^n)}{n}$$

exists, which is called the *stable commutator length* of x . Recall that G is called *perfect* if $G = [G, G]$ and *uniformly perfect* if there is a positive integer N such that every element x of G can be written as a product of at most N commutators ($c(x) \leq N$). Notice that if G is uniformly perfect, then $\|x\| = 0$ for all $x \in G$.

It is a well-known fact that for $g \geq 3$ the mapping class group Γ_g is perfect [41], i.e., every element in Γ_g is a product of commutators. In particular, one can talk about the stable commutator length of a Dehn twist. The mapping class group Γ_2 is not perfect; its first homology group is isomorphic to the cyclic group of order 10 generated by the class of a Dehn about a nonseparating simple closed curve [38]. Thus the tenth power of every Dehn twist is in the commutator subgroup.

Theorem 5.6 ([17], [29]). *Let a be a homotopically nontrivial simple closed curve on a closed connected oriented surface Σ_g of genus $g \geq 2$, and let t_a be the Dehn twist about a . Then $\|t_a\| \geq \frac{1}{18g-6}$ if $g \geq 3$ and $\|t_a^{10}\| \geq \frac{1}{3}$ if $g = 2$.* \square

In the proof of this theorem, the following symplectic Parshin–Arakelov inequality (proved by Tian–Jun Li) was used:

Theorem 5.7 ([34]). *Let X be a relatively minimal genus- g Lefschetz fibration over a genus- h surface. If X is not rational or ruled, then the square of the first Chern class*

$c_1(X) \in H^2(X; \mathbb{R})$ of X satisfies

$$c_1^2(X) \geq 2(g-1)(h-1). \quad \square$$

Recall that a 4-manifold is called *rational 4-manifold* if it is diffeomorphic either to $S^2 \times S^2$ or to some blow-up of the complex projective plane $\mathbb{C}P^2$. A 4-manifold is *ruled* if it is diffeomorphic to some blow-up of a sphere bundle over an orientable surface.

Theorem 5.8 ([9]). *Let a be a homotopically nontrivial simple closed curve on a closed connected oriented surface Σ_g of genus $g \geq 2$, and let t_a be the Dehn twist about a . Suppose that t_a^k with $k > 0$ can be written as a product of h commutators. Then*

$$h \geq 1 + \frac{k}{6(3g-1)}. \quad \square$$

We outline the proof of a slightly weaker result, namely that there are elements in the mapping class group for which the powers admit unbounded commutator length.

Theorem 5.9 (Endo–Kotschick [17], Korkmaz [29]). *Let c be a separating simple closed curve on a closed connected oriented surface Σ_g of genus $g \geq 2$. If $t_c^n = \prod_{i=1}^{k_n} [\alpha_i(n), \beta_i(n)]$ in Γ_g then the sequence $\{k_n\}$ cannot be bounded. In conclusion, the mapping class group Γ_g is not uniformly perfect.*

Proof. Notice that a commutator expression of the type of the theorem gives a relator which gives rise to a genus- g Lefschetz fibration $X_n \rightarrow \Sigma_{k_n}$ over a closed surface of genus k_n with n reducible vanishing cycles. Suppose that $\{k_n\}$ is bounded, say $k_n \leq K$. By adding trivial monodromies if necessary, this assumption provides a sequence $f_n: X_n \rightarrow \Sigma_K$ ($n \in \mathbb{N}$) of Lefschetz fibrations over the fixed base Σ_K . It is easy to see that the Euler characteristic of X_n is

$$\chi(X_n) = \chi(\Sigma_g)\chi(\Sigma_K) + n = 4(K-1)(g-1) + n,$$

while by Novikov additivity and the signature calculation for a regular neighborhood of a separating vanishing cycle we get

$$\sigma(X_n) = -n + \sigma\left(X_n - \bigcup_{i=1}^n \nu f_n^{-1}(q_i)\right),$$

where the points q_i denote the critical values of the Lefschetz fibration f_n . On the other hand, one can show that

$$\sigma\left(X_n - \bigcup_{i=1}^n \nu f_n^{-1}(q_i)\right) \leq C$$

for some constant C depending on K and g only, since the second Betti number of $X_n - \bigcup_{i=1}^n \nu f_n^{-1}(q_i)$ is bounded. This implies that

$$c_1^2(X_n) = 3\sigma(X_n) + 2\chi(X_n) \leq -3n + 2n + C' = -n + C',$$

where $C' = 3C + 8(K - 1)(g - 1)$, and hence for n large enough the expression $c_1^2(X_n)$ will be negative. This observation contradicts the result of [50] where it is proved that a relatively minimal Lefschetz fibration over a base of positive genus is minimal, hence by a famous result of Taubes [51] its c_1^2 invariant is nonnegative. The contradiction shows that the sequence k_n is unbounded, verifying the statement of the theorem. \square

As a consequence of Theorem 5.6, one can deduce the fact that the growth rate of a Dehn twist in the mapping class group is linear.

Theorem 5.10 ([20], [29]). *Let a be a simple closed curve not bounding a disk on a connected oriented surface of genus $g \geq 2$. Let $d(t_a^n, 1)$ denote the word length of t_a^n with respect to a fixed finite generating set of the mapping class group. Then the limit*

$$\lim_{n \rightarrow \infty} \frac{d(t_a^n, 1)}{n}$$

is positive. \square

The mapping class group of a closed oriented surface can be generated by two torsion elements [27]. For a mapping class f , let $\tau(f)$ denote the *torsion length* of f , the least number of factors needed to express f as a product of torsion elements. The limit

$$\|f\|_\tau = \lim_{n \rightarrow \infty} \frac{\tau(f^n)}{n}$$

is called the *stable torsion length* of f .

Theorem 5.11 ([28]). *Let a be a homotopically nontrivial simple closed curve on a closed connected oriented surface of genus at least three. Then the stable torsion length of the Dehn twist t_a is positive.* \square

5.4 Minimal number of singular fibers

For two nonnegative integers g and h , let $N(g, h)$ denote the minimal number of singular fibers in all relatively minimal Lefschetz fibrations of base genus h and fiber genus g , having at least one singular fiber. By taking the fiber sum with the trivial Lefschetz fibration, it is easy to see that $N(g, h) \geq N(g, h + 1)$.

Theorem 5.12. *For the number $N(g, h)$ the following holds.*

- (1) $N(g, h) = 1$ if and only if $g \geq 3$ and $h \geq 2$.
- (2) $N(g, 1) > 1$ for all $g \geq 1$.
- (3) $N(1, h) = 12$ for all $h \geq 0$.
- (4) $5 \leq N(2, h) \leq 8$ for all h .
- (5) $N(g, 0)$ is less than or equal to $2g + 4$ if g is even and $2g + 10$ if g is odd.

(6) $N(g, 0) \geq \frac{4}{5}(2g - 1)$. □

The parts (1)-(4) were proved by the first author and Burak Ozbagci in [32], (5) follows from [30] and (6) was proved by the second author in [49]. For most Lefschetz fibrations whose total space have $b_2^+ = 1$, we have the precise value of $N(g, 0)$.

Theorem 5.13 (Stipsicz [46]). *Suppose that $X \rightarrow S^2$ is a nontrivial genus- g Lefschetz fibration with $b_2^+ = 1$.*

- (1) *If g is even and $g \geq 6$ then $X \rightarrow S^2$ contains at least $2g + 4$ singular fibers. This lower bound is sharp.*
- (2) *If g is odd and $g \geq 15$ then $X \rightarrow S^2$ contains at least $2g + 10$ singular fibers. This lower bound is sharp.* □

Consider the closed oriented surface Σ_g shown in Figure 2. Let W_g denote the word $(t_{B_0}t_{B_1}t_{B_2} \dots t_{B_g}t_c)^2$ if g is even and $(t_{B_0}t_{B_1}t_{B_2} \dots t_{B_g}t_a^2t_b^2)^2$ if g is odd. It was shown in [30] that W_g represents the identity element in the mapping class group Γ_g (the $g = 2$ case of this fact was proved by Matsumoto in [36]). Let X_g be the total space of the Lefschetz fibration over S^2 with monodromy $W_g = 1$. Then $b_2^+(X_g) = 1$. Thus the lower bounds in (1) and (2) in Theorem 5.13 are attained by X_g . This also proves (5) in Theorem 5.12.

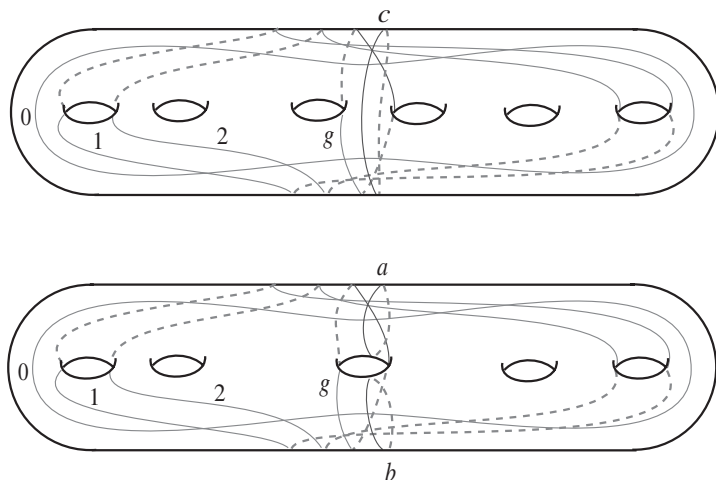


Figure 2. The simple closed curve labelled i on the surface Σ_g is B_i . The genus of the top surface is even and the genus of the bottom surface is odd.

We now show that $N(2, h) = 5$ for $h \geq 6$. This fact is new.

Lemma 5.14. *Let a, b, x, y be four simple closed curves on an oriented surface Σ . Suppose that a intersects x transversely at one point and b intersects y transversely at one point. Then the product $t_a t_b t_y^{-1} t_x^{-1}$ is a commutator in the mapping class group of Σ .*

Proof. In order to avoid double indices, let A, B, X, Y denote the Dehn twists t_a, t_b, t_x, t_y about a, b, x, y respectively. Since the curves $X(b) = t_x(b)$ and $X(y) = t_x(y)$ intersect transversely at one point, by the classification of surfaces there exists a diffeomorphism f of the surface such that $f(x) = X(b)$ and $f(a) = X(y)$. Then

$$\begin{aligned} t_a t_b t_y^{-1} t_x^{-1} &= A B Y^{-1} X^{-1} = A X^{-1} B^X (Y^{-1})^X = t_a t_x^{-1} t_{X(b)} t_{X(y)}^{-1} \\ &= t_a t_x^{-1} t_{f(x)} t_{f(a)}^{-1} = t_a t_x^{-1} f t_x t_a^{-1} f^{-1} = [t_a t_x^{-1}, f]. \quad \square \end{aligned}$$

Here, we use the notation g^h for the conjugation $h g h^{-1}$ in a group, and the fact that $f t_x f^{-1} = t_{f(x)}$ for any simple closed curve x and for any element f in the mapping class group.

Theorem 5.15. *Let c be a simple closed curve on a closed oriented surface Σ_2 of genus 2 splitting the surface into two genus-1 subsurfaces and let C denote the Dehn twist about c . Then C^5 is a product of six commutators.*

Proof. We can assume that c is as shown in Figure 2 for $g = 2$. Consider the simple closed curves B_0, B_1, B_2 on Σ_2 . Let X, Y, Z denote the Dehn twists $t_{B_0}, t_{B_1}, t_{B_2}$ respectively. The relation $W_2 = 1$ then gives

$$X Y Z C X Y Z C = 1.$$

Thus

$$X Y Z X^C Y^C Z^C C^2 = 1,$$

or

$$C^2 = Z_1^{-1} Y_1^{-1} X_1^{-1} Z^{-1} Y^{-1} X^{-1},$$

where $X_1 = X^C, Y_1 = Y^C, Z_1 = Z^C$. We also know that $C = (AB)^6$ where A and B are the Dehn twists about the curves labelled 1 and 2 respectively in Figure 1. Here, X^C denotes the conjugation $C X C^{-1}$, and similarly for the others. Hence,

$$\begin{aligned} C^5 &= (AB)^6 C^4 \\ &= ((AB)^3 C^2)^2 \\ &= ((AB)^3 Z_1^{-1} Y_1^{-1} X_1^{-1} Z^{-1} Y^{-1} X^{-1})^2. \end{aligned}$$

Now it suffices to show that $(AB)^3 Z_1^{-1} Y_1^{-1} X_1^{-1} Z^{-1} Y^{-1} X^{-1}$ is a product of three commutators:

$$\begin{aligned}
 & (AB)^3 Z_1^{-1} Y_1^{-1} X_1^{-1} Z^{-1} Y^{-1} X^{-1} \\
 &= ABABAB Z_1^{-1} Y_1^{-1} X_1^{-1} Z^{-1} Y^{-1} X^{-1} \\
 &= AABAAB Z_1^{-1} Y_1^{-1} X_1^{-1} Z^{-1} Y^{-1} X^{-1} \\
 &= (AB Z_1^{-1} Y_1^{-1})^{AAB A} (BA X_1^{-1} Z^{-1})^{AA} (AA Y^{-1} X^{-1}).
 \end{aligned}$$

Now the theorem follows from Lemma 5.14. \square

Corollary 5.16. For $h \geq 6$, $N(2, h) = 5$. \square

5.5 Fundamental groups of Lefschetz fibrations

Every finitely presented group is the fundamental group of some closed 4-manifold. Gompf proved in [22] that every finitely presented group is the fundamental group of some closed symplectic 4-manifold.

By Theorem 4.2, if X is a closed symplectic 4-manifold, then it admits a Lefschetz pencil structure. Thus, there exists an integer $k \geq 0$ such that the manifold $X \# k\overline{\mathbb{C}\mathbb{P}^2}$ admits the structure of a Lefschetz fibration (with base S^2). It follows that every finitely presented group is the fundamental group of the total space of some Lefschetz fibration. Another construction of a Lefschetz fibration whose fundamental group is a given finitely presented group was given by Amorós, Bogomolov, Katzarkov and Pantev in [3]. In their construction, the genus of the Lefschetz fibration is implicit and is quadratic in the number of generators and the lengths of relators of the given finitely presented group.

A further construction of a Lefschetz fibration with a given finitely presented group as its fundamental group was given by the first author in [31]. In this construction it was shown that given a finitely presented group there is a Lefschetz fibration over S^2 such that the fundamental group of the total space of the fibration is isomorphic to the given group. The genus of the Lefschetz fibration depends linearly on the number of generators and the sum of the syllable lengths (defined below) of the relators. Moreover, the monodromy of the fibration is described explicitly, which was not given in the construction in [3]. We will now describe this result of [31]. We start with giving some necessary definitions.

Let G be a finitely generated group generated by a set $A = \{a_1, a_2, \dots, a_n\}$. For an element $w \in G$, let us define the *syllable length* $\ell(w)$ of w to be

$$\ell(w) = \min\{s \mid w = z_1^{m_1} z_2^{m_2} \dots z_s^{m_s}, z_j \in A \cup A^{-1}, m_j \in \mathbb{Z}\}.$$

Suppose that G is a finitely presented group with a presentation

$$G = \langle a_1, a_2, \dots, a_n \mid r_1, r_2, \dots, r_k \rangle,$$

so that G is the quotient F/R , where F is the free group (nonabelian for $n \geq 2$) freely generated by $\{a_1, a_2, \dots, a_n\}$ and R is the normal subgroup of F generated normally by the elements r_1, r_2, \dots, r_k . That is, R is the subgroup of F generated by all conjugates of r_1, r_2, \dots, r_k . Define $\ell = \ell(r_1) + \ell(r_2) + \dots + \ell(r_k)$, which, of course, depends on the presentation.

Theorem 5.17 (Korkmaz [31]). *Let G be a finitely presented group with a presentation as above. Then for every $g \geq 2(n + \ell - k)$ there exists a genus- g Lefschetz fibration $X \rightarrow S^2$ such that $\pi_1(X)$ is isomorphic to G . \square*

Although every finitely presented group is the fundamental group of some Lefschetz fibration, the situation changes drastically under the absence of reducible vanishing cycles and the assumption that the monodromy is contained in the hyperelliptic mapping class group. Recall that the *hyperelliptic mapping class group* of a surface is defined to be the centralizer of a hyperelliptic involution. A Lefschetz fibration is called *hyperelliptic* if its monodromy is contained in the hyperelliptic mapping class group.

Theorem 5.18 (Siebert–Tian [42]). *Nontrivial hyperelliptic Lefschetz fibrations of genus $g \geq 2$ without reducible vanishing cycles are simply connected. \square*

5.6 Sections of Lefschetz fibrations

It is not clear which Lefschetz fibrations admit sections. Note that if $f: X' \rightarrow \mathbb{C}P^1$ originates from a Lefschetz pencil on some 4-manifold X and if the base locus of this pencil is nonempty, then f admits sections: the exceptional spheres of the blow-ups of points in the base locus intersect each fiber exactly once. Since these sections are exceptional spheres, their homological self-intersections are equal to (-1) . In fact, for the homological square of a section we have

Theorem 5.19 ([47], [48], [45]). *If $f: X \rightarrow S^2$ is a nontrivial Lefschetz fibration on a 4-manifold X with fiber genus $g > 0$ admitting a section $\sigma: S^2 \rightarrow X$, then $[\sigma(S^2)]^2 < 0$.*

Proof (sketch). If a Lefschetz fibration $f: X \rightarrow S^2$ admits a section σ with $[\sigma(S^2)] \geq 0$, then by fiber summing f with itself the nontriviality gives a symplectic 4-manifold $X \#_f X$ with $b_2^+(X \#_f X) > 1$ containing a symplectic sphere of nonnegative self-intersection. By Seiberg–Witten theory, however, symplectic 4-manifolds with $b_2^+ > 1$ contain homologically essential spheres of negative self-intersection only, providing the desired contradiction and concluding the proof. \square

Since sections determine special factorizations in mapping class groups of surfaces with marked points, the observation above can be interpreted as restriction on the existence of factorizations into right-handed Dehn twists in certain mapping class groups.

Corollary 5.20. *Let Σ be a compact connected oriented surface of genus g with $n \geq 1$ boundary components and let δ be a simple closed curve parallel to a boundary component. If k is a nonnegative integer then t_δ^{-k} cannot be written as a product of right-handed Dehn twists in the mapping class group Γ_g^n of Σ . In particular, a product of right-handed Dehn twists in the mapping class group Γ_g^n cannot be equal to the identity. \square*

5.7 Surface bundles over surfaces with nonzero signature

The Euler characteristic is multiplicative in fiber bundles. That is, the Euler characteristic of the total space of a fiber bundle is equal to the product of the Euler characteristics of the base and of the fiber. The signature is multiplicative in fiber bundles if the action of the fundamental group of the base is trivial on the cohomology of the fiber [11].

This is not true for every fiber bundle: the first examples of surface bundles over surfaces with nonzero signature were constructed by Atiyah [7] and independently by Kodaira [26]. If the fiber is a sphere or a torus, the signature vanishes. The signature vanishes if the genus of the base surface is 0 or 1 as well. Since $H_2(\Gamma_2; \mathbb{Q}) = 0$ by a theorem of Igusa, the signature also vanishes if the fiber genus is 2. Endo [16] proved that there exists a surface bundle over a surface of genus 111 with fiber genus 3 such that the signature of the total space is -4 . Using subtraction of Lefschetz fibrations, the following improvement was shown:

Theorem 5.21 ([18]). *For any $g \geq 3$ and $h \geq 9$ there is a genus- g surface bundle $f: X \rightarrow \Sigma_h$ such that the signature of X is nonzero. \square*

Bryan and Donagi constructed a surface bundle of nonzero signature over the surface of genus 2.

Theorem 5.22 ([10]). *There exists a 4-manifold X which carries a surface bundle structure $f: X \rightarrow \Sigma_2$ with fiber genus 25 and signature 16. \square*

5.8 Teichmüller spaces

Recall that a Lefschetz fibration $f: X \rightarrow \Sigma$ provides an almost complex structure on X . Suppose now that X is a closed manifold and consider such an almost complex structure J on X . The restriction of J to a fiber $F_t = f^{-1}(t)$ equips (the smooth part of) F_t with an almost complex structure, which (by simple dimension reason) is integrable on F_t . Hence by restricting f to the regular fibers we get a map from the set $\Sigma - Q = \Sigma - \{f(p_1), \dots, f(p_k)\}$ of regular values to the Teichmüller space \mathcal{T} of complex structures on F_t , which extends to a map

$$\Phi: \Sigma \rightarrow \overline{\mathcal{T}},$$

where $\overline{\mathcal{T}}$ is the Deligne–Mumford compactification of \mathcal{T} . The Lefschetz fibration is holomorphic if and only if this corresponding map Φ is holomorphic. Therefore the search for certain Lefschetz fibrations can be reduced to searching for complex (or symplectic) surfaces in $\overline{\mathcal{T}}$ of a given genus. In this picture the intersection of the surface with the compactifying divisor corresponds to the singular fibers of the fibration.

6 Variations of Lefschetz fibrations

There are modifications of the original concept of Lefschetz fibrations which turn out to be very useful in other geometric contexts.

6.1 Achiral Lefschetz fibrations

The most obvious generalization of Definition 2.2 is the relaxation of the orientation-preserving condition for the local charts near the singularities. This concept leads to the definition of *achiral Lefschetz fibrations*. Some parts of the discussion of the previous sections go through verbatim for these objects. A notable (and obvious) exception is the orientation of the surgery and the Dehn twist in the monodromy representation. In fact, the surgery coefficient corresponding to an achiral fiber is not (-1) with respect to the fiber framing, but $(+1)$. Similarly, the monodromy of the surface bundle around an achiral singular fiber is a *left-handed* Dehn twist along the vanishing cycle.

Surprisingly enough, the little modification of the definition destroys the nice geometric correspondence we have for Lefschetz fibrations. The total space of an achiral Lefschetz fibration does not necessarily support a symplectic structure; in fact, it is easy to construct examples when the total space does not admit an almost complex structure either. This can be verified, for example, by finding a genus-1 achiral Lefschetz fibration on the 4-sphere S^4 [24, Figure 8.38].

There are topological obstructions for a manifold to admit an achiral Lefschetz fibration, but by appropriately surgering the manifold along a circle these obstructions vanish, and the resulting 4-manifold will carry an achiral Lefschetz fibration structure (cf. [19]). The concept of achiral Lefschetz fibrations turned out to be extremely useful in studying contact 3-manifolds through their contact surgery presentations.

6.2 Lefschetz fibrations and Stein structures

Definition 6.1. Suppose now that X and Σ are given oriented 4- and 2-dimensional manifolds, with possibly nonempty boundary. The map $f: X \rightarrow \Sigma$ is a *Lefschetz fibration* if df is onto except at finitely many critical points $\{p_1, \dots, p_k\} = C \subset \text{int}X$, f has only Lefschetz critical points, and f is a locally trivial fiber bundle over $\Sigma - f(C)$.

Notice that by dropping the condition $f^{-1}(\partial\Sigma) = \partial X$ some boundary points of X might map to interior points of Σ . In this case the fiber is an oriented *bounded* surface, that is, a 2-manifold with nonempty boundary.

Suppose now that the interior $\text{int}X$ of the compact 4-manifold X (with nonempty boundary) admits a Stein structure with an exhausting plurisubharmonic function having finitely many critical points. Resting on Eliashberg's topological description of such Stein 4-manifolds, Loi and Piergallini [35] (and Akbulut–Ozbagci [2]) showed that such an X admits a Lefschetz fibration over the 2-disk D^2 . Conversely, again by the handle attachment scheme developed by Eliashberg, a Lefschetz fibration with bounded fibers admits a Stein structure.

The algebraic description of Lefschetz fibrations with closed surface fibers extend verbatim to the case of bounded fibers; in this latter case the factorization is considered in the mapping class group of the bounded fiber, that is, diffeomorphisms and isotopies are assumed to be the identity on the boundary. Also, the fibration might not be trivial along the boundary of the base surface, so instead of providing a factorization of the unit element, the fibration over D^2 gives a factorization (into right-handed Dehn twists) of the element given by the monodromy along the boundary S^1 . Using this correspondence, questions regarding topological properties of Stein 4-manifolds can be phrased purely in terms of algebraic properties of various mapping class groups, cf. Section 7.

6.3 Achiral Lefschetz fibrations and contact structures

Recall that a cooriented 2-plane field ξ on an oriented 3-manifold Y is a *contact structure* if ξ can be given as

$$\xi = \ker \alpha$$

for some 1-form $\alpha \in \Omega(Y)$ with the property that $\alpha \wedge d\alpha > 0$ everywhere. In the light of the famous theorem of Frobenius this condition means that ξ is a nowhere integrable 2-plane field in TY .

An *open book decomposition* on Y consists of a link $B \subset Y$ and a fibration $\varphi: Y - B \rightarrow S^1$ with the property that any fiber $\varphi^{-1}(t)$, $t \in S^1$, is a Seifert surface for B . According to the celebrated result of Giroux, an open book decomposition uniquely determines a contact structure (unique up to isotopy), and conversely, any contact structure arises in this way. In fact, the relation between open book decompositions defining isotopic contact structures is also clarified (but we will omit its description here). Notice that the fibration φ is determined by its monodromy h , which is now an element of the mapping class group of the fiber, a surface with boundary. It follows from the assumptions that the element h represents a diffeomorphism equal to the identity along the boundary. Conversely, such an element h determines $Y - B$, and by the triviality of h near the boundary of the fiber, the tori of $\partial(Y - \nu B)$ are canonically framed. Therefore there is a unique way to fill these boundary tori with copies of $S^1 \times D^2$ to get an open book, hence the monodromy uniquely determines the 3-mani-

fold Y , together with an open book decomposition on it. In conclusion, through the open book decomposition, h also determines a contact structure on Y .

As we explained above, a factorization of h into the product of right-handed Dehn twists determines a Lefschetz fibration, which also carries a Stein structure, providing a Stein filling for the contact structure determined by the open book decomposition. In fact, any Stein filling arises in this way, with the caveat that the fiber genus might need to be increased. This latter observation makes it somewhat hard to directly use this correspondence.

Achiral Lefschetz fibrations with bounded fibers can be defined on manifolds with boundary by dropping the orientation-preserving condition for the complex charts containing the Lefschetz singularities – just like in the case we passed from (ordinary) Lefschetz fibrations to (ordinary) achiral Lefschetz fibrations. It is an elementary fact that any mapping class group element factors as a product of Dehn twists (left- and right-handed). Having left-handed Dehn twists in a factorization ruins the possible Stein structure on the 4-manifold, but still provides some kind of filling: by a local analysis the anti-blow-ups (connected sums with complex projective planes) at the achiral Lefschetz critical points provides an almost complex filling of the given contact 3-manifold. Such a filling can be used to determine homotopic properties of contact structures, cf. [12]. A closer look also shows a way to prove the existence of a contact surgery presentation for the contact 3-manifold. We will not address these issues in the present chapter.

6.4 Further generalizations

There are further generalizations of Lefschetz fibrations proposed by Auroux–Donaldson–Katzarkov: we may allow a further type of singularity and get a structure called *broken Lefschetz fibrations*. Allowing achiral Lefschetz singularities as well, we get the notion of broken achiral Lefschetz fibrations. As it is shown in [21], any closed 4-manifold admits such a structure. Indeed, it was shown by Baykur [8] (and independently by Akbulut–Karakurt [1]) that any closed orientable 4-manifold admits a broken Lefschetz fibration. In the broken case the various fibers are not necessarily diffeomorphic (when crossing the additionally allowed singularities, the genus of the typical fiber might jump), hence these structures do not fall within the scope of the present overview.

7 Open problems

In this final section we collect a few open problems we consider to be of particular interest in the subject.

Problem 7.1. For a given symplectic 4-manifold X find the lowest genus Lefschetz pencil on X . What is the relation of this genus to other invariants of X ?

Problem 7.2. Prove (or disprove by a counterexample) that any Lefschetz fibration over S^2 having singular fibers admits a section.

Problem 7.3. Prove (or disprove by a counterexample) that any genus-2 Lefschetz fibration over S^2 has an abelian fundamental group.

Problem 7.4. Show that a Lefschetz fibration over the sphere S^2 has nonpositive signature. In fact, show that if $X \rightarrow S^2$ is a nontrivial Lefschetz fibration, then the signature of X satisfies $\sigma(X) < 0$.

Problem 7.5. More generally, show that for any $(g, h) \in \mathbb{N} \times \mathbb{N}$ there is an integer $S(g, h)$ such that for any genus- g Lefschetz fibration $X \rightarrow \Sigma_h$ over the genus- h surface Σ_h the 4-manifold X has signature $\sigma(X) \leq S(g, h)$. (If such $S(g, h)$ exists, then it is not hard to see that $S(g, 0) = S(g, 1) = 0$, providing an answer to the previous problem.)

Problem 7.6. Let $n > 0$ and let \mathcal{D} denote the submonoid of Γ_g^n consisting of products of right-handed Dehn twists. Verify that for each $h \in \mathcal{D}$ there is a bound T_h such that whenever h is written as a product of k nontrivial right-handed Dehn twists, then $k \leq T_h$. The existence of such a bound would provide a major step toward proving that the Euler characteristics of Stein fillings of a fixed Stein fillable contact 3-manifold are bounded.

Problem 7.7. Determine whether a genus-3 surface bundle over the genus-2 surface with nonzero signature exists or not. Notice that the existence of such a surface bundle would violate the Bogomolov–Miyaoaka–Yau inequality for symplectic 4-manifolds (an inequality known to hold for complex surfaces).

References

- [1] S. Akbulut and C. Karakurt, Every 4-manifold is BLF. Preprint, arxiv:math.0803.2297. [292](#)
- [2] S. Akbulut and B. Ozbagci, Lefschetz fibrations on compact Stein surfaces. *Geom. Topol.* 5 (2001), 319–334. [291](#)
- [3] J. Amorós, F. Bogomolov, L. Katzarkov and T. Pantev, Symplectic Lefschetz fibrations with arbitrary fundamental groups. *J. Differential Geom.* 54 (2000), 489–545. [277](#), [287](#)
- [4] D. Auroux, Fiber sums of genus 2 Lefschetz fibrations. *Turkish J. Math.* 27 (1) (2003), 1–10. [280](#), [282](#)
- [5] D. Auroux, A stable classification of Lefschetz fibrations. *Geom. Topol.* 9 (2005), 203–217. [281](#)
- [6] D. Auroux, S. Donaldson and L. Katzarkov, Singular Lefschetz pencils. *Geom. Topol.* 9 (2005), 1043–1114. [279](#)

- [7] M. F. Atiyah, The signature of fibre-bundles. In *Global Analysis, Papers in Honor of K. Kodaira*, Tokyo University Press, Tokyo 1969, 73–84. [289](#)
- [8] R. I. Baykur, Existence of broken Lefschetz fibrations. *Internat. Math. Res. Notices* 2008 (2008), Article ID rnn101. [292](#)
- [9] V. Braungart and D. Kotschick, Clustering of critical points in Lefschetz fibrations and the symplectic Szpiro inequality. *Trans. Amer. Math. Soc.* 355 (8) (2003), 3217–3226. [283](#)
- [10] J. Bryan and R. Donagi, Surface bundles over surfaces of small genus. *Geom. Topol.* 6 (2002), 59–67. [289](#)
- [11] S. S. Chern, F. Hirzebruch and J. P. Serre, On index of a fibered manifold. *Proc. Amer. Math. Soc.* 8 (1957), 587–596. [289](#)
- [12] F. Ding, H. Geiges and A. Stipsicz, Surgery diagrams for contact 3-manifolds. *Turkish J. Math.* 28 (2004) 41–74. [292](#)
- [13] S. K. Donaldson, Symplectic submanifolds and almost-complex geometry. *J. Differential Geom.* 44 (1996), 666–705. [271](#)
- [14] S. K. Donaldson, Lefschetz fibrations in symplectic geometry. In *Proceedings of the International Congress of Mathematicians* (Berlin, 1998), Vol. II, Doc. Math., Extra Vol. ICM Berlin, 1998, 309–314. [272](#)
- [15] S. K. Donaldson, Lefschetz pencils on symplectic manifolds. *J. Differential Geom.* 53 (1999), 205–236. [272](#)
- [16] H. Endo, A construction of surface bundles over surfaces with non-zero signature. *Osaka J. Math.* 35 (1998), 915–930. [289](#)
- [17] H. Endo and D. Kotschick, Bounded cohomology and non-uniform perfection of mapping class groups. *Invent. Math.* 144 (1) 2001, 169–175. [272](#), [282](#), [283](#)
- [18] H. Endo, M. Korkmaz, D. Kotschick, B. Ozbagci and A. Stipsicz, Commutators, Lefschetz fibrations and signatures of surface bundles. *Topology* 41 (2002), 961–977. [289](#)
- [19] J. Etnyre and T. Fuller, Realizing 4-manifolds as achiral Lefschetz fibrations. *Internat. Math. Res. Notices* 2006 (2006). [290](#)
- [20] B. Farb, A. Lubotzky and Y. Minsky, Rank one phenomena for mapping class groups. *Duke Math. J.* 106 (3) (2001), 581–597. [284](#)
- [21] D. Gay and R. Kirby, Constructing Lefschetz-type fibrations on four-manifolds. *Geom. Topol.* 11 (2007), 2075–2115. [292](#)
- [22] R. E. Gompf, A new construction of symplectic manifolds. *Ann. of Math.* (2) 142 (3) (1995), 527–595. [287](#)
- [23] R. Gompf, The topology of symplectic manifolds. *Turkish J. Math.* 25 (2001) 43–59. [272](#), [278](#)
- [24] R. E. Gompf and A. I. Stipsicz, *4-manifolds and Kirby calculus*. Grad. Stud. Math. 20, Amer. Math. Soc., Providence; R.I., 1999. [272](#), [278](#), [290](#)
- [25] P. de la Harpe, *Topics in geometric group theory*. Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 2000.
- [26] K. Kodaira, A certain type of irregular algebraic surfaces. *J. Analyse Math.* 19 (1967), 207–215. [289](#)

- [27] M. Korkmaz, Generating the surface mapping class group by two elements. *Trans. Amer. Math. Soc.* 357 (2005), 3299–3310. [284](#)
- [28] M. Korkmaz, On stable torsion length of a Dehn twist. *Math. Res. Lett.* 12 (2005), 335–339. [284](#)
- [29] M. Korkmaz, Stable commutator length of a Dehn twist. *Michigan Math. J.* 52 (2004), 23–31. [272](#), [282](#), [283](#), [284](#)
- [30] M. Korkmaz, Noncomplex smooth 4-manifolds with Lefschetz fibrations. *Internat. Math. Res. Notices* 2001 (2001), 115–128. [281](#), [285](#)
- [31] M. Korkmaz, Lefschetz fibrations and an invariant of finitely presented groups. *Internat. Math. Res. Notices*, to appear. [287](#), [288](#)
- [32] M. Korkmaz and B. Ozbagci, Minimal number of singular fibers in a Lefschetz fibration. *Proc. Amer. Math. Soc.* 129 (2001), 1545–1549. [285](#)
- [33] D. Kotschick, Four-manifold invariants of finitely presentable groups. In *Topology, geometry and field theory*, World Sci. Publishing, River Edge, NJ, 1994, 89–99.
- [34] T. J. Li, Symplectic Parshin-Arakelov inequality. *Internat. Math. Res. Notices* 2000 (2000), 941–954. [282](#)
- [35] A. Loi and R. Piergallini, Compact Stein surfaces with boundary as branched covers of B^4 . *Invent. Math.* 143 (2001), 325–348. [291](#)
- [36] Y. Matsumoto, Lefschetz fibrations of genus two—a topological approach. In *Topology and Teichmüller spaces* (Katinkulta, 1995), World Sci. Publishing, River Edge, NJ, 1996, 123–148. [285](#)
- [37] B. Moishezon, *Complex surfaces and connected sums of complex projective planes*. Lecture Notes in Math. 603, Springer-Verlag, Berlin 1977. [280](#)
- [38] D. Mumford, Abelian quotients of the Teichmüller modular group. *J. Analyse Math.* 18 (1967), 227–244. [282](#)
- [39] B. Ozbagci and A. I. Stipsicz, Noncomplex smooth 4-manifolds with genus-2 Lefschetz fibrations. *Proc. Amer. Math. Soc.* 128 (2000), 3125–3128. [274](#), [281](#)
- [40] B. Ozbagci and A. I. Stipsicz, Contact 3-manifolds with infinitely many Stein fillings. *Proc. Amer. Math. Soc.* 132 (2004), 1549–1558.
- [41] J. Powell, Two theorems on the mapping class group of a surface. *Proc. Amer. Math. Soc.* 68 (1978), 347–350. [282](#)
- [42] B. Siebert and G. Tian, On hyperelliptic C^∞ -Lefschetz fibrations of four manifolds. *Commun. Contemp. Math.* 1 (1999), 255–280. [288](#)
- [43] B. Siebert and G. Tian, On the holomorphicity of genus two Lefschetz fibrations. *Ann. of Math.* 161 (2005), 955–1016 [282](#)
- [44] I. Smith, Lefschetz fibrations and the Hodge bundle. *Geom. Topol.* 3 (1999), 211–233.
- [45] I. Smith, Geometric monodromy and the hyperbolic disc. *Q. J. Math.* 52 (2001), 217–228. [288](#)
- [46] A. I. Stipsicz, Singular fibers in Lefschetz fibrations on manifolds with $b_2^+ = 1$. *Topology Appl.* 117 (2002), 9–21. [285](#)
- [47] A. I. Stipsicz, Indecomposability of certain Lefschetz fibrations. *Proc. Amer. Math. Soc.* 129 (2001), 1499–1502. [288](#)

- [48] A. I. Stipsicz, Sections of Lefschetz fibrations and Stein fillings. *Turkish J. Math.* 25 (2001), 97–101. [288](#)
- [49] A. I. Stipsicz, On the number of vanishing cycles in Lefschetz fibrations. *Math. Res. Lett.* 6 (1999), 449–456. [277](#), [285](#)
- [50] A. I. Stipsicz, Chern numbers of certain Lefschetz fibrations. *Proc. Amer. Math. Soc.* 128 (6) (2000), 1845–1851. [284](#)
- [51] C. Taubes, $GR = SW$: counting curves and connections. *J. Differential Geom.* 52 (9) (1999), 453–609. [284](#)

Chapter 9

Introduction to measurable rigidity of mapping class groups

Yoshikata Kida

Contents

1	Introduction	297
2	Preliminaries	303
	2.1 Mapping class groups	303
	2.2 Discrete measured groupoids	308
	2.3 ME and isomorphism of groupoids	316
3	ME rigidity for mapping class groups	320
	3.1 Reduction of self ME couplings of mapping class groups	320
	3.2 Deriving ME rigidity from reduction of self ME couplings	327
	3.3 OE rigidity	329
4	Amenable discrete measured groupoids	331
5	Two types of subgroupoids: IA and reducible ones	340
	5.1 IA and reducible subgroupoids	341
	5.2 IA subgroupoids	343
	5.3 Reducible subgroupoids	346
6	Normal subgroupoids	349
	6.1 Generalities	349
	6.2 Normalizers of IA and reducible subgroupoids	350
7	Characterization of reducible subgroupoids	352
8	Concluding remarks	361
	References	364

1 Introduction

A subgroup of a locally compact second countable group G is said to be a *lattice* if it is discrete in G and has cofinite measure with respect to the Haar measure on G . Classically, it is a basic problem to consider whether a lattice in a Lie group determines its ambient Lie group. More precisely, when Γ and Λ are lattices in Lie groups G and H , respectively, it is an interesting problem to study when the existence of an isomorphism between Γ and Λ implies the existence of an isomorphism between G and H . The Mostow–Prasad–Margulis rigidity theorem gives a complete answer to this question for semisimple Lie groups in a more sophisticated form.

In this chapter, by a discrete group we mean a discrete and countable group. Let us consider the following question: Given two discrete groups, under what conditions can they be realized as lattices in the same locally compact second countable group? Suppose that discrete groups Γ, Λ are lattices in a locally compact second countable group G . We shall observe some consequences of this situation. Consider the action of $\Gamma \times \Lambda$ on G given by

$$(\gamma, \lambda)g = \gamma g \lambda^{-1}, \quad \gamma \in \Gamma, \lambda \in \Lambda, g \in G.$$

It is easy to check the following:

- The actions $\Gamma (\simeq \Gamma \times \{e\}) \curvearrowright G$ and $\Lambda (\simeq \{e\} \times \Lambda) \curvearrowright G$ are both measure-preserving with respect to the (left) Haar measure m on G . Note that the existence of a lattice in G implies the invariance of m under right multiplication by each element of G .
- The action $\Gamma \curvearrowright G$ is free and admits a *fundamental domain* of finite measure, i.e., a Borel subset $F \subset G$ such that $m(F) < \infty$, $\bigcup_{\gamma \in \Gamma} \gamma F = G$, and $m(\gamma_1 F \cap \gamma_2 F) = 0$ for any distinct $\gamma_1, \gamma_2 \in \Gamma$. We can say the same thing for the action $\Lambda \curvearrowright G$.

In a more general situation than the above one, Gromov introduced the notion of measure equivalence as follows.

Definition 1.1 ([23, 0.5.E]). We say that two discrete groups Γ and Λ are *measure equivalent (ME)* if there exists a measure-preserving action of $\Gamma \times \Lambda$ on a standard Borel space (Σ, m) with a σ -finite positive measure such that both of the actions $\Gamma (\simeq \Gamma \times \{e\}) \curvearrowright \Sigma$ and $\Lambda (\simeq \{e\} \times \Lambda) \curvearrowright \Sigma$ are essentially free and admit a fundamental domain of finite measure. The space (Σ, m) (equipped with the $(\Gamma \times \Lambda)$ -action) is then called an *ME coupling* of Γ and Λ .

A *standard Borel space* is a Borel space arising from a separable complete metric space (see [34] for details of standard Borel spaces). An action of a discrete group on a measure space is said to be *essentially free* if the stabilizers of almost all points are trivial.

It is easy to see that ME defines an equivalence relation among discrete groups (see Section 2 in [16] or Remark 3.8 in this chapter). In the study of ME, it is fundamental to classify various discrete groups up to ME and to determine completely the class consisting of all discrete groups ME to a given group. We give three typical examples of ME couplings.

Example 1.2. Let G be a locally compact second countable group equipped with the Haar measure and let Γ, Λ be lattices in G . The action of $\Gamma \times \Lambda$ on G given by

$$(\gamma, \lambda)g = \gamma g \lambda^{-1}, \quad \gamma \in \Gamma, \lambda \in \Lambda, g \in G$$

defines an ME coupling of Γ and Λ .

Example 1.3. This is a special case of the above example. Let Γ be a discrete group and let Λ be a finite index subgroup of Γ . The action of $\Gamma \times \Lambda$ on Γ given by

$$(\gamma, \lambda)\gamma' = \gamma\gamma'\lambda^{-1}, \quad \gamma, \gamma' \in \Gamma, \lambda \in \Lambda$$

defines an ME coupling of Γ and Λ , where the measure on Γ is the counting one.

Example 1.4. Let Γ be a discrete group and let N be a finite normal subgroup of Γ . Choose an essentially free, measure-preserving action of Γ on a standard Borel space X with a finite positive measure (e.g., the Bernoulli action $\Gamma \curvearrowright \prod_{\Gamma}[0, 1]$ when Γ is infinite). Then the action of $\Gamma \times (\Gamma/N)$ on $X \times (\Gamma/N)$ given by

$$(\gamma, \lambda)(x, \lambda') = (\gamma x, p(\gamma)\lambda'\lambda^{-1}), \quad \gamma \in \Gamma, \lambda, \lambda' \in \Gamma/N, x \in X$$

defines an ME coupling of Γ and Γ/N , where $p: \Gamma \rightarrow \Gamma/N$ is the quotient homomorphism. Note that we can find a fundamental domain F for the action $N \curvearrowright X$ since N is finite. It is easy to see that $F \times \{eN\} \subset \Gamma \times (\Gamma/N)$ is a fundamental domain for the action $\Gamma \times (\Gamma/N) \curvearrowright X \times (\Gamma/N)$.

Commensurability up to finite kernels is the equivalence relation for discrete groups defined by declaring two groups in an exact sequence $1 \rightarrow A \rightarrow B \rightarrow C \rightarrow 1$ of discrete groups to be equivalent if the third group is finite. It follows from the last two examples that two commensurable groups up to finite kernels are ME. In particular, all finite groups are ME. Conversely, it is easy to see that a discrete group ME to a finite group is also finite.

Measure equivalence can be viewed as a measure-theoretic analogue of quasi-isometry (QI) between finitely generated groups. It is known that two finitely generated groups are QI if and only if there exists a continuous $(\Gamma \times \Lambda)$ -action on some locally compact space Ω such that both of the actions of Γ ($\simeq \Gamma \times \{e\}$) and Λ ($\simeq \{e\} \times \Lambda$) on Ω are properly discontinuous and cocompact (see 0.2.C in [23]). On the other hand, there are examples of two ME groups which are not QI, and examples of two QI groups which are not ME. For example, \mathbb{Z} and \mathbb{Z}^2 are ME but not QI (see Theorem 1.8). It is known that Kazhdan's property (T) is invariant under ME (see Corollary 1.4 in [17]) and that there exist two finitely generated groups Γ_1, Γ_2 satisfying the following (see Section 3.6 in [6]): Γ_1 and Γ_2 are QI; and Γ_1 satisfies property (T), whereas Γ_2 does not satisfy property (T). Hence, Γ_1 and Γ_2 are not ME.

ME rigidity for mapping class groups. In this chapter, we study mapping class groups of compact orientable surfaces from the viewpoint of ME, and consider a locally compact second countable group containing a lattice isomorphic to mapping class groups.

Let M be a connected compact orientable surface of genus g and with p boundary components. The mapping class group $\Gamma(M)$ of M is defined to be the group of isotopy classes of all orientation-preserving diffeomorphisms of M . Let $\Gamma(M)^\diamond$ be the extended mapping class group of M , i.e., the group of isotopy classes of all

diffeomorphisms of M . The group $\Gamma(M)^\diamond$ contains $\Gamma(M)$ as a subgroup of index 2. Let $\kappa(M) = 3g + p - 4$ be the complexity of M and assume that $\kappa(M) > 0$. Let $C = C(M)$ be the curve complex of M , on which $\Gamma(M)^\diamond$ naturally acts (see Definition 2.1). We denote by $\text{Aut}(C)$ the automorphism group of the simplicial complex C . It is known that the kernel of the natural homomorphism $\pi : \Gamma(M)^\diamond \rightarrow \text{Aut}(C)$ and the index $[\text{Aut}(C) : \pi(\Gamma(M)^\diamond)]$ are both finite (see Theorem 2.3). Our first aim in this chapter is to survey the proof of the following rigidity theorem for $\Gamma(M)$, which completely determines the class of discrete groups ME to $\Gamma(M)$.

Theorem 1.5 ([36, Theorem 1.1]). *Let M be a surface with $\kappa(M) > 0$. If a discrete group Λ is ME to the mapping class group $\Gamma(M)$, then there exists a homomorphism $\rho : \Lambda \rightarrow \text{Aut}(C)$ such that the kernel of ρ and the index $[\text{Aut}(C) : \rho(\Lambda)]$ are both finite.*

Our second aim is to survey the proof of the following theorem, which determines all locally compact second countable groups containing a lattice isomorphic to mapping class groups. The idea of this work relies on Furman's paper [18] about the same problem for higher rank lattices. To the best of our knowledge, there exists no natural topological group containing the mapping class group as a lattice other than the mapping class group itself. The following theorem assures this observation. It has already been known that the mapping class group for a surface with positive complexity is not isomorphic to a lattice in any semisimple Lie group, by a result due to Kaimanovich and Masur [33]. (They also showed that any sufficiently large subgroup of the mapping class group is not isomorphic to a lattice in a semisimple Lie group with real rank at least 2.)

Theorem 1.6 ([36, Theorem 1.4]). *Let M be a surface with $\kappa(M) > 0$ and let Γ be a finite index subgroup of $\Gamma(M)^\diamond$. Let G be a locally compact second countable group and let $\sigma : \Gamma \rightarrow G$ be a lattice embedding, that is, σ is an injective homomorphism such that $\sigma(\Gamma)$ is a lattice in G . Then the following assertions hold:*

- (i) *There exists a continuous homomorphism $\Phi_0 : G \rightarrow \text{Aut}(C)$ such that $\Phi_0(\sigma(\gamma)) = \pi(\gamma)$ for any $\gamma \in \Gamma$, where $\pi : \Gamma(M)^\diamond \rightarrow \text{Aut}(C)$ is the natural homomorphism.*
- (ii) *Let K be the kernel of Φ_0 and let Γ act on K by conjugation via σ . Let further $\rho : \Gamma \rtimes K \rightarrow G$ be the homomorphism defined by $\rho(k) = k$ for $k \in K$ and $\rho(\gamma) = \sigma(\gamma)$ for $\gamma \in \Gamma$. Then the kernel of ρ and the index $[G : \rho(\Gamma \rtimes K)]$ are both finite.*

In particular, G admits infinitely many connected components, and $\sigma(\Gamma)$ is cocompact in G .

This theorem says that there exists no interesting lattice embedding of the mapping class group into a locally compact second countable group.

Amenability of the action $\Gamma(M)^\diamond \curvearrowright \partial C$. This property plays an important role in the proof of Theorems 1.5 and 1.6. Let M be a surface with $\kappa(M) \geq 0$. It is known that the curve complex $C = C(M)$ is a hyperbolic metric space in the sense of Gromov, by a result due to Masur and Minsky [43]. See also Hamenstädt's proof in Volume I of this Handbook [27]. Hence, we can construct the Gromov boundary ∂C of C , which is known to be non-empty. Then $\Gamma(M)^\diamond$ acts on ∂C continuously with respect to the topology on ∂C as the Gromov boundary of C . It can be shown that ∂C is a standard Borel space with respect to the σ -field of subsets of ∂C generated by this topology (see Proposition 3.10 in [35]). We refer to [39], [24], [27] for more details of the boundary ∂C . The action $\Gamma(M)^\diamond \curvearrowright \partial C$ admits the following remarkable property:

Theorem 1.7 ([35, Theorem 3.29]). *Let M be a surface with $\kappa(M) \geq 0$ and let C be the curve complex for M . Let μ be a probability measure on the Gromov boundary ∂C such that the action of $\Gamma(M)^\diamond$ on $(\partial C, \mu)$ is non-singular. Then the action $\Gamma(M)^\diamond \curvearrowright (\partial C, \mu)$ is amenable (in a measurable sense).*

Here, when we are given a Borel action of a discrete group Γ on a Borel space S equipped with a positive measure ν , we say that the action $\Gamma \curvearrowright (S, \nu)$ is *non-singular* if $\nu(\gamma A) = 0$ for any $\gamma \in \Gamma$ and any Borel subset A of S with $\nu(A) = 0$. Amenability of group actions on measure spaces was first introduced by Zimmer [62] as a generalization of amenability of groups. Once it is shown that some action of a group is amenable, there are many applications to the study of that group from various aspects (see Section 8). In Section 4, we discuss the notion of amenable actions of groups and Theorem 1.7. We will apply Theorem 1.7 in the proof of Theorem 5.10 to show that IA subgroupoids are amenable.

Short description of history. The first magnificent result on ME is due to Ornstein and Weiss. Following Dye's results [11], [12] on some amenable groups from the viewpoint of orbit equivalence, Ornstein and Weiss obtained the following result (see Section 4 for the definition and elementary facts about amenable groups).

Theorem 1.8 ([52]). *An infinite discrete group is ME to \mathbb{Z} if and only if it is amenable. In particular, all infinite solvable groups are ME to each other.*

It is natural to consider lattices in various Lie groups from the viewpoint of ME because of Example 1.2. Based on Zimmer's cocycle superrigidity theorem [63], Furman established the following rigidity result for higher rank lattices.

Theorem 1.9 ([16]). *Let G be a connected simple Lie group of non-compact type with finite center and real rank at least 2. Let Γ be a lattice in G . If a discrete group Λ is ME to Γ , then there exists a homomorphism $\rho: \Lambda \rightarrow \text{Aut}(\text{Ad } G)$ such that $\ker \rho$ is finite and $\rho(\Lambda)$ is a lattice in $\text{Aut}(\text{Ad } G)$.*

Note that the kernel of the natural composed map $G \rightarrow \text{Ad } G \rightarrow \text{Aut}(\text{Ad } G)$ and the index of the image of G in $\text{Aut}(\text{Ad } G)$ are both finite. Thanks to this result, the

class of discrete groups ME to a lattice in G is completely determined. At present, these two theorems and Theorem 1.5 are the only results which completely describe the class of discrete groups ME to a given infinite group. It is known that there exist continuously many discrete groups ME to a non-abelian free group (see Theorem 2.27 in [48]). Although no group-theoretic characterization of the class of such groups is known, some non-trivial examples of groups in that class are known (see [21]).

Gaboriau [20] proved that the sequence $\{\beta_n(\Gamma)\}_{n \in \mathbb{N}}$ of ℓ^2 -Betti numbers for a discrete group Γ is an invariant for ME in the following sense: If two discrete groups Γ and Λ are ME, then there exists a positive real number c such that $\beta_n(\Gamma) = c\beta_n(\Lambda)$ for all n . This fact leads to big progress in the classification problem of discrete groups up to ME because this numerical invariant is defined for all discrete groups and is computable for various discrete groups arising geometrically.

The theory of ME is deeply linked with the theory of orbit equivalence. In fact, Ornstein and Weiss's original theorem is formulated in terms of orbit equivalence. Moreover, orbit equivalence is closely related to the theory of von Neumann algebras. There are many noteworthy results around these fields. We recommend the reader to consult [21], [58], [60] and the references therein for recent development of these fields.

Organization of this chapter. In Section 2, we recall fundamentals of mapping class groups, groupoids, and ME. It is important to know the construction of an isomorphism between two discrete measured groupoids from an ME coupling. Thanks to this construction, we can handle the classification problem of ME as an algebraic problem of groupoids arising from measure-preserving actions of discrete groups on measure spaces. To analyze the groupoid arising from an action of the mapping class group, we study its subgroupoids. Many facts about subgroups of mapping class groups reviewed in this section will help us to proceed to the analysis of subgroupoids because a groupoid is a generalization of a group.

In Section 3, we give an outline of the proof of Theorems 1.5 and 1.6. The main step in the proof is to consider a self ME coupling of the mapping class group $\Gamma(M)$, i.e., an ME coupling of $\Gamma(M)$ and $\Gamma(M)$. This corresponds to considering an isomorphism between groupoids arising from two actions of $\Gamma(M)$. We first explain what we can say about such an isomorphism, which will be formulated in Theorem 3.6. Here, we give only its statement, and will explain its proof in subsequent sections. Assuming Theorem 3.6, we show that any self ME coupling of $\Gamma(M)$ can be reduced to a much simpler self ME coupling of $\Gamma(M)$. We explain how to deduce the rigidity results in Theorems 1.5 and 1.6 from such a reduction. As another direct application, we prove a rigidity result in terms of orbit equivalence.

In Section 4, we recall amenability of discrete measured groupoids. This notion is often utilized in the study of groupoids and plays an important role in this work.

From Section 5 to Section 7, we study subgroupoids of a groupoid \mathcal{G} arising from a measure-preserving action of $\Gamma(M)$ on a standard Borel space with a finite positive measure. In Section 5, we classify subgroupoids of \mathcal{G} , following the classification of subgroups of $\Gamma(M)$ due to McCarthy and Papadopoulos [45]. We introduce two

types of subgroupoids of \mathcal{G} , which are called IA and reducible ones, respectively. In Section 6, we recall the definition of normal subgroupoids of a discrete measured groupoid, and study the normalizers in \mathcal{G} of an IA or reducible subgroupoid. In Section 7, using results shown in the previous sections, we characterize various reducible subgroupoids in terms of amenability and normal subgroupoids. This characterization makes it possible to study an isomorphism between groupoids arising from two actions of $\Gamma(M)$ and to prove Theorem 3.6.

Finally, in Section 8, we briefly explain other related results shown in the series of papers [35], [36], [37], [38].

Acknowledgements. The author would like to express his deep gratitude to Athanase Papadopoulos and Charles Boubel for reading the first version of this chapter very carefully. Thanks to their valuable comments, this chapter was greatly improved. This chapter was written during the stay at Max Planck Institute for Mathematics in Bonn. The author wishes to thank the institute for its warm hospitality.

2 Preliminaries

2.1 Mapping class groups

In this subsection, we recall fundamental facts about mapping class groups and several geometric objects related to them. We refer the reader to [13], [30], [32] or Sections 3.1, 3.2, 4.3 and 4.5 in [35] and the references therein for the material of this subsection. Chapter 8 of Volume I of this handbook ([49]) also deals with this material.

Let $M = M_{g,p}$ be a connected, compact and orientable surface of type (g, p) , that is, of genus g and with p boundary components. Throughout the chapter, a surface is assumed to be connected, compact and orientable unless otherwise stated. Let $\Gamma(M)$ be the *mapping class group* of M , i.e., the group of isotopy classes of all orientation-preserving diffeomorphisms of M . The *extended mapping class group* $\Gamma(M)^\diamond$ of M is the group of isotopy classes of all diffeomorphisms of M , which contains $\Gamma(M)$ as a subgroup of index 2. Let $\kappa(M) = 3g + p - 4$ be the *complexity* of M . We recall two geometric objects, the curve complex and the Thurston boundary, on which $\Gamma(M)^\diamond$ naturally acts.

The curve complex C . This simplicial complex was introduced by Harvey [28] and plays an indispensable role in this chapter. We recall some fundamental properties.

Definition 2.1. For a surface M , let $V(C) = V(C(M))$ be the set of all non-trivial isotopy classes of non-peripheral simple closed curves on M . Here, a simple closed curve on M is said to be non-peripheral if it is not isotopic to any boundary component of M . Let $S(M)$ denote the set of all non-empty finite subsets of $V(C)$ which can be realized disjointly on M at the same time.

When $\kappa(M) > 0$, we define the *curve complex* $C = C(M)$ as a simplicial complex such that the set of vertices is $V(C)$, and the set of simplices is $S(M)$.

When $\kappa(M) = 0$, that is, when M is either of type $(1, 1)$ or $(0, 4)$, we define the *curve complex* $C = C(M)$ as the one-dimensional simplicial complex such that the set of vertices is $V(C)$, the set of edges is defined as follows: A pair $\{\alpha, \beta\}$ of two distinct elements of $V(C)$ forms an edge if α and β have the lowest possible intersection number, that is, 1 for $M_{1,1}$ and 2 for $M_{0,4}$.

When $M = M_{0,3}$, let $C = C(M)$ be the empty set. For other surfaces, we do not need to define curve complexes because such surfaces do not appear as components of the surface obtained by cutting a surface with non-negative complexity along disjoint and mutually non-isotopic curves.

We immediately see that the curve complex is locally infinite (if it is non-empty). Although it is often difficult to treat the curve complex because of this property, it admits the following remarkable property.

Theorem 2.2 ([43], [47]). *If M is a surface with $\kappa(M) \geq 0$, then the curve complex $C = C(M)$ is connected. Moreover, when C is equipped with the natural simplicial metric, it has infinite diameter and is hyperbolic in the sense of Gromov.*

Let $\text{Aut}(C)$ be the automorphism group of the simplicial complex C . Note that since $\Gamma(M)^\diamond$ acts on C simplicially, there is a natural homomorphism $\pi : \Gamma(M)^\diamond \rightarrow \text{Aut}(C)$. It is natural to ask whether this natural homomorphism is an isomorphism or not. The following theorem answers this question completely in the case of $\kappa(M) > 0$. We refer to [40], [42], [47] for the case of $\kappa(M) = 0$.

Theorem 2.3 ([31], [40], [42]). *Let M be a surface with $\kappa(M) > 0$.*

- (i) *If M is neither $M_{1,2}$ nor $M_{2,0}$, then π is an isomorphism.*
- (ii) *If $M = M_{1,2}$, then the image of π is a subgroup of $\text{Aut}(C)$ with index 5 and $\ker(\pi)$ is the subgroup generated by a hyperelliptic involution, which is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.*
- (iii) *If $M = M_{2,0}$, then π is surjective and $\ker(\pi)$ is the subgroup generated by a hyperelliptic involution, which is isomorphic to $\mathbb{Z}/2\mathbb{Z}$.*
- (iv) *The two simplicial complexes $C(M_{0,5})$ and $C(M_{1,2})$ (resp. $C(M_{0,6})$ and $C(M_{2,0})$) are isomorphic.*

The Thurston boundary \mathcal{PMF} . Here, we recall some important facts on the Thurston boundary. We recommend the reader to consult [13] for details and proofs of the following facts.

Let M be a surface with $\kappa(M) \geq 0$ and let $\mathcal{R}(M)$ be the set of all non-negative real valued functions on $V(C)$, endowed with the product topology. We denote by $\mathcal{P}\mathcal{R}(M)$ the quotient space of $\mathcal{R}(M) \setminus \{0\}$ by the natural diagonal action of the multiplicative

group $\mathbb{R}_{>0}^*$ of all positive real numbers. Let $i: V(C) \times V(C) \rightarrow \mathbb{N}$ be the minimal geometric intersection number among representatives of two elements of $V(C)$. In particular, $i(\alpha, \alpha) = 0$ for all $\alpha \in V(C)$. For each $\alpha \in V(C)$, we can define an element of $\mathcal{R}(M) \setminus \{0\}$ by the function $V(C) \ni \beta \mapsto i(\alpha, \beta)$. The induced map $V(C) \rightarrow \mathcal{R}(M)$ is then injective. The closure of $\mathbb{R}_{>0}^* \cdot V(C)$ in $\mathcal{R}(M)$ is denoted by $\mathcal{MF} = \mathcal{MF}(M)$, and it is called the space of *measured foliations* on M . This space \mathcal{MF} is homeomorphic to $\mathbb{R}^{6g-6+2p}$. In fact, it is known that each element of \mathcal{MF} can be identified with a foliation with some singularities on M equipped with a transverse measure. The way to identify an element of \mathcal{MF} and a measured foliation is not immediate. The reader should be referred to Exposé 5 in [13] for this identification.

Moreover, the composed map $V(C) \rightarrow \mathcal{R}(M) \setminus \{0\} \rightarrow \mathcal{PR}(M)$ is also injective. The closure of the image is denoted by $\mathcal{PMF} = \mathcal{PMF}(M)$, and it is called the *Thurston boundary* or the space of *projective measured foliations* on M . This space \mathcal{PMF} is homeomorphic to the sphere of dimension $6g - 7 + 2p$. It is known that $S(M)$ can also naturally be embedded into \mathcal{PMF} by using the minimal geometric intersection number $i: S(M) \times V(C) \rightarrow \mathbb{N}$ among representatives of elements of $S(M)$ and $V(C)$. This function i can be continuously extended to a function $\mathcal{MF} \times \mathcal{MF} \rightarrow \mathbb{R}_{\geq 0}$ which is $\mathbb{R}_{>0}^*$ -homogeneous in the following sense:

$$i(r_1 F_1, r_2 F_2) = r_1 r_2 i(F_1, F_2)$$

for any $r_1, r_2 \in \mathbb{R}_{>0}^*$ and $F_1, F_2 \in \mathcal{MF}$. Hence, for two elements $F_1, F_2 \in \mathcal{PMF}$, whether $i(F_1, F_2) = 0$ or $\neq 0$ makes sense. As $\mathcal{R}(M)$ is endowed with the product topology, the group $\Gamma(M)^\diamond$ acts continuously on both \mathcal{MF} and \mathcal{PMF} , and the equation

$$i(gF_1, gF_2) = i(F_1, F_2)$$

holds for any $g \in \Gamma(M)^\diamond$ and $F_1, F_2 \in \mathcal{MF}$ (or \mathcal{PMF}). Let

$$\mathcal{MIN} = \{F \in \mathcal{PMF} : i(F, \alpha) \neq 0 \text{ for any } \alpha \in V(C)\}$$

be the set of all *minimal measured foliations* on M , which is a $\Gamma(M)^\diamond$ -invariant Borel subset of \mathcal{PMF} .

Each point of the Teichmüller space $\mathcal{T} = \mathcal{T}(M)$ also defines an element of $\mathcal{R}(M) \setminus \{0\}$. Indeed, once chosen a hyperbolic metric on M , there is exactly one geodesic in each free homotopy class of closed, non-peripheral curves on M . The lengths of these geodesics give a map $\mathcal{T} \rightarrow \mathcal{R}(M) \setminus \{0\}$. The induced map $\mathcal{T} \rightarrow \mathcal{PR}(M)$ is then injective, and \mathcal{PMF} forms the boundary of the image of this map. The disjoint union $\overline{\mathcal{T}} = \mathcal{T} \cup \mathcal{PMF}$ is called the *Thurston compactification* of the Teichmüller space, which is homeomorphic to a closed Euclidean ball of dimension $6g - 6 + 2p$ whose boundary corresponds to \mathcal{PMF} .

For $g \in \Gamma(M)$, let us denote by

$$\text{Fix}(g) = \{x \in \overline{\mathcal{T}} : gx = x\}$$

the fixed point set of g . Each element $g \in \Gamma(M)$ is classified as follows in terms of its fixed points on $\overline{\mathcal{T}}$ (see Exposé 9, §V, Théorème and Exposé 11, §4, Théorème in [13]):

Theorem 2.4. *Let M be a surface with $\kappa(M) \geq 0$. Each element $g \in \Gamma(M)$ can be classified into the following three types:*

- (i) g has finite order and has a fixed point on \mathcal{T} .
- (ii) g is pseudo-Anosov, that is, $\text{Fix}(g)$ consists of exactly two points of $\mathcal{M}\mathcal{I}\mathcal{N}$.
- (iii) g has infinite order and is reducible, that is, there exists $\sigma \in S(M)$ such that $g\sigma = \sigma$.

Note that these three types are mutually exclusive. We say that $F \in \mathcal{PMF}$ is a *pseudo-Anosov foliation* if F is a fixed point for some pseudo-Anosov element. It is known that the set of all pseudo-Anosov foliations is dense in \mathcal{PMF} .

Dynamics of each element of $\Gamma(M)$ on \mathcal{PMF} . This information will help us to consider the problem of probability measures on \mathcal{PMF} which are invariant for the action of a subgroup of $\Gamma(M)$ (see Subsection 5.1). Let M be a surface with $\kappa(M) \geq 0$. A pseudo-Anosov element $g \in \Gamma(M)$ has the following remarkable dynamics on $\overline{\mathcal{T}}$.

Theorem 2.5 ([32, Theorem 7.3.A]). *Let M be a surface with $\kappa(M) \geq 0$ and let $g \in \Gamma(M)$ be a pseudo-Anosov element. Then the two fixed points $F_{\pm}(g) \in \mathcal{M}\mathcal{I}\mathcal{N}$ of g satisfy the following: If U is an open neighborhood of $F_+(g)$ in $\overline{\mathcal{T}}$ and if K is a compact subset of $\overline{\mathcal{T}} \setminus \{F_-(g)\}$, then there exists $N \in \mathbb{N}$ such that $g^n(K) \subset U$ for all $n \geq N$.*

We call $F_+(g)$ (resp. $F_-(g)$) the *unstable* (resp. *stable*) *foliation* for g .

We next consider the dynamics of a reducible element. We say that $g \in \Gamma(M)$ is *pure* if the isotopy class g contains a diffeomorphism φ of M satisfying the following condition (P).

We say that a diffeomorphism φ of M satisfies *Condition (P)* if there exists a closed one-dimensional submanifold c (may be empty) of M such that

- each component of c is neither homotopic on M to a point nor to ∂M ;
- φ is the identity on c , and it does not rearrange the components of $M \setminus c$. Moreover, φ induces on each component of the surface M_c obtained by cutting M along c a diffeomorphism isotopic to either a pseudo-Anosov or the identity diffeomorphism.

We may assume that c does not have *superfluous components*, that is, we cannot discard any component of c without violating Condition (P). Note that if some component of c is on the boundary of two components on which the action of φ is isotopic to the identity, then the action of φ on the union of these two components is not necessarily isotopic to the identity. There exists a finite index subgroup of $\Gamma(M)$ consisting of pure elements (see Theorem 2.8 (i)).

Let $g \in \Gamma(M)$ be a pure element and let c be a one-dimensional submanifold of M satisfying Condition (P) for some diffeomorphism in the isotopy class g . Let Q_1, \dots, Q_n be the components of M_c on which g induces a pseudo-Anosov element, and let $F_+^1, \dots, F_+^n, F_-^1, \dots, F_-^n \in \mathcal{MF}$ be some representatives of the corresponding unstable and stable foliations. Let $\alpha_1, \dots, \alpha_m$ be the isotopy classes of the components of c which are also boundary components of some Q_i . Let β_1, \dots, β_l be the isotopy classes of the remaining components of c . For $F \in \mathcal{MF}$, let $[F] \in \mathcal{PMF}$ denote the projection of F onto \mathcal{PMF} . Define two subsets Δ^u, Ψ^s of \mathcal{PMF} by

$$\Delta^u = \left\{ \left[\sum_{i=1}^n m_i F_+^i + \sum_{j=1}^m a_j \alpha_j + \sum_{k=1}^l b_k \beta_k \right] \in \mathcal{PMF} : \right. \\ \left. m_i, a_j, b_k \geq 0, \sum_{i=1}^n m_i + \sum_{j=1}^m a_j + \sum_{k=1}^l b_k > 0 \right\},$$

$$\Psi^s = \{ [F] \in \mathcal{PMF} : i(F, F_-^i) = i(F, \beta_k) = 0 \text{ for all } i, k \}.$$

See Section 2.4 in [30] for the sum of disjoint foliations. These subsets Δ^u, Ψ^s are closed in \mathcal{PMF} . Moreover, if g is a reducible element of infinite order, then both Δ^u and Ψ^s are contained in $\mathcal{PMF} \setminus \mathcal{MLN}$ (see Corollary 2.16 in [30]). The following gives the behavior of the dynamics of a pure reducible element on \mathcal{PMF} .

Theorem 2.6 ([30, Theorem 3.5]). *Let M be a surface with $\kappa(M) \geq 0$ and let $g \in \Gamma(M)$ be a pure element. Let U be an open subset and let K be a compact subset of \mathcal{PMF} such that $\Delta^u \subset U$ and $K \subset \mathcal{PMF} \setminus \Psi^s$. Then there exists $N \in \mathbb{N}$ such that $g^n(K) \subset U$ for all $n \geq N$.*

Classification of subgroups of $\Gamma(M)$. Let M be a surface with $\kappa(M) \geq 0$. Using the classification of elements of $\Gamma(M)$ in Theorem 2.4, McCarthy and Papadopoulos [45] classified subgroups of $\Gamma(M)$ as follows.

Theorem 2.7. *Let M be a surface with $\kappa(M) \geq 0$. Each subgroup Γ of $\Gamma(M)$ can be classified into the following four types:*

- (i) Γ is finite.
- (ii) *There exists a pseudo-Anosov element $g \in \Gamma$ such that $h\{F_{\pm}(g)\} = \{F_{\pm}(g)\}$ for any $h \in \Gamma$. In this case, Γ is virtually cyclic and we say that Γ is IA (= infinite, irreducible and amenable).*
- (iii) Γ is infinite and there exists $\sigma \in S(M)$ such that $g\sigma = \sigma$ for any $g \in \Gamma$.
- (iv) *There exist two pseudo-Anosov elements g_1 and g_2 in Γ such that $\{F_{\pm}(g_1)\} \cap \{F_{\pm}(g_2)\} = \emptyset$. In this case, Γ contains a non-abelian free subgroup and is said to be sufficiently large.*

Note that these four types are mutually exclusive (use Theorem 2.5). A subgroup of $\Gamma(M)$ is said to be *reducible* if it fixes some element of $S(M)$.

Some special subgroups of finite index in $\Gamma(M)$. We finally introduce some finite index subgroups of $\Gamma(M)$ which satisfy nice properties. Since a discrete group and its finite index subgroup are ME as seen in Example 1.3, we may consider such special subgroups instead of $\Gamma(M)$ in the problem of ME. Thanks to the nice properties, many arguments technically get much easier.

For $\sigma \in S(M)$, we often denote by M_σ for simplicity the surface obtained by cutting M along a realization of curves in σ when a realization of σ is not specified. It is well known that if $g \in \Gamma(M)$ satisfies the equation $g\sigma = \sigma$, then there exist a realization c of σ and a diffeomorphism φ of M whose isotopy class is g such that $\varphi(c) = c$ (see Theorem 5.2 in [41] for the proof). Then φ induces a diffeomorphism on the surface M_c obtained by cutting M along c . When φ preserves each component of M_c , we say that g preserves each component of M_σ . This definition depends only on the isotopy classes σ and g , and does not depend on the choice of c and φ . Likewise, we often identify an isotopy class and some representative of it for simplicity of the notation if no serious problem occurs. For an integer m , let $\Gamma(M; m)$ be the subgroup of $\Gamma(M)$ consisting of all elements which act trivially on the homology group $H_1(M; \mathbb{Z}/m\mathbb{Z})$. This subgroup has the following notable properties (see Theorem 1.2 and Corollaries 1.5, 1.8, 3.6 in [30]).

Theorem 2.8. *Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. Then the following assertions hold:*

- (i) $\Gamma(M; m)$ is a torsion-free subgroup of finite index in $\Gamma(M)$ and consists of pure elements.
- (ii) If $g \in \Gamma(M; m)$ and $F \in \mathcal{PMF}$ satisfy $g^n F = F$ for some $n \in \mathbb{Z} \setminus \{0\}$, then $gF = F$.
- (iii) If $g \in \Gamma(M; m)$ and $\sigma \in S(M)$ satisfy $g^n \sigma = \sigma$ for some $n \in \mathbb{Z} \setminus \{0\}$, then $g\alpha = \alpha$ for any $\alpha \in \sigma$, and g preserves each component of M_σ and preserves each component of the boundary of M .

2.2 Discrete measured groupoids

This subsection is a short review of the notion of a discrete measured groupoid. We refer to [4], [5] and Chapter XIII, §3 in [59] for more details.

Measure theory. We first recall some basic terminology in measure theory. A Borel space X is a set equipped with a distinguished σ -field of subsets of X . A subset in the σ -field is called a Borel subset. A map $f: X \rightarrow Y$ between Borel spaces X and Y is said to be Borel if $f^{-1}(A)$ is a Borel subset of X for any Borel subset A of Y . In this chapter, we always assume a Borel space to be standard. A Borel space is *standard* if as a Borel space, it is isomorphic to a Borel space associated with a separable complete metric space. The following facts are known:

Theorem 2.9. (i) If a Borel space X is a countable union of Borel subsets of X which are standard as a Borel space, then X is standard.

(ii) Any Borel subset of a standard Borel space is standard as a Borel space.

(iii) Any two standard Borel spaces with the same cardinality are isomorphic as a Borel space.

(iv) Let X, Y be standard Borel spaces and let $f: X \rightarrow Y$ be a Borel map such that $f^{-1}(y)$ is countable for each $y \in Y$. Then there exists a countable Borel partition $X = \bigsqcup_n X_n$ satisfying the following: Let f_n denote the restriction of f to X_n . The image $f_n(X_n)$ is a Borel subset of Y , and the map $f_n: X_n \rightarrow f_n(X_n)$ is a Borel isomorphism.

We refer to 13.4, 15.6 and 18.14 in [34] for Assertions (ii), (iii) and (iv), respectively. The reader should consult [34] for more details of standard Borel spaces. By a *standard measure space* we mean a standard Borel space X equipped with a σ -finite positive measure μ . If μ is finite, i.e., if $\mu(X) < \infty$, then we say that (X, μ) is a *standard finite measure space*.

Let μ be a positive measure on a Borel space X . We say that a Borel subset A of X is (μ) -null (resp. *conull*) if $\mu(A) = 0$ (resp. $\mu(X \setminus A) = 0$). A property of points of X which holds for all x outside some μ -null Borel subset of X is said to hold for (μ) -almost every (or *a.e.*) $x \in X$. A point $x \in X$ with $\mu(\{x\}) > 0$ is called an *atom* for the measure space (X, μ) . Two measures μ and ν on a Borel space X are said to be *equivalent* if the following holds: For a Borel subset A of X , $\mu(A) = 0$ if and only if $\nu(A) = 0$.

Let $(X, \mu), (Y, \nu)$ be Borel spaces with a positive measure. By a *measure space isomorphism* $f: (X, \mu) \rightarrow (Y, \nu)$ we mean a Borel isomorphism $f: X \rightarrow Y$ between conull Borel subsets $X' \subset X$ and $Y' \subset Y$ such that $f_*\mu$ and ν are equivalent.

Groupoids. A groupoid is a generalization of a group. Given a set X , a groupoid \mathcal{G} on X is, roughly speaking, the set of arrows whose end and initial points are in X satisfying several conditions. In the following definition, the maps $r, s: \mathcal{G} \rightarrow X$ assign to an arrow in \mathcal{G} its end and initial points in X , respectively.

Definition 2.10. If two non-empty sets \mathcal{G}, X are equipped with two maps $r, s: \mathcal{G} \rightarrow X$ and the following operations, then \mathcal{G} is called a *groupoid* on X :

(i) We put

$$\mathcal{G}^{(2)} = \{(\gamma_1, \gamma_2) \in \mathcal{G} \times \mathcal{G} : s(\gamma_1) = r(\gamma_2)\}.$$

There is a map $\mathcal{G}^{(2)} \ni (\gamma_1, \gamma_2) \mapsto \gamma_1\gamma_2 \in \mathcal{G}$ satisfying the two equations $r(\gamma_1\gamma_2) = r(\gamma_1)$ and $s(\gamma_1\gamma_2) = s(\gamma_2)$, and satisfying the associative law. The last condition means that the equation $(\gamma_1\gamma_2)\gamma_3 = \gamma_1(\gamma_2\gamma_3)$ holds for all $(\gamma_1, \gamma_2), (\gamma_2, \gamma_3) \in \mathcal{G}^{(2)}$.

(ii) There is a map $X \ni x \mapsto e_x \in \mathcal{G}$ satisfying the following equations: $r(e_x) = s(e_x) = x$; $\gamma e_x = \gamma$ for any $\gamma \in \mathcal{G}$ with $s(\gamma) = x$; and $e_x\gamma' = \gamma'$ for any $\gamma' \in \mathcal{G}$

with $r(\gamma') = x$. It is easy to see that for each $x \in X$, e_x is a unique element of \mathcal{G} satisfying these equations.

- (iii) There is a map $\mathcal{G} \ni \gamma \mapsto \gamma^{-1} \in \mathcal{G}$ satisfying the following equations: $r(\gamma^{-1}) = s(\gamma)$; $s(\gamma^{-1}) = r(\gamma)$; $\gamma\gamma^{-1} = e_{r(\gamma)}$; and $\gamma^{-1}\gamma = e_{s(\gamma)}$. It is easy to see that for each $\gamma \in \mathcal{G}$, γ^{-1} is a unique element of \mathcal{G} satisfying these equations.

In the above notation, X is called the *unit space*, and X is identified with the set of all units of \mathcal{G} via the map $x \mapsto e_x$. The maps $r, s: \mathcal{G} \rightarrow X$ are called the *range*, *source* maps, respectively. For $(\gamma_1, \gamma_2) \in \mathcal{G}^{(2)}$, the element $\gamma_1\gamma_2 \in \mathcal{G}$ is called the *product* of two elements γ_1, γ_2 . We refer to e_x as the *unit* on $x \in X$ and refer to γ^{-1} as the *inverse* of $\gamma \in \mathcal{G}$.

Consider a subset $\mathcal{H} \subset \mathcal{G}$ satisfying the following three conditions:

- If $(\gamma_1, \gamma_2) \in \mathcal{G}^{(2)} \cap (\mathcal{H} \times \mathcal{H})$, then $\gamma_1\gamma_2 \in \mathcal{H}$.
- If $\gamma \in \mathcal{H}$, then $\gamma^{-1} \in \mathcal{H}$.
- $e_x \in \mathcal{H}$ for all $x \in X$.

This subset \mathcal{H} admits the structure of a groupoid on X induced from the one for \mathcal{G} . This groupoid \mathcal{H} on X is called a *subgroupoid* of \mathcal{G} .

We say that a groupoid is *Borel* if all the associated spaces and maps are Borel. When we consider a Borel groupoid \mathcal{G} on a standard Borel space, we always assume \mathcal{G} to be also standard as a Borel space.

Notation. Let \mathcal{G} be a groupoid on the unit space X with the range and source maps $r, s: \mathcal{G} \rightarrow X$, respectively. We denote by

$$I: \mathcal{G} \ni \gamma \mapsto \gamma^{-1} \in \mathcal{G}.$$

the inverse map. We write $\mathcal{G}^x = r^{-1}(x)$ and $\mathcal{G}_x = s^{-1}(x)$ for $x \in X$. Note that $\mathcal{G}^x = I(\mathcal{G}_x)$ for each $x \in X$. We say that \mathcal{G} is *discrete* when \mathcal{G}^x is countable for each $x \in X$. For $x, y \in X$, we write

$$\mathcal{G}_y^x = \{\gamma \in \mathcal{G} : r(\gamma) = x, s(\gamma) = y\}.$$

It is easy to see that for each $x \in X$, \mathcal{G}_x^x admits the structure of a group induced from the structure of a groupoid on \mathcal{G} . This group \mathcal{G}_x^x is called the *isotropy group* on $x \in X$.

Example 2.11. Groups. Let G be a group. Then G can be seen as a groupoid on the set consisting of a single point. Conversely, any groupoid on the set consisting of a single point is a group.

Example 2.12. Equivalence relations. Let X be a non-empty set. Let \mathcal{R} be an equivalence relation on X , i.e., a subset of $X \times X$ satisfying the following three conditions:

- $(x, x) \in \mathcal{R}$ for all $x \in X$.
- If $(x, y) \in \mathcal{R}$, then $(y, x) \in \mathcal{R}$.

- If $(x, y), (y, z) \in \mathcal{R}$, then $(x, z) \in \mathcal{R}$.

Define two maps $r, s: \mathcal{R} \rightarrow X$ and the operations of products and inverses by

$$r(x, y) = x, \quad s(x, y) = y, \quad (x, y)(y, z) = (x, z), \quad (x, y)^{-1} = (y, x).$$

Then \mathcal{R} is a groupoid on X . If each equivalence class for \mathcal{R} is at most countable, then \mathcal{R} is a discrete groupoid.

Measures on discrete Borel groupoids. If we are given a discrete Borel groupoid and a positive measure on the unit space, then we can define a natural measure on the groupoid as follows.

Definition 2.13. Given a discrete Borel groupoid \mathcal{G} on a Borel space X , we say that a σ -finite positive measure μ on X is *quasi-invariant* for \mathcal{G} if the two measures $\tilde{\mu}$ and $I_*\tilde{\mu}$ on \mathcal{G} are equivalent. Here, the measure $\tilde{\mu}$ is defined by

$$\tilde{\mu}(A) = \int_X \sum_{\gamma \in \mathcal{G}_x} \chi_A(\gamma) d\mu(x),$$

for a Borel subset A of \mathcal{G} , where χ_A is the characteristic function on A . We say that μ is *invariant* for \mathcal{G} if $I_*\tilde{\mu} = \tilde{\mu}$. A discrete Borel groupoid \mathcal{G} equipped with a quasi-invariant measure μ on the unit space X is called a *discrete measured groupoid* on (X, μ) . Given a discrete measured groupoid \mathcal{G} on (X, μ) , we always equip \mathcal{G} with the measure $\tilde{\mu}$ defined above. This measure $\tilde{\mu}$ is a σ -finite positive measure on \mathcal{G} .

Notation. Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) . If A is a Borel subset of X , then we denote by $\mathcal{G}A$ the *saturation* of A , which is the Borel subset of X defined by

$$\mathcal{G}A = \{r(\gamma) \in X : \gamma \in \mathcal{G}, s(\gamma) \in A\} = \{s(\gamma) \in X : \gamma \in \mathcal{G}, r(\gamma) \in A\}.$$

It can be shown that $\mathcal{G}A$ is a Borel subset of X and that $\mu(\mathcal{G}A) = 0$ when $\mu(A) = 0$ (use Theorem 2.9 (iv)). If $\mathcal{G}A = A$, then A is said to be *\mathcal{G} -invariant*. Note that if X' is a conull Borel subset of X , then $X' \setminus \mathcal{G}(X \setminus X')$ is a conull \mathcal{G} -invariant Borel subset of X contained in X' .

Definition 2.14. Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) . If A is a Borel subset of X with positive measure, then the Borel subset

$$\{\gamma \in \mathcal{G} : r(\gamma), s(\gamma) \in A\}$$

has the natural structure of a groupoid on A induced from \mathcal{G} . This groupoid is called the *restriction* of \mathcal{G} to A and is denoted by $(\mathcal{G})_A$.

Definition 2.15. Let \mathcal{G}, \mathcal{H} be discrete measured groupoids on standard measure spaces $(X, \mu), (Y, \nu)$, respectively. By a groupoid homomorphism $f: \mathcal{G} \rightarrow \mathcal{H}$ we mean a Borel map $f: (\mathcal{G})_A \rightarrow \mathcal{H}$ for some conull \mathcal{G} -invariant Borel subset A of X satisfying the following two conditions:

- $f_*\mu$ and ν are equivalent;
- f preserves the operation of products, i.e., the equation $f(\gamma_1\gamma_2) = f(\gamma_1)f(\gamma_2)$ holds for all $(\gamma_1, \gamma_2) \in ((\mathcal{G})_A)^{(2)}$.

When X is identified with the set of all units of \mathcal{G} , the map $f: (\mathcal{G})_A \rightarrow \mathcal{H}$ induces a Borel map $f: A \rightarrow Y$.

We do not distinguish two groupoid homomorphisms $f_1, f_2: \mathcal{G} \rightarrow \mathcal{H}$ such that $f_1 = f_2$ on $(\mathcal{G})_A$ for some conull \mathcal{G} -invariant Borel subset A of X .

Remark 2.16. For $i \in \{1, 2, 3\}$, let \mathcal{G}_i be a discrete measured groupoid on a standard measure space (X_i, μ_i) . Let $f: \mathcal{G}_1 \rightarrow \mathcal{G}_2$ and $g: \mathcal{G}_2 \rightarrow \mathcal{G}_3$ be groupoid homomorphisms. For $i \in \{1, 2\}$, take a conull \mathcal{G}_i -invariant Borel subset A_i of X_i such that f_i is defined on $(\mathcal{G}_i)_{A_i}$. It can be easily seen that the Borel subset $A'_1 = (A_1 \cap f^{-1}(A_2)) \setminus \mathcal{G}_1(A_1 \setminus f^{-1}(A_2))$ of X_1 is conull and \mathcal{G}_1 -invariant. The composition of the two Borel maps $f: (\mathcal{G}_1)_{A'_1} \rightarrow \mathcal{G}_2$ and $g: (\mathcal{G}_2)_{A_2} \rightarrow \mathcal{G}_3$ is then defined. It is clear that this composition defines a groupoid homomorphism from \mathcal{G}_1 into \mathcal{G}_3 . We denote it by $g \circ f: \mathcal{G}_1 \rightarrow \mathcal{G}_3$.

Definition 2.17. Let \mathcal{G}, \mathcal{H} be discrete measured groupoids on standard measure spaces $(X, \mu), (Y, \nu)$, respectively. A groupoid homomorphism $f: \mathcal{G} \rightarrow \mathcal{H}$ is called an *isomorphism* if there exists a groupoid homomorphism $g: \mathcal{H} \rightarrow \mathcal{G}$ such that the compositions $g \circ f: \mathcal{G} \rightarrow \mathcal{G}$ and $f \circ g: \mathcal{H} \rightarrow \mathcal{H}$ coincide with the identity homomorphisms on \mathcal{G} and on \mathcal{H} , respectively. In this case, \mathcal{G} and \mathcal{H} are said to be *isomorphic*.

Though we often need to take \mathcal{G} -invariant Borel subsets of X in many situations in this chapter, we do not always mention it for simplicity of the notation.

As seen in Example 2.12, an equivalence relation on a set defines a groupoid on the set. We next introduce an equivalence relation on a Borel space which induces a discrete measured groupoid on the Borel space.

Definition 2.18. Let (X, μ) be a standard measure space. Let \mathcal{R} be a Borel subset of $X \times X$ such that

- \mathcal{R} defines an equivalence relation on X as in Example 2.12;
- for each $x \in X$, the equivalence class $\mathcal{R}_x = \{y \in X : (y, x) \in \mathcal{R}\}$ of x is at most countable.

Then \mathcal{R} is a discrete Borel groupoid on X with respect to the structure introduced in Example 2.12. If μ is quasi-invariant for this groupoid, then \mathcal{R} is called a *discrete measured equivalence relation* (or simply an *equivalence relation*) on (X, μ) .

Definition 2.19. Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) . It is easy to see that

$$\mathcal{R} = \{(r(\gamma), s(\gamma)) \in X \times X : \gamma \in \mathcal{G}\}$$

has the structure of a discrete measured groupoid on (X, μ) such that

$$r(x, y) = x, \quad s(x, y) = y, \quad (x, y)(y, z) = (x, z), \quad (x, y)^{-1} = (y, x).$$

This groupoid is called the *quotient equivalence relation* of \mathcal{G} . Note that if the isotropy group \mathcal{G}_x^x is trivial for a.e. $x \in X$, then \mathcal{G} and its quotient equivalence relation \mathcal{R} are isomorphic via the following isomorphism:

$$\mathcal{G} \ni \gamma \mapsto (r(\gamma), s(\gamma)) \in \mathcal{R}.$$

In this case, \mathcal{G} is said to be *principal*.

We give one typical example of discrete measured groupoids appearing in this chapter. We recommend the reader to see [5] for other examples of discrete measured groupoids.

Example 2.20. *Group actions.* Let G be a discrete group and assume that G admits a *non-singular* action on a standard measure space (X, μ) , which means that $\mu(A) = 0$ if and only if $\mu(gA) = 0$ for any $g \in G$ and for any Borel subset $A \subset X$. The direct product $G \times X$ then has the structure of a groupoid such that

$$r(g, x) = gx, \quad s(g, x) = x, \quad (g, hx)(h, x) = (gh, x), \quad (g, x)^{-1} = (g^{-1}, gx).$$

This groupoid is often written as $G \ltimes (X, \mu)$ or $G \ltimes X$. Since the action $G \curvearrowright (X, \mu)$ is non-singular, μ is quasi-invariant for $G \ltimes X$.

It is easy to see that μ is invariant for the action $G \curvearrowright (X, \mu)$ if and only if it is an invariant measure for the groupoid $G \ltimes X$. For a Borel subset $A \subset X$, the saturation $(G \ltimes X)A$ is equal to the saturation $GA = \bigcup_{g \in G} gA$.

The quotient equivalence relation

$$\mathcal{R} = \{(gx, x) \in X \times X : g \in G, x \in X\}$$

of $G \ltimes X$ admits the structure of a discrete measured groupoid on (X, μ) as seen in Definition 2.19. This \mathcal{R} can also be seen as a discrete measured equivalence relation on (X, μ) arising from the equivalence relation declaring that two points of X are equivalent if and only if they are in the same G -orbit.

Note that the action $G \curvearrowright (X, \mu)$ is *essentially free*, that is, the stabilizer of almost every point $x \in X$ is trivial if and only if $G \ltimes X$ is principal.

In this chapter, we mainly treat groupoids isomorphic to subgroupoids of a groupoid arising from a measure-preserving action of a discrete group on a standard finite measure space. In Section 4, we however treat discrete measured groupoids arising from non-singular actions of discrete groups which are never measure-preserving (see Theorems 4.20 and 4.21).

Conjugacy and orbit equivalence. Given two actions $G \curvearrowright (X, \mu)$ and $H \curvearrowright (Y, \nu)$, when are the two associated groupoids isomorphic? We shall give two equivalence relations for non-singular actions of discrete groups on measure spaces, called con-

jugacy and orbit equivalence. It is shown that when two actions are both essentially free, they are orbit equivalent if and only if the associated groupoids are isomorphic.

Definition 2.21. Let Γ, Λ be discrete groups and let $(X, \mu), (Y, \nu)$ be standard measure spaces. Consider non-singular actions $\Gamma \curvearrowright (X, \mu)$ and $\Lambda \curvearrowright (Y, \nu)$. The two actions are said to be *conjugate* if there exist an isomorphism $F: \Gamma \rightarrow \Lambda$ and a measure space isomorphism $f: (X, \mu) \rightarrow (Y, \nu)$ such that

$$f(\gamma x) = F(\gamma) f(x) \quad \text{for any } \gamma \in \Gamma \text{ and a.e. } x \in X.$$

More precisely, this means that we can take conull Borel subsets $X' \subset X$ and $Y' \subset Y$ and a Borel isomorphism $f: X' \rightarrow Y'$ satisfying the following: the two measures $f_*\mu$ and ν are equivalent; and for any $\gamma \in \Gamma$ and a.e. $x \in X'$, γx belongs to X' and the equation $f(\gamma x) = F(\gamma) f(x)$ holds.

Orbit equivalence is a weaker equivalence relation than conjugacy.

Definition 2.22. Let Γ, Λ be discrete groups and let $(X, \mu), (Y, \nu)$ be standard measure spaces. Consider non-singular actions $\Gamma \curvearrowright (X, \mu)$ and $\Lambda \curvearrowright (Y, \nu)$. The two actions are said to be *orbit equivalent (OE)* if there exists a measure space isomorphism $f: (X, \mu) \rightarrow (Y, \nu)$ such that

$$f(\Gamma x) = \Lambda f(x) \quad \text{for a.e. } x \in X.$$

More precisely, this means that we can take conull Borel subsets $X' \subset X$ and $Y' \subset Y$ and a Borel isomorphism $f: X' \rightarrow Y'$ satisfying the following: the two measures $f_*\mu$ and ν are equivalent; and for a.e. $x \in X'$, Γx is contained in X' and the equation $f(\Gamma x) = \Lambda f(x)$ holds. It is easy to see that this f induces an isomorphism between the two quotient equivalence relations of $\Gamma \ltimes X$ and $\Lambda \ltimes Y$, which is defined by $(x, y) \mapsto (f(x), f(y))$. Conversely, if an isomorphism between the two quotient equivalence relations of $\Gamma \ltimes X$ and $\Lambda \ltimes Y$ is given, then the associated map f between their unit spaces satisfies the above condition of OE.

If the action $\Lambda \curvearrowright (Y, \nu)$ is essentially free, then we can define a Borel map

$$\alpha: \Gamma \times X \rightarrow \Lambda \quad \text{so that } f(\gamma x) = \alpha(\gamma, x) f(x)$$

for $\gamma \in \Gamma$ and a.e. $x \in X$. This map α satisfies the following cocycle identity

$$\alpha(\gamma_1, \gamma_2 x) \alpha(\gamma_2, x) = \alpha(\gamma_1 \gamma_2, x)$$

for any $\gamma_1, \gamma_2 \in \Gamma$ and a.e. $x \in X$. Thus, α is a groupoid homomorphism from $\Gamma \ltimes X$ into Λ . We call α the *OE cocycle* associated with f .

The reader can check that when the actions $\Gamma \curvearrowright (X, \mu)$ and $\Lambda \curvearrowright (Y, \nu)$ are both essentially free, they are OE via f if and only if the two groupoids $\Gamma \ltimes X$ and $\Lambda \ltimes Y$ are isomorphic under the groupoid homomorphism $(\gamma, x) \mapsto (\alpha(\gamma, x), f(x))$ associated with f .

We next introduce a slightly weaker equivalence relation than OE, called weak orbit equivalence (WOE). It is known that two discrete groups are measure equivalent

(ME) if and only if the two groups admit ergodic, measure-preserving and essentially free actions which are WOE (see Corollary 2.34).

Definition 2.23. Let Γ, Λ be discrete groups and let $(X, \mu), (Y, \nu)$ be standard measure spaces. Consider non-singular actions $\Gamma \curvearrowright (X, \mu)$ and $\Lambda \curvearrowright (Y, \nu)$. The two actions are said to be *weakly orbit equivalent (WOE)* if there exist Borel subsets $A \subset X, B \subset Y$ and a Borel isomorphism $f: A \rightarrow B$ satisfying the following three conditions:

- (i) $\Gamma A = X, \Lambda B = Y$ up to null sets.
- (ii) The two measures $f_*(\mu|_A)$ and $\nu|_B$ are equivalent.
- (iii) $f(\Gamma x \cap A) = \Lambda f(x) \cap B$ for a.e. $x \in A$.

As in the case of OE, this f induces an isomorphism between the two quotient equivalence relations of $(\Gamma \times X)_A$ and $(\Lambda \times Y)_B$. Conversely, if an isomorphism between the two quotient equivalence relations of $(\Gamma \times X)_A$ and $(\Lambda \times Y)_B$ for Borel subsets A, B of X satisfying Condition (i) is given, then the associated map f between their unit spaces A and B satisfies Conditions (ii), (iii).

Groupoids of infinite type. In most sections of this chapter, we study a groupoid \mathcal{G} associated with a measure-preserving action of a discrete group Γ on a standard finite measure space and study its subgroupoids. In particular, we mainly study its subgroupoids of infinite type. When the unit space of \mathcal{G} consists of a single atom and \mathcal{G} is then isomorphic to Γ , subgroupoids of \mathcal{G} of infinite type correspond to infinite subgroups of Γ . Before defining the notion of groupoids of infinite type, we introduce recurrence of a discrete measured equivalence relation. Recall that a discrete measured equivalence relation can be seen as a discrete measured groupoid (see Definition 2.18).

Definition 2.24. Let \mathcal{R} be a discrete measured equivalence relation on a standard finite measure space (X, μ) with an invariant measure μ for \mathcal{R} . We say that \mathcal{R} is *recurrent* if a.e. equivalence class for \mathcal{R} is infinite, that is, for a.e. $x \in X$, the set $\mathcal{R}_x = \{y \in X : (y, x) \in \mathcal{R}\}$ is infinite.

Let \mathcal{R} be a discrete measured equivalence relation on a standard finite measure space (X, μ) with an invariant measure μ for \mathcal{R} . It is known that if \mathcal{R} is recurrent and A is a Borel subset of X with positive measure, then the restriction $(\mathcal{R})_A$ is also recurrent (see the proof of Lemma 2.5 in [2]). Moreover, it is shown that there exists an essentially unique Borel partition $X = A_1 \sqcup A_2$ such that

- $(\mathcal{R})_{A_1}$ is recurrent;
- $(\mathcal{R})_{A_2}$ admits a *fundamental domain*, i.e., there exists a Borel subset B of A_2 such that for a.e. $x \in A_2$, $(\mathcal{R})_{A_2}x \cap B$ consists of exactly one point, where $(\mathcal{R})_{A_2}x$ denotes the equivalence class for $(\mathcal{R})_{A_2}$ containing x .

See Lemma 2.12 in [2] for the proof of this fact. The reader can check that $\mathcal{R}A_1 = A_1$ and $\mathcal{R}A_2 = A_2$ up to null sets. It is easy to treat an equivalence relation which

admits a (Borel) fundamental domain because the space of orbits of the equivalence relations can be identified with its fundamental domain. The above fact means that any equivalence relation can be divided into an easy part and a non-trivial part. Hence, it is often enough to consider only recurrent equivalence relations in the study of equivalence relations. We refer to Section 2 in [2] for fundamental properties of discrete measured equivalence relations, where the recurrence of discrete measured equivalence relations such that μ is not necessarily invariant is also discussed. The notion of groupoids of infinite type is defined as follows.

Definition 2.25. Let \mathcal{G} be a discrete measured groupoid on a standard finite measure space (X, μ) with an invariant measure μ for \mathcal{G} . Then we say that \mathcal{G} is of *infinite type* if there exists a Borel partition $X = X_1 \sqcup X_2$ such that

- the isotropy group \mathcal{G}_x^x is infinite for a.e. $x \in X_1$;
- the quotient equivalence relation of $(\mathcal{G})_{X_2}$ is recurrent.

By definition, if \mathcal{G} is of infinite type, then $(\mathcal{G})_A$ is also of infinite type for any Borel subset $A \subset X$ with positive measure. The next proposition shows that a measure-preserving action of an infinite discrete group on a standard finite measure space always gives rise to a groupoid of infinite type.

Proposition 2.26. *Let Γ be an infinite discrete group and suppose that Γ admits a measure-preserving action on a standard finite measure space (X, μ) . Then the associated groupoid $\Gamma \times X$ is of infinite type, and thus so is the restriction $(\Gamma \times X)_A$ for any Borel subset $A \subset X$ with positive measure.*

Proof. Let \mathcal{R} be the quotient equivalence relation of $\Gamma \times X$. Choose a Borel partition $X = A_1 \sqcup A_2$ such that $(\mathcal{R})_{A_1}$ is recurrent and $(\mathcal{R})_{A_2}$ admits a fundamental domain $B \subset A_2$ (see the comment right after Definition 2.24). Both A_1 and A_2 are then invariant under the action $\Gamma \curvearrowright (X, \mu)$. For a.e. $x \in A_2$, the orbit Γx consists of only finitely many points because the action $\Gamma \curvearrowright (X, \mu)$ is measure-preserving and $\mu(A_2) < \infty$. Therefore, the stabilizer of a.e. $x \in A_2$ is infinite. \square

2.3 ME and isomorphism of groupoids

In this subsection, we construct from an ME coupling of discrete groups Γ and Λ an isomorphism of groupoids associated with some measure-preserving actions of Γ and Λ on standard finite measure spaces. This construction was essentially given in Section 3 in [17]. Thanks to this construction, we can reduce the problem of ME to an algebraic problem of groupoids arising from group actions.

Let (Σ, m) be an ME coupling of discrete groups Γ and Λ , i.e., (Σ, m) is a standard Borel space with a σ -finite positive measure, and there is a measure-preserving action $\Gamma \times \Lambda \curvearrowright (\Sigma, m)$ such that both of the actions $\Gamma (\simeq \Gamma \times \{e\}) \curvearrowright (\Sigma, m)$ and

$\Lambda (\simeq \{e\} \times \Lambda) \curvearrowright (\Sigma, m)$ are essentially free and admit a fundamental domain of finite measure (see Definition 1.1). Choose fundamental domains $Y \subset \Sigma$ for the action $\Gamma \curvearrowright \Sigma$, and $X \subset \Sigma$ for the action $\Lambda \curvearrowright \Sigma$. Remark that we have a natural measure-preserving action of Γ on X equipped with the restricted finite measure μ of m to X because X can be identified with the quotient space Σ/Λ as a Borel space. Similarly, we have a natural measure-preserving action of Λ on Y with a finite measure ν . In order to distinguish from the original actions of Γ and Λ on Σ , we denote the actions $\Gamma \curvearrowright X$ and $\Lambda \curvearrowright Y$ by $\gamma \cdot x, \lambda \cdot y$, respectively, using a dot.

Lemma 2.27. *In the above notation, one can choose X and Y so that $A = X \cap Y$ satisfies the following two conditions:*

- $\Gamma \cdot A = X$ up to null sets when A is regarded as a subset of X ;
- $\Lambda \cdot A = Y$ up to null sets when A is regarded as a subset of Y .

Proof. Let S be the set of all Borel subsets $B \subset \Sigma$ such that $m(\gamma_1 B \Delta \gamma_2 B) = 0$ for all distinct $\gamma_1, \gamma_2 \in \Gamma$, and $m(\lambda_1 B \Delta \lambda_2 B) = 0$ for all distinct $\lambda_1, \lambda_2 \in \Lambda$. Here, $C \Delta D$ denotes the symmetric difference of two sets C and D . If we find $A \in S$ such that the equation $(\Gamma \times \Lambda)A = \Sigma$ holds up to null sets, then choose fundamental domains $X \subset \Sigma$ for the action $\Lambda \curvearrowright \Sigma$, and $Y \subset \Sigma$ for the action $\Gamma \curvearrowright \Sigma$ such that $A \subset X$ and $A \subset Y$. The above two conditions are then satisfied for these X and Y . Hence, we will find $A \in S$ satisfying the equation $(\Gamma \times \Lambda)A = \Sigma$ up to null sets.

Put $M = \sup_{B \in S} m(B)$. Then $M < \infty$. Since we can always take fundamental domains of the actions of Γ and Λ on Σ whose intersection has positive measure, the number M is positive. Let $\{B_n\}_{n \in \mathbb{N}}$ be a sequence of elements of S such that $m(B_n) \rightarrow M$ as $n \rightarrow \infty$. Put $A_1 = B_1$ and define $A_n = (B_n \setminus (\Gamma \times \Lambda)A_{n-1}) \cup A_{n-1}$ for $n \geq 2$ inductively. Then $A_n \in S$ and $A = \bigcup_n A_n$ is also in S . It is easy to see that $m((\Gamma \times \Lambda)B_n \setminus (\Gamma \times \Lambda)A_n) = 0$. In particular, $m((\Gamma \times \Lambda)B_n \setminus (\Gamma \times \Lambda)A) = 0$ for all n .

We claim that $(\Gamma \times \Lambda)A = \Sigma$ up to null sets. If $\Sigma \setminus (\Gamma \times \Lambda)A$ had positive measure, then it would be an ME coupling of Γ and Λ . There exists a Borel subset $B \subset \Sigma \setminus (\Gamma \times \Lambda)A$ which is in S as a Borel subset of Σ and has positive measure. Take $n \in \mathbb{N}$ so that $M - m(B_n) < m(B)$. Then $B \cup B_n \in S$ and $m(B \cup B_n) = m(B) + m(B_n) > M$, which is a contradiction. \square

In what follows, suppose that X and Y satisfy the conditions of Lemma 2.27. Let $\mathcal{G} = \Gamma \times (X, \mu)$ (resp. $\mathcal{H} = \Lambda \times (Y, \nu)$) be the groupoid associated with the action $\Gamma \curvearrowright X$ (resp. $\Lambda \curvearrowright Y$). We can define a Borel map

$$\alpha: \Gamma \times X \rightarrow \Lambda \quad \text{so that } \gamma \cdot x = \alpha(\gamma, x)\gamma x \in Y$$

for any $\gamma \in \Gamma$ and a.e. $x \in X$ because X is a fundamental domain of the action $\Lambda \curvearrowright \Sigma$. Similarly, we can define a Borel map

$$\beta: \Lambda \times Y \rightarrow \Gamma \quad \text{so that } \lambda \cdot y = \beta(\lambda, y)\lambda y \in X$$

for $\lambda \in \Lambda$ and a.e. $y \in Y$.

Lemma 2.28. *The map $\alpha : \Gamma \times X \rightarrow \Lambda$ is a cocycle, that is, the cocycle identity*

$$\alpha(\gamma_1, \gamma_2 \cdot x)\alpha(\gamma_2, x) = \alpha(\gamma_1\gamma_2, x)$$

is satisfied for each $\gamma_1, \gamma_2 \in \Gamma$ and a.e. $x \in X$. The map $\beta : \Lambda \times Y \rightarrow \Gamma$ also satisfies a similar identity.

This cocycle identity implies that α is a groupoid homomorphism from $\Gamma \times X$ into Λ . We call α (resp. β) the *ME cocycle* associated with X (resp. Y).

Proof. This follows from the following equality:

$$\alpha(\gamma_1, \gamma_2 \cdot x)\alpha(\gamma_2, x)\gamma_1\gamma_2x = \alpha(\gamma_1, \gamma_2 \cdot x)\gamma_1(\gamma_2 \cdot x) = \gamma_1 \cdot (\gamma_2 \cdot x) = (\gamma_1\gamma_2) \cdot x$$

for $\gamma_1, \gamma_2 \in \Gamma$ and $x \in X$, where the right hand side is in X . □

Let $p : X \rightarrow Y$ and $q : Y \rightarrow X$ be the Borel maps defined by

$$p(x) = \Gamma x \cap Y, \quad q(y) = \Lambda y \cap X$$

for $x \in X$ and $y \in Y$. Note that both p and q are the identity on $A = X \cap Y$. Then we can show that

$$p(\gamma \cdot x) = \alpha(\gamma, x) \cdot p(x), \quad q(\lambda \cdot y) = \beta(\lambda, y) \cdot q(y)$$

for any $\gamma \in \Gamma, \lambda \in \Lambda$ and a.e. $x \in X, y \in Y$ as follows: Since $\gamma \cdot x = \alpha(\gamma, x)\gamma x$, there exists a unique $\gamma_1 \in \Gamma$ such that

$$p(\gamma \cdot x) = \gamma_1\alpha(\gamma, x)x \in Y. \tag{2.1}$$

Let $\gamma_2, \gamma_3 \in \Gamma$ be unique elements such that $p(x) = \gamma_2x \in Y$ and

$$\alpha(\gamma, x) \cdot p(x) = \gamma_3\alpha(\gamma, x)\gamma_2x \in Y. \tag{2.2}$$

Comparing (2.1) and (2.2), we see that $\gamma_1 = \gamma_3\gamma_2$ since Y is a fundamental domain of the action $\Gamma \curvearrowright \Sigma$. This proves the claim.

Define groupoid homomorphisms

$$\begin{aligned} f : (\mathcal{G})_A \ni (\gamma, x) &\mapsto (\alpha(\gamma, x), p(x)) \in (\mathcal{H})_A, \\ g : (\mathcal{H})_A \ni (\lambda, y) &\mapsto (\beta(\lambda, y), q(y)) \in (\mathcal{G})_A. \end{aligned}$$

Note that $\beta(\alpha(\gamma, x), x) = \gamma$ for any $\gamma \in \Gamma$ and a.e. $x \in A$ with $\gamma \cdot x \in A$ because $\gamma\alpha(\gamma, x)x = \gamma \cdot x \in A \subset Y$. Similarly, $\alpha(\beta(\lambda, y), y) = \lambda$ for any $\lambda \in \Lambda$ and a.e. $y \in A$ with $\lambda \cdot y \in A$. Therefore, we obtain the following

Proposition 2.29. *In the above notation, the groupoid homomorphisms*

$$f : (\mathcal{G})_A \rightarrow (\mathcal{H})_A, \quad g : (\mathcal{H})_A \rightarrow (\mathcal{G})_A$$

satisfy $g \circ f = \text{id}$ and $f \circ g = \text{id}$.

This is a construction of an isomorphism between two (restrictions of) groupoids generated by actions of Γ and Λ from an ME coupling of Γ and Λ . In particular, if we

can show that Γ and Λ admit no actions which generate isomorphic groupoids (even after restricting to any Borel subsets with positive measure), then this implies that Γ and Λ are *not* ME. Therefore, when we consider the problem of ME, it is effective to study algebraic properties of groupoids arising from specific groups. Such a groupoid $\Gamma \ltimes X$ often behaves like the group Γ . More precisely, suppose that Γ admits some “nice” action on a space S . The action is given by a homomorphism $\Gamma \rightarrow \text{Aut}(S)$. The projection $\Gamma \ltimes X \rightarrow \Gamma$ is a groupoid homomorphism. We can then view the groupoid homomorphism $\Gamma \ltimes X \rightarrow \Gamma \rightarrow \text{Aut}(S)$ as an action of $\Gamma \ltimes X$ on S , and we observe that this action of $\Gamma \ltimes X$ often gives rise to phenomena similar to the ones for the action of Γ on S . This idea greatly helps us to study the groupoid $\Gamma \ltimes X$ (see also the beginning of Section 5).

Return to the situation before Proposition 2.29 and consider the action of $\Gamma \times \Lambda$ on $X \times \Lambda$ defined by

$$(\gamma, \lambda)(x, \lambda') = (\gamma \cdot x, \alpha(\gamma, x)\lambda'\lambda^{-1}), \quad \gamma \in \Gamma, \lambda, \lambda' \in \Lambda, x \in X.$$

It is easy to check the following lemma, which means that we can reconstruct an ME coupling from the cocycle α .

Lemma 2.30. *In the above notation, the Borel map $\Sigma \rightarrow X \times \Lambda$ defined by $\lambda x \mapsto (x, \lambda^{-1})$ for $x \in X$ and $\lambda \in \Lambda$ is Borel isomorphic and $(\Gamma \times \Lambda)$ -equivariant.*

Note that Proposition 2.29 implies that the two actions of Γ on X and Λ on Y are WOE (see Definition 2.23). Conversely, the following theorem is known. This states that given WOE actions of Γ and Λ , we can construct the corresponding ME coupling of Γ and Λ .

Definition 2.31. For simplicity, by a *standard action* of a discrete group we mean an essentially free, measure-preserving Borel action of that group on a standard finite measure space.

Theorem 2.32 ([17, Theorem 3.3]). *Suppose that two discrete groups Γ and Λ admit ergodic standard actions on (X, μ) and (Y, ν) , respectively, which are WOE. Then we can construct an ME coupling (Σ, m) of Γ and Λ such that the Γ -actions on X and on $\Lambda \setminus \Sigma$ (resp. the Λ -actions on Y and on $\Gamma \setminus \Sigma$) are conjugate.*

In particular, if the two actions $\Gamma \curvearrowright (X, \mu)$ and $Y \curvearrowright (Y, \nu)$ are OE via a Borel isomorphism f between conull Borel subsets of X and Y , then we can construct the above ME coupling (Σ, m) so that the ME cocycle associated with some fundamental domain of the Λ -action on Σ , which is identified with X under the above conjugacy of the Γ -actions on X and on $\Lambda \setminus \Sigma$, is equal to the OE cocycle associated with f .

Remark 2.33. In the case of WOE, we can also define an associated WOE cocycle and prove a statement similar to the latter assertion in Theorem 2.32 (see Theorem 3.3 in [17]).

Corollary 2.34. *Two discrete groups are ME if and only if they admit ergodic standard actions on standard finite measure spaces which are WOE.*

Proof. The “if” part follows from Theorem 2.32. We prove the “only if” part. Let (Σ, m) be an ME coupling of discrete groups Γ and Λ . Let $\Gamma \curvearrowright (X, \mu)$ be a standard action. For example, the Bernoulli action $\Gamma \curvearrowright \prod_{\Gamma} [0, 1]$ given by

$$\gamma(x_g)_{g \in \Gamma} = (x_{\gamma^{-1}g})_{g \in \Gamma}, \quad \gamma \in \Gamma, \quad (x_g)_{g \in \Gamma} \in \prod_{\Gamma} [0, 1]$$

is standard (see Section 2 in [38]). The action of $\Gamma \times \Lambda$ on $\Sigma \times X$ given by

$$(\gamma, \lambda)(x, x') = ((\gamma, \lambda)x, \gamma x'), \quad \gamma \in \Gamma, \quad \lambda \in \Lambda, \quad x \in \Sigma, \quad x' \in X$$

defines an ME coupling of Γ and Λ such that the action $\Gamma \times \Lambda \curvearrowright (\Sigma \times X, m \times \mu)$ is essentially free. By utilizing the ergodic decomposition for the action $\Gamma \times \Lambda \curvearrowright (\Sigma \times X, m \times \mu)$, we can construct an ME coupling (Σ_0, m_0) of Γ and Λ such that the action $\Gamma \times \Lambda \curvearrowright (\Sigma_0, m_0)$ is essentially free and ergodic (see Lemma 2.2 in [16]). Thus, the two actions $\Gamma \curvearrowright \Lambda \setminus \Sigma_0$ and $\Lambda \curvearrowright \Gamma \setminus \Sigma_0$ are both ergodic and standard. Proposition 2.29 implies that the two actions are WOE. \square

3 ME rigidity for mapping class groups

In this section, we state two key theorems for the proof of the ME rigidity result of the mapping class group $\Gamma(M)$. The first one, Theorem 3.1, is reduction of a self ME coupling of $\Gamma(M)$ (i.e., an ME coupling of $\Gamma(M)$ and itself) to a simpler self ME coupling of $\Gamma(M)$. This reduction is a very important step for the proof of measurable rigidity. As stated in Proposition 2.29, a self ME coupling of $\Gamma(M)$ gives rise to an isomorphism between groupoids arising from two measure-preserving actions of $\Gamma(M)$. The second key theorem 3.6 states a certain important property of such an isomorphism. The proof of this theorem will be explained in subsequent sections. In Subsection 3.1, assuming Theorem 3.6, we show Theorem 3.1. To establish ME rigidity from these theorems, we need one more step, which is explained in Subsection 3.2. In Subsection 3.3, we give another immediate application of the reduction of self ME couplings. We prove an OE rigidity result for ergodic standard actions of the mapping class group.

3.1 Reduction of self ME couplings of mapping class groups

We first give an outline to prove Theorem 1.5, an ME rigidity result for the mapping class group, and give three steps (1), (2), (3) for the proof. This outline is similar to Furman’s one for the proof of Theorem 1.9, an ME rigidity result for higher rank lattices. Step (2) is devoted to the first key theorem 3.1, noted above. Here is the

most important and difficult step. We give three steps (a), (b), (c) for the proof of this theorem. Some of the steps are formulated in terms of groupoids arising from actions of mapping class groups, and it seems complicated for beginners of groupoids. Before giving an explicit formulation of these steps, we explain how the steps are formulated when the unit spaces of the groupoids consist of a single atom, i.e., when the groupoids are groups. Under this assumption, these steps are formulated in terms of groups, which is much easier to understand.

We shall give three steps (1), (2), (3) for the proof of Theorem 1.5. Let M be a surface with $\kappa(M) > 0$. Let (Σ, m) be an ME coupling of the mapping class group $\Gamma = \Gamma(M)$ and a discrete group Λ . Let $C = C(M)$ be the curve complex of M and let $\text{Aut}(C)$ be its automorphism group. Let $\pi : \Gamma(M)^\diamond \rightarrow \text{Aut}(C)$ be the natural homomorphism.

- (1) Let Ω be the quotient space of $\Sigma \times \Lambda \times \Sigma$ by the $(\Lambda \times \Lambda)$ -action on $\Sigma \times \Lambda \times \Sigma$ given by

$$(\lambda_1, \lambda_2)(x, \lambda, y) = (\lambda_1 x, \lambda_1 \lambda \lambda_2^{-1}, \lambda_2 y), \quad \lambda_1, \lambda_2, \lambda \in \Lambda, \quad x, y \in \Sigma.$$

We define a $(\Gamma \times \Gamma)$ -action on $\Sigma \times \Lambda \times \Sigma$ by

$$(\gamma_1, \gamma_2)(x, \lambda, y) = (\gamma_1 x, \lambda, \gamma_2 y), \quad \gamma_1, \gamma_2 \in \Gamma, \quad \lambda \in \Lambda, \quad x, y \in \Sigma.$$

This $(\Gamma \times \Gamma)$ -action then induces a $(\Gamma \times \Gamma)$ -action on Ω . It is easy to check that Ω is a self ME coupling of Γ , i.e., an ME coupling of Γ and Γ .

- (2) We construct an almost $(\Gamma \times \Gamma)$ -equivariant Borel map $\Phi : \Omega \rightarrow \text{Aut}(C)$, i.e.,

$$\Phi((\gamma_1, \gamma_2)z) = \pi(\gamma_1)\Phi(z)\pi(\gamma_2)^{-1}$$

for any $\gamma_1, \gamma_2 \in \Gamma$ and a.e. $z \in \Omega$.

- (3) Using the map Φ , we construct a representation ρ of the group Λ on $\text{Aut}(C)$. Moreover, this homomorphism $\rho : \Lambda \rightarrow \text{Aut}(C)$ has finite kernel, and $\rho(\Lambda)$ is a finite index subgroup of $\text{Aut}(C)$. This proves Theorem 1.5.

In this subsection, details of Step (2) are discussed. We explain Step (3) in Subsection 3.2. Step (2) is a consequence of the following theorem.

Theorem 3.1 ([36, Corollary 5.9]). *Let M be a surface with $\kappa(M) > 0$. Let Γ_1 and Γ_2 be finite index subgroups of $\Gamma(M)^\diamond$ and suppose that there is an ME coupling (Ω, ω) of Γ_1 and Γ_2 . Then there exists an almost $(\Gamma_1 \times \Gamma_2)$ -equivariant Borel map $\Phi : \Omega \rightarrow \text{Aut}(C)$, i.e.,*

$$\Phi((\gamma_1, \gamma_2)z) = \pi(\gamma_1)\Phi(z)\pi(\gamma_2)^{-1}$$

for any $\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2$ and a.e. $z \in \Omega$.

This theorem means that all ME coupling of Γ_1 and Γ_2 can be reduced to the simpler ME coupling $\text{Aut}(C)$ of Γ_1 and Γ_2 on which $\Gamma_1 \times \Gamma_2$ acts as follows:

$$(\gamma_1, \gamma_2)g = \pi(\gamma_1)g\pi(\gamma_2)^{-1}, \quad \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, g \in \text{Aut}(C).$$

By a technical lemma (see Lemma 5.8 in [36]), if we can construct an almost $(\Gamma'_1 \times \Gamma'_2)$ -equivariant Borel map $\Phi: \Omega \rightarrow \text{Aut}(C)$ for some finite index subgroups Γ'_i of Γ_i for $i \in \{1, 2\}$, then Φ is in fact an almost $(\Gamma_1 \times \Gamma_2)$ -equivariant. It follows that in the proof of Theorem 3.1, we may assume that both Γ_1 and Γ_2 are finite index subgroups of $\Gamma(M; m)$ with an integer $m \geq 3$ (see Theorem 2.8 for the subgroup $\Gamma(M; m)$ of $\Gamma(M)$). In what follows in this subsection, we always assume this condition (because the key theorem, Theorem 3.6, is proved under this assumption). To state an outline of the proof of Theorem 3.1, we fix the notation as follows.

Notation. We refer to the following assumption as (\bullet) :

- Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ_1 and Γ_2 be finite index subgroups of $\Gamma(M; m)$. Let (Ω, ω) be an ME coupling of Γ_1 and Γ_2 .
- Take fundamental domains $X_1 \subset \Omega$ for the Γ_2 -action on Ω , and $X_2 \subset \Omega$ for the Γ_1 -action on Ω . Recall that the natural actions $\Gamma_1 \curvearrowright X_1$ and $\Gamma_2 \curvearrowright X_2$ are denoted by $(\gamma, x) \mapsto \gamma \cdot x$ by using a dot. By Lemma 2.27, we can choose X_1, X_2 so that $Y = X_1 \cap X_2$ satisfies that for $i \in \{1, 2\}$, $\Gamma_i \cdot Y = X_i$ up to null sets when Y is regarded as a subset of X_i .
- For $i \in \{1, 2\}$, set $\mathcal{G}^i = \Gamma_i \times X_i$ and let $\rho_i: \mathcal{G}^i \rightarrow \Gamma_i$ be the projection, which is a groupoid homomorphism. By Proposition 2.29, there exists a groupoid isomorphism

$$f: (\mathcal{G}^1)_Y \rightarrow (\mathcal{G}^2)_Y.$$

Note that f is the identity on the unit space Y .

- For $i \in \{1, 2\}$ and $\alpha \in V(C)$, let D_α^i be the intersection of Γ_i with the subgroup of $\Gamma(M)$ generated by the Dehn twist $t_\alpha \in \Gamma(M)$ about α . Let \mathcal{G}_α^i be the subgroupoid of \mathcal{G}^i generated by the action of D_α^i , i.e.,

$$\mathcal{G}_\alpha^i = \{(\gamma, x) \in \mathcal{G}^i : \gamma \in D_\alpha^i, x \in X_i\}.$$

An outline of the proof of Theorem 3.1 is as follows.

- f preserves subgroupoids generated by Dehn twists up to a countable Borel partition (see Theorem 3.6 for a precise statement).
- Using Step (a), we construct a Borel map $\Psi: Y \rightarrow \text{Aut}(C)$ associated with f .
- The Borel map $Y \ni x \mapsto \Psi(x)^{-1} \in \text{Aut}(C)$ can be extended to an almost $(\Gamma_1 \times \Gamma_2)$ -equivariant Borel map $\Phi: X_1 \times \Gamma_2 \rightarrow \text{Aut}(C)$. Here, Y is identified with the Borel subset $Y \times \{e\}$ of $X_1 \times \Gamma_2$. Note that $(\Gamma_1 \times \Gamma_2)(Y \times \{e\}) = X_1 \times \Gamma_2$ up to null sets and that $X_1 \times \Gamma_2$ can be identified with Ω as an ME coupling of Γ_1 and Γ_2 (see Lemma 2.30).

In what follows, we explain an explicit statement of Step (a) and details of Steps (b) and (c).

Ivanov's argument. Before discussing details of Steps (a), (b) and (c), we study these steps in the degenerate case, that is, in the case where both X_1 and X_2 consist of a single atom. In this case, our groupoids \mathcal{G}^1 and \mathcal{G}^2 degenerate into the groups Γ_1 and Γ_2 , respectively. Therefore, the argument for these steps gets much easier and clearer.

In the above argument, we obtained an isomorphism between the two restricted groupoids $(\Gamma_1 \times X_1)_\gamma$ and $(\Gamma_2 \times X_2)_\gamma$ arising from an ME coupling of finite index subgroups Γ_1 and Γ_2 of the mapping class group. What happens if we assume that each of X_1 and X_2 consists of a single atom? In this case, we obtain a group isomorphism $f: \Gamma_1 \rightarrow \Gamma_2$. Conversely, if $f: \Gamma_1 \rightarrow \Gamma_2$ is an isomorphism, then the action of $\Gamma_1 \times \Gamma_2$ on Γ_2 given by

$$(\gamma_1, \gamma_2)\gamma = f(\gamma_1)\gamma\gamma_2^{-1}, \quad \gamma_1 \in \Gamma_1, \gamma_2, \gamma \in \Gamma_2$$

defines an ME coupling of Γ_1 and Γ_2 such that $\{e\}$ is a fundamental domain for both of the actions of Γ_1 and Γ_2 on Γ_2 , and the isomorphism between Γ_1 and Γ_2 given in Proposition 2.29 is equal to f . Ivanov showed the following theorem about an isomorphism between finite index subgroups of the mapping class group.

Theorem 3.2 ([32, Theorem 8.5.A]). *Let M be a surface with $\kappa(M) > 0$ and $M \neq M_{1,2}, M_{2,0}$. Let Γ_1 and Γ_2 be finite index subgroups of $\Gamma(M)^\diamond$. If $f: \Gamma_1 \rightarrow \Gamma_2$ is an isomorphism, then there exists a unique $g \in \Gamma(M)^\diamond$ such that $f(\gamma) = g\gamma g^{-1}$ for any $\gamma \in \Gamma_1$.*

Outline of the proof. The first step of the proof is to show that f maps sufficiently high powers of Dehn twists into powers of Dehn twists. Namely, for each $\alpha \in V(C)$, there exist non-zero integers N, M and $\beta \in V(C)$ such that $t_\alpha^N \in \Gamma_1, t_\beta^M \in \Gamma_2$ and $f(t_\alpha^N) = t_\beta^M$. This fact is a consequence of the following theorem, which characterizes a non-trivial power of a Dehn twist algebraically.

Theorem 3.3 ([32, Theorem 7.5.B]). *Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let G be a finite index subgroup of $\Gamma(M; m)$. An element $g \in G$ is a non-trivial power of some Dehn twist (i.e., there are $n \in \mathbb{Z} \setminus \{0\}$ and $\alpha \in V(C)$ such that $g = t_\alpha^n$) if and only if the center of the centralizer of g in G is isomorphic to \mathbb{Z} and is not equal to the centralizer of g in G .*

Note that in the notation of Theorem 3.3, if $g = t_\alpha^n$ for $n \in \mathbb{Z} \setminus \{0\}$ and $\alpha \in V(C)$, then

- the centralizer $C_G(g)$ of g in G is equal to the stabilizer of α in G , i.e., $\{h \in G : h\alpha = \alpha\}$;
- the center of $C_G(g)$ is equal to $G \cap \langle t_\alpha \rangle$, where $\langle t_\alpha \rangle$ is the subgroup of $\Gamma(M)$ generated by t_α .

Return to the proof of Theorem 3.2. It follows from Theorem 3.3 that the isomorphism $f: \Gamma_1 \rightarrow \Gamma_2$ induces a map $\varphi: V(C) \rightarrow V(C)$ determined by $f(t_\alpha^N) = t_{\varphi(\alpha)}^M$ for $\alpha \in V(C)$ and some non-zero integers N, M . Such an element $\varphi(\alpha)$ is uniquely determined by the following fact.

Lemma 3.4. *Let M be a surface with $\kappa(M) \geq 0$. For $\alpha, \beta \in V(C)$ and $(k, l) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$, if $t_\alpha^k = t_\beta^l$, then $\alpha = \beta$ and $k = l$.*

For the proof of Lemma 3.4, we use the following lemma, which is shown by using the dynamics of Dehn twists on the Thurston boundary.

Lemma 3.5 ([36, Lemma 5.3]). *Let M be a surface with $\kappa(M) \geq 0$.*

- (i) *If two curves $\alpha, \beta \in V(C)$ satisfy $i(\alpha, \beta) = 0$, then the subgroup of $\Gamma(M)$ generated by the Dehn twists $t_\alpha, t_\beta \in \Gamma(M)$ about them is a free abelian group of rank 2. In particular, it is amenable.*
- (ii) *On the other hand, if $i(\alpha, \beta) \neq 0$, then the subgroup of $\Gamma(M)$ generated by t_α^n and t_β^m is a non-abelian free group of rank 2 for all sufficiently large $n, m \in \mathbb{N}$.*

Proof of Lemma 3.4. Suppose that $t_\alpha^k = t_\beta^l$ for $\alpha, \beta \in V(C)$ and $(k, l) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$. It is enough to prove that $\alpha = \beta$ because any Dehn twist is an element of infinite order. If $i(\alpha, \beta) \neq 0$, then it would contradict Lemma 3.5 (ii). Thus, $i(\alpha, \beta) = 0$. When $\kappa(M) = 0$, two distinct elements of $V(C)$ always have non-zero geometric intersection number. This shows that $\alpha = \beta$. Suppose that $\kappa(M) > 0$. If $\alpha \neq \beta$, then there would exist $\gamma \in V(C)$ such that $i(\alpha, \gamma) = 0$ and $i(\beta, \gamma) \neq 0$. This also contradicts Lemma 3.5. □

Return to the proof of Theorem 3.2. Since f is an isomorphism, it is easy to see that the map $\varphi: V(C) \rightarrow V(C)$ is a bijection. By using Lemma 3.5, one can show that φ induces an automorphism of the curve complex C , which comes from some $g \in \Gamma(M)^\diamond$ by Theorem 2.3. Namely, for each $\alpha \in V(C)$, there exist non-zero integers N, M such that $f(t_\alpha^N) = t_{g(\alpha)}^M$. Note that

$$gt_\alpha g^{-1} = t_{g\alpha}^\varepsilon \quad \text{for } \alpha \in V(C) \text{ and } g \in \Gamma(M)^\diamond, \tag{3.1}$$

where ε is 1 if $g \in \Gamma(M)$, and -1 otherwise (see Lemma 4.1.C in [32]). Let $\gamma \in \Gamma_1$. For each $\alpha \in V(C)$, we have

$$f(\gamma t_\alpha^N \gamma^{-1}) = f(\gamma) f(t_\alpha^N) f(\gamma)^{-1} = f(\gamma) t_{g(\alpha)}^M f(\gamma)^{-1} = t_{f(\gamma)g(\alpha)}^{\varepsilon M}$$

for some non-zero integers N, M and $\varepsilon \in \{\pm 1\}$. On the other hand,

$$f(\gamma t_\alpha^{N'} \gamma^{-1}) = f(t_{\gamma\alpha}^{\varepsilon' N'}) = t_{g(\gamma\alpha)}^{M'}$$

for some non-zero integers N', M' and $\varepsilon' \in \{\pm 1\}$. These equations imply that $f(\gamma)g(\alpha) = g(\gamma\alpha)$ by Lemma 3.4, and thus $f(\gamma)\beta = g(\gamma g^{-1}(\beta))$ for any $\beta \in V(C)$. By Theorem 2.3, $f(\gamma) = g\gamma g^{-1}$ for any $\gamma \in \Gamma_1$. Uniqueness of g satisfying this equation follows from the fact that the center of Γ_1 is trivial (use the equation (3.1) and Theorem 2.3). This proves Theorem 3.2. □

To sum up, Ivanov’s proof of Theorem 3.2 is outlined as follows. Steps (A), (B) correspond to our Steps (a), (b), respectively. We use the notation of Theorem 3.2.

- (A) Characterize Dehn twists algebraically, and show that the isomorphism $f : \Gamma_1 \rightarrow \Gamma_2$ preserves Dehn twists.
- (B) Define a bijection $\varphi : V(C) \rightarrow V(C)$ by the equation $f(t_\alpha^N) = t_{\varphi(\alpha)}^M$ for each $\alpha \in V(C)$ and some non-zero integers N, M . Show that φ defines an element of $\text{Aut}(C)$. Let $g \in \Gamma(M)^\diamond \simeq \text{Aut}(C)$ be the corresponding element.
- (C) By a direct calculation, show the equation $f(\gamma) = g\gamma g^{-1}$ for all $\gamma \in \Gamma_1$.

The case of groupoids. Return to our situation. Our Step (a) corresponds to Ivanov’s Step (A) and is stated explicitly as follows.

Theorem 3.6. *Under Assumption (●), for each $\alpha \in V(C)$, there exist a countable Borel partition $Y = \bigsqcup Y_n$ and $\beta_n \in V(C)$ such that*

$$f((\mathcal{G}_\alpha^1)_{Y_n}) = (\mathcal{G}_{\beta_n}^2)_{f(Y_n)} \quad \text{for each } n.$$

As explained in Step (a), this equation means that f preserves subgroupoids generated by Dehn twists after taking some countable Borel partition of Y .

Remark 3.7. Note that if a countable Borel partition $Y = \bigsqcup Y'_m$ and $\beta'_m \in V(C)$ also satisfy the equation in Theorem 3.6 and if $Z = f(Y_n \cap Y'_m)$ has positive measure for some n and m , then $(\mathcal{G}_{\beta_n}^2)_Z = (\mathcal{G}_{\beta'_m}^2)_Z$. It follows from Proposition 2.26 that there exist non-zero integers N, M and $x \in Z$ such that $(t_{\beta_n}^N, x) = (t_{\beta'_m}^M, x) \in (\mathcal{G}_{\beta_n}^2)_Z = (\mathcal{G}_{\beta'_m}^2)_Z$, which implies that $\beta_n = \beta'_m$ by Lemma 3.4.

The subsequent sections of this chapter will be devoted to the proof of Theorem 3.6. In Section 7, our plan of the proof will be presented. In this section, assuming Theorem 3.6, we proceed to Step (b).

About Step (b). Assuming Theorem 3.6, we construct the map Ψ in Step (b). We use the notation in Assumption (●). Let $\Psi : Y \times V(C) \rightarrow V(C)$ be the Borel map defined by

$$\Psi(x, \alpha) = \beta_n \quad \text{if } x \in Y_n$$

for $\alpha \in V(C)$, where $Y = \bigsqcup Y_n$ and β_n are chosen for α as in Theorem 3.6. By Remark 3.7, this definition does not depend on the choice of the countable Borel partition of Y . It can be shown that the map $\Psi(x, \cdot) : V(C) \rightarrow V(C)$ is a bijection (because f is an isomorphism), and moreover $\Psi(x, \cdot)$ defines an automorphism of the curve complex C for a.e. $x \in Y$. Namely, $\Psi(x, \cdot)$ satisfies the following two conditions for a.e. $x \in Y$:

- If $\alpha, \beta \in V(C)$ satisfy $i(\alpha, \beta) = 0$, then $i(\Psi(x, \alpha), \Psi(x, \beta)) = 0$.
- If $\alpha, \beta \in V(C)$ satisfy $i(\alpha, \beta) \neq 0$, then $i(\Psi(x, \alpha), \Psi(x, \beta)) \neq 0$.

This fact can be shown by utilizing Lemma 3.5 and some elementary facts about amenable discrete measured groupoids. (In Section 4, amenability of a discrete measured groupoid will be introduced, which is an isomorphism invariant of discrete measured groupoids.)

Therefore, we can define a Borel map $\Psi: Y \rightarrow \text{Aut}(C)$ by $\Psi(x) = \Psi(x, \cdot)$ for $x \in Y$. Using the equation (3.1) in Ivanov's proof, we can show that this map Ψ satisfies the following equality:

$$\Psi(r(\delta)) = \pi \circ \rho_2(f(\delta))\Psi(s(\delta))\pi \circ \rho_1(\delta)^{-1},$$

or equivalently,

$$\Psi(\gamma \cdot x) = \pi \circ \rho_2(f(\gamma, x))\Psi(x)\pi(\gamma)^{-1} \quad (3.2)$$

for a.e. $\delta = (\gamma, x) \in (\mathcal{G}^1)_Y$ (see Lemma 5.5 in [36]), where $\pi: \Gamma(M)^\diamond \rightarrow \text{Aut}(C)$ is the natural homomorphism. Note that Ψ , $\pi \circ \rho_2(f(\delta))$ and $\pi \circ \rho_1(\delta)$ correspond to g , $f(\gamma)$ and γ in Ivanov's argument, respectively. Thus, the equation (3.2) corresponds to his conclusion $g = f(\gamma)g\gamma^{-1}$.

About Step (c). Recall that the action of $\Gamma_1 \times \Gamma_2$ on $X_1 \times \Gamma_2$ was defined by

$$(\gamma_1, \gamma_2)(x, \gamma) = (\gamma_1 \cdot x, \alpha(\gamma_1, x)\gamma\gamma_2^{-1}), \quad \gamma_1 \in \Gamma_1, \gamma_2, \gamma \in \Gamma_2, x \in X_1$$

(see Lemma 2.30). Here, $\alpha: \Gamma_1 \times X_1 \rightarrow \Gamma_2$ is the ME cocycle associated with a fundamental domain X_1 of the Γ_2 -action on Ω . Note that $\rho_2 \circ f = \alpha$ on $(\mathcal{G}^1)_Y$ (see the definition of f introduced right before Proposition 2.29). Recall that the equation $(\Gamma_1 \times \Gamma_2)(Y \times \{e\}) = X_1 \times \Gamma_2$ holds up to null sets. Therefore, we define a Borel map $\Phi: X_1 \times \Gamma_2 \rightarrow \text{Aut}(C)$ by

$$\Phi((\gamma_1, \gamma_2)(x, e)) = \pi(\gamma_1)\Psi(x)^{-1}\pi(\gamma_2)^{-1}$$

for $\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2$ and $x \in Y$. If it is well-defined, then it is easy to see that Φ is almost $(\Gamma_1 \times \Gamma_2)$ -equivariant. Take $\gamma_1, \gamma'_1 \in \Gamma_1, \gamma_2, \gamma'_2 \in \Gamma_2$ and $x, x' \in Y$ satisfying the equality

$$(\gamma_1, \gamma_2)(x, e) = (\gamma'_1, \gamma'_2)(x', e).$$

This equality implies that

$$(x, e) = (\gamma_1^{-1}\gamma'_1, \gamma_2^{-1}\gamma'_2)(x', e) = ((\gamma_1^{-1}\gamma'_1) \cdot x', \alpha(\gamma_1^{-1}\gamma'_1, x')(\gamma_2^{-1}\gamma'_2)^{-1}).$$

Hence, $(\gamma_1^{-1}\gamma'_1) \cdot x' = x \in Y$. By using the equation (3.2) and the equation $\rho_2 \circ f = \alpha$ on $(\mathcal{G}^1)_Y$, we see that

$$\begin{aligned} \Psi(x) &= \Psi((\gamma_1^{-1}\gamma'_1) \cdot x') = \pi \circ \rho_2(f(\gamma_1^{-1}\gamma'_1, x'))\Psi(x')\pi(\gamma_1^{-1}\gamma'_1)^{-1} \\ &= \pi \circ \alpha(\gamma_1^{-1}\gamma'_1, x')\Psi(x')\pi(\gamma_1^{-1}\gamma'_1)^{-1} = \pi(\gamma_2^{-1}\gamma'_2)\Psi(x')\pi(\gamma_1^{-1}\gamma'_1)^{-1}. \end{aligned}$$

This implies that $\pi(\gamma_1)\Psi(x)^{-1}\pi(\gamma_2)^{-1} = \pi(\gamma'_1)\Psi(x')^{-1}\pi(\gamma'_2)^{-1}$ and that the map Φ is well-defined. This shows Step (c).

Therefore, the remaining problem is to show Theorem 3.6, which will be explained in the subsequent sections. The first Step (A) for Ivanov's proof is to show that the isomorphism $f: \Gamma_1 \rightarrow \Gamma_2$ preserves powers of Dehn twists. To prove this, he characterized a power of a Dehn twist algebraically as in Theorem 3.3. In our case, to prove that the groupoid isomorphism $f: (\mathcal{G}^1)_Y \rightarrow (\mathcal{G}^2)_Y$ preserves subgroupoids generated by Dehn twists as in Theorem 3.6, we characterize such a subgroupoid algebraically in terms of discrete measured groupoids. However, we cannot expect a characterization similar to that of Theorem 3.3 because there is no notion corresponding to centralizers and centers in the theory of discrete measured groupoids. In the subsequent sections, we give a characterization of a subgroupoid generated by a Dehn twist from another point of view. This is formulated in terms of amenable, non-amenable subgroups and normal subgroups (see Propositions 7.7 and 7.8). Since amenability of a discrete measured groupoid and normality of a subgroupoid are invariant under isomorphism of groupoids, subgroupoids generated by Dehn twists are preserved by f thanks to this characterization.

In Sections 4, 5 and 6, we introduce many notions necessary for the formulation of this characterization. In Section 7, the characterization is given.

3.2 Deriving ME rigidity from reduction of self ME couplings

As an application of Theorem 3.1, we prove Theorems 1.5 and 1.6.

ME rigidity. The process to deduce ME rigidity from reduction of self ME couplings has already been developed by Furman [16], and Monod and Shalom [48]. We review their techniques here. Recall the following two operations to construct a new ME coupling from a given ME coupling.

An opposite coupling. Let (Σ, m) be an ME coupling of discrete groups Γ and Λ . Then an ME coupling $(\check{\Sigma}, \check{m})$ of Λ and Γ is defined as follows: As a measure space, $(\check{\Sigma}, \check{m}) = (\Sigma, m)$. The action of $\Lambda \times \Gamma$ on $(\check{\Sigma}, \check{m})$ is defined via the canonical isomorphism between $\Gamma \times \Lambda$ and $\Lambda \times \Gamma$.

A composed coupling. If (Σ, m) is an ME coupling of discrete groups Γ and Λ and if (Ω, n) is an ME coupling of discrete groups Λ and Δ , then an ME coupling $\Sigma \times_{\Lambda} \Omega$ of Γ and Δ is defined to be the quotient space of $\Sigma \times \Omega$ by the diagonal Λ -action, equipped with the induced action of $\Gamma \times \Delta$.

Remark 3.8. By using the above two associated couplings, we see that ME is an equivalence relation among discrete groups (see Section 2 in [16]). Note that a discrete group Γ is itself an ME coupling of Γ and Γ as in Example 1.3.

Let M be a surface with $\kappa(M) > 0$ and let $\Gamma = \Gamma(M)$ be the mapping class group. Let (Σ, m) be an ME coupling of Γ and an unknown group Λ . Construct the self ME

coupling

$$\Omega = \Sigma \times_{\Lambda} \Lambda \times_{\Lambda} \check{\Sigma}$$

of Γ . We denote by $[x, \lambda, y] \in \Omega$ the projection of $(x, \lambda, y) \in \Sigma \times \Lambda \times \check{\Sigma}$. By applying Theorem 3.1, we obtain an almost $(\Gamma \times \Gamma)$ -equivariant Borel map $\Phi: \Omega \rightarrow \text{Aut}(C)$, i.e.,

$$\Phi((\gamma_1, \gamma_2)z) = \pi(\gamma_1)\Phi(z)\pi(\gamma_2)^{-1}$$

for any $\gamma_1, \gamma_2 \in \Gamma$ and a.e. $z \in \Omega$, where $\pi: \Gamma(M)^{\circ} \rightarrow \text{Aut}(C)$ is the natural homomorphism. From this map, we want to construct a representation ρ of the unknown group Λ on $\text{Aut}(C)$. We first consider the following special case.

Example 3.9 ([16, Example 5.1]). Let G be a locally compact second countable group and let Γ and Λ be lattices in G . Then G equipped with its Haar measure is an ME coupling of Γ and Λ as in Example 1.2. Define a Borel map

$$\Phi: G \times_{\Lambda} \Lambda \times_{\Lambda} \check{G} \rightarrow G$$

by $\Phi([x, \lambda, y]) = x\lambda y^{-1}$. This map is $(\Gamma \times \Gamma)$ -equivariant. Observe that the map

$$\lambda \mapsto \Phi([x, \lambda, y])\Phi([x, e, y])^{-1} = (x\lambda y^{-1})(xy^{-1})^{-1} = x\lambda x^{-1}$$

does not depend on y , and defines a representation of Λ on G for a fixed x .

From this observation, in our case, we can also expect that the map

$$\lambda \mapsto \Phi([x, \lambda, y])\Phi([x, e, y])^{-1} \tag{3.3}$$

does not depend on y , and defines a representation of Λ on $\text{Aut}(C)$ for a.e. $x \in \Sigma$. In fact, we can show these claims by using the following notable fact.

Theorem 3.10 ([36, Theorem 2.6]). *Let C be the curve complex of a surface M with $\kappa(M) > 0$. Let Γ be a finite index subgroup of $\text{Aut}(C)$. Then the set*

$$\{\gamma g \gamma^{-1} \in \text{Aut}(C) : \gamma \in \Gamma\}$$

is infinite for any $g \in \text{Aut}(C) \setminus \{e\}$.

We do not here present how to use this theorem. It can be shown that the kernel of the representation

$$\rho_x: \Lambda \rightarrow \text{Aut}(C), \quad \rho_x(\lambda) = \Phi([x, \lambda, y])\Phi([x, e, y])^{-1}$$

and the index $[\text{Aut}(C) : \rho_x(\Lambda)]$ are both finite, which implies Theorem 1.5. To construct this representation, we do not use special properties of the mapping class group other than the one in Theorem 3.10. In fact, this construction can be applied to a more general setting (see Theorem 6.1 in [36]).

Lattice embeddings of mapping class groups. We briefly give an outline of the proof of Theorem 1.6. We explain only the construction of Φ_0 in the statement. The reader is referred to Section 8 in [36] for more details.

Let M be a surface with $\kappa(M) > 0$ and let Γ be a finite index subgroup of $\Gamma(M)^\diamond$. Let G be a locally compact second countable group and let $\sigma: \Gamma \rightarrow G$ be a lattice embedding, i.e., an injective homomorphism such that the image $\sigma(\Gamma)$ is a lattice in G . As in Example 1.2, G is a self ME coupling of Γ (via σ). By Theorem 3.1, there exists an almost $(\Gamma \times \Gamma)$ -equivariant Borel map $\Phi: G \rightarrow \text{Aut}(C)$. By using Theorem 3.10 and the fact that the self ME coupling G of Γ is not only a measure space but a group, we can show that $\Phi(g_1 g_2) = \Phi(g_1)\Phi(g_2)$ for a.e. $(g_1, g_2) \in G \times G$. Recall the following theorem.

Theorem 3.11 ([63, Theorems B.2, B.3]). *If H_1, H_2 are locally compact second countable groups and $f: H_1 \rightarrow H_2$ is a Borel map such that $f(hh') = f(h)f(h')$ for a.e. $(h, h') \in H_1 \times H_1$, then there exists a continuous homomorphism $f_0: H_1 \rightarrow H_2$ such that f and f_0 are equal a.e. on H_1 .*

It follows that there exists a continuous homomorphism $\Phi_0: G \rightarrow \text{Aut}(C)$ such that Φ and Φ_0 are equal a.e. on G . It is easy to check that $K = \ker \Phi_0$ admits a finite invariant measure. Therefore, K is compact. After several easy observations, we see that this Φ_0 is a desired homomorphism.

3.3 OE rigidity

In this subsection, we briefly give another application of Theorem 3.1. We prove a rigidity result for ergodic standard actions of mapping class groups in terms of OE.

Corollary 3.12. *Let M be a surface with $\kappa(M) > 0$ and $M \neq M_{1,2}, M_{2,0}$. Put $\Gamma = \Lambda = \Gamma(M)^\diamond$. Let $\Gamma \curvearrowright (X, \mu)$ and $\Lambda \curvearrowright (Y, \nu)$ be ergodic standard (i.e., measure-preserving and essentially free) actions on standard finite measure spaces. If the two actions are OE, then they are conjugate.*

Proof. Since the two actions are OE, there exists a measure space isomorphism $f: (X, \mu) \rightarrow (Y, \nu)$ such that

$$f(\Gamma x) = \Lambda f(x) \quad \text{for a.e. } x \in X.$$

One can then construct the OE cocycle $\alpha: \Gamma \times X \rightarrow \Lambda$ associated with f by the equation

$$f(\gamma x) = \alpha(\gamma, x)f(x) \quad \text{for } \gamma \in \Gamma \text{ and a.e. } x \in X.$$

By Theorem 2.32, we can construct an ME coupling (Ω, m) of Γ and Λ such that the ME cocycle associated with some fundamental domain of the Λ -action on Ω , which can be identified with X , is equal to α . In what follows, we denote the action $\Gamma \curvearrowright X$ by $(\gamma, x) \mapsto \gamma \cdot x$, using a dot.

It follows from Theorem 3.1 that there exists an almost $(\Gamma \times \Lambda)$ -equivariant Borel map $\Phi: \Omega \rightarrow G$, where $G = \Gamma(M)^\diamond$ and $\text{Aut}(C)$ are identified via the natural isomorphism $\pi: \Gamma(M)^\diamond \rightarrow \text{Aut}(C)$ (see Theorem 2.3). Let $\varphi: X \rightarrow G$ be the Borel map defined by $\varphi(x) = \Phi(x)$ for $x \in X$. Then

$$\begin{aligned} \varphi(\gamma \cdot x)\alpha(\gamma, x)\varphi(x)^{-1} &= \Phi(\gamma \cdot x)\alpha(\gamma, x)\Phi(x)^{-1} \\ &= \Phi(\gamma\alpha(\gamma, x)x)\alpha(\gamma, x)\Phi(x)^{-1} = \gamma\Phi(x)\Phi(x)^{-1} = \gamma \end{aligned}$$

for any $\gamma \in \Gamma$ and a.e. $x \in X$. Define a Borel map $f_\varphi: X \rightarrow Y$ by $f_\varphi(x) = \varphi(x)f(x)$ for $x \in X$. Then for any $\gamma \in \Gamma$ and a.e. $x \in X$,

$$\begin{aligned} f_\varphi(\gamma \cdot x) &= \varphi(\gamma \cdot x)f(\gamma \cdot x) = \varphi(\gamma \cdot x)\alpha(\gamma, x)f(x) \\ &= \gamma\varphi(x)f(x) = \gamma f_\varphi(x). \end{aligned}$$

Since the actions $\Gamma \curvearrowright (X, \mu)$ and $\Lambda \curvearrowright (Y, \nu)$ are both essentially free and f is a measure space isomorphism, the above equation implies that $f_\varphi: X \rightarrow Y$ is a measure space isomorphism. \square

Remark 3.13. We can show the following much stronger rigidity theorem than Corollary 3.12.

Theorem 3.14 ([37, Theorem 1.1]). *Let M be a surface with $\kappa(M) > 0$. If an ergodic standard action of a finite index subgroup Γ of $\Gamma(M)^\diamond$ and an ergodic standard action of a discrete group Λ are WOE, then the two actions are virtually conjugate.*

See Definition 1.3 in [37] for the definition of virtual conjugacy. In particular, the conclusion of this theorem implies that Γ and Λ are commensurable up to finite kernels. We refer to [17], [19], [37], [48], [55], [56], [60] for other rigidity results in terms of OE. See also the fourth remark in Section 8. These rigidity theorems and Theorem 3.14 sharply contrast with the following theorem due to Ornstein and Weiss.

Theorem 3.15 ([52]). *Let G_1 and G_2 be infinite amenable groups and suppose that G_i admits an ergodic standard action on a standard finite measure space (X_i, μ_i) for $i \in \{1, 2\}$. Then the two actions are OE.*

It is known that amenability of the acting group is preserved under OE. More precisely, let $G_i \curvearrowright (X_i, \nu_i)$ be an ergodic standard action of a discrete group G_i for $i \in \{1, 2\}$. If the two actions are OE and G_1 is amenable, then G_2 is also amenable (see Theorem 4.18 (i), (ii)). Therefore, Theorem 1.8 is a consequence of Corollary 2.34 and Theorem 3.15. Connes, Feldman, and Weiss [10] proved a generalization of Theorem 3.15 in terms of discrete measured equivalence relations (see Theorem 4.17).

It is well known that there are many non-conjugate ergodic standard actions of \mathbb{Z} as follows. Let Γ be a discrete group and let (X_0, μ_0) be a standard probability space, i.e., a standard Borel space with a probability measure. We assume that (X_0, μ_0) may contain an atom, whereas (X_0, μ_0) does not consist of a single atom.

The Bernoulli action of Γ associated with (X_0, μ_0) is the action of Γ on the product space $(X_0, \mu_0)^\Gamma = \prod_\Gamma (X_0, \mu_0)$ given by

$$\gamma(x_g)_{g \in \Gamma} = (x_{\gamma^{-1}g})_{g \in \Gamma}, \quad \gamma \in \Gamma, (x_g)_{g \in \Gamma} \in X_0^\Gamma.$$

It is a natural question to understand when Bernoulli actions of \mathbb{Z} arising from two different standard probability spaces are conjugate. Kolmogorov and Sinai introduced a conjugacy invariant for actions of \mathbb{Z} , called entropy, and showed that the entropy of Bernoulli actions of \mathbb{Z} can be computed in terms of (X_0, μ_0) and assumes all non-negative values. In particular, there exist continuously many conjugacy classes of ergodic actions of \mathbb{Z} . As the culmination of the study on this conjugacy problem, Ornstein [50], [51] proved that entropy is a complete invariant for Bernoulli actions of \mathbb{Z} , that is, two Bernoulli actions of \mathbb{Z} which have the same entropy are conjugate. Moreover, this theory of entropy was extended to the setting of Bernoulli actions of infinite amenable groups by Ornstein and Weiss [53].

4 Amenable discrete measured groupoids

In the study of discrete measured groupoids, amenability is one of the most important notions like amenability of groups. One can construct a discrete measured groupoid from a non-singular action of a discrete group on a standard measure space (see Example 2.20). If the groupoid associated with a non-singular action of a discrete group is amenable, then the action is said to be amenable. This notion was first introduced by Zimmer [62]. One advantage of studying amenability of a group action is that (the groupoid arising from) an amenable action of a group behaves like an amenable group even if the acting group is non-amenable. We can thus apply various techniques for amenable groups to study a non-amenable group via its amenable action. Another advantage of the study of amenable groupoids is that under a certain condition, we can easily decide whether a groupoid is amenable or not. Since amenability is invariant under isomorphism of groupoids, this property is often used to distinguish two groupoids.

In this section, we recall the definition of amenable discrete measured groupoids and some of their fundamental properties. References for the material of this section is [4], [5] and Chapter 4 in [63]. We recommend the reader to consult [54] for applications of amenable actions of groups.

As discussed in Section 3, our final goal is to prove Theorem 3.6. For this purpose, we analyze various subgroupoids of the groupoid arising from a measure-preserving action of the mapping class group. It will be often necessary to prove amenability of some subgroupoids. To prove it, we make use of the amenability (in a measurable sense) of the action of the mapping class group on the boundary ∂C of the curve complex C . This fact will be explained at the end of this section. We note here that in

this chapter, this amenability of the boundary action will be used only in the proof of Theorem 5.10.

Amenable groups. We first recall the notion of amenability of discrete groups. Although we can proceed to most parts of this section under the assumption that a group is locally compact and second countable, we always assume that a group is discrete for simplicity. We refer to Section 4.1 in [63] for amenability of locally compact second countable groups. Although there are many equivalent definitions of amenability of groups, we recall only the definition which motivated Zimmer to define amenability of a group action.

Let G be a discrete group. Let A be a non-empty compact convex subset in the closed unit ball of E^* , where E is a separable Banach space and its dual E^* is equipped with the weak*-topology. Suppose that G acts on E by isometric isomorphisms and that A is invariant for the induced action of G on E^* . Such an action of G is called an *affine action* on A .

Definition 4.1. Let G be a discrete group. We say that G is *amenable* if for every affine action of G on a space A like above, there exists a fixed point, that is, $a \in A$ such that $ga = a$ for any $g \in G$.

Example 4.2. We refer to Section 4 in [63] for the proof of the following facts.

- (i) Finite groups and abelian groups are both amenable.
- (ii) Let $1 \rightarrow A \rightarrow B \rightarrow C \rightarrow 1$ be an exact sequence of discrete groups. Then A and C are both amenable if and only if B is amenable. Hence, all solvable groups are amenable.
- (iii) If G is a discrete group and $\{H_i\}_{i \in I}$ is a directed set of amenable subgroups of G , then the union $\bigcup_{i \in I} H_i$ is also amenable. For example, the following groups are amenable:
 - The direct product $\bigoplus_{n \in \mathbb{N}} H_n$. Here, H_n is an amenable group.
 - The infinite symmetric group $\mathfrak{S}_\infty = \bigcup_{n \in \mathbb{N}} \mathfrak{S}_n$. Here, \mathfrak{S}_n is the symmetric group on n letters, and \mathfrak{S}_n is identified with the subgroup of \mathfrak{S}_{n+1} fixing the $(n+1)$ -st letter.
- (iv) Non-abelian free groups are typical examples of non-amenable groups. Therefore, every group containing a non-abelian free subgroup is non-amenable.

Example 4.3. Let G be an amenable group and suppose that G acts on a separable compact space K continuously. We denote by $M(K)$ the space of all probability measures on K with the weak*-topology, on which G acts continuously. Note that $M(K)$ is a weak*-closed, convex subset of the closed unit ball of $C(K)^*$, where $C(K)$ is the Banach space of \mathbb{C} -valued continuous functions on K with the sup norm. By the definition of amenability, there exists $\mu \in M(K)$ such that $g\mu = \mu$ for any $g \in G$.

If G is an infinite amenable subgroup of the mapping class group $\Gamma(M)$ for a surface M with $\kappa(M) \geq 0$, then it follows from Theorem 2.7 that there exists a non-empty finite subset $S \subset \mathcal{PMF}$ such that $gS = S$ for any $g \in G$. More explicitly, if G is IA, then we put $S = \{F_{\pm}(g)\}$ for some pseudo-Anosov element $g \in G$. If G is reducible, then there exists $\sigma \in S(M)$ fixed by all elements of G , and we put $S = \{\sigma\}$. The uniformly distributed probability measure on S is then a fixed point for the action of G on $M(\mathcal{PMF})$. Therefore, in this case, we can explicitly find an invariant probability measure on \mathcal{PMF} for each amenable subgroup of $\Gamma(M)$.

Amenable groupoids. Zimmer [62] defined amenability of a group action as an analogue of Definition 4.1. The following definition of an amenable discrete measured groupoid is introduced in Chapter 4 of [4], which is a generalization of Zimmer's definition. A precise definition of amenable discrete measured groupoids is somehow complicated. After giving it, we recall several fundamental facts. The readers unfamiliar with this notion are recommended to consult [5], where a survey of amenability of groupoids is given.

When we defined amenability of groups, we considered an action of it on a separable Banach space. In the definition of amenability of groupoids \mathcal{G} , it is necessary to consider measurable bundles over the unit space of \mathcal{G} whose fiber is an object appearing in the definition of amenability of groups. We first introduce an object on which a groupoid acts, called a measurable Banach bundle. A reference for the material in the following Definitions 4.4 and 4.6 is Chapter II in [15]. In the first definition, we shall recall basic terminology in measure theory. Recall that we refer to a standard Borel space X equipped with a σ -finite positive measure μ as a standard measure space. If $\mu(X) < \infty$, then we say that (X, μ) is a standard finite measure space.

Definition 4.4 ([15, II.1]). Let (X, μ) be a standard measure space. We denote by \mathcal{B} the set of all Borel subsets of X .

- (i) A subset A of X is μ -null if there exists a countable family $\{A_n\}_n$ of elements of \mathcal{B} such that $A \subset \bigcup_n A_n$ and $\mu(A_n) = 0$ for all n .
- (ii) A subset A of X is μ -measurable if the symmetric difference $A \Delta B$ is μ -null for some $B \in \mathcal{B}$.
- (iii) A property of points of X which holds for all x outside some μ -null subset of X is said to hold for μ -almost every (or μ -a.e.) x .
- (iv) A map $f: X \rightarrow Y$ into a standard Borel space Y is μ -measurable if $f^{-1}(A)$ is μ -measurable for any Borel subset A of Y .

The following lemma is an easy exercise. For the proof, note that the σ -field of Borel subsets of a standard Borel space is generated by countably many Borel subsets of it as a σ -field.

Lemma 4.5. *Let (X, μ) be a standard measure space. If $\varphi: X \rightarrow Y$ is a μ -measurable map into a standard Borel space Y , then there exist a Borel map $\psi: X \rightarrow Y$ and a Borel subset X' of X such that $\mu(X \setminus X') = 0$ and $\varphi(x) = \psi(x)$ for all $x \in X'$.*

We next introduce the notion of measurable Banach bundles over a standard measure space (X, μ) . Suppose that for each $x \in X$, we are given a Banach space E_x . We refer to a function f on X such that $f(x) \in E_x$ for each $x \in X$ as a *vector field* on X . We will define measurability of such a vector field. We equip the complex field \mathbb{C} with the structure of a standard Borel space associated with the usual topology of \mathbb{C} .

Definition 4.6 ([15, II.4]). In the above notation, a μ -measurable structure for the family $\{E_x\}_{x \in X}$ is a non-empty family \mathcal{M} of vector fields on X satisfying the following four conditions:

- (i) If $f, g \in \mathcal{M}$, then the vector field $x \mapsto f(x) + g(x)$ is also in \mathcal{M} .
- (ii) If $f \in \mathcal{M}$ and a map $\phi: X \rightarrow \mathbb{C}$ is μ -measurable, then the vector field $x \mapsto \phi(x)f(x)$ is also in \mathcal{M} .
- (iii) If $f \in \mathcal{M}$, then the function $x \mapsto \|f(x)\|$ is μ -measurable, where $\|\cdot\|$ is the norm on E_x .
- (iv) Suppose that f is a vector field on X such that there exists a sequence $\{g_n\}$ of elements of \mathcal{M} such that $g_n(x) \rightarrow f(x)$ in E_x as $n \rightarrow \infty$ for μ -a.e. $x \in X$. Then $f \in \mathcal{M}$.

The family $\{E_x\}_{x \in X}$ endowed with this structure \mathcal{M} is called a *measurable Banach bundle* over (X, μ) , and is denoted by E . We refer to an element of \mathcal{M} as a *measurable section* for the bundle E .

In the next definition, we introduce the notion of separability for a measurable Banach bundle.

Definition 4.7 ([4, Definition A.3.4]). Let $E = (\{E_x\}_{x \in X}, \mathcal{M})$ be a measurable Banach bundle over a standard measure space (X, μ) . We say that E is *separable* if there exists a sequence $\{g_n\}_n$ of elements of \mathcal{M} such that the set $\{g_n(x)\}_n$ is total in E_x for μ -a.e. $x \in X$, that is, the set of all finite \mathbb{C} -linear combinations of elements in $\{g_n(x)\}_n$ is dense in E_x .

Remark 4.8. Let $E = (\{E_x\}_{x \in X}, \mathcal{M})$ be a measurable Banach bundle over a standard measure space (X, μ) .

- (i) Let (Y, ν) be a standard measure space and suppose that we are given a Borel map $\pi: Y \rightarrow X$ such that $\pi_*\nu$ and μ are equivalent. The set $\mathcal{N} = \{f \circ \pi : f \in \mathcal{M}\}$ generates a ν -measurable structure $\pi^*\mathcal{M}$ for the family $\{E_{f(y)}\}_{y \in Y}$. We denote by π^*E the corresponding bundle over (Y, ν) and call it the *pull-back* of E by π . If E is separable, then so is π^*E (see Example (3) of Appendix A in [4]).
- (ii) Consider the family $\{E_x^*\}_{x \in X}$ of duals. We denote by \mathcal{M}^* the set of all vector fields φ for this family such that the function $x \mapsto \langle \varphi(x), f(x) \rangle$ is μ -measurable for all $f \in \mathcal{M}$. The following fact is known (see Lemma A.3.7 in [4]): If E is separable, then $E^* = (\{E_x^*\}_{x \in X}, \mathcal{M}^*)$ is a measurable Banach bundle.

- (iii) When E is separable, we denote by $L^\infty(X, E^*)$ the space of all $\varphi \in \mathcal{M}^*$ such that the function $x \mapsto \|\varphi(x)\|$ belongs to $L^\infty(X)$, and we denote by $\|\varphi\|_\infty$ the μ -essential supremum for this function. It is known that $L^\infty(X, E^*)$ is a Banach space with respect to the norm $\|\cdot\|_\infty$ (see Proposition A.3.9 in [4]).

We next define an action of a discrete measured groupoid on a measurable Banach bundle.

Definition 4.9 ([4, Definition 4.1.1]). Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) . A *measurable \mathcal{G} -bundle* over (X, μ) is a pair (E, L) , where $E = (\{E_x\}_{x \in X}, \mathcal{M})$ is a measurable Banach bundle over (X, μ) , and L is a linear isometric representation of \mathcal{G} on E . Namely,

- for each $\gamma \in \mathcal{G}$, L gives an isometric isomorphism $L(\gamma): E_{s(\gamma)} \rightarrow E_{r(\gamma)}$;
- L preserves products, i.e., $L(\gamma_1\gamma_2) = L(\gamma_1)L(\gamma_2)$ for all $(\gamma_1, \gamma_2) \in \mathcal{G}^{(2)}$;
- L is measurable in the sense that for each $f \in \mathcal{M}$, the vector field $\gamma \mapsto L(\gamma)f(s(\gamma))$ for the family $\{E_{r(\gamma)}\}_{\gamma \in \mathcal{G}}$ is in $r^*\mathcal{M}$, where $r: \mathcal{G} \rightarrow X$ is the range map.

Remark 4.10. In Definition 4.9, assume that E is separable. The pair (E^*, L^*) defined by the following equation then gives a measurable \mathcal{G} -bundle over (X, μ) , and we call it the *dual \mathcal{G} -bundle* of the \mathcal{G} -bundle E :

$$\langle L^*(\gamma)e^*, e \rangle = \langle e^*, L(\gamma^{-1})e \rangle \quad \text{for } \gamma \in \mathcal{G}, e^* \in E_{s(\gamma)}^*, e \in E_{r(\gamma)}.$$

The next definition introduces the notion corresponding to convex, weak*-closed subsets contained in the closed unit ball of the dual of a separable Banach space appearing in the definition of amenability of groups.

Definition 4.11 ([4, Definitions 4.2.1, 4.2.5]). Let (X, μ) be a standard measure space.

- (i) Let $E = (\{E_x\}_{x \in X}, \mathcal{M})$ be a separable measurable Banach bundle over (X, μ) . Suppose that for each $x \in X$, we are given a subset A_x of the closed unit ball of the dual E_x^* . We refer to the family $A = \{A_x\}_{x \in X}$ as a *measurable field* for the dual E^* if there exists a sequence $\{\psi_n\}_n$ of elements of $L^\infty(X, E^*)$ such that A_x is the closed convex hull of the set $\{\psi_n(x)\}_n$ for μ -a.e. $x \in X$.
- (ii) Let \mathcal{G} be a discrete measured groupoid on (X, μ) and let (E, L) be a separable measurable \mathcal{G} -bundle over (X, μ) . A measurable field $A = \{A_x\}_{x \in X}$ for the dual E^* is called a *\mathcal{G} -field* if $L^*(\gamma)A_{s(\gamma)} = A_{r(\gamma)}$ for $\tilde{\mu}$ -a.e. $\gamma \in \mathcal{G}$, where $\tilde{\mu}$ is the measure on \mathcal{G} introduced in Definition 2.13.

Finally, we define amenability of discrete measured groupoids as follows.

Definition 4.12. A discrete measured groupoid \mathcal{G} on a standard measure space (X, μ) is *amenable* if the following holds: For any separable measurable \mathcal{G} -bundle (E, L) over (X, μ) and for any \mathcal{G} -field $A = \{A_x\}_{x \in X}$ for the dual E^* , there exists $\varphi \in L^\infty(X, E^*)$ such that

- $\varphi(x) \in A_x$ for μ -a.e. $x \in X$;
- $L^*(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma))$ for $\tilde{\mu}$ -a.e. $\gamma \in \mathcal{G}$.

When the discrete measured groupoid arising from a non-singular action of a discrete group on a standard finite measure space is amenable, we say that the action is *amenable*.

Note that in Zimmer's definition of amenable actions of groups, only constant Banach bundles (i.e., bundles $\{E_x\}_{x \in X}$ such that E_x is the same for all $x \in X$) are considered instead of general Banach bundles as above. However, Zimmer's definition is equivalent to the above one. The proof of this fact is given in Theorem 4.2.7 in [4] and Section 3 in [3].

Compared with the definition of amenability of groups in Definition 4.1, the second condition for φ in Definition 4.12 can be phrased by saying that φ is a fixed point for the action L of the groupoid \mathcal{G} on E . In general, given a group G and a space S , we refer to a homomorphism $G \rightarrow \text{Aut}(S)$ as an action of G on S . Hence, given a groupoid \mathcal{G} , we should refer to a groupoid homomorphism $\mathcal{G} \rightarrow \text{Aut}(S)$ as an action of \mathcal{G} on S . We next define a fixed point for such an action of a groupoid. However, for a standard Borel space S , we know no natural Borel structure on $\text{Aut}(S)$, the group of Borel automorphisms of S . Hence, we consider nothing but the following special action of a groupoid when the groupoid admits a Borel structure. In what follows, a groupoid homomorphism from a discrete measured groupoid \mathcal{G} into a discrete group Γ is always assumed to be Borel as a map from \mathcal{G} into Γ .

Definition 4.13. Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) . Let S be a standard Borel space. Suppose that we are given a Borel action of a discrete group Γ on S and a groupoid homomorphism $\rho: \mathcal{G} \rightarrow \Gamma$. Then a Borel map $\varphi: X \rightarrow S$ satisfying the equation

$$\rho(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma)) \quad \text{for a.e. } \gamma \in \mathcal{G}$$

is called an *invariant Borel map* for \mathcal{G} . We say that φ is ρ -invariant for \mathcal{G} when we specify ρ .

More generally, if A is a Borel subset of X and if a Borel map $\varphi: A \rightarrow S$ satisfies the above equation for a.e. $\gamma \in (\mathcal{G})_A$, then we say for simplicity that φ is *invariant* for \mathcal{G} although we should say that φ is invariant for $(\mathcal{G})_A$.

Given an action of a groupoid, we often use amenability of the groupoid to obtain an invariant Borel map for the action as shown in the following proposition. Recall that for a separable compact space K , we denote by $M(K)$ the space of probability measures on K . This space is a convex, weak*-closed subset contained in the closed unit ball of the dual of $C(K)$, the Banach space of \mathbb{C} -valued continuous functions on K with the sup norm.

Proposition 4.14. Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) . Let Γ be a discrete group and suppose that Γ acts on a separable

compact space K continuously. Let $\rho: \mathcal{G} \rightarrow \Gamma$ be a groupoid homomorphism. If \mathcal{G} is amenable, then there exists a ρ -invariant Borel map $\varphi: X \rightarrow M(K)$, i.e., a Borel map satisfying the equation

$$\rho(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma)) \quad \text{for a.e. } \gamma \in \mathcal{G}.$$

Here, the action $\Gamma \curvearrowright M(K)$ is given by the induced one from the action $\Gamma \curvearrowright K$.

Proof. We put $E_x = C(K)$ for all $x \in X$. We define a μ -measurable structure \mathcal{M} for the family $\{E_x\}_{x \in X}$ as the one generated by the constant vector fields $x \mapsto e$ for all $e \in C(K)$. Then $E = (\{E_x\}_{x \in X}, \mathcal{M})$ is a separable measurable Banach bundle. For $\gamma \in \mathcal{G}$, we define an isometric isomorphism $L(\gamma): E_{s(\gamma)} \rightarrow E_{r(\gamma)}$ by $L(\gamma)e = \rho(\gamma)e$ for $e \in E_{s(\gamma)}$. The pair (E, L) is then a \mathcal{G} -bundle. Since the family $A = \{A_x\}_{x \in X}$ given by $A_x = M(K)$ defines a \mathcal{G} -field for E^* , we get $\varphi \in L^\infty(X, E^*)$ such that $\varphi(x) \in M(K)$ for μ -a.e. $x \in X$, and $L^*(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma))$ for $\tilde{\mu}$ -a.e. $\gamma \in \mathcal{G}$. This equation is equivalent to $\rho(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma))$. The proposition follows from Lemma 4.5. \square

Example 4.15. Let Γ be a discrete group and suppose that Γ admits a non-singular action on a standard measure space (X, μ) . We denote by \mathcal{G} the associated groupoid $\Gamma \ltimes X$. Then

$$\rho: \mathcal{G} \rightarrow \Gamma, \quad (g, x) \mapsto g$$

defines a groupoid homomorphism. Suppose that we are given a separable compact space K on which Γ acts continuously. It follows from Proposition 4.14 that if \mathcal{S} is an amenable subgroupoid of \mathcal{G} , then there exists a Borel map $\varphi: X \rightarrow M(K)$ such that $\rho(g, x)\varphi(x) = \varphi(gx)$, that is, $g\varphi(x) = \varphi(gx)$ for a.e. $(g, x) \in \mathcal{S}$.

We give fundamental properties of amenable discrete measured groupoids.

Theorem 4.16. *Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) .*

(i) *\mathcal{G} is amenable if and only if its quotient equivalence relation*

$$\{(r(\gamma), s(\gamma)) \in X \times X : \gamma \in \mathcal{G}\}$$

is amenable and for a.e. $x \in X$, the isotropy group $\mathcal{G}_x^x = \{\gamma \in \mathcal{G} : r(\gamma) = s(\gamma) = x\}$ is amenable.

(ii) *Any subgroupoid of an amenable discrete measured groupoid is amenable.*

(iii) *Let $A \subset X$ be a Borel subset with positive measure. If \mathcal{G} is amenable, then so is the restricted groupoid $(\mathcal{G})_A$. If $\mathcal{G}_A = X$ up to null sets, then the converse also holds.*

For Assertion (i), we refer to Corollary 5.3.33 in [4]. Assertion (ii) follows from hyperfiniteness of amenable equivalence relations shown in [10] and Assertion (i).

The former part of Assertion (iii) can be shown by using Assertion (ii) because $(\mathcal{G})_A$ is identified with the subgroupoid $(\mathcal{G})_A \cup \{e_x : x \in X \setminus A\}$ of \mathcal{G} , where $e_x \in \mathcal{G}$ denotes the unit on x . The latter part can also be proved directly by using this identification. The following is one of the most highlighted theorems about principal discrete measured groupoids, and it is a generalization of Theorem 3.15. Recall that \mathcal{G} is said to be principal if the isotropy group \mathcal{G}_x^x is trivial for each $x \in X$. A principal groupoid is isomorphic to its quotient equivalence relation.

Theorem 4.17 ([10]). *For $i \in \{1, 2\}$, let \mathcal{G}_i be ergodic principal discrete measured groupoids on a standard finite measure space (X_i, μ_i) . For $i \in \{1, 2\}$, we suppose that μ_i is invariant for \mathcal{G}_i and that μ_i has no atom, that is, there exists no point $x \in X_i$ with $\mu_i(\{x\}) > 0$. Then \mathcal{G}_1 and \mathcal{G}_2 are isomorphic.*

Here, a discrete measured groupoid \mathcal{G} on (X, μ) is said to be *ergodic* if the following holds: If a Borel subset $A \subset X$ satisfies the equation $\mathcal{G}A = A$ up to null sets, then either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$. In the next theorem, we particularly consider a groupoid arising from a group action.

Theorem 4.18. *Let (X, μ) and (Y, ν) be standard finite measure spaces.*

- (i) *Let G be a discrete group and suppose that we have a non-singular action of G on (X, μ) . Let \mathcal{G} be the associated groupoid. If G is amenable, then \mathcal{G} is amenable.*
- (ii) *Conversely, in Assertion (i), if the action $G \curvearrowright (X, \mu)$ is measure-preserving and \mathcal{G} is amenable, then G is amenable.*
- (iii) *Let G be a discrete group and suppose that we have non-singular actions $G \curvearrowright (X, \mu)$ and $G \curvearrowright (Y, \nu)$. If there exists a G -equivariant Borel map $f: X \rightarrow Y$ such that $f_*\mu = \nu$ and if the action $G \curvearrowright (Y, \nu)$ is amenable, then the action $G \curvearrowright (X, \mu)$ is also amenable.*

Assertion (i) follows from Propositions 4.2.2 and A.3.9 in [4]. For Assertions (ii) and (iii), we refer to Proposition 4.3.3 in [63] and [3], respectively. This subsection will end with several examples of amenable discrete measured groupoids.

Example 4.19. *Groupoids admitting fundamental domains.* Let G be a discrete group. Then the action of G on G by left multiplication is amenable, where a measure μ on G is given by $f \in \ell^1(G)$ such that $f(g) > 0$ for each $g \in G$. More generally, suppose that G admits a non-singular action on a standard finite measure space (X, μ) and suppose that the action admits a *fundamental domain*, that is, there exists a Borel subset $F \subset X$ such that $\mu(\bigcup_{g \in G} F) = \mu(X)$ and $Gx \cap F$ consists of a single point for a.e. $x \in X$. Then the action $G \curvearrowright (X, \mu)$ is amenable. Note that if G is infinite and G admits an essentially free, measure-preserving action on a standard finite measure space, then there exists no Borel fundamental domain for the action. A discrete measured equivalence relation which admits a fundamental domain is also amenable.

The following is an example of amenable actions of non-amenable groups.

Theorem 4.20 ([1], [4, Appendix B]). *Let Γ be an infinite hyperbolic group in the sense of Gromov. Let μ be a probability measure on the Gromov boundary $\partial\Gamma$ such that the action of Γ on $(\partial\Gamma, \mu)$ is non-singular. Then the action $\Gamma \curvearrowright (\partial\Gamma, \mu)$ is amenable.*

In the proof of this result, approximately invariant means for (the groupoid arising from) the boundary action of Γ are constructed. Recall that for a discrete group G , a sequence $\{f_n\}_{n \in \mathbb{N}}$ in $\ell^1(G)$ is called *approximately invariant means* for G if

- for each n , $f_n(g) \geq 0$ for all $g \in G$ and $\sum_{g \in G} f_n(g) = 1$;
- for each $g \in G$, $\sum_{h \in G} |f_n(g^{-1}h) - f_n(h)| \rightarrow 0$ as $n \rightarrow \infty$.

It is well known that a discrete group G is amenable if and only if G admits approximately invariant means. We can also define approximately invariant means for a discrete measured groupoid as an analogue of the above definition (see Chapter 3 in [4]), and we can show that a discrete measured groupoid is amenable if and only if there exist such means for it (see Theorem 4.2.7 in [4]). When we are given a group action and we want to show that it is amenable, we often prove that it admits approximately invariant means, for it is often difficult to prove the fixed point property in Definition 4.12 directly for concrete examples of group actions.

In Example 3.8 of [5] and Example 2.2 of [54], approximately invariant means for the boundary action of non-abelian free groups are constructed explicitly. This construction can be generalized to the case of hyperbolic groups by using the uniform thinness of all geodesic triangles on their Cayley graphs. Since the curve complex C for a surface M with $\kappa(M) \geq 0$ is hyperbolic (see Theorem 2.2), this proof motivates the following theorem. We denote by ∂C the Gromov boundary of C . It is known that ∂C is a non-empty standard Borel space (see Proposition 3.10 in [35]). We refer to [39], [24], [27] for details of the boundary ∂C .

Theorem 4.21 ([35, Theorem 3.29]). *Let M be a surface with $\kappa(M) \geq 0$ and let C be the curve complex for M . Let μ be a probability measure on the Gromov boundary ∂C such that the action of $\Gamma(M)^\diamond$ on $(\partial C, \mu)$ is non-singular. Then the action $\Gamma(M)^\diamond \curvearrowright (\partial C, \mu)$ is amenable.*

Since C is hyperbolic, we expect a construction of approximately invariant means for the action of $\Gamma(M)^\diamond$ on ∂C similar to the one for hyperbolic groups noted above. However, we can not apply the construction directly because C is locally infinite. To avoid this difficulty, we use the finiteness property of tight geodesics on the curve complex established by Masur and Minsky [44], and Bowditch [9]. A tight geodesic is a geodesic in C with a special property. Roughly speaking, the finiteness property of tight geodesics says that the set of tight geodesics behaves like the set of geodesics on a locally finite hyperbolic graph. Thanks to this property, we can construct approximately invariant means for the action of $\Gamma(M)^\diamond$ on ∂C as in the case of hyperbolic

groups. Geometric properties of the curve complex we use in the proof of Theorem 4.21 are only the hyperbolicity and this finiteness property. We omit the proof of Theorem 4.21 and note here that the finiteness property of tight geodesics will also be used in the construction of several natural Borel maps associated with the curve complex (see Remark 5.9).

In Section 5, we will use the following corollary, which is an immediate consequence of Theorem 4.21. We denote by $\partial_2 C$ the quotient space of $C \times C$ by the coordinate exchanging action of the symmetric group on two letters.

Corollary 4.22 ([35, Lemma 4.32]). *Let M be a surface with $\kappa(M) \geq 0$. Let μ be a probability measure on $\partial_2 C$ such that the action of $\Gamma(M)^\diamond$ on $(\partial_2 C, \mu)$ is non-singular. Then the action $\Gamma(M)^\diamond \curvearrowright (\partial_2 C, \mu)$ is amenable.*

5 Two types of subgroupoids: IA and reducible ones

Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ be a finite index subgroup of $\Gamma(M; m)$ (see Theorem 2.8 for the subgroup $\Gamma(M; m)$ of $\Gamma(M)$). Let \mathcal{G} be the discrete measured groupoid on a standard finite measure space (X, μ) which arises from a measure-preserving action $\Gamma \curvearrowright (X, \mu)$. The final goal of Sections 5, 6 and 7 is to prove Theorem 3.6. This theorem states that any isomorphism between such groupoids arising from actions of mapping class groups preserves subgroupoids generated by actions of Dehn twists. To characterize such subgroupoids algebraically in terms of groupoids, we introduce two types of subgroupoids of \mathcal{G} . The first one is called IA subgroupoids, which correspond to IA (= infinite, irreducible and amenable) subgroups in the classification theorem of subgroups of mapping class groups (see Theorem 2.7). The second one is called reducible subgroupoids, which correspond to infinite reducible subgroups.

Let $\rho: \mathcal{G} \rightarrow \Gamma$ be the groupoid homomorphism given by $(g, x) \mapsto g$ for $g \in \Gamma$ and $x \in X$. We denote by $M(\mathcal{PMF})$ the space of probability measures on the Thurston boundary \mathcal{PMF} . Each element $\gamma \in \mathcal{G}$ then acts on $M(\mathcal{PMF})$ via ρ . We can regard this assignment as the action of \mathcal{G} on $M(\mathcal{PMF})$. We define the above two classes of subgroupoids \mathcal{S} of \mathcal{G} in terms of Borel maps $\varphi: X \rightarrow M(\mathcal{PMF})$ which is ρ -invariant for \mathcal{S} , i.e., $\rho(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma))$ for a.e. $\gamma \in \mathcal{S}$. These ρ -invariant Borel maps play a role of fixed points for the action of \mathcal{G} on $M(\mathcal{PMF})$ (see Definition 4.13 and the comment right before it).

In Subsection 5.1, we characterize IA and reducible subgroups in terms of their fixed points in $M(\mathcal{PMF})$. This will help us to understand the motivation of the definition of IA and reducible subgroupoids. In Subsection 5.2, we analyze IA subgroupoids and study properties of Borel maps into $M(\mathcal{PMF})$ which are ρ -invariant for them. It is shown that IA subgroupoids are in fact amenable as groupoids. In Subsection 5.3, we study reducible subgroupoids \mathcal{S} and give the definition of canonical reduction systems for \mathcal{S} . This is an essentially unique Borel map into $S(M)$ which is ρ -invariant

for \mathcal{F} and which satisfies nice properties. This system is a generalization of canonical reduction systems for reducible subgroups introduced by Birman, Lubotzky, and McCarthy [8], and Ivanov [30].

5.1 IA and reducible subgroups

Let M be a surface with $\kappa(M) \geq 0$ and let Γ be an infinite subgroup of $\Gamma(M)$. Recall that Γ is said to be IA if there exists a pseudo-Anosov element $g \in \Gamma$ such that $\{F_{\pm}(g)\}$, the set of its pseudo-Anosov foliations, is fixed by all elements of Γ . In this case, Γ is virtually cyclic. If there exists $\sigma \in S(M)$ fixed by all elements of Γ , then Γ is said to be reducible. In the next two propositions, we characterize these two classes of subgroups in terms of their fixed points on the space $M(\mathcal{PMF})$ of probability measures on the Thurston boundary \mathcal{PMF} . We say that $\nu \in M(\mathcal{PMF})$ is *invariant* for a subgroup Γ of $\Gamma(M)$ if $g\nu = \nu$ for each $g \in \Gamma$.

Proposition 5.1. *Let M be a surface with $\kappa(M) \geq 0$ and let Γ be an infinite subgroup of $\Gamma(M)$. Then the following assertions hold:*

- (i) *The subgroup Γ is IA if and only if there exists an invariant measure $\nu \in M(\mathcal{PMF})$ for Γ such that $\nu(\mathcal{MLN}) = 1$.*
- (ii) *If Γ is IA, then any invariant measure $\nu \in M(\mathcal{PMF})$ for Γ satisfies that $\nu(\{F_{\pm}(g)\}) = 1$ for some pseudo-Anosov element $g \in \Gamma$.*

Proof. The “only if” part of Assertion (i) has already been seen in Example 4.3. Assertion (ii) follows from the dynamics of pseudo-Anosov elements on \mathcal{PMF} (see Theorem 2.5).

We may assume that Γ is a subgroup of $\Gamma(M; m)$ for an integer $m \geq 3$ to prove the “if” part of Assertion (i). Recall that $\Gamma(M; m)$ consists of pure elements and is torsion-free (see Theorem 2.8). Let $\nu \in M(\mathcal{PMF})$ be an invariant measure for Γ such that $\nu(\mathcal{MLN}) = 1$.

Assume that Γ contains a reducible element g of infinite order. Let Δ^u and Ψ^s be the subsets of \mathcal{PMF} associated with g as in the comment right before Theorem 2.6. We can choose a non-empty closed one-dimensional submanifold c of M which satisfies Property (P) for g and does not have superfluous components. Note that both Ψ^s and Δ^u are contained in $\mathcal{PMF} \setminus \mathcal{MLN}$. Let $\{U_n\}_n$ be a sequence of open subsets of \mathcal{PMF} such that $U_n \supset U_{n+1}$ for each n , and $\Delta^u = \bigcap_n U_n$. It follows from $\nu(\Delta^u) = 0$ that $\nu(U_n) \searrow 0$. Let K be any compact subset of $\mathcal{PMF} \setminus \Psi^s$. By Theorem 2.6, for each n , there exists N such that $g^N K \subset U_n$, and thus $\nu(K) = \nu(g^N K) \leq \nu(U_n)$. Therefore, $\nu(K) = 0$. Since $\mathcal{PMF} \setminus \Psi^s$ can be expressed as a countable union of compact subsets, this implies that $\nu(\mathcal{PMF} \setminus \Psi^s) = 0$, which is a contradiction.

Thus, Γ does not contain a reducible element of infinite order, and it consists of pseudo-Anosov elements and the trivial element. If g is a pseudo-Anosov element of Γ , then by the dynamics of g on \mathcal{PMF} (see Theorem 2.5), the support of ν is contained in $\{F_{\pm}(g)\}$. Since ν is invariant for Γ , this implies that Γ is an IA subgroup. \square

Proposition 5.2. *Let M be a surface with $\kappa(M) \geq 0$ and let Γ be an infinite subgroup of $\Gamma(M)$. Then Γ is reducible if and only if there exists an invariant measure $\nu \in M(\mathcal{PMF})$ for Γ such that $\nu(\mathcal{PMF} \setminus \mathcal{MLN}) = 1$.*

Proof. The “only if” part has already been seen in Example 4.3. To prove the “if” part, let $\nu \in M(\mathcal{PMF})$ be an invariant measure for Γ such that $\nu(\mathcal{PMF} \setminus \mathcal{MLN}) = 1$. It follows from Theorem 2.7 that Γ is either IA, reducible or sufficiently large. By Proposition 5.1 (ii), Γ is not IA. If Γ were sufficiently large, then there exist pseudo-Anosov elements $g_1, g_2 \in \Gamma$ such that $\{F_{\pm}(g_1)\} \cap \{F_{\pm}(g_2)\} = \emptyset$. Theorem 2.5 implies that any sufficiently large subgroup admits no invariant probability measure on \mathcal{PMF} . Therefore, Γ is reducible. \square

We next define IA and reducible subgroupoids, which is motivated by the above two propositions. We often use the following notation in what follows.

Notation. We refer to the following assumption as (\star) : Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ be a finite index subgroup of $\Gamma(M; m)$. Suppose that Γ admits a measure-preserving action on a standard finite measure space (X, μ) . We denote by \mathcal{G} the associated groupoid $\Gamma \times (X, \mu)$. Let $\rho: \mathcal{G} \rightarrow \Gamma$ be the groupoid homomorphism defined by $(g, x) \mapsto g$.

Propositions 5.1 and 5.2 imply that there exists no infinite subgroup of $\Gamma(M)$ which admits an invariant measure $\nu \in M(\mathcal{PMF})$ such that $0 < \nu(\mathcal{MLN}) < 1$. The following is a generalization of this fact.

Theorem 5.3. *Under Assumption (\star) , let Y be a Borel subset of X with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Suppose that there is an invariant Borel map $\varphi: Y \rightarrow M(\mathcal{PMF})$ for \mathcal{S} . Then there exists a Borel partition $Y = Y_1 \sqcup Y_2$ such that*

- $\varphi(x)(\mathcal{MLN}) = 1$ for a.e. $x \in Y_1$;
- $\varphi(x)(\mathcal{PMF} \setminus \mathcal{MLN}) = 1$ for a.e. $x \in Y_2$.

Recall that a Borel map $\varphi: Y \rightarrow M(\mathcal{PMF})$ is (ρ) -invariant for \mathcal{S} if the equation $\rho(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma))$ holds for a.e. $\gamma \in \mathcal{S}$.

Remark 5.4. In Theorem 5.3, let us assume that there is another invariant Borel map $\psi: Y \rightarrow M(\mathcal{PMF})$ for \mathcal{S} . It is easy to check that ψ also satisfies

- $\psi(x)(\mathcal{MLN}) = 1$ for a.e. $x \in Y_1$;
- $\psi(x)(\mathcal{PMF} \setminus \mathcal{MLN}) = 1$ for a.e. $x \in Y_2$.

for the same Y_1 and Y_2 as in the theorem. (Consider the invariant Borel map $(\varphi + \psi)/2$ for \mathcal{S} and apply the theorem.)

By this remark, the two subgroupoids $(\mathcal{S})_{Y_1}$ and $(\mathcal{S})_{Y_2}$ should be distinguished, and it is natural to define the following two classes of subgroupoids.

Definition 5.5. Under Assumption (\star) , let Y be a Borel subset of X with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type.

- (i) We say that \mathcal{S} is IA (= irreducible and amenable) if there is an invariant Borel map $\varphi: Y \rightarrow M(\mathcal{PMF})$ for \mathcal{S} such that $\varphi(x)(\mathcal{MLN}) = 1$ for a.e. $x \in Y$.
- (ii) We say that \mathcal{S} is *reducible* if there is an invariant Borel map $\varphi: Y \rightarrow M(\mathcal{PMF})$ for \mathcal{S} such that $\varphi(x)(\mathcal{PMF} \setminus \mathcal{MLN}) = 1$ for a.e. $x \in Y$.

It follows from Remark 5.4 that the classes of IA and reducible subgroupoids are mutually exclusive. The definition of reducible subgroupoids is also motivated by the following lemma.

Lemma 5.6. *Under Assumption (\star) , let Y be a Borel subset of X with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Then the following two conditions are equivalent:*

- (i) \mathcal{S} is reducible.
- (ii) There exists an invariant Borel map $Y \rightarrow S(M)$ for \mathcal{S} .

It is clear that Assertion (ii) implies Assertion (i) because there is a $\Gamma(M)^\diamond$ -equivariant embedding $\iota: S(M) \rightarrow \mathcal{PMF} \setminus \mathcal{MLN}$. To prove the converse implication, we construct a Borel map $H: \mathcal{PMF} \setminus \mathcal{MLN} \rightarrow S(M)$ which is equivariant for the action of $\Gamma(M)^\diamond$ and satisfies $H \circ \iota = \text{id}$ (see Subsection 4.2 in [35]).

5.2 IA subgroupoids

The following lemma is the first important observation about invariant Borel maps for IA subgroupoids. It is known that there exists a natural $\Gamma(M)^\diamond$ -equivariant map $\pi: \mathcal{MLN} \rightarrow \partial C$, which is continuous and surjective (see [39]). We can define a Borel structure on the set $M(\partial C)$ of all probability measures on ∂C by using a Borel section of $\pi: \mathcal{MLN} \rightarrow \partial C$, i.e., a Borel map $s: \partial C \rightarrow \mathcal{MLN}$ such that $\pi \circ s = \text{id}$ (see the comment right before Proposition 4.30 in [35]). For a technical reason, we study invariant Borel maps into $M(\partial C)$ for IA subgroupoids instead of ones into $M(\mathcal{PMF})$.

Lemma 5.7. *Under Assumption (\star) , let Y be a Borel subset of X with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Then the following assertions hold:*

- (i) \mathcal{S} is IA if and only if there exists an invariant Borel map $\varphi: Y \rightarrow M(\partial C)$ for \mathcal{S} .
- (ii) If \mathcal{S} is IA and $\varphi: Y \rightarrow M(\partial C)$ is an invariant Borel map for \mathcal{S} , then $\text{supp}(\varphi(x))$ consists of at most two points.

Here, for a measure ν , we denote by $\text{supp}(\nu)$ the support of ν . It is easy to see the “only if” part of Assertion (i) by using the map $\pi: \mathcal{MLN} \rightarrow \partial C$.

We denote by $\partial_2 C$ the quotient space of $\partial C \times \partial C$ by the coordinate exchanging action of the symmetric group on two letters. Then $\partial_2 C$ can be viewed as a Borel

subset of $M(\partial C)$ by regarding each element of $\partial_2 C$ as an atomic measure on ∂C such that each atom has measure 1 or $1/2$. We denote by $M(\mathcal{M}\mathcal{I}\mathcal{N})$ the Borel subset of $M(\mathcal{P}\mathcal{M}\mathcal{F})$ consisting of all measures ν such that $\nu(\mathcal{M}\mathcal{I}\mathcal{N}) = 1$. We can prove the following lemma by using Lemma 5.7.

Lemma 5.8. *Under Assumption (\star) , let Y be a Borel subset of X with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. If \mathcal{S} is IA, then there exists an essentially unique invariant Borel map $\varphi_0: Y \rightarrow \partial_2 C$ for \mathcal{S} satisfying the following condition: If $Y' \subset Y$ is a Borel subset with positive measure and $\varphi: Y' \rightarrow M(\partial C)$ is an invariant Borel map for \mathcal{S} , then*

$$\text{supp}(\varphi(x)) \subset \text{supp}(\varphi_0(x)) \quad \text{for a.e. } x \in Y'.$$

This unique invariant Borel map plays an important role when we study the normalizer of an IA subgroupoid (see Lemma 6.7).

Remark 5.9. In the proof of Propositions 5.1 and 5.2, it was important to observe the dynamics of each element of the mapping class group on $\mathcal{P}\mathcal{M}\mathcal{F}$. However, we cannot consider the dynamics of each element of a groupoid because powers γ^n of an element γ of a groupoid do not make sense in general. Hence, we cannot apply a similar argument in the proof of Theorem 5.3 and Lemma 5.7. As a different approach, by using the finiteness properties of tight geodesics in the curve complex (see Theorem 4.21 and the comment around it), we construct the following natural Borel maps: Put

$$\delta C = \{(a, b, c) \in (\partial C)^3 : a \neq b \neq c \neq a\}$$

and define an action of $\Gamma(M)^\diamond$ on δC by $g(a, b, c) = (ga, gb, gc)$. Let $\mathcal{F}'(C)$ be the set of all non-empty finite subsets of $V(C)$ whose diameters are at least three, on which $\Gamma(M)^\diamond$ naturally acts. We can then construct a $\Gamma(M)^\diamond$ -equivariant Borel map

$$MS: \delta C \rightarrow \mathcal{F}'(C)$$

(see Section 4.1 in [35]). A remarkable property of the set $\mathcal{F}'(C)$ is that the stabilizer of each element of $\mathcal{F}'(C)$ is finite (see Lemma 10 in [7]). Moreover, in the proof of Theorem 5.3, we construct a $\Gamma(M)^\diamond$ -equivariant Borel map

$$G: \partial_2 C \times V(C) \rightarrow \mathcal{F}'(C),$$

where the action of $\Gamma(M)^\diamond$ on $\partial_2 C \times V(C)$ is given by $g(a, x) = (ga, gx)$ (see Lemma 4.40 in [35]). In this chapter, we do not further mention the proof of these facts.

As observed in Proposition 5.1, if an infinite subgroup Γ of $\Gamma(M; m)$ has an invariant measure $\nu \in M(\mathcal{P}\mathcal{M}\mathcal{F})$ such that $\nu(\mathcal{M}\mathcal{I}\mathcal{N}) = 1$, then Γ is IA and in particular, Γ is amenable. Hence, we can expect any IA subgroupoid \mathcal{S} to be amenable, which is in fact shown in the following theorem. In the proof of this theorem, we use the

amenability of the action $\Gamma(M)^\diamond$ on ∂C shown in Theorem 4.21 and Corollary 4.22. We give the proof of this theorem to show how to use this amenable action of the mapping class group.

Theorem 5.10. *Under Assumption (\star) , let Y be a Borel subset of X with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. If \mathcal{S} is IA, then \mathcal{S} is amenable. Equivalently, if there is an invariant Borel map $Y \rightarrow \partial_2 C$ for \mathcal{S} , then \mathcal{S} is amenable.*

Note that \mathcal{S} is IA if and only if there exists an invariant Borel map $Y \rightarrow \partial_2 C$ for \mathcal{S} (see Lemma 5.7 (ii)). Let $\varphi: Y \rightarrow \partial_2 C$ be an invariant Borel map for \mathcal{S} . An important point of the proof is to construct a standard Borel space S on which Γ acts so that

- \mathcal{S} is isomorphic to a subgroupoid of $\Gamma \ltimes S$;
- we can construct a Γ -equivariant Borel map $S \rightarrow \partial_2 C$ by using φ .

If we can construct such a space S , then the theorem follows from Theorem 4.16 (ii), Theorem 4.18 (iii) and Corollary 4.22.

Proof of Theorem 5.10. We identify \mathcal{S} with the groupoid on (X, μ) defined by the union $\{e_x \in \mathcal{G} : x \in X \setminus Y\} \cup \mathcal{S}$. Extend φ to the map from X defined by $\varphi(x) = a_0$ for $x \in X \setminus Y$, where $a_0 \in \partial_2 C$ is some fixed point. We denote by the same symbol φ the extended map. The extended map φ is then also invariant for \mathcal{S} .

Consider the action of Γ on $X \times \Gamma$ given by

$$g(x, g_1) = (x, g_1 g^{-1}) \quad \text{for } x \in X, g, g_1 \in \Gamma.$$

The equivalence relation \mathcal{R}_1 on $X \times \Gamma$ defined by

$$(s(\gamma), g) \sim (r(\gamma), \rho(\gamma)g) \quad \text{for } \gamma \in \mathcal{G}, g \in \Gamma$$

admits a fundamental domain $F_1 = X \times \{e\}$, i.e., a Borel subset F_1 of the unit space $X \times \Gamma$ such that $\mathcal{R}_1 x \cap F_1$ consists of exactly one point for a.e. $x \in X \times \Gamma$, where $\mathcal{R}_1 x$ denotes the equivalence class containing x . Let \mathcal{R}_2 be the equivalence relation on $X \times \Gamma$ given by

$$(s(\gamma), g) \sim (r(\gamma), \rho(\gamma)g) \quad \text{for } \gamma \in \mathcal{S}, g \in \Gamma.$$

Since \mathcal{R}_2 is a subrelation of \mathcal{R}_1 , we can show that \mathcal{R}_2 also admits a fundamental domain $F_2 \subset X \times \Gamma$ (use Lemma 2.12 in [2]). Let S be the quotient space of $X \times \Gamma$ by \mathcal{R}_2 , which is identified with F_2 as a measure space via the projection $X \times \Gamma \rightarrow S$. Note that the action of Γ on $X \times \Gamma$ induces an action of Γ on S . Denote the projection of $(x, g) \in X \times \Gamma$ onto S by $[x, g] \in S$. Then \mathcal{S} can be identified with a Borel subgroupoid

$$\mathcal{H} = \{(\rho(\gamma), [s(\gamma), e]) \in \Gamma \ltimes S : \gamma \in \mathcal{S}\}$$

of $\Gamma \ltimes S$ via an isomorphism

$$\mathcal{S} \ni \gamma \mapsto (\rho(\gamma), [s(\gamma), e]) \in \mathcal{H}.$$

Using the invariant Borel map $\varphi: X \rightarrow \partial_2 C$ for \mathcal{S} , we construct a Borel map $\varphi': S \rightarrow \partial_2 C$ by the formula $[x, g] \mapsto g^{-1}\varphi(x)$. Then φ' is well-defined and Γ -equivariant. By Theorem 4.18 (iii) and Corollary 4.22, the groupoid $\Gamma \times S$ is amenable. Since \mathcal{H} is a subgroupoid of $\Gamma \times S$, it is also amenable. \square

5.3 Reducible subgroupoids

We shall recall Assumption (\star): Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ be a finite index subgroup of $\Gamma(M; m)$. Suppose that Γ admits a measure-preserving action on a standard finite measure space (X, μ) . We denote by \mathcal{G} the associated groupoid $\Gamma \times (X, \mu)$. Let $\rho: \mathcal{G} \rightarrow \Gamma$ be the groupoid homomorphism defined by $(g, x) \mapsto g$.

Let Y be a Borel subset of X with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Suppose that \mathcal{S} is reducible. By Lemma 5.6, we can construct an invariant Borel map $\varphi: Y \rightarrow S(M)$ for \mathcal{S} , i.e., $\rho(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma))$ for a.e. $\gamma \in \mathcal{S}$. In general, there are many such maps φ . The aim of this subsection is to construct an invariant Borel map $Y \rightarrow S(M)$ for \mathcal{S} with some nice properties, called the canonical reduction system (CRS) for \mathcal{S} . This is a generalization of the canonical reduction system (CRS) for a reducible subgroup, introduced by Birman, Lubotzky, and McCarthy [8], and Ivanov [30]. It is shown that the CRS exists essentially uniquely for each reducible subgroupoid. This uniqueness will be useful when we study the normalizer of a reducible subgroupoid (see Lemma 6.8).

We shall recall the definition and some fundamental facts of the CRS for a subgroup of the mapping class group. Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. We first define the CRS for a subgroup of $\Gamma(M; m)$.

Definition 5.11. Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. Let Γ be a subgroup of $\Gamma(M; m)$.

- (i) An element $\alpha \in V(C)$ is called an *essential reduction class* for Γ if the following two conditions are satisfied:
 - $g\alpha = \alpha$ for any $g \in \Gamma$;
 - If $\beta \in V(C)$ satisfies $i(\alpha, \beta) \neq 0$, then there exists $g \in \Gamma$ such that $g\beta \neq \beta$.
- (ii) The *canonical reduction system (CRS)* for Γ is defined to be the set of all essential reduction classes for Γ . We denote by $\sigma(\Gamma)$ the CRS for Γ .

It is easy to check that $\sigma(\Gamma) \in S(M) \cup \{\emptyset\}$. It can be shown that if Λ is a finite index subgroup of Γ , then $\sigma(\Lambda) = \sigma(\Gamma)$. Therefore, we can define the CRS for a general subgroup Γ of $\Gamma(M)$ as the CRS for $\Gamma \cap \Gamma(M; m)$, which is independent of m . We refer to Chapter 7 in [30] for more details. Note that if Γ is finite, then $\sigma(\Gamma) = \emptyset$ because $\sigma(\{e\}) = \emptyset$. The following is a fundamental fact on the CRS for an infinite subgroup of $\Gamma(M)$.

Theorem 5.12 ([30, Corollary 7.17]). *An infinite subgroup Γ of $\Gamma(M)$ is reducible if and only if $\sigma(\Gamma)$ is non-empty.*

In the next theorem, we give a geometric meaning of CRS's. We introduce the following notation.

Notation. Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. Let Γ be a subgroup of $\Gamma(M; m)$ and assume that each element of Γ fixes $\sigma \in S(M)$. By Theorem 2.8 (iii), there is a natural homomorphism

$$p_\sigma : \Gamma \rightarrow \prod_Q \Gamma(Q),$$

where Q runs through all components of M_σ , the surface obtained by cutting M along a realization of σ . For each component Q of M_σ , let $p_Q : \Gamma \rightarrow \Gamma(Q)$ be the composition of p_σ and the projection onto $\Gamma(Q)$.

Theorem 5.13 ([30, Theorem 7.16]). *Let M be a surface with $\kappa(M) \geq 0$ and let Γ be a subgroup of $\Gamma(M)$. Then there exists a unique $\sigma \in S(M) \cup \{\emptyset\}$ satisfying the following three conditions, and then σ is in fact the CRS for Γ :*

- (i) *All elements of Γ fix σ .*
- (ii) *Let $m \geq 3$ be an integer and put $\Gamma_0 = \Gamma \cap \Gamma(M; m)$. For each component Q of M_σ , the quotient group $p_Q(\Gamma_0)$ cannot be infinite reducible.*
- (iii) *$\sigma \in S(M)$ is the minimal one satisfying the above Conditions (i), (ii) (for some/any m).*

Example 5.14. We present some examples of reducible subgroups whose CRS can be computed. Let M be a surface with $\kappa(M) \geq 0$.

- (i) Let $\sigma \in S(M)$ and let D_σ be the subgroup of $\Gamma(M)$ generated by all Dehn twists about curves in σ , which is isomorphic to a free abelian group of rank $|\sigma|$. Then $\sigma(D_\sigma) = \sigma$.
- (ii) Let $g \in \Gamma(M)$ be a pure element and take a closed one-dimensional submanifold c (may be empty) of M such that Condition (P) is satisfied for c and some representative of g (see the comment right after Theorem 2.5). If we denote by $\sigma \in S(M) \cup \{\emptyset\}$ the isotopy class of c , then σ is the CRS for the cyclic subgroup of $\Gamma(M)$ generated by g .
- (iii) Take $\sigma \in S(M)$. If we denote by $\Gamma_\sigma = \{g \in \Gamma(M) : g\sigma = \sigma\}$ its stabilizer, then $\sigma(\Gamma_\sigma) = \sigma$.

In the same manner, we can define the canonical reduction system for a reducible subgroupoid as an invariant Borel map into $S(M)$ satisfying some special properties. In the following definition, a purely ρ -invariant pair corresponds to an essential reduction class. We shall recall Assumption (\star): Let M be a surface with $\kappa(M) > 0$ and let

$m \geq 3$ be an integer. Let Γ be a finite index subgroup of $\Gamma(M; m)$. Suppose that Γ admits a measure-preserving action on a standard finite measure space (X, μ) . We denote by \mathcal{G} the associated groupoid $\Gamma \ltimes (X, \mu)$. Let $\rho: \mathcal{G} \rightarrow \Gamma$ be the groupoid homomorphism defined by $(g, x) \mapsto g$.

Definition 5.15. Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Let A be a Borel subset of Y with positive measure and let $\alpha \in V(C)$.

- (i) We say that the pair (α, A) is ρ -invariant for \mathcal{S} if there exists a countable Borel partition $A = \bigsqcup_n A_n$ of A such that for each n , the constant map $A_n \rightarrow \{\alpha\}$ is invariant for \mathcal{S} , i.e., $\rho(\gamma)\alpha = \alpha$ for a.e. $\gamma \in (\mathcal{S})_{A_n}$.
- (ii) Suppose that (α, A) is ρ -invariant for \mathcal{S} . The pair (α, A) is said to be *purely ρ -invariant* for \mathcal{S} if (β, B) is not ρ -invariant for \mathcal{S} for any Borel subset B of A with positive measure and any $\beta \in V(C)$ with $i(\alpha, \beta) \neq 0$. (In [35], we refer to such a pair as an essential ρ -invariant one for \mathcal{S} .)

Remark 5.16. In the notation of Definition 5.15, it is easy to see the following:

- (i) If (α, A) is a ρ -invariant pair for \mathcal{S} , then so is the pair (α, B) for any Borel subset B of A with positive measure. The same statement is true for purely ρ -invariant pairs for \mathcal{S} .
- (ii) For each $n \in \mathbb{N}$, let A_n be a Borel subset of Y with positive measure. If (α, A_n) is a ρ -invariant pair for \mathcal{S} , then so is the pair $(\alpha, \bigcup_{n \in \mathbb{N}} A_n)$. The same statement is true for purely ρ -invariant pairs for \mathcal{S} .

Theorem 5.17 ([35, Theorem 4.50]). *Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. If \mathcal{S} is reducible, then there exists a purely ρ -invariant pair for \mathcal{S} .*

In the notation of Theorem 5.17, for $\alpha \in V(C)$, let \mathcal{M}_α be the set of all Borel subsets A of Y such that either $\mu(A) = 0$ or the pair (α, A) is purely ρ -invariant for \mathcal{S} . Put $m_\alpha = \sup_{A \in \mathcal{M}_\alpha} \mu(A)$. By Remark 5.16 (ii), there exists an essentially unique Borel subset Y_α of Y such that $\mu(Y_\alpha) = m_\alpha$. Theorem 5.17 implies the equation $Y = \bigcup_{\alpha \in V(C)} Y_\alpha$ up to null sets if \mathcal{S} is reducible. By the definition of purely ρ -invariant pairs, if $\alpha, \beta \in V(C)$ satisfy $\mu(Y_\alpha \cap Y_\beta) > 0$, then $i(\alpha, \beta) = 0$. We then define a Borel map $\varphi: Y \rightarrow S(M)$ by the formula

$$\varphi(x) = \{\alpha \in V(C) : x \in Y_\alpha\}$$

for x in a conull Borel subset of Y .

Definition 5.18. The map $\varphi: Y \rightarrow S(M)$ constructed above is called the *canonical reduction system (CRS)* for a reducible subgroupoid \mathcal{S} .

The following theorem states that the invariance and the uniqueness of the CRS for a reducible subgroupoid.

Theorem 5.19 ([35, Lemma 4.53]). *Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Suppose that \mathcal{S} is reducible. Then the CRS $\varphi: Y \rightarrow S(M)$ for \mathcal{S} is an essentially unique invariant Borel map for \mathcal{S} such that*

- if $\sigma \in S(M)$ satisfies $\mu(\varphi^{-1}(\sigma)) > 0$ and if $\alpha \in \sigma$, then $(\alpha, \varphi^{-1}(\sigma))$ is a purely ρ -invariant pair for \mathcal{S} ;
- if (α, A) is a purely ρ -invariant pair for \mathcal{S} , then

$$\mu(A \setminus \varphi^{-1}(\{\sigma \in S(M) : \alpha \in \sigma\})) = 0.$$

6 Normal subgroupoids

Feldman, Sutherland, and Zimmer [14] introduced the notion of normal subrelations of discrete measured equivalence relations. We define the notion of normal subgroupoids as its generalization, which is also a generalization of the notion of normal subgroups. It will be shown that if a subgroup of the mapping class group is IA (resp. infinite and reducible), then so is its normalizer. We prove a similar statement in the setting of groupoids. These facts will be used repeatedly in Section 7.

6.1 Generalities

Let G be a discrete group and let H be a subgroup of G . We refer to the subgroup $N_G(H) = \{g \in G : gHg^{-1} = H\}$ as the normalizer of H in G . If $N_G(H) = G$, then H is called a normal subgroup of G .

Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) and let $r, s: \mathcal{G} \rightarrow X$ be the range, source maps, respectively. Let \mathcal{S} be a subgroupoid of \mathcal{G} . (We mean by a subgroupoid of \mathcal{G} a Borel subgroupoid of \mathcal{G} whose unit space is the same as the one for \mathcal{G} .) We denote by $\text{End}_{\mathcal{G}}(\mathcal{S})$ the set of all Borel maps $\phi: \text{dom}(\phi) \rightarrow \mathcal{G}$ from a Borel subset $\text{dom}(\phi)$ of X such that

- $s(\phi(x)) = x$ for a.e. $x \in \text{dom}(\phi)$;
- for a.e. $\gamma \in (\mathcal{G})_{\text{dom}(\phi)}$, the following equivalence holds: $\gamma \in \mathcal{S}$ if and only if $\phi(r(\gamma))\gamma\phi(s(\gamma))^{-1} \in \mathcal{S}$.

If X consists of a single atom, i.e., if \mathcal{G} and \mathcal{S} are groups, then $\text{End}_{\mathcal{G}}(\mathcal{S})$ is equal to the normalizer $N_{\mathcal{G}}(\mathcal{S})$ of \mathcal{S} in \mathcal{G} .

Remark 6.1. Let $\phi \in \text{End}_{\mathcal{G}}(\mathcal{S})$. Note that the groupoid homomorphism $(\mathcal{S})_{\text{dom}(\phi)} \ni \gamma \mapsto \phi(r(\gamma))\gamma\phi(s(\gamma))^{-1} \in \mathcal{G}$ does not define an isomorphism onto its image when the map $\text{dom}(\phi) \ni x \mapsto r(\phi(x)) \in X$ is not injective. Hence, we use the symbol “End”.

Definition 6.2. Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) . A subgroupoid \mathcal{S} of \mathcal{G} is said to be *normal* in \mathcal{G} if the following condition is satisfied: There exists a countable family $\{\phi_n\}$ of elements of $\text{End}_{\mathcal{G}}(\mathcal{S})$ such that for a.e. $\gamma \in \mathcal{G}$, we can find ϕ_n in the family such that $r(\gamma) \in \text{dom}(\phi_n)$ and $\phi_n(r(\gamma))\gamma \in \mathcal{S}$. In this case, we write $\mathcal{S} \triangleleft \mathcal{G}$, and we call $\{\phi_n\}$ a family of normal choice functions for the pair $(\mathcal{G}, \mathcal{S})$.

Example 6.3. *Normal subgroups.* Let G be a discrete group and let H be a subgroup of G . When we regard G as a groupoid, $\text{End}_G(H) = N_G(H)$ as noted above. It is easy to see that H is normal in G in the sense of Definition 6.2 if and only if we can choose all representatives of G/H from $N_G(H)$, that is, $G = N_G(H)$. This means that H is a normal subgroup of G .

Lemma 6.4. *Suppose that we are given a non-singular action of a discrete group G on a standard measure space (X, μ) . We denote by \mathcal{G} the associated groupoid. Let H be a normal subgroup of G and let \mathcal{S} be the subgroupoid of \mathcal{G} associated with the action $H \curvearrowright (X, \mu)$. Then \mathcal{S} is normal in \mathcal{G} .*

Proof. For $g \in G$, let $\phi_g: X \rightarrow \mathcal{G}$ be the Borel map defined by $\phi_g(x) = (g, x)$. We show that $\phi_g \in \text{End}_{\mathcal{G}}(\mathcal{S})$. Let $(h, x) \in \mathcal{G}$. If $(h, x) \in \mathcal{S}$, then

$$\phi_g(hx)(h, x)\phi_g(x)^{-1} = (g, hx)(h, x)(g, g^{-1}x) = (ghg^{-1}, g^{-1}x) \in \mathcal{S}$$

since H is a normal subgroup of G . Conversely, if $(ghg^{-1}, g^{-1}x) \in \mathcal{S}$, then $ghg^{-1} \in H$, which implies that $h \in H$ and $(h, x) \in \mathcal{S}$. Thus, $\phi_g \in \text{End}_{\mathcal{G}}(\mathcal{S})$.

Since $\phi_{g^{-1}}(gx)(g, x) = e_x \in \mathcal{S}$ for $(g, x) \in \mathcal{G}$, $\{\phi_g\}_{g \in G}$ is a family of normal choice functions for the pair $(\mathcal{G}, \mathcal{S})$. \square

We omit the proof of the following lemma.

Lemma 6.5 ([36, Lemma 2.13]). *Let \mathcal{G} be a discrete measured groupoid on a standard measure space (X, μ) . Let \mathcal{S} be a normal subgroupoid of \mathcal{G} . If A is a Borel subset of X with positive measure, then $(\mathcal{S})_A$ is normal in $(\mathcal{G})_A$.*

6.2 Normalizers of IA and reducible subgroupoids

Let M be a surface with $\kappa(M) \geq 0$ and let N be an infinite subgroup of $\Gamma(M)$. If N is IA, then there exists a pseudo-Anosov element $g \in N$ such that $\{F_{\pm}(g)\}$, the set of pseudo-Anosov foliations of g , is fixed by all elements of N . If N is reducible, then N fixes an element of $S(M)$. The CRS $\sigma(N)$ for N is a special element of $S(M)$ fixed by N (see Subsection 5.3). By using these special fixed elements of N , we show the following

Proposition 6.6. *Let M be a surface with $\kappa(M) \geq 0$ and let Γ be an infinite subgroup of $\Gamma(M)$. Suppose that Γ contains an infinite normal subgroup N .*

- (i) *If N is IA (resp. reducible), then so is Γ .*
- (ii) *If N is reducible, then $\sigma(N) \subset \sigma(\Gamma)$.*

Proof. First assume that N is IA. By Theorem 2.7, there exists a pseudo-Anosov element $g \in N$ such that $h\{F_{\pm}(g)\} = \{F_{\pm}(g)\}$ for any $h \in N$. Let $\gamma \in \Gamma$. Then $\gamma^{-1}g\gamma\{F_{\pm}(g)\} = \{F_{\pm}(g)\}$ since $\gamma^{-1}g\gamma \in N$. Thus, $g\gamma\{F_{\pm}(g)\} = \gamma\{F_{\pm}(g)\}$. On the other hand, the fixed point set on \mathcal{PMF} for g consists of exactly the two points $F_{\pm}(g)$. Hence, $\gamma\{F_{\pm}(g)\} = \{F_{\pm}(g)\}$. This means that every $\gamma \in \Gamma$ fixes $\{F_{\pm}(g)\}$ and that Γ is IA by Theorem 2.7.

Next assume that N is reducible. For $\gamma \in \Gamma$, the equation $\gamma\sigma(N) = \sigma(\gamma N\gamma^{-1}) = \sigma(N)$ holds. The first equation follows by definition. Thus, Γ is reducible. By the definition of essential reduction classes for Γ , we see that $\sigma(N) \subset \sigma(\Gamma)$. \square

In this subsection, we prove a result similar to the above proposition in the framework of groupoids. We recall Assumption (\star) : Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ be an infinite subgroup of $\Gamma(M; m)$. Suppose that Γ admits a measure-preserving action on a standard finite measure space (X, μ) . Let \mathcal{G} be the associated groupoid $\Gamma \times (X, \mu)$. Define a cocycle $\rho: \mathcal{G} \rightarrow \Gamma$ by $(g, x) \mapsto g$.

Using the uniqueness of the invariant Borel maps for IA and reducible subgroupoids constructed in Subsections 5.2 and 5.3, we show that if a subgroupoid \mathcal{T} of \mathcal{G} contains an IA (resp. reducible) subgroupoid as a normal one, then \mathcal{T} is also IA (resp. reducible). We give only the proof of Lemma 6.7, where IA subgroupoids are dealt with. The proof of Lemma 6.8 for reducible ones is not given here. We refer to Lemma 4.60 in [35] for the proof, in which we assume that the action $\Gamma \curvearrowright (X, \mu)$ is essentially free. However, one can show Lemma 6.8 along the same line as in Lemma 4.60 in [35].

Lemma 6.7. *Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Suppose that \mathcal{S} is IA. Let \mathcal{T} be a subgroupoid of $(\mathcal{G})_Y$ with $\mathcal{S} \triangleleft \mathcal{T}$. Let $\varphi_0: Y \rightarrow \partial_2 C$ be the essentially unique Borel map constructed in Lemma 5.8. Then φ_0 is invariant for \mathcal{T} . In particular, \mathcal{T} is IA by Lemma 5.7 (i).*

Proof. Let $r: \mathcal{T} \rightarrow Y$ be the range map. Take $g \in \text{End}_{\mathcal{T}}(\mathcal{S})$. Recall that g is a Borel map from a Borel subset $\text{dom}(g)$ of Y into \mathcal{T} such that

- $s(g(x)) = x$ for a.e. $x \in \text{dom}(g)$;
- for a.e. $\gamma \in (\mathcal{T})_{\text{dom}(g)}$, the following equivalence holds: $\gamma \in \mathcal{S}$ if and only if $g(r(\gamma))\gamma g(s(\gamma))^{-1} \in \mathcal{S}$.

It is enough to show that $\rho(g(x)^{-1})\varphi_0(r(g(x))) = \varphi_0(x)$ for a.e. $x \in \text{dom}(g)$. By applying Theorem 2.9 (iv) to the composition $r \circ g: \text{dom}(g) \rightarrow Y$, we get a countable Borel partition $\text{dom}(g) = \bigsqcup_n Y_n$ satisfying the following: Let g_n denote the restriction

of g to Y_n . The image $r \circ g_n(Y_n)$ is a Borel subset of Y , and the map $r \circ g_n: Y_n \rightarrow r \circ g_n(Y_n)$ is a Borel isomorphism. Moreover, each g_n is an element of $\text{End}_{\mathcal{T}}(\mathfrak{g})$. We may therefore assume that $g \in \text{End}_{\mathcal{T}}(\mathfrak{g})$ satisfies that $r \circ g(\text{dom}(g))$ is a Borel subset of Y and the map $r \circ g: \text{dom}(g) \rightarrow r \circ g(\text{dom}(g))$ is a Borel isomorphism.

We define a Borel map h from $\text{dom}(h) = r \circ g(\text{dom}(g))$ into \mathcal{T} by $h(y) = g((r \circ g)^{-1}(y))^{-1}$ for $y \in \text{dom}(h)$. It is easy to see that $h \in \text{End}_{\mathcal{T}}(\mathfrak{g})$.

We put $\psi_g(x) = \rho(g(x)^{-1})\varphi_0(r(g(x)))$ for $x \in \text{dom}(g)$. Take $\gamma \in (\mathfrak{g})_{\text{dom}(g)}$ and put $x = r(\gamma)$ and $y = s(\gamma)$. Then the equation

$$\begin{aligned} \rho(\gamma)\psi_g(y) &= \rho(\gamma)\rho(g(y)^{-1})\varphi_0(r(g(y))) \\ &= \rho(g(x)^{-1})\rho(g(x))\rho(\gamma)\rho(g(y)^{-1})\varphi_0(r(g(y))) \\ &= \rho(g(x)^{-1})\varphi_0(r(g(x))) = \psi_g(x) \end{aligned}$$

holds since $g(x)\gamma g(y)^{-1} \in \mathfrak{g}$ and φ_0 is invariant for \mathfrak{g} . The map ψ_g is thus invariant for $(\mathfrak{g})_{\text{dom}(g)}$. By Lemma 5.8, we have

$$\text{supp}(\psi_g(x)) \subset \text{supp}(\varphi_0(x)) \quad \text{for a.e. } x \in \text{dom}(g).$$

By considering h instead of g , we have

$$\text{supp}(\rho(h(y)^{-1})\varphi_0(r(h(y)))) \subset \text{supp}(\varphi_0(y)) \quad \text{for a.e. } y \in \text{dom}(h).$$

By putting $y = r \circ g(x)$ in the above two inclusions, we get the equation $\psi_g(x) = \varphi_0(x)$ for a.e. $x \in \text{dom}(g)$. Therefore, φ_0 is invariant for \mathcal{T} . \square

Lemma 6.8. *Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure and let \mathfrak{g} be a subgroupoid of $(\mathfrak{g})_Y$ of infinite type. Suppose that \mathfrak{g} is reducible. Let \mathcal{T} be a subgroupoid of $(\mathfrak{g})_Y$ with $\mathfrak{g} \triangleleft \mathcal{T}$. Let $\varphi_0: Y \rightarrow S(M)$ be the CRS for \mathfrak{g} (see Definition 5.18). Then φ_0 is invariant for \mathcal{T} . In particular, \mathcal{T} is reducible.*

7 Characterization of reducible subgroupoids

In this section, we prove Theorem 3.6. This theorem states that any isomorphism between groupoids associated with measure-preserving actions of mapping class groups preserves subgroupoids generated by Dehn twists. To prove it, we characterize such subgroupoids algebraically in terms of discrete measured groupoids. As in the previous sections, we first investigate the case of groups. We give a complete proof in the case of groups, and give only some comments about the case of groupoids. Most theorems in the case of groupoids can be shown by an idea similar to the one in the case of groups.

Classification of components into three types. We first consider the action of a reducible subgroup on each component of the surface obtained by cutting along the CRS for the subgroup. We recall the following notation.

Notation. Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. Let Γ be a reducible subgroup of $\Gamma(M; m)$ and assume that each element of Γ fixes $\sigma \in S(M)$. By Theorem 2.8 (iii), there is a natural homomorphism

$$p_\sigma: \Gamma \rightarrow \prod_Q \Gamma(Q),$$

where Q runs through all components of M_σ , the surface obtained by cutting M along a realization of σ . For each component Q of M_σ , let $p_Q: \Gamma \rightarrow \Gamma(Q)$ be the composition of p_σ and the projection onto $\Gamma(Q)$.

In the following theorem, we consider the quotient groups $p_Q(\Gamma)$ when σ is the CRS for Γ (see also Theorem 5.13).

Theorem 7.1 ([30, Lemma 1.6, Corollary 7.18]). *Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. Let Γ be an infinite reducible subgroup of $\Gamma(M; m)$ and let $\sigma \in S(M)$ be the CRS for Γ . If Q is a component of M_σ , then the following assertions hold:*

- (i) $p_Q(\Gamma)$ is torsion-free.
- (ii) $p_Q(\Gamma)$ either is trivial or contains a pseudo-Anosov element of $\Gamma(Q)$.

If $p_Q(\Gamma)$ is trivial, infinite amenable or non-amenable, then we say that Q is *T*, *IA* or *IN* for Γ , respectively. Theorem 7.1 implies that any component Q of M_σ is either *T*, *IA* or *IN*, and the following assertions hold:

- (A) Q is *T* for Γ if and only if $p_Q(\Gamma)$ is trivial.
- (B) Q is *IA* for Γ if and only if $p_Q(\Gamma)$ is an *IA* subgroup of $\Gamma(Q)$.
- (C) Q is *IN* for Γ if and only if $p_Q(\Gamma)$ is a sufficiently large subgroup of $\Gamma(Q)$.

Remark 7.2. These three types of Q can be characterized in terms of fixed points for the action of $p_Q(\Gamma)$ on the space $M(\mathcal{PMF}(Q))$ of probability measures on $\mathcal{PMF}(Q)$ as follows:

- (a) Q is *T* for Γ if and only if either Q is a pair of pants (= $M_{0,3}$) or $p_Q(g)\alpha = \alpha$ for any $g \in \Gamma$ and for any/some $\alpha \in V(C(Q))$.
- (b) Q is *IA* for Γ if and only if the following three conditions are satisfied:
 - Q is not a pair of pants.
 - $p_Q(g)\alpha \neq \alpha$ for any $g \in \Gamma \setminus \{e\}$ and for any/some $\alpha \in V(C(Q))$.
 - There exists $\mu \in M(\mathcal{PMF}(Q))$ such that $p_Q(g)\mu = \mu$ for any $g \in \Gamma$ and $\mu(\mathcal{MLN}(Q)) = 1$.
- (c) Q is *IN* for Γ if and only if the following two conditions are satisfied:
 - Q is not a pair of pants.

- There exists no fixed point for the action of $p_Q(\Gamma)$ on $M(\mathcal{PMF}(Q))$.

In the setting of groupoids, motivated by the above characterization, we can define three types of components of the surface obtained by cutting M along the CRS for a reducible subgroupoid \mathcal{S} . They are defined in terms of invariant Borel maps into $M(\mathcal{PMF}(Q))$, etc. for \mathcal{S} . We refer to Theorems 5.6, 5.9 and Section 5.2 in [35] for a precise definition of them.

As an application of Theorem 7.1, we give a criterion for amenability of reducible subgroups.

Proposition 7.3. *Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. Let Γ be a reducible subgroup of $\Gamma(M; m)$ and let $\sigma \in S(M)$ be the CRS for Γ . Then Γ is amenable if and only if each component of M_σ is either T or IA for Γ .*

Theorem 7.1 (ii) implies the “only if” part because the quotient group $p_Q(\Gamma)$ is amenable for each component Q of M_σ if Γ is amenable. The “if” part follows since the intersection of the kernels of p_Q for all components Q of M_σ is amenable by the following proposition (see Lemma 2.1 (1) in [8] or Corollary 4.1.B, Lemma 4.1.C in [32]).

Proposition 7.4. *Let M be a surface with $\kappa(M) \geq 0$. Let G be a reducible subgroup of $\Gamma(M)$ and let $\sigma \in S(M)$ be an element such that $g\sigma = \sigma$ for any $g \in G$. Let $p: G \rightarrow \Gamma(M_c)$ be the natural homomorphism into the mapping class group of the disconnected surface M_c obtained by cutting M along a realization c of σ . Then the following assertions hold:*

- (i) $\ker p$ is contained in the subgroup D_σ of $\Gamma(M)$ generated by Dehn twists about curves in σ .
- (ii) All elements of D_σ belong to the center of $\ker p$.

Characterization of some subgroups. Let M be a surface with $\kappa(M) > 0$. The following is our plan to characterize subgroups of $\Gamma(M)$ generated by Dehn twists.

- (I) Characterize reducible subgroups of $\Gamma(M)$ in terms of amenability and normal subgroups.
- (II) Describe maximal reducible subgroups of $\Gamma(M; m)$ explicitly, where $m \geq 3$ is an integer.
- (III) Describe an infinite amenable normal subgroup N of a maximal reducible subgroup in Step (II). In fact, such a subgroup N is contained in the subgroup generated by the Dehn twist about some element of $V(C)$.

One important observation for Step (I) is the following lemma. This gives a sufficient condition for a subgroup to be reducible.

Lemma 7.5. *Let M be a surface with $\kappa(M) > 0$. Let G be a non-amenable subgroup of $\Gamma(M)$ and let N be an infinite normal subgroup of G . If N is amenable, then G is reducible.*

Proof. It follows from Theorem 2.7 that N is either IA or reducible. By Proposition 6.6 (i), if N is IA (resp. reducible), then so is G . Since G is non-amenable, G must be reducible. \square

Remark 7.6. When $\kappa(M) = 0$, there exists no non-amenable reducible subgroup of $\Gamma(M)$. This fact implies that we cannot characterize reducible subgroups of $\Gamma(M)$ as in Propositions 7.7 and 7.8.

We characterize infinite reducible subgroups in the next two propositions. Although it is not necessary to characterize infinite amenable reducible subgroups for our purpose (because maximal reducible subgroups in Step (II) are always non-amenable), we give it for completeness.

Proposition 7.7. *Let M be a surface with $\kappa(M) > 0$ and let Γ be an infinite amenable subgroup of $\Gamma(M)$. Then the following two assertions are equivalent:*

- (i) Γ is reducible.
- (ii) *There exist four subgroups $\Gamma_0, \Gamma', \Gamma''$ and Λ of Γ satisfying the following:*
 - (a) Γ_0 is a finite index subgroup of Γ ;
 - (b) Γ' is amenable and $\Gamma_0 < \Gamma'$;
 - (c) Γ'' is infinite and $\Gamma'' < \Gamma'$;
 - (d) Λ is non-amenable and $\Gamma'' \triangleleft \Lambda$.

Proof. We first show that Assertion (ii) implies Assertion (i). From Lemma 7.5 it follows that Γ'' and Λ are both reducible. By Theorem 2.7, Γ' must be either IA or reducible since Γ' is amenable. If Γ' were IA, then there would exist a finite index subgroup of Γ' which is cyclic and generated by a pseudo-Anosov element. This contradicts the assumption that Γ' contains the infinite reducible subgroup Γ'' . Thus, Γ' is reducible and so are both Γ_0 and Γ .

We next show that the converse holds. Put $\Gamma_0 = \Gamma \cap \Gamma(M; 3)$ and let $\sigma \in S(M)$ be the CRS for Γ_0 . Note that for each T component Q of M_σ for Γ_0 , the quotient $p_Q(\Gamma_0)$ is trivial by Theorem 7.1 (ii). For each IA component R of M_σ for Γ_0 , let $\{F_\pm^R\}$ be the pair of pseudo-Anosov foliations in $\mathcal{PMF}(R)$ such that

$$p_R(g)\{F_\pm^R\} = \{F_\pm^R\} \quad \text{for any } g \in \Gamma_0.$$

Let Γ' be the subgroup of $\Gamma(M; 3)$ consisting of all $g \in \Gamma(M; 3)$ satisfying the following three conditions:

- $g\sigma = \sigma$;

- $p_Q(g)\alpha = \alpha$ for all $\alpha \in V(C(Q))$ and all T components Q for Γ_0 which is not a pair of pants;
- $p_R(g)\{F_{\pm}^R\} = \{F_{\pm}^R\}$ for all IA components R for Γ_0 .

Then $\Gamma_0 < \Gamma'$. Moreover, the CRS for Γ' is σ and a component of M_σ is T (resp. IA) for Γ' if and only if so is for Γ_0 . In particular, Γ' is amenable since there is no IN component for Γ_0 and thus for Γ' (see Lemma 7.3).

In general, given $\tau \in S(M)$, we denote by D_τ the subgroup of $\Gamma(M)$ generated by Dehn twists about curves in τ . If $|\sigma| < \kappa(M) + 1$, then put $\Gamma'' = D_\sigma \cap \Gamma(M; 3)$. This subgroup is infinite. Let Λ be the stabilizer of σ in $\Gamma(M; 3)$, i.e., $\Lambda = \{g \in \Gamma(M; 3) : g\sigma = \sigma\}$. Since there is a component of M_σ which is not a pair of pants, Λ is non-amenable. (Recall that when we cut M along curves in $\tau \in S(M)$ and get the surface M_τ , all components of M_τ are pairs of pants if and only if $|\tau| = \kappa(M) + 1$.) Moreover, Γ'' is a normal subgroup of Λ by Proposition 7.4 (ii). These Γ'' , Λ satisfy the conditions of Assertion (ii).

If $|\sigma| = \kappa(M) + 1$, then we see that $\Gamma' = D_\sigma \cap \Gamma(M; 3)$. Choose $\alpha_0 \in \sigma$. Let $\sigma' = \sigma \setminus \{\alpha_0\}$ and put $\Gamma'' = D_{\sigma'} \cap \Gamma(M; 3)$. This subgroup is infinite and satisfies $\Gamma'' < \Gamma'$. If we define Λ to be the stabilizer of σ' in $\Gamma(M; 3)$, then these subgroups satisfy the conditions of Assertion (ii). □

Proposition 7.8. *Let M be a surface with $\kappa(M) > 0$ and let Γ be a non-amenable subgroup of $\Gamma(M)$. Then the following two assertions are equivalent:*

- (i) Γ is reducible.
- (ii) *There exist two subgroups Γ', Γ'' of $\Gamma(M)$ satisfying the following:*
 - (a) $\Gamma < \Gamma'$;
 - (b) Γ'' is infinite amenable and $\Gamma'' \triangleleft \Gamma'$.

Proof. We first show that Assertion (ii) implies Assertion (i). From Lemma 7.5 it follows that Γ' and Γ'' are both reducible. Thus, so is Γ .

We next show that the converse holds. Let $\sigma \in S(M)$ be the CRS for Γ . Let Γ' be the stabilizer of σ , i.e., $\Gamma' = \{g \in \Gamma(M) : g\sigma = \sigma\}$. Then Γ' contains Γ . Let Γ'' be the subgroup of $\Gamma(M)$ generated by Dehn twists about curves in σ . By Proposition 7.4 (ii), we see that $\Gamma'' \triangleleft \Gamma'$. □

Corollary 7.9. *Let M be a surface with $\kappa(M) > 0$. Let Γ_1, Γ_2 be finite index subgroups of $\Gamma(M)$. If $f : \Gamma_1 \rightarrow \Gamma_2$ is an isomorphism and Λ is an infinite reducible subgroup of Γ_1 , then $f(\Lambda)$ is an infinite reducible subgroup of Γ_2 .*

Notation. Let M be a surface with $\kappa(M) \geq 0$. Given a subgroup Γ of $\Gamma(M)$ and $\sigma \in S(M)$, we denote by

$$\Gamma_\sigma = \{g \in \Gamma : g\sigma = \sigma\}$$

the stabilizer of σ in Γ . When σ consists of only one element $\alpha \in V(C)$, we denote Γ_σ by Γ_α for simplicity.

In the next lemma, we explicitly describe maximal reducible subgroups.

Lemma 7.10. *Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. Let Γ be a finite index subgroup of $\Gamma(M; m)$ and let $\alpha \in V(C)$. Then the following assertions hold:*

- (i) Γ_α is a maximal reducible subgroup of Γ , that is, if Λ is a reducible subgroup of Γ with $\Gamma_\alpha < \Lambda$, then $\Lambda = \Gamma_\alpha$.
- (ii) Conversely, any reducible subgroup of Γ is contained in Γ_α for some $\alpha \in V(C)$.

Proof. Assertion (ii) follows from Theorem 2.8 (iii). We prove Assertion (i). One can show that α is the only class in $V(C)$ fixed by all elements of Γ_α . In fact, if $\beta \in V(C)$ satisfies $i(\alpha, \beta) \neq 0$, then some power of the Dehn twist about α is in Γ_α and does not fix β (see Theorem 2.6). Suppose that $\beta \in V(C)$ satisfies $\alpha \neq \beta$ and $i(\alpha, \beta) = 0$. Let M_α be the surface obtained by cutting M along a realization of α . Let Q be a component of M_α such that $\beta \in V(C(Q))$. Since Γ is a finite index subgroup of $\Gamma(M; m)$, the component Q is IN for Γ_α . Hence, $p_Q(\Gamma_\alpha)$ does not fix β . This proves the claim. Assertion (i) then follows because Λ fixes some curve in $V(C)$, which has to be α by this claim. \square

Finally, we give an algebraic characterization of subgroups generated by Dehn twists.

Lemma 7.11. *Let M be a surface with $\kappa(M) \geq 0$ and let $m \geq 3$ be an integer. Let Γ be a subgroup of finite index in $\Gamma(M; m)$ and let $\alpha \in V(C)$. We denote by D_α the intersection of Γ with the subgroup of $\Gamma(M)$ generated by the Dehn twist about α .*

- (i) Let N be an infinite amenable subgroup of Γ_α with $N \triangleleft \Gamma_\alpha$. Then N is contained in D_α .
- (ii) Conversely, any subgroup of D_α is amenable and is a normal one of Γ_α .

Proof. Assertion (ii) follows from Proposition 7.4 (ii). We show Assertion (i). When $\kappa(M) = 0$, Proposition 7.4 (i) implies that $\Gamma_\alpha = D_\alpha$, and Assertion (i) follows. We assume that $\kappa(M) > 0$. Let $\sigma \in S(M)$ be the CRS for N . Note that the CRS for Γ_α is $\{\alpha\}$ (see Example 5.14 (iii)). By Proposition 6.6 (ii), we see that $\sigma \subset \{\alpha\}$, which means that $\sigma = \{\alpha\}$. By Proposition 7.4 (i), it is enough to show that each component of M_α is T for N , which follows from the next Lemma 7.12. \square

Lemma 7.12. *Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ be a subgroup of $\Gamma(M; m)$ and let N be an infinite normal subgroup of Γ . Suppose that N is reducible and let $\sigma \in S(M)$ be the CRS for N . (Note that $\sigma \subset \sigma(\Gamma)$ by Proposition 6.6 (ii).) If a component Q of M_σ is IA for N , then Q is a component of $M_{\sigma(\Gamma)}$ and it is IA for Γ .*

Proof. Recall that σ is fixed by Γ by Proposition 6.6 (ii). If Q were not a component of $M_{\sigma(\Gamma)}$, then there would exist $\alpha \in \sigma(\Gamma) \setminus \sigma$ such that $\alpha \in V(C(Q))$ since $\sigma \subset \sigma(\Gamma)$. Then α is fixed by all elements of N . This contradicts the assumption that Q is IA for N . Thus, Q is a component of $M_{\sigma(\Gamma)}$. Since $p_Q(N)$ is IA and is a normal subgroup of $p_Q(\Gamma)$, we see that $p_Q(\Gamma)$ is also IA by Proposition 6.6 (i). \square

Corollary 7.13. *Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ_1, Γ_2 be finite index subgroups of $\Gamma(M; m)$ and let $f: \Gamma_1 \rightarrow \Gamma_2$ be an isomorphism. For $i \in \{1, 2\}$ and $\alpha \in V(C)$, let D_α^i be the intersection of Γ_i with the cyclic subgroup of $\Gamma(M)$ generated by the Dehn twist about α . Then for each $\alpha \in V(C)$, there exists $\beta \in V(C)$ such that $f(D_\alpha^1) = D_\beta^2$.*

Proof. For $i \in \{1, 2\}$ and $\alpha \in V(C)$, we denote by Γ_α^i the stabilizer of α in Γ_i . Let $\alpha \in V(C)$. By Lemma 7.10 (i), Γ_α^1 is a maximal reducible subgroup of Γ_1 . It follows from Corollary 7.9 that $f(\Gamma_\alpha^1)$ is also a maximal reducible subgroup of Γ_2 . Thus, there exists $\beta \in V(C)$ such that $f(\Gamma_\alpha^1) = \Gamma_\beta^2$ by Lemma 7.10 (ii). Since D_α^1 is a normal subgroup of Γ_α^1 , we see that $f(D_\alpha^1)$ is also a normal subgroup of Γ_β^2 . By Lemma 7.11 (i), $f(D_\alpha^1) < D_\beta^2$. Considering f^{-1} , we see that there exists $\alpha' \in V(C)$ such that $D_\beta^2 < f(D_{\alpha'}^1)$. Since D_α^1 and $D_{\alpha'}^1$ has non-trivial intersection, we obtain the equality $\alpha = \alpha'$ by Lemma 3.4. \square

By using this corollary in place of Theorem 3.3 in Ivanov’s argument in Subsection 3.1, we can show Theorem 3.2, which states that any isomorphism between finite index subgroups of the extended mapping class group $\Gamma(M)^\diamond$ with $\kappa(M) > 0$ and $M \neq M_{1,2}, M_{2,0}$ is equal to the inner conjugation by a unique element of $\Gamma(M)^\diamond$.

The case of groupoids. We first restate Theorem 3.6. Recall the following notation.

Notation. We refer to the following assumption as (\bullet) :

- Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ_1 and Γ_2 be finite index subgroups of $\Gamma(M; m)$. Let (Ω, ω) be an ME coupling of Γ_1 and Γ_2 .
- Take fundamental domains $X_1 \subset \Omega$ for the Γ_2 -action on Ω , and $X_2 \subset \Omega$ for the Γ_1 -action on Ω . Recall that the natural actions $\Gamma_1 \curvearrowright X_1$ and $\Gamma_2 \curvearrowright X_2$ are denoted by $(\gamma, x) \mapsto \gamma \cdot x$ by using a dot. By Lemma 2.27, we can choose X_1, X_2 so that $Y = X_1 \cap X_2$ satisfies that for $i \in \{1, 2\}$, $\Gamma_i \cdot Y = X_i$ up to null sets when Y is regarded as a subset of X_i .
- For $i \in \{1, 2\}$, set $\mathcal{G}^i = \Gamma_i \ltimes X_i$ and let $\rho_i: \mathcal{G}^i \rightarrow \Gamma_i$ be the projection, which is a groupoid homomorphism. By Proposition 2.29, there exists a groupoid isomorphism

$$f: (\mathcal{G}^1)_Y \rightarrow (\mathcal{G}^2)_Y.$$

Note that f is the identity on the unit space Y .

- For $i \in \{1, 2\}$ and $\alpha \in V(C)$, let D_α^i be the intersection of Γ_i with the subgroup of $\Gamma(M)$ generated by the Dehn twist $t_\alpha \in \Gamma(M)$ about α . Let \mathcal{G}_α^i be the subgroupoid of \mathcal{G}^i generated by the action of D_α^i , i.e.,

$$\mathcal{G}_\alpha^i = \{(\gamma, x) \in \mathcal{G}^i : \gamma \in D_\alpha^i, x \in X_i\}.$$

Theorem 3.6. *Under Assumption (\bullet) , for each $\alpha \in V(C)$, there exist a countable Borel partition $Y = \bigsqcup Y_n$ and $\beta_n \in V(C)$ such that*

$$f((\mathcal{G}_\alpha^1)_{Y_n}) = (\mathcal{G}_{\beta_n}^2)_{f(Y_n)} \quad \text{for each } n.$$

This theorem states that f preserves subgroupoids generated by Dehn twists up to a countable Borel partition. When each of X_1 and X_2 consists of a single atom, this theorem reduces to Corollary 7.13. To prove Theorem 3.6, we characterize subgroupoids generated by Dehn twists algebraically in terms of discrete measured groupoids. Our plan is the following:

- (1) We characterize reducible subgroupoids algebraically. It follows that the isomorphism f in Assumption (\bullet) preserves reducible subgroupoids.
- (2) By Step (1), f preserves maximal reducible subgroupoids. We explicitly describe such subgroupoids.
- (3) A subgroupoid generated by a Dehn twist can be characterized algebraically as an amenable normal subgroupoid of infinite type of some maximal reducible subgroupoid. This implies Theorem 3.6.

Note that the above steps correspond to the ones for the proof of Corollary 7.13 given right before Lemma 7.5. In this final part of Section 7, we give only precise statements and some comments for the above steps. Most statements can be proved along the same line as in the case of groups. We refer to Section 4 in [36] for the proof of them.

About Step (1). We shall recall Assumption (\star) : Let M be a surface with $\kappa(M) > 0$ and let $m \geq 3$ be an integer. Let Γ be a finite index subgroup of $\Gamma(M; m)$. Suppose that Γ admits a measure-preserving action on a standard finite measure space (X, μ) . Let \mathcal{G} be the associated groupoid $\Gamma \times (X, \mu)$. Define a groupoid homomorphism $\rho: \mathcal{G} \rightarrow \Gamma$ by $(g, x) \mapsto g$.

Proposition 7.14 ([36, Proposition 4.1]). *Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Suppose that \mathcal{S} is amenable. Then the following two assertions are equivalent:*

- \mathcal{S} is reducible.
- For any Borel subset A of Y with positive measure, there exist a Borel subset B of A with positive measure and the following three subgroupoids \mathcal{S}' , \mathcal{S}'' and \mathcal{T} of $(\mathcal{G})_B$:
 - \mathcal{S}' is amenable and $(\mathcal{S})_B < \mathcal{S}'$;

- (b) \mathcal{S}'' is of infinite type and $\mathcal{S}'' < \mathcal{S}'$;
- (c) \mathcal{T} is non-amenable and $\mathcal{S}'' \triangleleft \mathcal{T}$.

Proposition 7.15 ([36, Proposition 4.2]). *Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure and let \mathcal{S} be a subgroupoid of $(\mathcal{G})_Y$ of infinite type. Suppose that $(\mathcal{S})_{Y'}$ is not amenable for any Borel subset Y' of Y with positive measure. Then the following two assertions are equivalent:*

- (i) \mathcal{S} is reducible.
- (ii) *For any Borel subset A of Y with positive measure, there exist a Borel subset B of A with positive measure and the following two subgroupoids \mathcal{S}' and \mathcal{S}'' of $(\mathcal{G})_B$:*
 - (a) $(\mathcal{S})_B < \mathcal{S}'$;
 - (b) \mathcal{S}'' is an amenable subgroupoid of infinite type and $\mathcal{S}'' \triangleleft \mathcal{S}'$.

Along the same line as in the proof of Propositions 7.7 and 7.8, these propositions are proved by using invariant Borel maps developed in Section 5. Thanks to these algebraic characterizations, we obtain the following corollary.

Corollary 7.16. *Under Assumption (\bullet) , let A be a Borel subset of Y with positive measure and let \mathcal{S}^1 be a subgroupoid of $(\mathcal{G}^1)_A$ of infinite type. Then \mathcal{S}^1 is reducible if and only if the image $f(\mathcal{S}^1)$ is reducible.*

About Step (2). Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure. For a Borel map $\varphi: Y \rightarrow V(C)$, we write

$$\mathcal{S}_\varphi = \{\gamma \in (\mathcal{G})_Y : \rho(\gamma)\varphi(s(\gamma)) = \varphi(r(\gamma))\}.$$

This subgroupoid can be viewed as the stabilizer of φ in $(\mathcal{G})_Y$. As in the case of groups, we can show that \mathcal{S}_φ is a maximal reducible subgroupoid in $(\mathcal{G})_Y$, and conversely that any reducible subgroupoid is contained in \mathcal{S}_φ for some φ . Thus, we obtain the following

Corollary 7.17. *Under Assumption (\bullet) , let A_1 be a Borel subset of Y_1 with positive measure and let $\varphi_1: A_1 \rightarrow V(C)$ be a Borel map. Put $A_2 = f(A_1)$. Then there exists a Borel map $\varphi_2: A_2 \rightarrow V(C)$ such that $f(\mathcal{S}_{\varphi_1}^1) = \mathcal{S}_{\varphi_2}^2$, where $\mathcal{S}_{\varphi_i}^i = \{\gamma \in (\mathcal{G}^i)_{A_i} : \rho_i(\gamma)\varphi_i(s(\gamma)) = \varphi_i(r(\gamma))\}$ for $i \in \{1, 2\}$.*

About Step (3). Under Assumption (\star) , let $Y \subset X$ be a Borel subset with positive measure and let $\varphi: Y \rightarrow V(C)$ be a Borel map. As in Lemma 7.11, if \mathcal{S} is an amenable subgroupoid of \mathcal{S}_φ of infinite type with $\mathcal{S} \triangleleft \mathcal{S}_\varphi$, then we can show that there exists a countable Borel partition $Y = \bigsqcup Y_n$ of Y satisfying the following two conditions:

- (i) The map φ is constant a.e. on Y_n . Let $\alpha_n \in V(C)$ be its value on Y_n .
- (ii) $(\mathcal{S})_{Y_n} < (\mathcal{G}_{\alpha_n})_{Y_n} < (\mathcal{S}_\varphi)_{Y_n}$ for each n .

Here, for $\alpha \in V(C)$, we denote by \mathcal{G}_α the subgroupoid of \mathcal{G} generated by the intersection D_α of Γ with the cyclic subgroup of $\Gamma(M)$ generated by the Dehn twist about α .

In what follows, we prove Theorem 3.6. Under Assumption (\bullet) , let $\alpha \in V(C)$ and let $\varphi_1: Y \rightarrow V(C)$ be the constant map with value α . Since $\mathcal{S}_{\varphi_1}^1$ is the subgroupoid generated by the action of the stabilizer of α in Γ_1 , we see that $(\mathcal{G}_\alpha^1)_Y \triangleleft (\mathcal{S}_{\varphi_1}^1)_Y$ by Lemma 6.5. By Corollary 7.17 and the above fact, there exist a Borel map $\varphi_2: Y \rightarrow V(C)$ and a countable Borel partition $Y = \bigsqcup Y_n$ of Y satisfying the following two conditions:

- (i) The map φ_2 is constant a.e. on Y_n . Let $\beta_n \in V(C)$ be its value on Y_n .
- (ii) $f((\mathcal{G}_\alpha^1)_{Y_n}) < (\mathcal{G}_{\beta_n}^2)_{f(Y_n)} < (\mathcal{S}_{\varphi_2}^2)_{f(Y_n)}$ for each n .

By considering f^{-1} , we can show that $f((\mathcal{G}_\alpha^1)_{Y_n}) = (\mathcal{G}_{\beta_n}^2)_{f(Y_n)}$ for each n . This proves Theorem 3.6.

8 Concluding remarks

We present some comments about other related results shown in the series of papers [35], [36], [37], [38].

1. Classification of mapping class groups up to ME. For $i \in \{1, 2\}$, let M_i be a surface of type (g_i, p_i) , that is, of genus g_i and with p_i boundary components. When are the mapping class groups $\Gamma(M_1)$ and $\Gamma(M_2)$ ME? Note that for a surface M of type (g, p) , the mapping class group $\Gamma(M)$ is finite if and only if $\kappa(M) = 3g + p - 4 < 0$ and $(g, p) \neq (1, 0)$. We may exclude these cases. If $(g, p) = (1, 0), (1, 1)$, then $\Gamma(M)$ is isomorphic to $\mathrm{SL}(2, \mathbb{Z})$. If $(g, p) = (0, 4)$, then there exists a finite index subgroup of $\Gamma(M)$ isomorphic to $\mathrm{PSL}(2, \mathbb{Z})$ (see Section 7 in [29]). In particular, if $(g, p) = (1, 0), (0, 4), (1, 1)$, then $\Gamma(M)$ is ME to $\mathrm{SL}(2, \mathbb{Z})$ and is hyperbolic in the sense of Gromov. It follows from Theorem 1.5 that if $\kappa(M) > 0$ and if a discrete group Λ is ME to $\Gamma(M)$, then they are commensurable up to finite kernels. Thus, if $\kappa(M) > 0$, then $\Gamma(M)$ and $\mathrm{SL}(2, \mathbb{Z})$ are not ME since $\Gamma(M)$ is not hyperbolic. Hence, if we classify the case $\kappa(M) > 0$, then we obtain a complete classification. Thanks to Theorem 1.5, this remaining problem is reduced to a simple algebraic problem of mapping class groups. By Theorem 2 in [57], we obtain the following

Theorem 8.1 ([36, Theorem 1.2]). *Let M^1 and M^2 be distinct surfaces of type $(g_1, p_1), (g_2, p_2)$, respectively, such that $\kappa(M^1), \kappa(M^2) > 0$ and $g_1 \leq g_2$. Suppose that $\Gamma(M^1)$ and $\Gamma(M^2)$ are ME. Then we have the following only two possibilities: $((g_1, p_1), (g_2, p_2)) = ((0, 5), (1, 2)), ((0, 6), (2, 0))$.*

In Chapters 5 and 6 in [35], we obtain a weaker classification result by a complete different approach, using tools developed in Sections 5 and 6.

Gaboriau [20] proved that the sequence $\{\beta_n(\Gamma)\}_{n \in \mathbb{N}}$ of ℓ^2 -Betti numbers for a discrete group Γ is an invariant for ME in the following sense: If two discrete groups Γ and Λ are ME, then there exists a positive real number c such that $\beta_n(\Gamma) = c\beta_n(\Lambda)$ for all n . Combining this with results due to Gromov [22] and McMullen [46], we can calculate the ℓ^2 -Betti numbers of mapping class groups as follows: If M is a surface with $\kappa(M) \geq 0$, then

$$\beta_{\kappa(M)+1}(\Gamma(M)) > 0 \quad \text{and} \quad \beta_n(\Gamma(M)) = 0 \quad \text{for } n \neq \kappa(M) + 1.$$

Therefore, the value $\kappa(M)$ is invariant under ME. The reader is referred to Appendix D in [35] for more details, in which explicit values of $\beta_{\kappa(M)+1}(\Gamma(M))$ are also discussed.

2. Exactness of mapping class groups. We defined amenability of a group action in a measurable sense in Section 4. We can also define amenability of a group action in a topological sense. A discrete group is said to be *exact* if it admits an amenable action on some compact Hausdorff space in a topological sense (see [4], [54] for the definition). It is widely expected that the class of exact groups is huge. Indeed, all amenable groups, hyperbolic ones and linear ones are exact. Exactness is closed under taking subgroups, extensions, direct unions and amalgamated free products. Moreover, exactness has many equivalent conditions in terms of geometry of Cayley graphs and operator algebras, and has many applications to various research fields, the study of the Baum–Connes conjecture and the classification of group von Neumann algebras. We recommend the reader to consult [54], [61] and the references therein for more details.

As a byproduct of Theorem 4.21, we can show that if M is a surface with $\kappa(M) > 0$, then the action of $\Gamma(M)^\diamond$ on its Stone–Čech compactification is amenable in a topological sense (see Theorem C.5 in [35]). Hence, $\Gamma(M)^\diamond$ and all its subgroups are exact. Note that the action of $\Gamma(M)^\diamond$ on \mathcal{PMF} is not amenable because there exist non-amenable stabilizers and note that ∂C is not compact (see Proposition 3.8 in [35]). Hamenstädt [26] also proved that $\Gamma(M)^\diamond$ is exact by constructing an explicit compact space on which $\Gamma(M)^\diamond$ admits an amenable action in a topological sense.

3. Direct products of mapping class groups. We can also prove an ME rigidity result for finite direct products of mapping class groups.

Theorem 8.2 ([36, Theorem 1.3]). *Let n be a positive integer and let M_i be a surface with $\kappa(M_i) > 0$ for all $i \in \{1, \dots, n\}$. If a discrete group Λ is ME to the direct product $\Gamma(M_1) \times \cdots \times \Gamma(M_n)$, then there exists a homomorphism $\rho: \Lambda \rightarrow G = \text{Aut}(C(M_1)) \times \cdots \times \text{Aut}(C(M_n))$ such that the kernel of ρ and the index $[G : \rho(\Lambda)]$ are both finite.*

Let (Σ, m) be an ME coupling of $\Gamma = \Gamma(M_1) \times \cdots \times \Gamma(M_n)$ and an unknown group Λ . For the proof of Theorem 8.2, we first consider a self ME coupling $\Omega = \Sigma \times_\Lambda \Lambda \times_\Lambda \check{\Sigma}$ of Γ as in the proof of Theorem 1.5. We then construct an almost

$(\Gamma \times \Gamma)$ -equivariant Borel map

$$\Omega \rightarrow G = \text{Aut}(C(M_1)) \times \cdots \times \text{Aut}(C(M_n)) \quad (8.1)$$

for some $(\Gamma \times \Gamma)$ -action on G for which G is a self ME coupling of Γ (see Theorem 7.1 and Corollary 7.2 in [36] for a more explicit statement).

Monod and Shalom [48] introduced the class \mathcal{C} consisting of discrete groups Δ which admit a mixing unitary representation π on a Hilbert space such that the second bounded cohomology $H_b^2(\Delta, \pi)$ of Δ with coefficient π does not vanish. They studied self ME couplings of discrete groups of the form $\Delta_1 \times \cdots \times \Delta_n$ with $\Delta_i \in \mathcal{C}$ and $n \geq 2$ via the theory of bounded cohomology. They obtained many interesting measurable rigidity results on ergodic standard actions of such product groups. Whether a discrete group is in the class \mathcal{C} or not is invariant under ME, and \mathcal{C} contains all non-elementary hyperbolic groups in the sense of Gromov. Hamenstädt [25] proved that the mapping class group $\Gamma(M)$ with $\kappa(M) > 0$ is in \mathcal{C} . We apply these results to our situation, and construct the map in (8.1).

We note that a theorem similar to Theorem 1.6 can also be shown for direct products of mapping class groups (see Theorem 1.4 in [36]).

4. Construction of non-OE actions. Let M be a surface with $\kappa(M) > 0$ and put $\Gamma = \Gamma(M)^\circ$. In Corollary 3.12, we proved that if two ergodic standard (i.e., measure-preserving and essentially free) actions of Γ are OE, then they are conjugate. In the theory of OE, it is an interesting problem to construct (continuously) many ergodic standard actions of one specified group which are mutually non-OE. Thanks to Corollary 3.12, if we construct non-conjugate actions, then they are non-OE. In [38], we give a family of non-OE actions of Γ as shown in the following: Let $\alpha \in V(C)$ and consider its Γ -orbit $K = \Gamma\alpha$, on which Γ naturally acts. Let (X_0, μ_0) be a standard probability space, i.e., a standard Borel space with a probability measure. We assume that (X_0, μ_0) may contain atoms, whereas (X_0, μ_0) is non-trivial, i.e., it does not consist of a single atom. The generalized Bernoulli action of Γ on $(X_0, \mu_0)^K = \prod_K (X_0, \mu_0)$ is defined by

$$g(x_\beta)_{\beta \in K} = (x_{g^{-1}\beta})_{\beta \in K}, \quad (x_\beta)_{\beta \in K} \in X_0^K, \quad g \in \Gamma.$$

This action is ergodic and standard. We can show that for two non-trivial standard probability spaces (X_0, μ_0) and (Y_0, ν_0) , the two generalized Bernoulli actions of Γ on $(X_0, \mu_0)^K$ and $(Y_0, \nu_0)^K$ are conjugate if and only if (X_0, μ_0) and (Y_0, ν_0) are isomorphic, i.e., there exists a Borel isomorphism $f: X'_0 \rightarrow Y'_0$ between conull Borel subsets $X'_0 \subset X_0$ and $Y'_0 \subset Y_0$ such that $f_*\mu_0 = \nu_0$. Hence, this example gives a family of continuously many ergodic standard actions of Γ which are mutually non-OE.

References

- [1] S. Adams, Boundary amenability for word hyperbolic groups and an application to smooth dynamics of simple groups. *Topology* 33 (1994), 765–783. [339](#)
- [2] S. Adams, Indecomposability of equivalence relations generated by word hyperbolic groups. *Topology* 33 (1994), 785–798. [315](#), [316](#), [345](#)
- [3] S. Adams, G. A. Elliott, and T. Giordano, Amenable actions of groups. *Trans. Amer. Math. Soc.* 344 (1994), 803–822. [336](#), [338](#)
- [4] C. Anantharaman-Delaroche and J. Renault, *Amenable groupoids*. Monogr. Enseign. Math. 36, Enseignement Math., Geneva 2000. [308](#), [331](#), [333](#), [334](#), [335](#), [336](#), [337](#), [338](#), [339](#), [362](#)
- [5] C. Anantharaman-Delaroche and J. Renault, Amenable groupoids. In *Groupoids in analysis, geometry, and physics* (Boulder, CO, 1999), Contemp. Math. 282, Amer. Math. Soc., Providence, RI, 2001, 35–46. [308](#), [313](#), [331](#), [333](#), [339](#)
- [6] B. Bekka, P. de la Harpe, and A. Valette, *Kazhdan’s property (T)*. New Math. Monogr. 11, Cambridge University Press, Cambridge 2008. [299](#)
- [7] M. Bestvina and K. Fujiwara, Bounded cohomology of subgroups of mapping class groups. *Geom. Topol.* 6 (2002), 69–89. [344](#)
- [8] J. S. Birman, A. Lubotzky, and J. McCarthy, Abelian and solvable subgroups of the mapping class groups. *Duke Math. J.* 50 (1983), 1107–1120. [341](#), [346](#), [354](#)
- [9] B. H. Bowditch, Tight geodesics in the curve complexes. *Invent. Math.* 171 (2008), 281–300. [339](#)
- [10] A. Connes, J. Feldman, and B. Weiss. An amenable equivalence relation is generated by a single transformation. *Ergodic Theory Dynam. Systems* 1 (1981), 431–450. [330](#), [337](#), [338](#)
- [11] H. A. Dye, On groups of measure preserving transformation. I. *Amer. J. Math.* 81 (1959), 119–159. [301](#)
- [12] H. A. Dye, On groups of measure preserving transformations. II. *Amer. J. Math.* 85 (1963), 551–576. [301](#)
- [13] A. Fathi, F. Laudenbach and V. Poénaru (eds.), *Travaux de Thurston sur les surfaces*. *Astérisque* 66–67 (1979). [303](#), [304](#), [305](#), [306](#)
- [14] J. Feldman, C. E. Sutherland, and R. J. Zimmer, Subrelations of ergodic equivalence relations. *Ergodic Theory Dynam. Systems* 9 (1989), 239–269. [349](#)
- [15] J. M. G. Fell and R. S. Doran, *Representations of $*$ -algebras, locally compact groups, and Banach $*$ -algebraic bundles. Vol. 1. Basic representation theory of groups and algebras*. Pure and Appl. Math. 125, Academic Press, Inc., Boston, MA, 1988. [333](#), [334](#)
- [16] A. Furman, Gromov’s measure equivalence and rigidity of higher rank lattices. *Ann. of Math.* (2) 150 (1999), 1059–1081. [298](#), [301](#), [320](#), [327](#), [328](#)
- [17] A. Furman, Orbit equivalence rigidity. *Ann. of Math.* (2) 150 (1999), 1083–1108. [299](#), [316](#), [319](#), [330](#)
- [18] A. Furman, Mostow-Margulis rigidity with locally compact targets. *Geom. Funct. Anal.* 11 (2001), 30–59. [300](#)
- [19] A. Furman, On Popa’s cocycle superrigidity theorem. *Internat. Math. Res. Notices* 2007 (19) (2007). [330](#)

- [20] D. Gaboriau, Invariants ℓ^2 de relations d'équivalence et de groupes. *Publ. Math. Inst. Hautes Études Sci.* 95 (2002), 93–150. [302](#), [362](#)
- [21] D. Gaboriau, Examples of groups that are ME to the free group. *Ergodic Theory Dynam. Systems* 25 (2005), 1809–1827. [302](#)
- [22] M. Gromov, Kähler hyperbolicity and L_2 -Hodge theory. *J. Differential Geom.* 33 (1991), 263–292. [362](#)
- [23] M. Gromov, M. Gromov, *Asymptotic invariants of infinite groups*. London Math. Soc. Lecture Note Ser. 182, Cambridge University Press, Cambridge 1993. [298](#), [299](#)
- [24] U. Hamenstädt, Train tracks and the Gromov boundary of the complex of curves. In *Spaces of Kleinian groups*, London Math. Soc. Lecture Note Ser. 329, Cambridge University Press, Cambridge 2006, 187–207. [301](#), [339](#)
- [25] U. Hamenstädt, Bounded cohomology and isometry groups of hyperbolic spaces. *J. Eur. Math. Soc.* 10 (2008), 315–349. [363](#)
- [26] U. Hamenstädt, Geometry of the mapping class groups I: Boundary amenability. *Invent. Math.* 175 (2009), 545–609. [362](#)
- [27] U. Hamenstädt, Geometry of the complex of curves and of Teichmüller space. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 447–467. [301](#), [339](#)
- [28] W. J. Harvey, Boundary structure of the modular group. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference* (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 245–251. [303](#)
- [29] N. V. Ivanov, Automorphisms of Teichmüller modular groups. In *Topology and geometry—Rohlin Seminar*, Lecture Notes in Math. 1346, Springer-Verlag, Berlin 1988, 199–270. [361](#)
- [30] N. V. Ivanov, *Subgroups of Teichmüller modular groups*. Transl. Math. Monogr. 115, Amer. Math. Soc., Providence, RI, 1992. [303](#), [307](#), [308](#), [341](#), [346](#), [347](#), [353](#)
- [31] N. V. Ivanov, Automorphism of complexes of curves and of Teichmüller spaces. *Internat. Math. Res. Notices* 1997 (14) (1997), 651–666. [304](#)
- [32] N. V. Ivanov, Mapping class groups. In *Handbook of geometric topology*, North-Holland, Amsterdam 2002, 523–633. [303](#), [306](#), [323](#), [324](#), [354](#)
- [33] V. A. Kaimanovich and H. Masur, The Poisson boundary of the mapping class group. *Invent. Math.* 125 (1996), 221–264. [300](#)
- [34] A. S. Kechris, *Classical descriptive set theory*. Grad. Texts in Math. 156, Springer-Verlag, New York 1995. [298](#), [309](#)
- [35] Y. Kida, The mapping class group from the viewpoint of measure equivalence theory. *Mem. Amer. Math. Soc.* 196 (2008). [301](#), [303](#), [339](#), [340](#), [343](#), [344](#), [348](#), [349](#), [351](#), [354](#), [361](#), [362](#)
- [36] Y. Kida, Measure equivalence rigidity of the mapping class group. *Ann. of Math.*, to appear; arXiv:math.GR/0607600. [300](#), [303](#), [321](#), [322](#), [324](#), [326](#), [328](#), [329](#), [350](#), [359](#), [360](#), [361](#), [362](#), [363](#)
- [37] Y. Kida, Orbit equivalence rigidity for ergodic actions of the mapping class group. *Geom. Dedicata* 131 (2008), 99–109. [303](#), [330](#), [361](#)

- [38] Y. Kida, Classification of certain generalized Bernoulli actions of mapping class groups. Preprint. [303](#), [320](#), [361](#), [363](#)
- [39] E. Klarreich, The boundary at infinity of the curve complex and the relative Teichmüller space. Preprint, available at <http://nasw.org/users/klarreich/>. [301](#), [339](#), [343](#)
- [40] M. Korkmaz, Automorphisms of complexes of curves on punctured spheres and on punctured tori. *Topology Appl.* 95 (1999), 85–111. [304](#)
- [41] A. A. Kosinski, *Differential manifolds*. Pure Appl. Math. 138. Academic Press, Inc., Boston, MA, 1993. [308](#)
- [42] F. Luo, Automorphisms of the complex of curves. *Topology* 39 (2000), 283–298. [304](#)
- [43] H. A. Masur and Y. N. Minsky, Geometry of the complex of curves. I. Hyperbolicity. *Invent. Math.* 138 (1999), 103–149. [301](#), [304](#)
- [44] H. A. Masur and Y. N. Minsky, Geometry of the complex of curves. II. Hierarchical structure. *Geom. Funct. Anal.* 10 (2000), 902–974. [339](#)
- [45] J. McCarthy and A. Papadopoulos, Dynamics on Thurston’s sphere of projective measured foliations. *Comment. Math. Helv.* 64 (1989), 133–166. [302](#), [307](#)
- [46] C. T. McMullen, The moduli space of Riemann surfaces is Kähler hyperbolic. *Ann. of Math.* (2) 151 (2000), 327–357. [362](#)
- [47] Y. N. Minsky, A geometric approach to the complex of curves on a surface. In *Topology and Teichmüller spaces* (Katinkulta, 1995), World Sci. Publishing, River Edge, NJ, 1996, 149–158. [304](#)
- [48] N. Monod and Y. Shalom, Orbit equivalence rigidity and bounded cohomology, *Ann. of Math.* (2) 164 (2006), 825–878. [302](#), [327](#), [330](#), [363](#)
- [49] L. Mosher, Geometric survey of subgroups of mapping class groups. in *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 387–410. [303](#)
- [50] D. S. Ornstein, Bernoulli shifts with the same entropy are isomorphic. *Adv. Math.* 4 (1970), 337–352. [331](#)
- [51] D. S. Ornstein, Two Bernoulli shifts with infinite entropy are isomorphic. *Adv. Math.* 5 (1970), 339–348. [331](#)
- [52] D. S. Ornstein and B. Weiss, Ergodic theory of amenable group actions. I. The Rohlin lemma. *Bull. Amer. Math. Soc. (N.S.)* 2 (1980), 161–164. [301](#), [330](#)
- [53] D. S. Ornstein and B. Weiss, Entropy and isomorphism theorems for actions of amenable groups. *J. Analyse Math.* 48 (1987), 1–141. [331](#)
- [54] N. Ozawa, Amenable actions and applications. In *Proceedings of the International Congress of Mathematicians* (Madrid, 2006), Vol. II, EMS Publishing House, Zurich 2006, 1563–1580. [331](#), [339](#), [362](#)
- [55] S. Popa, Strong rigidity of II_1 factors arising from malleable actions of w -rigid groups, II. *Invent. Math.* 165 (2006), 409–451. [330](#)
- [56] S. Popa, Cocycle and orbit equivalence superrigidity for malleable actions of w -rigid groups. *Invent. Math.* 170 (2007), 243–295. [330](#)
- [57] K. J. Shackleton, Combinatorial rigidity in curve complexes and mapping class groups. *Pacific J. Math.* 230 (2007), 217–232. [361](#)

- [58] Y. Shalom, Measurable group theory. In *European Congress of Mathematics* (Stockholm 2004), EMS Publishing House, Zurich 2005, 391–423. [302](#)
- [59] Y. Takesaki, *Theory of Operator algebras. III*. Encyclopaedia Math. Sci. 127, Operator Algebras and Non-commutative Geometry 8, Springer-Verlag, Berlin 2003. [308](#)
- [60] S. Vaes, Rigidity results for Bernoulli actions and their von Neumann algebras (after Sorin Popa). *Astérisque* 311 (2007), 237–294. [302](#), [330](#)
- [61] G. Yu, Higher index theory of elliptic operators and geometry of groups, In *Proceedings of the International Congress of Mathematicians* (Madrid, 2006), Vol. II, EMS Publishing House, Zurich 2006, 1623–1639. [362](#)
- [62] R. J. Zimmer, Amenable ergodic group actions and an application to Poisson boundaries of random walks. *J. Functional Analysis* 27 (1978), 350–372. [301](#), [331](#), [333](#)
- [63] R. J. Zimmer, *Ergodic theory and semisimple groups*. Monogr. Math. 81, Birkhäuser Verlag, Basel 1984. [301](#), [329](#), [331](#), [332](#), [338](#)

Chapter 10

Affine groups of flat surfaces

Martin Möller

Contents

1	Introduction	369
2	Basic properties of affine groups	370
3	Thurston’s construction and implications for the trace field	374
3.1	Trace fields of affine groups	375
4	The Arnoux–Yoccoz surface as a multi-purpose counter-example	377
5	Large affine groups: Veech surfaces	379
6	More on large affine groups: infinitely generated	383
7	The size of the affine group compared to the size of the orbit closure	384
	References	386

1 Introduction

According to Thurston’s classification, a diffeomorphism of a closed oriented surface is either elliptic, reducible or a pseudo-Anosov diffeomorphism. The structure of pseudo-Anosov diffeomorphisms, e.g. their dilatation coefficients, and the corresponding measured foliations have been intensely investigated since. Rather than focusing on properties of a single diffeomorphism, the purpose of this chapter is to study the flat surfaces the pseudo-Anosov diffeomorphisms live on together with their whole group of affine diffeomorphisms.

A *flat surface* is a pair (X, ω) consisting of a Riemann surface X together with a holomorphic one-form $\omega \in \Gamma(X, \Omega_X^1)$. Equivalently, flat surfaces arise from gluing rational-angled planar polygons by parallel translations along their edges. Furthermore, flat surfaces naturally arise when studying the trajectories of a ball on a rational-angled billiard table.

One of the basic invariants of a flat surface (X, ω) is the *affine group* $\mathrm{SL}(X, \omega)$ (also called *Veech group*) defined as follows. Let $\mathrm{Aff}^+(X, \omega)$ be the group of orientation-preserving diffeomorphisms that are affine on the complement of the zero set of ω with respect to the charts defined by integrating ω . The linear part of the affine map is independent of the charts and provides a map

$$D: \mathrm{Aff}^+(X, \omega) \rightarrow \mathrm{SL}_2(\mathbb{R}).$$

The image of D is called the affine group $\mathrm{SL}(X, \omega)$. The interest in these groups stems from Veech's paper [38], where flat surfaces are constructed whose affine groups are non-arithmetic lattices in $\mathrm{SL}_2(\mathbb{R})$. We will discuss them and recent developments in Section 5.

This chapter touches the following aspects of affine groups. The affine group is often said to be trivial for a generic surface. If we define generic to be meant in the strata of a natural stratification of the space of flat surfaces, this is a little imprecise. Our first goal is to give a complete description of the generic affine group. Next we recall Thurston's construction of pseudo-Anosov diffeomorphisms using a pair of multicurves. This construction has a lot of flexibility and produces rather large affine groups. Results of McMullen resp. of Hubert–Lanneau show that in genus two all pseudo-Anosov diffeomorphisms arise in this way but that this holds no longer for $g \geq 3$.

In Sections 5 and 6 we review the known constructions of very large affine groups: lattices and infinitely generated affine groups. In the last section we discuss some relations between the size of the affine group and the closure of $\mathrm{SL}_2(\mathbb{R})$ -orbit of the corresponding flat surface in the moduli space of flat surfaces.

We remark that the affine group is similarly defined for pairs (X, q) of a Riemann surface X and a quadratic differential q . But such a surface admits a canonical double covering which is a flat surface. Hence up to passing to finite index subgroups all the information is contained in affine groups of flat surfaces.

The whole topic is not completely understood at the time of writing. Consequently, the content of this chapter simply reflects the present state of knowledge and almost all sections are concluded by an open problem.

The author thanks Erwan Lanneau for a helpful discussion on the proof of Theorem 2.1.

2 Basic properties of affine groups

Our first aim is to realize that for a general flat surface nothing exciting happens. In order to define what 'general' means, we define the parameterizing space of flat surfaces. Let M_g denote the moduli space of curves of genus g . Over M_g there is a vector bundle of rank g whose fiber over a point corresponding to the surface X is the vector space of holomorphic one-forms (or *abelian differentials*) on X . Let ΩM_g be the total space of this vector bundle minus the zero section. By construction, flat surfaces correspond to points of ΩM_g . The space ΩM_g is stratified into subspaces

$$\Omega M_g = \bigcup_{\sum_{i=1}^n k_i = 2g-2} \Omega M_g(k_1, \dots, k_n)$$

according to the number and multiplicities of the zeros of the holomorphic one-form ω . Some of the strata are not connected, see [18]. A component of a stratum is called a

hyperelliptic component if it consists exclusively of hyperelliptic curves, i.e. curves with a degree two map to the projective line.

The strata are complex orbifolds that carry a natural complex coordinate system, called period coordinates, whose definition will be recalled below. We say that (X, ω) is generic in its stratum if it lies outside a countable union of real codimension one submanifolds in its stratum.

Theorem 2.1. *For $g(X) \geq 2$, the affine group of a generic surface (X, ω) is $\mathbb{Z}/2\mathbb{Z}$ or trivial, depending on whether (X, ω) belongs to a hyperelliptic component or not.*

Before we can give the proof we need to recall some facts on flat surfaces and to classify affine diffeomorphisms in order to explain the notions in Thurston's theorem stated at the beginning of the introduction.

As stated in the introduction, a *flat surface* is a pair (X, ω) of a Riemann surface X together with a holomorphic one-form ω . A flat surface has a finite number of zeros of ω , called *singularities*. These correspond to points where the total angle with respect to $|\omega|$ exceeds 2π . On a flat surface we may talk of geodesics with respect to the metric $|\omega|$. Such a geodesic has a well-defined *direction* in $\mathbb{R}\mathbb{P}^1$. A geodesic joining two singularities or a singularity to itself is called a *saddle connection*.

Definition 2.2. A diffeomorphism φ of X is called *elliptic* if it is isotopic to a diffeomorphism of finite order. A diffeomorphism φ is called *reducible* if it is isotopic to a diffeomorphism fixing a (real) simple closed curve on X . If φ is neither reducible nor elliptic, then φ is called *pseudo-Anosov*.

We alert the reader that we follow the common abuse of the notion diffeomorphism for homeomorphisms that are C^1 outside a finite set of points ([7], Exposé V).

It is easy to see that an affine diffeomorphism φ of (X, ω) is elliptic if it is of finite order. In particular $D(\varphi)$ is of finite order. Conversely, if $D(\varphi)$ is of finite order, then φ is of finite order, since $\text{Ker}(D)$ consists of holomorphic diffeomorphisms of X and consequently $\text{Ker}(D)$ is finite by Hurwitz' theorem.

If φ is a pseudo-Anosov diffeomorphism, there exists a pair (X, q) such that φ is an affine diffeomorphism of (X, q) . As stated above, we will restrict to the case that $q = \omega^2$. Moreover, (X, ω) can be chosen such that φ stretches the horizontal lines by some factor $\lambda > 1$, called *dilatation coefficient*, and contracts the vertical lines by the same factor λ . Thus, $|\text{tr}D(\varphi)| > 2$ for an affine pseudo-Anosov diffeomorphism.

Consequently, an affine diffeomorphism φ with $|\text{tr}D(\varphi)| = 2$, i.e. such that $D(\varphi)$ is parabolic, is a reducible affine diffeomorphism. We briefly recall the structure of such a *parabolic diffeomorphism*. Say the horizontal direction is the eigendirection of $D(\varphi)$. Then some power of φ fixes all the finitely many horizontal saddle connections and the complement of these saddle connections has to consist of metric cylinders.

In order to define coordinates on a stratum of ΩM_g , fix locally on some open set U a basis of the integral homology $H_1(X, Z(\omega), \mathbb{Z})$ relative to $Z(\omega)$, the zeros of ω . The cardinality of the basis is $N = 2g - 1 + n$, where n is the number of zeros of ω . The

map $U \rightarrow \mathbb{C}^N$, that maps (X, ω) to the integrals of ω along the fixed basis, is a local diffeomorphism ([37], see [29] for an algebraic proof). The system of coordinates is called *period coordinates*.

There is a natural action of $\mathrm{GL}_2^+(\mathbb{R})$ on ΩM_g . In terms of period coordinates, consider

$$\mathbb{C}^N \cong \mathbb{R}^N \otimes_{\mathbb{R}} \mathbb{R}^2$$

and let $\mathrm{GL}_2^+(\mathbb{R})$ act naturally on \mathbb{R}^2 . This is equivalent to letting an element of $\mathrm{GL}_2^+(\mathbb{R})$ act on the local complex charts of X given by integration of ω as real linear map. This action leaves the affine group essentially unchanged, we have for $A \in \mathrm{GL}_2^+(\mathbb{R})$

$$\mathrm{SL}(A \cdot (X, \omega)) = A \cdot (\mathrm{SL}(X, \omega)) \cdot A^{-1}.$$

Proof of Theorem 2.1. For each of the countably many pseudo-Anosov diffeomorphisms φ in the mapping class group there is a unique flat surface (X, q) or (X, ω) up to the action of $\mathrm{GL}_2^+(\mathbb{R})$, such that φ is an affine diffeomorphism on (X, q) or (X, ω) respectively. Consequently, the set of flat surfaces whose affine group contains a pseudo-Anosov element is a countable union of real 4-dimensional subspaces. Since we assume $g(X) \geq 2$, the generic flat surface does not carry any affine pseudo-Anosov diffeomorphism.

Suppose that $\mathrm{SL}(X, \omega)$ contains a parabolic element. By the classification above, (X, ω) decomposes into metric cylinders in some direction. The boundaries of these cylinders consist of saddle connections and, since $g \geq 2$, at least two of them, say γ_1 and γ_2 , are not homologous, i.e. they are linearly independent elements of $H_1(X, Z(\omega), \mathbb{Z})$. Since the saddle connections are parallel, the periods of γ_1 and γ_2 are \mathbb{R} -linearly dependent. The locus of surfaces where γ_1 and γ_2 are linearly dependent is of \mathbb{R} -codimension at least one in period coordinates. Since the γ_i in question are two elements in the countable group $H_1(X, Z(\omega), \mathbb{Z})$, the generic flat surface does not contain an affine parabolic element.

The remaining discussion serves to prove that the number of elliptic affine diffeomorphisms is as small as claimed. Suppose the generic flat surface (X, ω) in a stratum contains such a diffeomorphism φ of finite order. Each stratum contains square-tiled surfaces and their affine group is a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ (see Section 5). In $\Omega \mathcal{T}_g$, the pullback bundle to Teichmüller space, the presence of an affine diffeomorphism of finite order is a closed condition. Consequently, a generic affine diffeomorphism must be reflected in the affine group of all square-tiled surfaces, hence $\mathrm{ord}(D(\varphi)) \in \{1, 2, 3, 4, 6\}$.

First suppose that $D(\varphi)$ is the identity or minus the identity, in particular $D(\varphi) \in \mathrm{SO}_2(\mathbb{R})$. Then φ is in fact an automorphism of X and fixes $q = \omega^{\otimes 2}$. We are faced with the problem of classifying strata of half-translation surfaces that consist entirely of pullbacks of half-translation surfaces of lower genus. This classification was solved in [19], although precisely the case of squares of abelian differential was excluded from the discussion in loc. cit. The difference between the cases is apparent whenever

the dimension count is involved, since

$$\dim Q_g(k_1, \dots, k_n) = 2g - 2 + n \quad \text{while} \quad \dim \Omega M_g(k_1, \dots, k_n) = 2g - 1 + n.$$

We let g_0 denote the genus of the quotient surface $X/\langle\varphi\rangle$ and we let d be the degree of the covering, i.e. the order of φ . Moreover, let p be the number of poles of q , let r be the number of zeros of q over which the covering $\pi: X \rightarrow X/\langle\varphi\rangle$ is unramified, and let m be the number of regular points of q , over which π is ramified. Finally, let n be the number of zeros of q over which π is ramified.

The first case is $D(\varphi) = \text{id}$. Then ω is the pullback of an abelian differential and we obtain, as in loc. cit. using the Riemann–Hurwitz formula, that

$$(d-1)(2g_0 - 2 + n + r) \leq -m.$$

We deduce $g_0 = 0$, which is absurd since the projective line carries no abelian differentials.

The second case is $D(\varphi) = -\text{id}$. Then ω is the pullback of a strictly quadratic differential. An analysis of the covering results this time in

$$(d-1)(2g_0 - 2 + m + n + r) \leq \begin{cases} m(d-2) + pd/2 - 1 & \text{if } d \text{ is even,} \\ m(d-2) + p(d-1)/2 - 1 & \text{if } d \text{ is odd.} \end{cases}$$

This implies $g_0 = 0$ and for $d = 2$ one obtains the hyperelliptic components. For $d \geq 3$ we deduce $n+r \leq 1$ and since $p \geq 4$ the case $n+r = 1$ is absurd. If $n+r = 0$, we conclude that $p = 4$, that d is even and that $m \in \{1, 2\}$. This case is excluded in the same way as the corresponding case in the proof of Theorem 1 in [19].

We finally have to treat the cases where $D(\varphi)$ has order 3, 4 or 6. In these cases $D(\varphi)$ is conjugate to an element in $\text{SO}_2(\mathbb{R})$ and φ is actually an automorphism if $D(\varphi) \in \text{SO}_2(\mathbb{R})$. Consequently, for each (X, ω) the $\text{SL}_2(\mathbb{R})$ -orbit contains a flat surface where the conjugate of φ is actually an automorphism. It thus suffices to prove that in each stratum the locus of flat surfaces with an automorphism of order 3, 4 and 6 is of codimension more than one.

We start with the case $\text{ord}(D(\varphi)) = 3$. Consider the surface $X/\langle\varphi\rangle$ marked with s images of the ramification points. We give the details in the case $\text{ord}(\varphi) = 3$, in all other cases even cruder dimension estimates suffice. The quotient surface has

$$3g_0 - 3 + s = g - 1$$

moduli by Riemann–Hurwitz. If (X, ω) lies in the generic stratum, the locus of flat surfaces with such an automorphism has dimension

$$g - 1 + g < (4g - 3) - 1 = \dim \Omega M_g(1, \dots, 1) - 1,$$

since $g > 1$. If (X, ω) lies in a non-generic stratum S , then the fiber of $S \rightarrow M_g$ has at most dimension $g - 1$ and again the locus of flat surfaces with such an automorphism has dimension less than $\dim(S) - 1$.

The case $\text{ord}(D(\varphi)) = 6$ is contained in the previous one by considering φ^2 . The same trick allows to reduce the case $\text{ord}(D(\varphi)) = 4$ to the hyperelliptic loci. Again the Riemann–Hurwitz formula yields that the quotient surface has not enough moduli. \square

Proposition 2.3 ([38], Proposition 2.7). *The group $\text{SL}(X, \omega)$ is a discrete subgroup of $\text{SL}_2(\mathbb{R})$.*

Proof. Let φ_n be a sequence of affine diffeomorphisms such that $D(\varphi_n)$ converges to the identity. By Arzela–Ascoli and after passing to a subsequence, we may suppose that φ_n converges to some affine diffeomorphism φ uniformly on X . Hence for large enough n , the composition $\varphi_n \varphi_{n+1}^{-1}$ is isotopic to the identity. Using Thurston’s classification of diffeomorphisms this is not possible unless $D(\varphi_n) = \text{id}$ for large enough n . \square

Concerning the existence of cyclic affine groups, the parabolic case is easy, while the hyperbolic case seems wide open at present. The proof of Proposition 2.4 will be given in the next section.

Proposition 2.4. *In every stratum there exist flat surfaces whose affine group is cyclic generated by a parabolic element.*

Question 2.5. Does there exist a flat surface (X, ω) whose affine group $\text{SL}(X, \omega)$ is cyclic generated by a hyperbolic element?

3 Thurston’s construction and implications for the trace field

The following construction first appears in Thurston’s famous 1976 preprint ([34]), see also [31], [38], [20], [11] and [25], and the chapter by Harvey in Volume I of this Handbook ([9]) for different versions and presentations.

A *multicurve* A on a surface Σ_g of genus g is a union of disjoint essential simple closed curves, no two of which bound an annulus. A pair (A, B) of multicurves *fills* (or *binds*) the surface if for each curve in A and each curve in B the geometric intersection number is minimal in their homotopy classes and if the complement $\Sigma_g \setminus (A \cup B)$ is a simply connected polygonal region with at least 4 sides.

We index the components of A and B such that $A = \bigcup_{i=1}^a \gamma_i$ and $B = \bigcup_{i=a+1}^{a+b} \gamma_i$ and let C be the (unsigned) intersection matrix of A and B , i.e. for $i \neq j$ we have $C_{ij} = |\gamma_i \cap \gamma_j|$ and $C_{jj} = 0$ for all j .

As additional input datum for the construction we fix a set of multiplicities $m_i \in \mathbb{N}$ for $i = 1, \dots, a + b$. Since (A, B) fills Σ_g , the intersection graph is connected and the matrix $(m_i C_{ij})$ is a Perron–Frobenius matrix. Hence there is a unique positive eigenvector (h_i) up to scale such that

$$\mu h_i = \sum_{j=1}^{j=a+b} m_i C_{ij} h_j \quad (3.1)$$

for some positive eigenvalue μ .

We now glue a surface X from rectangles $R_p = [0, h_i] \times [0, h_j] \subset \mathbb{C}$ for each intersection point $p \in \gamma_i \cap \gamma_j$. Namely, glue R_p to R_q along the vertical (resp. horizontal) sides whenever p and q are joined by an edge in A (resp. B) of the graph $A \cup B$. The differentials dz^2 on each rectangle glue to a global quadratic differential q on X .

Let τ_i be the Dehn twist around γ_i and define

$$\tau_A = \prod_{i=1}^a \tau_i^{m_i} \quad \text{and} \quad \tau_B = \prod_{i=a+1}^{a+b} \tau_i^{m_i}.$$

Theorem 3.1 ([34]). *The flat surface (X, q) constructed above contains affine diffeomorphisms τ_A and τ_B with derivatives*

$$D\tau_A = \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad D\tau_B = \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix}$$

In particular the elements $\tau_A^n \tau_B$ are pseudo-Anosov diffeomorphisms for n large enough.

Proof. By construction the modulus of the cylinder with core curve γ_i is m_i/μ . Hence the powers of the Dehn twists occurring in the definition of τ_A and τ_B have linear part as claimed. They fix the boundary of the horizontal resp. vertical cylinders and together define affine diffeomorphisms.

In order to check the last claim, one has to recall that an affine diffeomorphism is pseudo-Anosov if and only if the absolute value of its trace is greater than two. \square

Since we are dealing exclusively with flat surfaces in the sequel, we remark that the quadratic differential has a square root, i.e. $q = \omega^2$ if and only if for a suitable orientation of the γ_i their geometric and algebraic intersection numbers coincide.

3.1 Trace fields of affine groups

Given a pair (X, ω) resp. (X, q) we define the *trace field of the affine group* $\text{SL}(X, \omega)$ to be $K = \mathbb{Q}(\text{tr}(A), A \in \text{SL}(X, \omega))$. The notion of trace field is a useful invariant since it turns out to be stable under passing to a finite index subgroup.

Theorem 3.2 ([17], Appendix; [23]). *Let $A = D\varphi \in \text{SL}(X, \omega)$ be any hyperbolic element. Then the trace field of the affine group equals the trace field of φ . More precisely, if $\text{SL}(X, \omega)$ contains a hyperbolic element A , then the \mathbb{Q} -vector space generated by the periods of ω is a 2-dimensional K -vector space, where $K = \mathbb{Q}(\text{tr}(A))$.*

With this result we can obviously determine the trace fields of affine groups arising from Thurston's construction.

Corollary 3.3. *If φ is constructed using a pair of multicurves then $K = \mathbb{Q}(\mu^2)$, where μ is as in Equation (3.1).*

Hubert and Lanneau have shown that Thurston’s construction imposes a restriction on the trace field. We will see below (Corollary 4.1) that this property does not hold for all pseudo-Anosov diffeomorphisms.

Theorem 3.4 ([11]). *If (X, ω) is given by Thurston’s construction, then the trace field K of $\text{SL}(X, \omega)$ is totally real, i.e. all embeddings $K \rightarrow \mathbb{C}$ factor through \mathbb{R} . In particular, if $\text{SL}(X, \omega)$ contains two non-commuting parabolic elements then K is totally real.*

Proof of Theorem 3.4. Let D_m be the diagonal matrix with entries m_i . The square of the largest eigenvalue of the matrix C (as in Thurston’s construction) is the largest eigenvalue of the matrix C^2 . Hence we have to show that all the eigenvalues of $(D_m C)^2$ are real.

Suppose first for simplicity $m_i = 1$ for all i . Since for some matrix C_0 we have

$$D_m C = C = \begin{pmatrix} 0 & C_0 \\ C_0^\top & 0 \end{pmatrix}, \quad \text{hence } (D_m C)^2 = C^2 = \begin{pmatrix} C_0 C_0^\top & 0 \\ 0 & C_0^\top C_0 \end{pmatrix}$$

Since C^2 is symmetric, all its eigenvalues are real. Thus $\mathbb{Q}(\mu^2)$ is totally real.

If the m_i are no longer identically one, $(D_m C)^2$ is still similar to a symmetric matrix: Split D_m into two pieces D'_m and D''_m of size a resp. b and let $D'_{\sqrt{m}}$ resp. $D''_{\sqrt{m}}$ denote the diagonal matrix with entries $\sqrt{m_i}$. Then

$$(D_m C)^2 = \begin{pmatrix} D'_m C_0 D''_m C_0^\top & 0 \\ 0 & D''_m C_0^\top D'_m C_0 \end{pmatrix}.$$

The upper block decomposes as

$$D'_m C_0 D''_m C_0^\top = D'_{\sqrt{m}} (D'_{\sqrt{m}} C_0 D''_{\sqrt{m}}) (D'_{\sqrt{m}} C_0 D''_{\sqrt{m}})^\top (D'_{\sqrt{m}})^{-1}$$

and for the lower block the same trick works. The above conclusion about the eigenvalues thus still holds. □

We can now easily give a proof of a statement from the previous section.

Proof of Proposition 2.4. It is easy to construct in each stratum a flat surface (X, ω) that consists of only one cylinder horizontally. Consequently, $\text{SL}(X, \omega)$ contains a parabolic element φ irrespectively of the lengths of the horizontal saddle connections. Since $g(X) > 1$ by hypothesis, we may arrange that the periods of all horizontal saddle connections generate a K -vector space of dimension two or more, where K is real, but not totally real. By Theorem 3.2 and Theorem 3.4, the affine group $\text{SL}(X, \omega)$ does not contain two non-commuting parabolic elements. Suppose $\text{SL}(X, \omega)$ contains a hyperbolic or an elliptic element ψ . Then φ and $\psi \varphi \psi^{-1}$ are non-commuting parabolic elements. This contradiction completes the claim. □

Remark 3.5. Recent results on Thurston's construction can be found in [20]. E.g., the smallest dilatation coefficients of the pseudo-Anosov diffeomorphisms arising from Thurston's construction are determined there.

4 The Arnoux–Yoccoz surface as a multi-purpose counter-example

The following construction is a special case of what has become known as the construction of zippered rectangles ([36], [21]). We sketch the version in [2] where the construction is given for hyperelliptic curves of genus g . The construction appeared originally in [3]. It will be used below to refute many naive conjectures one could derive seeing only the construction from the preceding sections.

We first construct an *interval exchange transformation* $f: [0, 1) \rightarrow [0, 1)$, i.e. a map that consists of translations on a subdivision of intervals. Here we take α to be the real root of

$$\alpha^g + \alpha^{g-1} + \dots + \alpha = 1$$

and subdivide $[0, 1)$ into g pairs of subintervals I_k of lengths $\alpha^k/2$, $k=1, \dots, g$. We let f be the exchange of the pairs of same length composed with a half turn, where $[0, 1)$ is identified with a circle. Explicitly, $f = f_r \circ f_s$, where for $k = 0, \dots, g-1$ we let

$$f_s(x) := \begin{cases} x + \alpha^{k+1}/2 & \text{if } x \in [\sum_{i=1}^k \alpha^i, \alpha^{k+1}/2 + \sum_{i=1}^k \alpha^i), \\ x - \alpha^{k+1}/2 & \text{if } x \in [\alpha^{k+1}/2 + \sum_{i=1}^k \alpha^i, \sum_{i=1}^{k+1} \alpha^i), \end{cases}$$

and define

$$f_r(x) = \begin{cases} x + 1/2 & \text{if } x \in [0, 1/2), \\ x - 1/2 & \text{if } x \in [1/2, 1). \end{cases}$$

The interval exchange f enjoys a remarkable 'self-similarity' property, inherited from the form of the minimal polynomial of α . The map

$$\varphi_I : \begin{cases} [0, 1) \rightarrow [0, \alpha) \\ x \mapsto \begin{cases} \alpha x + (\alpha + \alpha^{g+1})/2 & \text{if } x \in [0, (1 - \alpha^g)/2), \\ \alpha x - (\alpha + \alpha^{g+1})/2 & \text{if } x \in [(1 - \alpha^g)/2, 1), \end{cases} \end{cases}$$

which shrinks the interval linearly by the factor α and then exchanges the two pieces (of unequal length), commutes with f resp. its induction on the subinterval $[0, \alpha)$, i.e.

$$\varphi_I \circ f = f|_{[0, \alpha)} \circ \varphi_I,$$

where

$$f|_{[0, \alpha)}(x) = f^n(x), \text{ where } n \in \mathbb{N}_{>0} \text{ is minimal such that } f^n(x) \in [0, \alpha).$$

Suppose we let each of the intervals I_i be the bottom of a rectangle of height h_i and glue the top of these rectangles to $[0, 1)$ according to f . Under some conditions (see e.g. [36] or the survey [39], §5, for details) the sides of the rectangles can be glued to yield a translation surface. Here we take

$$h_i = \sum_{j=1}^{g-k} \alpha^j \quad \text{for } i = 2k + 1, 2k + 2.$$

The resulting flat surface (X_{AY}, ω_{AY}) has genus g with two singularities of type $g - 1$, i.e. with angle $2g\pi$.

The main point in this choice of heights is that the self-similarity of the base interval is also reflected in the gluing of the vertical sides. That is, there is a map $\varphi: X_{AY} \rightarrow X_{AY}$ that restricts to φ_I on the segment $[0, 1)$ and which stretches the vertical side by α^{-1} . Obviously φ is a pseudo-Anosov diffeomorphism with dilatation α . Since $\mathbb{Q}(\alpha + 1/\alpha)$ is not totally real, we conclude:

Corollary 4.1 ([11]). *There exist flat surfaces with a pseudo-Anosov diffeomorphism whose trace field is not totally real. In particular, there exist flat surfaces with a pseudo-Anosov diffeomorphism that does not arise via Thurston’s construction.*

The Arnoux–Yoccoz surfaces were thought to be good candidates to answer Question 2.5 affirmatively. But at least for $g = 3$ this is not the case.

Theorem 4.2 ([16]). *For (X_{AY}, ω_{AY}) the Arnoux–Yoccoz surface with $g(X) = 3$ the group $\text{SL}(X, \omega)$ is not cyclic.*

We sketch the proof in order to illustrate a phenomenon that yet needs deeper investigation. First, there exist many (diagonal) directions on (X_{AY}, ω_{AY}) that topologically look like the horizontal one in Figure 1, called $2T2C$ -direction in [16]. The vague ‘looks like topologically’ can be made precise using numerical invariants of a given direction, like the widths, heights and twists and some finite data, called *combinatorics*. The reader may consult [16] for the definition of these invariants.

Second, many of these $2T2C$ -directions have the same combinatorics and the same projectivised tuple of numerical invariants. Consequently, for each such pair of directions p, q , there exists an affine diffeomorphism of (X, ω) , that maps p to q . Necessarily, such a diffeomorphism is pseudo-Anosov.

Finally, there exist pairs of $2T2C$ -directions on (X_{AY}, ω_{AY}) with the same projectivised invariants, such that the corresponding pseudo-Anosov diffeomorphism is not a power of the diffeomorphism φ constructed above. Consequently, the affine group is not cyclic.

Question 4.3. How large is the affine group of (X_{AY}, ω_{AY}) ? Is it finitely generated or infinitely generated?

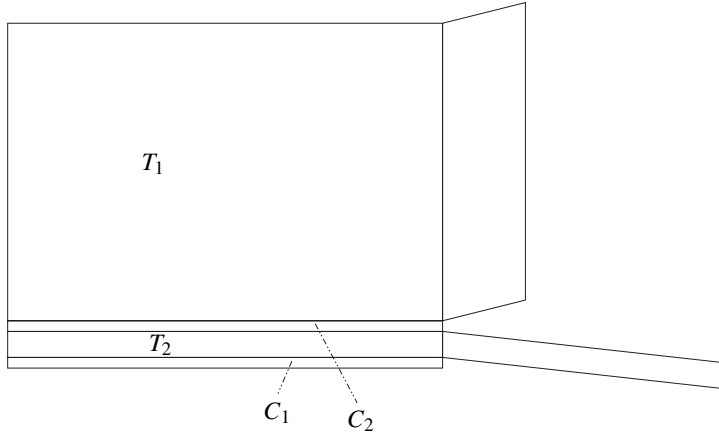


Figure 1. Topology of a special diagonal direction on the Arnoux–Yoccoz surface, displayed horizontally.

We finally mention three more questions, the Arnoux–Yoccoz surface provides a negative answer to. First, the dilatation coefficient of a pseudo-Anosov diffeomorphism of a surface of genus g is an algebraic number of degree r at most $2g$ over \mathbb{Q} . In Thurston’s original examples r turned out to be even, but the Arnoux–Yoccoz surface shows that odd r is possible too ([3]).

Second, the directional flow on a flat surface defines an interval exchange transformation (IET). An IET has an easily computable invariant, the SAF-invariant (compare [1] for the definition), that vanishes if the directional flow has periodic orbits only. For a surface of genus two the converse holds, but the Arnoux–Yoccoz surface shows that the converse does not hold in genus three ([1]).

The third question concerns $\mathrm{SL}_2(\mathbb{R})$ -orbit closures and will be dealt with in the last section.

5 Large affine groups: Veech surfaces

A flat surface (X, ω) is called a *Veech surface* if $\mathrm{SL}(X, \omega)$ is a lattice in $\mathrm{SL}_2(\mathbb{R})$. We do not want to address the dynamics of flat surfaces here, but we mention the most striking result, Veech’s dichotomy ([38]), for later use. If (X, ω) is a Veech surface then for each direction either

- all geodesics are uniformly distributed, in particular dense, or,
- all geodesics are closed or a saddle connection. Such directions are called *periodic*.

The presence of saddle connections on (X, ω) forces the lattice $\mathrm{SL}(X, \omega)$ to be non-cocompact. Up to coverings, all Veech surfaces known at the time of writing except for one arise from two fundamental constructions which we explain below:

- Quotients of cyclic coverings of the projective line branched at 4 points.
- Eigenforms for real multiplication by a quadratic field in genus two and the Prym varieties in genus $g \leq 5$.

From the point of view of affine groups we can state this result as follows:

Theorem 5.1 ([5]). *All triangle groups (m, n, ∞) for $1/m + 1/n < 1$ and $m, n \leq \infty$ arise as affine groups of Veech surfaces.*

Theorem 5.2 ([22], [6]). *All real quadratic fields arise as trace fields of lattice affine groups.*

Before we give the proofs we need to recall some background. Recall from Section 2 the action of $GL_2^+(\mathbb{R})$ on flat surfaces. By a theorem of Smillie (see [33] for a recent proof), the orbit of (X, ω) is closed in ΩM_g if and only if $SL(X, \omega)$ is a lattice in $SL_2(\mathbb{R})$. In this case the image C of the orbit in M_g is a complex, in fact algebraic curve, called a *Teichmüller curve*. We mention that such an algebraic curve is a totally geodesic subsurface for the Teichmüller metric, whence the name, but we won't need details on Teichmüller theory. Instead of considering a Teichmüller curve as a curve in the moduli space of curves, it is often useful to restrict the universal family over M_g to a family $f : \mathcal{X} \rightarrow C$ over the Teichmüller curve and to study f instead.

Proof of Theorem 5.1. We extract from [5] the special case where m, n are both odd, finite and coprime. This case illustrates almost all ideas, except for a fiber product construction needed to cover the general case.

The basic idea is to study a family of cyclic coverings ramified over the projective line at 4 points. There is a criterion ([27], we will apply the version [5] Theorem 1.2 (b)) that detects Teichmüller curves by the existence of an eigenspace of the relative de Rham cohomology, whose monodromy group is the affine group. For appropriate cyclic coverings, there is such an eigenspace whose monodromy group is the desired triangle group. But the family of cyclic coverings does not quite match the cohomological criterion as we shall see, so we need furthermore to find a suitable quotient family.

Consider the family of cyclic degree N covering

$$\mathcal{Y}_t : y^N = x^{a_1}(x - 1)^{a_2}(x - t)^{a_3}$$

of \mathbb{P}_x^1 with t varying in $\mathbb{P}_t^1 \setminus \{0, 1, \infty\}$, where $N = 2mn$ and

$$a_1 = 2mn - m + n, \quad a_2 = 2mn + m - n, \quad a_3 = 2mn + m + n, \quad a_4 = 2mn - m - n.$$

The coverings is ramified precisely over $x = 0, x = 1, x = t$ and $x = \infty$. Let $\mathbb{L}(i)$ denote the ζ_N^i -eigenspace of the relative de Rham cohomology for the automorphism $\varphi : (x, y) \mapsto (x, \zeta_N y)$. The local systems $\mathbb{L}(1), \mathbb{L}(-1), \mathbb{L}(mn + 1), \mathbb{L}(mn - 1)$ are isomorphic and the a_i are chosen such that the monodromy group is $\Delta(m, n, \infty)$.

We claim that we can lift the automorphism of \mathbb{P}_x^1 that interchanges the points $\{0, 1, t, \infty\}$ in pairs to an automorphism group H of \mathcal{Y}_t such that the stable model of

the fibers \mathcal{Y}_0/H and \mathcal{Y}_1/H are smooth. Let $\mathcal{X} = \mathcal{Y}/H$ be the quotient and denote by $f: \mathcal{X} \rightarrow \mathbb{P}_t^1$ the corresponding family of curves. Given the claim, the local system

$$\mathbb{L}(1) \oplus \mathbb{L}(-1) \oplus \mathbb{L}(\alpha) \oplus \mathbb{L}(-\alpha)$$

is made to be H -invariant. Thus, the de Rham cohomology of \mathcal{X}_t has a local subsystem of rank two with discrete monodromy group $\Delta(m, n, \infty)$ and the fibers of the family \mathcal{X}_t are smooth precisely over $\mathbb{H}/\Delta(m, n, \infty)$. We conclude using the characterization of Teichmüller curves given in [5] Theorem 1.2 (b).

To establish the claim, choose elements $t^{1/n}, (t - 1)^{1/m} \in \overline{\mathbb{C}(t)}$ and define

$$c = (t - 1)^{\sigma_2 + \sigma_3}, \quad d = t^{\sigma_1 + \sigma_3}. \tag{5.1}$$

We now define $H = \langle \sigma, \tau \rangle$ by

$$\sigma(z) = cd \frac{x(x - 1)}{y(x - t)} = cd \frac{-y}{(x - t)^2}$$

and for $\alpha \equiv 1 \pmod m$ and $\alpha \equiv -1 \pmod n$

$$\tau(z) = d \frac{y(\alpha)}{x^2}.$$

The fiber over $t = 0$ consists of two smooth components with affine charts $y^N = x^{a_1}(x - 1)^{a_3}$ (since branch points of type a_1 and a_3 have come together) and $y^N = x^{a_2}(x - 1)^{a_4}$. They meet in $\gcd(a_1 + a_3, N)$ points transversally. One of the elements in H exchanges the two components and, in fact, fixes the intersection points of the components. Consequently, the quotient is smooth as claimed. See [5] for the details. \square

Proof of Theorem 5.2. Consider a curve X of genus two, such that its Jacobian $\text{Jac}(X)$ has more endomorphisms than just multiplication by an integer, namely such that $\text{End}(\text{Jac}(X))$ is an order \mathfrak{o}_D in a real quadratic field $\mathbb{Q}(\sqrt{D})$. These endomorphisms act on the space of holomorphic one-forms of $\text{Jac}(X)$, which is in natural bijection with the space of holomorphic one-forms on X . Let $\mathcal{E}_D \subset \Omega M_2$ be the locus of flat surfaces (X, ω) , such that $\text{Jac}(X)$ has real multiplication by \mathfrak{o}_D and such that ω is an eigenform for the action of \mathfrak{o}_D on the space of holomorphic one-forms. Obviously, $\mathcal{E}_K \subset \Omega M_2$ is a closed subvariety and the main point is to show that \mathcal{E}_D is invariant under the action of $\text{SL}_2(\mathbb{R})$. Granted this, the intersection $W_D = \mathcal{E}_D \cap \Omega M_2(2)$ is again closed and $\text{SL}_2(\mathbb{R})$ -invariant. A local dimension count shows that the image of W_D in M_2 is a curve, by construction a Teichmüller curve.

We now single out the role of genus two rather than rigorously proving the main point. Let $A = \mathbb{C}^g/\Lambda$ be a g -dimensional abelian variety. An endomorphism of A consists of an endomorphism of the lattice Λ plus a linear map of \mathbb{C}^g with the obvious compatibility condition. Suppose that (X, ω) is an eigenform for real multiplication and T a generator of \mathfrak{o}_D . For any $M \in \text{GL}_2^+(\mathbb{R})$ let $M \cdot (X, \omega) = (Y, \eta)$. By definition of the $\text{GL}_2^+(\mathbb{R})$ -action, there is an affine diffeomorphism $\varphi_M: X \rightarrow Y$. The map $\varphi_M \circ T \circ \varphi_M^{-1}$ defines a map of Λ , where $\text{Jac}(Y) = \mathbb{C}^g/\Lambda$. Moreover this map

preserves the complex line

$$M \cdot \langle \Re\omega, \Im\omega \rangle = \langle \Re\eta, \Im\eta \rangle.$$

Since here $g = 2$, the orthogonal complement of $\langle \Re\omega, \Im\omega \rangle$ with respect to the symplectic form is also a complex line, also preserved by $\varphi_M \circ T \circ \varphi_M^{-1}$. Consequently, (Y, ω) has also real multiplication by σ_D . See [22] for the missing details. \square

In fact, the orbifold Euler characteristic of the quotients $\mathbb{H}/\mathrm{SL}(X, \omega)$ for (X, ω) as constructed in Theorem 5.2 has been determined by Bainbridge ([4]). The complete description of the structure (starting with the number of elliptic elements) of these Veech groups is an open question. Only for the 12 smallest examples, when $\mathbb{H}/\mathrm{SL}(X, \omega)$ is a rational curve, the affine groups are known by generators and relations. It would be interesting to have such a description of the Veech groups for the whole series.

Square-tiled surfaces, covering constructions

Let (Y, η) be a Veech surface. A point P on Y is called a *periodic point* if the orbit $\mathrm{SL}(X, \omega) \cdot P$ is finite. A covering surface $\pi : X \rightarrow Y$ provided with the flat structure $\omega = \pi^*\eta$ is again a Veech surface if and only if π is branched at most over periodic points ([8]).

For the rest of this section we suppose that Y is the torus. In this case, periodic points are precisely the torsion points on Y , if we normalize $0 \in Y$ to be one of the branch points. For more on periodic points on flat surfaces of higher genera, see [28]. Composition of $X \rightarrow Y$ with the multiplication on Y ensures that the composition map is ramified over the origin only. These flat surfaces are called *square-tiled surfaces*, sometimes also *origamis*. These surfaces are studied in Chapter 13 of this Handbook by Herrlich and Schmithüsen [10]. By [8] the affine group of a square-tiled surface is a subgroup of finite index in $\mathrm{SL}_2(\mathbb{Z})$. We mention two results indicating that many types of subgroups of $\mathrm{SL}_2(\mathbb{Z})$ arise as affine group.

Theorem 5.3 ([13], [32]). *With the exception of the covering consisting of three squares, the affine groups of square-tiled surfaces in ΩM_2 are non-congruence subgroups. In any genus $g \geq 2$ there are square-tiled surfaces, whose affine group is a non-congruence subgroup.*

Theorem 5.4 ([32]). *All congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ with possibly 5 exceptions occur as affine groups of square-tiled surfaces.*

Question 5.5. Is there a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ that is *not* the affine group of a square-tiled surface?

6 More on large affine groups: infinitely generated

There exist two constructions for infinitely generated affine groups. McMullen's construction ([23]) gives a complete description in genus two but the techniques apply to genus two only. On the other hand, the construction of Hubert and Schmidt ([14]) is a way to construct a flat surface with infinitely generated affine group starting from a Veech surface with a special point. The resulting surfaces have genus at least four.

We sketch both constructions and conclude with a number of open questions concerning the precise structure of the limit of these infinitely generated affine groups.

Theorem 6.1 ([23]). *Suppose that $(X, \omega) \in \Omega M_2(1, 1)$ has a hyperbolic element in its affine group, but (X, ω) is neither in the $\mathrm{GL}_2^+(\mathbb{R})$ -orbit of the regular decagon nor obtained as a covering of the torus. Then $\mathrm{SL}(X, \omega)$ is infinitely generated.*

Sketch of proof. Veech surfaces in $\Omega M_2(1, 1)$ are either in the orbit of the decagon or torus coverings ([28], [24]). Hence it suffices to show that once $\mathrm{SL}(X, \omega)$ contains a hyperbolic element the limit set of $\mathrm{SL}(X, \omega)$ is the whole S^1 . For that purpose it is enough to show that each direction s joining a zero and a Weierstraß point decomposes the surface into cylinders of commensurable moduli, since then the affine group contains a parabolic element in such a direction and since those directions are dense in S^1 .

In order to prove this, one first shows that the presence of the hyperbolic element implies that the SAF-invariant of the induced interval exchange transformation (IET) on a transverse interval to s vanishes. (The Galois flux used in [23] is a quantity equivalent to the SAF-invariant.) In genus two, due to the bad approximation of quadratic irrationals, this implies that the IET is not minimal. Topological considerations using the Weierstraß point imply that the direction s decomposes into cylinders. Using the presence of the hyperbolic element again, one checks that the moduli of the cylinders have to be commensurable. \square

Theorem 6.2 ([14]). *For $g \geq 4$ there exist flat surfaces (X, ω) whose affine group is infinitely generated.*

More precisely, take any of the Veech surfaces in genus two with trace field $K \neq \mathbb{Q}$ (see Theorem 5.2) and normalize it by $\mathrm{GL}_2^+(\mathbb{R})$ to have periods in $K(i)$. Then a covering ramified over a Weierstraß point and a non-Weierstraß point with coordinates in $K[i]$ has infinitely generated Veech group.

Proof. Recall the definition of a periodic point from Section 5. In order to ensure that $\mathrm{SL}(X, \omega)$ is infinitely generated, the branch points must not be exclusively periodic points on the one hand and not too general either for $\mathrm{SL}(X, \omega)$ might become trivial then. A *connection point* P on Y has the property that every straight line emanating from a singularity of Y and passing through P ends in a singularity, i.e. yields a saddle connection.

Suppose that (Y, η) admits a non-periodic connection point P . The subgroup $SL(P) \subset SL(X, \omega)$ that fixes P is not of finite index, since P is not periodic. On the other hand, for each direction of a geodesic from a singularity to P there is a parabolic element σ in $SL(X, \omega)$ by Veech dichotomy and the definition of a connection point. A suitable power of σ fixes all saddle connections, hence lies in $SL(P)$.

Since the set of directions joining P to a singularity is dense in S^1 , there is a dense set of directions in S^1 fixed by some parabolic element in $SL(P)$. Said differently, the limit set of $SL(P)$ is S^1 . Consequently, $SL(P)$ is infinitely generated.

It thus suffices to find a Veech surface with a non-periodic connection point. The periodic points of the Veech surfaces from Theorem 5.2 are precisely the Weierstraß points by [28]. A Veech surface normalized as in the second statement of the theorem is said to have *strong holonomy type* if the set of periodic directions is precisely $\mathbb{P}^1(K)$. It is straightforward to check that strong holonomy type implies that points with coordinates in $K[i]$ are connection points. Finally, [23] Theorem A.1 implies that all the Veech surfaces in question are of strong holonomy type. \square

We remark that the abundance of surfaces of strong holonomy type is a particular property of genus $g = 2$, too.

Some more results on the structure of infinitely generated affine groups are known ([15]). For example $\mathbb{H}/SL(X, \omega)$ has infinitely many cusps and infinitely many infinite ends. Yet, many questions concerning these infinitely generated groups both for the case of [14] and [23] remain open, in particular the convergence behavior of the associated Poincaré series.

7 The size of the affine group compared to the size of the orbit closure

In Section 5 we have encountered Veech surfaces. Their affine group is, by definition, large, and the $GL_2^+(\mathbb{R})$ -orbit is closed in ΩM_g , it projects to a Teichmüller curve. In genus two, actually the motto 'the larger the affine group the smaller the orbit closure' holds.

Theorem 7.1 ([26]). *Suppose that $g(X) = 2$ and that $SL(X, \omega)$ contains a hyperbolic element. Then the closure of the $GL_2^+(\mathbb{R})$ -orbit of $SL(X, \omega)$ projects to an orbifold of dimension one or two in M_2 .*

In fact, this projection is a Teichmüller curve if $(X, \omega) \in \Omega M_2(2)$. If $(X, \omega) \in \Omega M_2(1, 1)$, then the projection is a Teichmüller curve or it is the preimage of a Hilbert modular surface in the moduli space of abelian surfaces under the Torelli map.

In genus three this motto no longer holds:

Theorem 7.2 ([16]). *The affine group of the Arnoux–Yoccoz surface contains a hyperbolic element and nevertheless the orbit closure is as big as possible, namely the hyperelliptic locus \mathcal{L} in the connected component of $\Omega M_3(2, 2)$ which does not consist entirely of hyperelliptic surfaces.*

There is a $\mathrm{GL}_2^+(\mathbb{R})$ -equivariant map from \mathcal{L} to the stratum $Q(1, 1, 1, 1)$ of quadratic differentials in genus two with four simple zeros. Consequently, the above statement can be rephrased as follows: There is a surface in $Q(1, 1, 1, 1)$, whose affine group contains a hyperbolic element and whose $\mathrm{GL}_2^+(\mathbb{R})$ -orbit closure is the whole stratum $Q(1, 1, 1, 1)$.

We now explain the idea of proof of both theorems. The starting point is to reduce the orbit closure question for flat surfaces to a question in a homogeneous space, where Ratner’s theorem predicts how orbit closures look like. For that purpose, one needs to cut the surface along saddle connections in some fixed direction into tori and cylinders. In order to be able to do so in a neighborhood of the surface, too, the slitting configuration has to be stable under small deformation. This means that the saddle connections have to be homologous.

Such sets of homologous saddle connections are rather rare, but in genus two each surface admits such a set ([26]) and in the locus \mathcal{L} the generic surface does. In \mathcal{L} , the horizontal saddle connections in Figure 1 split the surface into two tori and two cylinders. Not all surfaces in \mathcal{L} admit such a $2T2C$ -direction, but the Arnoux–Yoccoz surface does.

Cut the surface in pieces along the homologous saddle connections. The difference between genus two and genus three becomes apparent in the application of Ratner’s theorem to the splitting pieces. In genus two, if $\mathrm{SL}(X, \omega)$ contains a hyperbolic element, then the two splitting pieces are isogenous tori and the orbit closure is a (‘small’) unipotent subgroup of $(\mathrm{GL}_2^+(\mathbb{R}))^2$. The major remaining step to complete the proof of Theorem 7.1 consists in showing that flat surfaces that split into isogenous tori have Jacobians with real multiplication. Consequently, compare to the proof of Theorem 5.2, the $\mathrm{GL}_2^+(\mathbb{R})$ -orbit closure is contained in the preimage of a Hilbert modular surface

In the case of the Arnoux–Yoccoz surface however, the splitting pieces are ‘as incommensurable as possible’ despite the presence of a hyperbolic element in $\mathrm{GL}_2^+(\mathbb{R})$. Consequently, an application of Ratner’s theorem yields a large orbit closure and a second application in a different $2T2C$ -direction implies that the orbit closure is the whole locus \mathcal{L} .

As a first, and maybe important, step towards extending a Ratner type theorem from genus two to genus three we are thus led to ask:

Question 7.3. How can one describe the locus of flat surfaces in \mathcal{L} that admit a $2T2C$ -direction?

References

- [1] P. Arnoux, Un invariant pour les échanges d’intervalles et les flots sur les surfaces. Thèse, Université de Reims, 1981. [379](#)
- [2] P. Arnoux, Un exemple de semi-conjugaison entre un échange d’intervalles et une translation sur le tore. *Bull. Soc. Math. France* 116 (1988), 489–500. [377](#)
- [3] P. Arnoux, J.-C. Yoccoz, Construction de difféomorphismes pseudo-Anosov. *C. R. Acad. Sci.* 292 (1981), 75–78. [377](#), [379](#)
- [4] M. Bainbridge, Euler characteristics of Teichmüller curves in genus two. *Geom. Topol.* 11 (2007), 1887–2073. [382](#)
- [5] I. Bouw, M. Möller, Teichmüller curves, triangle groups, and Lyapunov exponents. Preprint, 2005, *Ann. of Math.*, to appear. [380](#), [381](#)
- [6] K. Calta, Veech surfaces and complete periodicity in genus two. *J. Amer. Math. Soc.* 17 (2004), 871–908. [380](#)
- [7] A. Fathi, F. Laudenbach, V. Poénaru (eds.), *Travaux de Thurston sur les surfaces. Astérisque* 66–67 (1979). [371](#)
- [8] E. Gutkin, C. Judge, Affine mappings of translation surfaces. *Duke Math. J.* 103 (2000), 191–212. [382](#)
- [9] W. Harvey, Teichmüller spaces, triangle groups and Grothendieck’s dessins. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 249–292. [374](#)
- [10] F. Herrlich, G. Schmithüsen, Dessins d’enfants and Origami curves. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 767–809. [382](#)
- [11] P. Hubert, E. Lanneau, Veech groups without parabolic elements. *Duke Math. J.* 133 (2006), 335–346. [374](#), [376](#), [378](#)
- [12] J. Hubbard, H. Masur, Quadratic differentials and foliations. *Acta Math.* 142 (1979), 221–274.
- [13] P. Hubert, S. Lelièvre, Prime arithmetic Teichmüller discs in $H(2)$. *Israel J. Math.* 151 (2006), 281–321. [382](#)
- [14] P. Hubert, T. Schmidt, T., Infinitely generated Veech groups. *Duke Math. J.* 123 (2004), 49–69. [383](#), [384](#)
- [15] P. Hubert, T. Schmidt, Geometry of infinitely generated Veech groups. *Conform. Geom. Dyn.* 10 (2006), 1–20. [384](#)
- [16] P. Hubert, E. Lanneau, M. Möller, The Arnoux-Yoccoz Teichmüller disc. Preprint, 2006. [378](#), [385](#)
- [17] R. Kenyon, J. Smillie, Billiards on rational-angled triangles. *Comment. Math. Helv.* 75 (2000), 65–108. [375](#)
- [18] M. Kontsevich, A. Zorich, Connected components of the moduli space of abelian differentials with prescribed singularities. *Invent. Math.* 153 (3) (2003), 631–678. [370](#)
- [19] E. Lanneau, Hyperelliptic components of the moduli spaces of quadratic differentials with prescribed singularities. *Comment. Math. Helv.* 79 (2004), 471–501. [372](#), [373](#)
- [20] C. Leininger, On groups generated by two positive multitwists: Teichmüller curves and Lehmer’s number. *Geom. Topol.* 8 (2004), 1301–1359. [374](#), [377](#)

- [21] H. Masur, Interval exchange transformations and measured foliations. *Ann. of Math.* 115 (1982), 169–200. [377](#)
- [22] C. McMullen, Billiards and Teichmüller curves on Hilbert modular surfaces. *J. Amer. Math. Soc.* 16 (2003), 857–885. [380](#), [382](#)
- [23] C. McMullen, Teichmüller geodesics of infinite complexity. *Acta Math.* 191 (2003), mbox191–223. [375](#), [383](#), [384](#)
- [24] C. McMullen, Teichmüller curves in genus two: Torsion divisors and ratios of sines. *Invent. Math.* 165 (2006), 651–672. [383](#)
- [25] C. McMullen, Prym varieties and Teichmüller curves. *Duke Math. J.* 133 (2006), 569–590. [374](#)
- [26] C. McMullen, C., Dynamics of $SL_2(\mathbb{R})$ over moduli space in genus two. *Ann. of Math.* 165 (2007), 397–456. [384](#), [385](#)
- [27] M. Möller, Variations of Hodge structures of Teichmüller curves. *J. Amer. Math. Soc.* 19 (2006), 327–344. [380](#)
- [28] M. Möller, Periodic points on Veech surfaces and the Mordell–Weil group over a Teichmüller curve. *Invent. Math.* 165 (2006), 633–649. [382](#), [383](#), [384](#)
- [29] M. Möller, Linear manifolds in the moduli space of one-forms. *Duke Math. J.* 144 (2008), 447–487. [372](#)
- [30] A. Papadopoulos, Difféomorphismes pseudo-Anosov et automorphismes symplectiques de l’homologie. *Ann. Sci. École Norm. Sup.* 15 (1982), 543–546.
- [31] R. C. Penner, A construction of pseudo-Anosov homeomorphisms. *Trans. Amer. Math. Soc.* 310 (1988), 179–197. [374](#)
- [32] G. Schmithüsen, Veech groups of origamis. Dissertation, Karlsruhe, 2005. [382](#)
- [33] J. Smillie, B. Weiss, Veech’s dichotomy and the lattice property. *Ergodic Theory Dynam. Systems* 28 (2008), 1959–1972. [380](#)
- [34] W. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Am. Math. Soc.* 19 (1988), 417–431. [374](#), [375](#)
- [35] S. Vasilyev, Genus two Veech surfaces arising from general quadratic differentials. Preprint, 2005.
- [36] W. Veech, Gauss measures on the space of interval exchange maps. *Ann. of Math.* 115 (1982), 201–242. [377](#), [378](#)
- [37] W. Veech, The Teichmüller geodesic flow. *Ann. of Math.* 124 (1986), 441–530. [372](#)
- [38] W. Veech, Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. Math.* 97 (1989), 533–583. [370](#), [374](#), [379](#)
- [39] J. C. Yoccoz, Continued fraction algorithm for interval exchange maps: an introduction. In *Frontiers in geometry, number theory and physics* (Les Houches, 2003), Springer-Verlag, Berlin 2006, 401–435. [378](#)

Chapter 11

Braid groups and Artin groups

Luis Paris

Contents

1	Introduction	389
2	Braid groups	392
	2.1 Braids	392
	2.2 Configuration spaces	395
	2.3 Mapping class groups	398
	2.4 Automorphisms of free groups	401
3	Artin groups	403
	3.1 Definitions and examples	403
	3.2 Coxeter groups	407
	3.3 Artin monoids	410
	3.4 Artin groups	411
4	Garside groups	412
	4.1 Garside monoids	412
	4.2 Reversing processes and presentations	415
	4.3 Normal forms and automatic structures	417
	4.4 The conjugacy problem	419
5	Cohomology and the Salvetti complex	421
	5.1 Cohomology	421
	5.2 Salvetti complex	424
6	Linear representations	430
	6.1 Algebraic approach	430
	6.2 Topological approach	432
7	Geometric representations	434
	7.1 Definitions and examples	434
	7.2 Presentations	438
	7.3 Classification	441
	References	444

1 Introduction

Braids go back to several centuries and were universally used for ornamental purposes or even practical ones, for example in the fashioning of ropes. Today, they are described

by means of abstract models which occur in the “theory of braids”. The theory of braids studies the concept of braids (such as we imagine them) as well as various generalizations arising from various branches of the mathematics. The idea is that braids form a group. The number of strands must be fixed so that the operation is well-defined. So, we have a braid group on two strands, a braid group on three strands, and so on. The braid group on one strand is trivial because a string cannot be braided (although it can be knotted).

We generally make the mathematical study of braids go back to an article of Emile Artin [7] dating from 1925, in which is described the notion of braids under various aspects, one being that obvious, like a “series of tended and interlaced strings”, and others more conceptual but equally deep, such as a presentation by generators and relations, or a presentation as the mapping class group of a punctured disk.

Since the 1930s, strong links between braids and links (and knots) were established by people such as Alexander and Markov (see [19]). This link is at the origin in the 1980s of a deep revival in the theory of knots with the work of Jones and his invariant defined from the theory of braids (see [101], [102], [83], and [138]).

Later, interesting relations with algebraic geometry and the theory of finite groups generated by reflections were established, in particular by Arnol’d [5], [3], [6] and Brieskorn [29], [31]. These relations become particularly interesting when we extend the notion of braid group to that of Artin group of spherical type, also called generalized braid groups. Although Artin groups were introduced by Tits [146] as extensions of Coxeter groups, their study really began in the seventies with the works of Brieskorn [30], [31], Saito [32] and Deligne [71], where different aspects of these groups are studied, such as their combinatorics, as well as their link with hyperplane arrangements and singularities.

Some problems in group theory, often very close to algorithmics, such as the word and conjugacy problems, have experienced a renewal of interest not only through their applications in other domains, but also because the notion of mathematical proof is changing. Indeed, we distinguish now the notion of demonstration from the notion of effective proof, the one which builds up the solution. Such a proof gives rise to an algorithm, and its complexity (calculation time) is of importance. Algorithmics in the theory of braid groups is especially active. Decision problems such as the conjugacy problem were solved by Garside [85] in 1969 with methods which are now the source of numerous works on braid groups. In [70] is introduced a more formal and more general framework to study algorithmic problems on braid groups: the Garside groups. The idea is to isolate certain combinatorial properties of braid groups, in particular those emphasized by Garside [85]. It is a less restrictive model which uses tools from language theory (monoids, rewriting systems) and combinatorics (ordered sets), tools that are especially adapted to treat algorithmic problems. Today, the major part of the algorithmic problems on braid groups are studied within the framework of Garside groups. Let us also indicate that the Artin groups of spherical type are Garside groups.

This survey is written from these viewpoints but also maintaining two other objectives: (1) to make a survey understandable by non-specialists; (2) to make as often as possible the link with mapping class groups.

The second section is about the “classical” theory of braid groups. Various aspects as well as some of their properties are presented. The third section is an introduction to Artin groups, and the fourth is an introduction to Garside groups. There, the reader will find algorithms to solve some decision problems such as the conjugacy one for braid groups (and Garside groups).

The fifth section is about the cohomology of Artin groups, although the exposition goes beyond by explaining the Salvetti complexes. These are tools originally from the theory of hyperplane arrangements that turn out to be useful in the context of braid groups.

The sixth section is about the linear representations of braid groups studied by Bigelow [17] and Krammer [105], [106], as well as about its various generalizations (to Artin groups). The algebraic aspect and the topological aspect of these representations are both explained. Other linear representations of braid groups have been studied and are also interesting but, for lack of space and for reason of coherence, these will not be treated in this text. We refer to [21] for a survey on the other linear representations.

The seventh section is about geometric representations of Artin groups. (By a geometric representation we simply mean a homomorphism in a mapping class group.) This subject is less popular than the previous ones but I strongly believe in its future. In particular, Subsection 7.3, where are explained the results of Castel [40], shows all the power of such a study.

Finally, I would like to indicate two aspects of braid groups which are not in this survey and which “should be in any survey on the braid groups”.

The first aspect is the link between braids and links and knots. This is very important in the theory but amply explained in all books and almost all surveys on the subject. So, I voluntarily ignore this aspect in order to be able to treat in more detail the other ones. The reader will find in [19], [93], [128], [103] detailed expositions on this aspect and on braid groups in general.

I would have wanted to make an eighth section to explain the second aspect: the orders in braid groups. But, unfortunately, this chapter is long enough and there is no more room for another section. Inspired by problems of set theory, Dehornoy [65] founded an explicit construction of a total ordering invariant by left multiplication in the braid group. The fact that the braid group is orderable may be not completely new, in the sense that it results from Nielsen theory [129], but Dehornoy’s ordering is interesting in itself. In my opinion, it is an important tool to understand the braid groups, and I augur numerous developments in this direction. Artin groups of type B_n and A_n embed into braid groups (see Section 3) thus they are also orderable. Artin groups of type D_n embed into mapping class groups of surfaces with boundary (see Section 7), and, by [139], such a group is orderable. We do not know whether the other Artin groups are orderable or not. We encourage the reader to consult [68] for a detailed discussion on this subject.

2 Braid groups

2.1 Braids

Let $n \geq 1$ be an integer, and let P_1, \dots, P_n be n distinct points in the plane \mathbb{R}^2 (except for mention of the contrary, we will always assume $P_k = (k, 0)$ for all $1 \leq k \leq n$). Define a *braid on n strands* to be an n -tuple $\beta = (b_1, \dots, b_n)$ of paths, $b_k: [0, 1] \rightarrow \mathbb{R}^2$, such that

- $b_k(0) = P_k$ for all $1 \leq k \leq n$;
- there exists a permutation $\chi = \theta(\beta) \in \text{Sym}_n$ such that $b_k(1) = P_{\chi(k)}$ for all $1 \leq k \leq n$;
- $b_k(t) \neq b_l(t)$ for all $k \neq l$ and all $t \in [0, 1]$.

Two braids α and β are said to be *homotopic* if there exists a continuous family $\{\gamma_s\}_{s \in [0, 1]}$ of braids such that $\gamma_0 = \alpha$ and $\gamma_1 = \beta$. Note that $\theta(\alpha) = \theta(\beta)$ if α and β are homotopic.

We represent graphically a homotopy class of braids as follows. For $1 \leq k \leq n$, let I_k be a copy of the interval $[0, 1]$. Take a braid $\beta = (b_1, \dots, b_n)$ and define the *geometric braid*

$$\beta^g: I_1 \sqcup \dots \sqcup I_n \rightarrow \mathbb{R} \times [0, 1]$$

by $\beta^g(t) = (b_k(t), t)$ for all $t \in I_k$ and all $1 \leq k \leq n$. Let $\text{proj}: \mathbb{R}^2 \times [0, 1] \rightarrow \mathbb{R} \times [0, 1]$ be the projection defined by

$$\text{proj}(x, y, t) = (x, t).$$

Up to homotopy, we can assume that $\text{proj} \circ \beta^g$ is a smooth immersion with only transversal double points that we call *crossings*. In each crossing we indicate graphically like in Figure 1 which strand goes over the other. Such a representation of β is called a *braid diagram* of β . An example is illustrated in Figure 2.

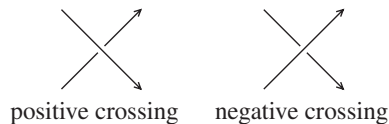


Figure 1. Crossings in a braid diagram.

The *product* of two braids $\alpha = (a_1, \dots, a_n)$ and $\beta = (b_1, \dots, b_n)$ is defined to be the braid

$$\alpha \cdot \beta = (a_1 b_{\chi(1)}, \dots, a_n b_{\chi(n)}),$$

where $\chi = \theta(\alpha)$. An example is illustrated in Figure 3.

Let \mathcal{B}_n denote the set of homotopy classes of braids on n strands. It is easily seen that the above defined multiplication of braids induces an operation on \mathcal{B}_n . Moreover, we have the following.

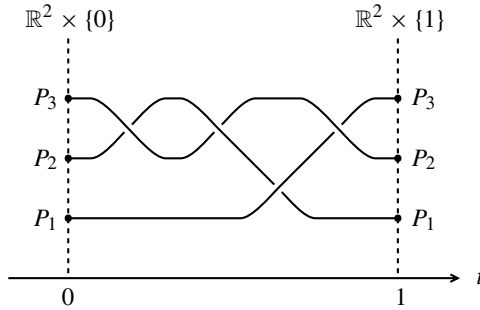


Figure 2. A braid diagram.

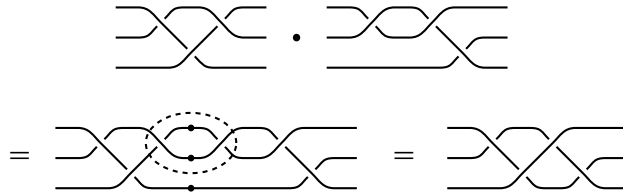


Figure 3. Product of two braids.

Proposition 2.1. *The set \mathcal{B}_n endowed with this operation is a group.*

From now on, except for mention of the contrary, by a braid we will mean a homotopy class of braids. The group \mathcal{B}_n of Proposition 2.1 is called the *braid group on n strands*. The identity is the *constant braid* $\text{Id} = (\text{Id}_1, \dots, \text{Id}_n)$, where, for $1 \leq k \leq n$, Id_k denotes the constant path on P_k . The inverse of a braid β is its *mirror* as illustrated in Figure 4.

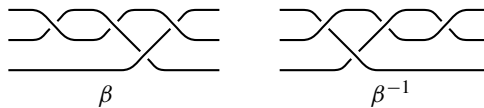


Figure 4. Inverse of a braid.

Recall that if two braids α, α' are homotopic, then $\theta(\alpha) = \theta(\alpha')$. Hence, the map θ from the set of braids on n strands to Sym_n induces a map $\theta: \mathcal{B}_n \rightarrow \text{Sym}_n$. It is easily checked that this map is an epimorphism. Its kernel is called the *pure braid group on n strands* and is denoted by $\mathcal{P}\mathcal{B}_n$. It plays an important role in the theory.

Let σ_k be the braid illustrated in Figure 5. One can easily verify that $\sigma_1, \dots, \sigma_{n-1}$ generate the braid group \mathcal{B}_n and satisfy the relations

$$\begin{aligned} \sigma_k \sigma_l &= \sigma_l \sigma_k & \text{if } |k - l| \geq 2, \\ \sigma_k \sigma_l \sigma_k &= \sigma_l \sigma_k \sigma_l & \text{if } |k - l| = 1. \end{aligned}$$

(See Figure 6.) These relations suffice to define the braid group, namely:

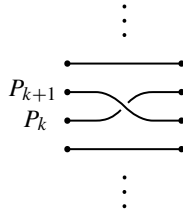


Figure 5. The braid σ_k .

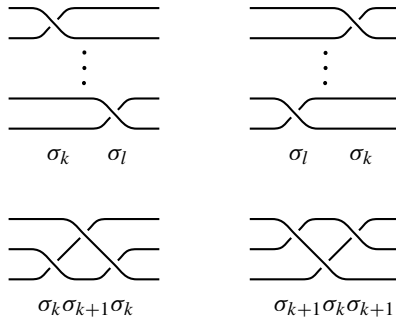


Figure 6. Relations in \mathcal{B}_n .

Theorem 2.2 (Artin [7], [8], Magnus [118]). *The group \mathcal{B}_n has a presentation with generators $\sigma_1, \dots, \sigma_{n-1}$ and relations*

$$\begin{aligned} \sigma_k \sigma_l &= \sigma_l \sigma_k && \text{if } |k - l| \geq 2, \\ \sigma_k \sigma_l \sigma_k &= \sigma_l \sigma_k \sigma_l && \text{if } |k - l| = 1. \end{aligned}$$

Theorem 2.3 (Burau [35], Markov [123]). *For $1 \leq k < l \leq n$, let*

$$\delta_{kl} = \sigma_{l-1} \dots \sigma_{k+1} \sigma_k^2 \sigma_{k+1}^{-1} \dots \sigma_{l-1}^{-1}.$$

Then the pure braid group $\mathcal{P}\mathcal{B}_n$ has a presentation with generators

$$\delta_{kl}, \quad 1 \leq k < l \leq n,$$

and relations

$$\begin{aligned} \delta_{rs} \delta_{kl} \delta_{rs}^{-1} &= \delta_{kl} && \text{if } 1 \leq r < s < k < l \leq n, \\ &&& \text{or } 1 \leq k < r < s < l \leq n, \\ \delta_{rk} \delta_{kl} \delta_{rk}^{-1} &= \delta_{kl}^{-1} \delta_{rl}^{-1} \delta_{kl} \delta_r l \delta_{kl} && \text{if } 1 \leq r < k < l \leq n, \end{aligned}$$

$$\begin{aligned} \delta_{rk}\delta_{rl}\delta_{rk}^{-1} &= \delta_{kl}^{-1}\delta_{rl}\delta_{kl} && \text{if } 1 \leq r < k < l \leq n, \\ \delta_{rs}\delta_{kl}\delta_{rs}^{-1} &= \delta_{sl}^{-1}\delta_{rl}^{-1}\delta_{sl}\delta_{rl}\delta_{kl}\delta_{rl}^{-1}\delta_{sl}^{-1}\delta_{rl}\delta_{sl} && \text{if } 1 \leq r < k < s < l \leq n. \end{aligned}$$

Note. Most of the proofs of Theorems 2.2 and 2.3 that can be found in the literature proceed as follows. Given an exact sequence

$$1 \rightarrow K \longrightarrow G \longrightarrow H \rightarrow 1,$$

there is a machinery to compute a presentation of G from presentations of K and H . We start with the observation that $\mathcal{P}\mathcal{B}_2 \simeq \mathbb{Z}$ and with the exact sequence

$$1 \rightarrow F_n \longrightarrow \mathcal{P}\mathcal{B}_{n+1} \longrightarrow \mathcal{P}\mathcal{B}_n \rightarrow 1, \tag{2.1}$$

where F_n is a free group of rank n , to prove Theorem 2.3 by induction on n . (The exact sequence (2.1) will be explained in Subsection 2.2.) Then we use the exact sequence

$$1 \rightarrow \mathcal{P}\mathcal{B}_n \longrightarrow \mathcal{B}_n \longrightarrow \text{Sym}_n \rightarrow 1$$

to prove Theorem 2.2 from Theorem 2.3. Another proof which, as far as I know, is not in the literature but is known to experts, consists in extracting the presentation of Theorem 2.2 from the Salvetti complex of \mathcal{B}_n . This is a cellular complex which is a $K(\mathcal{B}_n, 1)$ (see Section 5).

2.2 Configuration spaces

We identify \mathbb{R}^2 with \mathbb{C} and P_k with $k \in \mathbb{C}$ for all $1 \leq k \leq n$. For $1 \leq k < l \leq n$ we denote by H_{kl} the linear hyperplane of \mathbb{C}^n defined by the equation $z_k = z_l$. The *big diagonal* of \mathbb{C}^n is defined to be

$$\text{Diag}_n = \bigcup_{1 \leq k < l \leq n} H_{kl}.$$

The *space of ordered configurations of n points* in \mathbb{C} is defined to be

$$M_n = \mathbb{C}^n \setminus \text{Diag}_n.$$

This is the space of n -tuples $\mathbf{z} = (z_1, \dots, z_n)$ of complex numbers such that $z_k \neq z_l$ for $k \neq l$. The symmetric group Sym_n acts freely on M_n . The quotient

$$N_n = M_n / \text{Sym}_n$$

is called the *space of configurations of n points* in \mathbb{C} . This is the space of unordered n -tuples $\mathbf{z} = \{z_1, \dots, z_n\}$ of complex numbers such that $z_k \neq z_l$ for $k \neq l$.

Proposition 2.4. *Let $P_0 = (1, 2, \dots, n) \in M_n$. Then $\pi_1(M_n, P_0) = \mathcal{P}\mathcal{B}_n$.*

Proof. For a pure braid $\beta = (b_1, \dots, b_n)$ we set

$$\begin{aligned} \varphi(\beta): [0, 1] &\rightarrow M_n, \\ t &\mapsto (b_1(t), \dots, b_n(t)). \end{aligned}$$

Clearly, $\varphi(\beta)$ is a loop based at P_0 . Moreover, two pure braids α and α' are homotopic if and only if $\varphi(\alpha)$ and $\varphi(\alpha')$ are homotopic. Thus φ induces a bijection $\varphi_*: \mathcal{P}\mathcal{B}_n \rightarrow \pi_1(M_n, P_0)$ which turns out to be a homomorphism. \square

For $z \in M_n$, we denote by $[z]$ the element of $N_n = M_n/\text{Sym}_n$ represented by z .

Proposition 2.5. $\pi_1(N_n, [P_0]) = \mathcal{B}_n$.

Proof. For a braid $\beta = (b_1, \dots, b_n)$ we set

$$\begin{aligned} \hat{\varphi}(\beta): [0, 1] &\rightarrow N_n \\ t &\mapsto [b_1(t), \dots, b_n(t)]. \end{aligned}$$

Clearly, $\hat{\varphi}(\beta)$ is a loop based at $[P_0]$. It is easily checked that $\hat{\varphi}$ induces a homomorphism $\hat{\varphi}_*: \mathcal{B}_n \rightarrow \pi_1(N_n, [P_0])$, and that the following diagram commutes:

$$\begin{array}{ccccccc} 1 & \longrightarrow & \mathcal{P}\mathcal{B}_n & \longrightarrow & \mathcal{B}_n & \longrightarrow & \text{Sym}_n \longrightarrow 1 \\ & & \downarrow \simeq \varphi_* & & \downarrow \hat{\varphi}_* & & \downarrow \text{Id} \\ 1 & \longrightarrow & \pi_1(M_n, P_0) & \longrightarrow & \pi_1(N_n, [P_0]) & \longrightarrow & \text{Sym}_n \longrightarrow 1. \end{array}$$

The first row is exact by definition, and the second one is associated to the regular covering $M_n \rightarrow N_n = M_n/\text{Sym}_n$, so it is exact, too. We conclude by the five lemma that $\hat{\varphi}_*$ is an isomorphism. \square

Let $f, g \in \mathbb{C}[x]$ be two non-constant polynomials. Set

$$\begin{aligned} f &= a_0x^m + a_1x^{m-1} + \dots + a_m, & a_0 &\neq 0, \\ g &= b_0x^n + b_1x^{n-1} + \dots + b_n, & b_0 &\neq 0. \end{aligned}$$

The *Sylvester matrix* of f and g is defined to be

$$\text{Sylv}(f, g) = \begin{pmatrix} a_0 & 0 & \dots & 0 & b_0 & 0 & \dots & 0 \\ a_1 & a_0 & \ddots & \vdots & b_1 & b_0 & \ddots & \vdots \\ \vdots & a_1 & \ddots & 0 & \vdots & b_1 & \ddots & 0 \\ a_m & \vdots & \ddots & a_0 & b_n & \vdots & \ddots & b_0 \\ 0 & a_m & & a_1 & 0 & b_n & & b_1 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_m & 0 & \dots & 0 & b_n \end{pmatrix}$$

$\underbrace{\hspace{15em}}_{n \text{ columns}}$
 $\underbrace{\hspace{15em}}_{m \text{ columns}}$

The *resultant* of f and g is defined to be

$$\text{Res}(f, g) = \det(\text{Sylv}(f, g)).$$

The following is classical in algebraic geometry (see [54], for example).

Theorem 2.6. *Let $f, g \in \mathbb{C}[x]$ be two non-constant polynomials. Then f and g have a common root if and only if $\text{Res}(f, g) = 0$.*

Corollary 2.7. *Let $f \in \mathbb{C}[x]$ be a polynomial of degree $d \geq 2$. Then f has a multiple root if and only if $\text{Res}(f, f') = 0$.*

The number $\text{Res}(f, f')$ is called the *discriminant* of f and is denoted by $\text{Disc}(f)$. For instance, if $f = ax^2 + bx + c$, then $\text{Disc}(f) = b^2 - 4ac$.

Let $n \geq 2$ and let $\mathbb{C}_n[x]$ be the set of monic polynomials of degree n . In particular, $\mathbb{C}_n[x]$ is isomorphic to \mathbb{C}^n . The map $\text{Disc}: \mathbb{C}_n[x] \rightarrow \mathbb{C}$ is clearly a polynomial function, thus

$$\mathcal{D} = \{f \in \mathbb{C}_n[x]; f \text{ has a multiple root}\} = \{f \in \mathbb{C}_n[x]; \text{Disc}(f) = 0\}$$

is an algebraic hypersurface called the *n -th discriminant*. It is related to the braid group by the following.

Proposition 2.8. $N_n = \mathbb{C}_n[x] \setminus \mathcal{D}$.

Proof. Let $\Phi: M_n \rightarrow \mathbb{C}_n[x] \setminus \mathcal{D}$ be the map defined by

$$\Phi(z_1, \dots, z_n) = (x - z_1) \dots (x - z_n).$$

Then Φ is surjective and we have $\Phi(\mathbf{u}) = \Phi(\mathbf{v})$ if and only if there exists $\chi \in \text{Sym}_n$ such that $\mathbf{v} = \chi(\mathbf{u})$. Thus $\mathbb{C}_n[x] \setminus \mathcal{D} \simeq M_n / \text{Sym}_n = N_n$. \square

Now, recall the homotopy long exact sequence of a fiber bundle (see [96], for example).

Theorem 2.9. *Let $p: M \rightarrow B$ be a locally trivial fiber bundle. Let $b_0 \in B$, let $F = p^{-1}(b_0)$, and let $P_0 \in F$. Assume that F is connected. Then there is a long exact sequence of homotopy groups*

$$\begin{aligned} \cdots \rightarrow \pi_{k+1}(B, b_0) \rightarrow \pi_k(F, P_0) \rightarrow \pi_k(M, P_0) \rightarrow \pi_k(B, b_0) \rightarrow \cdots \\ \cdots \rightarrow \pi_2(B, b_0) \rightarrow \pi_1(F, P_0) \rightarrow \pi_1(M, P_0) \rightarrow \pi_1(B, b_0) \rightarrow 1. \end{aligned}$$

There are two cases where this long exact sequence becomes a short exact sequence: when $\pi_2(B, b_0) = \{0\}$, and when p admits a cross-section $\kappa: B \rightarrow M$. In the latter case the short exact sequence splits. It turns out that both situations hold in the study of M_n .

Theorem 2.10 (Fadell, Neuwirth [80]). *Let $p: M_{n+1} \rightarrow M_n$ be defined by*

$$p(z_1, \dots, z_n, z_{n+1}) = (z_1, \dots, z_n).$$

Then p is a locally trivial fiber bundle which admits a cross-section $\kappa: M_n \rightarrow M_{n+1}$.

Let $b_0 = (1, 2, \dots, n)$. Then the fiber $p^{-1}(b_0)$ is naturally homeomorphic to $\mathbb{C} \setminus \{1, 2, \dots, n\}$ whose fundamental group is the free group F_n of rank n . A cross-section of p is the map $\kappa: M_n \rightarrow M_{n+1}$ defined by

$$\kappa(z_1, \dots, z_n) = (z_1, \dots, z_n, |z_1| + \dots + |z_n| + 1).$$

Corollary 2.11. *Let $n \geq 2$. Then there is a split exact sequence*

$$1 \longrightarrow F_n \longrightarrow \mathcal{P}\mathcal{B}_{n+1} \begin{array}{c} \xrightarrow{P_*} \\ \xleftarrow{\kappa_*} \end{array} \mathcal{P}\mathcal{B}_n \longrightarrow 1.$$

A connected CW-complex X is called $K(\pi, 1)$ if its universal cover is contractible. Equivalently, X is $K(\pi, 1)$ if $\pi_k(X) = \{0\}$ for all $k \geq 2$. In particular, a connected space X is $K(\pi, 1)$ if and only if one of its connected covers Y is $K(\pi, 1)$. The notion of $K(\pi, 1)$ spaces is of importance in the calculation of the (co)homology of groups. We refer to [34] for detailed explanations on the subject.

It is easily seen that $\mathbb{C} \setminus \{1, \dots, n\}$ is $K(\pi, 1)$, thus, from Theorems 2.9 and 2.10 follows:

Corollary 2.12. *The spaces M_n and N_n are $K(\pi, 1)$.*

It is also known that the fundamental group of a finite dimensional $K(\pi, 1)$ space is torsion free (see [34]), thus:

Corollary 2.13. *$\mathcal{B}_n = \pi_1(N_n)$ is torsion free.*

2.3 Mapping class groups

Let Σ be an oriented compact surface, possibly with boundary. Let $\mathcal{P} = \{P_1, \dots, P_n\}$ be a collection of n punctures in the interior of Σ . Let $\text{Homeo}^+(\Sigma, \mathcal{P})$ denote the group of homeomorphisms $h: \Sigma \rightarrow \Sigma$ which preserve the orientation, which pointwise fix the boundary of Σ , and such that $h(\mathcal{P}) = \mathcal{P}$. We assume $\text{Homeo}^+(\Sigma, \mathcal{P})$ endowed with the compact-open topology. Let $\text{Homeo}_0^+(\Sigma, \mathcal{P})$ denote the connected component of the identity in $\text{Homeo}(\Sigma, \mathcal{P})$. The *mapping class group* of the pair (Σ, \mathcal{P}) is defined to be

$$\mathcal{M}(\Sigma, \mathcal{P}) = \pi_0(\text{Homeo}^+(\Sigma, \mathcal{P})) = \text{Homeo}^+(\Sigma, \mathcal{P})/\text{Homeo}_0^+(\Sigma, \mathcal{P}).$$

A *braid* of Σ based at \mathcal{P} is defined to be a n -tuple $\beta = (b_1, \dots, b_n)$ of paths, $b_k: [0, 1] \rightarrow \Sigma$, such that

- $b_k(0) = P_k$ for all $1 \leq k \leq n$;
- there exists a permutation $\chi = \theta(\beta) \in \text{Sym}_n$ such that $b_k(1) = P_{\chi(k)}$ for all $1 \leq k \leq n$;
- $b_k(t) \neq b_l(t)$ for all $k \neq l$ and all $t \in [0, 1]$.

The homotopy classes of braids based at \mathcal{P} form a group denoted by $\mathcal{B}_n(\Sigma, \mathcal{P})$ and called the *braid group of Σ on n strands based at \mathcal{P}* . It does not depend up to isomorphism on the choice of \mathcal{P} but only on the cardinality $n = |\mathcal{P}|$. So, we may often write $\mathcal{B}_n(\Sigma)$ in place of $\mathcal{B}_n(\Sigma, \mathcal{P})$. If $\Sigma = \mathbb{D}$ is a disk, then $\mathcal{B}_n(\Sigma)$ is naturally isomorphic to the braid group \mathcal{B}_n .

For $1 \leq k < l \leq n$, we denote by $H_{kl}(\Sigma)$ the set of n -tuples $\mathbf{x} = (x_1, \dots, x_n) \in \Sigma^n$ such that $x_k = x_l$. The *big diagonal* of Σ^n is defined to be

$$\text{Diag}_n(\Sigma) = \bigcup_{1 \leq k < l \leq n} H_{kl}(\Sigma).$$

The *space of ordered configurations of n points in Σ* is defined to be

$$M_n(\Sigma) = \Sigma^n \setminus \text{Diag}_n(\Sigma).$$

This is the space of n -tuples $\mathbf{x} = (x_1, \dots, x_n)$ in Σ^n such that $x_k \neq x_l$ for all $1 \leq k \neq l \leq n$. The symmetric group Sym_n acts freely on $M_n(\Sigma)$, and the quotient

$$N_n(\Sigma) = M_n(\Sigma)/\text{Sym}_n$$

is called the *space of configurations of n points in Σ* . This is the space of unordered n -tuples $\mathbf{x} = \{x_1, \dots, x_n\}$ of elements of Σ such that $x_k \neq x_l$ for all $1 \leq k \neq l \leq n$.

Set $\mathbf{P}_0 = (P_1, \dots, P_n) \in M_n(\Sigma)$. For $\mathbf{x} \in M_n(\Sigma)$, we denote by $[\mathbf{x}]$ the element of $N_n(\Sigma)$ represented by \mathbf{x} . The following can be proved in the same way as Proposition 2.5.

Proposition 2.14. $\pi_1(N_n(\Sigma), [\mathbf{P}_0]) \simeq \mathcal{B}_n(\Sigma)$.

Now, the surface braid groups and the mapping class groups are related by the following exact sequence.

Theorem 2.15 (Birman [18]). *Suppose Σ is neither a sphere, nor a torus. Then we have the exact sequence*

$$1 \rightarrow \mathcal{B}_n(\Sigma, \mathcal{P}) \rightarrow \mathcal{M}(\Sigma, \mathcal{P}) \rightarrow \mathcal{M}(\Sigma) \rightarrow 1.$$

Note. Let

$$\begin{aligned} \Phi: \text{Homeo}^+(\Sigma) &\rightarrow N_n(\Sigma), \\ \varphi &\mapsto \{\varphi(P_1), \dots, \varphi(P_n)\}. \end{aligned}$$

Then Φ is a locally trivial fiber bundle, and the fiber of Φ over $\mathcal{P} = [\mathbf{P}_0]$ is $\text{Homeo}^+(\Sigma, \mathcal{P})$. Furthermore, it is known that $\pi_1(\text{Homeo}^+(\Sigma)) = \{1\}$ (see [92]).

Thus, by the homotopy long exact sequence of a fiber bundle (see [96]), we have the short exact sequence

$$1 \rightarrow \pi_1(N_n(\Sigma), \mathcal{P}) \longrightarrow \pi_0(\text{Homeo}^+(\Sigma, \mathcal{P})) \longrightarrow \pi_0(\text{Homeo}^+(\Sigma)) \rightarrow 1,$$

which is the same as the exact sequence of Theorem 2.15.

It is known that $\mathcal{M}(\mathbb{D}) = \{1\}$ (see [1]), thus, by Theorem 2.15:

Theorem 2.16 (Artin [7], [8]). *Let $\mathcal{P} = \{P_1, \dots, P_n\}$ be a collection of n punctures in the interior of the disk \mathbb{D} . Then $\mathcal{M}(\mathbb{D}, \mathcal{P}) \simeq \mathcal{B}_n$.*

The isomorphism $\Phi: \mathcal{M}(\mathbb{D}, \mathcal{P}) \rightarrow \mathcal{B}_n$ can be easily described as follows. Let $\varphi \in \text{Homeo}^+(\mathbb{D}, \mathcal{P})$. We know by [1] that $\pi_0(\text{Homeo}^+(\mathbb{D})) = \{1\}$. Thus, there exists a continuous path $\{\varphi_t\}_{t \in [0,1]}$ in $\text{Homeo}^+(\mathbb{D})$ such that $\varphi_0 = \text{Id}$ and $\varphi_1 = \varphi$. Let $\beta = (b_1, \dots, b_n)$ be the braid defined by

$$b_k(t) = \varphi_t(P_k), \quad 1 \leq k \leq n \text{ and } t \in [0, 1].$$

Then $\Phi(\varphi)$ is the homotopy class of β .

The inverse isomorphism $\Phi^{-1}: \mathcal{B}_n \rightarrow \mathcal{M}(\mathbb{D}, \mathcal{P})$ is more complicated to describe, but the images of the standard generators can easily be defined in terms of braid twists as follows.

We come back to the situation where Σ is an oriented compact surface and $\mathcal{P} = \{P_1, \dots, P_n\}$ is a collection of n punctures in the interior of Σ . Let $P_k, P_l \in \mathcal{P}, k \neq l$. An *essential arc* joining P_k to P_l is defined to be an embedding $a: [0, 1] \rightarrow \Sigma$ such that $a(0) = P_k, a(1) = P_l, a((0, 1)) \cap \mathcal{P} = \emptyset$, and $a([0, 1]) \cap \partial \Sigma = \emptyset$. Two essential arcs a and a' are said to be *isotopic* if there is a continuous family $\{a_t\}_{t \in [0,1]}$ of essential arcs such that $a_0 = a$ and $a_1 = a'$. Isotopy of essential arcs is an equivalence relation that we denote by $a \sim a'$.

Let a be an essential arc joining P_k to P_l . Let $\mathbb{D} = \{z \in \mathbb{C}; |z| \leq 1\}$ be the standard disk, and let $A: \mathbb{D} \rightarrow \Sigma$ be an embedding such that

- $a(t) = A(t - \frac{1}{2})$ for all $t \in [0, 1]$;
- $A(\mathbb{D}) \cap \mathcal{P} = \{P_k, P_l\}$.

Let $T \in \text{Homeo}^+(\Sigma, \mathcal{P})$ be defined by

$$(T \circ A)(z) = A(e^{2i\pi|z|}z), \quad z \in \mathbb{D},$$

and T is the identity outside the image of A (see Figure 7). The *braid twist along a* is defined to be the element $\tau_a \in \mathcal{M}(\Sigma, \mathcal{P})$ represented by T , that is, the isotopy class of T . Note that:

- the definition of τ_a does not depend on the choice of $A: \mathbb{D} \rightarrow \Sigma$;
- if a is isotopic to a' , then $\tau_a = \tau_{a'}$.

Now, we view the disk \mathbb{D} as the disk in \mathbb{C} of radius $\frac{n+1}{2}$ centered at $\frac{n+1}{2}$, and we set $P_k = k$ for $1 \leq k \leq n$. Let $a_k: [0, 1] \rightarrow \mathbb{D}$ be the arc defined by

$$a_k(t) = k + t, \quad t \in [0, 1].$$

(See Figure 8.) Then:

Lemma 2.17. *The inverse isomorphism $\Phi^{-1}: \mathcal{B}_n \rightarrow \mathcal{M}(\mathbb{D}, \mathcal{P})$ is defined by*

$$\Phi^{-1}(\sigma_k) = \tau_{a_k}, \quad 1 \leq k \leq n - 1.$$

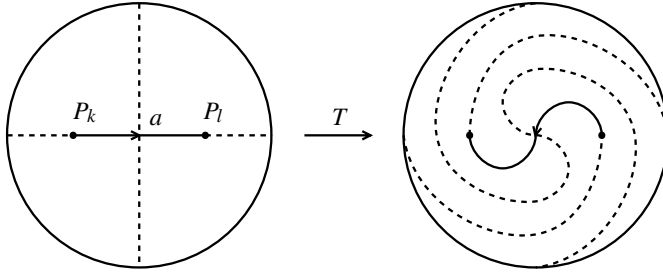


Figure 7. Braid twist.

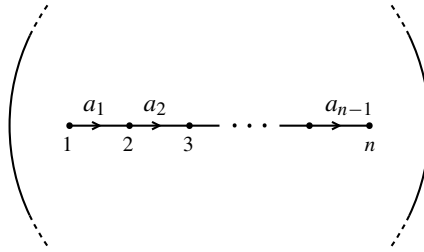


Figure 8. The standard generators of $\mathcal{M}(\mathbb{D}, \mathcal{P}) = \mathcal{B}_n$.

2.4 Automorphisms of free groups

For a group G , we denote by $\text{Aut}(G)$ the group of automorphisms of G , by $\text{Inn}(G)$ the group of inner automorphisms of G , and by $\text{Out}(G) = \text{Aut}(G)/\text{Inn}(G)$ the group of outer automorphisms of G .

Let $F_n = F(x_1, \dots, x_n)$ be the free group of rank n . For $1 \leq k \leq n - 1$, let $\tau_k: F_n \rightarrow F_n$ be the automorphism defined by

$$\tau_k: \begin{cases} x_k \mapsto x_k^{-1} x_{k+1} x_k, \\ x_{k+1} \mapsto x_k, \\ x_l \mapsto x_l \quad \text{if } l \neq k, k + 1. \end{cases}$$

One can easily show the following.

Proposition 2.18. *The mapping $\sigma_k \mapsto \tau_k$, $1 \leq k \leq n - 1$, determines a representation $\rho: \mathcal{B}_n \rightarrow \text{Aut}(F_n)$.*

The above representation $\rho: \mathcal{B}_n \rightarrow \text{Aut}(F_n)$ is called the *Artin representation*. It is faithful, more precisely:

Theorem 2.19 (Artin [7], [8]). (1) *The Artin representation $\rho: \mathcal{B}_n \rightarrow \text{Aut}(F_n)$ is faithful.*

(2) *An automorphism $\alpha \in \text{Aut}(F_n)$ belongs to $\text{Im} \rho$ if and only if $\alpha(x_n \dots x_2 x_1) = x_n \dots x_2 x_1$ and there exists a permutation $\chi \in \text{Sym}_n$ such that $\alpha(x_k)$ is conjugate to $x_{\chi(k)}$ for all $1 \leq k \leq n$.*

In particular, \mathcal{B}_n can be viewed as a subgroup of $\text{Aut}(F_n)$. This has some consequences on \mathcal{B}_n itself such as the two properties defined below.

A group G is called *residually finite* if for all $g \in G \setminus \{1\}$ there exists a homomorphism $\varphi: G \rightarrow H$ such that H is finite and $\varphi(g) \neq 1$. A group G is called *Hopfian* if every epimorphism $\varphi: G \rightarrow G$ is an isomorphism. It is known that the subgroups of $\text{Aut}(F_n)$ are both residually finite and Hopfian (see [119]). Thus, by Theorem 2.19:

Corollary 2.20. *The braid group \mathcal{B}_n is residually finite and Hopfian.*

There are several ways to describe geometrically the Artin representation. The first way uses the Fadell-Neuwirth fiber bundle $p: M_{n+1} \rightarrow M_n$ of Theorem 2.10. Let Sym_n act on M_n and on M_{n+1} . The second action is on the first n coordinates, that is,

$$\chi(z_1, \dots, z_n, z_{n+1}) = (z_{\chi^{-1}(1)}, \dots, z_{\chi^{-1}(n)}, z_{n+1}), \quad \text{for } \chi \in \text{Sym}_n.$$

The map $p: M_{n+1} \rightarrow M_n$ induces a map $\bar{p}: M_{n+1}/\text{Sym}_n \rightarrow M_n/\text{Sym}_n = N_n$ which turns out to be a locally trivial fiber bundle. The fiber is again homeomorphic to $\mathbb{C} \setminus \{1, 2, \dots, n\}$, and $\bar{p}: M_{n+1}/\text{Sym}_n \rightarrow N_n$ has also a cross-section $\bar{\kappa}: N_n \rightarrow M_{n+1}/\text{Sym}_n$. So, from the homotopy long exact sequence of a fiber bundle (see Theorem 2.9) we obtain the following split exact sequence

$$1 \longrightarrow F_n \longrightarrow \pi_1(M_{n+1}/\text{Sym}_n) \begin{array}{c} \xrightarrow{\bar{p}_*} \\ \xleftarrow{\bar{\kappa}_*} \end{array} \pi_1(N_n) = \mathcal{B}_n \longrightarrow 1$$

where $F_n = \pi_1(\mathbb{C} \setminus \{1, \dots, n\})$, which is a free group of rank n . The action of $\mathcal{B}_n = \pi_1(N_n)$ on F_n derived from the above split exact sequence is exactly the Artin representation.

Another way to represent the Artin representation is by using the isomorphism $\mathcal{B}_n \simeq \mathcal{M}(\mathbb{D}, \{P_1, \dots, P_n\})$. Fix a basepoint $P_0 \in \partial\mathbb{D}$. Then it is easily shown that $\mathcal{M}(\mathbb{D}, \{P_1, \dots, P_n\})$ acts on $\pi_1(\mathbb{D} \setminus \{P_1, \dots, P_n\}, P_0) = F_n$, and that this action is the Artin representation.

The latter point of view of the Artin representations can be extended to all mapping class groups. In this setting, it is known as the Dehn–Nielsen–Baer theorem. Here is a version of this theorem.

Theorem 2.21 (Dehn, Nielsen [129], Baer [9], Magnus [118]). *Let Σ be a closed oriented surface, and let $\mathcal{P} = \{P_1, \dots, P_n\}$ be a collection of n punctures in Σ . Then*

the natural homomorphism $\rho: \mathcal{M}(\Sigma, \mathcal{P}) \rightarrow \text{Out}(\pi_1(\Sigma \setminus \mathcal{P}))$ is injective. Moreover, if $\mathcal{P} = \emptyset$, then the image of ρ is an index 2 subgroup of $\text{Out}(\pi_1(\Sigma))$.

We refer to [99] and [127] for detailed expositions on the Dehn–Nielsen–Baer theorem that include other versions of it.

Note. There are some variants of the Artin representations introduced in [151] and [59] that lead to invariants of links.

3 Artin groups

3.1 Definitions and examples

Let S be a finite set. A *Coxeter matrix* over S is a square matrix $M = (m_{st})_{s,t \in S}$ indexed by the elements of S such that

- $m_{ss} = 1$ for all $s \in S$;
- $m_{st} = m_{ts} \in \{2, 3, 4, \dots, +\infty\}$ for all $s, t \in S, s \neq t$.

A Coxeter matrix $M = (m_{st})_{s,t \in S}$ is usually represented by its *Coxeter graph*, $\Gamma = \Gamma(M)$. This is a labeled graph defined by the following data.

- S is the set of vertices of Γ .
- Two vertices $s, t \in S, s \neq t$, are joined by an edge if $m_{st} \geq 3$. This edge is labeled by m_{st} if $m_{st} \geq 4$.

Let Γ be a Coxeter graph. Define the *Coxeter system of type Γ* to be the pair (W, S) , where $W = W_\Gamma$ is the group presented by the generating set S and the relations

$$\begin{aligned} s^2 &= 1 \quad \text{for all } s \in S, \\ (st)^{m_{st}} &= 1 \quad \text{for all } s, t \in S, s \neq t, \text{ and } m_{st} \neq +\infty, \end{aligned}$$

where $M = (m_{st})_{s,t \in S}$ is the Coxeter matrix of Γ . The group $W = W_\Gamma$ is called the *Coxeter group of type Γ* .

If a, b are two letters and $m \in \mathbb{N}$, then $\text{prod}(a, b : m)$ denotes the word

$$\text{prod}(a, b : m) = \begin{cases} (ab)^{\frac{m}{2}} & \text{if } m \text{ is even,} \\ (ab)^{\frac{m-1}{2}} a & \text{if } m \text{ is odd.} \end{cases}$$

Let $\Sigma = \{\sigma_s; s \in S\}$ be an abstract set in one-to-one correspondence with S . Define the *Artin system of type Γ* to be the pair (G, Σ) , where $G = G_\Gamma$ is the group presented by the generating set Σ and the relations

$$\text{prod}(\sigma_s, \sigma_t : m_{st}) = \text{prod}(\sigma_t, \sigma_s : m_{st}) \quad \text{for } s, t \in S, s \neq t, \text{ and } m_{st} \neq +\infty.$$

The group G is called the *Artin group of type Γ* .

It is easily checked that the group W_Γ is also presented by the generating set S and the relations

$$s^2 = 1 \quad \text{for all } s \in S,$$

$$\text{prod}(s, t : m_{st}) = \text{prod}(t, s : m_{st}) \quad \text{for all } s, t \in S, s \neq t \text{ and } m_{st} \neq +\infty.$$

This shows that the mapping $\Sigma \rightarrow S, \sigma_s \mapsto s$, induces a *canonical epimorphism* $\theta: G_\Gamma \rightarrow W_\Gamma$.

If $m_{st} = 2$, then

$$\sigma_s \sigma_t = \text{prod}(\sigma_s, \sigma_t : m_{st}) = \text{prod}(\sigma_t, \sigma_s : m_{st}) = \sigma_t \sigma_s,$$

that is, σ_s and σ_t commute. So, if $\Gamma_1, \dots, \Gamma_l$ are the connected components of Γ , then

$$G_\Gamma = G_{\Gamma_1} \times G_{\Gamma_2} \times \dots \times G_{\Gamma_l}.$$

Similarly, we have

$$W_\Gamma = W_{\Gamma_1} \times W_{\Gamma_2} \times \dots \times W_{\Gamma_l}.$$

We say that G_Γ (or W_Γ) is *irreducible* if Γ is connected. We say that Γ (or G_Γ) is of *spherical type* if W_Γ is finite.

Example 1. Suppose that Γ is the graph A_n of Figure 9. Then $W_\Gamma = \text{Sym}_{n+1}$ is the symmetric group of $\{1, \dots, n+1\}$, and the Coxeter generators are the transpositions $s_1 = (1, 2), s_2 = (2, 3), \dots, s_n = (n, n+1)$. The Artin group G_Γ is the braid group \mathcal{B}_{n+1} on $n+1$ strands, and the Artin generators are the standard generators of \mathcal{B}_{n+1} given in Theorem 2.2. The canonical epimorphism coincides with the epimorphism described in Subsection 2.1.

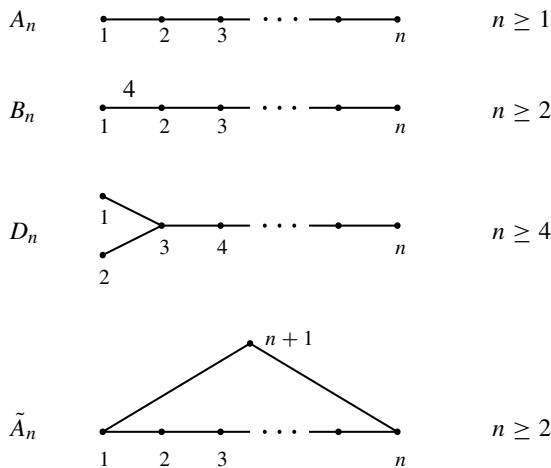


Figure 9. The Coxeter graphs A_n, B_n, D_n , and \tilde{A}_n .

Example 2. Suppose that Γ is the Coxeter graph B_n of Figure 9. Let $C_2 = \{\pm 1\}$ denote the cyclic group of order 2. Set $\text{Cub}_n = C_2^n \rtimes \text{Sym}_n$, where Sym_n acts on C_2^n by permutation of the coordinates. This is the group of isometries of a regular n -cube (see [94], for example). The group Cub_n is the Coxeter group of type B_n , and the Coxeter generators are

$$s_1 = (-1, 1, \dots, 1) \in C_2^n, \quad s_i = (i-1, i) \in \text{Sym}_n \quad \text{for } 2 \leq i \leq n.$$

Recall the Artin representation $\rho: \mathcal{B}_n \rightarrow \text{Aut}(F_n)$ defined in Subsection 2.4. Set $G = F_n \rtimes_{\rho} \mathcal{B}_n$. Recall also the action of Sym_n on M_{n+1} defined in Subsection 2.4. It follows from the exact sequence (2.4) that $G = \pi_1(M_{n+1}/\text{Sym}_n)$. In particular, G is an index $n+1$ subgroup of $\pi_1(M_{n+1}/\text{Sym}_{n+1}) = \pi_1(N_{n+1}) = \mathcal{B}_{n+1} = G_{A_n}$. Now, G is the Artin group of type B_n , and the Artin generators are

$$\tau_1 = x_1 \in F_n, \quad \tau_i = \sigma_{i-1} \in \mathcal{B}_n \quad \text{for } 2 \leq i \leq n.$$

(See [60]).

Example 3. Suppose that Γ is the Coxeter graph D_n of Figure 9, where $n \geq 4$. Let $\text{sgn}: C_2^n \rightarrow C_2$ be the homomorphism defined by

$$\text{sgn}(\varepsilon_1, \dots, \varepsilon_n) = \prod_{i=1}^n \varepsilon_i,$$

and let K be the kernel of sgn . The subgroup K is invariant under the action of Sym_n , thus one can consider the subgroup $W = K \rtimes \text{Sym}_n$ of $\text{Cub}_n = C_2^n \rtimes \text{Sym}_n$. This is the Coxeter group of type D_n , and the Coxeter generators are

$$s_1 = (-1, -1, 1, \dots, 1) \cdot (1, 2), \quad s_i = (1, 1, 1, \dots, 1) \cdot (i-1, i) \quad \text{for } 2 \leq i \leq n.$$

(See [94], for example).

Let $F_{n-1} = F(y_1, \dots, y_{n-1})$ be a free group of rank $n-1$. Let $\rho_{D,1}: F_{n-1} \rightarrow F_{n-1}$ be the automorphism defined by

$$\rho_{D,1}: \begin{cases} y_1 \mapsto y_1, \\ y_j \mapsto y_1^{-1} y_j \quad \text{if } j \geq 2. \end{cases}$$

For $2 \leq i \leq n-1$, let $\rho_{D,i}: F_{n-1} \rightarrow F_{n-1}$ be the automorphism defined by

$$\rho_{D,i}: \begin{cases} y_{i-1} \mapsto y_i, \\ y_i \mapsto y_i y_{i-1}^{-1} y_i, \\ y_j \mapsto y_j \quad \text{if } j \neq i-1, i. \end{cases}$$

One can easily show the following.

Lemma 3.1. *The mapping $\sigma_i \mapsto \rho_{D,i}$, $1 \leq i \leq n-1$, determines a representation $\rho_D: \mathcal{B}_n \rightarrow \text{Aut}(F_{n-1})$.*

The following is implicit in [135] and explicit in [60].

Theorem 3.2 (Perron, Vannier [135]). *The representation $\rho_D: \mathcal{B}_n \rightarrow \text{Aut}(F_{n-1})$ is faithful, and the semidirect product $F_{n-1} \rtimes_{\rho_D} \mathcal{B}_n$ is isomorphic to the Artin group G_{D_n} of type D_n .*

Note. It was shown by Allcock [2] that the Artin group G_{D_n} of type D_n can be also presented as an index 2 subgroup of the n -strand braid group of a plane with a single orbifold point of degree 2.

Example 4. Suppose that Γ is the graph \tilde{A}_n of Figure 9. Let Sym_{n+1} act on \mathbb{Z}^{n+1} by permutation of the coordinates. Set $\Lambda = \{(x_1, \dots, x_{n+1}) \in \mathbb{Z}^{n+1}; \sum_{i=1}^{n+1} x_i = 0\}$, and observe that $\Lambda \simeq \mathbb{Z}^n$ is invariant under the action of Sym_{n+1} . Then $\Lambda \rtimes \text{Sym}_{n+1}$ is the Coxeter group of type Γ (see [28]).

Let $\Phi: G_{B_{n+1}} \rightarrow \mathbb{Z}$ be the homomorphism defined by

$$\Phi(\sigma_1) = 1, \quad \Phi(\sigma_i) = 0 \quad \text{for } 2 \leq i \leq n.$$

It was observed by several authors [2], [47], [73], [104], that the kernel of Φ is isomorphic to the Artin group $G_{\tilde{A}_n}$ of type \tilde{A}_n . In particular, $G_{\tilde{A}_n}$ is a subgroup of \mathcal{B}_{n+2} .

Viewed as a subgroup of $\mathcal{B}_{n+2} = \mathcal{M}(\mathbb{D}, \{P_1, P_2, \dots, P_{n+2}\})$, the Artin generators of $G_{\tilde{A}_n}$ can be described in terms of braid twists as follows. We place P_1, \dots, P_{n+2} in the interior of \mathbb{D} like in Figure 10. For $1 \leq i \leq n + 1$, let τ_i denote the braid twist along the arc a_i . Then $\tau_1, \dots, \tau_{n+1}$ are the Artin generators of $G_{\tilde{A}_n}$.

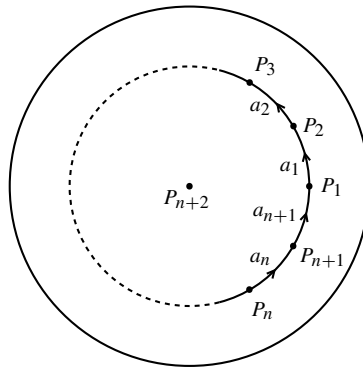


Figure 10. Standard generators of $G_{\tilde{A}_n}$.

Note. For a group G we denote by $Z(G)$ the center of G . If $\Gamma = A_n, B_n$, or \tilde{A}_n , then $G_\Gamma/Z(G_\Gamma)$ can be viewed as a finite index subgroup of the mapping class group of a punctured sphere. This has been cleverly exploited to study the group G_Γ itself, in particular, to compute the group of automorphisms of G_Γ (see [44], [10]). Note that the center of G_{A_n} and G_{B_n} is an infinite cyclic group (see [71], [32]), and the center of $G_{\tilde{A}_n}$ is trivial (see [100]).

3.2 Coxeter groups

Coxeter groups were introduced by Tits [148] in a manuscript which was recently published, and whose results appeared in the seminal Bourbaki's book [28]. The present subsection is a brief survey on these groups with a special emphasis on the results that are needed to study Artin groups. Standard references for the subject are [28], [97].

Let Γ be a Coxeter graph, let $M = (m_{st})_{s,t \in S}$ be its associated Coxeter matrix, and let (W, S) be the Coxeter system of type Γ .

Let $\Pi = \{e_s; s \in S\}$ be an abstract set in one-to-one correspondence with S , whose elements are called *simple roots*. We denote by V the real vector space having Π as a basis, and by $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ the symmetric bilinear form defined by

$$\langle e_s, e_t \rangle = \begin{cases} -\cos(\frac{\pi}{m_{st}}) & \text{if } m_{st} \neq +\infty, \\ -1 & \text{if } m_{st} = +\infty. \end{cases}$$

For $s \in S$ we define the reflection $r_s: V \rightarrow V$ by

$$r_s(\mathbf{x}) = \mathbf{x} - 2\langle \mathbf{x}, e_s \rangle e_s, \quad \mathbf{x} \in V.$$

Theorem 3.3 (Tits [148]). *The mapping $s \mapsto r_s$, $s \in S$, determines a faithful linear representation $\rho: W \rightarrow \text{GL}(V)$.*

The above linear representation is called the *canonical representation* of (W, S) . Note that the bilinear form $\langle \cdot, \cdot \rangle$ is invariant under the action of W .

The *root system* Φ of (W, S) is defined to be the orbit of Π under the action of W , that is,

$$\Phi = \{w \cdot e_s; w \in W, s \in S\}.$$

Let $f \in \Phi$. Write $f = \sum_{s \in S} \lambda_s e_s$, where $\lambda_s \in \mathbb{R}$ for all $s \in S$. We say that f is a *positive root* (resp. a *negative root*) if $\lambda_s \geq 0$ (resp. $\lambda_s \leq 0$) for all $s \in S$. The set of positive roots (resp. negative roots) is denoted by Φ_+ (resp. by Φ_-). The following is proved in [28] for finite root systems, but the same proof works in general (see also [97], [72]).

Proposition 3.4. *We have the disjoint union $\Phi = \Phi_+ \sqcup \Phi_-$.*

Let A be a finite set that we call an *alphabet*. Let A^* denote the set of finite sequences of elements of A that we call *words on A* . We define an operation on A^* by

$$(a_1, \dots, a_p) \cdot (b_1, \dots, b_q) = (a_1, \dots, a_p, b_1, \dots, b_q).$$

Clearly, A^* endowed with this operation is a monoid which is called the *free monoid on A* . The unit in A^* is the empty word $\epsilon = ()$.

Each element w in the Coxeter group W can be written in the form $w = s_1 s_2 \dots s_l$, where $s_1, s_2, \dots, s_l \in S$. If l is as small as possible, then l is called the *word length* of w

and is denoted by $l = \text{lg}_S(w)$. If $w = s_1 s_2 \dots s_l$, then the word $\omega = (s_1, s_2, \dots, s_l)$ is called an *expression* of w . If in addition $l = \text{lg}_S(w)$, then ω is called a *reduced expression* of w .

For $w \in W$ we set

$$\Phi_w = \{f \in \Phi_+; w^{-1}f \in \Phi_-\}.$$

Then the word length and the root systems are related by the following.

Proposition 3.5 (Bourbaki [28]). *We have $|\Phi_w| = \text{lg}_S(w)$ for all $w \in W$.*

Let G be a group. A subset $S \subset G$ is called a *positive generating set* of G if it generates G as a monoid. Let S be a positive generating set of G . for $\omega \in S^*$, we denote by $\bar{\omega}$ the element of G represented by ω . A *solution to the word problem* for G is an algorithm which, given $\omega \in S^*$, decides whether $\bar{\omega}$ is trivial or not.

We turn now to describe Tits' solution to the word problem for Coxeter groups.

Let $\omega, \omega' \in S^*$. We say that ω is *transformable to ω' by an M -operation of type I* if there exist $\omega_1, \omega_2 \in S^*$ and $s \in S$ such that

$$\omega = \omega_1 \cdot (s, s) \cdot \omega_2 \quad \text{and} \quad \omega' = \omega_1 \cdot \omega_2.$$

We say that ω is *transformable to ω' by an M -operation of type II* if there exist $\omega_1, \omega_2 \in S^*$ and $s, t \in S$ such that $s \neq t, m_{st} \neq +\infty$,

$$\omega = \omega_1 \cdot \text{prod}(s, t : m_{st}) \cdot \omega_2 \quad \text{and} \quad \omega' = \omega_1 \cdot \text{prod}(t, s : m_{st}) \cdot \omega_2.$$

Note that an M -operation of type I shortens the length of the word, but not an M -operation of type II. An M -operation of type II is reversible, but not an M -operation of type I. If ω is transformable to ω' by an M -operation, then $\bar{\omega} = \bar{\omega}'$.

A word ω is called *M -reduced* if its length cannot be reduced by means of M -operations.

Theorem 3.6 (Tits [147]). (1) *A word $\omega \in S^*$ is reduced if and only if it is M -reduced.*

(2) *Let $\omega, \omega' \in S^*$ be two reduced words. We have $\bar{\omega} = \bar{\omega}'$ if and only if one can pass from ω to ω' with a finite sequence of M -operations of type II.*

Now, we introduce a partial order on the Coxeter group W whose role is of importance in the study of the associated Artin group and monoid.

For $u, v \in W$, we set $u \leq_L v$ if there exists $w \in W$ such that $v = uw$ and $\text{lg}_S(v) = \text{lg}_S(u) + \text{lg}_S(w)$.

Proposition 3.7 (Bourbaki [28]). (1) *Let $u, v \in W$. There exists a unique $w^o \in W$ such that $w^o \leq_L u, w^o \leq_L v$, and $w \leq_L w^o$ whenever $w \leq_L u$ and $w \leq_L v$.*

(2) *Suppose that W is finite. Let $u, v \in W$. There exists a unique $w_o \in W$ such that $u \leq_L w_o, v \leq_L w_o$, and $w_o \leq_L w$ whenever $u \leq_L w$ and $v \leq_L w$.*

The element w^o of Proposition 3.7 is denoted by $w^o = u \wedge_L v$, and the element w_o is denoted by $w_o = u \vee_L v$ (if it exists). Note that, by the above, (W, \leq_L) is a lattice if W is finite. In that case, W has a greatest element which is often denoted by w_0 .

We finish the subsection with the classification of the spherical type Coxeter graphs.

Recall that, if $\Gamma_1, \dots, \Gamma_l$ are the connected components of a Coxeter graph Γ , then

$$W_\Gamma = W_{\Gamma_1} \times W_{\Gamma_2} \times \dots \times W_{\Gamma_l}.$$

In particular, Γ is of spherical type if and only if all the components $\Gamma_1, \dots, \Gamma_l$ are of spherical type. So, we only need to classify the connected Coxeter graphs of spherical type.

Theorem 3.8 (Coxeter [55], [56]). (1) *A Coxeter graph Γ is of spherical type if and only if the canonical bilinear form $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ is positive definite.*

(2) *The connected spherical type Coxeter graphs are the Coxeter graphs listed in Figure 11.*

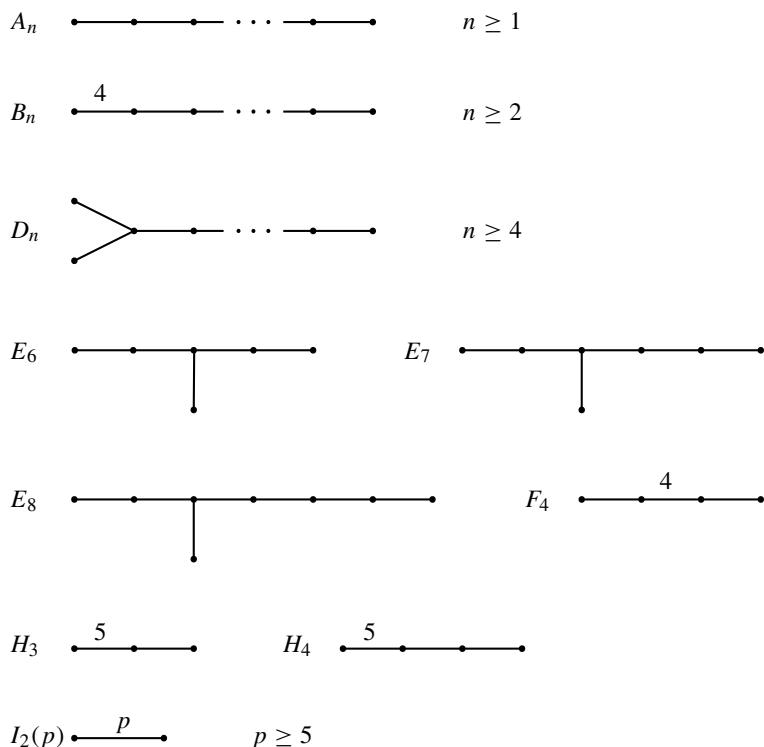


Figure 11. The connected spherical type Coxeter graphs.

3.3 Artin monoids

Let Γ be a Coxeter graph, let (W, S) be the Coxeter system of type Γ , and let (G, Σ) be the Artin system of type Γ . Define the *Artin monoid of type Γ* to be the monoid $G^+ = G_\Gamma^+$ presented as a monoid by the generating set $\Sigma = \{\sigma_s; s \in S\}$ and the relations

$$\begin{aligned} \text{prod}(\sigma_s, \sigma_t : m_{st}) = \text{prod}(\sigma_t, \sigma_s : m_{st}) \quad \text{for all } s, t \in S, \\ s \neq t \text{ and } m_{st} \neq +\infty. \end{aligned}$$

Theorem 3.9 (Paris [133]). *The natural homomorphism $G_\Gamma^+ \rightarrow G_\Gamma$ is injective.*

Recall the homomorphism $\theta: G_\Gamma \rightarrow W_\Gamma, \sigma_s \mapsto s$. We denote by $\theta^+: G_\Gamma^+ \rightarrow W_\Gamma$ the restriction of θ to G_Γ^+ . We define a set-section $\kappa: W_\Gamma \rightarrow G_\Gamma^+$ of θ^+ as follows. Let $w \in W$, and let $\omega = (s_1, s_2, \dots, s_l)$ be a reduced expression of w . Then

$$\kappa(w) = \sigma_{s_1} \sigma_{s_2} \dots \sigma_{s_l}.$$

By Theorem 3.6, the definition of $\kappa(w)$ does not depend on the choice of the reduced expression of w .

Observe also that the defining relations of G_Γ^+ are homogeneous, thus G_Γ^+ has a well-defined word length $\text{lg}: G_\Gamma^+ \rightarrow \mathbb{N}, \sigma_{s_1} \dots \sigma_{s_l} \mapsto l$. This word length satisfies the following properties:

- $\text{lg}(\alpha) = 0$ if and only if $\alpha = 1$;
- $\text{lg}(\alpha\beta) = \text{lg}(\alpha) + \text{lg}(\beta)$ for all $\alpha, \beta \in G_\Gamma^+$.

We define partial orders \leq_L and \leq_R on G_Γ^+ by

- $\alpha \leq_L \beta$ if there exists $\gamma \in G_\Gamma^+$ such that $\alpha\gamma = \beta$;
- $\alpha \leq_R \beta$ if there exists $\gamma \in G_\Gamma^+$ such that $\gamma\alpha = \beta$.

The following is again a direct consequence of Theorem 3.6.

Lemma 3.10. *Let $u, v \in W$. We have $u \leq_L v$ if and only if $\kappa(u) \leq_L \kappa(v)$.*

The set $\mathcal{S} = \{\kappa(w); w \in W\}$ is called the set of *simple elements* of G_Γ^+ . If W is finite and w_0 is the greatest element of W , then $\kappa(w_0)$ is called the *Garside element* of G_Γ^+ and is denoted by $\Delta = \kappa(w_0)$.

The following Theorems 3.11 and 3.12 are key results in the study of Artin monoids and groups. They are implicit in the work of Brieskorn and Saito [32], and explicit for the spherical type Artin groups in the work of Deligne [71]. Complete and detailed proofs of them can be found in [126].

Theorem 3.11. *Let $\alpha \in G_\Gamma^+$. Set*

$$E(\alpha) = \{a \in \mathcal{S}; a \leq_L \alpha\}.$$

Then $E(\alpha)$ has a greatest element. That is, there exists $a_0 \in E(\alpha)$ such that $E(\alpha) = \{a \in \mathcal{S}; a \leq_L a_0\}$.

For $\alpha \in G_\Gamma^+$ we denote by $\delta(\alpha)$ the greatest element of $E(\alpha)$.

Theorem 3.12. *Let $\alpha, \beta \in G_\Gamma^+$. Then $\delta(\alpha\beta) = \delta(\alpha\delta(\beta))$.*

Theorems 3.11 and 3.12 have the following consequences whose significance will become clear in the next section.

Theorem 3.13. (1) *Let $\alpha, \beta \in G_\Gamma^+$. There exists a unique $\gamma^o \in G_\Gamma^+$ such that $\gamma^o \leq_L \alpha$, $\gamma^o \leq_L \beta$, and $\gamma \leq_L \gamma^o$ whenever $\gamma \leq_L \alpha$ and $\gamma \leq_L \beta$.*

(2) *Suppose that Γ is of spherical type. Let $\alpha, \beta \in G_\Gamma^+$. There exists a unique $\gamma_o \in G_\Gamma^+$ such that $\alpha \leq_L \gamma_o$, $\beta \leq_L \gamma_o$, and $\gamma_o \leq_L \gamma$ whenever $\alpha \leq_L \gamma$ and $\beta \leq_L \gamma$.*

The element γ^o of Theorem 3.13 is denoted by $\gamma^o = \alpha \wedge_L \beta$, and the element γ_o is denoted by $\gamma_o = \alpha \vee_L \beta$ (if it exists). Note that the same result is valid if we replace \leq_L by \leq_R .

Proof. We prove (1) by induction on $\lg(\alpha) + \lg(\beta)$. By Proposition 3.7 and by Lemma 3.10, $\alpha \wedge_L \beta$ exists if $\alpha, \beta \in \mathcal{S}$.

Let $\alpha, \beta \in G_\Gamma^+$. Set $a = \delta(\alpha) \wedge_L \delta(\beta)$ (a exists by the above observation). If $a = 1$, then we must have $\gamma^o = \alpha \wedge_L \beta = 1$. Suppose $a \neq 1$. Let $\alpha', \beta' \in G_\Gamma^+$ such that $\alpha = a\alpha'$ and $\beta = a\beta'$. The element $\alpha' \wedge_L \beta'$ exists by induction. Then $\gamma^o = a \cdot (\alpha' \wedge_L \beta')$ (the proof of this equality is left to the reader).

Now, we assume that Γ is of spherical type and turn to prove (2). Let w_0 be the greatest element of W , and let $\Delta = \kappa(w_0)$ be the Garside element of G_Γ^+ . It is shown in [28] that $w_0^{-1} = w_0$ and $w_0 S w_0 = S$. This implies that $\Delta \cdot \Sigma \cdot \Delta^{-1} = \Sigma$, and, consequently, there exists a permutation $\tau: \mathcal{S} \rightarrow \mathcal{S}$ such that $\Delta\alpha = \tau(\alpha)\Delta$ for all $\alpha \in G_\Gamma^+$.

Let $\alpha \in G_\Gamma^+$. Set $\alpha = a_1 a_2 \dots a_r$, where $a_i = \delta(a_i a_{i+1} \dots a_r) \in \mathcal{S}$ for all $1 \leq i \leq r$. Using the above observation, it is easily shown that $\alpha \leq_L \Delta^r$.

Let $\alpha, \beta \in G_\Gamma^+$. Set $\mathcal{E} = \{\gamma \in G_\Gamma^+; \alpha \leq_L \gamma \text{ and } \beta \leq_L \gamma\}$. We have $\mathcal{E} \neq \emptyset$ since, by the above, it contains an element of the form Δ^r . Let γ_o be the smallest element of \mathcal{E} (this element exists by (1)). Then $\gamma_o = \alpha \vee_L \beta$. \square

3.4 Artin groups

We turn now to present a geometrical interpretation of Artin groups which extends the interpretation of braid groups in term of configuration spaces. We focus our presentation on spherical type Artin groups, but many of the results stated in this subsection can be extended in some sense to the other Artin groups.

Let Γ be a spherical type Coxeter graph, let (W, S) be the Coxeter system of type Γ , and let (G, Σ) be the Artin system of type Γ . Recall the set $\Pi = \{e_s; s \in S\}$ of simple roots, the vector space $V = \bigoplus_{s \in S} \mathbb{R}e_s$, and the canonical bilinear form $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$, which, by Theorem 3.8, is positive definite. We assume that W is embedded in $\text{GL}(V)$ via the canonical representation.

Let \mathcal{R} be the set of reflections in W . For each $r \in \mathcal{R}$, let H_r be the hyperplane of V fixed by r . Then W acts freely on the complement of $\bigcup_{r \in \mathcal{R}} H_r$ (see [28]). Complexifying the action, we get an action of W on $V_{\mathbb{C}} = \mathbb{C} \otimes V$ which is free on the complement of $\bigcup_{r \in \mathcal{R}} \mathbb{C} \otimes H_r$. Set

$$M_{\Gamma} = V_{\mathbb{C}} \setminus \left(\bigcup_{r \in \mathcal{R}} \mathbb{C} \otimes H_r \right), \quad N_{\Gamma} = M_{\Gamma}/W.$$

By a theorem of Chevalley [49], Shephard, and Todd [144], $V_{\mathbb{C}}/W$ is isomorphic to \mathbb{C}^n , thus N_{Γ} is the complement in \mathbb{C}^n of an algebraic set, $(\bigcup_{r \in \mathcal{R}} \mathbb{C} \otimes H_r)/W$, called the *discriminant* of type Γ .

Theorem 3.14 (Brieskorn [30]). $\pi_1(N_{\Gamma}) \simeq G_{\Gamma}$.

Note. Infinite Coxeter groups also act as reflection groups on \mathbb{R}^n . However, to extend Theorem 3.14 to these groups we should replace V by the Tits cone $U \subset V$ (see [28]), and $V_{\mathbb{C}}$ by $(U + iV) \subset V_{\mathbb{C}}$. Then W acts freely on $(U + iV) \setminus (\bigcup_{r \in \mathcal{R}} \mathbb{C} \otimes H_r) = M_{\Gamma}$, and it was shown by Van der Lek [116] that $\pi_1(N_{\Gamma}) \simeq G_{\Gamma}$, where $N_{\Gamma} = M_{\Gamma}/W$.

An extension of Corollary 2.12 to the spherical type Artin groups is:

Theorem 3.15 (Deligne [71]). *Let Γ be a spherical type Coxeter graph. Then N_{Γ} and M_{Γ} are $K(\pi, 1)$.*

Note. It is an open problem to know whether N_{Γ} is $K(\pi, 1)$ if Γ is not of spherical type. The answer is yes for the so-called FC-type Artin groups and 2-dimensional Artin groups [45], and also for few affine type Artin groups (see [47], [38]).

Note. We may replace W by a finite complex reflection group acting on \mathbb{C}^n , and M_{Γ} by $M(W) = \mathbb{C}^n \setminus (\bigcup_{r \in \mathcal{R}} H_r)$, where \mathcal{R} is the set of reflections in W , and H_r denotes the hyperplane fixed by r . Here again, the group W acts freely on $M(W)$ and, by [49] and [144], $N(W) = M(W)/W$ is isomorphic to the complement in \mathbb{C}^n of an algebraic set. It was recently proved by Bessis [15] that $N(W)$ is always $K(\pi, 1)$. A classification of the finite complex reflection groups was obtained by Shephard and Todd [144], and a nice presentation of $\pi_1(N(W))$ is known for all these groups but four exceptional cases (see [33], [16]).

4 Garside groups

4.1 Garside monoids

A monoid M is called *atomic* if there exists a function $\nu: M \rightarrow \mathbb{N}$ such that

- $\nu(\alpha) = 0$ if and only if $\alpha = 1$;
- $\nu(\alpha\beta) \geq \nu(\alpha) + \nu(\beta)$ for all $\alpha, \beta \in M$.

Such a function ν is called a *norm* on M . An element $\alpha \in M$ is called an *atom* if it is indecomposable, that is, if $\alpha = \beta\gamma$, then either $\beta = 1$ or $\gamma = 1$.

The following is proved in [70].

Lemma 4.1. *Let M be an atomic monoid. A subset $S \subset M$ generates M if and only if it contains all the atoms. In particular, M is finitely generated if and only if it contains finitely many atoms.*

Let M be an atomic monoid. We define on M two partial orders \leq_L and \leq_R as follows.

- Set $\alpha \leq_L \beta$ if there exists $\gamma \in M$ such that $\alpha\gamma = \beta$.
- Set $\alpha \leq_R \beta$ if there exists $\gamma \in M$ such that $\gamma\alpha = \beta$.

The orders \leq_L and \leq_R are called the *left and right divisibility orders*, respectively.

A monoid M is called a *Garside monoid* if

- M is atomic and finitely generated;
- M is cancelative (that is, if $\alpha\beta\gamma = \alpha\beta'\gamma$, then $\beta = \beta'$, for all $\alpha, \beta, \beta', \gamma \in M$);
- (M, \leq_L) and (M, \leq_R) are lattices;
- there exists an element $\Delta \in M$, called a *Garside element*, such that the sets $L(\Delta) = \{\alpha \in M; \alpha \leq_L \Delta\}$ and $R(\Delta) = \{\alpha \in M; \alpha \leq_R \Delta\}$ are equal and generate M .

If M is a Garside monoid, then the lattice operations of (M, \leq_L) (resp. of (M, \leq_R)) are denoted by \vee_L and \wedge_L (resp. by \vee_R and \wedge_R).

Let M be a monoid. The *group of fractions* of M is defined to be the group $G(M)$ presented with the generating set M and the relations $\alpha \cdot \beta = \gamma$ if $\alpha\beta = \gamma$ in M . Such a group has the universal property that if $\varphi: M \rightarrow H$ is a homomorphism and H is a group, then there exists a unique homomorphism $\hat{\varphi}: G(M) \rightarrow H$ such that $\varphi = \hat{\varphi} \circ \iota$, where $\iota: M \rightarrow G(M)$ is the natural homomorphism. Note that the latter homomorphism $\iota: M \rightarrow G(M)$ is not injective in general.

A *Garside group* is defined to be the group of fractions of a Garside monoid.

Note. Garside monoids and groups were introduced in [70] in a slightly restricted sense, and in [67] in the larger sense which is now generally used. This notion was extended to the notion of quasi-Garside monoids [75], [13], to study some non-spherical Artin groups. Quasi-Garside monoids have the same definition as the Garside monoids except they are not required to be finitely generated. Recently, this notion was extended to the notion of Garside categories [107], [108], [76], [14], which, in some sense, has to be considered as a geometric object more than as an algebraic one. Garside categories are a central concept in Bessis' solution to the $K(\pi, 1)$ problem for complex reflection arrangements (see [15]).

Motivating examples of Garside groups are the Artin groups of spherical type:

Theorem 4.2. *Let Γ be a spherical type Coxeter graph. Then G_Γ^+ is a Garside monoid. In particular, G_Γ is a Garside group.*

Note that Theorem 4.2 is essentially a restatement of Theorem 3.13.

Other interesting examples of Garside groups include all torus link groups (see [137]) and some generalized braid groups associated to complex reflection groups (see [15]).

Note. Two different Garside monoids can have the same group of fractions. In particular, the Artin groups of spherical type are groups of fractions of other Garside monoids, called *dual Artin monoids*, introduced by Birman, Ko, and Lee [27] for the braid groups, and by Bessis [12] for the other ones.

Note. A Garside element is not unique. For instance, if Δ is a Garside element, then Δ^k is a Garside element for all $k \geq 1$ (see [67]).

We say that a monoid M satisfies the *Ore conditions* if

- M is cancelative;
- for all $\alpha, \beta \in M$, there exist $\alpha', \beta' \in M$ such that $\alpha\alpha' = \beta\beta'$.

It is well-known that a monoid which satisfies the Ore conditions embeds in its group of fractions. On the other hand, a Garside monoid clearly satisfies the Ore conditions. Thus:

Proposition 4.3. *Let M be a Garside monoid. Then the natural homomorphism $\iota: M \rightarrow G(M)$ is injective.*

Let M be a Garside monoid and let $G = G(M)$ be the group of fractions of M . Then the partial orders \leq_L and \leq_R can be extended to G as follows.

- Set $\alpha \leq_L \beta$ if $\alpha^{-1}\beta \in M$.
- Set $\alpha \leq_R \beta$ if $\beta\alpha^{-1} \in M$.

One can easily verify that (G, \leq_L) and (G, \leq_R) are lattices. This can be used, for example, to prove the following.

Proposition 4.4. *A Garside group is torsion free.*

Proof. Let $\alpha \in G$ such that $\alpha^n = 1$ for some $n \geq 1$. Set $\beta = 1 \vee_L \alpha \vee_L \dots \vee_L \alpha^{n-1}$. It is easily seen that \leq_L is invariant by left multiplication. This implies that $\alpha\beta = \beta$, hence $\alpha = 1$. □

Note. Let G be a Garside group. Finite dimensional $K(G, 1)$ (that is, $K(\pi, 1)$ spaces having G as fundamental group) were described in [69] and [46]. This implies that G is torsion free, but also more.

4.2 Reversing processes and presentations

Let Σ be a finite set. Let Σ^* be the free monoid on Σ . Recall that the elements of Σ^* are the finite sequences of elements of Σ that are called *words* on Σ . Recall that a congruence on Σ^* is defined to be an equivalence relation invariant by left and right multiplication. If \equiv is a congruence on Σ^* and $M = (\Sigma^*/\equiv)$, then we denote by $\Sigma^* \rightarrow M$, $\omega \mapsto \bar{\omega}$ the natural epimorphism.

Define a *complement* on Σ to be a map $f: \Sigma \times \Sigma \rightarrow \Sigma^*$ such that $f(x, x) = \epsilon$ for all $x \in \Sigma$, where ϵ denotes the empty word. To a complement f we associate two monoids:

$$M_L^f = \langle \Sigma \mid xf(x, y) = yf(y, x) \text{ for all } x, y \in \Sigma \rangle^+;$$

$$M_R^f = \langle \Sigma \mid f(y, x)x = f(x, y)y \text{ for all } x, y \in \Sigma \rangle^+.$$

For $u, v \in \Sigma^*$, we use the notation $u \equiv_L^f v$ (resp. $u \equiv_R^f v$) to mean that $\bar{u} = \bar{v}$ in M_L^f (resp. in M_R^f).

Example. Let Γ be a Coxeter graph and let $M = (m_{st})_{s,t \in S}$ be the Coxeter matrix of Γ . Suppose that $m_{st} \neq +\infty$ for all $s, t \in S$, $s \neq t$. Let $\Sigma = \{\sigma_s; s \in S\}$. Let $f: \Sigma \times \Sigma \rightarrow \Sigma^*$ be the complement defined by

$$f(\sigma_s, \sigma_t) = \text{prod}(\sigma_t, \sigma_s : m_{st} - 1).$$

Then $G_\Gamma^+ = M_L^f$.

Suppose we are given a complement $f: \Sigma \times \Sigma \rightarrow \Sigma^*$. Let $\Sigma^{-1} = \{x^{-1}; x \in \Sigma\}$ be the set of *inverses* of elements of Σ . Let $\omega, \omega' \in (\Sigma \sqcup \Sigma^{-1})^*$. We say that ω is *f-reversible on the left in one step* to ω' if there exist $\omega_1, \omega_2 \in (\Sigma \sqcup \Sigma^{-1})^*$ and $x, y \in \Sigma$ such that

$$\omega = \omega_1 x^{-1} y \omega_2 \quad \text{and} \quad \omega' = \omega_1 \cdot f(x, y) \cdot f(y, x)^{-1} \cdot \omega_2.$$

Note that y can be equal to x in the above definition. In that case we have $\omega = \omega_1 x^{-1} x \omega_2$ and $\omega' = \omega_1 \omega_2$. Also note that $\bar{\omega} = \bar{\omega}'$ in $G(M_L^f)$ if ω is *f-reversible on the left in one step* to ω' .

Let $p \geq 0$. We say that ω is *f-reversible on the left in p steps* to ω' if there exists a sequence $\omega = \omega_0, \omega_1, \dots, \omega_p = \omega'$ in $(\Sigma \sqcup \Sigma^{-1})^*$ such that ω_{i-1} is *f-reversible on the left in one step* to ω_i for all $1 \leq i \leq p$. The property that ω is *f-reversible on the left* to ω' is denoted by $\omega \mapsto_L^f \omega'$.

We define the *f-reversibility on the right* in the same way, replacing subwords of the form yx^{-1} by their corresponding words $f(x, y)^{-1} \cdot f(y, x)$. The property that ω is *f-reversible on the right* to ω' is denoted by $\omega \mapsto_R^f \omega'$.

A word $\omega \in (\Sigma \sqcup \Sigma^{-1})^*$ is said to be *f-reduced on the left* (resp. *f-reduced on the right*) if it is of the form $\omega = vu^{-1}$ (resp. $\omega = u^{-1}v$) with $u, v \in \Sigma^*$.

It is shown in [66] that a reversing process is confluent, namely:

Proposition 4.5 (Dehornoy [66]). *Let $f: \Sigma \times \Sigma \rightarrow \Sigma^*$ be a complement, and let $\omega \in (\Sigma \sqcup \Sigma^{-1})^*$. Suppose that there exist $p \geq 0$ and an f -reduced word vu^{-1} on the left such that ω is f -reversible on the left in p steps to vu^{-1} . Then any sequence of left f -reversing transformations starting from ω converges to vu^{-1} in p steps.*

Let $u, v \in \Sigma^*$. Suppose there exist $u', v' \in \Sigma^*$ such that $u^{-1}v \mapsto_L^f v'(u')^{-1}$. By Proposition 4.5, the words u' and v' are unique. Moreover, it is easily checked that we also have $v^{-1}u \mapsto_L^f u'(v')^{-1}$. In this case we set

$$u' = C_L^f(v, u) \quad \text{and} \quad v' = C_L^f(u, v).$$

Similarly, if there exist $u', v' \in \Sigma^*$ such that $vu^{-1} \mapsto_R^f (u')^{-1}(v')$, then we have $uv^{-1} \mapsto_R^f (v')^{-1}(u')$, u' and v' are unique, and we set

$$u' = C_R^f(u, v) \quad \text{and} \quad v' = C_R^f(v, u).$$

Lemma 4.6 (Dehornoy [66]). *Let $f: \Sigma \times \Sigma \rightarrow \Sigma^*$ be a complement. Let $u, v \in \Sigma^*$. Suppose that $C_L^f(u, v)$ and $C_L^f(v, u)$ exist. Then*

$$u \cdot C_L^f(u, v) \equiv_L^f v \cdot C_L^f(v, u).$$

A complement $f: \Sigma \times \Sigma \rightarrow \Sigma^*$ is said to be *coherent on the left* if for all $x, y, z \in \Sigma$, $C_L^f(f(x, y), f(x, z))$ and $C_L^f(f(y, x), f(y, z))$ exist and are \equiv_L^f -equivalent. Similarly, we say that f is *coherent on the right* if for all $x, y, z \in \Sigma$, $C_R^f(f(z, x), f(y, x))$ and $C_R^f(f(z, y), f(x, y))$ exist and are \equiv_R^f -equivalent.

Theorem 4.7 (Dehornoy, Paris [70], [67]). *Let M be a finitely generated monoid and let Σ be a finite generating set of M . Then M is a Garside monoid if and only if it satisfies the following three conditions.*

- M is atomic.
- There exist a complement $f: \Sigma \times \Sigma \rightarrow \Sigma^*$ coherent on the left and a complement $g: \Sigma \times \Sigma \rightarrow \Sigma^*$ coherent on the right such that $M = M_L^f = M_R^g$.
- There exists an element $\Delta \in M$ such that the sets $L(\Delta) = \{\alpha \in M; \alpha \leq_L \Delta\}$ and $R(\Delta) = \{\alpha \in M; \alpha \leq_R \Delta\}$ are equal and generate M .

We refer to [70] and [67] for more “algorithmic” conditions to detect a Garside monoid in terms of complements and presentations, and turn to explain some applications of the reversing processes.

Let M be a Garside monoid and let $G = G(M)$ be its group of fractions. Let $f: \Sigma \times \Sigma \rightarrow \Sigma^*$ and $g: \Sigma \times \Sigma \rightarrow \Sigma^*$ be complements such that $M = M_L^f = M_R^g$.

First, the complements f and g lead to algorithms:

Proposition 4.8 (Dehornoy, Paris [70], [67]). (1) *The complement f is coherent on the left, and the complement g is coherent on the right.*

(2) *Let $\omega \in (\Sigma \sqcup \Sigma^{-1})^*$. There exist a (unique) f -reduced word vu^{-1} on the left, and a (unique) g -reduced word $(u')^{-1}(v')$ on the right, such that $\omega \mapsto_L^f vu^{-1}$ and $\omega \mapsto_R^g (u')^{-1}(v')$.*

This can be used to solve the word problem:

Proposition 4.9 (Dehornoy, Paris [70], [67]). *Let $\omega \in (\Sigma \sqcup \Sigma^{-1})^*$. Let $u, v \in \Sigma$ such that $\omega \mapsto_L^f vu^{-1}$ (see Proposition 4.8). Then $\bar{\omega} = 1$ in $G = G(M)$ if and only if $u^{-1}v \mapsto_L^f \epsilon$, where ϵ denotes the empty word.*

This can be also used to compute the lattice operations of (M, \leq_L) and (M, \leq_R) .

Proposition 4.10 (Dehornoy, Paris [70], [67]). *Let $u, v \in \Sigma^*$. Set $u' = C_L^f(u, v)$ and $v' = C_L^f(v, u)$. Then $\bar{u} \vee_L \bar{v}$ is represented by*

$$uu' \equiv_L^f vv',$$

and $\bar{u} \wedge_L \bar{v}$ is represented by

$$C_R^g(u, C_R^g(v', u')) \equiv_L^f C_R^g(v, C_R^g(u', v')).$$

4.3 Normal forms and automatic structures

Let M be a Garside monoid, let $G = G(M)$ be the group of fractions of M , and let Δ be a fixed Garside element of M . Define the set of *simple elements* to be

$$\mathcal{S} = \{a \in M ; a \leq_L \Delta\} = \{a \in M ; a \leq_R \Delta\}.$$

By definition, \mathcal{S} is finite and generates M .

Let $\alpha \in M$. Then α can be uniquely written in the form

$$\alpha = a_1 a_2 \dots a_l,$$

where $a_1, a_2, \dots, a_l \in \mathcal{S}$, and

$$a_i = \Delta \wedge_L (a_i a_{i+1} \dots a_l) \quad \text{for all } 1 \leq i \leq l.$$

Such an expression of α is called the *normal form* of α .

Let $\alpha \in G$. Then α can be written in the form $\alpha = \beta^{-1}\gamma$, where $\beta, \gamma \in M$ (see Proposition 4.8, for instance). Obviously, we can also assume that $\beta \wedge_L \gamma = 1$. In that case β and γ are unique. Let $\beta = b_1 b_2 \dots b_p$ be the normal form of β and let $\gamma = c_1 c_2 \dots c_q$ be the normal form of γ . Then the expression

$$\alpha = b_p^{-1} \dots b_2^{-1} b_1^{-1} c_1 c_2 \dots c_q$$

is called the *normal form* of α .

There is another notion of normal form for the elements of G , called Δ -normal form, which is used, in particular, in several solutions to the conjugacy problem for G . It is defined as follows.

It is easily seen that there exists a permutation $\tau: \mathcal{S} \rightarrow \mathcal{S}$ such that $\Delta a \Delta^{-1} = \tau(a)$ for all $a \in \mathcal{S}$. Moreover, for all $a \in \mathcal{S}$, there exists $a^* \in \mathcal{S}$ such that $a^* a = \Delta$ (i.e. $a^{-1} = \Delta^{-1} a^*$). These two observations show that every $\alpha \in G$ can be written in the form $\alpha = \Delta^p \beta$, where $p \in \mathbb{Z}$ and $\beta \in M$. One can choose p to be maximal, and, in that case, β is unique. Let $b_1 b_2 \dots b_r$ be the normal form of β . Then the expression

$$\alpha = \Delta^p b_1 b_2 \dots b_r$$

is called the Δ -normal form of α .

Definition 4.11. A *finite state automaton* is a quintuple $\mathcal{A} = (Q, \mathcal{S}, T, A, q_0)$, where

- Q is a finite set, called the set of *states*;
- \mathcal{S} is a finite set, called the *alphabet*;
- T is a map $T: Q \times \mathcal{S} \rightarrow Q$, called the *transition function*;
- A is a subset of Q , called the set of *accepted states*;
- q_0 is an element of Q , called the *initial state*.

The *iterated transition function* is the map $T^*: Q \times \mathcal{S}^* \rightarrow Q$ defined by induction on the length of the second component as follows.

$$\begin{aligned} T^*(q, \epsilon) &= q, \\ T^*(q, x_1 x_2 \dots x_l) &= T(T^*(q, x_1 \dots x_{l-1}), x_l). \end{aligned}$$

The set

$$\mathcal{L}_{\mathcal{A}} = \{\omega \in \mathcal{S}^*; T^*(q_0, \omega) \in A\}$$

is called the *language recognized* by \mathcal{A} . A *regular language* is a language recognized by a finite state automaton.

Let G be a group generated by a finite set \mathcal{S} . Define the *word length* of an element $\alpha \in G$, denoted by $\text{lg}_{\mathcal{S}}(\alpha)$, to be the shortest length of a word in $(\mathcal{S} \sqcup \mathcal{S}^{-1})^*$ which represents α . The *distance* between two element $\alpha, \beta \in G$, denoted by $d_{\mathcal{S}}(\alpha, \beta)$, is the length of $\alpha^{-1}\beta$.

Let $\mathcal{L} \subset (\mathcal{S} \sqcup \mathcal{S}^{-1})^*$ be a language. We say that \mathcal{L} *represents* G if every element of G is represented by an element of \mathcal{L} . We say, furthermore, that \mathcal{L} has the *uniqueness property* if every element of G is represented by a unique element of \mathcal{L} . We say that \mathcal{L} is *symmetric* if $\mathcal{L}^{-1} = \mathcal{L}$, where $\mathcal{L}^{-1} = \{\omega^{-1}; \omega \in \mathcal{L}\}$. We say that \mathcal{L} is *geodesic* if $\text{lg}(\omega) = \text{lg}_{\mathcal{S}}(\bar{\omega})$ for all $\omega \in \mathcal{L}$. Let $\omega = x_1^{\epsilon_1} \dots x_l^{\epsilon_l} \in (\mathcal{S} \sqcup \mathcal{S}^{-1})^*$. For $t \in \mathbb{N}$ we set

$$\bar{\omega}(t) = \begin{cases} 1 & \text{if } t = 0, \\ \overline{x_1^{\epsilon_1} \dots x_t^{\epsilon_t}} & \text{if } 1 \leq t \leq l, \\ \bar{\omega} & \text{if } t \geq l. \end{cases}$$

Let c be a positive integer. We say that \mathcal{L} has the c -fellow traveler property if

$$d_{\mathcal{S}}(\bar{u}(t), \bar{v}(t)) \leq c \cdot d_{\mathcal{S}}(\bar{u}, \bar{v})$$

for all $u, v \in \mathcal{L}$ and all $t \in \mathbb{N}$.

A group G is said to be *automatic* if there exist a finite generating set $\mathcal{S} \subset G$, a regular language $\mathcal{L} \subset (\mathcal{S} \sqcup \mathcal{S}^{-1})^*$, and a constant $c > 0$, such that \mathcal{L} represents G and has the c -fellow traveler property. If, in addition, \mathcal{L}^{-1} has also the c -fellow traveler property, then G is said to be *biautomatic*. We say that G is *fully biautomatic* if \mathcal{L} is symmetric, and that G is *geodesically automatic* if \mathcal{L} is geodesic.

Biautomatic groups have many attractive properties. For instance, they have soluble word and conjugacy problems, and they have quadratic isoperimetric inequalities. We refer to [79] for a general exposition on the subject.

Theorem 4.12 (Charney [43], Dehornoy, Paris [70]). *Let M be a Garside monoid, and let $G = G(M)$ be the group of fractions of M . Let $\mathcal{L} \subset (\mathcal{S} \sqcup \mathcal{S}^{-1})^*$ be the language of normal forms. Then \mathcal{L} is regular, represents G , has the uniqueness property, has the 5-fellow traveler property, is symmetric, and is geodesic.*

Corollary 4.13. *Garside groups are fully geodesically biautomatic.*

Note. The language of Δ -normal forms is also regular and satisfies some fellow traveler property, and the language of inverses of Δ -normal forms satisfies the same fellow traveler property. So, Δ -normal forms determine another biautomatic structure on G . This was proved by Thurston [79] for the braid groups and by Charney [42] for all the spherical type Artin groups, and the same proof works in general for all Garside groups.

4.4 The conjugacy problem

Let G be a group and let \mathcal{S} be a finite generating set of G . A *solution to the conjugacy problem* for G is an algorithm which, for given $u, v \in (\mathcal{S} \sqcup \mathcal{S}^{-1})^*$, decides whether \bar{u} and \bar{v} are conjugate or not, where, for $w \in (\mathcal{S} \sqcup \mathcal{S}^{-1})^*$, \bar{w} denotes the element of G represented by w .

The first solution to the conjugacy problem for the braid groups was obtained by Garside [85]. Garside's algorithm was improved by El-Rifai and Morton [78], and this improvement was extended to Garside groups by Picantin [136]. Picantin's algorithm was improved by Franco and González-Meneses [81], then by Gebhardt [86], and now by Gebhardt and González-Meneses [87]. The algorithm that we present here is not the optimal one, but is probably the simplest one. It is based on the algorithm of [87].

Note. In addition to the above mentioned papers, there are several recent papers where the algorithms are analyzed, in particular to obtain the best possible complexity (see [22], [23], [24], [90], [115], [113], [114]). These analyses often lead to new and unexpected results on braid groups and, more generally, on Garside groups.

Let M be a Garside monoid, let $G = G(M)$ be its group of fractions, let Δ be a fixed Garside element, and let $\mathcal{S} = \{a \in M; a \leq_L \Delta\}$ be the set of simple elements. Recall that, for every $a \in \mathcal{S}$, there exists a unique $a^* \in \mathcal{S}$ such that $aa^* = \Delta$. Recall also that there exists a permutation $\tau: \mathcal{S} \rightarrow \mathcal{S}$ such that $\Delta a \Delta^{-1} = \tau(a)$ for all $a \in \mathcal{S}$.

Let $\alpha \in G$. Let $\alpha = \Delta^p a_1 a_2 \dots a_r$ be the Δ -normal form of α . The number p is called the *infimum* of α and is denoted by $\inf(\alpha)$, $p + r$ is called the *supremum* and is denoted by $\sup(\alpha)$, and r is called the *canonical length* and is denoted by $\|\alpha\|$. The above terminology comes from the fact that p is the greatest number n such that $\Delta^n \leq_L \alpha$, and $p + r$ is the smallest number n such that $\alpha \leq_L \Delta^n$. The (simple) element $\tau^p(a_1)$ is called the *initial factor* of α and is denoted by $i(\alpha)$, and a_r is called the *terminal factor* and is denoted by $t(\alpha)$. It is easily checked that $i(\alpha^{-1}) = t(\alpha)^*$. Let

$$\pi(\alpha) = i(\alpha) \wedge_L t(\alpha)^* = i(\alpha) \wedge_L i(\alpha^{-1}).$$

Define the *sliding* of α to be

$$S(\alpha) = \pi(\alpha)^{-1} \cdot \alpha \cdot \pi(\alpha).$$

Observe that $\|S(\alpha)\| \leq \|\alpha\|$.

For $\alpha, \beta \in G$, we use the notation $\alpha \sim \beta$ to mean that α is conjugate to β . Let $\alpha \in G$. Define the *sliding circuits* of α to be

$$SC(\alpha) = \{\beta \in G; \beta \sim \alpha \text{ and } S^m(\beta) = \beta \text{ for some } m \geq 1\}.$$

It is shown in [87] that the elements of $SC(\alpha)$ have minimal canonical length in the conjugacy class of α , but not all the elements of the conjugacy class of minimal canonical length belong to $SC(\alpha)$.

Clearly, if $\alpha \sim \beta$, then $SC(\alpha) = SC(\beta)$, and if $\alpha \not\sim \beta$, then $SC(\alpha) \cap SC(\beta) = \emptyset$. So, our solution to the conjugacy problem for G follows the following stages.

Input. Two elements $\alpha, \beta \in G$.

Stage 1. Calculate an element $\alpha_0 \in SC(\alpha)$ and an element $\beta_0 \in SC(\beta)$.

Stage 2. Calculate the whole set $SC(\alpha) = SC(\alpha_0)$ from α_0 .

Output. YES if $\beta_0 \in SC(\alpha)$, and NO otherwise.

In order to find an element of $SC(\alpha)$ we use the following which is easy to prove.

Lemma 4.14. *Let $\alpha \in G$. There exists $m, k \geq 1$ such that $S^{m+k}(\alpha) = S^k(\alpha)$. In particular, $S^k(\alpha) \in SC(\alpha)$.*

The key result for Stage 2 is the following.

Theorem 4.15 (Gebhardt, González-Meneses [87]). *Let $\alpha, \beta \in G$ and let $\gamma_1, \gamma_2 \in M$. If $\beta, \gamma_1^{-1}\beta\gamma_1$, and $\gamma_2^{-1}\beta\gamma_2$ are elements of $SC(\alpha)$, then $(\gamma_1 \wedge_L \gamma_2)^{-1}\beta(\gamma_1 \wedge_L \gamma_2)$ is also an element of $SC(\alpha)$.*

Corollary 4.16. *Let $\alpha, \beta, \gamma \in G$ such that β and $\gamma^{-1}\beta\gamma$ are elements of $\text{SC}(\alpha)$. Let $\gamma = \Delta^p c_1 c_2 \dots c_r$ be the Δ -normal form of γ . Set $\beta_0 = \Delta^{-p}\beta\Delta^p$, and $\beta_i = c_i^{-1}\beta_{i-1}c_i$ for $1 \leq i \leq r$. Then $\beta_i \in \text{SC}(\alpha)$ for all $0 \leq i \leq r$.*

Proof. We prove that $\beta_i \in \text{SC}(\alpha)$ by induction on i . It is easily seen that, if $\beta \in \text{SC}(\alpha)$, then $\Delta^{-1}\beta\Delta \in \text{SC}(\alpha)$. In particular, we have $\beta_0 = \Delta^{-p}\beta\Delta^p \in \text{SC}(\alpha)$.

Let $i > 0$. By induction, $\beta_{i-1} \in \text{SC}(\alpha)$. By the above observation, we have $\Delta^{-1}\beta_{i-1}\Delta \in \text{SC}(\alpha)$. On the other hand, we have

$$\gamma^{-1}\beta\gamma = (c_i c_{i+1} \dots c_r)^{-1} \beta_{i-1} (c_i c_{i+1} \dots c_r) \in \text{SC}(\alpha).$$

By definition of a normal form, we have $\Delta \wedge_L (c_i c_{i+1} \dots c_r) = c_i$. We conclude by Theorem 4.14 that $\beta_i = c_i^{-1}\beta_{i-1}c_i \in \text{SC}(\alpha)$. \square

From Corollary 4.15 we obtain the following which, together with Lemma 4.13, provides an algorithm to compute $\text{SC}(\alpha)$.

Corollary 4.17. *Let $\alpha \in G$. Let Ω_α be the graph defined by the following data.*

- *The set of vertices of Ω_α is $\text{SC}(\alpha)$.*
- *Two vertices $\beta, \beta' \in \text{SC}(\alpha)$ are joined by an edge if there exists $a \in \mathcal{S}$ such that $\beta' = a^{-1}\beta a$.*

Then Ω_α is connected.

5 Cohomology and the Salvetti complex

5.1 Cohomology

Let Γ be a Coxeter graph, let (W_Γ, S) be the Coxeter system of type Γ , and let (G_Γ, Σ) be the Artin system of type Γ . Let Γ_{ab} be the graph defined by the following data.

- S is the set of vertices of Γ ;
- two vertices $s, t \in S$ are joined by an edge if $m_{st} \neq +\infty$ and m_{st} is odd.

The following is easy to prove from the presentation of G_Γ .

Proposition 5.1. *Let d be the number of connected components of Γ_{ab} . Then the abelianization of G_Γ is a free abelian group of rank d . In particular, $H^1(G_\Gamma, \mathbb{Z}) \simeq \mathbb{Z}^d$.*

Now, assume that Γ is of spherical type, and recall the space N_Γ defined in Subsection 3.4. Except for Proposition 5.1, all the known results on the cohomology of G_Γ use the fact that $\pi_1(N_\Gamma) = G_\Gamma$ (see Theorem 3.14), and that N_Γ is a $K(\pi, 1)$ space (see Theorem 3.15). Recall that these two results imply that $H^*(G_\Gamma, A) = H^*(N_\Gamma, A)$ for any G_Γ -module A .

In [4] Arnol'd established the following properties on the cohomology of braid groups.

Theorem 5.2 (Arnol'd [4]). *Let $n \geq 2$.*

- (1) $H^0(\mathcal{B}_n, \mathbb{Z}) = H^1(\mathcal{B}_n, \mathbb{Z}) = \mathbb{Z}$, $H^q(\mathcal{B}_n, \mathbb{Z})$ is finite for all $q \geq 2$, and $H^q(\mathcal{B}_n, \mathbb{Z}) = 0$ for all $q \geq n$.
- (2) If n is even, then $H^q(\mathcal{B}_n, \mathbb{Z}) = H^q(\mathcal{B}_{n+1}, \mathbb{Z})$ for all $q \geq 0$.
- (3) $H^q(\mathcal{B}_n, \mathbb{Z}) = H^q(\mathcal{B}_{2q-2}, \mathbb{Z})$ for all $q \leq \frac{1}{2}n + 1$.

The study of the cohomology of the braid groups was continued by Fuks [84] who calculated the cohomology of \mathcal{B}_n with coefficients in $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$. Let $\mathcal{B}_\infty = \varinjlim \mathcal{B}_n$, where the limit is taken relative to the natural embeddings $\mathcal{B}_n \hookrightarrow \mathcal{B}_{n+1}$, $n \geq 2$.

Theorem 5.3 (Fuks [84]). (1) $H^*(\mathcal{B}_\infty, \mathbb{F}_2)$ is the exterior \mathbb{F}_2 -algebra generated by $\{a_{m,k}; m \geq 1 \text{ and } k \geq 0\}$ where $\deg a_{m,k} = 2^k(2^m - 1)$.

(2) The natural embedding $\mathcal{B}_n \hookrightarrow \mathcal{B}_\infty$ induces a surjective homomorphism $H^*(\mathcal{B}_\infty, \mathbb{F}_2) \rightarrow H^*(\mathcal{B}_n, \mathbb{F}_2)$ whose kernel is generated by the monomials

$$a_{m_1, k_1} a_{m_2, k_2} \dots a_{m_t, k_t}$$

such that

$$2^{m_1 + \dots + m_t + k_1 + \dots + k_t} > n.$$

Later on, the cohomology with coefficients in $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ (where p is an odd prime number) and the cohomology with coefficients in \mathbb{Z} were calculate by Cohen [53], Segal [143], and Vainštejn [149].

Theorem 5.4 (Cohen [53], Segal [143], Vainštejn [149]). (1) $H^*(\mathcal{B}_\infty, \mathbb{F}_p)$ is the tensor product of a polynomial algebra generated by $\{x_i; i \geq 0\}$, where $\deg x_i = 2p^{i+1} - 2$, and an exterior algebra generated by $\{y_j; j \geq 0\}$, where $\deg y_j = 2p^j - 1$.

(2) The natural embedding $\mathcal{B}_n \hookrightarrow \mathcal{B}_\infty$ induces a surjective homomorphism $H^*(\mathcal{B}_\infty, \mathbb{F}_p) \rightarrow H^*(\mathcal{B}_n, \mathbb{F}_p)$, whose kernel is generated by the monomials

$$x_{i_1} x_{i_2} \dots x_{i_s} y_{j_1} y_{j_2} \dots y_{j_t}$$

such that

$$2(p^{i_1+1} + \dots + p^{i_s+1} + p^{j_1} + \dots + p^{j_t}) > n.$$

Let $\beta_2: H^*(\mathcal{B}_n, \mathbb{F}_2) \rightarrow H^*(\mathcal{B}_n, \mathbb{F}_2)$ be the homomorphism defined by

$$\beta_2(a_{m,k}) = a_{m+1,0} a_{m,1} \dots a_{m,k-1}.$$

For an odd prime number p , let $\beta_p: H^*(\mathcal{B}_n, \mathbb{F}_p) \rightarrow H^*(\mathcal{B}_n, \mathbb{F}_p)$ be the homomorphism defined by

$$\beta_p(x_i) = y_{i+1}, \quad \beta_p(y_j) = 0.$$

Theorem 5.5 (Cohen [53], Vainštejn [149]). *Let $q \geq 2$. Then*

$$H^q(\mathcal{B}_n, \mathbb{Z}) = \bigoplus_p \beta_p(H^{q-1}(\mathcal{B}_n, \mathbb{F}_p)),$$

where the sum is over all primes p .

The integral cohomology of the Artin groups of type B and D were calculated by Goryunov [91] in terms of the cohomology groups of the braid groups.

Theorem 5.6 (Goryunov [91]). (1) Let $n \geq 2$, and let $q \geq 2$. Then

$$H^q(G_{B_n}, \mathbb{Z}) = \bigoplus_{i=0}^n H^{q-i}(\mathcal{B}_{n-i}, \mathbb{Z}).$$

(2) Let $n \geq 4$, and let $q \geq 2$. Then

$$H^q(G_{D_n}, \mathbb{Z}) = H^q(\mathcal{B}_n, \mathbb{Z}) \oplus \left(\bigoplus_{i=0}^{+\infty} \text{Ker } \gamma_{n-2i}^{q-2i} \right) \oplus \left(\bigoplus_{j=0}^{+\infty} H^{q-2j-3}(\mathcal{B}_{n-3j-3}, \mathbb{F}_2) \right),$$

where, for $k \geq 2$ and $j \geq 0$, $\gamma_k^j : H^j(\mathcal{B}_k, \mathbb{Z}) \rightarrow H^j(\mathcal{B}_{k-1}, \mathbb{Z})$ denotes the homomorphism induced by the inclusion $\mathcal{B}_{k-1} \hookrightarrow \mathcal{B}_k$.

Finally, the integral cohomology of the remainder irreducible Artin groups of spherical type were calculate by Salvetti in [142].

Theorem 5.7 (Salvetti [142]). The integral cohomology of the Artin groups of type $I_2(p)$ ($p = 2q \geq 5$), H_3 , H_4 , F_4 , E_6 , E_7 , and E_8 is given in the tables below.

	H^0	H^1	H^2	H^3	H^4
$I_2(2q)$	\mathbb{Z}	\mathbb{Z}^2	\mathbb{Z}	0	0
$I_2(2q + 1)$	\mathbb{Z}	\mathbb{Z}	0	0	0
H_3	\mathbb{Z}	\mathbb{Z}	\mathbb{Z}	\mathbb{Z}	0
H_4	\mathbb{Z}	\mathbb{Z}	0	$\mathbb{Z} \times \mathbb{Z}_2$	\mathbb{Z}
F_4	\mathbb{Z}	\mathbb{Z}^2	\mathbb{Z}^2	\mathbb{Z}^2	\mathbb{Z}
E_6	\mathbb{Z}	\mathbb{Z}	0	\mathbb{Z}_2	\mathbb{Z}_2
E_7	\mathbb{Z}	\mathbb{Z}	0	\mathbb{Z}_2	$\mathbb{Z}_2 \times \mathbb{Z}_2$
E_8	\mathbb{Z}	\mathbb{Z}	0	\mathbb{Z}_2	\mathbb{Z}_2

	H^5	H^6	H^7	H^8
E_6	\mathbb{Z}_6	\mathbb{Z}_3	0	0
E_7	$\mathbb{Z}_6 \times \mathbb{Z}_6$	$\mathbb{Z}_3 \times \mathbb{Z}_6 \times \mathbb{Z}$	\mathbb{Z}	0
E_8	$\mathbb{Z}_2 \times \mathbb{Z}_6$	$\mathbb{Z}_3 \times \mathbb{Z}_6$	$\mathbb{Z}_2 \times \mathbb{Z}_6 \times \mathbb{Z}$	\mathbb{Z}

Cohomology of the spherical type Artin groups.

Note. It is a direct consequence of [140] that N_Γ has the same homotopy type as a CW-complex of dimension n , where $n = |S|$. This implies that the cohomological dimension of G_Γ is $\leq n$, and, therefore, that $H^q(G_\Gamma, \mathbb{Z}) = 0$ for all $q > n$.

Note. Recall the space M_Γ of Subsection 3.4. The cohomology $H^*(M_\Gamma, \mathbb{Z})$ was calculated by Brieskorn in [31]. In particular, $H^*(M_\Gamma, \mathbb{Z})$ is torsion free and $H^n(M_\Gamma, \mathbb{Z}) \neq 0$. Let CG_Γ be the kernel of the canonical epimorphism $\theta: G_\Gamma \rightarrow W_\Gamma$. By [71] we have $H^n(M_\Gamma, \mathbb{Z}) = H^n(CG_\Gamma, \mathbb{Z})$, thus, by the above, $\text{cd}(G_\Gamma) = \text{cd}(CG_\Gamma) \geq n$, where $\text{cd}(G_\Gamma)$ denotes the cohomological dimension of G_Γ . We already know that $\text{cd}(G_\Gamma) \leq n$, thus $\text{cd}(G_\Gamma) = n$.

Note. The ring structure of $H^*(G_\Gamma, \mathbb{Z})$, where Γ is a Coxeter graph in the list of Theorem 5.7, was calculated in [111]. Some cohomologies with twisted coefficients were also considered. An interesting case is the cohomology over the module of Laurent polynomials $\mathbb{Q}[q^{\pm 1}]$ (resp. $\mathbb{Z}[q^{\pm 1}]$), because it determines the rational (resp. integral) cohomology of the Milnor fiber of the discriminant of type Γ (see [36]). For the case $\Gamma = A_n$ (i.e. G_Γ is the braid group \mathcal{B}_{n+1}), the $\mathbb{Q}[q^{\pm 1}]$ -cohomology was calculated by several people in several ways (see [82], [122], [52], [61]), and the $\mathbb{Z}[q^{\pm 1}]$ -cohomology was calculated by Callegaro in [37]. The $\mathbb{Q}[q^{\pm 1}]$ -cohomology for the other spherical type Artin groups was calculated in [62]. The $\mathbb{Z}[q^{\pm 1}]$ -cohomology for the exceptional cases was calculated in [39], and the top $\mathbb{Z}[q^{\pm 1}]$ -cohomology for all cases was calculated in [64].

Note. The cohomology of the non-spherical Artin groups is badly understood. Some calculations for the type \tilde{A}_n were done in [38].

We refer to [150] for a more detailed exposition on the cohomology of the braid groups and the Artin groups of spherical type, and turn to present the Salvetti complex (of a real hyperplane arrangement). This is the main tool in Salvetti's calculations of the cohomology of Artin groups (see [142]), but it can be used for other purposes. For instance, it can be also used to prove Theorems 3.14 and 3.15 (see [141] and [131]), and to produce a free resolution of \mathbb{Z} by $\mathbb{Z}[G_\Gamma]$ -modules (see Theorem 5.15).

5.2 Salvetti complex

Define a (*real*) *hyperplane arrangement* to be a finite family \mathcal{A} of linear hyperplanes of \mathbb{R}^n . For every $H \in \mathcal{A}$ we denote by $H_{\mathbb{C}}$ the hyperplane of \mathbb{C}^n having the same equation as H (i.e. $H_{\mathbb{C}} = \mathbb{C} \otimes H$), and we set

$$M(\mathcal{A}) = \mathbb{C}^n \setminus \left(\bigcup_{H \in \mathcal{A}} H_{\mathbb{C}} \right).$$

Note that $M(\mathcal{A})$ is an open connected subvariety of \mathbb{C}^n .

The arrangement \mathcal{A} subdivides \mathbb{R}^n into *facets*. We denote by $\mathcal{F}(\mathcal{A})$ the set of all facets. The *support* of a facet $F \in \mathcal{F}(\mathcal{A})$ is the linear subspace $\langle F \rangle$ spanned by F .

We denote by \bar{F} the closure of a facet F . We order $\mathcal{F}(\mathcal{A})$ by $F \leq G$ if $F \subset \bar{G}$. The set $\mathcal{F}(\mathcal{A})$ has a unique minimal element: $\bigcap_{H \in \mathcal{A}} H$. The maximal elements of $\mathcal{F}(\mathcal{A})$ are the facets of codimension 0, and they are called *chambers*. The set of all chambers is denoted by $\mathcal{C}(\mathcal{A})$.

Set

$$\mathcal{X} = \{(F, C) \in \mathcal{F}(\mathcal{A}) \times \mathcal{C}(\mathcal{A}) ; F \leq C\}.$$

We partially order \mathcal{X} as follows. For $F \in \mathcal{F}(\mathcal{A})$ we set $\mathcal{A}_F = \{H \in \mathcal{A}; H \supset F\}$. For $F \in \mathcal{F}(\mathcal{A})$ and $C \in \mathcal{C}(\mathcal{A})$ we denote by C_F the chamber of \mathcal{A}_F which contains C . We set

$$(F_1, C_1) \leq (F_2, C_2) \quad \text{if} \quad F_1 \leq F_2 \quad \text{and} \quad (C_1)_{F_2} = (C_2)_{F_2}.$$

(See Figure 12.)

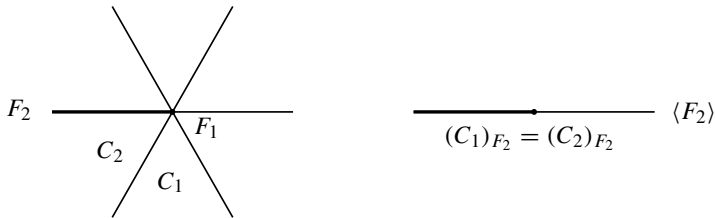


Figure 12. Order in \mathcal{X} .

Define the *Salvetti complex* $\text{Sal}(\mathcal{A})$ of \mathcal{A} to be the (geometric realization of the) flag complex of (\mathcal{X}, \leq) . That is, to every chain $X_0 < X_1 < \dots < X_d$ in \mathcal{X} corresponds a simplex $\Delta(X_0, X_1, \dots, X_d)$ of $\text{Sal}(\mathcal{A})$, and every simplex of $\text{Sal}(\mathcal{A})$ is of this form.

Theorem 5.8 (Salvetti [140]). *The simplicial complex $\text{Sal}(\mathcal{A})$ is homotopy equivalent to $M(\mathcal{A})$.*

We turn now to describe a cellular decomposition of $\text{Sal}(\mathcal{A})$ which is the version which is usually used in the literature.

Without loss of generality, we can and do assume that \mathcal{A} is *essential*, that is, $\bigcap_{H \in \mathcal{A}} H = \{0\}$. Consider the unit sphere $\mathbb{S}^{n-1} = \{\mathbf{x} \in \mathbb{R}^n; \|\mathbf{x}\| = 1\}$. The arrangement \mathcal{A} determines a cellular decomposition of \mathbb{S}^{n-1} : to each facet $F \in \mathcal{F}(\mathcal{A}) \setminus \{0\}$ corresponds the open cell $F \cap \mathbb{S}^{n-1}$, and each cell is of this form. This cellular decomposition is regular in the sense that the closure of a cell is a closed disk. Hence, one can consider the barycentric subdivision. For each facet $F \in \mathcal{F}(\mathcal{A}) \setminus \{0\}$ we fix a point $\mathbf{x}(F) \in F \cap \mathbb{S}^{n-1}$. To each chain $\{0\} \neq F_0 < F_1 < \dots < F_d$ in $\mathcal{F}(\mathcal{A}) \setminus \{0\}$ corresponds a simplex $\Delta(F_0, F_1, \dots, F_d)$ whose vertices are $\mathbf{x}(F_0), \mathbf{x}(F_1), \dots, \mathbf{x}(F_d)$, and every simplex of \mathbb{S}^{n-1} is of this form. So, the simplicial decomposition of \mathbb{S}^{n-1} is the flag complex of $(\mathcal{F}(\mathcal{A}) \setminus \{0\}, \leq)$.

We extend the above simplicial decomposition of \mathbb{S}^{n-1} to a simplicial decomposition of the n -disk $\mathbb{B}^n = \{\mathbf{x} \in \mathbb{R}^n; \|\mathbf{x}\| \leq 1\}$, adding a single vertex $\mathbf{x}(0) = 0$. That

is, we view \mathbb{B}^n as the cone of \mathbb{S}^{n-1} . Now, to any chain $F_0 < F_1 < \dots < F_d$ in $\mathcal{F}(\mathcal{A})$ corresponds a simplex $\Delta(F_0, F_1, \dots, F_d)$ of \mathbb{B}^n (here we may have $F_0 = 0$), and every simplex of \mathbb{B}^n is of this form. Note that this simplicial decomposition of \mathbb{B}^n is the flag complex of $(\mathcal{F}(\mathcal{A}), \leq)$.

Let $F_b \in \mathcal{F}(\mathcal{A})$ be a facet. It can be easily checked that the union of the simplices of the form $\Delta(F_0, F_1, \dots, F_d)$ with $F_b = F_0 < F_1 < \dots < F_d$ is a closed disk whose dimension is equal to $\text{codim } F_b$. Its interior is denoted by $U(F_b)$. So, the set $\{U(F); F \in \mathcal{F}(\mathcal{A})\}$ forms a cellular decomposition of \mathbb{B}^n called the *dual decomposition*.

Example. Let \mathcal{A} be a collection of 3 lines in \mathbb{R}^2 (see Figure 13). The poset $\mathcal{F}(\mathcal{A})$ contains 6 chambers, 6 facets of dimension 1 (half-lines), and 0. The dual decomposition of $\mathbb{B}^2 = \mathbb{D}$ has 6 vertices, 6 edges, and one 2-cell.

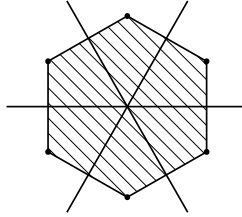


Figure 13. A dual decomposition.

Let $X_b = (F_b, C_b)$ be contained in \mathcal{X} . We denote by $\bar{U}(X_b)$ the union of the simplices $\Delta(X_0, X_1, \dots, X_d)$ of $\text{Sal}(\mathcal{A})$ such that $X_b = X_0 < X_1 < \dots < X_d$. One can show (with some effort) that, for every $F \geq F_b$, there exists a unique chamber $C \in \mathcal{C}(\mathcal{A})$ such that $F \leq C$ and $(F_b, C_b) \leq (F, C)$. This implies that $\bar{U}(X_b)$ is homeomorphic to $\bar{U}(F_b)$ via the map $(F, C) \mapsto x(F), F_b \leq F$. Hence, $\bar{U}(X_b)$ is a closed disk whose dimension is equal to $\text{codim } F_b$. We denote by $U(X_b)$ the interior of $\bar{U}(X_b)$. So, $\{U(X); X \in \mathcal{X}\}$ forms a (regular) cell decomposition of $\text{Sal}(\mathcal{A})$.

0-skeleton. For $C \in \mathcal{C}(\mathcal{A})$, we set $\omega(C) = U(C, C) = \bar{U}(C, C)$. Then the 0-skeleton of $\text{Sal}(\mathcal{A})$ is

$$\text{Sal}_0(\mathcal{A}) = \{\omega(C) ; C \in \mathcal{C}(\mathcal{A})\}.$$

1-skeleton. Let $F \in \mathcal{F}(\mathcal{A})$ be a facet of codimension 1. There are exactly two chambers $C, D \in \mathcal{C}(\mathcal{A})$ such that $F \leq C$ and $F \leq D$. Then there are two edges, $U(F, C)$ and $U(F, D)$, joining $\omega(C)$ and $\omega(D)$ in the 1-skeleton of $\text{Sal}(\mathcal{A})$ (see Figure 14). We use the convention that $U(F, C)$ is endowed with an orientation which goes from $\omega(C)$ to $\omega(D)$.

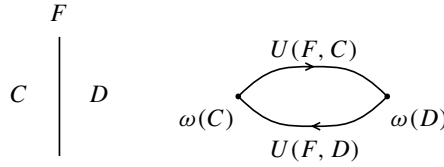


Figure 14. 1-skeleton of $\text{Sal}(\mathcal{A})$.

2-skeleton. Let $F_b \in \mathcal{F}(\mathcal{A})$ be a facet of codimension 2, and let $C_b \in \mathcal{C}(\mathcal{A})$ such that $F_b \leq C_b$. Let $C_0 = D_0 = C_b, C_1, \dots, C_l = D_l, \dots, D_l$ be the chambers $C \in \mathcal{C}(\mathcal{A})$ such that $F_b \leq C$, arranged like in Figure 15. Let $F_1, \dots, F_l, G_1, \dots, G_l$ be the facets $F \in \mathcal{F}(\mathcal{A})$ of codimension 1 such that $F_b \leq F$, arranged like in Figure 15. Set $a_i = U(F_i, C_{i-1})$ and $b_i = U(G_i, D_{i-1})$ for $1 \leq i \leq l$. Then $U(F_b, C_b)$ is a 2-disk whose boundary is $(a_1 a_2 \dots a_l)(b_1 b_2 \dots b_l)^{-1}$.

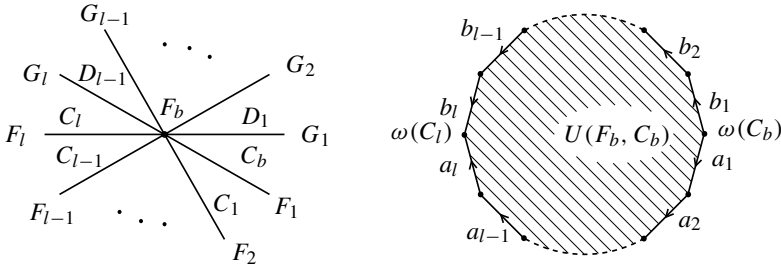


Figure 15. 2-skeleton of $\text{Sal}(\mathcal{A})$.

Let Γ be a Coxeter graph of spherical type, let (W_Γ, S) be the Coxeter system of type Γ , and let (G_Γ, Σ) be the Artin system of type Γ . Recall the set $\Pi = \{e_s; s \in S\}$ of simple roots, the linear space $V = \bigoplus_{s \in S} \mathbb{R}e_s$, and the canonical bilinear form $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$. Recall also from Theorem 3.8 that $\langle \cdot, \cdot \rangle$ is positive definite, and that $W = W_\Gamma$ can be viewed as a finite subgroup of $O(V) = O(V, \langle \cdot, \cdot \rangle)$ generated by reflections.

Let \mathcal{A}_Γ denote the set of reflecting hyperplanes of W . Then $M_\Gamma = M(\mathcal{A}_\Gamma)$, the group W_Γ acts freely on M_Γ , $N_\Gamma = M_\Gamma / W_\Gamma$, and $\pi_1(N_\Gamma) = G_\Gamma$ (see Subsection 3.4).

Fix a (base) chamber $C_b \in \mathcal{C}(\mathcal{A}_\Gamma)$. A hyperplane $H \in \mathcal{A}_\Gamma$ is called a *wall* of C_b if $\text{codim}(\overline{C_b} \cap H) = 1$. The following is proved in [28].

Proposition 5.9. (1) C_b is a simplicial cone.

(2) Let H_1, \dots, H_n be the walls of C_b , and, for $1 \leq i \leq n$, let s_i be the orthogonal reflection with respect to H_i . Then, up to conjugation, $S = \{s_1, \dots, s_n\}$ is the Coxeter generating set of W .

For $T \subset S$ we denote by W_T the subgroup of W generated by T , and by Γ_T the full subgraph of Γ spanned by T . It is a well-know fact (see [28], for example) that (W_T, T) is the Coxeter system of type Γ_T . The *Coxeter complex* of (W, S) is defined to be the set

$$\text{Cox}_\Gamma = \{wW_T ; T \subset S \text{ and } w \in W\}$$

ordered by the reverse inclusion (i.e. $w_1W_{T_1} \leq w_2W_{T_2}$ if $w_1W_{T_1} \supset w_2W_{T_2}$).

We fix a base chamber C_b and we take $S = \{s_1, \dots, s_n\}$ as in Proposition 5.9. For each $s \in S$ we denote by H_s the hyperplane fixed by s . So, $\{H_s; s \in S\}$ is the set of walls of C_b . Since C_b is a simplicial cone, for every $T \subset S$ there exists a unique facet $F(T) \in \mathcal{F}(\mathcal{A}_\Gamma)$ such that $F(T) \leq C_b$ and $\langle F(T) \rangle = \bigcap_{s \in T} H_s$. The proof of the following can be found in [28].

Proposition 5.10. *The map*

$$\begin{aligned} \psi : \text{Cox}_\Gamma &\rightarrow \mathcal{F}(\mathcal{A}_\Gamma), \\ wW_T &\mapsto wF(T) \end{aligned}$$

is well-defined and is an isomorphism of ordered sets.

Now, the following Lemmas 5.11 and 5.12 are used to describe the poset \mathcal{X} in terms of Coxeter complexes.

Lemma 5.11 (Bourbaki [28]). *Let $T \subset S$ and $w \in W$. Then wW_T has a smallest element u for the order \leq_L (defined in Subsection 3.2). That is, for all $w' \in wW_T$ there exists a unique $v' \in W_T$ such that $w' = uv'$ and $\text{lg}_S(w') = \text{lg}_S(u) + \text{lg}_S(v')$.*

The smallest element of wW_T is denoted by $u = \min_T(w)$, and such an element is called *T-minimal*. The set of *T-minimal* elements is denoted by $\text{Min}(T)$. For $w \in W$, we denote by $\pi_T(w)$ the element $v \in W_T$ such that $w = \min_T(w) \cdot v$.

The proof of the following is left to the reader.

Lemma 5.12. *Let C_b be a base chamber, let $T \subset S$, and let $F = F(T)$. Let $w_1, w_2 \in W$. We have $(w_1C_b)_F = (w_2C_b)_F$ if and only if $\pi_T(w_1) = \pi_T(w_2)$.*

Set

$$\widehat{\text{Cox}}_\Gamma = \{(T, w) ; w \in W \text{ and } T \subset S\}.$$

Let \leq be the partial order on $\widehat{\text{Cox}}_\Gamma$ defined by

$$\begin{aligned} (T_1, w_1) \leq (T_2, w_2) \quad &\text{if } T_1 \supset T_2, \min_{T_1}(w_1) = \min_{T_1}(w_2), \\ &\text{and } \pi_{T_2}(w_1) = \pi_{T_2}(w_2). \end{aligned}$$

Note that the conditions “ $T_1 \supset T_2$ and $\min_{T_1}(w_1) = \min_{T_1}(w_2)$ ” are equivalent to the condition $w_1W_{T_1} \supset w_2W_{T_2}$, and, by Lemma 5.12, the condition $\pi_{T_2}(w_1) = \pi_{T_2}(w_2)$ is equivalent to the condition $(w_1C_b)_{F(T_2)} = (w_2C_b)_{F(T_2)}$. So:

Theorem 5.13. *The map*

$$\begin{aligned} \hat{\psi} : \widehat{\text{Cox}}_\Gamma &\rightarrow \mathcal{X}(\mathcal{A}_\Gamma), \\ (T, w) &\mapsto (wF(T), wC_b) \end{aligned}$$

is well-defined and is an isomorphism of posets.

For $(T, w) \in \widehat{\text{Cox}}_\Gamma$ we set $U(T, w) = U(\hat{\psi}(T, w))$. So, $\{U(T, w); (T, w) \in \widehat{\text{Cox}}_\Gamma\}$ is a cellular decomposition of $\text{Sal}(\mathcal{A}_\Gamma)$. Moreover, the dimension of $U(T, w)$ is $|T|$ for all $(T, w) \in \widehat{\text{Cox}}_\Gamma$.

The Coxeter group W acts on $\widehat{\text{Cox}}_\Gamma$ by

$$u \cdot (T, w) = (T, uw) \quad \text{for } (T, w) \in \widehat{\text{Cox}}_\Gamma \text{ and } u \in W.$$

It turns out that this action preserves the order of $\widehat{\text{Cox}}_\Gamma$ and induces a cellular action on $\text{Sal}(\mathcal{A}_\Gamma)$ defined by

$$u \cdot U(T, w) = U(T, uw) \quad \text{for } (T, w) \in \widehat{\text{Cox}}_\Gamma \text{ and } u \in W.$$

Theorem 5.14 (Salvetti [142]). *There exists an embedding $\text{Sal}(\mathcal{A}_\Gamma) \hookrightarrow M_\Gamma$ and a (strong) retracting deformation of M_Γ onto $\text{Sal}(\mathcal{A}_\Gamma)$ that are equivariant under the action of W . In particular, there exists an embedding $\text{Sal}(\mathcal{A}_\Gamma)/W \hookrightarrow M_\Gamma/W = N_\Gamma$ and a (strong) retracting deformation of N_Γ onto $\text{Sal}(\mathcal{A}_\Gamma)/W$.*

To each $T \subset S$ corresponds a unique cell $U_N(T)$ of $\text{Sal}(\mathcal{A}_\Gamma)/W$ of dimension $|T|$. This cell is the orbit of $U(T, w)$ for all $w \in W$. Every cell of $\text{Sal}(\mathcal{A}_\Gamma)/W$ is of this form.

The 0-skeleton of $\text{Sal}(\mathcal{A}_\Gamma)/W$ contains a unique vertex, $\omega_N = U_N(\emptyset)$. For every $s \in S$ there is an edge $U_N(s)$ in $\text{Sal}(\mathcal{A}_\Gamma)/W$ and each edge is of this form. For every pair $\{s, t\} \subset S$ there is a 2-cell $U_N(s, t)$ in $\text{Sal}(\mathcal{A}_\Gamma)/W$ whose boundary is

$$\text{prod}(U_N(s), U_N(t) : m_{st}) \cdot \text{prod}(U_N(t), U_N(s) : m_{st})^{-1},$$

and every 2-cell is of this form. Note that the 2-skeleton of $\text{Sal}(\mathcal{A}_\Gamma)/W$ is equal to the 2-cell complex associated to the standard presentation of G_Γ . This gives an alternative proof to Theorems 2.2 and 3.14.

For $0 \leq q \leq |S|$, set

$$C_q(G_\Gamma) = \bigoplus_{\substack{T \subset S \\ |T|=q}} \mathbb{Z}[G_\Gamma] \cdot E_T,$$

the free $\mathbb{Z}[G_\Gamma]$ -module freely spanned by $\{E_T; T \subset S \text{ and } |T| = q\}$. We fix a total order $S = \{s_1, \dots, s_n\}$ on S and we define $d : C_q(G_\Gamma) \rightarrow C_{q-1}(G_\Gamma)$ as follows. Let $T = \{s_{i_1}, \dots, s_{i_q}\} \subset S, i_1 < \dots < i_q$. Then

$$dE_T = \sum_{j=1}^q (-1)^{j-1} \left(\sum_{\substack{u \in W_T \\ u \in \text{Min}(T \setminus \{s_{i_j}\})}} (-1)^{\text{lg}_S(u)} \kappa(u) \right) \cdot E_{T \setminus \{s_{i_j}\}},$$

where $\kappa: W \rightarrow G_\Gamma$ is the set-section of the canonical epimorphism $\theta: G_\Gamma \rightarrow W$ defined in Subsection 3.3.

Theorem 5.15 (De Concini, Salvetti [63], Squier [145]). *The complex $(C_*(G_\Gamma), d)$ is a free resolution of \mathbb{Z} by $\mathbb{Z}[G_\Gamma]$ -modules.*

Note. Squier's proof of Theorem 5.15 does not use the Salvetti complexes at all and is independent from the proof of De Concini and Salvetti.

6 Linear representations

The existence (or non-existence) of faithful linear representations of the braid groups was one of the major problems in the field. This problem was solved by Bigelow [17] and Krammer [106] in 2000. Their representation, which is known now as the LKB representation, was right afterwards extended to the Artin groups of type D_n ($n \geq 4$) and E_k ($k = 6, 7, 8$) by Digne [74], Cohen, and Wales [51], and to all Artin groups of small type in [133]. The representations of Digne, Cohen and Wales were proved to be faithful. Hence, since any spherical type Artin group embeds in a direct product of Artin groups of type A_n ($n \geq 1$), D_n ($n \geq 4$), and E_k ($k = 6, 7, 8$) (see [57]), any Artin group of spherical type is linear. The extension to the non-spherical type Artin groups gives rise to a linear representation over an infinite dimensional vector space, so it cannot be used for proving that these groups are linear. However, these representations are useful tools to study the non-spherical type Artin groups. In particular, they are the main tool in the proof of Theorem 3.9.

In Subsection 6.1 we present the algebraic approach to the LKB representations as constructed in [133] for the Artin groups of small type. Subsection 6.2 is dedicated to the topological construction of the LKB representations. Curiously, this topological point of view is known only for the braid groups.

6.1 Algebraic approach

Let Γ be a Coxeter graph, let $M = (m_{st})_{s,t \in S}$ be the Coxeter matrix of Γ , let (W_Γ, S) be the Coxeter system of type Γ , let (G_Γ, Σ) be the Artin system of type Γ , and let G_Γ^+ be the Artin monoid of type Γ .

We say that Γ is of *small type* if $m_{st} \leq 3$ for all $s, t \in S$, $s \neq t$, and we say that Γ is *without triangle* if there is no triple $\{s, t, r\}$ in S such that $m_{st} = m_{tr} = m_{rs} = 3$. We assume from now on that Γ is of small type and without triangle.

Recall from Subsection 3.2 the set $\Pi = \{e_s; s \in S\}$ of simple roots, the space $V = \bigoplus_{s \in S} \mathbb{R}e_s$, the canonical bilinear form $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$, and the root system $\Phi = \{we_s; s \in S \text{ and } w \in W\}$. Recall also that we have the disjoint union $\Phi = \Phi_+ \sqcup \Phi_-$, where Φ_+ is the set of positive roots and Φ_- is the set of negative roots (see Proposition 3.4).

Set $\mathcal{E} = \{u_f; f \in \Phi_+\}$ an abstract set in one-to-one correspondence with Φ_+ , and $\mathbb{K} = \mathbb{Q}(x, y)$. Note that \mathcal{E} is finite if and only if Γ is of spherical type. We denote by \mathcal{V} the \mathbb{K} -vector space having \mathcal{E} as a basis.

For all $s \in S$ we define a linear transformation $\varphi_s: \mathcal{V} \rightarrow \mathcal{V}$ by

$$\varphi_s(u_f) = \begin{cases} 0 & \text{if } f = e_s, \\ u_f & \text{if } \langle e_s, f \rangle = 0, \\ y \cdot u_{f-ae_s} & \text{if } \langle e_s, f \rangle = a > 0 \text{ and } f \neq e_s, \\ (1-y) \cdot u_f + u_{f+ae_s} & \text{if } \langle e_s, f \rangle = -a < 0. \end{cases}$$

The following is easy to prove.

Lemma 6.1. *The mapping $\sigma_s \mapsto \varphi_s$, $s \in S$, induces a homomorphism of monoids $\varphi: G_\Gamma^+ \rightarrow \text{End}(\mathcal{V})$.*

For all $s \in S$ and all $f \in \Phi_+$ we choose a polynomial $T(s, f) \in \mathbb{Q}[y]$ and we define $\Phi_s: \mathcal{V} \rightarrow \mathcal{V}$ by

$$\Phi_s(u_f) = \varphi_s(u_f) + x \cdot T(s, f) \cdot u_{e_s}.$$

Now, we have:

Theorem 6.2 (Paris [133]). *There exists a choice of polynomials $T(s, f)$, $s \in S$ and $f \in \Phi_+$, such that the mapping $\sigma_s \mapsto \Phi_s$, $s \in S$, induces a homomorphism $\Phi: G_\Gamma^+ \rightarrow \text{GL}(\mathcal{V})$.*

Theorem 6.3 (Paris [133]). *The above defined homomorphism $\Phi: G_\Gamma^+ \rightarrow \text{GL}(\mathcal{V})$ is injective.*

Corollary 6.4 (Paris [133]). *The natural homomorphism $\iota: G_\Gamma^+ \rightarrow G_\Gamma$ is injective.*

Proof. Since G_Γ is the group of fractions of G_Γ^+ , there exists a unique homomorphism $\hat{\Phi}: G_\Gamma \rightarrow \text{GL}(\mathcal{V})$ such that $\Phi = \hat{\Phi} \circ \iota$. Since Φ is injective, we conclude that ι is also injective. \square

Corollary 6.5 (Bigelow [17], Krammer [106], Digne [74], Cohen, Wales [51]). *Suppose that Γ is of spherical type. Let $\hat{\Phi}: G_\Gamma \rightarrow \text{GL}(\mathcal{V})$ be the homomorphism induced by Φ . Then $\hat{\Phi}$ is injective.*

Proof. Let $\alpha \in \text{Ker } \hat{\Phi}$. By Proposition 4.8, α can be written in the form $\alpha = \beta^{-1}\gamma$, with $\beta, \gamma \in G_\Gamma^+$. We have $1 = \hat{\Phi}(\alpha) = \Phi(\beta)^{-1}\Phi(\gamma)$, thus $\Phi(\beta) = \Phi(\gamma)$. Since Φ is injective, it follows that $\beta = \gamma$, thus $\alpha = \beta^{-1}\gamma = 1$. \square

Note. It is shown in [133] that any Artin monoid G_Γ^+ can be embedded in an Artin monoid G_Ω^+ , where Ω is of small type without triangle. Moreover, if Γ is of spherical

type, then Ω can be chosen to be of spherical type (see also [57], [89], [58], [41]). So, Corollary 6.4 implies that $\iota: G_\Gamma^+ \rightarrow G_\Gamma$ is injective for all Coxeter graphs Γ , and Corollary 6.5 implies that all the Artin groups of spherical type are linear.

Note. It is shown in [121] that if Γ is of type A_n, D_n, E_k ($k = 6, 7, 8$), then the image of $\hat{\Phi}$ is Zariski dense in $GL(\mathcal{V})$. In particular, this shows that $\hat{\Phi}$ is irreducible (see also [155], [120], [50]).

Note. The proof of Theorem 6.3 given in [133] is largely inspired by Krammer’s proof of the same theorem for the braid groups [106]. A new, short, and elegant proof can be found now in [95].

6.2 Topological approach

Now, we give a topological interpretation of the representation $\hat{\Phi}: G_\Gamma \rightarrow GL(\mathcal{V})$ in the case $\Gamma = A_{n-1}$, that is, when $G_\Gamma = \mathcal{B}_n$ is the braid group on n strands. Such an interpretation is unknown for the other Artin groups.

Let M be a connected CW-complex, let $G = \pi_1(M)$, and let R be a (right) $\mathbb{Z}[G]$ -module. Let \tilde{M} be the universal cover of M . The action of G on \tilde{M} induces an action of G on the group $C_q(\tilde{M})$ of (cellular) q -chains of \tilde{M} , and this action makes $C_q(\tilde{M})$ a module over the group ring $\mathbb{Z}[G]$. It is also easily seen that the boundary maps $\partial: C_q(\tilde{M}) \rightarrow C_{q-1}(\tilde{M})$ are $\mathbb{Z}[G]$ -module homomorphisms. We define $C_q(M, R)$ to be $R \otimes_{\mathbb{Z}[G]} C_q(\tilde{M})$. These groups form a chain complex with boundary map $\text{Id} \otimes \partial$. The homology groups $H_q(M, R)$ of this chain complex are the *homology groups of M with local coefficients R* .

Now, for $n \geq 1$, M_n denotes the space of ordered configurations of n points in \mathbb{C} , and $N_n = M_n/\text{Sym}_n$ denotes the space of (unordered) configurations of n points in \mathbb{C} (see Section 2). Let $n, m \geq 2$. By [80], the map

$$p_{n,m}: M_{n+m} \rightarrow M_n$$

$$(z_1, \dots, z_n, z_{n+1}, \dots, z_{n+m}) \mapsto (z_1, \dots, z_n)$$

is a locally trivial fiber bundle which admits a cross-section. The fiber of $p_{n,m}$ is as follows. Set

$$H_{i j} = \{w \in \mathbb{C}^m ; w_i = w_j\} \quad \text{for } 1 \leq i < j \leq m,$$

$$K_{i k} = \{w \in \mathbb{C}^m ; w_i = k\} \quad \text{for } 1 \leq i \leq m \text{ and } 1 \leq k \leq n.$$

Set

$$X_{n,m} = \mathbb{C}^m \setminus \left(\left(\bigcup_{i < j} H_{i j} \right) \cup \left(\bigcup_{\substack{1 \leq i \leq m \\ 1 \leq k \leq n}} K_{i k} \right) \right).$$

Then

$$p_{n,m}^{-1}(1, 2, \dots, n) = \{(1, 2, \dots, n)\} \times X_{n,m}.$$

Let $\text{Sym}_n \times \text{Sym}_m$ act on M_{n+m} , Sym_n acting by permutations on the first n coordinates, and Sym_m acting on the last m ones. Set

$$\begin{aligned} N_{n,m} &= M_{n+m}/(\text{Sym}_n \times \text{Sym}_m), \\ Y_{n,m} &= X_{n,m}/\text{Sym}_m. \end{aligned}$$

Then $p_{n,m}$ induces a locally trivial fiber bundle $\bar{p}_{n,m}: N_{n,m} \rightarrow N_n$ whose fiber is $Y_{n,m}$.

For $z \in \mathbb{C}^n$ we set

$$\|z\|_\infty = \max\{|z_i|; 1 \leq i \leq n\}.$$

It is easily checked that the map

$$\begin{aligned} \kappa: M_n &\rightarrow M_{n+m}, \\ z &\mapsto (z, \|z\|_\infty + 1, \|z\|_\infty + 2, \dots, \|z\|_\infty + m) \end{aligned}$$

is a well-defined cross-section of $p_{n,m}$ which is equivariant by the action of Sym_n , thus it induces a cross-section $\bar{\kappa}: N_n \rightarrow N_{n,m}$ of $\bar{p}_{n,m}$. By the homotopy long exact sequence of a fiber bundle (see Theorem 2.9), we conclude that $\pi_1(N_{n,m})$ can be written as a semi-direct product $\pi_1(N_{n,m}) = \pi_1(Y_{n,m}) \rtimes \mathcal{B}_n$.

Set $G_{n,m} = \pi_1(Y_{n,m})$. We consider $G_{n,m}$ as a subgroup of $\pi_1(N_{n,m})$ which, in its turn, is viewed as a subgroup of $\pi_1(N_{n+m}) = \mathcal{B}_{n+m}$. It is easily seen that $G_{n,m}$ is generated by the set

$$\{\sigma_k; n+1 \leq k \leq n+m\} \cup \{\delta_{i,k}; 1 \leq i \leq n \text{ and } n+1 \leq k \leq n+m\},$$

where $\delta_{i,k}$ is the pure braid defined in Theorem 2.3. Let b be the homology class of σ_{n+1} in $H_1(G_{n,m}) = H_1(Y_{n,m})$, and let a_i be the homology class of $\delta_{i,n+1}$, $1 \leq i \leq n$. The proof of the following is left to the reader.

Proposition 6.6. $H_1(Y_{n,m}) = H_1(G_{n,m})$ is a free abelian group freely generated by $\{b, a_1, a_2, \dots, a_n\}$.

Let $\bar{\rho}: H_1(G_{n,m}) \rightarrow \mathbb{Q}(x, y)^*$ be the homomorphism which sends a_i to x for all $1 \leq i \leq n$, and sends b to y . Let $\rho: G_{n,m} \rightarrow \mathbb{Q}(x, y)^*$ be the composition of the natural projection $G_{n,m} \rightarrow H_1(G_{n,m})$ with $\bar{\rho}$. This homomorphism makes $\mathbb{Q}(x, y)$ a $\mathbb{Z}[G_{n,m}]$ -module that we denote by Γ_ρ .

The proof of the following is also left to the reader.

Proposition 6.7. The kernel of ρ is invariant under the action of \mathcal{B}_n , and \mathcal{B}_n acts trivially on the quotient $G_{n,m}/\text{Ker } \rho \simeq \mathbb{Z} \times \mathbb{Z}$.

From Proposition 6.7 follows that the fibration $\bar{p}_{n,m}: N_{n,m} \rightarrow N_n$ induces a monodromy representation $\Phi_{n,m}: \mathcal{B}_n \rightarrow \text{Aut}_{\mathbb{Q}(x,y)}(H_*(Y_{n,m}, \Gamma_\rho))$.

The following was announced by Krammer [105], [106], and proved in [17] (see also [130]).

Theorem 6.8 (Bigelow [17]). *The homomorphism*

$$\Phi_{n,2}: \mathcal{B}_n \rightarrow \text{Aut}_{\mathbb{Q}(x,y)}(H_2(Y_{n,2}, \Gamma_\rho))$$

coincides with the representation $\hat{\Phi}: G_{A_{n-1}} \rightarrow \text{GL}(\mathcal{V})$ defined in Subsection 6.1.

Note. The representation $\hat{\Phi}: G_{A_{n-1}} \rightarrow \text{GL}(\mathcal{V})$ also coincides with the representation studied by Lawrence in [112]. Lawrence’s construction is also geometric. It slightly differs from the one presented above, but I do not know exactly how to relate them without the formulas.

Note. It is announced in [153] that $\Phi_{n,m}$ is faithful for all $m \geq 2$, and it is announced in [48] that $\Phi_{n,m}: \mathcal{B}_n \rightarrow \text{Aut}_{\mathbb{Q}(x,y)}(H_m(Y_{n,m}, \Gamma_\rho))$ is irreducible for all $m \geq 2$.

7 Geometric representations

7.1 Definitions and examples

Let Σ be an oriented compact surface, possibly with boundary, and let \mathcal{P} be a finite collection of punctures in the interior of Σ . Let $\mathcal{M}(\Sigma, \mathcal{P})$ denote the mapping class group of the pair (Σ, \mathcal{P}) , as defined in Subsection 2.3. Let Γ be a Coxeter graph, and let G_Γ be the Artin group of type Γ . Define a *geometric representation* of G_Γ in $\mathcal{M}(\Sigma, \mathcal{P})$ to be a homomorphism from G_Γ to $\mathcal{M}(\Sigma, \mathcal{P})$.

The main tools for constructing geometric representations of Artin groups are the Dehn twists and the braid twists. The braid twists are defined in Subsection 2.3, and the Dehn twists are defined as follows.

An *essential circle* is an embedding $a: \mathbb{S}^1 \hookrightarrow \Sigma \setminus \mathcal{P}$ of the circle whose image is contained in the interior of Σ and does not bound any disk in Σ containing 0 or 1 puncture. Two essential circles a, a' are *isotopic* if there exists a continuous family $\{a_t\}_{t \in [0,1]}$ of essential circles such that $a = a_0$ and $a' = a_1$. Isotopy of essential circles is an equivalence relation that we denote by $a \sim a'$.

Let $a: \mathbb{S}^1 \rightarrow \Sigma \setminus \mathcal{P}$ be an essential circle. Take an embedding $A: [0, 1] \times \mathbb{S}^1 \rightarrow \Sigma \setminus \mathcal{P}$ of the annulus such that $A(\frac{1}{2}, z) = a(z)$ for all $z \in \mathbb{S}^1$, and define $T \in \text{Homeo}^+(\Sigma, \mathcal{P})$ by

$$(T \circ A)(t, z) = A(t, e^{2i\pi t} z),$$

and T is the identity outside the image of A (see Figure 16). The *Dehn twist* along a , denoted by σ_a , is defined to be the element of $\mathcal{M}(\Sigma, \mathcal{P})$ represented by T . Note that

- the definition of σ_a does not depend on the choice of the map A ;
- if a is isotopic to a' , then $\sigma_a = \sigma_{a'}$.

Recall that, for an essential arc a of (Σ, \mathcal{P}) , τ_a denotes the braid twist along a . The Dehn twists and the braid twists satisfy the following relations (see [20], [110]).

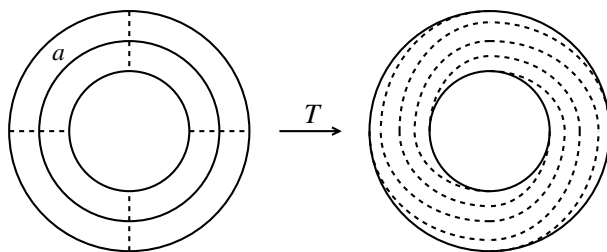


Figure 16. Dehn twist.

Proposition 7.1. (1) Let a, b be two essential circles that intersect transversely. Then

$$\begin{aligned}\sigma_a \sigma_b &= \sigma_b \sigma_a && \text{if } a \cap b = \emptyset, \\ \sigma_a \sigma_b \sigma_a &= \sigma_b \sigma_a \sigma_b && \text{if } |a \cap b| = 1.\end{aligned}$$

(2) Let a, b be two essential arcs of (Σ, \mathcal{P}) . Then

$$\begin{aligned}\tau_a \tau_b &= \tau_b \tau_a && \text{if } a \cap b = \emptyset, \\ \tau_a \tau_b \tau_a &= \tau_b \tau_a \tau_b && \text{if } a(0) = b(1) \text{ and } a \cap b = \{a(0)\}.\end{aligned}$$

(3) Let a be an essential arc, and let b be an essential circle which intersects a transversely. Then

$$\begin{aligned}\tau_a \sigma_b &= \sigma_b \tau_a && \text{if } a \cap b = \emptyset, \\ \tau_a \sigma_b \tau_a \sigma_b &= \sigma_b \tau_a \sigma_b \tau_a && \text{if } |a \cap b| = 1.\end{aligned}$$

Example 1. Suppose $\Sigma = \mathbb{D}$ is a disk, and $\mathcal{P}_n = \{P_1, \dots, P_n\}$ is a collection of n punctures in the interior of Σ . Then the Artin isomorphism $\Phi: \mathcal{B}_n \rightarrow \mathcal{M}(\mathbb{D}, \mathcal{P}_n)$ of Theorem 2.16 is a geometric representation of $G_{A_{n-1}} = \mathcal{B}_n$.

Example 2. Let $n \geq 3$. Suppose that, if n is odd, then Σ is a surface of genus $\frac{n-1}{2}$ with one boundary component, and if n is even, then Σ is a surface of genus $\frac{n-2}{2}$ with two boundary components. Let a_1, \dots, a_{n-1} be the essential circles of Σ pictured in Figure 17. By Proposition 7.1, the mapping $\sigma_i \mapsto \sigma_{a_i}$, $1 \leq i \leq n-1$, induces a representation $\rho_M: \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma)$ called the *monodromy representation* of \mathcal{B}_n . This geometric representation was introduced by Birman and Hilden in [25], where it is proved that ρ_M is faithful and its image consists of mapping classes arising from homeomorphisms symmetric with respect to a hyperelliptic involution (see also [26], [154], and [117]). It is also the geometric monodromy of the simple singularity of type A_{n-1} (see [135]). Let $P_0 \in \partial\Sigma$ be a base-point. Then ρ_M induces a homomorphism $\rho_{M*}: \mathcal{B}_n \rightarrow \text{Aut}(\pi_1(\Sigma, P_0))$ which turns out to coincide with the homomorphism $\rho_D: \mathcal{B}_n \rightarrow \text{Aut}(F_{n-1})$ defined in Subsection 3.1 (see [60]).

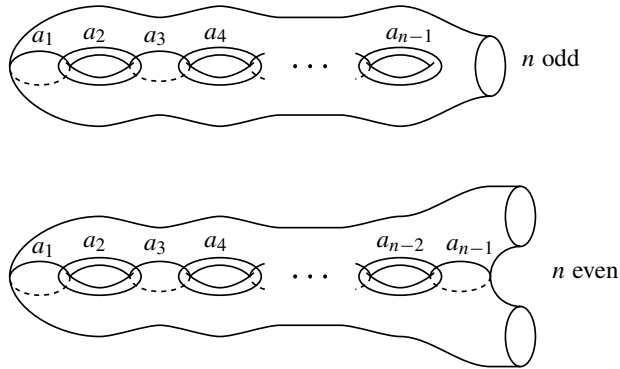


Figure 17. Monodromy representation of \mathcal{B}_n .

Example 3. Let $\mathbb{D}^2 = \{z \in \mathbb{C}; |z| \leq 1\}$ be the standard disk. A *chord diagram* in \mathbb{D}^2 is defined to be a collection $\{S_1, \dots, S_n\}$ of segments in \mathbb{D}^2 such that

- the endpoints of S_i belong to $\partial\mathbb{D}^2$ and its interior is contained in the interior of \mathbb{D}^2 , for all $1 \leq i \leq n$;
- either S_i and S_j are disjoint, or they intersect transversely in a unique point in the interior of \mathbb{D}^2 , for all $1 \leq i \neq j \leq n$.

From this data one can define a Coxeter matrix $M = (m_{ij})_{1 \leq i, j \leq n}$ setting $m_{ij} = 2$ if S_i and S_j are disjoint, and $m_{ij} = 3$ if they intersect. The Coxeter graph Γ of M is called the *intersection diagram* of the chord diagram.

From this data one can also define a surface Σ by attaching to \mathbb{D}^2 a handle H_i which joins the extremities of S_i , for all $1 \leq i \leq n$ (see Figure 18). Let a_i be the essential circle of Σ made with S_i and the central arc of H_i . Then, by Proposition 7.1,

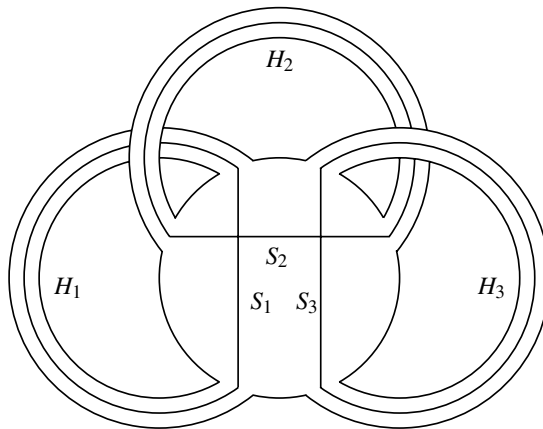


Figure 18. Chord diagram and associated surface.

the mapping $\sigma_i \mapsto \sigma_{a_i}$, $1 \leq i \leq n$, induces a geometric representation $\rho_{PV}: G_\Gamma \rightarrow \mathcal{M}(\Sigma)$, called *Perron–Vannier representation*.

The Perron–Vannier representations were introduced in [135]. If $\Gamma = A_{n-1}$, then ρ_{PV} is equal to the monodromy representation ρ_M defined in Example 2. More generally, if Γ is A_n ($n \geq 1$), D_n ($n \geq 4$), or E_k ($k = 6, 7, 8$), then ρ_{PV} is the geometric monodromy of the simple singularity of type Γ (see [135]). For a connected graph Γ , the representation ρ_{PV} is faithful if and only if either $\Gamma = A_n$ for some $n \geq 1$, or $\Gamma = D_n$ for some $n \geq 4$ (see [135], [109], [152]).

Example 4. This example comes from [58]. Recall that a Coxeter graph Γ is of *small type* if $m_{st} \leq 3$ for all $s, t \in S$, where $M = (m_{st})_{s,t \in S}$ is the Coxeter matrix of Γ . Let Γ be a small type Coxeter graph. We choose (arbitrarily) a total order $<$ on S . For $s \in S$, we set $St_s = \{t \in S; m_{st} = 3\} \cup \{s\}$. Write $St_s = \{t_1, t_2, \dots, t_k\}$ such that $t_1 < t_2 < \dots < t_k$, and suppose that $s = t_j$. For $1 \leq i \leq k$, the difference $i - j$ is called the *relative position* of t_i with respect to s and is denoted by $\text{pos}(t_i : s)$. In particular, $\text{pos}(s : s) = 0$.

Let $s \in S$ and let $k = |St_s|$. Let An_s denote the annulus $An_s = (\mathbb{R}/2k\mathbb{Z}) \times [0, 1]$. We define the surface $\Sigma = \Sigma_\Gamma$ by

$$\Sigma = \left(\bigsqcup_{s \in S} An_s \right) / \sim,$$

where \sim is the equivalence relation defined as follows. Let $s, t \in S$ such that $s < t$ and $m_{st} = 3$. Set $p = \text{pos}(t : s) > 0$ and $q = \text{pos}(s : t) < 0$. For all $(x, y) \in [0, 1] \times [0, 1]$ the relation \sim identifies the point $(2p + x, y)$ of An_s with the point $(2q + 1 - y, x)$ of An_t (see Figure 19).

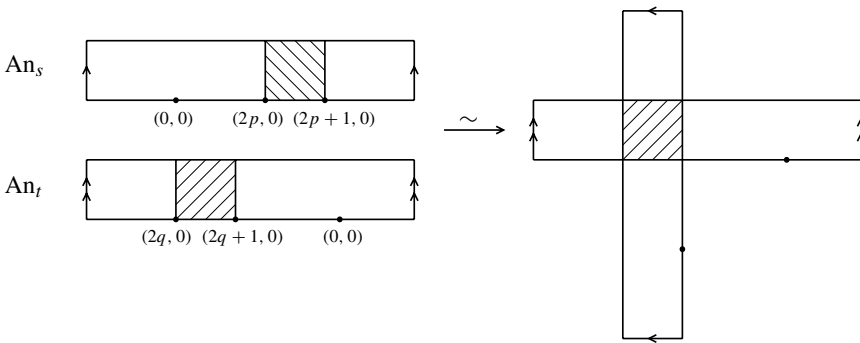


Figure 19. Identification of annuli.

We identify each annulus An_s with its image in Σ , and we denote by a_s its central curve. Note that a_s is an essential circle, $a_s \cap a_t = \emptyset$ if $m_{st} = 2$, and $|a_s \cap a_t| = 1$ if $m_{st} = 3$. So, by Proposition 7.1, the mapping $\sigma_s \mapsto \sigma_{a_s}$, $s \in S$, induces a geometric representation $\rho_{CP}: G_\Gamma \rightarrow \mathcal{M}(\Sigma)$.

We have $\rho_{CP} = \rho_{PV}$ if Γ is a tree. (Note that it may happen that ρ_{PV} is not defined if Γ is not a tree.) If $\Gamma = \tilde{A}_n$, then ρ_{CP} is faithful (while, by [109], ρ_{PV} is not faithful in this case).

7.2 Presentations

Let $\Sigma_{g,r}$ be a surface of genus $g \geq 1$ with $r \geq 0$ boundary components, and let \mathcal{P}_n be a collection of n punctures in the interior of $\Sigma_{g,r}$, where $n \geq 0$.

First assume that $r \geq 1$. Consider the essential circles $a_0, a_1, \dots, a_r, b_1, b_2, \dots, b_{2g-1}, c, d_1, \dots, d_{r-1}$, and the essential arcs e_1, e_2, \dots, e_{n-1} drawn in Figure 20. Note that there is no c if $g = 1$, there is no d_i if $r = 1$, there is no a_r if $n = 0$, and

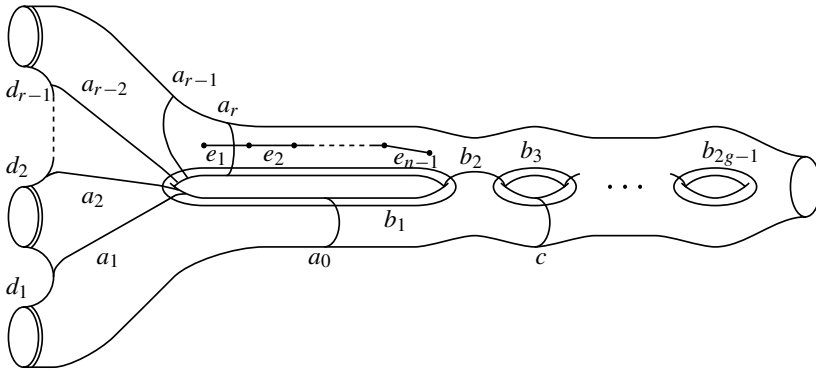


Figure 20. Generators of $\mathcal{M}(\Sigma_{g,r}, \mathcal{P}_n)$.

there is no e_i if $n = 0$ or 1 . Let $\Gamma(g, r, n)$ be the Coxeter graph drawn in Figure 21. One can show that the set

$$\{\sigma_{a_0}, \sigma_{a_1}, \dots, \sigma_{a_r}, \sigma_{b_1}, \sigma_{b_2}, \dots, \sigma_{b_{2g-1}}, \sigma_c, \sigma_{d_1}, \sigma_{d_2}, \dots, \sigma_{d_{r-1}}, \tau_{e_1}, \dots, \tau_{e_{n-1}}\}$$

generates $\mathcal{M}(\Sigma_{g,r}, \mathcal{P}_n)$. On the other hand, by Proposition 7.1, the mapping

$$\begin{aligned} x_i &\mapsto \sigma_{a_i} \quad (0 \leq i \leq r), & y_i &\mapsto \sigma_{b_i} \quad (1 \leq i \leq 2g - 1), & z &\mapsto \sigma_c \\ u_i &\mapsto \sigma_{d_i} \quad (1 \leq i \leq r - 1), & v_j &\mapsto \tau_{e_j} \quad (1 \leq j \leq n - 1), \end{aligned}$$

induces a homomorphism $\rho: G_{\Gamma(g,r,n)} \rightarrow \mathcal{M}(\Sigma_{g,r}, \mathcal{P}_n)$. So, in order to obtain a presentation for $\mathcal{M}(\Sigma_{g,r}, \mathcal{P}_n)$, it suffices to find normal generators for $\text{Ker } \rho$. This was done in [124] for $r = 1$ and $n = 0$, and in [110] for the other cases.

One can use the same kind of arguments for the case $r = 0$. Consider the essential circles $a_0, a_1, b_1, b_2, \dots, b_{2g-1}, c$, and the essential arcs e_1, e_2, \dots, e_{n-1} drawn in Figure 22. Then the set

$$\{\sigma_{a_0}, \sigma_{a_1}, \sigma_{b_1}, \sigma_{b_2}, \dots, \sigma_{b_{2g-1}}, \sigma_c, \tau_{e_1}, \tau_{e_2}, \dots, \tau_{e_{n-1}}\}$$

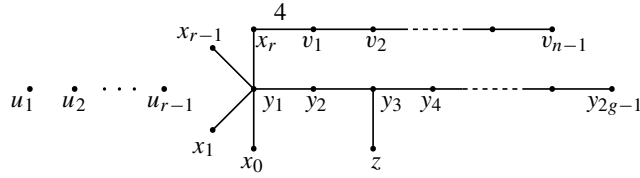


Figure 21. The Coxeter graph $\Gamma(g, r, n)$.

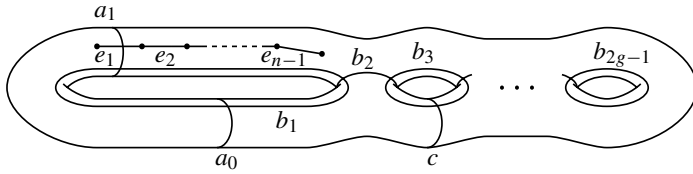


Figure 22. Generators of $\mathcal{M}(\Sigma_{g,0}, \mathcal{P}_n)$.

generates $\mathcal{M}(\Sigma_{g,0}, \mathcal{P}_n)$, and the mapping

$$\begin{aligned} x_i &\mapsto \sigma_{a_i} \quad (i = 0, 1), & y_i &\mapsto \sigma_{b_i} \quad (1 \leq i \leq 2g - 1), \\ z &\mapsto \sigma_c, & v_j &\mapsto \tau_{e_j} \quad (1 \leq j \leq n - 1), \end{aligned}$$

induces a homomorphism $\rho: G_{\Gamma(g,1,n)} \rightarrow \mathcal{M}(\Sigma_{g,0}, \mathcal{P}_n)$. Here again, the kernel of ρ was calculated in [124] for $n = 0$, and in [110] for $n \geq 1$.

In order to state the results of [124] and [110], we need the following notations. Let Γ be a Coxeter graph, let $M = (m_{st})_{s,t \in S}$ be the Coxeter matrix of Γ , and let (G, Σ) be the Artin system of type Γ . For $X \subset S$ (where S is the set of vertices of Γ), we denote by Γ_X the full subgraph of Γ generated by X , we set $\Sigma_X = \{\sigma_s; s \in X\}$, and we denote by G_X the subgroup of G generated by Σ_X . By [116], (G_X, Σ_X) is the Artin system of type Γ_X (see also [132]). If Γ_X is of spherical type, then we denote by $\Delta(X)$ the Garside element of (G_X, Σ_X) , viewed as an element of G .

Theorem 7.2 (Matsumoto [124]). (1) $\mathcal{M}(\Sigma_{g,1})$ is isomorphic to the quotient of $G_{\Gamma(g,1,0)}$ by the following relations

$$(R1) \quad \Delta(y_1, y_2, y_3, z)^4 = \Delta(x_0, y_1, y_2, y_3, z)^2 \quad \text{if } g \geq 2,$$

$$(R2) \quad \Delta(y_1, y_2, y_3, y_4, y_5, z)^2 = \Delta(x_0, y_1, y_2, y_3, y_4, y_5, z) \quad \text{if } g \geq 3.$$

(2) $\mathcal{M}(\Sigma_{g,0})$ is isomorphic to the quotient of $G_{\Gamma(g,1,0)}$ by the above relations (R1) and (R2) together with

$$(R3) \quad (x_0 y_1)^6 = 1 \quad \text{if } g = 1,$$

$$x_0^{2g-2} = \Delta(y_2, y_3, z, y_4, \dots, y_{2g-1}) \quad \text{if } g \geq 2.$$

Theorem 7.3 (Labruère, Paris [110]). *Let $g \geq 1$, $r \geq 1$, and $n \geq 0$. Then $\mathcal{M}(\Sigma_{g,r}, \mathcal{P}_n)$ is isomorphic to the quotient of $G_{\Gamma(g,r,n)}$ by the following relations.*

- *Relations from $\mathcal{M}(\Sigma_{g,1})$.*

$$(R1) \quad \Delta(y_1, y_2, y_3, z)^4 = \Delta(x_0, y_1, y_2, y_3, z)^2 \quad \text{if } g \geq 2,$$

$$(R2) \quad \Delta(y_1, y_2, y_3, y_4, y_5, z)^2 = \Delta(x_0, y_1, y_2, y_3, y_4, y_5, z) \quad \text{if } g \geq 3.$$

- *Relations of commutation.*

$$(R3) \quad \begin{aligned} & x_k \cdot \Delta(x_{i+1}, x_j, y_1)^{-1} x_i \Delta(x_{i+1}, x_j, y_1) \\ &= \Delta(x_{i+1}, x_j, y_1)^{-1} x_i \Delta(x_{i+1}, x_j, y_1) \cdot x_k \\ & \quad \text{if } 0 \leq k < j < i \leq r-1, \end{aligned}$$

$$(R4) \quad \begin{aligned} & y_2 \cdot \Delta(x_{i+1}, x_j, y_1)^{-1} x_i \Delta(x_{i+1}, x_j, y_1) \\ &= \Delta(x_{i+1}, x_j, y_1)^{-1} x_i \Delta(x_{i+1}, x_j, y_1) \cdot y_2 \\ & \quad \text{if } 0 \leq j < i \leq r-1 \text{ and } g \geq 2 \end{aligned}$$

- *Expressions of the u_i 's.*

$$(R5) \quad \begin{aligned} u_1 &= \Delta(x_0, x_1, y_1, y_2, y_3, z) \cdot \Delta(x_1, y_1, y_2, y_3, z)^{-2} \\ & \quad \text{if } g \geq 2, \end{aligned}$$

$$(R6) \quad \begin{aligned} u_{i+1} &= \Delta(x_i, x_{i+1}, y_1, y_2, y_3, z) \cdot \Delta(x_{i+1}, y_1, y_2, y_3, z)^{-2} \\ & \cdot \Delta(x_0, x_{i+1}, y_1)^2 \cdot \Delta(x_0, x_i, x_{i+1}, y_1)^{-1} \\ & \quad \text{if } 1 \leq i \leq r-2 \text{ and } g \geq 2. \end{aligned}$$

- *Other relations.*

$$(R7) \quad \Delta(x_{r-1}, x_r, y_1, v_1) = \Delta(x_r, y_1, v_1)^2 \quad \text{if } n \geq 2,$$

$$(R8a) \quad \begin{aligned} \Delta(x_0, x_1, y_1, y_2, y_3, z) &= \Delta(x_1, y_1, y_2, y_3, z)^2 \\ & \quad \text{if } n \geq 1, g \geq 2, \text{ and } r = 1, \end{aligned}$$

$$(R8b) \quad \begin{aligned} \Delta(x_{r-1}, x_r, y_1, y_2, y_3, z) \cdot \Delta(x_r, y_1, y_2, y_3, z)^{-2} \\ &= \Delta(x_0, x_{r-1}, x_r, y_1) \cdot \Delta(x_0, x_r, y_1)^{-2} \\ & \quad \text{if } n \geq 1, g \geq 2, \text{ and } r \geq 2. \end{aligned}$$

Note that only the relations (R1), (R2), (R7), and (R8a) remain in the presentation if $r = 1$, and (R8a) must be replaced by (R8b) if $r \geq 2$. Note also that, if $g \geq 2$, then u_1, \dots, u_{r-1} can be removed from the generating set. However, to do so, one must add new long relations.

Theorem 7.4 (Labruère, Paris [110]). *Let $g \geq 1$ and $n \geq 1$. Then $\mathcal{M}(\Sigma_{g,0}, \mathcal{P}_n)$ is isomorphic to the quotient of $G_{\Gamma(g,1,n)}$ by the following relations.*

- Relations from $\mathcal{M}(\Sigma_{g,1}, \mathcal{P}_n)$.

$$(R1) \quad \Delta(y_1, y_2, y_3, z)^4 = \Delta(x_0, y_1, y_2, y_3, z)^2 \quad \text{if } g \geq 2,$$

$$(R2) \quad \Delta(y_1, y_2, y_3, y_4, y_5, z)^2 = \Delta(x_0, y_1, y_2, y_3, y_4, y_5, z) \quad \text{if } g \geq 3,$$

$$(R7) \quad \Delta(x_0, x_1, y_1, v_1) = \Delta(x_1, y_1, v_1)^2 \quad \text{if } n \geq 2,$$

$$(R8a) \quad \Delta(x_0, x_1, y_1, y_2, y_3, z) = \Delta(x_1, y_1, y_2, y_3, z)^2 \\ \text{if } n \geq 1 \text{ and } g \geq 2.$$

- Other relations.

$$(R9a) \quad x_0^{2g-n-2} \cdot \Delta(x_1, v_1, \dots, v_{n-1}) = \Delta(z, y_2, \dots, y_{2g-1})^2 \quad \text{if } g \geq 2,$$

$$(R9b) \quad x_0^n = \Delta(x_1, v_1, \dots, v_{n-1}) \quad \text{if } g = 1,$$

$$(R9c) \quad \Delta(x_0, y_1)^4 = \Delta(v_1, \dots, v_{n-1})^2 \quad \text{if } g = 1.$$

Note. Presentations of $\mathcal{M}(\Sigma_{g,r})$, also in terms of Artin groups, with more generators but simpler relations, were obtained by Gervais in [88]. On the other hand, a unified proof of all these presentations can be found in [11].

7.3 Classification

This subsection is an account of Castel's results [40] on the geometric representations of the braid group \mathcal{B}_n in mapping class groups of surfaces of genus $g \leq \frac{n-1}{2}$.

First suppose that n is odd, $n \geq 5$. Write $n = 2k + 1$, where $k \geq 2$. Let $r \geq 0$. We present the surface $\Sigma_{k,r}$ as the union of three subsurfaces, Ω_0 , \mathbb{A} , and Ω_1 , where Ω_0 is a surface of genus k with one boundary component, c , Ω_1 is a surface of genus 0 with $r + 1$ boundary components, c', d_1, \dots, d_r , and \mathbb{A} is an annulus bounded by c and c' (see Figure 23). Consider the essential circles a_1, a_2, \dots, a_{2k} drawn in Figure 23. Then, by Proposition 7.1, there exists a homomorphism $\rho_M: \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{k,r})$ which sends σ_i to σ_{a_i} for all $1 \leq i \leq n - 1 = 2k$.

The statement of Castel's classification of the geometric representations of \mathcal{B}_n in $\mathcal{M}(\Sigma_{k,r})$ involves the centralizer of $\text{Im } \rho_M$ in $\mathcal{M}(\Sigma_{k,r})$. That is why we start with a description of the latter.

The inclusion of Ω_1 in $\Sigma_{k,r}$ induces a homomorphism $\mathcal{M}(\Omega_1) \rightarrow \mathcal{M}(\Sigma_{k,r})$ which is injective (see [134]). It is easily checked that the image of this homomorphism is contained in the centralizer of $\text{Im } \rho_M$. Another element of the centralizer is the element $u \in \mathcal{M}(\Sigma_{k,r})$ represented by the homeomorphism $U: \Sigma_{k,r} \rightarrow \Sigma_{k,r}$ which is the axial symmetry relative to the axis D on Ω_0 , a half-twist which pointwise fixes c' on the annulus \mathbb{A} , and the identity on Ω_1 .

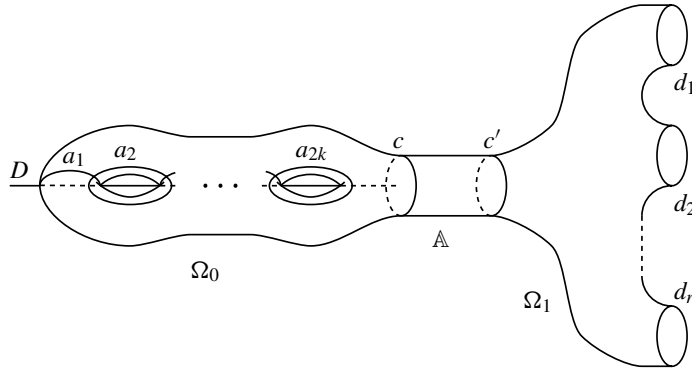


Figure 23. Decomposition of $\Sigma_{k,r}$ (n odd).

Proposition 7.5 (Castel [40]). *The centralizer of $\text{Im } \rho_M$ in $\mathcal{M}(\Sigma_{k,r})$ is generated by $\mathcal{M}(\Omega_1) \cup \{u\}$.*

If $r = 0$, then $\mathcal{M}(\Omega_1) = \{1\}$, u is of order 2, and $Z_{\mathcal{M}(\Sigma_{k,r})}(\text{Im } \rho_M) = \langle u \rangle$ is cyclic of order 2. If $r = 1$, then $\mathcal{M}(\Omega_1) = \langle \tau_c \rangle$, $u^2 = \tau_c$, and $Z_{\mathcal{M}(\Sigma_{k,r})}(\text{Im } \rho_M) = \langle u \rangle$ is an infinite cyclic group. If $r = 2$, then $Z_{\mathcal{M}(\Sigma_{k,r})}(\text{Im } \rho_M)$ is a free abelian group of rank 3 freely generated by $\{u, \sigma_{d_1}, \sigma_{d_2}\}$. If $r \geq 3$, then $Z_{\mathcal{M}(\Sigma_{k,r})}(\text{Im } \rho_M)$ is more complicated.

For $\varepsilon \in \{\pm 1\}$ and $z \in Z_{\mathcal{M}(\Sigma_{k,r})}(\text{Im } \rho_M)$, the mapping $\sigma_i \mapsto \sigma_{a_i}^\varepsilon z$, $1 \leq i \leq n - 1$, induces a homomorphism $\rho_M(\varepsilon, z) : \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{k,r})$ called the *transvection of ρ_M by (ε, z)* . On the other hand, a homomorphism $\varphi : \mathcal{B}_n \rightarrow G$, where G is a group, is called *cyclic* if there exists $\alpha \in G$ such that $\varphi(\sigma_i) = \alpha$ for all $1 \leq i \leq n - 1$.

Theorem 7.6 (Castel [40]). *Suppose n odd, $n \geq 5$, and set $n = 2k + 1$. Let $g \geq 0$ and $r \geq 0$.*

- (1) *If $g < k$, then all the homomorphisms $\varphi : \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{g,r})$ are cyclic.*
- (2) *All the non-cyclic homomorphisms $\varphi : \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{k,r})$ are conjugate to transvections of ρ_M .*
- (3) *The homomorphism $\rho_M : \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{k,r})$ is injective if and only if $r \geq 1$.*

Now, we suppose that n is even, $n \geq 6$, and we set $n = 2k + 2$. We choose $r_1, r_2 \geq 0$ such that $r_1 + r_2 = r$ and we represent the surface $\Sigma_{k,r}$ as the union of three subsurfaces, a surface Ω_0 of genus k with two boundary components, c_1 and c_2 , a surface Ω_1 of genus 0 with $r_1 + 1$ boundary components c_1, d_1, \dots, d_{r_1} , and a surface Ω_2 of genus 0 with $r_2 + 1$ boundary components $c_2, d_{r_1+1}, \dots, d_{r_1+r_2}$ (see Figure 24). Consider the essential circles a_1, \dots, a_{n-1} drawn in Figure 24. Then, by Proposition 7.1, there exists a homomorphism $\rho_M(r_1, r_2) : \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{k,r})$ which sends σ_i to σ_{a_i} for all $1 \leq i \leq n - 1$.

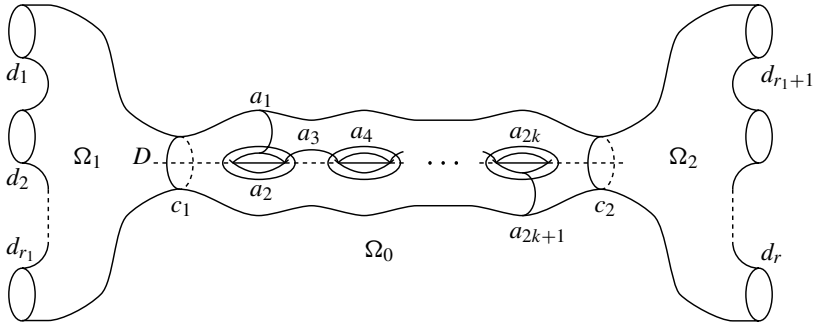


Figure 24. Decomposition of $\Sigma_{k,r}$ (n even).

The inclusions $\Omega_1, \Omega_2 \subset \Sigma_{k,r}$ induce a homomorphism $\mathcal{M}(\Omega_1) \times \mathcal{M}(\Omega_2) \rightarrow \mathcal{M}(\Sigma_{k,r})$ which is injective (see [134]), and we have:

Proposition 7.7 (Castel [40]). (1) If $r > 0$, then the centralizer of $\text{Im } \rho_M(r_1, r_2)$ in $\mathcal{M}(\Sigma_{k,r})$ is $\mathcal{M}(\Omega_1) \times \mathcal{M}(\Omega_2)$.

(2) If $r = 0$, then the centralizer of $\text{Im } \rho_M(r_1, r_2)$ in $\mathcal{M}(\Sigma_{k,r})$ is a cyclic group of order 2 generated by an element represented by the axial symmetry relative to the axis D of Figure 24.

For $\varepsilon \in \{\pm 1\}$ and $z \in Z_{\mathcal{M}(\Sigma_{k,r})}(\text{Im } \rho_M(r_1, r_2))$, the mapping $\sigma_i \mapsto \sigma_{a_i}^\varepsilon z$, $1 \leq i \leq n - 1$, induces a homomorphism $\rho_M(r_1, r_2, \varepsilon, z): \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{k,r})$ called the transvection of $\rho_M(r_1, r_2)$ by (ε, z) .

Theorem 7.8 (Castel [40]). Suppose n even, $n \geq 6$, and set $n = 2k + 2$. Let $g \geq 0$ and $r \geq 0$.

- (1) If $g < k$, then all the homomorphisms $\varphi: \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{g,r})$ are cyclic.
- (2) If $\varphi: \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{k,r})$ is a non-cyclic homomorphism, then there exist $r_1, r_2 \geq 0$ such that $r_1 + r_2 = r$ and φ is conjugate to a transvection of $\rho_M(r_1, r_2)$.
- (3) Let $r_1, r_2 \geq 0$ such that $r_1 + r_2 = r$. The homomorphism $\rho_M(r_1, r_2): \mathcal{B}_n \rightarrow \mathcal{M}(\Sigma_{k,r})$ is injective if and only if $r_1 \geq 1$ and $r_2 \geq 1$.

Recall that, for a group G , $\text{Out}(G)$ denotes the group of outer automorphisms of G . Now, Theorems 7.6 and 7.8 can be used for new proofs of the following two theorems.

Theorem 7.9 (Dyer, Grossman [77]). We have $\text{Out}(\mathcal{B}_n) = \mathbb{Z}/2\mathbb{Z}$ if $n \geq 5$.

Theorem 7.10 (Ivanov [98], McCarthy [125]). Let $g \geq 2$ and $r \geq 0$. Then

$$\text{Out}(\mathcal{M}(\Sigma_{g,r})) = \begin{cases} \{1\} & \text{if } r \geq 1, \\ \mathbb{Z}/2\mathbb{Z} & \text{if } r = 0 \text{ and } g \geq 3, \\ \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} & \text{if } r = 0 \text{ and } g = 2. \end{cases}$$

References

- [1] J. W. Alexander, Deformations of an n -cell. *Proc. Nat. Acad. Sci. USA* 9 (1923), 406–407. [400](#)
- [2] D. Allcock, Braid pictures for Artin groups. *Trans. Amer. Math. Soc.* 354 (9) (2002), 3455–3474. [406](#)
- [3] V. I. Arnol'd, The cohomology classes of algebraic functions that are preserved under Tschirnhausen transformations, *Funkcional. Anal. i Priložen* 4 (1) (1970), 84–85; English transl. *Functional Anal. Appl.* 4 (1970), 74–75. [390](#)
- [4] V. I. Arnol'd, Certain topological invariants of algebraic functions. *Trudy Moskov. Mat. Obšč.* 21 (1970), 27–46. [421](#), [422](#)
- [5] V. I. Arnol'd, Topological invariants of algebraic functions. II. *Funkcional. Anal. i Priložen* 4 (2) (1970), 1–9; English transl. *Functional Anal. Appl.* 4 (1970), 91–98. [390](#)
- [6] V. I. Arnol'd, Matrices depending on parameters, *Uspehi Mat. Nauk* 26 (2) (1971), 101–114; English transl. *Russian Math. Surveys* 26 (2) (1971), 29–43. [390](#)
- [7] E. Artin, Theorie der Zöpfe. *Abhandlungen Hamburg* 4 (1925), 47–72. [390](#), [394](#), [400](#), [402](#)
- [8] E. Artin, Theory of braids. *Ann. of Math. (2)* 48 (1947), 101–126. [394](#), [400](#), [402](#)
- [9] R. Baer, Isotopie von Kurven auf orientierbaren, geschlossenen Flächen und ihr Zusammenhang mit der topologischen Deformation der Flächen. *J. f. Math.* 159 (1928), 101–116. [402](#)
- [10] R. W. Bell, D. Margalit, Injections of Artin groups. *Comment. Math. Helv.* 82 (2007), 725–751. [406](#)
- [11] S. Benvenuti, Finite presentations for the mapping class group via the ordered complex of curves. *Adv. Geom.* 1 (3) (2001), 291–321. [441](#)
- [12] D. Bessis, The dual braid monoid. *Ann. Sci. École Norm. Sup* (4) 36 (5) (2003), 647–683. [414](#)
- [13] D. Bessis, A dual braid monoid for the free group. *J. Algebra* 302 (1) (2006), 55–69. [413](#)
- [14] D. Bessis, Garside categories, periodic loops and cyclic sets. Preprint, 2006. [413](#)
- [15] D. Bessis, Finite complex reflection arrangements are $K(\pi, 1)$. Preprint, 2006. [412](#), [413](#), [414](#)
- [16] D. Bessis, J. Michel, Explicit presentations for exceptional braid groups. *Experiment. Math.* 13 (3) (2004), 257–266. [412](#)
- [17] S. J. Bigelow, Braid groups are linear. *J. Amer. Math. Soc.* 14 (2) (2001), 471–486. [391](#), [430](#), [431](#), [433](#), [434](#)
- [18] J. S. Birman, Mapping class groups and their relationship to braid groups. *Comm. Pure Appl. Math.* 22 (1969), 213–238. [399](#)
- [19] J. S. Birman, *Braids, links, and mapping class groups*. Ann. of Math. Stud. 82, Princeton University Press, Princeton, N.J., 1974. [390](#), [391](#)
- [20] J. S. Birman. Mapping class groups of surfaces. In *Braids* (Santa Cruz, CA, 1986), Contemp. Math. 78, Amer. Math. Soc., Providence, RI, 1988, 13–43. [434](#)

- [21] J. S. Birman, T. E. Brendle, Braids: a survey. In *Handbook of knot theory*, Elsevier B.V., Amsterdam 2005, 19–103. [391](#)
- [22] J. S. Birman, V. Gebhardt, J. González-Meneses, Conjugacy in Garside groups I: cyclings, powers, and rigidity. *Groups Geom. Dyn.* 1 (2007), 221–279. [419](#)
- [23] J. S. Birman, V. Gebhardt, J. González-Meneses, Conjugacy in Garside groups II: Structure of the ultra summit set. *Groups Geom. Dyn.* 2 (2008), 13–61. [419](#)
- [24] J. S. Birman, V. Gebhardt, J. González-Meneses, Conjugacy in Garside groups III: Periodic braids. *J. Algebra* 316 (2007), 746–776. [419](#)
- [25] J. S. Birman, H. M. Hilden, On the mapping class groups of closed surfaces as covering spaces. In *Advances in the theory of Riemann surfaces* (Proc. Conf., Stony Brook, N.Y., 1969), Ann. of Math. Stud. 66, Princeton University Press, Princeton, N.J., 1971, 81–115. [435](#)
- [26] J. S. Birman, H. M. Hilden, On isotopies of homeomorphisms of Riemann surfaces. *Ann. of Math.* (2) 97 (1973), 424–439. [435](#)
- [27] J. S. Birman, K. H. Ko, S. J. Lee, A new approach to the word and conjugacy problems in the braid groups. *Adv. Math.* 139 (2) (1998), 322–353. [414](#)
- [28] N. Bourbaki, *Groupes et algèbres de Lie. Chapitres IV, V et VI*. Hermann, Paris 1968. [406](#), [407](#), [408](#), [411](#), [412](#), [427](#), [428](#)
- [29] E. Brieskorn, Singular elements of semi-simple algebraic groups. In *Actes du Congrès International des Mathématiciens* (Nice, 1970), Tome 2, Gauthier-Villars, Paris 1971, 279–284. [390](#)
- [30] E. Brieskorn, Die Fundamentalgruppe des Raumes der regulären Orbits einer endlichen komplexen Spiegelungsgruppe. *Invent. Math.* 12 (1971), 57–61. [390](#), [412](#)
- [31] E. Brieskorn, Sur les groupes de tresses (d’après V.I. Arnol’d). In *Séminaire Bourbaki, 24ème année (1971/1972)*, Exp. No. 401, Lecture Notes in Math. 317, Springer-Verlag, Berlin 1973, 21–44. [390](#), [424](#)
- [32] E. Brieskorn, K. Saito, Artin-Gruppen und Coxeter-Gruppen. *Invent. Math.* 17 (1972), 245–271. [390](#), [406](#), [410](#)
- [33] M. Broué, G. Malle, R. Rouquier, Complex reflection groups, braid groups, Hecke algebras. *J. Reine Angew. Math.* 500 (1998), 127–190. [412](#)
- [34] K. S. Brown, *Cohomology of groups*. Grad. Texts in Math. 87, Springer-Verlag, New York 1982. [398](#)
- [35] W. Burau, Über Zopf invarianten. *Abh. Math. Semin. Hamb. Univ.* 9 (1932), 117–124. [394](#)
- [36] F. Callegaro, On the cohomology of Artin groups in local systems and the associated Milnor fiber. *J. Pure Appl. Algebra* 197 (1–3) (2005), 323–332. [424](#)
- [37] F. Callegaro, The homology of the Milnor fiber for classical braid groups. *Algebr. Geom. Topol.* 6 (2006), 1903–1923. [424](#)
- [38] F. Callegaro, D. Moroni, M. Salvetti, The $K(\pi, 1)$ problem for the affine Artin group of type \tilde{B}_n and its cohomology. Preprint, 2007. [412](#), [424](#)
- [39] F. Callegaro, M. Salvetti, Integral cohomology of the Milnor fibre of the discriminant bundle associated with a finite Coxeter group. *C. R. Math. Acad. Sci. Paris* 339 (8) (2004), 573–578. [424](#)

- [40] F. Castel, Ph. D. Thesis, Université de Bourgogne, in preparation. [391](#), [441](#), [442](#), [443](#)
- [41] A. Castella, Automorphismes et admissibilité dans les groupes de Coxeter et les monoïdes d'Artin-Tits. Ph. D. Thesis, Université de Paris-Sud, 2006. [432](#)
- [42] R. Charney, Artin groups of finite type are biautomatic. *Math. Ann.* 292 (4) (1992), 671–683. [419](#)
- [43] R. Charney, Geodesic automation and growth functions for Artin groups of finite type. *Math. Ann.* 301 (2) (1995), 307–324. [419](#)
- [44] R. Charney, J. Crisp, Automorphism groups of some affine and finite type Artin groups. *Math. Res. Lett.* 12 (2–3) (2005), 321–333. [406](#)
- [45] R. Charney, M. W. Davis, The $K(\pi, 1)$ -problem for hyperplane complements associated to infinite reflection groups. *J. Amer. Math. Soc.* 8 (3) (1995), 597–627. [412](#)
- [46] R. Charney, J. Meier, K. Whittlesey, Bestvina's normal form complex and the homology of Garside groups. *Geom. Dedicata* 105 (2004), 171–188. [414](#)
- [47] R. Charney, D. Peifer, The $K(\pi, 1)$ -conjecture for the affine braid groups. *Comment. Math. Helv.* 78 (3) (2003), 584–600. [406](#), [412](#)
- [48] Z. Chen, Topological representations of the braid groups and the noncommutative polynomial rings. In preparation. [434](#)
- [49] C. Chevalley, Invariants of finite groups generated by reflections. *Amer. J. Math.* 77 (1955), 778–782. [412](#)
- [50] A. M. Cohen, D. A. H. Gijssbers, D. B. Wales, BMW algebras of simply laced type. *J. Algebra* 286 (1) (2005), 107–153. [432](#)
- [51] A. M. Cohen, D. B. Wales, Linearity of Artin groups of finite type. *Israel J. Math.* 131 (2002), 101–123. [430](#), [431](#)
- [52] D. C. Cohen, A. I. Suciú, Homology of iterated semidirect products of free groups. *J. Pure Appl. Algebra* 126 (1–3) (1998), 87–120. [424](#)
- [53] F. R. Cohen, The homology of C_{n+1} -spaces, $n \geq 0$. In *The homology of iterated loop spaces*. Lecture Notes in Math. 533, Springer-Verlag, Berlin 1976, 207–353. [422](#)
- [54] D. Cox, J. Little, D. O'Shea, *Ideals, varieties, and algorithms*. Third edition, Undergrad. Texts Math., Springer-Verlag, New York 2007. [397](#)
- [55] H. S. M. Coxeter, Discrete groups generated by reflections. *Ann. of Math.* (2) 35 (3) (1934), 588–621. [409](#)
- [56] H. S. M. Coxeter, The complete enumeration of finite groups of the form $R_i^2 = (R_i R_j)^{k_{ij}} = 1$. *J. London Math. Soc.* 10 (1935), 21–25. [409](#)
- [57] J. Crisp, Injective maps between Artin groups. In *Geometric group theory down under* (Canberra, 1996), 119–137, Walter de Gruyter, Berlin 1999. [430](#), [432](#)
- [58] J. Crisp, L. Paris, The solution to a conjecture of Tits on the subgroup generated by the squares of the generators of an Artin group. *Invent. Math.* 145 (1) (2001), 19–36. [432](#), [437](#)
- [59] J. Crisp, L. Paris, Representations of the braid group by automorphisms of groups, invariants of links, and Garside groups. *Pacific J. Math.* 221 (1) (2005), 1–27. [403](#)
- [60] J. Crisp, L. Paris, Artin groups of type B and D . *Adv. Geom.* 5 (4) (2005), 607–636. [405](#), [435](#)

- [61] C. De Concini, C. Procesi, M. Salvetti, Arithmetic properties of the cohomology of braid groups. *Topology* 40 (4) (2001), 739–751. [424](#)
- [62] C. De Concini, C. Procesi, M. Salvetti, F. Stumbo, Arithmetic properties of the cohomology of Artin groups. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* 28 (4) (1999), 695–717. [424](#)
- [63] C. De Concini, M. Salvetti, Cohomology of Artin groups. *Math. Res. Lett.* 3 (2) (1996), 293–297. [430](#)
- [64] C. De Concini, M. Salvetti, F. Stumbo, The top-cohomology of Artin groups with coefficients in rank-1 local systems over \mathbb{Z} . *Topology Appl.* 78 (1–2) (1997), 5–20. [424](#)
- [65] P. Dehornoy, Braid groups and left distributive operations. *Trans. Amer. Math. Soc.* 345 (1) (1994), 115–150. [391](#)
- [66] P. Dehornoy, Groups with a complemented presentation. *J. Pure Appl. Algebra* 116 (1–3) (1997), 115–137. [415](#), [416](#)
- [67] P. Dehornoy, Groupes de Garside. *Ann. Sci. École Norm. Sup. (4)* 35 (2) (2002), 267–306. [413](#), [414](#), [416](#), [417](#)
- [68] P. Dehornoy, I. Dynnikov, D. Rolfsen, B. Wiest, *Why are braids orderable?* Panoramas et Synthèses 14, Soc. Math. France, Paris 2002. [391](#)
- [69] P. Dehornoy, Y. Lafont, Homology of Gaussian groups. *Ann. Inst. Fourier (Grenoble)* 53 (2) (2003), 489–540. [414](#)
- [70] P. Dehornoy, L. Paris, Gaussian groups and Garside groups, two generalisations of Artin groups. *Proc. London Math. Soc.* (3) 79 (3) (1999), 569–604. [390](#), [413](#), [416](#), [417](#), [419](#)
- [71] P. Deligne, Les immeubles des groupes de tresses généralisés. *Invent. Math.* 17 (1972), 273–302. [390](#), [406](#), [410](#), [412](#), [424](#)
- [72] V. V. Deodhar, On the root system of a Coxeter group. *Comm. Algebra* 10 (6) (1982), 611–630. [407](#)
- [73] T. tom Dieck, Categories of rooted cylinder ribbons and their representations. *J. Reine Angew. Math.* 494 (1998), 35–63. [406](#)
- [74] F. Digne, On the linearity of Artin braid groups. *J. Algebra* 268 (1) (2003), 39–57. [430](#), [431](#)
- [75] F. Digne, Présentations duales des groupes de tresses de type affine \tilde{A} . *Comment. Math. Helv.* 81 (1) (2006), 23–47. [413](#)
- [76] F. Digne, J. Michel, Garside and locally Garside categories. Preprint, 2006. [413](#)
- [77] J. L. Dyer, E. K. Grossman, The automorphism groups of the braid groups. *Amer. J. Math.* 103 (6) (1981), 1151–1169. [443](#)
- [78] E. A. El-Rifai, H. R. Morton, Algorithms for positive braids. *Quart. J. Math. Oxford Ser.* (2) 45 (1994), no. 180, 479–497. [419](#)
- [79] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. T. Levy, M. S. Paterson, W. P. Thurston, *Word processing in groups*. Jones and Bartlett Publishers, Boston, MA, 1992. [419](#)
- [80] E. Fadell, L. Neuwirth, Configuration spaces. *Math. Scand.* 10 (1962), 111–118. [398](#), [432](#)
- [81] N. Franco, J. González-Meneses, Conjugacy problem for braid groups and Garside groups. *J. Algebra* 266 (1) (2003), 112–132. [419](#)

- [82] È. V. Frenkel', Cohomology of the commutator subgroup of the braid group. *Funktsional Anal. i Prilozhen* 22 (3) (1988), 91–92; English transl. *Funct. Anal. Appl.* 22 (3) (1988), 248–250. [424](#)
- [83] P. Freyd, D. Yetter, J. Hoste, W. B. R. Lickorish, K. Millett, A. Ocneanu, A new polynomial invariant of knots and links. *Bull. Amer. Math. Soc. (N.S.)* 12 (2) (1985), 239–246. [390](#)
- [84] D. B. Fuks, Cohomology of the braid group mod 2. *Funktsional Anal. i Prilozhen.* 4 (2) (1970), 62–73; English transl. *Functional Anal. Appl.* 4 (1970), 143–151. [422](#)
- [85] F. A. Garside, The braid group and other groups. *Quart. J. Math. Oxford Ser. (2)* 20 (1969), 235–254. [390](#), [419](#)
- [86] V. Gebhardt, A new approach to the conjugacy problem in Garside groups. *J. Algebra* 292 (1) (2005), 282–302. [419](#)
- [87] V. Gebhardt, J. González-Meneses, Solving the conjugacy problem in Garside groups by cyclic slidings. In preparation. [419](#), [420](#)
- [88] S. Gervais, A finite presentation of the mapping class group of a punctured surface. *Topology* 40 (4) (2001), 703–725. [441](#)
- [89] E. Godelle, Morphismes injectifs entre groupes d'Artin-Tits. *Algebr. Geom. Topol.* 2 (2002), 519–536. [432](#)
- [90] J. González-Meneses, V. Gebhardt, On the cycling operation in braid groups. *Discrete Appl. Math.* 156 (2008), 3072–3090. [419](#)
- [91] V. V. Goryunov, The cohomology of braid groups of series C and D and certain stratifications. *Funktsional Anal. i Prilozhen.* 12 (2) (1978), 76–77; English transl. *Functional Anal. Appl.* 12 (2) (1978), 139–149. [423](#)
- [92] M.-E. Hamstrom, Homotopy groups of the space of homeomorphisms of a 2-manifold. *Illinois J. Math.* 10 (1966), 563–573. [399](#)
- [93] V. L. Hansen, *Braids and coverings: selected topics*. London Math. Soc. Student Texts 18, Cambridge University Press, Cambridge 1989. [391](#)
- [94] P. de la Harpe, An invitation to Coxeter groups. In *Group theory from a geometrical viewpoint* (Trieste, 1990), World Scientific Publishing, River Edge, NJ, 1991, 193–253. [405](#)
- [95] J.-Y. Hée, Sur la fidélité de la représentation de Krammer-Paris. Preprint, 2007. [432](#)
- [96] S.-T. Hu, *Homotopy theory*. Pure and Applied Math. VIII, Academic Press, New York-London, 1959. [397](#), [400](#)
- [97] J. E. Humphreys, *Reflection groups and Coxeter groups*. Cambridge Stud. Adv. Math. 29, Cambridge University Press, Cambridge 1990. [407](#)
- [98] N. V. Ivanov, Algebraic properties of the Teichmüller modular group. *Dokl. Akad. Nauk SSSR* 275 (4) (1984), 786–789; English transl. *Soviet Math. Dokl.* 29 (2) (1984), 288–291. [443](#)
- [99] N. V. Ivanov, Mapping class groups. In *Handbook of geometric topology*, North-Holland, Amsterdam 2002, 523–633. [403](#)
- [100] D. L. Johnson, M. A. Albar, The center of the circular braid group. *Math. Japon.* 30 (4) (1985), 641–645. [406](#)
- [101] V. F. R. Jones, A polynomial invariant for knots via von Neumann algebras. *Bull. Amer.*

- Math. Soc. (N.S.)* 12 (1) (1985), 103–111. [390](#)
- [102] V. F. R. Jones, Hecke algebra representations of braid groups and link polynomials. *Ann. of Math. (2)* 126 (2) (1987), 335–388. [390](#)
- [103] C. Kassel, V. Turaev, *Braid groups*. Grad. Texts in Math. 247, Springer-Verlag, New York 2008. [391](#)
- [104] R. P. Kent, D. Peifer, A geometric and algebraic description of annular braid groups. *Internat. J. Algebra Comput.* 12 (1–2) (2002), 85–97. [406](#)
- [105] D. Krammer, The braid group B_4 is linear. *Invent. Math.* 142 (3) (2000), 451–486. [391](#), [433](#)
- [106] D. Krammer, Braid groups are linear. *Ann. of Math. (2)* 155 (1) (2002), 131–156. [391](#), [430](#), [431](#), [432](#), [433](#)
- [107] D. Krammer, A class of Garside groupoid structures on the pure braid group. *Trans. Amer. Math. Soc.* 360 (8) (2008), 4029–4061. [413](#)
- [108] D. Krammer, The braid group of \mathbb{Z}^n . *J. London Math. Soc. (2)* 76 (2) (2007), 293–312. [413](#)
- [109] C. Labruère, Generalized braid groups and mapping class groups. *J. Knot Theory Ramifications* 6 (5) (1997), 715–726. [437](#), [438](#)
- [110] C. Labruère, L. Paris, Presentations for the punctured mapping class groups in terms of Artin groups. *Algebr. Geom. Topol.* 1 (2001), 73–114. [434](#), [438](#), [439](#), [440](#), [441](#)
- [111] C. Landi, Cohomology rings of Artin groups. *Atti. Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* 11 (1) (2000), 41–65. [424](#)
- [112] R. J. Lawrence, Homological representations of the Hecke algebra. *Comm. Math. Phys.* 135 (1) (1990), 141–191. [434](#)
- [113] E. K. Lee, S. J. Lee, Translation numbers in a Garside group are rational with uniformly bounded denominators. *J. Pure Appl. Algebra* 211 (3) (2007), 732–743. [419](#)
- [114] E. K. Lee, S. J. Lee, Some power of an element in a Garside group is conjugate to a periodically geodesic element. *Bull. London Math. Soc.* 40 (4) (2008), 593–603. [419](#)
- [115] S. J. Lee, Garside groups are strongly translation discrete. *J. Algebra* 309 (2) (2007), 594–609. [419](#)
- [116] H. van der Lek, The homotopy type of complex hyperplane complements. Ph. D. Thesis, Nijmegen, 1983. [412](#), [439](#)
- [117] C. Maclachlan, W. J. Harvey, On mapping-class groups and Teichmüller spaces. *Proc. London Math. Soc. (3)* 30 (4) (1975), 496–512. [435](#)
- [118] W. Magnus, Über Automorphismen von Fundamentalgruppen berandeter Flächen. *Math. Ann.* 109 (1934), 617–646. [394](#), [402](#)
- [119] W. Magnus, A. Karrass, D. Solitar, *Combinatorial group theory. Presentations of groups in terms of generators and relations*. Reprint of the 1976 second edition, Dover Publications, Inc., Mineola, NY, 2004. [402](#)
- [120] I. Marin, Représentations linéaires des tresses infinitésimales. Ph. D. Thesis, Université Paris-Sud, 2001. [432](#)
- [121] I. Marin, Sur les représentations de Krammer génériques. *Ann. Inst. Fourier (Grenoble)* 57 (6) (2007), 1883–1925. [432](#)

- [122] N. S. Markaryan, Homology of braid groups with nontrivial coefficients. *Mat. Zametki* 59 (6) (1996), 846–854; English transl. *Math. Notes* 59 (5–6) (1996), 611–617. [424](#)
- [123] A. Markoff, *Foundations of the algebraic theory of tresses*. Trav. Inst. Math. Stekloff 16 (1945). [394](#)
- [124] M. Matsumoto, A presentation of mapping class groups in terms of Artin groups and geometric monodromy of singularities. *Math. Ann.* 316 (3) (2000), 401–418. [438](#), [439](#)
- [125] J. D. McCarthy, Automorphisms of surface mapping class groups. A recent theorem of N. Ivanov. *Invent. Math.* 84 (1) (1986), 49–71. [443](#)
- [126] J. Michel, A note on words in braid monoids. *J. Algebra* 215 (1) (1999), 366–377. [410](#)
- [127] S. Morita, Introduction to mapping class groups of surfaces and related groups. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 353–386. [403](#)
- [128] K. Murasugi, B. I. Kurpita, *A study of braids*. Math. Appl. 484, Kluwer Academic Publishers, Dordrecht 1999. [391](#)
- [129] J. Nielsen, Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen. *Acta Math.* 50 (1927), 189–358. [391](#), [402](#)
- [130] L. Paoluzzi, L. Paris, A note on the Lawrence-Krammer-Bigelow representation. *Algebr. Geom. Topol.* 2 (2002), 499–518. [433](#)
- [131] L. Paris, Universal cover of Salvetti’s complex and topology of simplicial arrangements of hyperplanes. *Trans. Amer. Math. Soc.* 340 (1) (1993), 149–178. [424](#)
- [132] L. Paris, Parabolic subgroups of Artin groups. *J. Algebra* 196 (2) (1997), 369–399. [439](#)
- [133] L. Paris, Artin monoids inject in their groups. *Comment. Math. Helv.* 77 (3) (2002), 609–637. [410](#), [430](#), [431](#), [432](#)
- [134] L. Paris, D. Rolfsen, Geometric subgroups of mapping class groups. *J. Reine Angew. Math.* 521 (2000), 47–83. [441](#), [443](#)
- [135] B. Perron, J.P. Vannier, Groupe de monodromie géométrique des singularités simples. *Math. Ann.* 306 (2) (1996), 231–245. [405](#), [406](#), [435](#), [437](#)
- [136] M. Picantin, The conjugacy problem in small Gaussian groups. *Comm. Algebra* 29 (3) (2001), 1021–1039. [419](#)
- [137] M. Picantin, Automatic structures for torus link groups. *J. Knot Theory Ramifications* 12 (6) (2003), 833–866. [414](#)
- [138] J. H. Przytycki, P. Traczyk, Invariants of links of Conway type. *Kobe J. Math.* 4 (2) (1988), 115–139. [390](#)
- [139] C. Rourke, B. Wiest, Order automatic mapping class groups. *Pacific J. Math.* 194 (1) (2000), 209–227. [391](#)
- [140] M. Salvetti, Topology of the complement of real hyperplanes in \mathbb{C}^N . *Invent. Math.* 88 (3) (1987), 603–618. [424](#), [425](#)
- [141] M. Salvetti, On the homotopy theory of complexes associated to metrical-hemisphere complexes. *Discrete Math.* 113 (1–3) (1993), 155–177. [424](#)
- [142] M. Salvetti, The homotopy type of Artin groups. *Math. Res. Lett.* 1 (5) (1994), 565–577. [423](#), [424](#), [429](#)

- [143] G. Segal, Configuration-spaces and iterated loop-spaces. *Invent. Math.* 21 (1973), 213–221. [422](#)
- [144] G. C. Shephard, J. A. Todd, Finite unitary reflection groups. *Canad. J. Math.* 6 (1954), 274–304. [412](#)
- [145] C. C. Squier, The homological algebra of Artin groups. *Math. Scand.* 75 (1) (1994), 5–43. [430](#)
- [146] J. Tits, Normalisateurs de tores. I. Groupes de Coxeter étendus. *J. Algebra* 4 (1966), 96–116. [390](#)
- [147] J. Tits, Le problème des mots dans les groupes de Coxeter. In *Symposia Mathematica* (INDAM, Rome, 1967/68), Vol. 1, Academic Press, London 1969, 175–185. [408](#)
- [148] J. Tits, Groupes et géométries de Coxeter. In *Wolf Prize in Mathematics*, Vol. 2, S. S. Chern and F. Hirzebruch, eds., World Scientific Publishing, River Edge, NJ, 2001, 740–754. [407](#)
- [149] F. V. Vainšteĭn, The cohomology of braid groups. *Funktsional. Anal. i Prilozhen.* 12 (2) (1978), 72–73; English transl. *Functional Anal. Appl.* 12 (2) (1978), 135–137. [422](#)
- [150] V. V. Vershinin, Braid groups and loop spaces. *Uspekhi Mat. Nauk* 54 (2) (1999), 3–84; English transl. *Russian Math. Surveys* 54 (2) (1999), 273–350. [424](#)
- [151] M. Wada, Group invariants of links. *Topology* 31 (2) (1992), 399–406. [403](#)
- [152] B. Wajnryb, Artin groups and geometric monodromy. *Invent. Math.* 138 (3) (1999), 563–571. [437](#)
- [153] H. Zheng, Faithfulness of the Lawrence representation of braid groups. Preprint, 2005. [434](#)
- [154] H. Zieschang, On the homeotopy group of surfaces. *Math. Ann.* 206 (1973), 1–21. [435](#)
- [155] M. G. Zinno, On Kramer’s representation of the braid group. *Math. Ann.* 321 (1) (2001), 197–211. [432](#)

Part C

Representation spaces and geometric structures, 1

Chapter 12

Complex projective structures

David Dumas

Contents

1	Introduction	455
2	Basic definitions	458
3	The Schwarzian parameterization	461
3.1	The Schwarzian derivative	461
3.2	Schwarzian parameterization of a fiber	462
3.3	Schwarzian parameterization of $\mathcal{P}(S)$	464
4	The Grafting parameterization	466
4.1	Definition of grafting	467
4.2	Thurston's theorem	469
4.3	The Thurston metric	473
4.4	Conformal grafting maps	475
5	Holonomy	477
5.1	Representations and characters	477
5.2	The holonomy map	479
5.3	Holonomy and bending	480
5.4	Fuchsian holonomy	483
5.5	Quasi-Fuchsian holonomy	485
5.6	Discrete holonomy	488
5.7	Holonomy in fibers	489
6	Comparison of parameterizations	491
6.1	Compactifications	491
6.2	Quadratic differentials and measured laminations	492
6.3	Limits of fibers	494
6.4	Limits of the Schwarzian	498
6.5	Infinitesimal compatibility	501
	References	502

1 Introduction

In this chapter we discuss the theory of complex projective structures on compact surfaces and its connections with Teichmüller theory, 2- and 3-dimensional hyperbolic geometry, and representations of surface groups into $\mathrm{PSL}_2(\mathbb{C})$. Roughly speaking, a

complex projective structure is a type of 2-dimensional geometry in which Möbius transformations play the role of geometric congruences (this is made precise below). Such structures are abundant – hyperbolic, spherical, and Euclidean metrics on surfaces all provide examples of projective structures, since each of these constant-curvature 2-dimensional geometries has a model in which its isometries are Möbius maps. However, these examples are not representative of the general situation, since most projective structures are not induced by locally homogeneous Riemannian metrics.

Developing a more accurate picture of a general projective structure is the goal of the first half of the chapter (§§2–4). After some definitions and preliminary discussion (in §2), we present the complex-analytic theory of projective structures in §3. This theory has its roots in the study of automorphic functions and differential equations by Klein [67, Part 1], Poincaré [95], Riemann [100], and others in the late nineteenth century (see [48], [47, §1] for further historical discussion and references), while its more recent history is closely linked to developments in Teichmüller theory and deformations of Fuchsian and Kleinian groups (e.g. [28], [45], [43], [47] [51], [74], [75], [76], [77]).

In this analytic approach, a projective structure is represented by a holomorphic quadratic differential on a Riemann surface, which is extracted from the geometric data using a Möbius-invariant differential operator, the *Schwarzian derivative*. The inverse of this construction describes every projective structure in terms of holomorphic solutions to a linear ordinary differential equation (the *Schwarzian equation*). In this way, many properties of projective structures and their moduli can be established using tools from complex function theory. However, in spite of the success of these techniques, the analytic theory is somewhat detached from the underlying geometry. In particular, the analytic parameterization of projective structures does not involve an explicit geometric construction, such as one has in the description of hyperbolic surfaces by gluing polygons.

In §4 we describe a more direct and geometric construction of complex projective structures using *grafting*, a gluing operation on surfaces which is also suggested by the work of the nineteenth-century geometers (e.g. [68]), but whose significance in complex projective geometry has only recently been fully appreciated. Grafting was used by Maskit [83], Hejhal [47], and Sullivan–Thurston [109] to construct certain deformations of Fuchsian groups, and in later work of Thurston (unpublished, see [64]) it was generalized to give a universal construction of complex projective surfaces starting from basic hyperbolic and Euclidean pieces.

This construction provides another coordinate system for the moduli space of projective structures, and it reveals an important connection between these structures and convex geometry in 3-dimensional hyperbolic space. However, the explicit geometric nature of complex projective grafting comes at the price of a more complicated parameter space, namely, the piecewise linear manifold of *measured geodesic laminations* on hyperbolic surfaces. In particular, the lack of a differentiable structure in this coordinate system complicates the study of variations of complex projective

structures, though there has been some progress in this direction using a weak notion of differentiability due to Thurston [115] and Bonahon [10].

After developing the analytic and geometric coordinates for the moduli space of projective structures, the second half of the chapter is divided into two major topics: In §5, we describe the relation between projective structures and the $\mathrm{PSL}_2(\mathbb{C})$ -representations of surface groups, their deformations, and associated problems in hyperbolic geometry and Kleinian groups. The key to these connections is the *holonomy representation* of a projective structure, which records the topological obstruction to analytically continuing its local coordinate charts over the entire surface. After constructing a parameter space for such representations and the holonomy map for projective structures, we survey various developments that center around two basic questions:

- Given a projective structure, described in either analytic or geometric terms, what can be said about its holonomy representation?
- Given a $\mathrm{PSL}_2(\mathbb{C})$ -representation of a surface group, what projective structures have this as their holonomy representation, if any?

We discuss partial answers to these general questions, along with much more detailed information about certain classes of holonomy representations (e.g. Fuchsian groups).

Finally, in §6 we take up the question of relating the analytic and geometric coordinate systems for the space of projective structures, or equivalently, studying the interaction between the Schwarzian derivative and complex projective grafting. We describe asymptotic results that relate compactifications of the analytic and geometric parameter spaces using the geometry of measured foliations on Riemann surfaces. Here a key tool is the theory of harmonic maps between Riemann surfaces and from Riemann surfaces to \mathbb{R} -trees, and the observation that two geometrically natural constructions in complex projective geometry (the *collapsing* and *co-collapsing* maps) are closely approximated by harmonic maps. We close with some remarks concerning infinitesimal compatibility between the geometric and analytic coordinate systems, once again using the limited kind of differential calculus that applies to the grafting parameter space.

Scope and approach. Although this chapter covers a range of topics in complex projective geometry, it is not intended to be a comprehensive guide to the subject. Rather, we have selected several important aspects of the theory (the Schwarzian derivative, grafting, and holonomy) and concentrated on describing their interrelationships while providing references for further reading and exploration. As a result, some major areas of research in complex projective structures are not mentioned at all (circle packings [70], [69], the algebraic-geometric aspects of the theory [36, §11], and generalizations to punctured or open Riemann surfaces [75], [81], to name a few) and others are only discussed in brief.

We have also included some detail on the basic analytic and geometric constructions in an attempt to make this chapter a more useful “invitation” to the theory. However, where we discuss more advanced topics and results of recent research, it has been

necessary to refer to many concepts and results that are not thoroughly developed here.

Finally, while we have attempted to provide thorough and accurate references to the literature, the subject of complex projective structures is broad enough (and connected to so many other areas of research) that we do not expect these references to cover every relevant source of additional information. We hope that the references included below are useful, and regret any inadvertent omissions.

Acknowledgments. The author thanks Richard Canary, George Daskalopoulos, William Goldman, Brice Loustau, Albert Marden, Athanase Papadopoulos, Richard Wentworth, and Michael Wolf for helpful discussions and suggestions related to this work, and Curt McMullen for introducing him to the theory of complex projective structures. This work was partially supported by a NSF postdoctoral research fellowship.

2 Basic definitions

Projective structures. Let S be an oriented surface. A *complex projective structure* Z on S is a maximal atlas of charts mapping open sets in S into \mathbb{CP}^1 such that the transition functions are restrictions of Möbius transformations. For brevity we also call these *projective structures* or \mathbb{CP}^1 -*structures*.

We often treat a projective structure Z on S as a surface in its own right – a *complex projective surface*. Differentiably, Z is the same as S , but Z has the additional data of a restricted atlas of projective charts.

Two projective structures Z_1 and Z_2 on S are *isomorphic* if there is an orientation-preserving diffeomorphism $\iota: Z_1 \rightarrow Z_2$ that pulls back the projective charts of Z_2 to projective charts of Z_1 , and *marked isomorphic* if furthermore ι is homotopic to the identity.

Our main object of study is the space $\mathcal{P}(S)$ of marked isomorphism classes of projective structures on a compact surface S . Thus far, we have only defined $\mathcal{P}(S)$ as a set, but later we will equip it with the structure of a complex manifold.

Non-hyperbolic cases. Projective structures on compact surfaces are most interesting when S has genus $g \geq 2$: The sphere has a unique projective structure (by $S^2 \simeq \mathbb{CP}^1$) up to isotopy, while a projective structure on a torus is always induced by an affine structure [43, §9, pp. 189–191]. We therefore make the assumption that S has genus $g \geq 2$ unless stated otherwise.

First examples. The projective structure of \mathbb{CP}^1 itself (using the identity for chart maps) also gives a natural projective structure on any open set $U \subset \mathbb{CP}^1$. If U is preserved by a group Γ of Möbius transformations acting freely and properly discontinuously, then the quotient surface $X = U/\Gamma$ has a natural projective structure in which the charts are local inverses of the covering $U \rightarrow X$.

In particular any Fuchsian group $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ gives rise to a projective structure on the quotient surface \mathbb{H}/Γ and a Kleinian group $\Gamma \subset \mathrm{PSL}_2(\mathbb{C})$ gives a projective structure on the quotient of its domain of discontinuity $\Omega(\Gamma)/\Gamma$. Rephrasing the latter example, the ideal boundary of a hyperbolic 3-manifold has a natural projective structure.

Locally Möbius maps. A map $f: Z \rightarrow W$ between complex projective surfaces is *locally Möbius* if for every sufficiently small open set $U \subset Z$, the restriction $f|_U$ is a Möbius transformation with respect to projective coordinates on U and $f(U)$. Examples of such maps include isomorphisms and covering maps of projective surfaces (where the cover is given the pullback projective structure) and inclusions of open subsets of surfaces.

Developing maps. A projective structure Z on a surface S lifts to a projective structure \tilde{Z} on the universal cover \tilde{S} . A *developing map* for Z is an immersion $f: \tilde{S} \rightarrow \mathbb{CP}^1$ such that the restriction of f to any sufficiently small open set in \tilde{S} is a projective chart for \tilde{Z} . Such a map is also called a *geometric realization* of Z (e.g. [45, §6]) or a *fundamental membrane* [47].

Developing maps always exist, and are essentially unique – two developing maps for a given structure differ by post-composition with a Möbius transformation. Concretely, a developing map can be constructed by analytic continuation starting from any basepoint $z_0 \in \tilde{Z}$ and any chart defined on a neighborhood U of z_0 . Another chart $V \rightarrow \mathbb{CP}^1$ that overlaps U can be adjusted by a Möbius transformation so as to agree on the overlap, gluing to give a map $(U \cup V) \rightarrow \mathbb{CP}^1$. Continuing in this way one defines a map on successively larger subsets of \tilde{Z} , and the limit is a developing map $\tilde{Z} \rightarrow \mathbb{CP}^1$. The simple connectivity of \tilde{Z} is essential here, as nontrivial homotopy classes of loops in the surface create obstructions to unique analytic continuation of a projective chart.

For a fixed projective structure, we will speak of *the* developing map when the particular choice is unimportant or implied.

Holonomy representation. The developing map $f: \tilde{S} \rightarrow \mathbb{CP}^1$ of a projective structure Z on S has an equivariance property with respect to the action of $\pi_1(S)$ on \tilde{S} : For any $\gamma \in \pi_1(S)$, the composition $f \circ \gamma$ is another developing map for Z . Thus there exists $A_\gamma \in \mathrm{PSL}_2(\mathbb{C})$ such that

$$f \circ \gamma = A_\gamma \circ f. \quad (2.1)$$

The map $\gamma \mapsto A_\gamma$ is a homomorphism $\rho: \pi_1(S) \rightarrow \mathrm{PSL}_2(\mathbb{C})$, the *holonomy representation* (or *monodromy representation*) of the projective structure.

Development-holonomy pairs. The developing map and holonomy representation form the *development-holonomy pair* (f, ρ) associated to the projective structure Z .

This pair determines Z uniquely, since restriction of f determines a covering of S by projective charts. Post-composition of the developing map with $A \in \mathrm{PSL}_2(\mathbb{C})$

conjugates ρ , and therefore the pair (f, ρ) is uniquely determined by Z up to the action of $\mathrm{PSL}_2(\mathbb{C})$ by

$$(f, \rho) \mapsto (A \circ f, \rho^A) \quad \text{where } \rho^A(\gamma) = A\rho(\gamma)A^{-1}.$$

Conversely, any pair (f, ρ) consisting of an immersion $f: \tilde{S} \rightarrow \mathbb{CP}^1$ and a homomorphism $\rho: \pi_1(S) \rightarrow \mathrm{PSL}_2(\mathbb{C})$ that satisfy (2.1) defines a projective structure on S in which lifting $U \subset S$ to \tilde{S} and applying f gives a projective chart (for all sufficiently small open sets U).

Thus we have an alternate definition of $\mathcal{P}(S)$ as the quotient of the set of development-holonomy pairs by the $\mathrm{PSL}_2(\mathbb{C})$ action and by precomposition of developing maps with orientation-preserving diffeomorphisms of S homotopic to the identity. We give the set of pairs of maps (f, ρ) the compact-open topology, and $\mathcal{P}(S)$ inherits a quotient topology. We will later see that $\mathcal{P}(S)$ is homeomorphic to \mathbb{R}^{12g-12} .

Relation to (G, X) -structures. There is a very general notion of a geometric structure defined by a Lie group G acting by diffeomorphisms on a manifold X . A (G, X) -structure on a manifold M is an atlas of charts mapping open subsets of M into X such that the transition maps are restrictions of elements of G .

In this language, complex projective structures are $(\mathrm{PSL}_2(\mathbb{C}), \mathbb{CP}^1)$ -structures. Some of the properties of projective structures we develop, such as developing maps, holonomy representations, deformation spaces, etc., can be applied in the more general setting of (G, X) -structures. See [40] for a survey of (G, X) -structures and analysis of several low-dimensional examples.

Circles. Because Möbius transformations map circles to circles, there is a natural notion of a circle on a surface with a projective structure Z : A smooth embedded curve $\alpha \subset Z$ is a *circular arc* if the projective charts map (subsets of) α to circular arcs in \mathbb{CP}^1 . Equivalently, the embedded curve α is a circular arc if the developing map sends any connected component of the preimage of α in \tilde{Z} to a circular arc in \mathbb{CP}^1 . A closed circular arc on Z is a *circle*.

Small circles are ubiquitous in any projective structure: For any $z \in Z$ there is a projective chart mapping a contractible neighborhood of z to an open set $V \in \mathbb{CP}^1$. The preimage of any circle contained in V is a homotopically trivial circle for the projective structure Z . Circles that bound disks on a projective surface have an important role in Thurston's projective grafting construction (see §4.1).

Circles on a projective surface can also be homotopically nontrivial. For example any simple closed geodesic on a hyperbolic surface on X is a circle, because its lifts to $\tilde{X} \simeq \mathbb{H}$ are half-circles or vertical lines in the upper half-plane. The analysis of circles on more general projective surfaces would be a natural starting point for the development of synthetic complex projective geometry; Wright's study of circle chains and Schottky-type dynamics in the Maskit slice of punctured tori is an example of work in this direction [120].

Forgetful map. Since Möbius transformations are holomorphic, a projective structure $Z \in \mathcal{P}(S)$ also determines a complex structure, making S into a compact Riemann surface. In this way, marked isomorphism of projective structures corresponds to marked isomorphism of Riemann surfaces, and so there is a natural (and continuous) *forgetful map*

$$\pi: \mathcal{P}(S) \rightarrow \mathcal{T}(S)$$

where $\mathcal{T}(S)$ is the Teichmüller space of marked isomorphism classes of complex structures on S . (See e.g. [79], [56], [52] and the other chapters of this Handbook for background on Teichmüller spaces.) As a matter of terminology, if Z is a projective structure with $\pi(Z) = X$, we say Z is a projective structure *on the Riemann surface X* .

The forgetful map is surjective: By the uniformization theorem, every complex structure $X \in \mathcal{T}(S)$ arises as the quotient of \mathbb{H} by a Fuchsian group Γ_X , and the natural projective structure on \mathbb{H}/Γ_X is a preimage of X by π . We call this the *standard Fuchsian structure* on X . The standard Fuchsian structures determine a continuous section

$$\sigma_0: \mathcal{T}(S) \rightarrow \mathcal{P}(S).$$

One might expect the fibers of π to be large, since isomorphism of projective structures is a much stronger condition than isomorphism of complex structures. Our next task is to describe the fibers explicitly.

3 The Schwarzian parameterization

3.1 The Schwarzian derivative

Let $\Omega \subset \mathbb{C}$ be a connected open set. The *Schwarzian derivative* of a locally injective holomorphic map $f: \Omega \rightarrow \mathbb{C}\mathbb{P}^1$ is the holomorphic quadratic differential

$$S(f) = \left[\left(\frac{f''(z)}{f'(z)} \right)' - \frac{1}{2} \left(\frac{f''(z)}{f'(z)} \right)^2 \right] dz^2.$$

Two key properties make the Schwarzian derivative useful in the theory of projective structures:

- (1) *Cocycle property.* If f and g are locally injective holomorphic maps such that the composition $f \circ g$ is defined, then

$$S(f \circ g) = g^*S(f) + S(g).$$

- (2) *Möbius invariance.* For any $A \in \mathrm{PSL}_2(\mathbb{C})$, we have

$$S(A) \equiv 0,$$

and conversely, if $S(f) \equiv 0$, then f is the restriction of a Möbius transformation.

Note that the pullback $g^*S(f)$ uses the definition of the Schwarzian as a quadratic differential. In classical complex analysis, the Schwarzian was regarded as a complex-valued function, with $g^*S(f)$ replaced by $g'(z)^2S(f)(g(z))$.

An elementary consequence of these properties is that the map f is almost determined by its Schwarzian derivative; if $S(f) = S(g)$, then the locally defined map $f \circ g^{-1}$ satisfies $S(f \circ g^{-1}) \equiv 0$, and so we have $f = A \circ g$ for some $A \in \text{PSL}_2(\mathbb{C})$.

Further discussion of the Schwarzian derivative can be found in e.g. [79, Chapter 2], [52, §6.3].

Osculation. Intuitively, the Schwarzian derivative measures the failure of a holomorphic map to be the restriction of a Möbius transformation. Thurston made this intuition precise as follows (see [116, §2], [2, §2.1]): For each $z \in \Omega$, there is a unique Möbius transformation that has the same 2-jet as f at z , called the *osculating Möbius transformation* $\text{osc}_z f$.

The *osculation map* $G: \Omega \rightarrow \text{PSL}_2(\mathbb{C})$ given by $G(z) = \text{osc}_z f$ is holomorphic, and its Darboux derivative (see [105]) is the holomorphic $\mathfrak{sl}_2(\mathbb{C})$ -valued 1-form

$$\omega(z) = G^{-1}(z) dG(z).$$

An explicit computation shows that ω only depends on f through its Schwarzian derivative; if $S(f) = \phi(z)dz^2$, then

$$\omega(z) = -\frac{1}{2}\phi(z) \begin{pmatrix} z & -z^2 \\ 1 & -z \end{pmatrix} dz.$$

3.2 Schwarzian parameterization of a fiber

Fibers over Teichmüller space. For any marked complex structure $X \in \mathcal{T}(S)$, let $P(X) = \pi^{-1}(X) \subset \mathcal{P}(S)$ denote the set of marked complex projective structures with underlying complex structure X . The Schwarzian derivative can be used to parameterize the fiber $P(X)$ as follows:

Fix a conformal identification $\tilde{X} \simeq \mathbb{H}$, whereby $\pi_1(S)$ acts on \mathbb{H} as a Fuchsian group. Abusing notation, we use the same symbol for $\gamma \in \pi_1(S)$ and for its action on \mathbb{H} by a real Möbius transformation.

Given $Z \in P(X)$, we regard the developing map as a meromorphic function f on \mathbb{H} . The Schwarzian derivative $\tilde{\phi} = S(f)$ is therefore a holomorphic quadratic differential on \mathbb{H} . Combining the equivariance property (2.1) of f and the properties of the Schwarzian derivative, we find

$$\tilde{\phi} = S(A_\gamma \circ f) = S(f \circ \gamma) = \gamma^* \tilde{\phi}.$$

Thus we have $\tilde{\phi} = \gamma^* \tilde{\phi}$ for all $\gamma \in \pi_1(S)$, and $\tilde{\phi}$ descends to a holomorphic quadratic differential ϕ on X . We call ϕ the *Schwarzian of the projective structure* Z .

Let $Q(X)$ denote the vector space of holomorphic quadratic differentials on the marked Riemann surface $X \in \mathcal{T}(S)$. By the Riemann–Roch theorem, we have

$Q(X) \simeq \mathbb{C}^{3g-3}$ (see [62]). The Schwarzian defines a map $P(X) \rightarrow Q(X)$. We will now show that this map is bijective by constructing its inverse.

Inverting the Schwarzian. Let $\phi(z)$ be a holomorphic function defined on a contractible open set $\Omega \subset \mathbb{C}$. Then the linear ODE (the *Schwarzian equation*)

$$u''(z) + \frac{1}{2}\phi(z)u(z) = 0 \quad (3.1)$$

has a two-dimensional vector space V of holomorphic solutions on Ω . Let $u_1(z)$ and $u_2(z)$ be a basis of solutions. The Wronskian $W(z)$ of u_1 and u_2 satisfies $W'(z) = 0$, so it is a nonzero constant function, and u_1 and u_2 cannot vanish simultaneously.

This ODE construction inverts the Schwarzian derivative in the sense that the meromorphic function $f(z) = u_1(z)/u_2(z)$ satisfies $S(f) = \phi(z)dz^2$ (see [91]). Note that changing the basis for V will alter f by composition with a Möbius transformation (and leave $S(f)$ unchanged). Furthermore, since

$$f'(z) = \frac{u_1'(z)u_2(z) - u_1(z)u_2'(z)}{u_2(z)^2} = \frac{-W(z)}{u_2(z)^2},$$

it follows that the holomorphic map $f: \Omega \rightarrow \mathbb{C}\mathbb{P}^1$ is locally injective except possibly on $\{u_2(z) = 0\} = f^{-1}(\infty)$. Applying similar considerations to $1/f(z)$, we find that f is locally injective away from $\{u_1(z) = 0\}$, and thus everywhere.

The existence of a holomorphic map with a given Schwarzian derivative can also be understood in terms of maps to the Lie group $\mathrm{PSL}_2(\mathbb{C})$ and the definition of the Schwarzian in terms of osculation (described in §3.1). Here the quadratic differential ϕ is interpreted as a $\mathfrak{sl}_2(\mathbb{C})$ -valued 1-form, which satisfies the integrability condition $d\phi + \frac{1}{2}[\phi, \phi] = 0$ because there are no holomorphic 2-forms on a Riemann surface. The integrating map to $\mathrm{PSL}_2(\mathbb{C})$ is the osculation map of a holomorphic function f satisfying $S(f) = \phi$. See [2, §2.2.3, Corollary 2.20] for details.

Parameterization of a fiber. Given a quadratic differential $\phi \in Q(X)$, lift to the universal cover $\tilde{X} \simeq \mathbb{H}$ to obtain $\tilde{\phi} = \tilde{\phi}(z) dz^2$. Applying the ODE construction to $\tilde{\phi}(z)$ yields a holomorphic immersion $f_\phi: \mathbb{H} \rightarrow \mathbb{C}\mathbb{P}^1$.

For any $\gamma \in \pi_1(S)$ we have $S(f_\phi \circ \gamma) = \gamma^*\tilde{\phi} = \tilde{\phi} = S(f_\phi)$, and thus $f_\phi \circ \gamma = A_\gamma \circ f_\phi$ for some $A_\gamma \in \mathrm{PSL}_2(\mathbb{C})$. We set $\rho_\phi(\gamma) = A_\gamma$. Then (f_ϕ, ρ_ϕ) determine a development-holonomy pair, and thus a projective structure X_ϕ on S . Since f is holomorphic, we also have $\pi(X_\phi) = X$.

The map $Q(X) \rightarrow P(X)$ given by $\phi \mapsto X_\phi$ is inverse to the Schwarzian map $P(X) \rightarrow Q(X)$ because the ODE construction is inverse to the Schwarzian derivative. In particular, each fiber of $\pi: \mathcal{P}(S) \rightarrow \mathcal{T}(S)$ is naturally parameterized by a complex vector space.

Affine naturality. The identification $Q(X) \simeq P(X)$ defined above depends on a choice of coordinates on the universal cover of X . Specifically, we computed the Schwarzian using the coordinate z of the upper half plane.

A coordinate-independent statement is that the Schwarzian derivative is a measure of the *difference* between a pair of projective structures on X , which we can see as follows: Given $Z_1, Z_2 \in P(X)$, let U be a sufficiently small open set on S so that there are projective coordinate charts $z_i : U \rightarrow \mathbb{C}\mathbb{P}^1$ of Z_i for $i = 1, 2$. We can assume that $\infty \notin z_i(U)$.

The quadratic differential $z_1^* S(z_2 \circ z_1^{-1})$ on U is holomorphic with respect to the Riemann surface structure X . Covering S by such sets, it follows from the cocycle property that these quadratic differentials agree on overlaps and define an element $\phi \in Q(X)$, which is the *Schwarzian of Z_2 relative to Z_1* . Abusing notation, we write $Z_2 - Z_1 = \phi$.

Thus $P(X)$ has a natural structure of an *affine space* modeled on the vector space $Q(X)$. The choice of a basepoint $Z_0 \in P(X)$ gives an isomorphism $P(X) \rightarrow Q(X)$, namely $Z \mapsto (Z - Z_0)$. See [51, §2] for details.

From this perspective, the previous identification $P(X) \rightarrow Q(X)$ using the Schwarzian of the developing map on \mathbb{H} is simply $Z \mapsto (Z - \sigma_0(X))$, that is, it is the Schwarzian relative to the standard Fuchsian structure. Complex-analytically, this is not the most natural way to choose a basepoint in each fiber, though this will be remedied below (§3.3).

The realization of $P(X)$ as an affine space modeled on a vector space of differential forms can also be understood in terms of Čech cochains on X with a fixed coboundary [45, §3], or in terms of connections on a principal $\mathrm{PSL}_2(\mathbb{C})$ -bundle of *projective frames* [2, §2.2] (and the related notions of the *graph* of a projective structure [40, §2] and of *\mathfrak{sl}_2 -opers* [35, §8.2]).

3.3 Schwarzian parameterization of $\mathcal{P}(S)$

Identification of bundles. There is a complex vector bundle $\mathcal{Q}(S) \rightarrow \mathcal{T}(S)$ over Teichmüller space whose total space consists of pairs (X, ϕ) , where $X \in \mathcal{T}(S)$ and $\phi \in Q(X)$. In Teichmüller theory, this bundle is identified with the holomorphic cotangent bundle of Teichmüller space (see e.g. [56], [52]). Since Teichmüller space is diffeomorphic to \mathbb{R}^{6g-6} , the bundle $\mathcal{Q}(S)$ is diffeomorphic to \mathbb{R}^{12g-12} .

Using a section $\sigma : \mathcal{T}(S) \rightarrow \mathcal{P}(S)$ to provide basepoints for the fibers, we can form a bijective *Schwarzian parameterization*

$$\begin{aligned} \mathcal{P}(S) &\rightarrow \mathcal{Q}(S), \\ Z &\mapsto (\pi(Z), Z - \sigma(\pi(Z))) \end{aligned}$$

which is compatible with the maps of these spaces to $\mathcal{T}(S)$. This correspondence identifies the zero section of $\mathcal{Q}(S)$ with the section σ of $\mathcal{P}(S)$. A different section σ will result in a parameterization that differs by a translation in each fiber.

Compatibility. The topology on $\mathcal{P}(S)$ defined using development-holonomy pairs is compatible with the topology of $\mathcal{Q}(S)$, in that the bijection induced by any continuous section $\sigma: \mathcal{T}(S) \rightarrow \mathcal{P}(S)$ is a homeomorphism. Continuity in one direction is elementary complex analysis – uniformly close holomorphic developing maps have uniformly close derivatives (on a smaller compact set), and therefore uniformly close Schwarzian derivatives, making $\mathcal{P}(S) \rightarrow \mathcal{Q}(S)$ continuous. On the other hand, continuity of $\mathcal{Q}(S) \rightarrow \mathcal{P}(S)$ follows from continuous dependence of solutions to the ODE (3.1) on its parameter ϕ .

Holomorphic structure. The bundle $\mathcal{Q}(S)$ is a complex manifold, and a holomorphic vector bundle over $\mathcal{T}(S)$. The Schwarzian parameterization given by a section $\sigma: \mathcal{T}(S) \rightarrow \mathcal{P}(S)$ transports these structures to $\mathcal{P}(S)$. However, two sections σ_1 and σ_2 induce the same complex structure on $\mathcal{P}(S)$ if and only if $(\sigma_1 - \sigma_2)$ is a holomorphic section of $\mathcal{Q}(S)$.

There is also a natural complex structure on $\mathcal{P}(S)$ that is defined without reference to its parameterization by $\mathcal{Q}(S)$: The tangent space $T_Z\mathcal{P}(S)$ can be identified with the cohomology group $H^1(Z, \mathcal{V}_{\text{proj}})$, where $\mathcal{V}_{\text{proj}}$ is the sheaf of projective vector fields over Z , i.e. vector fields that in a local projective coordinate are restrictions of infinitesimal Möbius transformations. This cohomology group is a complex vector space, which gives an integrable almost complex structure $J: T_Z\mathcal{P}(S) \rightarrow T_Z\mathcal{P}(S)$. (Compare the construction of [51, Propositions 1, 2].)

Quasi-Fuchsian sections. Using deformations of Kleinian surface groups, we can construct a class of sections of $\mathcal{P}(S)$ that transport the complex structure of $\mathcal{Q}(S)$ to the natural complex structure on $\mathcal{P}(S)$. Given $X, Y \in \mathcal{T}(S)$, let $Q(X, Y)$ denote the quasi-Fuchsian group (equipped with an isomorphism $\pi_1(S) \simeq Q(X, Y)$) that simultaneously uniformizes X and Y (see e.g. [56, Chapter 6]). This means that $Q(X, Y)$ has domain of discontinuity $\Omega_+ \sqcup \Omega_-$ with marked quotient Riemann surfaces

$$\Omega_+/Q(X, Y) \simeq X, \quad \Omega_-/Q(X, Y) \simeq \bar{Y}$$

where \bar{Y} is the complex conjugate Riemann surface of Y , which appears in the quotient because the induced orientation on the marked surface $\Omega_-/Q(X, Y)$ is opposite that of S .

As a quotient of a domain by a Kleinian group, the surface $\Omega_+/Q(X, Y)$ also has a natural projective structure, which we denote by $\Sigma_Y(X)$. By definition, the underlying Riemann surface of $\Sigma_Y(X)$ is X , so for any fixed $Y \in \mathcal{T}(S)$ this defines a *quasi-Fuchsian section*

$$\Sigma_Y: \mathcal{T}(S) \rightarrow \mathcal{P}(S).$$

These quasi-Fuchsian sections induce the natural complex structure on $\mathcal{P}(S)$. We sketch two ways to see this: First, Hubbard uses a cohomology computation to show that a section induces the canonical complex structure if and only if it can be represented by a *relative projective structure* on the universal curve over $\mathcal{T}(S)$ [51, Propositions 1, 2]. The quasi-Fuchsian groups provide such a structure due to the analytic dependence

of the solution of the Beltrami equation on its parameters [1], and the associated construction of the *Bers fiber space* [6].

Alternatively, one can show (as in the respective computations of Hubbard [51] and Earle [28]) that both the canonical complex structure on $\mathcal{P}(S)$ and the complex structure coming from a quasi-Fuchsian section make the holonomy map (discussed in §5) a local biholomorphism, and therefore they are holomorphically equivalent.

Norms. A norm on the vector space $Q(X)$ induces a natural measure of the “complexity” of a projective structure on X (relative to the standard Fuchsian structure), or of the difference between two projective structures. There are several natural choices for such a norm.

The hyperbolic L^∞ norm $\|\phi\|_\infty$ is the supremum of the function $|\phi|/\rho^2$, where ρ^2 is the area element of the hyperbolic metric on X . Lifting ϕ to the universal cover and identifying $\tilde{X} \rightarrow \Delta$, we have

$$\|\phi\|_\infty = \|\tilde{\phi}\|_\infty = \frac{1}{4} \sup_{z \in \Delta} |\tilde{\phi}(z)|(1 - |z|^2)^2.$$

As a result of Nehari’s theorem on univalent functions, a holomorphic immersion $f: \Delta \rightarrow \mathbb{C}\mathbb{P}^1$ satisfying $\|S(f)\|_\infty \leq \frac{1}{2}$ is injective, while any injective map satisfies $\|S(f)\|_\infty \leq \frac{3}{2}$ (see [91], also [96], [79]). More generally, the norm $\|S(f)\|_\infty$ gives a coarse estimate of the size of hyperbolic balls in Δ on which f is univalent [74, §3], [77, Lemma 5.1]. Thus, when applied to projective structures, the L^∞ norm reflects the geometry and valence of the developing map.

In Teichmüller theory, it is more common to use the L^1 norm $\|\phi\|_1$, which is the area of the surface X with respect to the singular Euclidean metric $|\phi|$. This norm is conformally natural, since it does not depend on the choice of a Riemannian metric on X . However, the intrinsic meaning of the L^1 norm of the Schwarzian derivative is less clear.

More generally, given any background Riemannian metric on X compatible with its conformal structure, there is an associated L^p norm on $Q(X)$. These norms, with $p \in (1, \infty)$ and especially $p = 2$, can be used to apply PDE estimates to the study of projective structures, as discussed in §6.4 below.

Note that while any two norms on the finite-dimensional vector space $Q(X)$ are bi-Lipschitz equivalent, the bi-Lipschitz constants between the L^∞ , L^1 , and hyperbolic L^p norms on $Q(X)$ diverge as $X \rightarrow \infty$ in Teichmüller space.

4 The Grafting parameterization

4.1 Definition of grafting

Grafting is a geometric operation that can be used to build an arbitrary projective structure by gluing together simple pieces. We start by defining grafting in a restricted setting, and then work toward the general definition.

Grafting simple geodesics. Equip a Riemann surface $X \in \mathcal{T}(S)$ with its hyperbolic metric. Let γ be a simple closed hyperbolic geodesic on X . The basic grafting construction replaces γ with the cylinder $\gamma \times [0, t]$ to obtain a new surface $\text{gr}_{t\gamma} X$, the *grafting of X by $t\gamma$* , as shown in Figure 1. The natural metric on this surface is partially hyperbolic (on $X - \gamma$) and partially Euclidean (on the cylinder), and underlying this metric is a well-defined conformal structure on $\text{gr}_{t\gamma} X$. Let \mathcal{S} denote the set of free

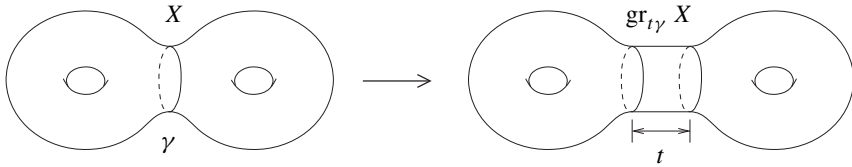


Figure 1. Grafting along a simple closed curve.

homotopy classes of homotopically nontrivial simple closed curves on S . Then \mathcal{S} is canonically identified with the set of simple closed geodesics for any hyperbolic structure on S , and we can regard grafting as a map

$$\text{gr}: \mathcal{S} \times \mathbb{R}^+ \times \mathcal{T}(S) \rightarrow \mathcal{T}(S).$$

When it is important to distinguish this construction from the projective version defined below, we will call this *conformal grafting*, since the result is a conformal structure.

Projective grafting. The Riemann surface X has a standard Fuchsian projective structure in which the holonomy of a simple closed geodesic γ is conjugate to $z \mapsto e^\ell z$, where $\ell = \ell(\gamma, X)$ is the hyperbolic length of γ .

For any $t < 2\pi$, let \tilde{A}_t denote a sector of angle t in the complex plane, with its vertex at 0. The quotient $A_t = \tilde{A}_t / \langle z \mapsto e^\ell z \rangle$ is an annulus equipped with a projective structure, which as a Riemann surface is isomorphic to the Euclidean product $\gamma \times [0, t]$.

There is a natural projective structure on the grafted surface $\text{gr}_{t\gamma} X$ that is obtained by gluing the standard Fuchsian projective structure of X to A_t ; these structures are compatible due to the matching holonomy around the gluing curves. In the universal cover of X , this corresponds to inserting a copy of \tilde{A}_t in place of each lift of γ (see Figure 2), applying Möbius transformations to \tilde{A}_t and the complementary regions of γ in \tilde{X} (which are bounded by circular arcs) so that they fit together. For sufficiently small t , this produces a Jordan domain in $\mathbb{C}\mathbb{P}^1$ that is the image of the developing

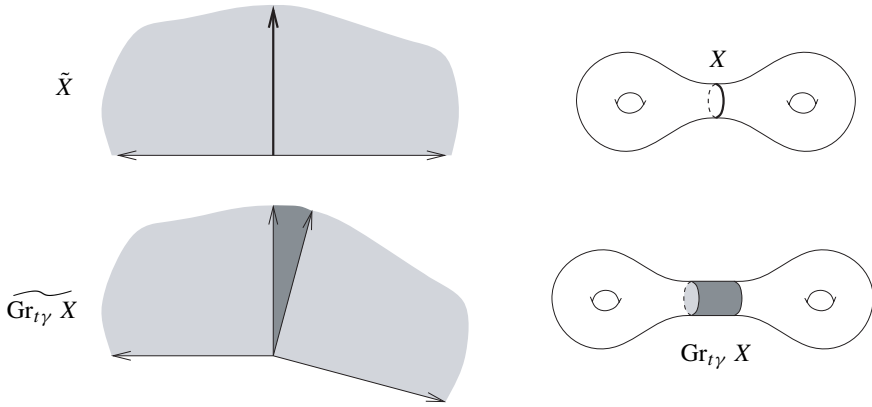


Figure 2. Projective grafting: Gluing a cylinder into the surface along a geodesic corresponds to inserting a sector or lune into each lift of the geodesic. Only one lift is shown here, but the gluing construction is repeated equivariantly in $\widetilde{\text{Gr}}_{t\gamma} X$.

map, while for large t the developing image is all of $\mathbb{C}P^1$. We denote the resulting projective structure by $\text{Gr}_{t\gamma} X$.

Applying a generic Möbius transformation to the sector \tilde{A}_t will map it to a t -lune, the intersection of two round disks with interior angle t . Thus the projective structure $\text{Gr}_{t\gamma} X$ corresponds to a decomposition of its universal cover into t -lunes and regions bounded by circular arcs.

The restriction to small values of t in this construction is not necessary; for $t > 2\pi$ we simply interpret \tilde{A}_t as a “sector” that wraps around the punctured plane \mathbb{C}^* some number of times. Alternatively, we could define A_t for $t \geq 2\pi$ by gluing n copies of $A_{t/n}$ end-to-end, for a sufficiently large $n \in \mathbb{N}$.

Therefore we have a *projective grafting map*,

$$\text{Gr}: \mathcal{S} \times \mathbb{R}^+ \times \mathcal{T}(S) \rightarrow \mathcal{P}(S)$$

which is a lift of grafting through the forgetful map $\pi: \mathcal{P}(S) \rightarrow \mathcal{T}(S)$, i.e. $\pi \circ \text{Gr} = \text{gr}$.

Variations on simple grafting. Grafting along a simple geodesic with weight $t = 2\pi$ was originally used by Maskit [83], Hejhal [47], and Sullivan–Thurston [109] to construct examples of exotic Fuchsian projective structures (discussed in §5.4 below). Grafting with weight 2π is special because it does not change the holonomy representation of the Fuchsian projective structure (see §5).

It is possible to extend this holonomy-preserving grafting operation to certain simple curves which are not geodesic, and to projective structures that are not standard Fuchsian (see [66, Chapter 7]); this generalization has been important to some applications in Kleinian groups and hyperbolic geometry (e.g. [13], [11, §5]), and it will appear again in our description of quasi-Fuchsian projective structures (§5.5). How-

ever, our main focus in this chapter is a different extension of grafting, defined by Thurston, which leads to a geometric model for the entire moduli space $\mathcal{P}(S)$.

Extension to laminations. Projective grafting is compatible with the natural completion of $\mathbb{R}^+ \times \mathcal{S}$ to the space $\mathcal{ML}(S)$ of *measured laminations*. An element $\lambda \in \mathcal{ML}(S)$ is realized on a hyperbolic surface $X \in \mathcal{T}(S)$ as a foliation of a closed subset of X by complete, simple hyperbolic geodesics (some of which may be closed), equipped with a transverse measure of full support. A piecewise linear coordinate atlas for $\mathcal{ML}(S)$ is obtained by integrating transverse measures over closed curves, making $\mathcal{ML}(S)$ into a PL manifold homeomorphic to \mathbb{R}^{6g-6} . See [113, Chapter 8–9], [18], [94, Chapter 3] for detailed discussion of measured laminations.

There is continuous extension $\text{Gr}: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{P}(S)$ of projective grafting, which is uniquely determined by the simple grafting construction because weighted simple closed curves are dense in $\mathcal{ML}(S)$. Similarly, there is an extension of the grafting map $\text{gr}: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{T}(S)$ defined by $\text{gr} = \pi \circ \text{Gr}$. These extensions were defined by Thurston (unpublished), and are discussed in detail in [64].

For a lamination $\lambda \in \mathcal{ML}(S)$ that is supported on a finite set of disjoint simple closed curves, i.e. $\lambda = \sum_{i=1}^n t_i \gamma_i$, the grafting $\text{gr}_\lambda X$ defined by this extension procedure agrees with the obvious generalization of grafting along simple closed curves, wherein the geodesics $\gamma_1, \dots, \gamma_n$ are simultaneously replaced with cylinders.

For a general measured lamination $\lambda \in \mathcal{ML}(S)$, one can think of $\text{gr}_\lambda X$ as a Riemann surface obtained from X by thickening the leaves of the lamination λ in a manner dictated by the transverse measure. This intuition is made precise by the definition of a canonical stratification of $\text{gr}_\lambda X$ in the next section.

4.2 Thurston's theorem

Projective grafting is a universal construction – every projective structure can be obtained from it, and in exactly one way:

Theorem 4.1 (Thurston [unpublished]). *The projective grafting map*

$$\text{Gr}: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{P}(S)$$

is a homeomorphism.

The proof of Theorem 4.1 proceeds by explicitly constructing the inverse map Gr^{-1} using complex projective and hyperbolic geometry. We will now sketch this construction; details can be found in [64].

The embedded case. First suppose that $Z \in \mathcal{P}(S)$ is a projective surface whose developing map is an embedding (an *embedded projective structure*). The image of the developing map is a domain $\Omega \subset \mathbb{CP}^1$ invariant under the action of $\pi_1(S)$ by the

holonomy representation ρ . In this case, we will describe the inverse of projective grafting in terms of convex hulls in hyperbolic space. See [32] for details on these hyperbolic constructions.

Considering \mathbb{CP}^1 as the ideal boundary of hyperbolic space \mathbb{H}^3 , let $\text{Pl}(Z)$ denote the boundary of the hyperbolic convex hull of $(\mathbb{CP}^1 - \Omega)$. Then $\text{Pl}(Z)$ is a *convex pleated plane* in \mathbb{H}^3 invariant under the action of $\pi_1(S)$ by isometries.

When equipped with the path metric, the pleated plane $\text{Pl}(Z)$ is isometric to \mathbb{H}^2 , and by this isometry, the action of $\pi_1(S)$ on \mathbb{H}^3 corresponds to a discontinuous cocompact action on \mathbb{H}^2 . Let $Y \in \mathcal{T}(S)$ denote the marked quotient surface.

The pleated plane $\text{Pl}(Z)$ consists of totally geodesic pieces (*plaques* or *facets*) meeting along geodesic *bending lines*. Applying the isometry $\text{Pl}(Z) \simeq \mathbb{H}^2$ to the union of the bending lines yields a geodesic lamination, which has a natural transverse measure recording the amount of bending of $\text{Pl}(Z)$. The lamination and measure are $\pi_1(S)$ -invariant, and therefore descend to the quotient, defining an element $\lambda \in \mathcal{ML}(S)$.

Thus, starting from an embedded projective structure Z , we obtain a hyperbolic structure Y and a measured lamination λ . To show that we have inverted the projective grafting map, we must check that $\text{Gr}_\lambda Y = Z$.

Nearest-point projection. There is a *nearest-point projection map* $\kappa: \Omega \rightarrow \text{Pl}(Z)$ that sends $z \in \Omega$ to the first point on $\text{Pl}(Z)$ that is touched by an expanding family of horoballs in \mathbb{H}^3 based at z . Convexity of $\text{Pl}(Z)$ ensures that this point is well-defined. In fact, from each $z \in \Omega$ we obtain not just a nearest point on $\text{Pl}(Z)$, but also a *support plane* H_z which contains $\kappa(z)$ and whose normal vector at that point defines a geodesic ray with ideal endpoint z . This gives a map $\hat{\kappa}: \Omega \rightarrow \mathcal{H}^{2,1}$, where $\mathcal{H}^{2,1}$ is the space of planes in \mathbb{H}^3 (the *de Sitter space*).

The *canonical stratification* of Ω is the decomposition into fibers of the map $\hat{\kappa}$. Strata are of two types:

- 1-dimensional strata – circular arcs that map homeomorphically by κ onto bending lines of $\text{Pl}(Z)$, and
- 2-dimensional strata – regions with nonempty interior bounded by circular arcs which map homeomorphically by κ to the totally geodesic pieces of $\text{Pl}(Z)$.

If λ is supported on a single closed geodesic (or on a finite union of them), the 1-dimensional strata and the boundary geodesics of the 2-dimensional strata in $\Omega \simeq \tilde{Z}$ fill out a collection of lunes, and the interiors of the 2-dimensional strata correspond by κ to the complementary regions of the lift of λ , realized geodesically on Y , to $\tilde{Y} \simeq \mathbb{H}^2$. See Figure 3 for an example of this type. This is the arrangement of lunes and circular polygons giving the projective structure of $\text{Gr}_\lambda Y$, and so $Z = \text{Gr}_\lambda Y$. A limiting argument shows the same holds for general λ .

The general case. The key to inverting the projective grafting in the embedded case is the construction of the convex pleated plane $\text{Pl}(Z)$. For general $Z \in \mathcal{P}(S)$, this is replaced by a *locally convex pleated plane* defined using the projective geometry of Z itself, rather than its developed image.

Let $f: \tilde{Z} \rightarrow \mathbb{CP}^1$ be the developing map of Z . A *round disk* in \tilde{Z} is an open subset U such that f is injective on U and $f(U)$ is an open disk in \mathbb{CP}^1 . The round disks in \tilde{Z} are partially ordered with respect to inclusion. A maximal element for this ordering is a *maximal round disk*.

Each maximal round disk U in \tilde{Z} corresponds to a disk in \mathbb{CP}^1 , and thus to an oriented plane H_U in \mathbb{H}^3 . Allowing U to vary over all maximal round disks in \tilde{Z} gives a family of oriented planes, and the envelope of this family is a locally convex pleated plane $\text{Pl}(Z)$.

The rest of the convex hull construction generalizes as follows: The intrinsic geometry of $\text{Pl}(Z)$ is hyperbolic, with quotient Y , and the bending of $\text{Pl}(Z)$ is recorded by a measured lamination λ . In place of the nearest-point projection and support planes, we have a *collapsing map* $\kappa: \tilde{Z} \rightarrow \text{Pl}(Z)$ and a *co-collapsing map* $\hat{\kappa}: \tilde{Z} \rightarrow \mathcal{H}^{2,1}$ (see also [25, §2, §7]). The fibers of $\hat{\kappa}$ induce a canonical stratification of \tilde{Z} , and separating the 1- and 2-dimensional strata describes Z as the projective grafting $\text{Gr}_\lambda Y$.

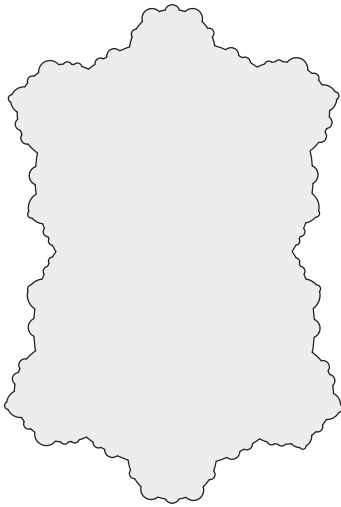
Note that the canonical stratification of \tilde{Z} is $\pi_1(S)$ -invariant, and therefore we have a corresponding decomposition of Z into 1- and 2-dimensional pieces. We will also refer to this as the canonical stratification. Similarly, the collapsing map descends to a map $\kappa: Z \rightarrow Y$ between quotient surfaces, which sends the union of 1-dimensional strata and boundary geodesics of 2-dimensional strata onto the bending lamination $\lambda \subset Y$.

The canonical stratification for complex projective structures is discussed further in [64, §1.2], where it is also generalized to n -manifolds equipped with *flat conformal structure* (see also [78], [102]).

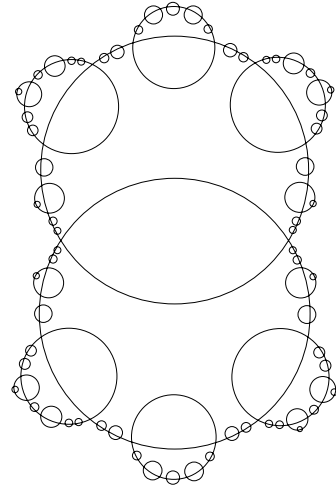
Dual trees. When grafting along a simple closed curve γ with weight t , each bending line of the associated pleated plane in \mathbb{H}^3 has a one-parameter family of support planes (see Figure 4). These give an interval in the image of $\hat{\kappa}$, and the angle between support planes gives a metric on this interval, making it isometric to $[0, t] \subset \mathbb{R}$. Alternatively, this metric could be defined as the restriction of the Lorentzian metric of $\mathcal{H}^{2,1}$, where the restriction is positive definite because any pair of support planes of a given bending line intersect (see [102, §5], [78, §3, §6.5]).

The intervals corresponding to different bending lines meet at vertices corresponding to support planes of flat pieces. This gives $\hat{\kappa}$ the structure of a metric tree, the *dual tree* of the weighted curve $t\gamma$, denoted $T_{t\gamma}$. As this notation suggests, this tree depends only on $t\gamma$ (through the bending lines, their bending angles, and the adjacency relationship between bending lines and flat pieces) and not on the quotient hyperbolic structure of the pleated plane. The equivariance of the pleated plane with respect to $\pi_1(S)$ determines an isometric action of $\pi_1(S)$ on $T_{t\gamma}$.

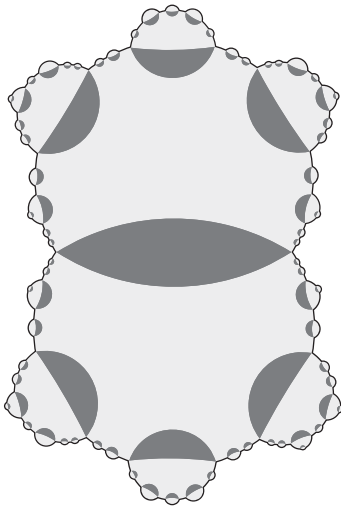
For a general grafting lamination $\lambda \in \mathcal{ML}(S)$, the image of $\hat{\kappa}$ has the structure of an \mathbb{R} -tree (see [101, Chapter 9], [78, §6, §11]), a geodesic metric space in which each pair of points is joined by a unique geodesic which is isometric to an interval in \mathbb{R} [90, Chapter 2]. This *dual \mathbb{R} -tree of λ* , denoted T_λ , is also equipped with an isometric action of $\pi_1(S)$.



(a) Developed image



(b) Maximal disks



(c) Lunes and ideal polygons

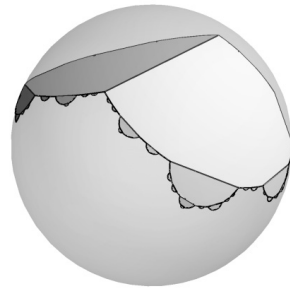
(d) Pleated plane in \mathbb{H}^3 (Klein model)

Figure 3. Four views of a projective structure lifted to the universal cover of a surface. The example shown here is an approximation of an embedded structure on a surface of genus 2, obtained by grafting along a separating simple closed curve. The approximation includes only a few of the maximal disks.

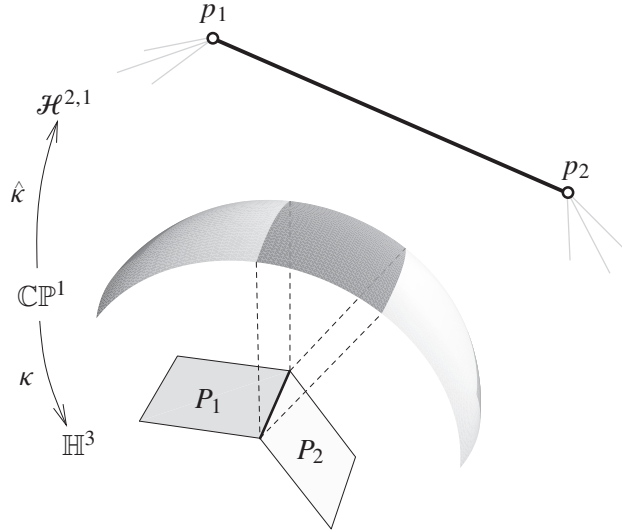


Figure 4. A lune between two maximal disks collapses to a bending line between two planes (P_1, P_2) , and co-collapses to an interval between two points (p_1, p_2) .

4.3 The Thurston metric

We have seen that when grafting along a simple closed curve, the resulting projective surface $\text{Gr}_{T,Y} X$ has a natural conformal metric that combines the hyperbolic structure of X and the Euclidean structure of the cylinder. This is the *Thurston metric* (or *projective metric*) on the projective surface.

This definition can be extended to arbitrary projective surfaces by taking limits of the metrics gained from an appropriate sequence of simple closed curves; however, we will prefer an intrinsic description of the metric based on complex projective geometry.

Kobayashi construction. The Kobayashi metric on a complex manifold is defined by a norm on each tangent space, where the length of a vector v is the infimum of lengths given to it by holomorphically immersed disks (each of which is equipped with its hyperbolic metric). For a surface Z with a projective structure, there is a variant of the Kobayashi metric in which one minimizes length over the smaller class of *projectively immersed disks*, that is, immersions $\Delta \rightarrow \tilde{Z}$ that are locally Möbius with respect to the projective structure on Δ as a subset of $\mathbb{C}\mathbb{P}^1$. The resulting “projective Kobayashi metric” is the *Thurston metric* of Z [110, §2.1].

Relation to grafting. This intrinsic definition of the Thurston metric is related to grafting as follows: for each $z \in \tilde{Z}$, there is a unique maximal round disk $U \subset \tilde{Z}$ such that the (lifted) Thurston metric at z agrees with the hyperbolic metric on U .

Furthermore, the set of points in \tilde{Z} that correspond to a given maximal disk U is a stratum in the canonical stratification of Z . Thus the Thurston metric is built from 2-dimensional hyperbolic regions with geodesic boundary and 1-dimensional geodesic strata. For $Z = \text{Gr}_\lambda X$, the union of the hyperbolic strata covers a subset of Z isometric to $(X - \lambda)$, where λ is realized geodesically on the hyperbolic surface X . When $\lambda = t\gamma$ is supported on a simple closed curve, the 1-dimensional strata sweep out Euclidean strips in $\widetilde{\text{Gr}_{t\gamma} X}$, which cover a Euclidean cylinder in $\text{Gr}_{t\gamma} X$, recovering the synthetic description of the Thurston metric in this case.

Conformal metrics and regularity. The Thurston metric on Z is a nondegenerate Riemannian metric compatible with the underlying complex structure $\pi(Z)$, i.e. it is a *conformal metric* on the Riemann surface. In local complex coordinates, the line element of such a metric has the form $\rho(z)|dz|$, where $\rho(z)$ is the real-valued *density function*. In the case of simple grafting, the density function of the Thurston metric is smooth on the hyperbolic and Euclidean pieces, but it is only C^1 on the interface between them. (The discontinuity in its second derivative is necessary since the curvature changes along the interface.) In general, the Thurston metric of a projective surface is $C^{1,1}$, meaning that its density function has Lipschitz derivatives, with Lipschitz constant locally bounded on $\mathcal{P}(S)$ [78].

Variation of metrics. The Thurston metric is a continuous function of the projective structure $Z \in \mathcal{P}(S)$ with respect to the topology of locally uniform convergence of density functions: For a sequence $Z_n \rightarrow Z \in \mathcal{P}(S)$, the Lipschitz bound on the derivatives of the Thurston metrics shows that uniform convergence follows from pointwise convergence, which in turn follows from the locally uniform convergence of the developing maps $f_n: \Delta \rightarrow \mathbb{CP}^1$ (or from the continuous variation of the associated locally convex pleated surfaces).

Area. A conformal metric on a Riemann surface with density function ρ induces an area measure by integration of $\rho^2 = \rho(z)^2|dz|^2$.

The total area of $\text{Gr}_\lambda X$ with respect to the Thurston metric is $4\pi(g-1) + \ell(\lambda, X)$, where $\ell(\lambda, X)$ is the length of the measured lamination λ with respect to the hyperbolic metric of X . The two terms correspond to the two types of strata: The union of the 2-dimensional strata has area $4\pi(g-1)$, because it is isometric to the complement of a geodesic lamination (a null set) in the hyperbolic surface X . The union of the 1-dimensional strata has area $\ell(\lambda, X)$, which is the continuous extension to $\mathcal{ML}(S)$ of the function $t\ell(\gamma, X)$ giving the area of the Euclidean cylinder $\gamma \times [0, t]$ in the case of simple grafting.

Curvature. The Gaussian curvature K and curvature 2-form Ω of a smooth conformal metric are related to its density function ρ by

$$\begin{aligned} K &= -\frac{1}{\rho^2} \Delta \log \rho, \\ \Omega &= K \rho^2 = -\Delta \log \rho. \end{aligned} \tag{4.1}$$

In particular, such a metric has nonpositive Gaussian curvature if and only if $\log \rho$ is a subharmonic function.

The Thurston metric is not smooth everywhere, but it is nonpositively curved (NPC), meaning that its geodesic triangles are thinner than triangles in Euclidean space with the same edge lengths. As in the smooth case, this implies that $\log \rho$ is subharmonic, so we have a nonpositive measure $\Omega = -\Delta \log(\rho)$ that generalizes the curvature 2-form [99] (see also [98], [55], [54], [87]). For the Thurston metric, Ω is absolutely continuous, $\Omega = K \rho^2$ where K is the (a.e. defined) Gaussian curvature function. By a generalization of the Gauss–Bonnet theorem, the total mass of Ω (which is the integral of K) is $-4\pi(g - 1)$ [54].

Hyperbolic and Euclidean. Since the Gaussian curvature of the Thurston metric is -1 in the interior of each 2-dimensional stratum, and these have total area $4\pi(g - 1)$, the curvature of the Thurston metric is almost everywhere 0 in the union of the 1-dimensional strata. In this sense, grafting along a general lamination can be seen as the operation of inserting a Euclidean “surface” in place of a geodesic lamination, generalizing the case of closed leaves.

4.4 Conformal grafting maps

Having discussed the projective grafting construction and its inverse, we turn our attention to properties of the conformal grafting map $\text{gr}: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{T}(S)$.

Using techniques from the theory of harmonic maps between surfaces (see §6.3), Tanigawa showed that this map is proper when either one of the coordinates is fixed:

Theorem 4.2 (Tanigawa [110]). *For each $\lambda \in \mathcal{ML}(S)$, the λ -grafting map*

$$\text{gr}_\lambda: \mathcal{T}(S) \rightarrow \mathcal{T}(S)$$

is a proper smooth map. For each $X \in \mathcal{T}(S)$, the X -grafting map

$$\text{gr}_\bullet X: \mathcal{ML}(S) \rightarrow \mathcal{T}(S)$$

is a proper continuous map.

Properness allows global properties of these maps to be derived from local considerations. For example, Scannell and Wolf showed that the λ -grafting map is an immersion, and therefore it is a local diffeomorphism. Since a proper local diffeomorphism is a covering map, this result and Theorem 4.2 give:

Theorem 4.3 (Scannell–Wolf [103]). *For each $\lambda \in \mathcal{ML}(S)$, the λ -grafting map $\text{gr}_\lambda: \mathcal{T}(S) \rightarrow \mathcal{T}(S)$ is a diffeomorphism.*

Earlier, Tanigawa had shown that gr_λ is a diffeomorphism when $\lambda \in \mathcal{ML}(S)$ is supported on a finite set of simple closed curves with weights that are integral multiples of 2π [110]. This follows from Theorem 4.2 because holonomy considerations (see §5) imply that gr_λ is a local diffeomorphism in this case.

In the general case, Scannell and Wolf analyze the Thurston metric and conformal grafting map through the interaction of two differential equations: The Liouville equation, which relates a Riemannian metric to its curvature, and the Jacobi equation, which determines the variation of a geodesic with respect to a family of Riemannian metrics. Analytic estimates for these equations are used to show that a 1-parameter family of graftings $t \mapsto \text{gr}_\lambda X_t$ cannot be conformally equivalent to first order unless $(d/dt)X_t|_{t=0} = 0$, which gives injectivity of the derivative of gr_λ .

As a consequence of Theorem 4.3, for any $\lambda \in \mathcal{ML}(S)$ the set of projective structures with grafting lamination λ projects homeomorphically to $\mathcal{T}(S)$ by the forgetful map. That is, the set of such projective structures forms a smooth section $\sigma_\lambda: \mathcal{T}(S) \rightarrow \mathcal{P}(S)$ of π , which is given by

$$\sigma_\lambda(X) = \text{Gr}_\lambda(\text{gr}_\lambda^{-1}(X)). \tag{4.2}$$

Note that this is compatible with our previous definition of the standard Fuchsian structure $\sigma_0(X)$, since this is the unique projective structure on X with zero grafting lamination. As with Theorem 4.3, in the special case of 2π -integral weighted multicurves, the existence of these smooth sections follows from the earlier work of Tanigawa.

Fixing X and varying $\lambda \in \mathcal{ML}(S)$, we can also use Theorem 4.3 to parameterize the fiber $P(X)$; that is,

$$\lambda \mapsto \sigma_\lambda(X)$$

gives a homeomorphism $\mathcal{ML}(S) \rightarrow P(X)$ (compare [26, §4]). It is the inverse of the map which sends $\text{Gr}_\lambda Y \in P(X)$ to λ .

Building on the Scannell–Wolf result, the author and Wolf showed that the X -grafting map is also a local homeomorphism, leading to:

Theorem 4.4 (Dumas and Wolf [27]). *For each $X \in \mathcal{T}(S)$, the X -grafting map $\text{gr}_\bullet X: \mathcal{ML}(S) \rightarrow \mathcal{T}(S)$ is a homeomorphism. Furthermore, this homeomorphism is bitangentiable.*

The last claim in this theorem involves the regularity of the grafting map as λ is varied. Let $f: U \rightarrow V$ be a continuous map, where $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open sets. The *tangent map* of f at x , denoted $T_x f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, is defined by

$$T_x f(v) = \lim_{\epsilon \rightarrow 0^+} \frac{f(x + \epsilon v) - f(x)}{\epsilon}.$$

The map f is *tangentiable* if this limit exists for all $(x, v) \in U \times \mathbb{R}^n$, and if the convergence is locally uniform in v when x is fixed. Intuitively, a tangentiable map is one which has one-sided derivatives everywhere. These notions generalize naturally to maps between smooth manifolds (e.g. $\mathcal{T}(S)$), piecewise linear manifolds (e.g. $\mathcal{ML}(S)$), or manifolds defined by an atlas of charts with tangentiable transition functions. Tangentiable maps and manifolds are discussed in [10].

A homeomorphism f is called *bitangentiable* if both f and f^{-1} are tangentiable, and if every tangent map of f or f^{-1} is a homeomorphism. Thus a bitangentiable homeomorphism is the analogue of a diffeomorphism in the tangentiable category.

The connection between grafting, projective structures, and tangentiability was studied by Bonahon, following work of Thurston on the infinitesimal structure of the space $\mathcal{ML}(S)$ [115]; the fundamental result, which strengthens Thurston's theorem, is

Theorem 4.5 (Bonahon [10]). *The projective grafting map $\text{Gr}: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{P}(S)$ is a bitangentiable homeomorphism.*

The proof of Theorem 4.4 uses Theorem 4.3, the above result of Bonahon, and a further complex linearity property of the tangent map of projective and conformal grafting (see [10], [9, §10], also [27, §3]). This complex linearity provides a “duality” between variation of $\text{gr}_\lambda X$ under changes in X and λ ; in a certain sense, grafting behaves like a holomorphic function, where X and λ are the real and imaginary parts of its parameter, respectively. This allows infinitesimal injectivity of $\text{gr}_\bullet X_0$ near λ_0 to be derived from the infinitesimal injectivity of gr_{λ_0} near X_0 .

After applying some additional tangentiable calculus, this infinitesimal injectivity is converted to local injectivity of $\text{gr}_\bullet X$, from which Theorem 4.4 follows by properness (Theorem 4.2).

5 Holonomy

We now turn our attention to the holonomy representations of projective structures in relation to the grafting and Schwarzian coordinate systems for $\mathcal{P}(S)$. General references for these matters include [47], [46], [28], [51], [36].

5.1 Representations and characters

Let $\mathcal{R}(S) = \text{Hom}(\pi_1(S), \text{PSL}_2(\mathbb{C}))$ denote the set of homomorphisms (representations) from $\pi_1(S)$ to $\text{PSL}_2(\mathbb{C})$, which is an affine \mathbb{C} -algebraic variety (as a subset of $(\text{PSL}_2(\mathbb{C}))^N \simeq (\text{SO}_3(\mathbb{C}))^N$). The group $\text{PSL}_2(\mathbb{C})$ acts algebraically on $\mathcal{R}(S)$ by conjugation, and there is a quotient *character variety*

$$\mathcal{X}(S) = \mathcal{R}(S) // \text{PSL}_2(\mathbb{C})$$

in the sense of geometric invariant theory. Concretely, the points of $\mathcal{X}(S)$ are in one-to-one correspondence with the set of *characters*, i.e. \mathbb{C} -valued functions on $\pi_1(S)$ of

the form

$$\gamma \mapsto \operatorname{tr}^2(\rho(\gamma))$$

where $\rho \in \mathcal{R}(S)$. Mapping a character to its values on an appropriate finite subset of $\pi_1(S)$ gives an embedding of $\mathcal{X}(S)$ as an affine variety in \mathbb{C}^n . See [49] for a discussion of $\operatorname{PSL}_2(\mathbb{C})$ character varieties, building on the work of Culler–Shalen in the $\operatorname{SL}_2(\mathbb{C})$ case [20]. Algebraic and topological properties of character varieties are also studied in [44, §9], [41], [97].

Liftability. The variety $\mathcal{X}(S)$ splits into two irreducible components according to whether or not the associated representations lift from $\operatorname{PSL}_2(\mathbb{C})$ to $\operatorname{SL}_2(\mathbb{C})$ (see [41], [97]). Denote these by $\mathcal{X}_0(S)$ and $\mathcal{X}_1(S)$, where the former consists of liftable characters. Each of these components has complex dimension $6g - 6$, which agrees with the “expected dimension”, i.e. $6g - 6 = (\dim \operatorname{PSL}_2(\mathbb{C}))(N_{\text{gens}} - N_{\text{relators}} - 1)$.

Elementary and non-elementary. When working with the character variety, complications may arise because the invariant-theoretic quotient $\mathcal{X}(S)$ is singular, or because it is not the same as the quotient set $\mathcal{R}(S)/\operatorname{PSL}_2(\mathbb{C})$. However we can avoid most of these difficulties by restricting attention to a subset of characters (which contains those that arise from projective structures).

A representation $\rho \in \mathcal{R}(S)$ is *elementary* if its action on \mathbb{H}^3 by isometries fixes a point or an ideal point, or if it preserves an unoriented geodesic, otherwise it is *non-elementary*.

A non-elementary representation is determined up to conjugacy by its character, so there is a one-to-one correspondence between the set of conjugacy classes of non-elementary representations and the set $\mathcal{X}'(S) \subset \mathcal{X}(S)$ of characters of non-elementary representations.

The subset $\mathcal{X}'(S)$ is open and lies in the smooth locus of the character variety [44], [46], [38]. Thus $\mathcal{X}'(S)$ is a complex manifold of dimension $6g - 6$, and is the union of the open and closed subsets $\mathcal{X}'_i(S) = \mathcal{X}'(S) \cap \mathcal{X}_i(S)$, $i = 1, 2$.

Fuchsian and quasi-Fuchsian spaces. The character variety $\mathcal{X}(S)$ contains the space $\mathcal{QF}(S)$ of conjugacy classes of quasi-Fuchsian representations of $\pi_1(S)$ as an open subset of $\mathcal{X}'_0(S)$. The parameterization of $\mathcal{QF}(S)$ by the pair of quotient conformal structures gives a holomorphic embedding

$$\mathcal{T}(S) \times \overline{\mathcal{T}(S)} \rightarrow \mathcal{X}(S),$$

where $\overline{\mathcal{T}(S)}$ represents the surface S with the opposite orientation (see [85, §4.3]). In this embedding, the diagonal $\{(X, \overline{X}) \mid X \in \mathcal{T}(S)\}$ corresponds to the set $\mathcal{F}(S)$ of Fuchsian representations, giving an identification $\mathcal{F}(S) \simeq \mathcal{T}(S)$. Note that this is *not* a holomorphic embedding of Teichmüller space into the character variety; the image is a totally real submanifold.

5.2 The holonomy map

Since the holonomy representation $\rho \in \mathcal{R}(S)$ of a projective structure Z is determined up to conjugacy, the associated character $[\rho] \in \mathcal{X}(S)$ is uniquely determined. Considering $[\rho]$ as a function of Z gives the *holonomy map*

$$\text{hol}: \mathcal{P}(S) \rightarrow \mathcal{X}(S).$$

In fact, the image of hol lies in $\mathcal{X}'_0(S)$: A lift to $\text{SL}_2(\mathbb{C})$ is given by the linear monodromy of the Schwarzian ODE (3.1). The holonomy representation is non-elementary because S does not admit an affine or spherical structure; for details, see [5, pp. 297–304], [63, Theorem 3.6], [46, §2], [43, Theorem 19, Corollary 3].

Holonomy theorem. For hyperbolic structures on compact manifolds, the holonomy representation determines the geometric structure. For projective structures on surfaces, the same is true *locally*:

Theorem 5.1 (Hejhal [47], Earle [28], Hubbard [51]). *The holonomy map*

$$\text{hol}: \mathcal{P}(S) \rightarrow \mathcal{X}(S)$$

is a local biholomorphism.

Originally, Hejhal showed that the holonomy map is a local homeomorphism using a cut-and-paste argument. Earle and Hubbard gave alternate proofs of this result, along with differential calculations showing that the map is locally biholomorphic. Recall that when considering $\mathcal{P}(S)$ as a complex manifold, we are using the complex structure induced by the quasi-Fuchsian sections.

A more general holonomy theorem for (G, X) structures is discussed in [40].

Negative results. Despite the simple local behavior described by Theorem 5.1, the global behavior of the holonomy map is quite complicated:

Theorem 5.2. (1) *The holonomy map is not injective. In fact, all of the fibers of the holonomy map are infinite.*

(2) *The holonomy map is not a covering of its image.*

The non-injectivity in (1) follows from the discussion of 2π -grafting in §5.4 below. Hejhal established (2) by showing that the path lifting property of coverings fails for the holonomy map [47]. The infinite fibers of the holonomy map arise from the existence of *admissible curves* that can be used to alter a projective structure while preserving its holonomy [36], [66, Chapter 7]; this is similar to the “constructive approach” discussed in §5.5 below.

Further pathological behavior of the holonomy map is discussed in [59, §5].

Surjectivity. Of course one would like to know which representations arise from the holonomy of projective structures. We have seen that in order to arise from a projective structure, a character must be non-elementary and liftable (i.e. $\text{hol}(\mathcal{P}(S)) \subset \mathcal{X}'_0(S)$). These necessary conditions are also sufficient:

Theorem 5.3 (Gallo, Kapovich, and Marden [36]). *Every non-elementary liftable $\text{PSL}_2(\mathbb{C})$ -representation of $\pi_1(S)$ arises from the holonomy of a projective structure on S . Equivalently, we have $\text{hol}(P(S)) = \mathcal{X}'_0(S)$.*

In the same article it is also shown that the non-elementary non-liftable representations arise from *branched* projective structures. In both cases, the developing map of a projective structure with holonomy representation ρ is constructed by gluing together simpler projective surfaces that can be analyzed directly. A key technical result that enables this construction is:

Theorem 5.4 ([36]). *Let $\rho: \pi_1(S) \rightarrow \text{PSL}_2(\mathbb{C})$ be a homomorphism with non-elementary image. Then there exists a pants decomposition of S such that the restriction of ρ to any component of the decomposition is a marked rank-2 classical Schottky group. In particular, the image of every curve in the decomposition is loxodromic.*

Projective structures on pairs of pants with loxodromic boundary holonomy are analyzed in [36, §§6–7].

Holonomy deformations. We have seen that projective structures on a Riemann surface X form an affine space modeled on $Q(X)$ (§3.2). Thus, given a non-elementary representation $\rho \in \mathcal{X}'_0(S)$, projective structures provide deformations of ρ as follows: Find $Z \in \mathcal{P}(S)$ with $\text{hol}(Z) = \rho$, which is possible by Theorem 5.3, and consider the family of holonomy representations $\{\text{hol}(Z + \phi) \mid \phi \in Q(X)\}$. This gives a holomorphic embedding of \mathbb{C}^{3g-3} into $\mathcal{X}(S)$, a family of *projective deformations* of ρ . (Compare [74], [75], where Kra refers to a projective structure on X as a *deformation* of the Fuchsian group uniformizing X .)

These deformations could be compared with the classical quasi-conformal deformation theory of Kleinian groups. Projective deformations are especially interesting because they are insensitive to the discreteness of the image of a representation, and because they apply to quasiconformally rigid Kleinian groups. On the other hand, it is difficult to describe the global behavior of a projective deformation explicitly, and there is often no canonical choice for the preimage of ρ under the holonomy map.

5.3 Holonomy and bending

The holonomy map for projective structures is related to the grafting coordinate system through the notion of *bending deformations*. We now describe these deformations,

mostly following Epstein and Marden [32]). In doing so, we are essentially re-creating the projective grafting construction of §4.1 while working entirely in hyperbolic 3-space, and starting with a Fuchsian representation rather than a hyperbolic surface.

Bending Fuchsian groups. We begin with an algebraic description of bending. A primitive element $\gamma \in \pi_1(S)$ representing a simple closed curve that separates the surface S determines a \mathbb{Z} -amalgamated free product decomposition

$$\pi_1(S) = \pi_1(S_1) *_{\langle \gamma \rangle} \pi_1(S_2)$$

where $(S - \gamma) = S_1 \sqcup S_2$. Note that the representative γ determines an orientation of the closed geodesic, and using this orientation, we make the convention that S_2 lies to the right of the curve. Given a homomorphism $\rho: \pi_1(S) \rightarrow \mathrm{PSL}_2(\mathbb{C})$ and an element $A \in \mathrm{PSL}_2(\mathbb{C})$ centralizing $\rho(\gamma)$, there is a deformed homomorphism ρ' uniquely determined by

$$\rho'(x) = \begin{cases} \rho(x) & \text{if } x \in \pi_1(S_1), \\ A\rho(x)A^{-1} & \text{if } x \in \pi_1(S_2). \end{cases} \quad (5.1)$$

Similarly, a nonseparating curve γ corresponds to a presentation of $\pi_1(S)$ as an HNN extension, and again each centralizing element gives a deformation of ρ . See [42, §3] for further discussion of this deformation procedure.

When ρ is a Fuchsian representation and A is an elliptic element having the same axis as $\rho(\gamma)$, the homomorphism ρ' is a *bending deformation* of ρ . When A rotates by angle t about the axis of $\rho(\gamma)$, clockwise with respect to the orientation, we denote the deformed representation by $\beta_{t\gamma}(\rho) = \rho'$. Up to conjugacy, this deformation depends only on the angle t and the curve γ , not on the representative in $\pi_1(S)$ or the induced orientation.

The “bending” terminology refers to the geometry of the action of $\pi_1(S)$ on \mathbb{H}^3 by $\beta_{t\gamma}(\rho)$. The Fuchsian representation ρ preserves a plane $\mathbb{H}^2 \subset \mathbb{H}^3$, whereas we will see that the bending deformation $\beta_{t\gamma}(\rho)$ preserves a locally convex pleated (or *bent*) plane.

In terms of characters, the Fuchsian representation ρ_0 is a point in $\mathcal{F}(S) \simeq \mathcal{T}(S)$ and bending defines a map

$$\beta: \mathcal{S} \times \mathbb{R}^+ \times \mathcal{T}(S) \rightarrow \mathcal{X}(S).$$

Like grafting, this map extends continuously to measured laminations [32, Theorem 3.11.5], giving

$$\beta: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{X}(S).$$

Note that while the bending path $t \mapsto \beta_{t\gamma}(X)$ is 2π -periodic, there is no apparent periodicity when bending along a general measured lamination.

Earthquakes and quakebends. The centralizer of a hyperbolic Möbius transformation $\gamma \in \mathrm{PSL}_2(\mathbb{C})$ contains all of the elliptic and hyperbolic transformations

with the same axis as γ , but in defining bending we have only considered the elliptic transformations. The deformation corresponding (by formula (5.1)) to a pure translation is known as an *earthquake*, and the common generalization of a bending or earthquake deformation (corresponding to the full centralizer) is a *quakebend* or *complex earthquake*. For further discussion of these deformations, see e.g. [32], [114], [86].

Bending cocycles. An alternate definition of the bending deformation makes the geometric content of the construction more apparent. Realize the simple closed curve γ as a hyperbolic geodesic on the surface $X \in \mathcal{T}(S)$, and consider the full preimage $\tilde{\gamma} \subset \mathbb{H}^2$ of γ in the universal cover; thus $\tilde{\gamma}$ consists of infinitely many complete geodesics, the *lifts* of γ . By analogy with the terminology for a pleated plane in \mathbb{H}^3 , the connected components of $\mathbb{H}^2 - \tilde{\gamma}$ will be called *plaques*. For the purposes of this discussion we regard \mathbb{H}^2 as a plane in \mathbb{H}^3 , stabilized by $\text{PSL}_2(\mathbb{R}) \subset \text{PSL}_2(\mathbb{C})$.

Given $x, y \in (\mathbb{H}^2 - \tilde{\gamma})$, let (g_1, \dots, g_n) be the set of lifts of γ that separate x from y , ordered according to the way they intersect the oriented geodesic segment from x to y , with g_1 closest to x . Orient each geodesic g_i so that y lies to the right. For any $t \in \mathbb{R}$, define the *bending cocycle* $B(x, y) \in \text{PSL}_2(\mathbb{C})$ by

$$B(x, y) = E(g_1, t)E(g_2, t) \dots E(g_n, t),$$

where $E(g, t)$ is an elliptic Möbius transformation with fixed axis g and clockwise rotation angle t .

In case x and y lie in a facet, this empty product is understood to be the identity. This construction defines a map $B : (\mathbb{H}^2 - \tilde{\gamma}) \times (\mathbb{H}^2 - \tilde{\gamma}) \rightarrow \text{PSL}_2(\mathbb{C})$. Clearly we have $B(x, x) = I$ and $B(x, y)$ only depends on the plaques containing x and y . Furthermore, the map B satisfies the *cocycle relation*

$$B(x, y)B(y, z) = B(x, z) \quad \text{for all } x, y, z \in \mathbb{H}^2 - \tilde{\gamma}, \tag{5.2}$$

and the *equivariance relation*

$$B(\gamma x, \gamma y) = \rho_0(\gamma)B(x, y)\rho_0(\gamma)^{-1} \quad \text{for all } \gamma \in \pi_1(S) \tag{5.3}$$

where $\rho_0 \in \mathcal{F}(S)$ represents Y .

The connection between the bending cocycle and the bending deformation described above is as follows (compare [32, Lemma 3.7.1]).

Lemma 5.5. *Given $Y \in \mathcal{T}(S)$, a simple closed curve γ , and $t \in \mathbb{R}$, choose a basepoint $O \in (\mathbb{H}^2 - \tilde{\gamma})$ and define*

$$\rho(\gamma) = B(O, \gamma O)\rho_0(\gamma),$$

where $\rho_0 \in \mathcal{F}(S)$ represents Y and B is the bending cocycle associated to Y, γ , and t . Then ρ is a homomorphism, and it lies in the same conjugacy class as the bending deformation $\beta_{t\gamma}(Y)$.

In other words, the bending cocycle records the “difference” between a Fuchsian character ρ_0 and the deformed character $\beta_{t\gamma}(\rho_0)$. The bending cocycle and this lemma extend naturally to measured laminations [32, §3.5.3].

Bending and grafting. The key observation relating bending and grafting is that the bending deformation $\beta_\lambda(Y): \pi_1(S) \rightarrow \mathrm{PSL}_2(\mathbb{C})$ preserves the locally convex pleated plane in \mathbb{H}^3 with intrinsic hyperbolic structure Y and bending lamination λ . In exploring this connection, let us suppose that $\lambda = t\gamma$ is supported on a simple closed curve. The pleating map $\mathrm{Pl}: \mathbb{H}^2 \rightarrow \mathbb{H}^3$ can be defined in terms of the bending cocycle as

$$\mathrm{Pl}(x) = B(O, x)x$$

where as before $O \in (\mathbb{H}^2 - \tilde{\gamma})$ is a base point. Equivariance of this map with respect to $\pi_1(S)$ then follows from Lemma 5.5 and the properties (5.2)–(5.3) of the bending cocycle. As written, this pleating map is only defined on $\mathbb{H}^2 - \tilde{\gamma}$, however it extends continuously to \mathbb{H}^2 because on the two sides of a lift $g \subset \tilde{\gamma}$, the values of $B(O, \bullet)$ differ by an elliptic Möbius transformation that fixes g pointwise.

The same reasoning shows that the image of Pl is a locally convex pleated plane: Since B is locally constant away from $\tilde{\gamma}$, the plaques map into planes in \mathbb{H}^3 , and when two such plaques share a boundary geodesic g , the images of the plaques in \mathbb{H}^3 meet along a geodesic $\mathrm{Pl}(g)$ with bending angle t (which is to say, their enveloping planes are related by an elliptic Möbius transformation fixing their line of intersection, with rotation angle t).

We have seen that the holonomy of the projective structure $Z = \mathrm{Gr}_\lambda Y$ also preserves the equivariant pleated plane in \mathbb{H}^3 constructed by bending $\tilde{Y} \simeq \mathbb{H}^2$ along λ . This leads to the fundamental relationship between grafting, bending and the holonomy map (see [86, §2]):

$$\mathrm{hol}(\mathrm{Gr}_\lambda Y) = \beta_\lambda(Y). \tag{5.4}$$

For laminations supported on simple closed curves, this is simply the observation that the processes of inserting lunes into $\mathbb{H} \subset \mathbb{CP}^1$ (which gives projective grafting) and bending $\mathbb{H}^2 \subset \mathbb{H}^3$ along geodesics (which gives the bending deformation) are related to one another by the convex hull construction of §4.2. The general equality follows from this case by continuity of hol , Gr , and β .

Using (5.4) we can think of projective grafting as a “lift” of the bending map $\beta: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{X}(S)$ through the locally diffeomorphic holonomy map $\mathrm{hol}: \mathcal{P}(S) \rightarrow \mathcal{X}(S)$ (which is not a covering).

5.4 Fuchsian holonomy

Let $\mathcal{P}_{\mathcal{F}}(S) = \mathrm{hol}^{-1}(\mathcal{F}(S))$ denote the set of all projective structures with Fuchsian holonomy.

We can construct examples of projective structures in $\mathcal{P}_{\mathcal{F}}(S)$ using grafting. Because of the 2π -periodicity of bending along a simple closed geodesic γ , the projective structures $\{\text{Gr}_{2\pi n\gamma}Y \mid n \in \mathbb{N}\}$ all have the same Fuchsian holonomy representation ρ_0 (up to conjugacy), which is the representation uniformizing Y . Of course $n = 0$ gives the standard Fuchsian structure on Y .

For $n > 0$ these projective structures have underlying Riemann surfaces of the form $\text{gr}_{2\pi n\gamma}Y$, and due to the 2π -lunes inserted in the projective grafting construction, their developing maps are surjective. This construction of “exotic” Fuchsian projective structures is due independently to Maskit [83], Hejhal [47, Theorem 4], and Sullivan–Thurston [109].

Goldman’s classification. Let $\mathcal{ML}_{\mathbb{Z}}(S)$ denote the countable subset of $\mathcal{ML}(S)$ consisting of disjoint collections of simple closed geodesics with positive integral weights. Generalizing the case of a single geodesic, every projective structure of the form $\text{Gr}_{2\pi\lambda}Y$ with $\lambda \in \mathcal{ML}_{\mathbb{Z}}(S)$ has Fuchsian holonomy. Goldman showed that all Fuchsian projective structures arise in this way:

Theorem 5.6 (Goldman [39]). *Let $Z \in \mathcal{P}_{\mathcal{F}}(S)$ and let $Y = \mathbb{H}^2/\text{hol}(Z)(\pi_1(S))$ be the hyperbolic surface associated to the holonomy representation. Then $Z = \text{Gr}_{2\pi\lambda}Y$ for some $\lambda \in \mathcal{ML}_{\mathbb{Z}}$.*

In terms of the holonomy map $\text{hol}: \mathcal{P}(S) \rightarrow \mathcal{T}(S)$, this result shows that we can identify $\mathcal{P}_{\mathcal{F}}(S)$ with countably many copies of Teichmüller space,

$$\text{Gr}^{-1}: \mathcal{P}_{\mathcal{F}}(S) \xrightarrow{\cong} (2\pi \mathcal{ML}_{\mathbb{Z}}(S)) \times \mathcal{T}(S),$$

and the restriction of the holonomy map to any one of these spaces $\{2\pi\lambda\} \times \mathcal{T}(S)$ gives the natural isomorphism $\mathcal{T}(S) \simeq \mathcal{F}(S)$.

Alternatively, using Theorem 4.3 in combination with Theorem 5.6, we can characterize $\mathcal{P}_{\mathcal{F}}(S)$ as the union of countably many sections of π ,

$$\mathcal{P}_{\mathcal{F}}(S) = \bigcup_{\lambda \in \mathcal{ML}_{\mathbb{Z}}} \sigma_{2\pi\lambda}(\mathcal{T}(S)).$$

Note the difference between these two descriptions of $\mathcal{P}_{\mathcal{F}}(S)$: In describing it as a union of sections, we see that the intersection of $\mathcal{P}_{\mathcal{F}}(S)$ with a fiber $P(X) = \pi^{-1}(X)$ consists of a countable discrete set naturally identified with $\mathcal{ML}_{\mathbb{Z}}(S)$, whereas in the holonomy picture we describe the intersection of $\mathcal{P}_{\mathcal{F}}(S)$ with $\text{hol}^{-1}(Y)$ in similar terms.

Describing $\mathcal{P}_{\mathcal{F}}(S)$ as a union of the smooth sections $\sigma_{2\pi\lambda}(\mathcal{T}(S))$ of π also allows us to conclude that each intersection between $\mathcal{P}_{\mathcal{F}}(S)$ and a fiber $P(X)$ is transverse. Previously, Faltings established this transversality result in the greater generality of *real holonomy*, that is, the projective structures in $\text{hol}^{-1}(\mathcal{X}_{\mathbb{R}}(S))$ where $\mathcal{X}_{\mathbb{R}}(S) \subset \mathcal{X}(S)$ consists of real-valued characters of homomorphisms of $\pi_1(S)$ into $\text{PSL}_2(\mathbb{C})$.

Theorem 5.7 (Faltings [33]). *Let $Z \in P(X)$ be a projective structure with real holonomy. Then $\text{hol}(P(X))$ is transverse to $\mathcal{X}_{\mathbb{R}}(S)$ at $\text{hol}(Z)$.*

The characters in $\mathcal{X}_{\mathbb{R}}(S)$ correspond to homomorphisms that are conjugate into $\text{SU}(2)$ or $\text{PSL}_2(\mathbb{R})$. Both cases include many non-Fuchsian characters, as a homomorphism $\rho: \pi_1(S) \rightarrow \text{PSL}_2(\mathbb{R})$ is Fuchsian if and only if its Euler class is maximal, $e(\rho) = 2g - 2$. Goldman describes the projective structures with real holonomy in terms of grafting in [39, §2.14], [40, pp. 14–15].

5.5 Quasi-Fuchsian holonomy

Let $\mathcal{P}_{\mathcal{QF}}(S) = \text{hol}^{-1}(\mathcal{F}(S))$ denote the set of all projective structures with quasi-Fuchsian holonomy, which is an open subset of $\mathcal{P}(S)$.

Goldman's proof of Theorem 5.6 involves a study of the topology and geometry of developing maps of Fuchsian projective structures. The topological arguments apply equally well to projective structures with quasi-Fuchsian holonomy, and the information they provide can be summarized as follows:

Theorem 5.8 (Goldman [39]). *Let $Z \in \mathcal{P}_{\mathcal{QF}}(S)$ have developing map $f: \tilde{Z} \rightarrow \mathbb{CP}^1$, and let $\Lambda \subset \mathbb{CP}^1$ be the limit set of the holonomy group, a Jordan curve with complementary regions Ω_{\pm} . Then:*

- (1) *The quotient of the developing preimage of the limit set, denoted $\Lambda(Z) = f^{-1}(\Lambda)/\pi_1(S)$, consists of a finite collection of disjoint simple closed curves.*
- (2) *The quotient of the developing preimage of Ω_- , denoted $Z_- = f^{-1}(\Omega_-)/\pi_1(S)$ consists of a finite collection of disjoint homotopically essential annuli bounded by the curves in $\Lambda(Z)$. In particular, the curves in $\Lambda(Z)$ are naturally grouped into isotopic pairs.*

Recall that among the two domains of discontinuity, Ω_+ is distinguished by the fact that the orientation of its quotient marked Riemann surface agrees with that of S , while that of the quotient of Ω_- is opposite.

The topology of a typical (surjective) quasi-Fuchsian developing map is represented schematically in Figure 5.

Wrapping invariant. Given this description of the preimage of the limit set, there is a natural \mathbb{Z} -weighted multicurve associated to a quasi-Fuchsian projective structure Z : Suppose the collection of annuli Z_- represents homotopy classes $\gamma_1, \dots, \gamma_n$, and that there are n_i parallel annuli homotopic to γ_i . Define the *wrapping invariant*

$$\text{wr}(Z) = \sum_i n_i \gamma_i \in \mathcal{ML}_{\mathbb{Z}}(S).$$

Note that we could have also defined this using the family of curves $\Lambda(Z)$, since $2n_i$ is the number of parallel curves homotopic to γ_i .

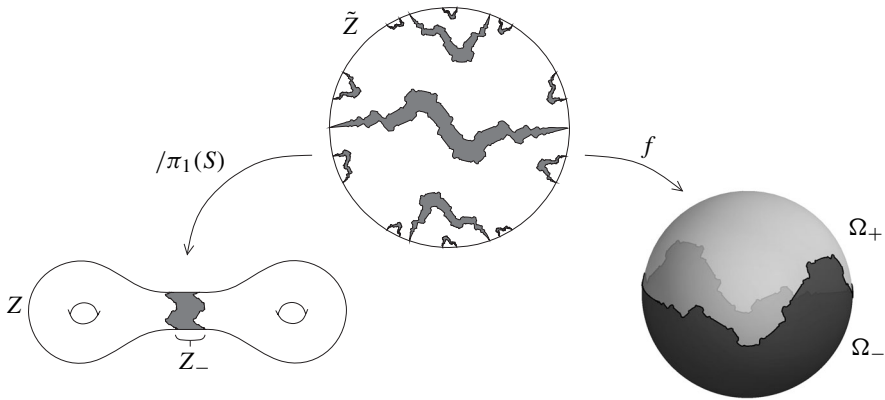


Figure 5. The relationship between the developing map f and the domains of discontinuity Ω_{\pm} for a projective structure with quasi-Fuchsian holonomy. In this example, the open set Z_- is an annulus, so the wrapping invariant is a simple closed curve with unit weight.

Theorem 5.6 is derived from Theorem 5.8 by showing that for a Fuchsian projective structure Z , we have

$$Z = \text{Gr}_{2\pi \text{wr}(Z)} Y,$$

where Y is the quotient of \mathbb{H}^2 by the holonomy group, as above. In other words, for Fuchsian projective structures, the wrapping invariant is the grafting lamination (up to a multiple of 2π).

Quasi-Fuchsian components. Because limit sets vary continuously in $\mathcal{QF}(S)$, the wrapping invariant is a locally constant function on $\mathcal{P}_{\mathcal{QF}}(S)$. Thus $\mathcal{P}_{\mathcal{QF}}(S)$ breaks into countably many subsets

$$\mathcal{P}_{\mathcal{QF}}(S) = \bigcup_{\lambda \in \mathcal{ML}_{\mathbb{Z}}(S)} \mathcal{P}_{\lambda}(S) \quad \text{where } \mathcal{P}_{\lambda}(S) = \text{wr}^{-1}(\lambda).$$

We will refer to these as *components* of $\mathcal{P}_{\mathcal{QF}}(S)$.

The quasi-Fuchsian component with zero wrapping invariant, $\mathcal{P}_0(S)$, consists of *standard* quasi-Fuchsian structures. The holonomy map gives a diffeomorphism

$$\text{hol}: \mathcal{P}_0(S) \rightarrow \mathcal{QF}(S),$$

where the inverse map associates to $\rho \in \mathcal{QF}(S)$ the induced projective structure on the quotient $\Omega_+/\rho(\pi_1(S))$ of one domain of discontinuity. The developing map of a standard quasi-Fuchsian projective structure is a Riemann map $f: \mathbb{H} \xrightarrow{\cong} \Omega_+$.

The other components $\mathcal{P}_{\lambda}(S)$, with $\lambda \neq 0$, consist of *exotic* quasi-Fuchsian projective structures; as in the Fuchsian case, these have surjective developing maps. Unlike

the Fuchsian case, however, the components $\mathcal{P}_\lambda(S)$ do not have a simple description in terms of the grafting coordinates on $\mathcal{P}(S)$. Nevertheless, when restricted to one of these components, the holonomy map

$$\text{hol}: \mathcal{P}_\lambda(S) \rightarrow \mathcal{QF}(S),$$

is again a diffeomorphism. The inverse $\mathcal{QF}(S) \rightarrow \mathcal{P}_\lambda(S)$ can be constructed by either of two methods:

- (1) *Constructive approach.* In a generalization of 2π -integral projective grafting, one starts with a standard quasi-Fuchsian projective structure Z and glues annuli into the surface to produce a new projective structure which has the same holonomy but which has wrapping invariant λ . Allowing the starting structure to vary gives a map $\mathcal{QF}(S) \simeq \mathcal{P}_0(S) \rightarrow \mathcal{P}_\lambda(S)$ that is inverse to hol . See [39, §1.2], [60, §2.4] [66, Chapter 7] for details.
- (2) *Deformation approach.* Starting with a fixed Fuchsian representation $\rho_0 \in \mathcal{F}(S)$, any quasi-Fuchsian representation ρ can be obtained by a ρ_0 -equivariant quasiconformal deformation. By pulling back the quasiconformal deformation through a developing map, one can simultaneously deform a Fuchsian projective structure Z_0 with holonomy ρ_0 to obtain a quasi-Fuchsian structure Z with holonomy ρ . This deformation does not change the wrapping invariant, so starting with $Z_0 = \text{Gr}_{2\pi\lambda}X$ and considering all quasiconformal deformations gives the desired map $\mathcal{QF}(S) \rightarrow \mathcal{P}_\lambda(S)$. See [107, §3], [57, §2.5].

Thus the structure of $\mathcal{P}_{\mathcal{QF}}(S)$ is similar to that of $\mathcal{P}_{\mathcal{F}}(S)$ described above: It consists of countably many connected components $\mathcal{P}_\lambda(S)$, each of which is diffeomorphic to $\mathcal{QF}(S)$ by the holonomy map (compare [66, §7.2], [57, §§2.5–2.6]).

Bumping of quasi-Fuchsian components. We say that two components $\mathcal{P}_\lambda(S)$ and $\mathcal{P}_\mu(S)$ *bump* if their closures intersect, i.e. if $\overline{\mathcal{P}_\lambda(S)} \cap \overline{\mathcal{P}_\mu(S)} \neq \emptyset$; an element of the intersection is called a *bumping point*. A component $\mathcal{P}_\lambda(S)$ *self-bumps* at $Z \in \mathcal{P}(S)$ if $U \cap \mathcal{P}_\lambda(S)$ is disconnected for all sufficiently small neighborhoods U of Z . These terms are adapted from similar phenomena in the theory of deformation spaces of Kleinian groups (surveyed in [16], see also [3], [4], [15], [50]).

The bumping of quasi-Fuchsian components has been studied by McMullen [86], Bromberg–Holt [14], and Ito [57], [60]. The basic problem of determining which component pairs bump is resolved by:

Theorem 5.9 (Ito [60]). (1) For any $\lambda, \mu \in \mathcal{ML}_{\mathbb{Z}}(S)$, the components $\mathcal{P}_\lambda(S)$ and $\mathcal{P}_\mu(S)$ bump.

(2) For any $\lambda \in \mathcal{ML}_{\mathbb{Z}}(S)$, the component $\mathcal{P}_\lambda(S)$ self-bumps at a point in $\overline{\mathcal{P}_0(S)}$.

The bumping points constructed in the proof of this theorem are all derived from a construction of Anderson–Canary that illustrates the difference between algebraic and geometric convergence for Kleinian groups [3]. This construction was first applied

to projective structures by McMullen to give an example of bumping between $\mathcal{P}_\lambda(S)$ and $\mathcal{P}_0(S)$ [86]. The holonomy representations for these bumping examples have accidental parabolics but are not quasiconformally rigid; recently, Brock, Bromberg, Canary, and Minsky have shown that these conditions are necessary for bumping [12] (compare [92]).

5.6 Discrete holonomy

Let $\mathcal{D}(S) \subset \mathcal{X}(S)$ denote the set of characters of discrete representations, and let $\mathcal{P}_\mathcal{D}(S)$ denote the set of projective structures with discrete holonomy. Since $\mathcal{F}(S) \subset \mathcal{QF}(S) \subset \mathcal{D}(S)$, we have corresponding inclusions

$$\mathcal{P}_\mathcal{F}(S) \subset \mathcal{P}_{\mathcal{QF}}(S) \subset \mathcal{P}_\mathcal{D}(S).$$

Because hol is a local diffeomorphism, topological properties of $\mathcal{D}(S)$ correspond to those of $\mathcal{P}_\mathcal{D}(S)$. For example, $\mathcal{D}(S)$ is closed (see [61], [19]), and its interior is the set $\mathcal{QF}(S)$ of quasi-Fuchsian representations [108], [7]. Thus $\mathcal{P}_\mathcal{D}(S)$ is a closed subset of $\mathcal{P}(S)$ with interior $\mathcal{P}_{\mathcal{QF}}(S)$.

If $Z \in \mathcal{P}_\mathcal{D}(S)$ has holonomy ρ , then the associated pleated plane $\text{Pl}(Z) : \mathbb{H}^2 \rightarrow \mathbb{H}^3$ is invariant under the holonomy group $\Gamma = \rho(\pi_1(S))$ and descends to a *locally convex pleated surface* in the quotient hyperbolic manifold $M = \mathbb{H}^3 / \Gamma$:

$$\begin{array}{ccc} \mathbb{H}^2 & \xrightarrow{\text{Pl}(Z)} & \mathbb{H}^3 \\ \downarrow / \pi_1(S) & & \downarrow / \Gamma \\ Y & \longrightarrow & M. \end{array}$$

Here $Y \in \mathcal{T}(S)$ is the hyperbolic surface such that $Z = \text{Gr}_\lambda Y$ for some $\lambda \in \mathcal{ML}(S)$.

The pleated surface arising from a projective structure Z with discrete holonomy may be one of the connected components of the boundary of the *convex core* of the associated hyperbolic manifold M . If so, the projective surface Z is the component of the ideal boundary of M on the “exterior” side of the pleated surface. Conversely, the ideal boundary and convex core boundary surfaces in a complete hyperbolic manifold are related by grafting (see [103, §5.1] [86, §2.8]).

For more general projective structures with discrete holonomy, the pleated surface need not be embedded in the quotient manifold, however it must lie within the convex core (see [17, §5.3.11]).

In addition to the Fuchsian and quasi-Fuchsian cases described above, projective structures with other classes of discrete holonomy representations have found application in Kleinian groups and hyperbolic geometry. For example, projective structures with degenerate holonomy are used in Bromberg’s approach to the Bers density conjecture [13], and those with Schottky holonomy are used in Ito’s study of sequences of Schottky groups accumulating on Bers’ boundary of Teichmüller space [58].

5.7 Holonomy in fibers

In contrast to the complicated global properties of the holonomy map, its restriction to a fiber is very well-behaved:

Theorem 5.10. *For each $X \in \mathcal{T}(S)$, the restriction $\text{hol}|_{P(X)}$ is a proper holomorphic embedding, whose image $\text{hol}(P(X))$ is a complex-analytic subvariety of $\mathcal{X}(S)$.*

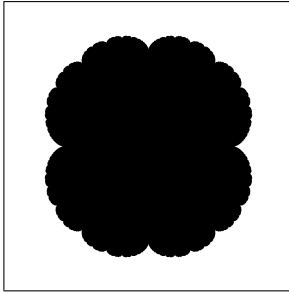
As stated, this theorem incorporates several related but separate results: Working in the context of systems of linear ODE on a fixed Riemann surface, Poincaré showed that the holonomy map is injective [5, p. 310] (see also [76], [47, Theorem 15]). Gallo, Kapovich, and Marden showed that the image is a complex-analytic subvariety [36], following an outline given by Kapovich [65]; when combined with injectivity, this implies properness. Tanigawa gave a more geometric argument establishing properness of $\text{hol}|_{P(X)}$ when considered as a map into the space $\mathcal{X}'(S)$ of non-elementary characters [111]. Tanigawa's argument relies on the existence of loxodromic pants decompositions (Theorem 5.4), which was announced in [65] and proved in [36].

Fuchsian and quasi-Fuchsian holonomy in fibers. For any $X \in \mathcal{T}(S)$, let $P_{\mathcal{D}}(X) = P(X) \cap \mathcal{P}_{\mathcal{D}}(S)$ denote the set of projective structures with discrete holonomy and with underlying complex structure X . Similarly, we define $P_{\mathcal{QF}}(X)$ and $P_{\mathcal{F}}(X)$ as the subsets of $P(X)$ having quasi-Fuchsian and Fuchsian holonomy, respectively.

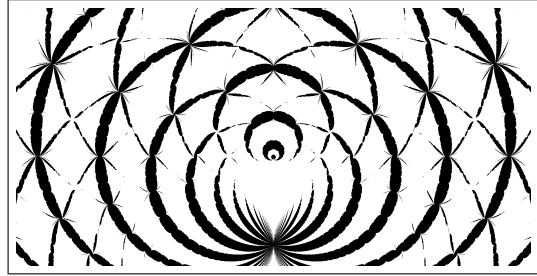
We have already seen (in §5.4) that the $P_{\mathcal{F}}(X)$ consists of the countable discrete set of projective structures $\{\sigma_{2\pi\lambda}(X) \mid \lambda \in \mathcal{ML}_{2\pi\mathbb{Z}}(S)\}$. Since the holonomy map is continuous, and $\mathcal{QF}(S)$ is an open neighborhood of $\mathcal{F}(S)$ in $\mathcal{X}(S)$, each of these Fuchsian points has a neighborhood in $P(X)$ consisting of quasi-Fuchsian projective structures with the same wrapping invariant. Elements of $P_{\mathcal{F}}(X)$ are sometimes called *Fuchsian centers* (or *centers of grafting* [2]), because they provide distinguished center points within these “islands” of quasi-Fuchsian holonomy (see [26, §13], [82, Theorem 6.6.10]).

Using the Schwarzian parameterization, the intersection $\mathcal{P}_0(S) \cap P(X)$, consisting of the standard quasi-Fuchsian projective structures on X , can be considered as an open set $B_X \subset \mathcal{Q}(X) \simeq \mathbb{C}^{3g-3}$. This set is the image of the holomorphic *Bers embedding* of Teichmüller space [106], and in particular it is connected and contractible. We also have $B(1/2) \subset B_X \subset B(3/2)$, where $B(r) = \{\phi \in \mathcal{Q}(X) \mid \|\phi\|_{\infty} < r\}$, as a consequence of Nehari's theorem [91]. See Figure 6 for examples of Bers embeddings of the Teichmüller space of punctured tori.

For $\lambda \neq 0$, it is not known whether the set $\mathcal{P}_{\lambda}(S) \cap P(X)$ is connected (or bounded), though experimental evidence in the punctured case suggests that it often has many connected components, and that the structure of the connected components changes with X (see Figure 7). Of course, only one component contains the Fuchsian structure $\sigma_{2\pi\lambda}(X)$.



(a) The Bers embedding of the square punctured torus.



(b) In this larger view, the Bers embedding of a punctured torus with a short geodesic appears as a small dot (center) surrounded by many islands of exotic quasi-Fuchsian projective structures.

Figure 6. Islands of quasi-Fuchsian holonomy in $P(X) \simeq \mathbb{C}$ (where X is a punctured torus) exhibit complicated structure at small and large scales. These images were created using the software package *Bear* [24].

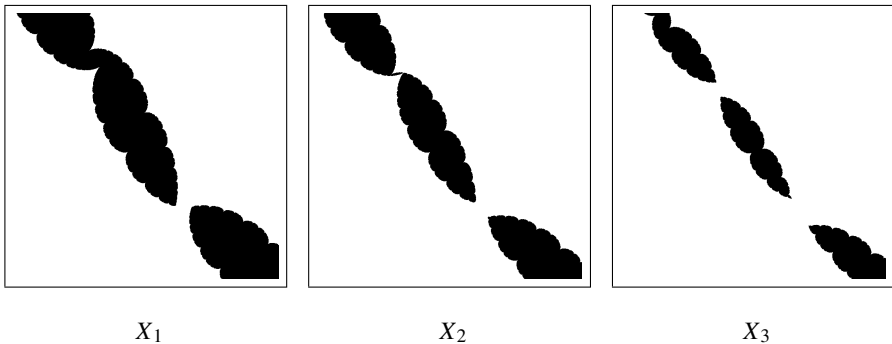


Figure 7. Islands of quasi-Fuchsian holonomy in $P(X)$ appear to break apart as the complex structure X is changed, suggesting that some islands do not contain Fuchsian centers. Each image shows a small square in $P(X_i) \simeq \mathbb{C}$, where $\{X_1, X_2, X_3\}$ are closely-spaced points in the Teichmüller space of the punctured torus.

Quasi-Fuchsian versus discrete in a fiber. In the space of all projective structures, the quasi-Fuchsian structures form the interior of the set with discrete holonomy. The same relationship holds for $P_{\mathcal{QF}}(X)$ and $P_{\mathcal{D}}(X)$.

Theorem 5.11 (Shiga and Tanigawa [107], Matsuzaki [84]). *For any $X \in \mathcal{T}(S)$, we have $P_{\mathcal{QF}}(X) = \text{int}(P_{\mathcal{D}}(X))$.*

In comparing these sets, one inclusion is immediate: Since $\text{int}(\mathcal{P}_{\mathcal{D}}(S)) = \mathcal{P}_{\mathcal{QF}}(S)$, we have $\text{int}(P_{\mathcal{D}}(X)) \supset P_{\mathcal{QF}}(X)$. The opposite inclusion is more subtle. Each component of the interior of $P_{\mathcal{D}}(X)$ necessarily consists of quasiconformally conjugate, discrete, faithful representations without accidental parabolics. However there exist $(3g - 3)$ -dimensional holomorphic families of *singly degenerate* surface groups in $\mathcal{X}(S)$ which satisfy these conditions, but which are not quasi-Fuchsian. Such a family could account for an open subset of $P_{\mathcal{D}}(X)$ (in either of two topologically distinct ways [84]), and a key step in the proof of the theorem is to exclude this possibility.

6 Comparison of parameterizations

6.1 Compactifications

Compactification of $\mathcal{ML}(S)$. The space of measured laminations has the structure of a cone: The group \mathbb{R}^+ acts by scaling the transverse measure ($\lambda \mapsto t\lambda$, $t \in \mathbb{R}^+$) and the empty lamination $0 \in \mathcal{ML}(S)$ is the unique fixed point of this action. The orbit of a nonzero lamination is a *ray* in $\mathcal{ML}(S)$. The space of rays,

$$\mathbb{P}\mathcal{ML}(S) = (\mathcal{ML}(S) - \{0\})/\mathbb{R}^+,$$

or *projective measured laminations* forms a natural boundary for $\mathcal{ML}(S)$. We say that a sequence $\lambda_i \in \mathcal{ML}(S)$ converges to $[\lambda] = \mathbb{R}^+ \cdot \lambda \in \mathbb{P}\mathcal{ML}(S)$ if there exists a sequence of positive real numbers c_i such that $c_i \rightarrow 0$ and $c_i \lambda_i \rightarrow \lambda$ in $\mathcal{ML}(S)$. The induced compactification

$$\overline{\mathcal{ML}(S)} = \mathcal{ML}(S) \cup \mathbb{P}\mathcal{ML}(S)$$

is homeomorphic to a closed ball, with interior $\mathcal{ML}(S) \simeq \mathbb{R}^{6g-6}$ and boundary $\mathbb{P}\mathcal{ML}(S) \simeq S^{6g-7}$. See [94, Chapter 3] for further discussion of the spaces $\mathcal{ML}(S)$ and $\mathbb{P}\mathcal{ML}(S)$, and [34] for related discussion of the space of measured foliations, which is naturally identified with $\mathcal{ML}(S)$ (as described in [80], [66, §11.8–11.9]).

Compactification of $\mathcal{T}(S)$. Recall that \mathcal{S} denotes the set of isotopy classes of simple closed curves on S , or equivalently, the simple closed geodesics of any hyperbolic structure on S . Thurston defined a compactification of $\mathcal{T}(S)$ using the hyperbolic length map

$$\begin{aligned} \mathbb{L}: \mathcal{T}(S) &\rightarrow \mathbb{R}^{\mathcal{S}}, \\ X &\mapsto (\ell(\gamma, X))_{\gamma \in \mathcal{S}}. \end{aligned}$$

This map is an embedding, as is its projectivization

$$\mathbb{P}\mathbb{L}: \mathcal{T}(S) \rightarrow \mathbb{P}^+\mathbb{R}^{\mathcal{S}} = (\mathbb{R}^{\mathcal{S}} - \{0\})/\mathbb{R}^+,$$

and in each case, a suitable finite subset of \mathcal{S} suffices to determine the image of a point. The boundary $\partial\mathbb{P}\mathcal{L}(\mathcal{T}(S))$ coincides with the image of $\mathbb{P}\mathcal{ML}(S)$ under the projectivization of the embedding

$$\begin{aligned} \mathcal{ML}(S) &\rightarrow \mathbb{R}^{\mathcal{S}}, \\ \lambda &\mapsto (i(\gamma, \lambda))_{\gamma \in \mathcal{S}} \end{aligned}$$

where $i(\lambda, \gamma)$ denotes the total mass of γ with respect to the transverse measure of λ . This gives the *Thurston compactification*

$$\overline{\mathcal{T}(S)} = \mathcal{T}(S) \cup \mathbb{P}\mathcal{ML}(S)$$

which has the topology of a closed $(6g - 6)$ -ball. Concretely, a sequence $X_n \rightarrow \infty$ in Teichmüller space converges to $[\lambda] \in \mathbb{P}\mathcal{ML}(S)$ if for every pair of simple closed curves $\alpha, \beta \in \mathcal{S}$ we have

$$\frac{\ell(\alpha, X_i)}{\ell(\beta, X_i)} \rightarrow \frac{i(\alpha, \lambda)}{i(\beta, \lambda)}$$

whenever the right hand side is well-defined (i.e. $i(\beta, \lambda) \neq 0$). A detailed discussion of the Thurston compactification can be found in [34, Exposé 7–8] (see also [112], [117], [8], [66, Chapter 11], [82, §5.9]).

Compactification of $Q(X)$. Since the vector space $Q(X)$ has an action of \mathbb{R}^+ by scalar multiplication, it supports a natural compactification analogous to that of $\mathcal{ML}(S)$; in this case, the boundary is the space of rays

$$\mathbb{P}^+ Q(X) = (Q(X) - \{0\})/\mathbb{R}^+$$

and we obtain $\overline{Q(X)} = Q(X) \cup \mathbb{P}^+ Q(X)$ which is homeomorphic to a closed ball.

6.2 Quadratic differentials and measured laminations

The Hubbard–Masur theorem. For any $X \in \mathcal{T}(S)$, there is a natural map

$$\Lambda: Q(X) \rightarrow \mathcal{ML}(S)$$

which is defined by a two-step procedure: First, a quadratic differential ϕ has an associated *horizontal foliation* $\mathcal{F}(\phi)$, a singular foliation on X which integrates the distribution of vectors $v \in TX$ such that $\phi(v) \geq 0$. This foliation is equipped with a transverse measure, induced by integration of $|\operatorname{Im} \sqrt{\phi}|$. In a local coordinate where $\phi = dz^2$, the foliation is induced by the horizontal lines in \mathbb{C} , with transverse measure $|dy|$. Zeros of ϕ correspond to singularities of the foliation, where three or more half-leaves emanate from a point. See e.g. [66, §5.3, §11.3], [37, §2.2, Chapter 11] for a discussion of quadratic differentials and their measured foliations.

Now lift the horizontal foliation of ϕ to the universal cover $\tilde{X} \simeq \mathbb{H}^2$. Each non-singular leaf of the lifted foliation is a uniform quasi-geodesic, so it is a bounded distance from unique hyperbolic geodesic. The hyperbolic geodesics obtained in this

way – the *straightening* of \mathcal{F} – form the lift of a geodesic lamination on X , and the transverse measure of the foliation induces a transverse measure on this lamination in a natural way [80]. The result is a measured lamination $\Lambda(\phi) \in \mathcal{ML}(S)$, which we call the *horizontal lamination* of ϕ .

The same constructions can be applied to the distribution of vectors satisfying $\phi(v) \leq 0$, which gives the *vertical foliation* and *vertical lamination* of ϕ . The former is induced by the foliation of \mathbb{C} by vertical lines in local coordinates such that $\phi = dz^2$. Note that multiplication by -1 in $Q(X)$ exchanges vertical and horizontal: for example, the horizontal lamination of $-\phi$ is the vertical lamination of ϕ .

The strong connection between quadratic differentials and measured laminations is apparent in:

Theorem 6.1 (Hubbard and Masur [53]). *For each $X \in \mathcal{T}(S)$, the map $\Lambda: Q(X) \rightarrow \mathcal{ML}(S)$ is a homeomorphism. In particular, every measured lamination is realized by a unique quadratic differential on X .*

Note that Hubbard and Masur work with measured foliations rather than measured laminations; the statement above incorporates the aforementioned straightening procedure to identify the two notions.

We call the inverse of Λ the *foliation map*, denoted $\phi_F: \mathcal{ML}(S) \rightarrow Q(X)$. Note that the definition of both Λ and ϕ_F depend on the choice of a fixed conformal structure X , but we suppress this dependence in the notation.

Since the transverse measure of $\Lambda(\phi)$ is obtained by integrating $|\operatorname{Im} \sqrt{\phi}|$, these maps have the following homogeneity properties:

$$\begin{aligned}\Lambda(c\phi) &= c^{\frac{1}{2}} \Lambda(\phi), \\ \phi_F(c\lambda) &= c^2 \phi_F(\lambda)\end{aligned}$$

for all $c \in \mathbb{R}^+$. Therefore Λ and ϕ_F descend to mutually inverse homeomorphisms between the spaces of rays $\mathbb{P}\mathcal{ML}(S)$ and $\mathbb{P}^+Q(X)$, and we also use Λ and ϕ_F to denote these induced maps.

Orthogonality and the antipodal map. Given $X \in \mathcal{T}(S)$, a pair of measured laminations $\lambda, \mu \in \mathcal{ML}(S)$ is *orthogonal with respect to X* if there exists $\phi \in Q(X)$ such that

$$\begin{aligned}\Lambda(\phi) &= \lambda, \\ \Lambda(-\phi) &= \mu.\end{aligned}$$

That is, λ and μ appear as the horizontal and vertical laminations of a single holomorphic quadratic differential on X . (Compare the torus case shown in Figure 8.)

By Theorem 6.1, two laminations λ and μ are orthogonal with respect to X if and only if

$$\phi_F(\lambda) = -\phi_F(\mu) \in Q(X).$$

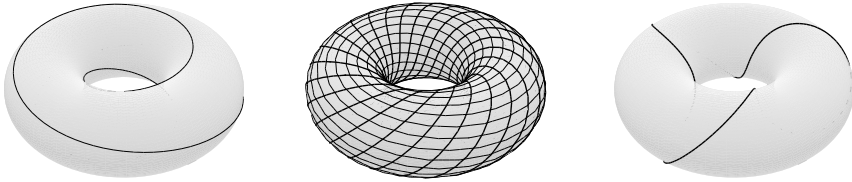


Figure 8. A pair of closed curves on a compact Riemann surface of genus 1 (as seen here on the far left and right) are “orthogonal” if they are isotopic to leaves of a pair of orthogonal geodesic foliations of the Euclidean metric (center). This situation is non-generic; typically, at least one of the two foliations will have dense leaves. For surfaces of higher genus, orthogonality of measured laminations is defined similarly, however there are many distinct singular Euclidean metrics.

Thus the homeomorphism $\phi_F: \mathcal{ML}(S) \rightarrow Q(X)$ turns orthogonal pairs into opposite quadratic differentials, and the set of X -orthogonal pairs is the graph of the *antipodal involution* $i_X: \mathcal{ML}(S) \rightarrow \mathcal{ML}(S)$ defined by

$$i_X(\lambda) = \Lambda(-\phi_F(\lambda)).$$

By homogeneity of Λ and ϕ_F , the antipodal map descends to $i_X: \mathbb{P}\mathcal{ML}(S) \rightarrow \mathbb{P}\mathcal{ML}(S)$. We say $[\lambda], [\mu] \in \mathbb{P}\mathcal{ML}(S)$ are orthogonal with respect to X if $i_X([\lambda]) = [\mu]$. See [25] for further discussion of the antipodal map and orthogonality.

6.3 Limits of fibers

Using the projective grafting homeomorphism $\text{Gr}: \mathcal{ML}(S) \times \mathcal{T}(S) \rightarrow \mathcal{P}(S)$, we can regard $\overline{\mathcal{ML}(S)} \times \overline{\mathcal{T}(S)}$ as a compactification of $\mathcal{P}(S)$. This is the *grafting compactification*.

Given $X \in \mathcal{T}(S)$, the fiber $P(X) \subset \mathcal{P}(S)$ corresponds to a set of pairs $\text{Gr}^{-1}(X) = \{(\lambda, Y) \mid \text{gr}_\lambda Y = X\}$ in the grafting coordinates. Since $P(X)$ is a distinguished subset of the Schwarzian parameterization of $\mathcal{P}(S)$, studying its behavior in the grafting parameterization is one way to study the relationship between these two coordinate systems. The asymptotic behavior of $P(X)$ can be described in terms of orthogonality:

Theorem 6.2 (Dumas [25]). *Let $(\lambda_n, Y_n) \in \mathcal{ML}(S) \times \mathcal{T}(S)$ be a divergent sequence such that $\text{Gr}_{\lambda_n} Y_n \in P(X)$ for all n . Then*

$$\lim_{n \rightarrow \infty} \lambda_n = [\lambda] \quad \text{if and only if} \quad \lim_{n \rightarrow \infty} Y_n = i_X([\lambda]),$$

where these limits are taken in $\overline{\mathcal{ML}(S)}$ and $\overline{\mathcal{T}(S)}$, respectively.

In particular, the boundary of $P(X)$ in the grafting compactification of $\mathcal{P}(S)$ is the graph of the antipodal involution $i_X: \mathbb{P}\mathcal{ML}(S) \rightarrow \mathbb{P}\mathcal{ML}(S)$.

This theorem can be considered as evidence of compatibility between the grafting coordinates for $\mathcal{P}(S)$ and the foliation of $\mathcal{P}(S)$ by fibers of π . For example, we have:

Corollary 6.3. *The closure of $P(X)$ in $\overline{\mathcal{ML}(S)} \times \overline{\mathcal{T}(S)}$ is homeomorphic to a closed ball of dimension $6g - 6$.*

The proof of Theorem 6.2 in [25] is essentially a study of the collapsing and co-collapsing maps of a complex projective structure, and their relation to the harmonic maps variational problem. We now describe this variational technique, and then outline the main steps in the proof.

Harmonic maps. Let (M, g) and (N, h) be complete Riemannian manifolds, and assume that M is compact. If $f: M \rightarrow N$ is a smooth map, the *energy* of f is defined by

$$\mathcal{E}(f) = \frac{1}{2} \int_M \|df(x)\|_2^2 dg(x).$$

The map f is *harmonic* if it is a critical point of the energy functional. If N is also compact and h has negative sectional curvature, then any nontrivial homotopy class of maps $M \rightarrow N$ contains a harmonic map, and this map is an absolute minimum of the energy functional in the homotopy class [31]. Furthermore, the harmonic map is unique in its homotopy class, unless the image of M is a closed geodesic in N , in which case there is a 1-parameter family of harmonic maps obtained by rotation. General references for the theory of harmonic maps include [29], [30], [104], with particular applications to Teichmüller theory surveyed in [22].

Equivariant harmonic maps. If $\pi_1(M)$ acts by isometries on a Riemannian manifold \hat{N} , then we can define the energy of an equivariant map $\tilde{M} \rightarrow \hat{N}$ by integration of $\|df\|_2^2$ over a fundamental domain for the action of $\pi_1 M$ by deck transformations. This generalizes the energy of smooth maps $M \rightarrow N$, because the action of $\pi_1 M$ on \hat{N} need not have a Hausdorff quotient. Existence of harmonic maps is more delicate in this case, but can sometimes be recovered under additional restrictions on the group action. For example if M is a surface and $N = \mathbb{H}^3$ is equipped with the isometric action coming from a non-elementary representation $\rho: \pi_1(S) \rightarrow \mathrm{PSL}_2(\mathbb{C})$, then there is a unique equivariant harmonic map $h: \tilde{S} \rightarrow \mathbb{H}^3$ [23].

Singular targets. Korevaar and Schoen developed a deep generalization of the theory of harmonic maps in which the Riemannian manifold N is replaced by a nonpositively curved (NPC, also known as locally CAT(0)) metric space [72], [73], [71]. Here the energy functional is approximated by the average squared distance between the image of a point $x \in M$ and the image of a small sphere centered at x . Inequalities comparing distances in NPC spaces to those in Euclidean space have an essential role in the development of this theory.

Generalizing the Riemannian case, we have the following equivariant existence and uniqueness results: If \hat{N} is a locally compact NPC space on which $\pi_1(M)$ acts by isometries without fixing any equivalence class of rays, then there is an equivariant harmonic map $h: \tilde{M} \rightarrow \hat{N}$, which is Lipschitz and energy-minimizing [73]. If furthermore \hat{N} is negatively curved (locally CAT(κ), for some $\kappa < 0$), then the harmonic map is unique unless its image is a geodesic [89].

Harmonic maps from surfaces. When M is 2-dimensional, the energy functional depends only on the conformal class of the metric g , so it makes sense to consider harmonic maps from Riemann surfaces to Riemannian manifolds and nonpositively curved metric spaces. An important invariant of a harmonic map $f: X \rightarrow (N, h)$ from a Riemann surface is its *Hopf differential*

$$\Phi(f) = [f^*(h)]^{2,0} \quad (6.1)$$

which is a holomorphic quadratic differential. In the Riemannian case, the holomorphicity of $\Phi(f)$ is a consequence of the Euler–Lagrange equation of the energy functional [29, §10]. With a suitable generalization of the pullback metric (see [72, §2.3]), a holomorphic Hopf differential is also obtained from a harmonic map to an NPC metric space (compare [88, §5]).

We can use the same formula (6.1) to define a Hopf differential for any smooth map $X \rightarrow (N, h)$, which can be further generalized to maps with L^2 distributional derivatives, and to finite-energy maps to NPC metric spaces [72, Theorem 2.3.1]. The result is a L^1 measurable quadratic differential that is not necessarily holomorphic.

Harmonic maps and dual trees. Recall from §4.2 that for each $\lambda \in \mathcal{ML}(S)$ we have a dual \mathbb{R} -tree T_λ . This tree is an NPC metric space (even CAT(κ) for all $\kappa < 0$) equipped with an isometric action of $\pi_1(S)$. The Hubbard–Masur construction of a quadratic differential on $X \in \mathcal{T}(S)$ with lamination λ can be described in terms of an equivariant harmonic map $X \rightarrow T_\lambda$.

Theorem 6.4 (Wolf [119], Daskalopoulos–Dostoglou–Wentworth [21]). *Let $h: \tilde{X} \rightarrow T_\lambda$ be an equivariant harmonic map to the dual \mathbb{R} -tree of $\lambda \in \mathcal{ML}(S)$. Then $\phi_F(\lambda) = -4\Phi(h)$.*

Harmonic maps and the Thurston compactification. The Thurston compactification of Teichmüller space can also be characterized in terms of Hopf differentials of harmonic maps from a fixed Riemann surface as follows:

Theorem 6.5 (Wolf [118]). *Fix $X \in \mathcal{T}(S)$ and let $Y_n \rightarrow \infty$ be a divergent sequence in $\mathcal{T}(S)$. Let $\Phi_n = \Phi(h_n)$ be the Hopf differential of the harmonic map $h_n: X \rightarrow Y_n$ compatible with the markings. Then*

$$\Lambda(-\Phi_n) \rightarrow [\lambda] \in \mathbb{P}\mathcal{ML}(S) \quad \text{if and only if} \quad Y_n \rightarrow [\lambda] \in \mathbb{P}\mathcal{ML}(S).$$

Collapsing, co-collapsing, and harmonic maps. Using the harmonic maps results presented above, we now describe the main steps of the proof of Theorem 6.2 in [25]. For simplicity, we will suppose that $\text{Gr}_{\lambda_n} Y_n \in P(X)$ and that both grafting coordinates have limits in $\mathbb{P}\mathcal{ML}(S)$, i.e.

$$\lim_{n \rightarrow \infty} \lambda_n = [\lambda], \quad \lim_{n \rightarrow \infty} Y_n = [\mu],$$

and we outline a proof that $i_X([\lambda]) = [\mu]$. The stronger statement of the theorem is derived from the same set of ideas.

Outline of proof of Theorem 6.2.

- (1) Both the collapsing maps $\kappa_n : X \rightarrow Y_n$ and the co-collapsing maps $\hat{\kappa}_n : \tilde{X} \rightarrow T_{\lambda_n}$ are *C-almost harmonic*, meaning that their energies exceed the minimum energies in their homotopy classes by at most C . Here C is a constant that depends only on the topology of S . (Compare [110].)
- (2) The maps κ_n and $\hat{\kappa}_n$ have an orthogonality relationship: their derivatives have rank 1 in the same subset of X (the Euclidean part of the Thurston metric), and in this set, the collapsed directions of κ_n and $\hat{\kappa}_n$ are orthogonal. This orthogonality relationship is expressed in terms of their Hopf differentials as

$$\Phi(\kappa_n) + \Phi(\hat{\kappa}_n) = 0. \tag{6.2}$$

- (3) Let $h_n : X \rightarrow Y_n$ and $\hat{h}_n : \tilde{X} \rightarrow T_{\lambda_n}$ denote the harmonic maps homotopic to κ_n and $\hat{\kappa}_n$, respectively. Then the projective limit of Hopf differentials $[\Phi] = \lim_{n \rightarrow \infty} \Phi(h_n)$ satisfies $[\Lambda(-\Phi)] = [\mu]$ by Theorem 6.5. Similarly, by Theorem 6.4, the projective limit $[\hat{\Phi}] = \lim_{n \rightarrow \infty} \Phi(\hat{h}_n) = \lim_{n \rightarrow \infty} (-\phi_F(\lambda_n)/4)$ satisfies $[\Lambda(-\hat{\Phi})] = [\lambda]$.
- (4) Since the pair of almost harmonic maps κ_n and $\hat{\kappa}_n$ have opposite Hopf differentials, one might expect that the associated harmonic maps h_n and \hat{h}_n have “almost opposite” Hopf differentials. Suppose that this is true in the sense of projective limits, i.e. that

$$[\Phi] = [-\hat{\Phi}] \in \mathbb{P}^+ Q(X). \tag{6.3}$$

Then we would have $[\Lambda(\Phi)] = [\lambda]$ and $[\Lambda(-\Phi)] = [\mu]$, or equivalently, that $i_X([\lambda]) = [\mu]$, completing the proof. Thus we need only derive (6.3).

- (5) The norm of the difference between the pullback metric of a C -almost harmonic map f to an NPC space and that of its homotopic harmonic map h is $O(C^{1/2} \mathcal{E}(h)^{1/2})$ as $\mathcal{E}(h) \rightarrow \infty$ (by an estimate of Korevaar and Schoen, see [72, §2.6]). Phrasing this in terms of Hopf differentials, which are the (2, 0) parts of the pullback metrics, and using that $|\mathcal{E}(h) - 2\|\Phi(h)\|| = O(1)$, we have

$$\|\Phi(f) - \Phi(h)\|_1 \leq C'(1 + \|\Phi(h)\|_1^{\frac{1}{2}}).$$

In particular the norm of the difference is much smaller than either term as $\|\Phi(h)\| \rightarrow \infty$, and so the Hopf differentials of any sequence of C -almost harmonic maps with energy tending to infinity has the same projective limit as the

Hopf differentials of the harmonic maps. Applying this to the collapsing and co-collapsing maps, and using (6.2), we have

$$[\Phi] = \lim_{n \rightarrow \infty} \Phi(\kappa_n) = \lim_{n \rightarrow \infty} (-\Phi(\hat{\kappa}_n)) = [-\hat{\Phi}],$$

and (6.3) follows. □

6.4 Limits of the Schwarzian

We now connect the previous discussion of asymptotics of grafting coordinates for $P(X)$ with the complex-analytic parameterization of $\mathcal{P}(S)$. Let $\overline{P(X)}$ denote the Schwarzian compactification of $P(X)$ obtained by attaching $\mathbb{P}^+ Q(X)$ using the limiting behavior of the Schwarzian derivative, i.e. a sequence $Z_n \in P(X)$ converges to $[\phi]$ if $(Z_n - Z_0) \rightarrow [\phi]$ in the topology of $\overline{Q(X)}$. Here Z_0 denotes an arbitrary basepoint, which is used to identify $P(X)$ with $Q(X)$; the limit of a sequence in $\mathbb{P}^+ Q(X)$ does not depend on this choice. Note that this construction only compactifies the individual fibers of $\mathcal{P}(S)$, but does not compactify $\mathcal{P}(S)$ itself.

There is a natural guess for the relationship between the Schwarzian compactification and the closure of $P(X)$ in the grafting compactification: The boundary of the latter is the set of X -antipodal pairs in $\mathbb{P}\mathcal{ML}(S) \times \mathbb{P}\mathcal{ML}(S)$, and each X -antipodal pair arises from a ray in the space of quadratic differentials, so one might expect a boundary point $[\phi] \in \mathbb{P}^+ Q(X)$ to correspond to the pair consisting of its vertical and horizontal laminations. The following makes this intuition precise:

Theorem 6.6 (Dumas [26]). *The grafting and Schwarzian compactifications of $P(X)$ are naturally homeomorphic, and the boundary map $\mathbb{P}^+ Q(X) \rightarrow \mathbb{P}\mathcal{ML}(S) \times \mathbb{P}\mathcal{ML}(S)$ is given by*

$$[\phi] \mapsto ([\Lambda(-\phi)], [\Lambda(\phi)]).$$

That is, for a divergent sequence in $P(X)$, the limit of the vertical (resp. horizontal) laminations of Schwarzian differentials is equal to the limit of the measured laminations (resp. hyperbolic structures) in the grafting coordinates.

This result about compactifications involves a comparison between two homeomorphisms $\mathcal{ML}(S) \rightarrow Q(X)$. One of these we have already seen – the foliation map ϕ_F which sends $\lambda \in \mathcal{ML}(S)$ to a quadratic differential whose horizontal foliation has straightening λ (§6.2). The other homeomorphism is derived from the Schwarzian parameterization of projective structures as follows. Recall (from §4.4) that there is a homeomorphism $\sigma_\bullet(X) : \mathcal{ML}(S) \rightarrow P(X)$ with the property that $\sigma_\lambda(X) \in P(X)$ is a projective structure with grafting lamination λ . Using $\sigma_0(X)$ as a basepoint, we compose with the Schwarzian parameterization $P(X) \simeq Q(X)$ to obtain the Thurston map:

$$\begin{aligned} \phi_T : \mathcal{ML}(S) &\rightarrow Q(X) \\ \lambda &\mapsto (\sigma_\lambda(X) - \sigma_0(X)). \end{aligned}$$

The Thurston map is a homeomorphism, and it satisfies $\phi_T(0) = 0$, but unlike the foliation map there is no *a priori* reason for ϕ_T to map rays in $\mathcal{ML}(S)$ to rays in $Q(X)$. However, the Thurston map does preserve rays in an asymptotic sense:

Theorem 6.7 ([26]). *For any $X \in \mathcal{T}(S)$, the foliation and Thurston maps are asymptotically proportional. Specifically, there exists a constant $C(X)$ such that*

$$\|\phi_F(\lambda) + 2\phi_T(\lambda)\|_1 \leq C(X) \left(1 + \|\phi_F(\lambda)\|_1^{\frac{1}{2}}\right)$$

for all $\lambda \in \mathcal{ML}(S)$.

Before discussing the proof of Theorem 6.7, we explain the connection with compactifications. In terms of the Thurston map, Theorem 6.6 asserts that if $\phi_T(\lambda_n) = \text{Gr}_{\lambda_n} Y_n$ is a divergent sequence in $P(X)$, then we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \lambda_n &= \lim_{n \rightarrow \infty} \Lambda(-\phi_T(\lambda_n)) \in \overline{\mathcal{ML}(S)} \quad \text{and} \\ \lim_{n \rightarrow \infty} Y_n &= \lim_{n \rightarrow \infty} \Lambda(\phi_T(\lambda_n)) \in \overline{\mathcal{T}(S)}. \end{aligned} \tag{6.4}$$

Theorem 6.2 has already given a similar characterization in terms of the map ϕ_F ; we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \lambda_n &= \lim_{n \rightarrow \infty} \Lambda(\phi_F(\lambda_n)) \in \overline{\mathcal{ML}(S)} \quad \text{and} \\ \lim_{n \rightarrow \infty} Y_n &= \lim_{n \rightarrow \infty} \Lambda(-\phi_F(\lambda_n)) = \lim_{n \rightarrow \infty} i_X(\lambda_n) \in \overline{\mathcal{T}(S)}, \end{aligned} \tag{6.5}$$

where the first line is trivial since $\Lambda \circ \phi_F = \text{Id}$, and the second line follows from the definition of the antipodal map (§6.2). However, since ϕ_F and ϕ_T are asymptotically proportional by a negative constant (Theorem 6.7), the limit characterizations (6.4) and (6.5) are equivalent, and Theorem 6.6 follows. See [26, §14] for details.

Thurston metrics and the Schwarzian. We now sketch the main ideas involved in the proof of Theorem 6.7. The proof is essentially a study of the Thurston metric on a complex projective surface (see §4.3). Recall that the goal is to show that $\|\phi_F(\lambda) + 2\phi_T(\lambda)\|_1 \leq C(X)\varepsilon(\lambda)$ where $\varepsilon(\lambda)$ is defined by

$$\varepsilon(\lambda) = 1 + \|\phi_F(\lambda)\|_1^{\frac{1}{2}}.$$

Outline of proof of Theorem 6.7.

- (1) The functions $\varepsilon(\lambda)$ and $\lambda \mapsto \|\phi_F(\lambda) + 2\phi_T(\lambda)\|_1$ are continuous on $\mathcal{ML}(S)$. Since weighted simple closed geodesics are dense in $\mathcal{ML}(S)$, it suffices establish an inequality relating these functions for such weighted geodesics, and the general case follows by continuity. Thus we will assume λ is a weighted simple closed geodesic for the rest of the proof.
- (2) Associated to such λ we have the following objects:
 - The Thurston metric ρ_λ of the projective structure $\sigma_\lambda(X) \in P(X)$

- The decomposition $X = X_0 \sqcup X_{-1}$ of X into Euclidean and hyperbolic parts of ρ_λ . Here X_0 is an open cylinder, the union of the 1-dimensional strata in the canonical stratification.
 - The collapsing map $\kappa : X \rightarrow Y_\lambda = \text{gr}_\lambda^{-1}(X)$ and its Hopf differential $\Phi(\kappa)$, which is a measurable (non-holomorphic) quadratic differential supported on X_0 .
 - The ratio of conformally equivalent metrics ρ_λ/ρ_0 , a well-defined positive function on X . Here ρ_0 is the hyperbolic metric.
- (3) The Schwarzian derivative $\phi_T(\lambda)$ of the projective structure $\sigma_\lambda(X)$ decomposes as a sum of two terms,

$$\phi_T(\lambda) = -2\Phi(\kappa) + 2\mathcal{B}(\log(\rho_\lambda/\rho_0)), \quad (6.6)$$

where the second-order differential operator \mathcal{B} is defined by

$$\mathcal{B}(\eta) = [\text{Hess}(\eta) - d\eta \otimes d\eta]^{2,0}.$$

In this expression, the Hessian is computed using the hyperbolic metric ρ_0 . This decomposition follows from the cocycle property for a generalization of the Schwarzian derivative introduced by Osgood and Stowe [93].

- (4) The harmonic map estimate from the proof of Theorem 6.2 shows that the first term of the decomposition (6.6) is approximately proportional to $\phi_F(\lambda)$. Specifically, we have

$$\|\phi_F(\lambda) - 4\Phi(\kappa)\|_1 \leq C\varepsilon(\lambda). \quad (6.7)$$

Therefore it suffices to show that the L^1 norm of $\beta = \mathcal{B}(\log(\rho_\lambda/\rho_0))$ is also bounded by a multiple of $\varepsilon(\lambda)$.

- (5) By the definition of \mathcal{B} and the Cauchy–Schwartz inequality, the L^1 norm of β is bounded by the L^2 norms of the Hessian and gradient of $\log(\rho_\lambda/\rho_0)$ with respect to the hyperbolic metric. By standard elliptic theory, these are in turn bounded by the L^2 norms of $\log(\rho_\lambda/\rho_0)$ and its Laplacian.
- (6) The Laplacian of $\log(\rho_\lambda/\rho_0)$ is essentially the difference of the curvature 2-forms of ρ_λ and ρ_0 (compare (4.1) above, also [55]). For large grafting, the surface X is dominated by its Euclidean part, forcing most of the curvature of ρ_λ to concentrate near a finite set of points.
- (7) This curvature concentration phenomenon provides a bound for the norm $\|\Delta \log(\rho_\lambda/\rho_0)\|_{L^2(D)}$ on a hyperbolic disk $D \subset X$ of definite size. A bound on $\|\log(\rho_\lambda/\rho_0)\|_{L^2(D)}$ follows using a weak Harnack inequality, completing the local estimate $\|\beta\|_{L^1(D)} < C(X)$.
- (8) Finally, we make the local estimate global: If β were holomorphic, then we would have $\|\beta\|_{L^1(X)} \leq C'(X)\|\beta\|_{L^1(D)}$ by compactness of the unit sphere in $Q(X)$. While β is not holomorphic, the decomposition (6.6) and the estimate (6.7) show that β is close to a holomorphic quadratic differential, with difference

of order $\varepsilon(\lambda)$. Combining this with the holomorphic case, we obtain $\|\beta\|_{L^1(X)} \leq C(X)\varepsilon(\lambda)$, completing the proof. \square

6.5 Infinitesimal compatibility

In this final section we discuss infinitesimal aspects of the map between the grafting and analytic coordinate systems for $\mathcal{P}(S)$.

The forgetful projection $\pi: \mathcal{P}(S) \rightarrow \mathcal{T}(S)$ can be thought of as a coordinate function in the Schwarzian parameterization of $\mathcal{P}(S)$. The other “coordinate” in this parameterization is an element of the fiber $Q(X)$ of the bundle of quadratic differentials, but lacking a canonical trivialization for this bundle, there is no associated global coordinate map.

On the other hand, in the grafting coordinate system, we have a pair of well-defined coordinate maps $p_{\mathcal{ML}}: \mathcal{P}(S) \rightarrow \mathcal{ML}(S)$ and $p_{\mathcal{T}}: \mathcal{P}(S) \rightarrow \mathcal{T}(S)$, which are defined by the property that the inverse of projective grafting is $\text{Gr}^{-1}(Z) = (p_{\mathcal{ML}}(Z), p_{\mathcal{T}}(Z)) \in \mathcal{ML}(S) \times \mathcal{T}(S)$.

The fiber of $p_{\mathcal{ML}}$ over λ consists of the projective structures $\{\text{Gr}_\lambda Y \mid Y \in \mathcal{T}(S)\}$. Since $\text{Gr}_\lambda: \mathcal{T}(S) \rightarrow \mathcal{P}(S)$ is a smooth map, these fibers are smooth submanifolds of $\mathcal{P}(S)$.

The fiber of $p_{\mathcal{T}}$ over Y consists of the projective structures $\{\text{Gr}_\lambda Y \mid \lambda \in \mathcal{ML}(S)\}$. Bonahon showed that $\lambda \mapsto \text{Gr}_\lambda Y$ includes $\mathcal{ML}(S)$ into $\mathcal{P}(S)$ tangentially (see Theorem 4.5). However, the fibers of $p_{\mathcal{T}}$ have even more regularity than one might expect from this tangential parameterization:

Theorem 6.8 (Bonahon [10, Theorem 3, Lemma 13]). *For each $Y \in \mathcal{T}(S)$, the set $p_{\mathcal{T}}^{-1}(Y)$ is a C^1 submanifold of $\mathcal{P}(S)$.*

Compare [27, §4].

Note that each of the three coordinate maps π , $p_{\mathcal{ML}}$, $p_{\mathcal{T}}$ projects $\mathcal{P}(S)$ onto a space of half its real dimension, i.e. each has both range and fibers of real dimension $6g - 6$. Thus one might expect that for any two of these maps, the pair of fibers intersecting at a generic point $Z \in \mathcal{P}(S)$ would have transverse tangent spaces that span $T_Z \mathcal{P}(S)$. In fact, this is true at every point, and furthermore we have:

Theorem 6.9 (Dumas and Wolf [27]). (1) *The maps π , $p_{\mathcal{ML}}$, $p_{\mathcal{T}}$ have pairwise transverse fibers.*

(2) *The fiber of any one of them projects homeomorphically by each of the others. Moreover, such a projection is a C^1 diffeomorphism whenever its range is $\mathcal{T}(S)$, and is a bitangentible homeomorphism when the range is $\mathcal{ML}(S)$.*

(3) *The product of any two of these maps gives a homeomorphism from $\mathcal{P}(S)$ to a product of two spaces of real dimension $6g - 6$.*

As before, we refer to Bonahon (see [10, §2]) for details about tangentiality, while limiting our focus to its geometric consequences. Also note that statement (1)

of the theorem does *not* involve tangentiability, and only makes sense for fibers of $p_{\mathcal{T}}$ due to Theorem 6.8.

We sketch the proof of this theorem; the details we omit can be found in [27, Theorems 1.2, 4.1, 4.2, Corollary 4.3].

Sketch of proof of Theorem 6.9. Statement (3) follows because the inverse map for each pair of coordinates can be written explicitly in terms of Gr , gr_{λ} , and $\text{gr}_{\bullet}X$ and their inverses (which exist by Theorems 4.1, 4.3, and 4.4, respectively). For example, $p_{\mathcal{T}} \times \pi : \mathcal{P}(S) \rightarrow \mathcal{T}(S) \times \mathcal{T}(S)$ is a homeomorphism with inverse

$$(X, Y) \mapsto \text{Gr}_{(\text{gr}_{\bullet}X)^{-1}(Y)}X.$$

Similarly, the map $(\lambda, X) \mapsto \sigma_{\lambda}(X)$ is inverse to $p_{\mathcal{M}\mathcal{L}} \times \pi$.

Statement (3) also shows that the restrictions of maps considered in statement (2) are homeomorphisms. To show that each case with target $\mathcal{T}(S)$ is actually a diffeomorphism, it is enough to show that the derivative of the restriction has no kernel (by the inverse function theorem). This kernel is the intersection of tangent spaces to fibers of two coordinate maps, thus this case will follow from statement (1). Similar reasoning applies in cases with target $\mathcal{M}\mathcal{L}(S)$, where one deduces bitangentiability from transversality using a criterion of Bonahon [10, Lemma 4].

Thus the proof is reduced to the transversality statement (1), which has one case for each pair of coordinate maps. The pair $(p_{\mathcal{M}\mathcal{L}}, p_{\mathcal{T}})$ follows easily from Thurston's theorem and the tangentiability of grafting (Theorems 4.1 and 4.5). For $(\pi, p_{\mathcal{M}\mathcal{L}})$ or $(\pi, p_{\mathcal{T}})$, a vector in the intersection of tangent spaces lies in the kernel of a tangent map of either gr_{λ} or $\text{gr}_{\bullet}X$, which must therefore be zero, by Theorems 4.3 and 4.4. \square

References

- [1] L. Ahlfors and L. Bers, Riemann's mapping theorem for variable metrics. *Ann. of Math.* (2) 72 (1960), 385–404. [466](#)
- [2] C. Anderson, Projective Structures on Riemann Surfaces and Developing Maps to \mathbb{H}^3 and $\mathbb{C}\mathbb{P}^n$. PhD Thesis, University of California at Berkeley, 1998. [462](#), [463](#), [464](#), [489](#)
- [3] J. Anderson and R. Canary, Algebraic limits of Kleinian groups which rearrange the pages of a book. *Invent. Math.* 126(2) (1996), 205–214. [487](#)
- [4] J. Anderson, R. Canary, and D. McCullough, The topology of deformation spaces of Kleinian groups. *Ann. of Math.* (2) 152 (3) (2000), 693–741. [487](#)
- [5] P. Appell, É. Goursat, and P. Fatou, *Théorie des Fonctions Algébriques, Tome 2: Fonctions Automorphes*. Gauthier-Villars, Paris 1930. [479](#), [489](#)
- [6] L. Bers, Fiber spaces over Teichmüller spaces. *Acta. Math.* 130 (1973), 89–126. [466](#)
- [7] L. Bers, Holomorphic families of isomorphisms of Möbius groups. *J. Math. Kyoto Univ.* 26 (1) (1986), 73–76. [488](#)

- [8] F. Bonahon, The geometry of Teichmüller space via geodesic currents. *Invent. Math.* 92 (1) (1988), 139–162. [492](#)
- [9] F. Bonahon, Shearing hyperbolic surfaces, bending pleated surfaces and Thurston’s symplectic form. *Ann. Fac. Sci. Toulouse Math.* (6) 5 (2) (1996), 233–297. [477](#)
- [10] F. Bonahon, Variations of the boundary geometry of 3-dimensional hyperbolic convex cores. *J. Differential Geom.* 50 (1) (1998), 1–24. [457](#), [477](#), [501](#), [502](#)
- [11] J. Brock and K. Bromberg, On the density of geometrically finite Kleinian groups. *Acta Math.* 192 (1) (2004), 33–93. [468](#)
- [12] J. Brock, K. Bromberg, R. Canary, and Y. Minsky, in preparation. [488](#)
- [13] K. Bromberg, Projective structures with degenerate holonomy and the Bers density conjecture. *Ann. of Math.* (2) 166 (1) (2007), 77–93. [468](#), [488](#)
- [14] K. Bromberg and J. Holt, Bumping of exotic projective structures. Preprint. [487](#)
- [15] K. Bromberg and J. Holt, Self-bumping of deformation spaces of hyperbolic 3-manifolds. *J. Differential Geom.* 57 (1) (2001), 47–65. [487](#)
- [16] R. Canary, Introductory bumponomics: The topology of deformation spaces of hyperbolic 3-manifolds. Preprint, 2007. [487](#)
- [17] R. Canary, D. Epstein, and P. Green, Notes on notes of Thurston. In *Analytical and geometric aspects of hyperbolic space* (Coventry/Durham, 1984), London Math. Soc. Lecture Note Ser. 111, Cambridge University Press, Cambridge 1987, 3–92. [488](#)
- [18] A. Casson and S. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*. London Math. Soc. Student Texts 9, Cambridge University Press, Cambridge 1988. [469](#)
- [19] V. Chuckrow, On Schottky groups with applications to Kleinian groups. *Ann. of Math.* (2) 88 (1968), 47–61. [488](#)
- [20] M. Culler and P. Shalen, Varieties of group representations and splittings of 3-manifolds. *Ann. of Math.* (2) 117 (1) (1983), 109–146. [478](#)
- [21] G. Daskalopoulos, S. Dostoglou, and R. Wentworth, Character varieties and harmonic maps to \mathbf{R} -trees. *Math. Res. Lett.* 5 (4) (1998), 523–533. [496](#)
- [22] G. Daskalopoulos and R. Wentworth, Harmonic maps and Teichmüller theory. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 33–110. [495](#)
- [23] S. Donaldson, Twisted harmonic maps and the self-duality equations. *Proc. London Math. Soc.* (3) 55 (1) (1987), 127–131. [495](#)
- [24] D. Dumas, Bear: A tool for studying Bers slices of punctured tori. Free software, available for download from <http://bear.sourceforge.net/>. [490](#)
- [25] D. Dumas, Grafting, pruning, and the antipodal map on measured laminations. *J. Differential Geometry* 74 (2006), 93–118; Erratum. *ibid.* 77 (2007), 175–176. [471](#), [494](#), [495](#), [497](#)
- [26] D. Dumas, The Schwarzian derivative and measured laminations on Riemann surfaces. *Duke Math. J.* 140 (2) (2007), 203–243. [476](#), [489](#), [498](#), [499](#)
- [27] D. Dumas and M. Wolf, Projective structures, grafting, and measured laminations. *Geom. Topol.* 12 (1) (2008), 351–386. [476](#), [477](#), [501](#), [502](#)

- [28] C. Earle, On variation of projective structures. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference* (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 87–99. [456](#), [466](#), [477](#), [479](#)
- [29] J. Eells and L. Lemaire, A report on harmonic maps. *Bull. London Math. Soc.* 10 (1) (1978), 1–68. [495](#), [496](#)
- [30] J. Eells and L. Lemaire, Another report on harmonic maps. *Bull. London Math. Soc.* 20 (5) (1988), 385–524. [495](#)
- [31] J. Eells and J. Sampson, Harmonic mappings of Riemannian manifolds. *Amer. J. Math.* 86 (1964), 109–160. [495](#)
- [32] D. Epstein and A. Marden, Convex hulls in hyperbolic space, a theorem of Sullivan, and measured pleated surfaces. In *Analytical and geometric aspects of hyperbolic space* (Coventry/Durham, 1984), London Math. Soc. Lecture Note Ser. 111, Cambridge University Press, Cambridge 1987, 113–253. [470](#), [481](#), [482](#), [483](#)
- [33] G. Faltings, Real projective structures on Riemann surfaces. *Compositio Math.* 48 (2) (1983), 223–269. [485](#)
- [34] A. Fathi, F. Laudenbach and V. Poénaru (eds.), *Travaux de Thurston sur les surfaces*. *Astérisque* 66–67 (1979). [491](#), [492](#)
- [35] E. Frenkel and D. Ben-Zvi, *Vertex algebras and algebraic curves*. Second edition, Math. Surveys Monogr. 88, Amer. Math. Soc., Providence, RI, 2004. [464](#)
- [36] D. Gallo, M. Kapovich, and A. Marden, The monodromy groups of Schwarzian equations on closed Riemann surfaces. *Ann. of Math.* (2) 151 (2) (2000), 625–704. [457](#), [477](#), [479](#), [480](#), [489](#)
- [37] F. Gardiner, *Teichmüller theory and quadratic differentials*. Pure Appl. Math., John Wiley & Sons Inc., New York 1987. [492](#)
- [38] W. Goldman, The symplectic nature of fundamental groups of surfaces. *Adv. in Math.* 54 (2) (1984), 200–225. [478](#)
- [39] W. Goldman, Projective structures with Fuchsian holonomy. *J. Differential Geom.* 25 (3) (1987), 297–326. [484](#), [485](#), [487](#)
- [40] W. Goldman, Geometric structures on manifolds and varieties of representations. In *Geometry of group representations* (Boulder, CO, 1987), Contemp. Math. 74, Amer. Math. Soc., Providence, RI, 1988, 169–198. [460](#), [464](#), [479](#), [485](#)
- [41] W. Goldman, Topological components of spaces of representations. *Invent. Math.* 93 (3) (1988), 557–607. [478](#)
- [42] W. Goldman, Ergodic theory on moduli spaces. *Ann. of Math.* (2) 146 (3) (1997), 475–507. [481](#)
- [43] R. Gunning, *Lectures on Riemann surfaces*. Princeton Mathematical Notes, Princeton University Press, Princeton, N.J., 1966. [456](#), [458](#), [479](#)
- [44] R. Gunning, *Lectures on vector bundles over Riemann surfaces*. University of Tokyo Press, Tokyo 1967. [478](#)
- [45] R. Gunning, Special coordinate coverings of Riemann surfaces. *Math. Ann.* 170 (1967), 67–86. [456](#), [459](#), [464](#)

- [46] R. Gunning, Affine and projective structures on Riemann surfaces. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference* (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 225–244. [477](#), [478](#), [479](#)
- [47] D. Hejhal, Monodromy groups and linearly polymorphic functions. *Acta Math.* 135 (1) (1975), 1–55. [456](#), [459](#), [468](#), [477](#), [479](#), [484](#), [489](#)
- [48] D. Hejhal, Monodromy groups and Poincaré series. *Bull. Amer. Math. Soc.* 84 (3) (1978), 339–376. [456](#)
- [49] M. Heusener and J. Porti, The variety of characters in $PSL_2(\mathbb{C})$. *Bol. Soc. Mat. Mexicana* (3) 10 (Special Issue) (2004), 221–237. [478](#)
- [50] J. Holt, Some new behaviour in the deformation theory of Kleinian groups. *Comm. Anal. Geom.* 9 (4) (2001), 757–775. [487](#)
- [51] J. Hubbard, The monodromy of projective structures. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference* (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 257–275. [456](#), [464](#), [465](#), [466](#), [477](#), [479](#)
- [52] J. Hubbard, *Teichmüller theory and applications to geometry, topology, and dynamics*. Vol. 1, Matrix Editions, Ithaca, NY, 2006. [461](#), [462](#), [464](#)
- [53] J. Hubbard and H. Masur, Quadratic differentials and foliations. *Acta Math.* 142 (3–4) (1979), 221–274. [493](#)
- [54] A. Huber, On subharmonic functions and differential geometry in the large. *Comment. Math. Helv.* 32 (1957), 13–72. [475](#)
- [55] A. Huber, Zum potentialtheoretischen Aspekt der Alexandrowschen Flächentheorie. *Comment. Math. Helv.* 34 (1960), 99–126. [475](#), [500](#)
- [56] Y. Iwayoshi and M. Taniguchi, *An introduction to Teichmüller spaces*. Springer-Verlag, Tokyo 1992. [461](#), [464](#), [465](#)
- [57] K. Ito, Exotic projective structures and quasi-Fuchsian space. *Duke Math. J.* 105 (2) (2000), 185–209. [487](#)
- [58] K. Ito, Schottky groups and Bers boundary of Teichmüller space. *Osaka J. Math.* 40 (3) (2003), 639–657. [488](#)
- [59] K. Ito, Grafting and components of quasi-Fuchsian projective structures. In *Spaces of Kleinian groups*, London Math. Soc. Lecture Note Ser. 329, Cambridge University Press, Cambridge 2006, 355–373. [479](#)
- [60] K. Ito, Exotic projective structures and quasi-Fuchsian space, II. *Duke Math. J.* 140 (1) (2007), 85–109. [487](#)
- [61] T. Jørgensen, On discrete groups of Möbius transformations. *Amer. J. Math.* 98 (3) (1976), 739–749. [488](#)
- [62] J. Jost, *Compact Riemann surfaces*. Third edition, Universitext, Springer-Verlag, Berlin 2006. [463](#)
- [63] Y. Kamishima, Conformally flat manifolds whose development maps are not surjective. I. *Trans. Amer. Math. Soc.* 294 (2) (1986), 607–623. [479](#)
- [64] Y. Kamishima and S. Tan, Deformation spaces on geometric structures. In *Aspects of low-dimensional manifolds*, Adv. Stud. Pure Math. 20, Kinokuniya, Tokyo 1992, 263–299. [456](#), [469](#), [471](#)

- [65] M. Kapovich, On monodromy of complex projective structures. *Invent. Math.* 119 (2) (1995), 243–265. [489](#)
- [66] M. Kapovich, *Hyperbolic manifolds and discrete groups*. Progr. Math. 183, Birkhäuser, Boston, MA, 2001. [468](#), [479](#), [487](#), [491](#), [492](#)
- [67] F. Klein, *Ausgewählte Kapitel aus der Theorie der linearen Differentialgleichungen zweiter Ordnung*. Volume 1, Göttingen 1891. [456](#)
- [68] F. Klein, *Vorlesungen über die Hypergeometrische Funktion*. Grundlehren Math. Wiss. 39, Springer-Verlag, Berlin 1933. [456](#)
- [69] S. Kojima, Circle packing and Teichmüller spaces. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 509–531. [457](#)
- [70] S. Kojima, S. Mizushima, and S. Tan, Circle packings on surfaces with projective structures: a survey. In *Spaces of Kleinian groups*, London Math. Soc. Lecture Note Ser. 329, Cambridge University Press, Cambridge 2006, 337–353. [457](#)
- [71] N. Korevaar and R. Schoen, Global existence theorems for harmonic maps: finite rank spaces and an approach to rigidity for smooth actions. Preprint. [495](#)
- [72] N. Korevaar and R. Schoen, Sobolev spaces and harmonic maps for metric space targets. *Comm. Anal. Geom.* 1 (3–4) (1993), 561–659. [495](#), [496](#), [497](#)
- [73] N. Korevaar and R. Schoen, Global existence theorems for harmonic maps to non-locally compact spaces. *Comm. Anal. Geom.* 5 (2) (1997), 333–387. [495](#), [496](#)
- [74] I. Kra, Deformations of Fuchsian groups. *Duke Math. J.* 36 (1969), 537–546. [456](#), [466](#), [480](#)
- [75] I. Kra, Deformations of Fuchsian groups. II. *Duke Math. J.* 38 (1971), 499–508. [456](#), [457](#), [480](#)
- [76] I. Kra, A generalization of a theorem of Poincaré. *Proc. Amer. Math. Soc.* 27 (1971), 299–302. [456](#), [489](#)
- [77] I. Kra and B. Maskit, Remarks on projective structures. In *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference* (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 343–359. [456](#), [466](#)
- [78] R. Kulkarni and U. Pinkall, A canonical metric for Möbius structures and its applications. *Math. Z.* 216 (1) (1994), 89–129. [471](#), [474](#)
- [79] O. Lehto, *Univalent functions and Teichmüller spaces*, Grad. Texts in Math. 109, Springer-Verlag, New York 1987. [461](#), [462](#), [466](#)
- [80] G. Levitt, Foliations and laminations on hyperbolic surfaces. *Topology* 22 (2) (1983), 119–135. [491](#), [493](#)
- [81] F. Luo, Monodromy groups of projective structures on punctured surfaces. *Invent. Math.* 111 (3) (1993), 541–555. [457](#)
- [82] A. Marden, *Outer Circles: An Introduction to Hyperbolic 3-Manifolds*. Cambridge University Press, Cambridge 2007. [489](#), [492](#)
- [83] B. Maskit, On a class of Kleinian groups. *Ann. Acad. Sci. Fenn. Ser. A I* 442 (1969), 8p. [456](#), [468](#), [484](#)

- [84] K. Matsuzaki, The interior of discrete projective structures in the Bers fiber. *Ann. Acad. Sci. Fenn. Math.* 32 (1) (2007), 3–12. [490](#), [491](#)
- [85] K. Matsuzaki and M. Taniguchi, *Hyperbolic manifolds and Kleinian groups*. Oxford Math. Monogr., The Clarendon Press Oxford University Press, New York 1998. [478](#)
- [86] C. McMullen, Complex earthquakes and Teichmüller theory. *J. Amer. Math. Soc.* 11 (2) (1998), 283–320. [482](#), [483](#), [487](#), [488](#)
- [87] C. Mese, Harmonic maps between surfaces and Teichmüller spaces. *Amer. J. Math.* 124 (3) (2002), 451–481. [475](#)
- [88] C. Mese, Harmonic maps into spaces with an upper curvature bound in the sense of Alexandrov. *Math. Z.* 242 (4) (2002), 633–661. [496](#)
- [89] C. Mese, Uniqueness theorems for harmonic maps into metric spaces. *Commun. Contemp. Math.* 4 (4) (2002), 725–750. [496](#)
- [90] J. Morgan and P. Shalen, Valuations, trees, and degenerations of hyperbolic structures. I. *Ann. of Math.* (2) 120 (3) (1984), 401–476. [471](#)
- [91] Z. Nehari, The Schwarzian derivative and schlicht functions. *Bull. Amer. Math. Soc.* 55 (1949), 545–551. [463](#), [466](#), [489](#)
- [92] K. Ohshika, Divergence, exotic convergence, and self-bumping in quasi-Fuchsian spaces. In preparation. [488](#)
- [93] B. Osgood and D. Stowe, The Schwarzian derivative and conformal mapping of Riemannian manifolds. *Duke Math. J.* 67 (1) (1992), 57–99. [500](#)
- [94] R. Penner and J. Harer, *Combinatorics of train tracks*. Ann. of Math. Stud. 125, Princeton University Press, Princeton, NJ, 1992. [469](#), [491](#)
- [95] H. Poincaré, Sur les groupes des équations linéaires. *Acta Math.* 4 (1) (1884), 201–312. [456](#)
- [96] C. Pommerenke, *Univalent functions*. Studia Mathematica/Mathematische Lehrbücher XXV, Vandenhoeck & Ruprecht, Göttingen 1975. [466](#)
- [97] A. Rapinchuk, V. Benyash-Krivetz, and V. Chernousov, Representation varieties of the fundamental groups of compact orientable surfaces. *Israel J. Math.* 93 (1996), 29–71. [478](#)
- [98] Y. Reshetnyak, Two-dimensional manifolds of bounded curvature. In *Geometry. IV: Nonregular Riemannian geometry*, Encyclopaedia Math. Sci., Springer-Verlag, Berlin 1993, 3–163; translation of *Geometry, 4*, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow 1989. [475](#)
- [99] Y. Reshetnyak, On the conformal representation of Alexandrov surfaces. In *Papers on analysis*, Rep. Univ. Jyväskylä Dep. Math. Stat. 83, Univ. Jyväskylä, Jyväskylä 2001, 287–304. [475](#)
- [100] B. Riemann, Vorlesungen über die hypergeometrische Reihe. In *Gesammelte mathematische Werke, wissenschaftlicher Nachlass und Nachträge*, Springer-Verlag, Berlin 1990, 667–692. [456](#)
- [101] K. Scannell, Flat conformal structures and causality in de Sitter manifolds. PhD Thesis, University of California, Los Angeles, 1996. [471](#)
- [102] K. Scannell, Flat conformal structures and the classification of de Sitter manifolds. *Comm. Anal. Geom.* 7 (2) (1999), 325–345. [471](#)

- [103] K. Scannell and M. Wolf, The grafting map of Teichmüller space. *J. Amer. Math. Soc.* 15 (4) (2002), 893–927. [476](#), [488](#)
- [104] R. Schoen, Analytic aspects of the harmonic map problem. In *Seminar on nonlinear partial differential equations* (Berkeley, Calif., 1983), Math. Sci. Res. Inst. Publ. 2, Springer-Verlag, New York 1984, 321–358. [495](#)
- [105] R. Sharpe, *Differential geometry. Cartan's generalization of Klein's Erlangen program*. Grad. Texts in Math. 166, Springer-Verlag, New York 1997. [462](#)
- [106] H. Shiga, Projective structures on Riemann surfaces and Kleinian groups. *J. Math. Kyoto Univ.* 27 (3) (1987), 433–438. [489](#)
- [107] H. Shiga and H. Tanigawa, Projective structures with discrete holonomy representations. *Trans. Amer. Math. Soc.* 351 (2) (1999), 813–823. [487](#), [490](#)
- [108] D. Sullivan, Quasiconformal homeomorphisms and dynamics. II. Structural stability implies hyperbolicity for Kleinian groups. *Acta Math.* 155 (3–4) (1985), 243–260. [488](#)
- [109] D. Sullivan and W. Thurston, Manifolds with canonical coordinate charts: some examples. *Enseign. Math.* (2) 29 (1–2) (1983), 15–25. [456](#), [468](#), [484](#)
- [110] H. Tanigawa, Grafting, harmonic maps and projective structures on surfaces. *J. Differential Geom.* 47 (3) (1997), 399–419. [473](#), [475](#), [476](#), [497](#)
- [111] H. Tanigawa, Divergence of projective structures and lengths of measured laminations. *Duke Math. J.* 98 (2) (1999), 209–215. [489](#)
- [112] W. Thurston, Hyperbolic Structures on 3-manifolds, II: Surface groups and 3-manifolds which fiber over the circle. Preprint. [492](#)
- [113] W. Thurston, Geometry and topology of three-manifolds. Princeton Lecture Notes, 1979. [469](#)
- [114] W. Thurston, Earthquakes in two-dimensional hyperbolic geometry. In *Low-dimensional topology and Kleinian groups* (Coventry/Durham, 1984), London Math. Soc. Lecture Note Ser. 112, Cambridge University Press, Cambridge 1986, 91–112. [482](#)
- [115] W. Thurston, Minimal stretch maps between hyperbolic surfaces. Unpublished preprint, 1986. [457](#), [477](#)
- [116] W. Thurston, Zippers and univalent functions. In *The Bieberbach conjecture* (West Lafayette, Ind., 1985), Math. Surveys Monogr. 21, Amer. Math. Soc., Providence, RI, 1986, 185–197. [462](#)
- [117] W. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)* 19 (2) (1988), 417–431. [492](#)
- [118] M. Wolf, The Teichmüller theory of harmonic maps. *J. Differential Geom.* 29 (2) (1989), 449–479. [496](#)
- [119] M. Wolf, On realizing measured foliations via quadratic differentials of harmonic maps to \mathbf{R} -trees. *J. Anal. Math.* 68 (1996), 107–120. [496](#)
- [120] D. Wright, The shape of the boundary of the Teichmüller space of once-punctured tori in Maskit's embedding. Unpublished preprint, 1987. [460](#)

Chapter 13

Circle packing and Teichmüller space

Sadayoshi Kojima

Contents

1	Introduction	509
2	Circle packing	511
	2.1 Circle on the Riemann sphere	511
	2.2 Projective geometry	511
	2.3 Projective Riemann surface	512
	2.4 Circle packing on surfaces	515
3	Rigidity	517
	3.1 Rigidity theorems	517
	3.2 Unified proof	519
	3.3 Density	521
4	Flexibility	524
	4.1 Constructing moduli	524
	4.2 Thurston coordinates	527
	4.3 Speculation	528
	4.4 Evidence	529
	References	530

1 Introduction

A circle in the complex plane \mathbb{C} is defined as either a euclidean circle or a euclidean straight line. This definition is very understandable if we look at a circle on the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. The Riemann sphere can be identified with the complex projective line, and every projective transformation on $\hat{\mathbb{C}}$ through this identification sends a circle to a circle. Conversely, a transformation of $\hat{\mathbb{C}}$ which sends a circle to a circle turns out to be projective. Thus 1-dimensional complex projective geometry fits well with the concept of a circle.

To globalize the notion of circle, consider a surface S , that is, a real 2-dimensional manifold, locally modeled on $\hat{\mathbb{C}}$ such that any coordinate change is the restriction of a projective transformation. Such a geometric structure is referred to as a projective structure, and we call a surface with a projective structure simply a projective Riemann surface. A 1-dimensional subset of S is said to be a circle if its developed image on $\hat{\mathbb{C}}$ is a circle.

Now, we are interested in a circle packing on a projective Riemann surface, which will be a particular configuration of circles such that all complementary regions are curvilinear triangles. It enjoys both rigid and flexible properties in connection with Teichmüller spaces. The main purpose of this chapter is to discuss such interesting properties from geometric viewpoints.

Recall that there are three typical Riemannian geometries in dimension 2, that is, the spherical, euclidean and hyperbolic geometries. They are geometries with constant curvature 1, 0 and -1 respectively, and they are regarded as subgeometries of the 1-dimensional complex projective geometry. In particular, a constant curvature surface is a projective Riemann surface.

The rigidity we discuss here has been worked out by Koebe [9], Andreev [1] and Thurston [18] for realization of a circle packing on a constant curvature surface. More specifically, we describe in a rather uniform way how a combinatorial adjacency data of circles determines uniquely a constant curvature surface which supports a geometric packing with prescribed data.

This rigidity motivated Brooks [3], [4] to analyze the flexibility phenomenon when we allow curvilinear quadrilateral complementary regions. He succeeded to parametrize the deformation space in terms of continued fractional type numerical invariants, and deduced the density of packable constant curvature surfaces in Teichmüller space. We discuss Brooks' idea briefly, and see how his parameters work through quasi-conformal deformation theory.

On the other hand, extending the problem Koebe–Andreev–Thurston settled on constant curvature surfaces, one may ask what the set of projective Riemann surfaces supporting a circle packing with a common combinatorial data looks like. It leads us to analyze flexibility of the object in question. Following [10], [11], [12], we present here a construction of the moduli space of pairs of such projective Riemann surfaces with circle packings, and see its basic properties. In particular, we discuss our belief that the moduli space provides a sort of uniformization in terms of circle packing. We state it as a conjecture in more explicit form, and report some progress towards it.

The organization of this chapter is as follows. After reviewing the basics of the subject in the next section, we discuss rigidity results together with density on constant curvature surfaces due to Koebe, Andreev, Thurston and Brooks. We then discuss flexibility by constructing moduli spaces, and formulate a conjecture along with some supporting evidence.

Acknowledgement. This work was partially supported by JSPS Grant A No.15204004.

2 Circle packing

2.1 Circle on the Riemann sphere

The complex projective line, which is the space of complex lines through the origin in \mathbb{C}^2 , can be identified with the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ by assigning slopes. Then, $\hat{\mathbb{C}}$ can be identified by stereographic projection with the unit sphere \mathbb{S}^2 in the 3-dimensional euclidean space \mathbb{E}^3 . These are the main playgrounds on which the circle is placed. Note that $\hat{\mathbb{C}}$ admits a natural orientation coming from its complex structure.

Given four complex numbers a, b, c, d such that $ad - bc \neq 0$, we obtain a linear fractional transformation,

$$z \mapsto \frac{az + b}{cz + d},$$

which acts on $\hat{\mathbb{C}}$ as a projective transformation. Such projective transformations form a group isomorphic to $\text{PGL}(2, \mathbb{C}) \cong \text{PSL}(2, \mathbb{C})$ by identifying a transformation with a matrix consisting of these four numbers. The action of the projective linear group preserves the orientation.

Another important transformation on $\hat{\mathbb{C}}$ is complex conjugation,

$$z \mapsto \bar{z},$$

which fixes $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ and reverses the orientation. Complex conjugation and $\text{PGL}(2, \mathbb{C})$ generate a group Möb of Möbius transformations which fits into a split short exact sequence,

$$1 \longrightarrow \text{PGL}(2, \mathbb{C}) \longrightarrow \text{Möb} \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow 0.$$

A circle is, by definition, the image of $\hat{\mathbb{R}}$ by a projective transformation. Through the identification of $\hat{\mathbb{C}}$ with \mathbb{S}^2 , a circle on $\hat{\mathbb{C}}$ projects to either a straight line or a circle in the usual sense in \mathbb{C} .

A circle in $\hat{\mathbb{C}} = \mathbb{S}^2$ can be defined also as a metric circle with respect to the spherical metric. Note however that a projective transformation which sends a circle to a circle does not preserve the spherical metric in general.

2.2 Projective geometry

The pair $(\text{PGL}(2, \mathbb{C}), \hat{\mathbb{C}})$ of the 2×2 projective linear group and the Riemann sphere is called the 1-dimensional complex projective geometry in the spirit of Felix Klein's Erlangen program. The projective geometry contains three typical 2-dimensional geometries as subgeometry.

The unitary group $U(2)$ in $GL(2, \mathbb{C})$ becomes $PU(2) \subset \text{PGL}(2, \mathbb{C})$ in the quotient. It is isomorphic to $SO(3)$, and the pair $(PU(2), \hat{\mathbb{C}}) = (SO(3), \mathbb{S}^2)$ is the spherical geometry.

The upper triangular group $UT(2)$ in $GL(2, \mathbb{C})$ yields $PUT(2) \subset PGL(2, \mathbb{C})$ whose action on $\hat{\mathbb{C}}$ fixes $\{\infty\}$. It is isomorphic to the 1-dimensional complex affine transformation group $A(1)$, and the pair $(PUT(2), \mathbb{C}) = (A(1), \mathbb{C})$ is the complex affine geometry. Under the canonical identification of \mathbb{C} with the euclidean plane \mathbb{E}^2 , the orientation preserving euclidean isometry group $Isom_+ \mathbb{E}^2$ can be embedded in $A(1)$ and the pair $(Isom_+ \mathbb{E}^2, \mathbb{E}^2) \subset (A(1), \mathbb{C})$ is the euclidean geometry.

The action of the subgroup $U(1, 1)$ of $GL(2, \mathbb{C})$, which preserves a Hermitian form of signature $(1, 1)$, leaves the unit disk \mathbb{D} of \mathbb{C} invariant. The projectivisation defines the hyperbolic geometry $(PU(1, 1), \mathbb{D})$, where the action of $PU(1, 1)$ preserves the orientation.

The spherical, euclidean and hyperbolic geometries have compact stabilizers and each admits a Riemannian metric of constant curvature, say $1, 0, -1$ respectively, invariant under the action of transformation groups. A circle on $\hat{\mathbb{C}}$ contained in the domain of these three geometries will be a metric circle in their own metrics.

2.3 Projective Riemann surface

Let Σ_g be a compact oriented surface of genus $g \geq 0$. A complex 1-dimensional projective structure, or simply a projective structure, on Σ_g is a system of local coordinates compatible with the orientation modeled on the Riemann sphere such that on any two overlapping coordinate patches, the change of coordinates is a restriction of a projective transformation. In modern language, it is a geometric structure modeled on $(PGL(2, \mathbb{C}), \hat{\mathbb{C}})$.

Throughout this chapter, “projective” means “complex projective”. Some of the basics of complex projective geometry of surfaces, and its relation to Teichmüller theory, are presented in the chapter written by David Dumas, [6].

Since projective transformations are holomorphic, every projective structure determines an underlying complex structure, and hence a surface with a projective structure can be regarded as a Riemann surface. For short, a surface with a projective structure will be called a *projective Riemann surface*. Notice that the notion of projective structure is finer than the notion of complex structure, and different projective Riemann surfaces can share the same underlying complex structure.

Also, since projective geometry contains spherical, euclidean and hyperbolic geometries as subgeometry, any constant curvature surface is a projective Riemann surface and in particular has an underlying complex structure.

Let S be a projective Riemann surface homeomorphic to Σ_g . We always attach to S an orientation preserving homeomorphism,

$$h: \Sigma_g \longrightarrow S,$$

which we call a marking. Two marked projective Riemann surfaces, say (S_1, h_1) and (S_2, h_2) , are considered to be projectively equivalent if there exists a projective isomorphism,

$$\varphi: S_1 \longrightarrow S_2,$$

such that $\varphi \circ h_1$ is homotopic to h_2 . Since $\varphi \circ h_1$ is required to be homotopic to h_2 , the marking determines the homotopy class of a projective isomorphism.

To each projective Riemann surface S , one assigns a developing map,

$$D: \tilde{S} \longrightarrow \hat{\mathbb{C}},$$

defined as an analytic continuation of a preferred local coordinate, where \tilde{S} is the universal cover of S . It is well defined up to composition with projective transformations. One also assigns to S a holonomy representation,

$$\rho: \pi_1(S) \longrightarrow \text{PGL}(2, \mathbb{C}),$$

defined so that the equivariance condition,

$$D(\gamma x) = \rho(\gamma)D(x),$$

holds for all $\gamma \in \pi_1(S)$ and $x \in \tilde{S}$, where $\pi_1(S)$ acts as deck transformations on \tilde{S} . It is well defined up to conjugation by projective transformations.

Let \mathcal{T}_g be the Teichmüller space of Σ_g , namely the space of all complex structures on Σ_g up to marked biholomorphic equivalence. \mathcal{T}_g is homeomorphic to a real euclidean space of dimension 0, 2, $6g - 6$ according to whether $g = 0, 1$ or otherwise. By the uniformization theorem, every Riemann surface is biholomorphic to a constant curvature surface, and hence every biholomorphic class is represented by a projective Riemann surface, but not uniquely. To see how many projective structures can share the same complex structure, we introduce the analytic viewpoint of projective structures below.

A holomorphic quadratic differential, $q = q(z)dz^2$, on a Riemann surface R is an assignment of a holomorphic function $q(z)$ to each local coordinate z such that if z_1 and z_2 are local coordinates with common domain, then

$$q_1(z_1) = q_2(z_2) \left(\frac{dz_2}{dz_1} \right)^2.$$

In other words, it is a holomorphic section of the square of the holomorphic cotangent bundle (the canonical line bundle) of R . The set of all holomorphic quadratic differentials on R becomes a complex vector space of complex dimension 0, 1, $3g - 3$ according to whether $g = 0, 1$ or otherwise. The dimension count is deduced from the Riemann–Roch theorem.

Suppose we have a holomorphic quadratic differential q on R . In a local coordinate z , the solutions of the Schwarzian differential equation

$$2w''(z) + \frac{1}{2}q(z)w(z) = 0$$

form a two-dimensional complex vector space. Then the ratio ϕ of two linearly independent solutions satisfies the identity,

$$\left(\frac{\phi''(z)}{\phi'(z)} \right)' - \frac{1}{2} \left(\frac{\phi''(z)}{\phi'(z)} \right)^2 = q(z),$$

where the left-hand side is called Schwarzian derivative of ϕ . By the standard existence and uniqueness of solutions to systems of holomorphic differential equations, ϕ extends by analytic continuation to a holomorphic map,

$$D: \tilde{R} \longrightarrow \hat{\mathbb{C}},$$

of the universal cover of R unique up to composition with a projective transformation. This map can be seen as a developing map of a projective structure on R associated with q . Thus we have obtained a projective structure from a pair (R, q) of a Riemann surface R and a holomorphic quadratic differential q on R . Conversely, if we are given a projective Riemann surface S , then the Schwarzian derivative of its developing map defines a holomorphic quadratic differential q with respect to the underlying complex structure on S .

Hence the set of projective structures on Σ_g corresponds bijectively to the set of all pairs (R, q) where R is a Riemann surface homeomorphic to Σ_g and q is a holomorphic quadratic differential on R .

Let \mathcal{P}_g be the space of all projective structures on Σ_g up to marked projective equivalence, in other words, the set of all projective Riemann surfaces homeomorphic to Σ_g with marking. We have a natural projection,

$$\pi: \mathcal{P}_g \longrightarrow \mathcal{T}_g,$$

by assigning the underlying complex structure to each projective Riemann surface. This is a vector bundle of complex rank 0, 1, $3g - 3$ according to whether $g = 0$, 1 or otherwise.

When $g = 0$, \mathcal{P}_0 and \mathcal{T}_0 both consist of a single point and the situation is quite simple. When $g \geq 1$, by the uniformization theorem, for each biholomorphic class of projective Riemann surfaces homeomorphic to Σ_g , there is a unique representative by either a euclidean torus or a hyperbolic surface according to whether $g = 1$ or $g \geq 2$. Hence, we obtain a natural section,

$$s: \mathcal{T}_g \longrightarrow \mathcal{P}_g,$$

to the projection $\pi: \mathcal{P}_g \rightarrow \mathcal{T}_g$ by assigning a corresponding constant curvature surface with marking.

In the case $g = 1$, there is a slight difference between complex affine structures and projective structures on Σ_1 . Let \mathcal{A}_1 be the space of all complex affine structures on Σ_1 . The image of a holonomy representation of a complex affine structure on Σ_1 which is not a euclidean structure is contained in the subgroup of $A(1)$ whose action fixes $\{0, \infty\}$. This subgroup is invariant under an involutive conjugation induced by the transformation $z \mapsto 1/z$. The action defines a double cover,

$$\mathcal{A}_1 \longrightarrow \mathcal{P}_1,$$

branched along $s(\mathcal{T}_1)$. Hence the correspondence between complex affine structures and projective structures on the torus is generically two to one.

2.4 Circle packing on surfaces

A projective Riemann surface S would be the most general underlying space when we discuss circle packings on a compact surface, since the group of projective transformations is the maximal group which sends a circle on the Riemann sphere to a circle.

Definition 2.1. A *circle* on a projective Riemann surface S will be a homotopically trivial simple closed curve on S of which a lift in the universal cover is mapped by the developing map to a circle on $\hat{\mathbb{C}}$.

A homotopically trivial simple closed curve on Σ_g always bounds a disk. When a projective Riemann surface S has genus $g \geq 1$, any circle on S bounds a unique disk. However a circle on the Riemann sphere bounds disks in both sides. Hence in this particular case, we need to choose a bounding disk to each circle.

Definition 2.2. A *circle configuration* C on a projective Riemann surface S is a collection of circles such that there is an assignment of bounding disks to each member which are disjoint. Note that the assignment is unique if any when C contains more than one member. We use the notation C also for a subset of S .

Two circles on $\hat{\mathbb{C}}$ bounding disjoint open disks either touch each other at a single point or coincide unless they are disjoint. However this is only for $\hat{\mathbb{C}}$. For circle configurations on a projective Riemann surface of genus $g \geq 1$, two circles may touch at several points, and even a single circle may have self contacts.

Definition 2.3. To each circle configuration C on a projective Riemann surface S , we assign a graph τ on S and simultaneously on Σ_g through a marking where vertices correspond to the circles of C and two (or possibly one) vertices are joined by an edge for each point of tangency. We call τ a *nerve* of C .

A circle configuration C determines an isotopy class of a graph τ on S and therefore on Σ_g through a marking. We use only this topological property for τ and we are not concerned with any geometric properties which τ may have.

Suppose we are given a circle configuration C on some projective Riemann surface. If the complement of the union of bounding disks contains a non simply connected component, we can insert finitely many circles without changing the original configuration to make the complement simply connected. If a simply connected complementary region is bounded by a curvilinear polygon with more than four sides, then again we can insert finitely many circles to make the configuration have the property that the complementary regions consists of only curvilinear triangles and quadrilaterals. In this case, the nerve τ defines a cell decomposition of Σ_g by triangles and quadrilaterals. This simplification of the shape of complementary regions is achieved elementarily. Thus from now on, we will always suppose that a circle configuration has this property unless otherwise stated.

On the other hand, most of curvilinear quadrilateral regions cannot be filled by finitely many circles with only triangular complementary regions. In other words, every projective Riemann surface admits a circle configuration so that the complementary regions are only curvilinear triangles or quadrilaterals, but not every one admits a configuration with only triangular complementary regions. Brooks gave a special name for a circle configuration with this strong property in [4].

Definition 2.4. A circle configuration is said to be a *circle packing* if the complementary regions all are triangular, namely the nerve defines a cell decomposition of Σ_g only by triangles.

The nerve here defines a triangulation of Σ_g in the most general sense, namely some 1-simplex and hence 2-simplex may be immersed and not embedded. Here are a few examples of circle packings illustrated in figures.

Figure 1 is the stereographic image of a circle packing on $\hat{\mathbb{C}}$ whose nerve decomposes $\hat{\mathbb{C}}$ as a tetrahedron. The circle packing P on \mathbb{E}^2 pictured in Figure 2 is called a hexagonal packing. It can be seen also as a universal cover of some circle packing on the torus by taking a quotient of the group action generated by appropriate parallel translations preserving P . If we choose the maximal such group, we get the circle packing on the hexagonal torus by one circle with three self contact points.

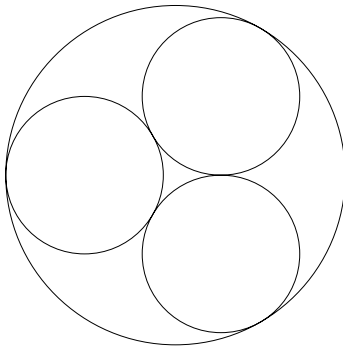


Figure 1. Tetrahedral packing.

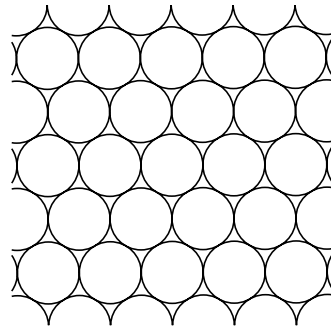


Figure 2. Hexagonal packing.

A circle packing by one circle is realized also on a hyperbolic surface of genus $g \geq 2$. It has $3(2g - 1)$ self contact points. Figure 3 represents a universal cover of such a packing when $g = 2$. As we will discuss in §3, such circle packings admit deformations. Figure 4 illustrates a small deformation of the circle packing in Figure 3 with the same combinatorics.

The graph which appears as a nerve of some circle configuration with polygonal complementary regions on a projective Riemann surface has the property that it defines a cell decomposition of the universal cover such that every closed cell is embedded.

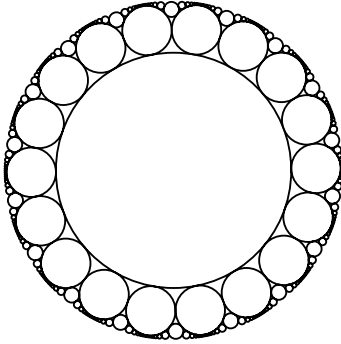


Figure 3. Universal cover of a circle packing on a hyperbolic surface.

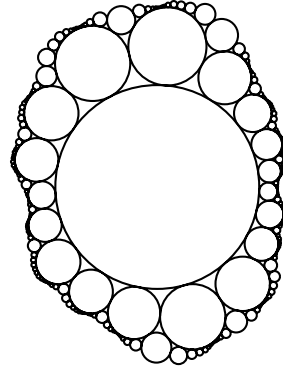


Figure 4. Deformation of the circle packing in Figure 3.

For instance, if all cells are triangular, it defines a honest triangulation in the universal cover. We give a special name for such graphs.

Definition 2.5. A graph on Σ_g is *simple* if it defines a cell decomposition in the universal cover such that each closed cell is embedded. In other words, if it defines a simple planar graph on the universal cover in the graph theoretic sense, namely no loops and no multiple edges.

A fundamental problem in the study of circle configurations on compact surfaces would be to understand the moduli space of the pairs (S, C) of a projective Riemann surface S and a circle configuration C on S under combinatorial control coming from the nerve. The pairs (S, C) and (S', C') will be equivalent if there is a projective isomorphism $\varphi: S \rightarrow S'$ compatible with marking such that $\varphi(C) = C'$.

Problem 2.6. Given a simple graph τ on Σ_g , find the moduli space of all pairs (S, C) of a projective Riemann surface S and a circle configuration C on S with a nerve isotopic to τ through a marking up to equivalence.

Our main concern will be when τ defines a cell decomposition by only triangles and quadrilaterals.

3 Rigidity

3.1 Rigidity theorems

Consider a circle packing P with nerve τ on the Riemann sphere $\hat{\mathbb{C}}$. Since any projective transformation φ of $\hat{\mathbb{C}}$ is isotopic to the identity and sends a circle to a

circle, $P' = \varphi(P)$ defines also a circle packing on $\hat{\mathbb{C}}$ with nerve isotopic to τ . Hence, $(\hat{\mathbb{C}}, P)$ is equivalent to $(\hat{\mathbb{C}}, P')$. Thus roughly speaking, the circle packing on $\hat{\mathbb{C}}$ dominated by τ has a complex 3-dimensional freedom to move on $\hat{\mathbb{C}}$. But this will be the only freedom. The projective rigidity of circle packings on the Riemann sphere was proved originally by Koebe and then rediscovered by Andreev.

Theorem 3.1 (Koebe [9], Andreev [1]). *Suppose a simple graph τ on Σ_0 defines a honest triangulation. Then there is a circle packing P on the Riemann sphere $\hat{\mathbb{C}}$, such that the nerve of P is isotopic to τ . Moreover, for any two such packings P and P' , there is a projective transformation $\varphi: \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ such that $\varphi(P) = P'$.*

Since every orientation preserving self-homeomorphism of the Riemann sphere is isotopic to the identity, the map φ in the above theorem may not be unique. In fact, every graph automorphism of τ which extends to an orientation preserving self-homeomorphism of $\hat{\mathbb{C}}$ can be realized by the restriction of a projective transformation.

When $g \geq 1$, the marking will be involved in the rigidity. Remember that there is a preferred homotopy class of a projective isomorphism between two surfaces with markings. Theorem 3.1 was generalized for higher genus surfaces by Thurston as marked projective rigidity within constant curvature structures. To see clearly the difference of the results between the cases $g = 1$ and $g \geq 2$, we split the statement into two theorems.

Theorem 3.2 (Thurston [18]). *Suppose a simple graph τ on Σ_1 defines a cell decomposition by triangles. Then there is a euclidean torus S with marking and a circle packing P on S , such that the nerve of P is isotopic to τ . Moreover, for any two such realizations (S, P) and (S', P') , there is a projective isomorphism $\varphi: S \rightarrow S'$ compatible with marking such that $\varphi(P) = P'$.*

A projective isomorphism between euclidean tori with markings is either a contraction, an expansion or a parallel translation. Hence the realization of a marked euclidean structure on the torus here is unique up to scaling.

When we fix a euclidean structure, the parallel transformation, which is isotopic to the identity, moves a circle packing. Hence the circle packing on a euclidean torus controlled by τ has a complex 1-dimensional freedom to move.

Also, φ in the above theorem may not be unique. In fact as in the spherical case, every graph automorphism of τ which extends to a self-homeomorphism of Σ_1 isotopic to the identity can be realized by the restriction of a projective transformation.

When $g \geq 2$, we have

Theorem 3.3 (Thurston [18]). *Suppose a simple graph τ on Σ_g ($g \geq 2$) defines a cell decomposition by triangles. Then there is a unique hyperbolic surface S with marking and a unique circle packing P on S , such that the nerve of P is isotopic to τ .*

Thus, in this case, the combinatorial structure of τ completely determines the hyperbolic surface S with marking and the location of a circle packing P on S .

If we combine the three theorems above, the rigidity will be established for the constant curvature surfaces up to marked projective equivalence. The unified statement is as follows.

Theorem 3.4. *Suppose a simple graph τ on Σ_g ($g \geq 0$) defines a cell decomposition by triangles. Then there is a constant curvature surface S with marking and a circle packing P on S such that the nerve of P is isotopic to τ . Moreover, for any two such realizations (S, P) and (S', P') , there is a projective isomorphism $\varphi: S \rightarrow S'$ compatible with markings such that $\varphi(P) = P'$.*

3.2 Unified proof

The theorems in the previous subsection can be proved uniformly by an argument due to Thurston [18]. The hyperbolic case is the source.

To see this, let τ be a simple graph on Σ_g ($g \geq 2$) which gives a cell decomposition by triangles, and V_τ the set of vertices. Let

$$r: V_\tau \longrightarrow \mathbb{R}_+$$

be any positive real valued function on the vertex set. This function will turn out to be an assignment of radii to each vertex. Given $r \in \mathbb{R}^{V_\tau}$, and suppose three vertices $u, v, w \in V_\tau$ span a triangle, then $r(u) + r(v)$, $r(v) + r(w)$, $r(w) + r(u)$ satisfy the triangle inequality. Assigning a hyperbolic triangle with those side lengths to each triple u, v, w which span a triangle on Σ_g , pasting these triangles along edges according to a cell decomposition defined by τ , we get a hyperbolic surface homeomorphic to Σ_g with cone singularities at vertices. It admits a circle packing with centers at V_τ and radii r .

To each vertex, assign the curvature concentrated at that point, and we obtain a curvature concentration map

$$\varrho_r: V_\tau \longrightarrow \mathbb{R},$$

where the value at $v \in V_\tau$ is equal to $2\pi -$ sum of angles meeting at v . The value at v is 0 if and only if v is not singular. Thus if the trivial map denoted by 0 in \mathbb{R}^{V_τ} , which has no curvature concentration for any vertices, is uniquely attained by some radii assignment, then we are done.

Thurston regards the correspondence $r \mapsto \varrho_r$ as a map

$$\mu: \mathbb{R}_+^{V_\tau} \longrightarrow \mathbb{R}^{V_\tau},$$

and sets up the problem more globally. He shows that μ is injective onto its image by comparing the images of different r 's based on the Gauss–Bonnet formula. Moreover he shows that the image of μ contains 0 by looking at the asymptotic nature of μ together with an invariance of domain argument. Thurston's analysis actually provides much more information about the map μ , but in particular it established that $\mu^{-1}(0)$ gives the unique nonsingular hyperbolic surface with a circle packing whose nerve is isotopic to τ .

When $g = 1$, an assignment of radii $r: V_\tau \rightarrow \mathbb{R}_+$, a curvature concentration $\varrho_r: V_\tau \rightarrow \mathbb{R}$ and the map $\mu: \mathbb{R}_+^{V_\tau} \rightarrow \mathbb{R}^{V_\tau}$ can be defined without any change. However, since homothetic radii give the same curvature concentration, μ will not be injective. Also the sum of curvature concentration at each vertex must be zero because of the Gauss–Bonnet formula, and thus μ can never be locally surjective.

To rule out such redundancy coming from expansion and contraction on the source and the target, we let

$$\Lambda_1 = \{r \in \mathbb{R}_+^{V_\tau} \mid \sum_{v \in V_\tau} r(v) = 1\}.$$

Then by an argument similar to the argument used in the case $g \geq 2$, μ restricted to Λ_1 is shown to be injective onto its image in $\{\varrho \in \mathbb{R}^{V_\tau} \mid \sum_{v \in V_\tau} \varrho(v) = 0\}$. Moreover, the image contains 0.

Finally, the spherical case is reduced to the euclidean case. Choose three vertices $u, v, w \in V_\tau$ which span a triangle in $\hat{\mathbb{C}}$, and locate u, v, w so that they span an equilateral triangle in \mathbb{C} with side length 2 and the other part of τ is contained in this triangle. Letting

$$\Lambda_0 = \{r \in \mathbb{R}_+^{V_\tau} \mid r(u) = r(v) = r(w) = 1\},$$

do the same construction of a singular euclidean surface for each $r \in \Lambda_0$ with a fixed triangle boundary spanned by u, v and w . Let V_0 be the set of vertices other than u, v, w , namely $V_0 = V_\tau - \{u, v, w\}$. Then, the map

$$\mu: \Lambda_0 \longrightarrow \mathbb{R}^{V_0}$$

in this case is also injective and the image contains 0. The circle configuration on \mathbb{C} corresponding to $\mu^{-1}(0)$ is pulled back to a circle packing on $\hat{\mathbb{C}}$ with nerve τ by stereographic projection.

Remark 3.5. The setup by Thurston above has led to a variational approach to find the solution $\mu^{-1}(0)$ with respect to the sup norm of ϱ_r by Colin de Verdière in [5]. Bennett and Luo took another variational viewpoint in [2] based on the combinatorial Ricci flow.

Since for a given τ , the realization of the pair (S, P) of a constant curvature surface S and a circle packing P on S is unique up to marked projective equivalence, we give a special name to this pair.

Definition 3.6. We call the unique pair (S, P) provided by the rigidity theorems in the previous subsection a *KAT solution* (the three letters stand for Koebe, Andreev and Thurston).

3.3 Density

One of the conclusions of the rigidity results is that the number of constant curvature surfaces which admit a circle packing is at most countable, because the number of cell decompositions by triangles on Σ_g is countable. In contrast with such a sparse situation, Brooks showed that such structures are dense in Teichmüller space if the combinatorial control is ignored. We here briefly review his idea.

Brooks starts with a circle configuration on the Riemann sphere by four circles with two quadrilateral complementary regions. Let Q be one quadrilateral complementary region normalized as located in a bounded part in Figure 5. The other quadrilateral region is unbounded in this normalization. Then one adds a unique circle which is tangent to either the top, left and bottom sides, or is tangent to the left, top and right sides. The dotted circle in Figure 5 is the one we add. Brooks calls the former case as in Figure 5 a horizontal circle and the latter a vertical circle. Notice that adding this new circles cuts out two new triangles and one new quadrilateral, except in the rare case where this circle is tangent to all four sides.

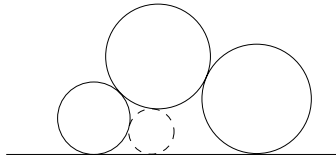


Figure 5. Normalized configuration by four circles

Now, iterate this process, each time adding a circle to the new quadrilateral created in the previous step. Denote by n_1 the number of horizontal circles obtained until one adds a vertical circle, n_2 the number of vertical circles then obtained until one adds a horizontal circles, and so on, and consider the continued fraction expansion

$$c(Q) = n_1 + \frac{1}{n_2 + \frac{1}{n_3 + \dots}}$$

Note that the value $c(Q)$ depends on which circle we put on the horizontal line in the normalized picture. Hence, when we discuss about $c(Q)$, we remember the reference circle for Q .

In [3], the number $c(Q)$ is shown to vary continuously as the original four circles are varied. A rational number corresponds to a continued fraction which terminates in finitely many steps. Geometrically, this means that a circle configuration can be completed by a circle packing in Q by inserting finitely many circles. Otherwise, the process of inserting circles in Q never ends. The number $c(Q)$ is a primitive numerical projective invariant for a curvilinear quadrilateral on $\hat{\mathbb{C}}$, which Brooks call a *continued fractional parameter*.

In [3], Brooks globalizes the argument above to a circle configuration on the Riemann sphere $\hat{\mathbb{C}}$. Let τ be a simple graph on $\hat{\mathbb{C}}$ which defines a cell decomposition by triangles and quadrilaterals, and Q_τ the set of quadrilateral cells in the cell decomposition defined by τ . Now, let \mathcal{C}_τ be the set of circle configurations on $\hat{\mathbb{C}}$ with nerve isotopic to τ up to projective equivalence, endowed with a natural topology. The nerve τ is a combinatorial object as before and \mathcal{C}_τ is the moduli space. By Theorem 3.1, if there are no quadrilateral complementary regions in the circle configuration, \mathcal{C}_τ is just a point. In general, we have

Theorem 3.7 (Brooks [3]). *Suppose a simple graph τ on $\hat{\mathbb{C}}$ defines a cell decomposition by triangles and quadrilaterals, and let Q_τ be the set of quadrilateral cells. Then the map*

$$c_\tau : \mathcal{C}_\tau \longrightarrow \mathbb{R}_+^{Q_\tau},$$

assigning to each configuration $C \in \mathcal{C}_\tau$ the continued fractional parameters of each member of Q_τ , is a homeomorphism.

This theorem can be understood in the language of the quasi-conformal deformation theory. To see this, let us quickly review the theory. A finitely generated discrete subgroup Γ of Möb is called a Kleinian group. The maximal region of $\hat{\mathbb{C}}$ on which Γ acts properly discontinuously is called the domain of discontinuity of Γ and denoted by Ω_Γ . By Ahlfors' finiteness theorem, the quotient Ω_Γ/Γ is a Riemann surface of finite type. A Kleinian group Γ' is quasi-conformally equivalent to Γ if there is a quasi-conformal map $\psi : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ such that $\Gamma' = \psi\Gamma\psi^{-1}$. Note here that Γ' is assumed to be a Kleinian group and thus a discrete subgroup of Möb. A culminating result of the extensive quasi-conformal deformation theory developed by Ahlfors, Bers, Maskit, Marden and many others gives an explicit description of the deformation space as follows, see [15], [16] for the final form due to Sullivan.

Theorem 3.8. *The set $\mathcal{QC}(\Gamma)$ of quasi-conformal deformations of Γ is homeomorphic to the Teichmüller space of the underlying topological surface of Ω_Γ/Γ , in other words, the space of Riemann surfaces homeomorphic to Ω_Γ/Γ with marking up to marked biholomorphic equivalence.*

Let us come back to the circle packing problem. We have a circle configuration C on $\hat{\mathbb{C}}$ with only triangular and quadrilateral complementary regions. Let Γ be a group generated by reflections about members in C . The group Γ will be a Kleinian group and Ω_Γ/Γ in this case consists of finitely many triangles and quadrilaterals which can be identified with the complementary regions of C . The Teichmüller space of each region, namely, the space of all complex structures on the region with marked points on the boundary, is homeomorphic to either a point or \mathbb{R} according to whether it is triangular or quadrilateral. Thus $\mathcal{QC}(\Gamma)$ in this case is homeomorphic to the euclidean space of dimension equal to the number of quadrilateral complementary regions.

Now, notice that there is a rigidity property of reflections in quasi-conformal deformations. If $\gamma \in \Gamma$ is a reflection, then the corresponding element $\gamma' \in \Gamma' = \psi\Gamma\psi^{-1}$

must also be a reflection, since a Möbius transformation which fixes a 1-dimensional set is a reflection.

Thus any quasi-conformal deformation $\Gamma' \in \mathcal{QC}(\Gamma)$ starting from a circle configuration C defines a circle configuration C' whose nerve is isotopic to τ . Therefore, the theory provides a homeomorphic correspondence

$$\delta: \mathcal{QC}(\Gamma) \longrightarrow \mathcal{C}_\tau$$

and gives a new parameterization of the Teichmüller space of quadrilateral regions, which is the underlying topological surface of Ω_Γ/Γ , in terms of continued fractions.

Corollary 3.9 (Brooks [3]). *The composition $c_\tau \circ \delta: \mathcal{QC}(\Gamma) \rightarrow \mathbb{R}_+^{\mathcal{Q}\tau}$ is a homeomorphism.*

This whole story can be extended to the study of circle packings on compact projective Riemann surfaces of genus $g \geq 2$. Let S be a compact hyperbolic surface, Δ the image of a holonomy representation of $\pi_1(S)$ in $\mathrm{PGL}(2, \mathbb{C})$ and C a circle configuration on S with only triangular and quadrilateral complementary regions. Notice that Δ is a Fuchsian group. A circle configuration C defines \tilde{C} on the universal cover $\tilde{S} \subset \hat{\mathbb{C}}$ with infinitely many circles, but finitely many conjugacy classes of circles with respect to the action of $\pi_1(S)$. Also, since S is a hyperbolic surface, the boundary of the universal cover \tilde{S} defines a circle $C_0 \subset \hat{\mathbb{C}}$. Let Γ be a group generated by Δ and reflections about circles of \tilde{C} and C_0 . Since the number of conjugacy classes of circles in \tilde{C} by the action of $\pi_1(S)$ is finite, Γ is a finitely generated Kleinian group.

By the theory of quasi-conformal deformations, the deformation space $\mathcal{QC}(\Gamma)$ is homeomorphic to the Teichmüller space of the union of quadrilateral complementary regions of $C \subset S$, which is parameterized by continued fractional parameters by an equivariant version of Theorem 3.7.

Since the quasi-conformal deformation $\Gamma' \in \mathcal{QC}(\Gamma)$ contains a reflection about the boundary of the universal cover of a deformed surface, the result becomes again a hyperbolic surface. We can choose a rational valued continued fractional parameter arbitrary close to the parameter of C . Then the quadrilateral regions in the deformed configuration C' on a hyperbolic surface S' can be completed by a circle packing by inserting finitely many circles. Since a hyperbolic surface S' could be chosen arbitrarily close to the original S , we get a twofold result by regarding \mathcal{T}_g as the space of Riemann surfaces with marking up to biholomorphic equivalence, and also as the space of hyperbolic surfaces with marking up to isometry through the identification by the section $s: \mathcal{T}_g \rightarrow \mathcal{P}_g$.

Theorem 3.10 (Brooks [4]). *The set of Riemann surfaces which admit a circle packing is dense in \mathcal{T}_g . Equivalently, the set of hyperbolic surfaces which admit a circle packing is dense in $s(\mathcal{T}_g)$.*

The argument so far depends on the well-developed theory of Kleinian groups, that is, discrete subgroups of Möb. On the other hand, the image of the holonomy

representation of a projective Riemann surface is not discrete in general, and some difficulty for analyzing density arises.

When $g = 1$, as a byproduct of the study of one circle packing on complex affine tori in [13], Mizushima proved the density of circle packing structures in \mathcal{A}_1 . Since \mathcal{A}_1 doubly covers \mathcal{P}_1 branched along $s(\mathcal{T}_1)$, we have

Theorem 3.11 (Mizushima [13]). *The set of projective Riemann tori which admit a circle packing is dense in \mathcal{P}_1 .*

We may ask

Question 3.12. Is the set of projective Riemann surfaces which admit a circle packing dense in \mathcal{P}_g for $g \geq 2$?

4 Flexibility

4.1 Constructing moduli

Despite of the rigidity discussed in the previous section, circle packings on projective Riemann surfaces with combinatorics controlled by τ have a more flexible nature. One expects to see the structure of a moduli space for the pairs (S, P) of a projective Riemann surface S and a circle packing P on S such that the nerve of P is isotopic to τ . To do this, a projective invariant of a circle packing on projective Riemann surfaces based on the cross ratio is introduced in [10]. In this section, we briefly review it based on the description in [12].

Suppose that (S, P) is a pair consisting of a projective Riemann surface S and a circle packing P on S . To each edge e of the nerve τ of P , we choose a lift \tilde{e} in $\tilde{\tau}$ and associate a configuration of four circles on $\hat{\mathbb{C}}$ in the developed image about $D(\tilde{e})$, see Figure 6. Recall that the cross ratio of four distinct ordered points in $\hat{\mathbb{C}}$ is given by

$$(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}.$$

It is the value of the image of z_1 under the projective transformation which takes z_2, z_3 and z_4 to 1, 0 and ∞ respectively. The value assigned to the edge e will be the imaginary part of the cross ratio of the four contact points $(p_{14}, p_{23}, p_{12}, p_{13})$ of the configuration chosen as in Figure 6 with orientation convention. The cross ratio of these four points is always purely imaginary with positive imaginary part.

Since the cross ratio is a projective invariant, the value does not depend on the choice of the lift \tilde{e} and on the developing map. Collecting the values for each edge, we obtain the map x of the edge set E_τ of τ ,

$$x: E_\tau \longrightarrow \mathbb{R},$$

which is called a *cross ratio parameter*. The cross ratio of the edge e determines the position of the circle C_4 in Figure 6 once the positions of C_1, C_2 and C_3 are fixed, and if the cross ratio of e approaches ∞ , then C_4 approaches p_{13} .

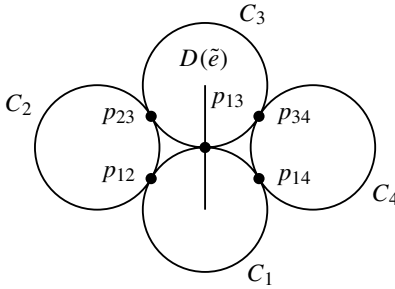


Figure 6. Four circle configuration.

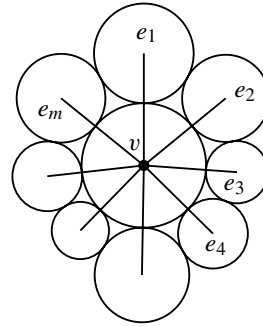


Figure 7. Surrounding circles.

Obviously, not all real valued maps of E_τ can be cross ratio parameters for some circle packing. To obtain necessary conditions, consider a normalized picture of a circle with its surrounding circles. The normalization we chose maps the central circle to the real line and one of the adjoining interstices to the standard interstice with vertices at $\infty, 0$ and $\sqrt{-1}$. This leads one to introduce an associated matrix $A \in \text{SL}(2, \mathbb{R})$ to each edge $e \in E_\tau$. If the value of a cross ratio parameter at e is x , A is defined to be $\begin{pmatrix} 0 & 1 \\ -1 & x \end{pmatrix} \in \text{SL}(2, \mathbb{R})$.

Then a simple computation shows that the associated matrix A represents a transformation which sends the left triangular interstice of this configuration to the right triangular interstice.

Let v be a vertex of τ with valence m . We read off the edges e_1, \dots, e_m incident to v in a clockwise direction to obtain a sequence of assigned values x_1, \dots, x_m of cross ratio parameters. Let

$$W_j = A_1 A_2 \dots A_j = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}, \quad j = 1, \dots, m,$$

where A_i is the matrix $\begin{pmatrix} 0 & 1 \\ -1 & x_i \end{pmatrix}$ associated to e_i . Then, it was shown in [10] that for each vertex v of τ , we have

$$W_v = A_1 A_2 \dots A_m = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{4.1}$$

and

$$\begin{cases} a_j, c_j < 0, b_j, d_j > 0 & \text{for } 1 \leq j \leq m - 1, \\ & \text{except for } a_1 = d_{m-1} = 0. \end{cases} \tag{4.2}$$

The first condition comes from the fact that the chain of circles surrounding the circle corresponding to v closes up. The second condition excludes overwinding, that is, it eliminates the case where the chain surrounds the central circle more than once. Notice here that the associated matrices are in $\mathrm{SL}(2, \mathbb{R})$ and not in $\mathrm{PSL}(2, \mathbb{R})$, so that the inequalities of (4.2) make sense.

On the other hand, given a real valued map \mathbf{x} of E_τ satisfying (4.1) and (4.2) for each vertex of τ , it is relatively routine to construct a pair (S, P) consisting of a projective Riemann surface S and a circle packing P on S so that its cross ratio parameter is \mathbf{x} (see [10] for details). Thus set

$$\mathcal{C}_\tau = \{\mathbf{x} : E_\tau \rightarrow \mathbb{R} \mid \mathbf{x} \text{ satisfies (4.1) and (4.2) for each vertex}\},$$

and call it the *cross ratio parameter space*.

Remark 4.1. In §3.3, we defined \mathcal{C}_τ as the moduli space of circle configurations on $\hat{\mathbb{C}}$ with nerve isotopic to τ up to projective equivalence, endowed with a natural topology. When τ defines a cell decomposition by triangles, the space \mathcal{C}_τ is a point. A nontrivial moduli space appears only when the decomposition contains a quadrilateral cell. Here we use the same notation since \mathcal{C}_τ is naturally identified with the moduli space of circle packings on projective Riemann surfaces of genus $g \geq 1$ controlled by τ . It will be nontrivial even if τ defines a cell decomposition by only triangles.

Since Condition (4.1) gives a set of polynomial equations for the x_i 's and (4.2) are polynomial inequalities in the x_i 's, the moduli space \mathcal{C}_τ is a semi-algebraic set, and we define the topology on \mathcal{C}_τ to be the one induced by the tautological inclusion $\iota : \mathcal{C}_\tau \rightarrow \mathbb{R}^{E_\tau}$. It turns out that this naive construction gives a correct parameterization of the moduli space of pairs (S, P) where S is a projective Riemann surface and P is a circle packing on S with nerve τ .

Lemma 4.2. *If $g \geq 1$, and a simple graph τ on Σ_g defines a triangulation in the universal cover, then we have the following:*

- (1) (Lemma 2.17 in [10]) *A moduli space \mathcal{C}_τ corresponds bijectively to the set of all pairs (S, P) where S is a projective Riemann surface and P is a circle packing on S with nerve τ , up to marked projective equivalence.*
- (2) (Lemma 3.2 in [11]) *The tautological inclusion $\iota : \mathcal{C}_\tau \rightarrow \mathbb{R}^{E_\tau}$ is proper.*

In view of the above results, \mathcal{C}_τ is naturally identified with the moduli space of all pairs (S, P) with nerve τ . The study of the moduli space then reduces to the study of its semi-algebraic representative \mathcal{C}_τ .

4.2 Thurston coordinates

To each pair (S, P) in \mathcal{C}_τ , assigning its first component, we obtain the *forgetful map*

$$f: \mathcal{C}_\tau \longrightarrow \mathcal{P}_g.$$

The image $f(\mathcal{C}_\tau)$ consists of all projective Riemann surfaces which admit a circle packing with nerve τ . The projective rigidity implies that $f(\mathcal{C}_\tau)$ intersects $s(\mathcal{T}_g)$ only at $f(\{\text{KAT}\})$ and furthermore, the rigidity of the circle packing on $f(\{\text{KAT}\})$ means that the inverse image of this point under f consists of exactly one point. We discuss here the description of $f(\mathcal{C}_\tau)$ with respect to Thurston coordinates of \mathcal{P}_g which we will describe shortly.

In this subsection, we assume that the surface Σ_g has genus $g \geq 2$. Roughly speaking, a measured lamination on Σ_g is a closed subset of Σ_g , locally homeomorphic to a product of a totally disconnected subset of the interval with an interval, together with a transverse measure. We refer to the other chapters of the Handbook for a precise definition, in particular to Chapter 12 ([6]). Moreover, we restrict ourselves to the case where every leaf is homotopic to a geodesic with respect to some (and hence any) hyperbolic metric on Σ_g . A noncontractible simple closed curve on Σ_g with counting measure for transverse arcs is an elementary, but important and fundamental example of a measured lamination. The space of isotopy classes of measured laminations on Σ_g ($g \geq 2$) with the weak* topology on measures will be denoted by \mathcal{ML}_g . The set of weighted homotopically nontrivial simple closed curves is dense in \mathcal{ML}_g . Also \mathcal{ML}_g is known to be homeomorphic to \mathbb{R}^{6g-6} . See [18], [19] for details.

Although a measured lamination is a topological concept, once we put a hyperbolic metric on Σ_g , its support is canonically realized as a disjoint union of simple geodesics which forms a closed subset on the surface. Such a lamination is called a geodesic lamination with transverse measure.

Thurston has assigned to each projective Riemann surface a hyperbolic surface with a measured geodesic lamination. Following [8], we briefly review his idea. Start with a projective Riemann surface S which is not a hyperbolic surface. A maximal disk in the universal cover \tilde{S} is a maximal one under the inclusion. We consider the set of maximal disks in the universal cover. Each maximal disk is naturally endowed with the hyperbolic metric, the boundary of each disk intersects the ideal boundary of \tilde{S} in two or more points and we can take the convex hull of these ideal boundary points. It can be shown that this gives a stratification of \tilde{S} by ideal polygons, and ideal bigons foliated by “parallel lines” joining the two ideal vertices of the bigons. The polygonal parts support a canonical hyperbolic metric. Collapsing each bigon foliated by parallel lines in \tilde{S} to a line and taking the quotient of the result by the action of the fundamental group, we obtain a hyperbolic surface H . This defines a hyperbolization map

$$\alpha: \mathcal{P}_g \longrightarrow s(\mathcal{T}_g) \subset \mathcal{P}_g.$$

The stratification also defines a geodesic lamination λ on H by taking the union of collapsed lines. Moreover, identifying $\hat{\mathbb{C}}$ with the boundary of 3-dimensional hy-

perbolic space \mathbb{H}^3 , and using the convex hull of the ideal points of the maximal disk not on the disk itself but in \mathbb{H}^3 , we can assign a transverse bending measure supported on λ . This defines a pleating map

$$\beta: \mathcal{P}_g \longrightarrow \mathcal{ML}_g.$$

Theorem 4.3 (Thurston, see [8]). *The product of these maps*

$$(\alpha, \beta): \mathcal{P}_g \longrightarrow s(\mathcal{T}_g) \times \mathcal{ML}_g, \quad (4.3)$$

is a homeomorphism.

We call the parameterization of \mathcal{P}_g by the target of (4.3) Thurston coordinates. It is known by Tanigawa [17] that the restriction of $\pi: \mathcal{P}_g \rightarrow \mathcal{T}_g$ to any slice $s(\mathcal{T}) \times \{*\}$ by the first factor is proper, and by Scannell and Wolf [14] that this map is locally injective. In particular, this map is a diffeomorphism. Dumas and Wolf also proved that the same is true for the slice $\{*\} \times \mathcal{ML}_g$ by the second factor in [7]. Further information on Thurston's coordinates is contained in Chapter 12 of this volume [6].

On the other hand, we have:

Lemma 4.4 (Lemma 4.1 in [11]). *If $g \geq 2$, then the composition $\beta \circ f: \mathcal{C}_\tau \rightarrow \mathcal{ML}_g$ of a forgetful map $f: \mathcal{C}_\tau \rightarrow \mathcal{P}_g$ with the pleating map $\beta: \mathcal{P}_g \rightarrow \mathcal{ML}_g$ has bounded image.*

This is a property for projective Riemann surfaces admitting a circle packing dominated by a single graph τ , and it is proved by observing how the developed image of a projective Riemann surface is controlled by the combinatorial data of τ .

4.3 Speculation

To expect a deeper understanding of the moduli space \mathcal{C}_τ and its image in \mathcal{P}_g under f , we formulate a conjecture which relates the moduli space with Teichmüller space.

Conjecture 4.5. Let τ be a simple graph on Σ_g which defines a cell decomposition by triangles. Then the composition $\pi \circ f: \mathcal{C}_\tau \rightarrow \mathcal{T}_g$ of the forgetful map $f: \mathcal{C}_\tau \rightarrow \mathcal{P}_g$ with the projection $\pi: \mathcal{P}_g \rightarrow \mathcal{T}_g$ is a homeomorphism.

The motivation goes back to the result of Mizushima in [13] which we discuss in the next subsection. Here are some expected implications of the affirmative solution to Conjecture 4.5, which have been verified in certain special cases.

- (1) *Topology of \mathcal{C}_τ .* The moduli space \mathcal{C}_τ would be homeomorphic to the euclidean space of dimension 2 or $6g - 6$ according to whether $g = 1$ or $g \geq 2$.
- (2) *Rigidity for circle packings.* The forgetful map $f: \mathcal{C}_\tau \rightarrow \mathcal{P}_g$ would be injective. Thus the rigidity of circle packings holds for all projective Riemann surfaces in $f(\mathcal{C}_\tau)$, that is, each projective Riemann surface S admits at most one circle packing with nerve τ up to projective automorphisms isotopic to the identity.

- (3) *New section to π .* The image $f(\mathcal{C}_\tau)$ of the forgetful map would define a new natural section or a slice to $\pi: \mathcal{P}_g \rightarrow \mathcal{T}_g$. It means for example that for each biholomorphic class of a Riemann surface, there exists a unique projective Riemann surface which admits a circle packing with nerve τ .

4.4 Evidence

4.4.1. Consider the moduli space \mathcal{C}_τ of circle packings by one circle on Σ_1 . The nerve τ in this case consists of one vertex v and 3 edges e_1, e_2 and e_3 with cross ratios $x > 0$, $y > 0$ and $z > 0$ respectively and associated matrices $X = \begin{pmatrix} 0 & 1 \\ -1 & x \end{pmatrix}$, $Y = \begin{pmatrix} 0 & 1 \\ -1 & y \end{pmatrix}$ and $Z = \begin{pmatrix} 0 & 1 \\ -1 & z \end{pmatrix}$. The word associated to the vertex is given by $W = XYZXYZ$, and $XYZXYZ = I$ implies

$$xyz = x + y + z.$$

Note that three equations derived from the matrix identity reduce to just one equation in this case. By an easy computation, we can see that the cross ratio parameter space \mathcal{C}_τ is given by

$$\mathcal{C}_\tau = \{(x, y, z) \in \mathbb{R}^3 \mid xyz = x + y + z, x, y, z > 0\},$$

which is homeomorphic to a convex domain in the xy -plane,

$$\{(x, y) \in \mathbb{R}^2 \mid xy - 1 > 0, x, y > 0\},$$

by the projection.

Mizushima studied the moduli space of complex affine structures on the torus in [13]. When we translate his result in our language, his moduli space is a doubly branched cover of the moduli space and provides an affirmative solution to Conjecture 4.5 for this very special case.

Theorem 4.6 (Mizushima [13]). *If a simple graph τ on Σ_1 has only one vertex, then the composition $\pi \circ f: \mathcal{C}_\tau \rightarrow \mathcal{T}_1$ of the forgetful map $f: \mathcal{C}_\tau \rightarrow \mathcal{P}_1$ with the projection $\pi: \mathcal{P}_1 \rightarrow \mathcal{T}_1$ is a homeomorphism.*

4.4.2. The argument developed by Brooks, described in §3.3, is extendable to projective Riemann surfaces such that a developing map extends to an embedding of the closure of the universal cover. Such a surface is here called *strongly uniformizable*. Let τ be a simple graph on Σ_g which defines a cell decomposition with only triangular and quadrilateral cells, S a strongly uniformizable surface, Δ the image of a holonomy representation in $\mathrm{PGL}(2, \mathbb{C})$ and C a circle configuration on S whose nerve is isotopic to τ . In this case, the universal cover \tilde{S} is embedded in $\hat{\mathbb{C}}$, but the boundary $\partial\tilde{S}$ would not be a round circle and is in general a quasi-circle on $\hat{\mathbb{C}}$. Adding to Δ only reflections about members of \tilde{C} without C_0 in §3.3, we get a Kleinian group Γ . Then the quasi-conformal deformation theory tells us that $\mathcal{QC}(\Gamma)$ is homeomorphic

to the product of the Teichmüller spaces of the quadrilateral complementary regions of $C \subset S$ and \mathcal{T}_g which corresponds to the outside region of the closure of \tilde{S} in \hat{C} .

If C is a circle packing, namely, if all complementary regions are triangular, then $\mathcal{QC}(\Gamma)$ is homeomorphic to \mathcal{T}_g . This shows that there is a family of projective Riemann surfaces parametrized by \mathcal{T}_g which admit a circle packing with nerve isotopic to τ , and in fact we have an embedding

$$\mathcal{T}_g \cong \mathcal{QC}(\Gamma) \longrightarrow \mathcal{QC}(\Delta) \cong \mathcal{T}_g \times \mathcal{T}_g \subset \mathcal{P}_g$$

This observation establishes the local structure of \mathcal{C}_τ at the KAT solution for $g \geq 2$.

When $g = 1$, hyperbolic Dehn surgery theory developed by Thurston [18] plays a role similar to the one of quasi-deformation theory in a small neighborhood of the KAT solution. Thus one may establish

Theorem 4.7 (Theorem 1 in [10]). *Let τ be a simple graph on Σ_g ($g \geq 1$) which defines a cell decomposition by triangles. Then there is a neighborhood U of the KAT solution in \mathcal{C}_τ such that*

- (1) *U is homeomorphic to the euclidean space of dimension 2 or $6g - 6$ according to whether $g = 1$ or $g \geq 2$,*
- (2) *the restriction of f to U is injective.*

4.4.3. The restriction for τ to have only one vertex as in Mizushima's setting simplifies the situation even for the case $g \geq 2$. In fact, \mathcal{C}_τ is defined by just one matrix equation and the set of inequalities corresponding to (4.1) and (4.2) respectively. This rather simple setting enable us to prove for example,

Theorem 4.8. *Let τ be a simple graph on Σ_g ($g \geq 2$) with only one vertex, then*

- (1) (Theorem 2 in [10]) \mathcal{C}_τ is homeomorphic to \mathbb{R}^{6g-6} ;
- (2) (Lemma 5.1 in [10]) $f: \mathcal{C}_\tau \rightarrow \mathcal{P}_g$ is injective;
- (3) (Theorem 1.1 in [11]) $\pi \circ f: \mathcal{C}_\tau \rightarrow \mathcal{T}_g$ is proper.

Theorem 4.8 comes fairly close to affirmatively answering Conjecture 4.5 for the one circle packing case. What is missing is a proof that p restricted to $f(\mathcal{C}_\tau)$ is locally injective.

References

- [1] E. M. Andreev, Convex polyhedra of finite volume in Lobacevskii space. *Mat. Sb.* (N.S.) 83 (1970), 256–260; English transl. *Math. USSR-Sb.* 12 (1970), 255–259. [510](#), [518](#)
- [2] C. Bennett and F. Luo, Combinatorial Ricci flow on surfaces. *J. Differential Geom.* 63 (2003), 97–129. [520](#)

- [3] R. Brooks, On the deformation theory of classical Schottky groups. *Duke Math. J.* 52 (1985), 1009–1024. [510](#), [521](#), [522](#), [523](#)
- [4] R. Brooks, Circle packings and co-compact extensions of Kleinian groups. *Invent. Math.* 86 (1986), 461–469. [510](#), [516](#), [523](#)
- [5] Y. Colin de Verdière, Un principe variationnel pour les empilements de cercles. *Invent. Math.* 104 (1991), 655–669. [520](#)
- [6] D. Dumas, Complex projective structures. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 455–508. [512](#), [527](#), [528](#)
- [7] D. Dumas and M. Wolf, Projective structures, grafting, and measured laminations. *Geom. Topol.* 12 (2008), 351–386. [528](#)
- [8] Y. Kamishima and S. P. Tan, Deformation spaces associated to geometric structures. In *Aspects of Low Dimensional Manifolds*, Adv. Stud. Pure Math. 20, Kinokuniya, Tokyo 1992, 263–300. [527](#), [528](#)
- [9] P. Koebe, Kontaktprobleme der konformen Abbildung. *Ber. Sachs. Akad. Wiss. Leipzig, Math-Phys. Klasse* 88 (1936), 141–164. [510](#), [518](#)
- [10] S. Kojima, S. Mizushima and S. P. Tan, Circle packings on surfaces with projective structures. *J. Differential Geom.* 63 (2003), 349–397. [510](#), [524](#), [525](#), [526](#), [530](#)
- [11] S. Kojima, S. Mizushima and S. P. Tan, Circle packings on surfaces with projective structures and uniformization. *Pacific J. Math.* 225 (2006), 287–300. [510](#), [526](#), [528](#), [530](#)
- [12] S. Kojima, S. Mizushima and S. P. Tan, Circle Packings on Surfaces with Projective Structures: A survey. London Math. Soc. Lecture Note Ser. 329, Cambridge University Press, Cambridge 2006, 337–353. [510](#), [524](#)
- [13] S. Mizushima, Circle packings on complex affine tori. *Osaka J. Math.* 37 (2000), 873–881. [524](#), [528](#), [529](#)
- [14] K. P. Scannell and M. Wolf, The grafting map of Teichmüller space. *J. Amer. Math. Soc.* 15 (2002), 893–927. [528](#)
- [15] D. Sullivan, On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions. *Ann. Math. Stud.* 97 (1981), 465–496. [522](#)
- [16] D. Sullivan, Quasi-conformal homomorphism and dynamics II: structural stability implies hyperbolicity for Kleinian groups. *Acta Math.* 155 (1985), 243–260. [522](#)
- [17] H. Tanigawa, Grafting, harmonic maps, and projective structures on surfaces. *J. Differential Geom.* 47 (1997), 399–419. [528](#)
- [18] W. P. Thurston, *The geometry and topology of 3-manifolds*. Lecture Notes, Princeton University, 1977/78. [510](#), [518](#), [519](#), [527](#), [530](#)
- [19] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc.* 19 (1988), 417–431. [527](#)

Chapter 14

(2 + 1) Einstein spacetimes of finite type

Riccardo Benedetti and Francesco Bonsante

Contents

1	Introduction	534
2	3-dimensional gravity	538
	2.1 General background	538
	2.2 (2 + 1)-spacetimes	539
3	The space $\mathcal{ML}(S)$	543
	3.1 The Teichmüller space $\tilde{\mathcal{T}}(S)$	543
	3.2 The convex-core map	544
	3.3 The space of measured geodesic laminations	550
	3.4 The subspace $\mathcal{ML}_c(S)$	552
	3.5 Enhanced bundle $\mathcal{ML}_c(S)^\#$	554
	3.6 Grafting, bending, earthquakes	559
4	Wick rotation-rescaling theory	566
	4.1 Cosmological time	567
	4.2 Grafting and Lorentzian grafting	567
	4.3 Wick rotation-rescaling set-up	571
	4.4 Flat spacetimes classification	573
	4.5 Wick rotation: flat Lorentzian vs. hyperbolic geometry	577
	4.6 Flat vs. de Sitter Lorentzian geometry	583
	4.7 Flat vs. Anti de Sitter Lorentzian geometry	585
5	Causal AdS spacetimes, earthquakes and black holes	594
	5.1 On holonomy information	594
	5.2 Canonical causal AdS spacetimes with prescribed holonomy	595
	5.3 AdS bending and Earthquake Theorems	597
	5.4 Convex core of $\Omega(h)$ and black holes	598
	5.5 (Broken) T -symmetry	601
6	Including particles	602
	6.1 Maximal globally hyperbolic spacetimes with particles	604
	6.2 Earthquakes on hyperbolic cone surfaces	605
	References	606

1 Introduction

A surface S is said to be of *finite type* if the following holds:

- (1) It is of the form

$$S = \hat{S} \setminus V$$

where (\hat{S}, V) is a closed oriented surface of genus $g \geq 0$, with a set of $r \geq 0$ marked points $V = \{p_1, \dots, p_r\}$.

- (2) The fundamental group of S is *non Abelian*, equivalently $2 - 2g - r < 0$.

The main aim of this chapter is to describe, for every S of finite type, and for every $\kappa = 0, \pm 1$, the geometry of 3-dimensional *maximal globally hyperbolic* Lorentzian spacetimes of *constant curvature* κ that contain a *complete* Cauchy surface homeomorphic to S . We call them generically *Einstein MGH spacetimes of finite type*. The (3-dimensional) general relativity background will be briefly recalled in Section 2. These spacetimes are supported by the product $S \times \mathbb{R}$. Considered up to Lorentzian isometry homotopic to the identity of $S \times \mathbb{R}$, they form, for every κ , a *Teichmüller-like space* denoted by

$$\mathcal{MGH}_{\kappa}(S).$$

Clearly these notions make sense also if S is not necessarily of finite type. In the monograph [15] we have developed a *canonical Wick rotation-rescaling* theory on such general MGH spacetimes. It is easy to see that $\mathcal{MGH}_{\kappa}(S) \neq \emptyset$ for every κ , if and only if the universal covering of S is homeomorphic to the open disk D^2 . In [15] we have actually analyzed $\mathcal{MGH}_{\kappa}(D^2)$, by developing also an *equivariant version* of the theory, with respect to the action of *any* discrete isometry group. Wick rotation-rescaling theory includes a wide generalization of Mess' classification [55] (completed by Scannell [60] for $\kappa = 1$) of MGH spacetimes with *compact* Cauchy surfaces (*i.e.* $V = \emptyset$). Moreover, it establishes explicit geometric correlations between spacetimes of different curvatures, and between spacetimes and *complex projective structures* on S . In particular, this gives a clear geometric explanation of the occurrence of a certain "*universal*" *parameter space*

$$\mathcal{ML}(S)$$

shared by all $\mathcal{MGH}_{\kappa}(S)$, $\kappa = 0, \pm 1$, and by $\mathcal{P}(S)$, the Teichmüller-like space of complex projective structures on S .

A large part of this chapter just reports on such a theory, by specializing it to the case of a surface S of finite type. This class is large enough to display the main features of the theory; on the other hand, spacetimes of finite type are possibly easier to describe than completely general ones. In fact we will spell out several specific statements that are quite implicit in the general treatment given in [15]. Hence the present chapter represents an actual complement to that monograph. Moreover, there are in this case direct relations between $\mathcal{ML}(S)$ and the more familiar *Teichmüller spaces of hyperbolic structures on S* and, to some extent, with their corresponding *tangent bundles* (see Section 3). For example, when S is compact $\mathcal{ML}(S)$ coincides with

the (topologically trivialized) bundle $\mathcal{T}_g \times \mathcal{ML}_g$ of *measured geodesic laminations* on hyperbolic structures on S . In general we will deal with hyperbolic structures F on S whose completions F^c have (non necessarily compact) geodesic boundary, and with a kind of measured geodesic laminations λ on F^c . In fact, another goal is to convince a reader familiar with such topics of hyperbolic geometry that not only these topics provide some important tools for studying Einstein spacetimes; in the reverse direction, via Lorentzian geometry we get a new insight into several fundamental hyperbolic constructions such as *grafting*, (3-dimensional hyperbolic) *bending* and *earthquakes* along laminations. To support this claim we just mention here the “AdS proof” of *Thurston’s Earthquake Theorem* for hyperbolic structures on compact surfaces S , that Mess obtained in [55] as a by-product of his classification of spacetimes in $\mathcal{MGH}_{-1}(S)$. An AdS look at earthquake theory *beyond the compact case* will be a theme of Section 5.

Finally, we note that spacetimes of finite type occur (via canonical Wick rotation) as “ending spacetimes” of geometrically finite hyperbolic 3-manifolds, which furnish basic examples for a bordism category supporting (2 + 1) QFT pertinent to 3-dimensional gravity (see Section 1.11 of [15], and [11], [12], [13]).

In Section 5, we focus on the AdS case that displays the richer phenomenology, mostly referring (besides [15]) to [7], [8] and [29]. In particular we will describe the common maximal *causal extension* $\Omega(h)$ of the MGH spacetimes of finite type that share a given AdS holonomy h . We will see that $\Omega(h)$ is still supported by the product $S \times \mathbb{R}$ but in general is not globally hyperbolic. This is a particularly interesting case, because we can detect a specific one among the maximal globally hyperbolic spacetimes contained in $\Omega(h)$ that can be truly considered as a *black hole*. The analysis of the causal extension is also important to achieve a proof of the Earthquake Theorem.

Finally, in Section 6 we will outline (by following [17] and mostly [28]) how the Wick rotation-rescaling theory (partially) extends to MGH spacetimes of finite type that include world lines of “particles” (*i.e.* inextensible timelike lines of spacelike *conical singularities*).

We stress that this chapter is not intended to be exhaustive of the subject. We have made a few partial and subjective choices, organized around our favorite Wick rotation-rescaling view point. Nevertheless, we hope that this would be enough to show that 3-dimensional gravity is a fairly non-trivial and beautiful “toy model”. In particular, we have neglected a classical analytic approach to the classification of constant curvature MGH spacetimes in terms of solutions of the Gauss–Codazzi equation at a Cauchy surface, possibly imposing some supplementary conditions to such solutions, that translates some geometric property of the embedding of S as Cauchy surface (see also Section 2). A widely studied possibility requires that the surface has *constant mean curvature* (see for instance [56], [4], [10], [49]). At least for compact S , the classical Teichmüller space of *conformal* structures on S and its complex cotangent bundle arise in this way towards the classification. This approach also selects a distinguished *global time* on MGH spacetimes, that basically coincides with the mean curvature of its level surfaces.

Wick rotation-rescaling theory is based on a rather different more geometric approach, initiated by Mess in [55]. It turns out that a key ingredient is another canonical time, the so called *cosmological time*. Every MGH spacetime is in a sense determined by the “asymptotic states” of the corresponding level surfaces, rather than the embedding data of some Cauchy surface. The Wick rotation-rescaling mechanism is ultimately based on the fact that MGH spacetimes (of different curvatures) can be associated in such a way that the intrinsic geometry of these level surfaces does not depend on the curvature, up to some scaling factor.

Acknowledgement. F. B. gratefully acknowledges partial support by A.N.R. through the project GEODYCOS.

Notation

- \mathcal{A} : r -uple of cone-angles at the marked points V of \hat{S} ;
- AdS: acronym of “Anti de Sitter”;
- α, β : horizontal and vertical rescaling functions;
- B_λ, B^λ : hyperbolic or AdS bending cocycle;
- B_λ^L : left-quake cocycle;
- $\beta: \tilde{\mathcal{T}}(S) \rightarrow \mathcal{T}(S)$, the natural retraction;
- $\beta^\#$: the quake-flow on $\mathcal{ML}_c^\#(S)$;
- $B(h), W(h)$: black or white holes in $\Omega(h)$;
- $(d_{\mathcal{P}}^\lambda, h_{\mathcal{P}}^\lambda), (d_{\mathbb{H}}^\lambda, h_{\mathbb{H}}^\lambda)$: developing map and holonomy of $\mathfrak{m}_{\mathcal{P}}(\lambda)$, and of its hyperbolic H -hull;
- C_i : see $\bar{\Sigma}, \Sigma$;
- $\mathfrak{C}(S)$: the subspace of $\tilde{\mathcal{H}}(S)$ of hyperbolic surfaces of finite area and such that all boundary components of the completion are closed geodesics;
- \mathcal{D} : a pant decomposition of $\bar{\Sigma}$;
- $\mathcal{E} = H/h, \mathcal{E}_\infty$: a crown and the ideal part of H ;
- F^c : see $\tilde{\mathcal{H}}(S)$;
- $[\hat{F}]$: the image of $[F]$ via the natural retraction β ;
- $\text{Gr}_\lambda(F)$: the grafting of F along λ ;
- $\tilde{\mathcal{H}}(S)$: the space of non-necessarily complete hyperbolic surfaces F homeomorphic to S , such that their completions F^c have geodesic boundary;
- $\mathcal{H}(S)$: the subspace of $\tilde{\mathcal{H}}(S)$ of complete surfaces;
- I : the marked measure spectrum;
- $\mathcal{K}: \tilde{\mathcal{T}}(S) \rightarrow \mathcal{T}_c(S)$, the convex-core map;

- $\overline{\mathcal{K}}(F)$, $\mathcal{K}(F)$: the convex core of the hyperbolic surface F , and its interior;
- $\mathcal{K}(h)$: the convex core of $\Omega(h)$;
- $\mathcal{K}(C)$: the convex core of the standard AdS spacetime $\mathcal{Y}(C)$;
- L : the marked length spectrum;
- \mathcal{L}_S , \mathcal{L}_W : the simplicial and weighted parts of a measured geodesic lamination $\lambda = (\mathcal{L}, \mu)$;
- $\lambda = (\mathcal{L}, \mu)$: a measured geodesic lamination;
- MGH: acronym of “maximal globally hyperbolic”;
- $\mathcal{MGH}_\kappa(S)$: the Teichmüller-like space of MGH spacetimes of constant curvature $\kappa = 0, \pm 1$, and with a complete Cauchy surface homeomorphic to S ;
- $\mathcal{ML}(S)$: the space of measured geodesic laminations on surfaces of $\tilde{\mathcal{T}}(S)$;
- \mathcal{ML}_g : the space of measured geodesic laminations on a closed surface of genus $g \geq 2$;
- $\mathcal{ML}_c(S)$: the subspace of $\mathcal{ML}(S)$ over $\mathcal{T}_c(S)$;
- $\mathcal{ML}_c(S)^0$: the subspace of $\mathcal{ML}_c(S)$ of laminations that do not enter any cusp;
- $\mathcal{ML}(F)$, $\mathcal{ML}_c(F)$: fibres of the projections \mathfrak{p} , \mathfrak{p}_c ;
- $\mathcal{ML}_c^\#(S)$, $I^\#$: the space of enhanced laminations and the enhanced marked measure spectrum;
- $\mathcal{MGH}(h)$: the set of all spacetimes in $\mathcal{MGH}_{-1}(S)$ with prescribed holonomy h ;
- $\mathfrak{m}_\mathcal{P}: \mathcal{ML}(S) \rightarrow \mathcal{P}(S)$, $\mathfrak{m}_\kappa: \mathcal{ML}(S) \rightarrow \mathcal{MGH}_\kappa(S)$: the materialization maps;
- $\mathcal{MGH}_c(h)$: the set of all spacetimes in $\mathfrak{m}_{-1}(\mathcal{ML}_c(S))$ with prescribed holonomy h ;
- $\Omega(h)$, $\tilde{\Omega}(h)$: the largest causal AdS spacetime with prescribed holonomy h , and its Lorentzian universal covering;
- $\mathfrak{p}: \mathcal{ML}(S) \rightarrow \tilde{\mathcal{T}}(S)$, $\mathfrak{p}_c: \mathcal{ML}_c(S) \rightarrow \tilde{\mathcal{T}}_c(S)$: the natural projections;
- $\mathcal{P}(C)$: the past part of $\mathcal{Y}(C)$;
- $\mathcal{P}(S)$: the Teichmüller-like space of complex projective structures on S ;
- R_λ : the ray of measured geodesic laminations determined by the lamination λ ;
- S : a fixed surface a finite type of the form $S = \hat{S} \setminus V$, where \hat{S} is closed of genus g , V is a set of r marked points on \hat{S} , and $2 - 2g - r < 0$;
- \hat{S} : see S ;
- $\overline{\Sigma}$, Σ : the compact surface with boundary components C_1, \dots, C_r , obtained from \hat{S} by removing r small open disks around each marked point in V , the interior of $\overline{\Sigma}$;
- \mathfrak{C} : the set of isotopy classes of essential simple closed curves on $\overline{\Sigma}$, not isotopic to any boundary components;

- $\mathfrak{S}_{\mathcal{D}}$: a distinguished set of simple closed curves on $\bar{\Sigma}$ associated to the pant decomposition \mathcal{D} ;
- $\sigma(\lambda): V_{\mathcal{H}} \rightarrow \{\pm 1\}$: the signature of a lamination λ ;
- \mathcal{T}_g : the classical Teichmüller space of a closed surface of genus g ;
- $\mathcal{T}_{g,r}$: the classical Teichmüller space of complete hyperbolic surfaces of finite area, genus g with r punctures;
- $\mathcal{T}(S)$: the restricted Teichmüller space associated to $\mathcal{H}(S)$;
- $\mathcal{T}_{\mathfrak{c}}(S)$: the restricted Teichmüller space associated to $\mathfrak{C}(S)$;
- $\tilde{\mathcal{T}}(S) = \tilde{\mathcal{H}}(S)/\text{Diff}^0$: the Teichmüller space associated to $\tilde{\mathcal{H}}(S)$;
- $\tilde{\mathcal{T}}^\theta(S)$: the subspace of $\tilde{\mathcal{T}}(S)$ of surfaces F such that $\beta(F)$ has fixed type θ ;
- $\mathcal{T}_{\mathfrak{c}}(S)^\#, L^\#$: the enhanced Teichmüller space and the enhanced length spectrum;
- $T_\kappa^\lambda, \Sigma_\kappa^\lambda$: the cosmological time and the initial singularity of Y_κ^λ ;
- $\tau: M \rightarrow (0, +\infty]$: the cosmological function;
- V : the set of marked points on \hat{S} ;
- $V = V_{\mathcal{P}} \cup V_{\mathcal{H}}, r_{\mathcal{P}}, r_{\mathcal{H}}$: the partition by types of the ends of $F \in \tilde{\mathcal{T}}(S), r_* = |V_*|$;
- \mathbb{X}_κ : the 3-dimensional Minkowski ($\kappa = 0$), de Sitter ($\kappa = 1$), Anti de Sitter ($\kappa = -1$) spaces;
- $Y_\kappa^\lambda, \mathcal{U}_\kappa^\lambda, (d_\kappa^\lambda, h_\kappa^\lambda)$: the spacetimes associated to the lamination λ via the maps m_κ , its Lorentzian universal covering, its developing map and holonomy;
- $\mathcal{Y}(C)$: the standard AdS spacetime given as the Cauchy development of an achronal meridian of $\partial\mathbb{X}_{-1}$.

2 3-dimensional gravity

2.1 General background

For the basic notions of global Lorentzian geometry and causality we refer for instance to [14], [44].

An $(n+1)$ spacetime consists of an $(n+1)$ -manifold M equipped with a Lorentzian metric h and with a *time orientation*, so that the *causal past/future* of every event $p \in (M, h)$ is determined. We also stipulate that M is oriented.

Roughly speaking, the general problem of gravity can be stated as follows. Given an $(n+1)$ -manifold M , a symmetric $(0, 2)$ -Tensor T on M and a constant Λ (called the *cosmological constant*), find out all spacetimes (M, h) such that:

- (a) The metric h satisfies the *Einstein equation*

$$\text{Ric}_h + \left(\Lambda - \frac{1}{2}R_h\right)h = T$$

where Ric_h is the Ricci tensor of h and R_h is its scalar curvature.

(b) The global *causal structure* of (M, h) satisfies determined conditions.

These spacetime structures are considered up to diffeomorphism of M that preserves the tensor T . Both features of the tensor T and of the causality conditions are determined by physical (even logical) considerations. Normally, they also impose some constraints on the topology of M . Requirements in (a) and (b) are basically of independent nature.

The *pure gravity* case is when $T = 0$. In such a case, the solutions of the Einstein equation coincide with the so called *Einstein metrics*: $\text{Ric}_h = \frac{2\Lambda}{n-1}h$.

The basic causality condition is that (M, h) is *chronological (causal)*, that is, it does not contain any *closed* timelike (causal) curve c . A curve is said timelike (causal) if its velocity field $v(t)$ is timelike (nowhere spacelike): $h(v(t), v(t)) < 0$ ($h(v(t), v(t)) \leq 0$).

The strongest causality condition is that (M, h) contains a *Cauchy surface* S ; this means that S is a *spacelike* hypersurface of M (the restriction of h to S is Riemannian), such that every causal inextensible line of (M, h) intersects S exactly once. In this case we say that (M, h) is *globally hyperbolic*. If (M, h) is globally hyperbolic then M turns out to be a product manifold $M \cong S \times \mathbb{R}$ so that (up to diffeomorphism of M) the Cauchy surface S coincides with $S \times \{0\}$, and every slice $S \times \{t\}$ is h -spacelike (indeed we can also require that every such slice is a Cauchy surface of (M, h)). Such a picture is coherent with the intuitive idea of *a space evolving in time*. Globally hyperbolic spacetimes naturally arise as *dependence domains* $(D(S), h|_{D(S)})$ of spacelike hypersurfaces S in arbitrary spacetimes (M, h) ; S turns to be a Cauchy surface of $D(S)$. Hence, globally hyperbolic spacetimes constitute a fundamental sector of gravity theory.

2.2 (2 + 1)-spacetimes

3D gravity is much simpler than the higher dimensional case because in dimension three the Riemann tensor is determined by the Ricci tensor. In particular 3D Einstein metrics actually have *constant (sectional) curvature*. The sign of the curvature coincides with the sign of the cosmological constant. We will be mainly concerned with (2 + 1) globally hyperbolic Einstein spacetimes (M, h) (*i.e.* of constant curvature κ).

We recall two possible ways of studying such spacetimes. The first *analytic* one is based on the important fact that the germ of the metric h at a Cauchy surface S determines, in a sense that we will make precise, the whole spacetime. This leads to consider the pairs (g, b) of a Riemannian metric, g , on the surface S and a g -symmetric endomorphism, b , of TS , that verify the Gauss–Codazzi equation

$$\begin{aligned}d^\nabla b &= 0, \\ \det b &= \kappa - \kappa_g\end{aligned}$$

where d^∇ is the differential with respect to the Levi-Civita connection of g , κ is a constant and κ_g is the Gauss curvature of g .

It is possible to associate with such a pair (g, b) a Lorentzian metric h on $M = S \times \mathbb{R}$ of constant curvature κ , such that $S = S \times \{0\}$ is a Cauchy surface, the *first fundamental form* of S in (M, h) is g and the *shape operator* is b . (Recall that the shape operator of a spacelike surface F in a Lorentzian or Riemannian manifold M is the endomorphism of TF that coincides with the covariant derivative of the normal field of F in M .) A priori the pair (g, b) determines only the germ of h around $S \times \{0\}$. On the other hand, it is proved in [32] that there exists a unique (up to isometry) such globally hyperbolic spacetime (M_{\max}, h_{\max}) that is *maximal* in the following sense:

Given any globally hyperbolic spacetime (M, h) as above, there exists an isometric embedding $(M, h) \rightarrow (M_{\max}, h_{\max})$ that is the identity on $S \times \{0\}$ (and preserves the orientations).

At first sight, this definition of “maximality” involves the choice of a Cauchy surface (i.e. $S \times \{0\}$). On the other hand, one can see that it is equivalent to the following one:

Every isometric embedding of (M_{\max}, h_{\max}) into an Einstein spacetime (N, k) that sends any Cauchy surface of M_{\max} onto a Cauchy surface of N actually is a global isometry.

This last property gives a good definition of the class of *maximal globally hyperbolic (MGH) Einstein spacetimes*, that has an intrinsic sense, not depending on the analytic approach we are outlining. It is reasonable to restrict to this class in order to get a classification.

Continuing with the analytic approach, a well-defined map eventually associates to every pair (g, b) as above the (isotopy class of the) maximal globally hyperbolic spacetime $(M_{\max}, h_{\max})_{(g,b)}$. Such map is surjective, but not injective. In fact it establishes a bijective correspondence between pairs (g, b) and spacetimes with a marked Cauchy surface. To get rid of this excess of degrees of freedom, some additional condition on (g, b) has to be imposed, possibly translating some geometric property of the Cauchy surface embedding. A widely investigated possibility consists in requiring that the trace of b is constant, that is, $S \times \{0\}$ is a surface of *constant mean curvature*.

The second *geometric* approach makes use of the (G, X) -structure technology. Indeed any $(2 + 1)$ Einstein spacetime M is a $(\mathbb{X}_\kappa, \text{Isom}(\mathbb{X}_\kappa))$ -manifold, where \mathbb{X}_κ is a suitable *isotropic model of constant curvature* κ .

Denote by \tilde{M} a universal covering of M . A very general “analytic continuation” procedure allows to associate to every (G, X) -manifold, M , a *compatible pair* (d, h) , where d is a *developing map*, that is, a local isomorphism

$$d: \tilde{M} \rightarrow X$$

and h is a *holonomy representation*

$$h: \pi_1(M) \rightarrow G$$

such that (π_1 -equivariance)

$$d(\gamma x) = h(\gamma)d(x)$$

(where $\pi_1(M)$ is identified with the covering transformation group of \tilde{M}). The developing map is determined up to post-composition by any element of G , whereas the holonomy is determined up to conjugation by the same element.

Conversely, a local diffeomorphism $d: \tilde{M} \rightarrow X$ equivariant with respect to a representation $h: \pi_1(M) \rightarrow G$ produces a well-defined (G, X) -structure on M .

In this chapter we will mainly focus on this second geometric approach. For this reason we will briefly recall the principal features of the isotropic models of constant curvature κ , that we will normalize to be $\kappa = 0, 1, -1$.

Minkowski space. The isotropic model of flat spacetimes, \mathbb{X}_0 , is the Minkowski space, that is, \mathbb{R}^3 equipped with the flat metric $-dx_0^2 + dx_1^2 + dx_2^2$. Isometries of \mathbb{X}_0 are affine transformations whose linear part preserves the Minkowski product (that is $\text{Isom}(\mathbb{X}_0) = \text{O}(2, 1) \ltimes \mathbb{R}^3$). We consider the time-orientation on \mathbb{X}_0 such that the x_0 -component of future-directed timelike vectors is positive. The set of future directed unit timelike vectors is a hypersurface of \mathbb{X}_0 that inherits from \mathbb{X}_0 a Riemannian metric. This is the hyperboloid model of the hyperbolic plane \mathbb{H}^2 . The isometric action of $\text{SO}^+(2, 1)$ on it induces an identification between $\text{SO}^+(2, 1)$ and $\text{PSL}(2, \mathbb{R}) \cong \text{Isom}^+(\mathbb{H}^2)$ (by using also the Poincaré half-plane model of \mathbb{H}^2). The main advantage of the hyperboloid model is that geodesics are just obtained by intersecting \mathbb{H}^2 with timelike planes. In particular, the duality between linear planes and linear straight lines given by the orthogonality relation induces an identification between the set of geodesics of \mathbb{H}^2 and the set of unoriented spacelike directions of \mathbb{X}_0 . The projection of \mathbb{H}^2 in the projective plane $\mathbb{P}(\mathbb{R}^3)$ is injective and the image is the set of timelike directions. Notice that in this projective (Klein) model geodesics are just projective lines. Moreover, the set of lightlike directions is the boundary of \mathbb{H}^2 and the endpoints of a geodesic l in \mathbb{H}^2 are the two lightlike directions contained in the plane of \mathbb{X}_0 containing l .

By using the 4-dimensional Minkowski space in a similar way, we get the different models of the hyperbolic space \mathbb{H}^3 .

De Sitter space. The set of unit spacelike vectors in 4-dimensional Minkowski space is a Lorentzian submanifold of constant curvature 1. It is called de Sitter spacetime and will be denoted by $\hat{\mathbb{X}}_1$. The isometric action of $\text{SO}(3, 1)$ on $\hat{\mathbb{X}}_1$ shows that this model is isotropic and that its isometry group coincides with $\text{SO}(3, 1)$. Also in this model geodesics are obtained by intersecting $\hat{\mathbb{X}}_1$ with linear planes of Minkowski space. In particular spacelike geodesics are closed with length equal to 2π , whereas timelike geodesics are embedded lines with infinite length.

It is often convenient to consider the projection of $\hat{\mathbb{X}}_1$ into the projective space $\mathbb{P}(\mathbb{R}^4)$. Notice that the image, \mathbb{X}_1 , is the set of spacelike directions, that is, it is the exterior of \mathbb{H}^3 into $\mathbb{P}(\mathbb{R}^4)$. Clearly the projection $\hat{\mathbb{X}}_1 \rightarrow \mathbb{X}_1$ is a 2-to-1 covering, so \mathbb{X}_1

is not simply connected. On the other hand, since $\mathbb{X}_1 = \hat{\mathbb{X}}_1/\{\pm\text{Id}\}$, and $\{\pm\text{Id}\}$ is the center of $\text{SO}(3, 1)$, also \mathbb{X}_1 is an isotropic model of the de Sitter geometry. Its isometry group is $\text{SO}(3, 1)/\{\pm\text{Id}\} \cong \text{SO}^+(3, 1)$. An advantage in using this model is that \mathbb{X}_1 and \mathbb{H}^3 share the same asymptotic boundary and their isometry groups actually coincide. By means of the duality between geodesic planes of \mathbb{H}^3 and spacelike directions of Minkowski space, \mathbb{X}_1 can be regarded as the set of unoriented geodesic planes of \mathbb{H}^3 .

Anti de Sitter space. Consider on \mathbb{R}^4 a scalar product η with signature $(2, 2)$. Then, the set of unit timelike vectors is a Lorentzian submanifold $\hat{\mathbb{X}}_{-1}$ of constant curvature -1 . Let \mathbb{R}^4 be identified with the set of 2×2 matrices, and consider the form η such that $\eta(X, X) = -\det X$. The signature of η is $(2, 2)$, so an explicit model of $\hat{\mathbb{X}}_{-1}$ is $\text{SL}(2, \mathbb{R})$ equipped with its Killing form. The isometric action of $\text{SL}(2, \mathbb{R}) \times \text{SL}(2, \mathbb{R})$ on $\text{SL}(2, \mathbb{R})$ by left and right multiplication shows that $\hat{\mathbb{X}}_{-1}$ is isotropic and that its isometry group is $\text{SL}(2, \mathbb{R}) \times \text{SL}(2, \mathbb{R})/(-\text{Id}, -\text{Id})$.

As in the previous case, the projection of $\hat{\mathbb{X}}_{-1}$ into the projective space $\mathbb{P}(\mathbb{R}^4)$ is a 2-to-1 covering map on an open set \mathbb{X}_{-1} of $\mathbb{P}(\mathbb{R}^4)$. Thus, $\text{PSL}(2, \mathbb{R})$ turns out to be an explicit model of this space. Since the covering transformations of $\hat{\mathbb{X}}_{-1} \rightarrow \mathbb{X}_{-1}$ are $\pm\text{Id}$ it follows that \mathbb{X}_{-1} inherits from $\hat{\mathbb{X}}_{-1}$ an isotropic Lorentzian metric of constant curvature -1 . The isometry group of \mathbb{X}_{-1} turns out to be $\text{PSL}(2, \mathbb{R}) \times \text{PSL}(2, \mathbb{R})$.

Topologically \mathbb{X}_{-1} is a solid torus and its boundary in $\mathbb{P}(\mathbb{R}^4)$ can be identified with the projective classes of rank 1 matrices. The Segre embedding produces a double foliation on $\partial\mathbb{X}_{-1}$ by projective lines (actually it induces a product structure $\partial\mathbb{X}_{-1} = \mathbb{P}^1 \times \mathbb{P}^1$). Isometries of \mathbb{X}_{-1} extend to the boundary: left multiplication preserves each leaf of the left foliation and permutes those of the right foliation, whereas right multiplication preserves each leaf of the right foliation and permutes those of the left foliation. Notice that the product structure on the boundary can be regarded as a conformal Lorentzian structure.

Geodesics and geodesic planes of \mathbb{X}_{-1} are the intersection of \mathbb{X}_{-1} with projective lines and projective planes of $\mathbb{P}(\mathbb{R}^4)$. In particular projective lines contained in \mathbb{X}_{-1} are timelike geodesics of length π , projective lines tangent to the boundary are lightlike lines and projective lines intersecting the boundary in two points are spacelike geodesics of infinite length. Notice that spacelike geodesics are determined by their endpoints on the boundary. Conversely, given two points on the boundary that do not lie on the same left nor right leaf, there exists a unique spacelike geodesic connecting them. Moreover, since lightlike rays are projective lines tangent to $\partial\mathbb{X}_{-1}$, the lightlike cone at $x \in \mathbb{X}_{-1}$ is the tangent cone from x to $\partial\mathbb{X}_{-1}$.

Projective planes intersecting \mathbb{X}_{-1} along compression disks are spacelike planes and turn out to be isometric to \mathbb{H}^2 . Points of \mathbb{X}_{-1} bijectively corresponds to spacelike planes via the duality induced by η between points of $\mathbb{P}(\mathbb{R}^4)$ and projective planes. Namely, the plane dual to a point $x \in \mathbb{X}_{-1}$ is the projective plane $P(x)$ containing the contact conic between the quadric $\partial\mathbb{X}_{-1}$ and the tangent cone with vertex at x .

There is a geometric interpretation of such a duality: given a point $x \in \mathbb{X}_{-1}$, its dual plane $P(x)$ is the set of points at distance $\pi/2$ from x along some timelike geodesic.

Conversely, given a spacelike plane, its normal geodesics intersect at the dual point of the plane. Given a spacelike geodesic line l , the points x such that $l \subset P(x)$ form another spacelike line l^* , that is the *dual* geodesic of l .

3 The space $\mathcal{ML}(S)$

This section is entirely set in the framework of (2-dimensional) *hyperbolic geometry*, and several facts that we are going to recall are well known. However, we will give later a new insight (if not an outline of a foundation) to many constructions and concepts in terms of *Lorentzian geometry*.

Let us fix once and for all some *base* surfaces that will support several geometric structures:

$$(\hat{S}, V)$$

is a compact closed oriented surface of genus $g \geq 0$, with a set of $r \geq 0$ *marked points* $V = \{p_1, \dots, p_r\}$, and

$$S = \hat{S} \setminus V.$$

$\bar{\Sigma}$ is obtained by removing from \hat{S} a small open disk around each point p_j . Hence $\bar{\Sigma}$ is compact with r boundary components C_1, \dots, C_r . We denote by Σ the interior of $\bar{\Sigma}$. We also fix a continuous map

$$\bar{\phi}: \bar{\Sigma} \rightarrow \hat{S}$$

such that for every j , $\bar{\phi}(C_j) = p_j$, and the restriction $\phi: \Sigma \rightarrow S$ is an oriented diffeomorphism that is the identity outside a regular neighbourhood of the boundary of $\bar{\Sigma}$. In this way, we will often tacitly identify S and Σ . We will also assume that S is *not elementary*, that is, its fundamental group is *non-Abelian*, equivalently $2 - 2g - r < 0$. Such a surface S is said to be *of finite type*.

3.1 The Teichmüller space $\tilde{\mathcal{T}}(S)$

We denote by

$$\tilde{\mathcal{H}}(S)$$

the space of non-necessarily complete hyperbolic structures F on S whose completion F^c is a *complete hyperbolic surface with geodesic boundary*. Note that we do not require that the boundary components of F^c are closed geodesics. Denote by Diff^0 the group of diffeomorphisms of S homotopic to the identity. Set

$$\tilde{\mathcal{T}}(S) = \tilde{\mathcal{H}}(S)/\text{Diff}^0.$$

In other words, two hyperbolic structures in $\tilde{\mathcal{H}}(S)$ are identified up to isometries homotopic to the identity. This is the “full” Teichmüller space we will deal with.

3.2 The convex-core map

Let us point out some distinguished subspaces of $\tilde{\mathcal{T}}(S)$. We denote by

$$\mathcal{H}(S) \subset \tilde{\mathcal{H}}(S)$$

the space of *complete* hyperbolic structures on S (i.e. $F = F^c$). Hence every $F \in \mathcal{H}(S)$ can be realized as the quotient \mathbb{H}^2/Γ by a discrete, torsion free subgroup $\Gamma \subset \text{Isom}^+(\mathbb{H}^2) \cong \text{PSL}(2, \mathbb{R})$, isomorphic to $\pi_1(S)$. The corresponding quotient space

$$\mathcal{T}(S) \subset \tilde{\mathcal{T}}(S)$$

can be identified with the space of conjugacy classes of such subgroups of $\text{PSL}(2, \mathbb{R})$. Further,

$$\mathcal{C}(S) \subset \tilde{\mathcal{H}}(S)$$

denotes the set of hyperbolic structures F of *finite area and such that all boundary components of F^c are closed geodesics*, and

$$\mathcal{T}_c(S) \subset \tilde{\mathcal{T}}(S)$$

is the corresponding quotient space.

Clearly, if S is compact ($V = \emptyset$), then

$$\mathcal{T}_g := \mathcal{T}_c(S) = \mathcal{T}(S) = \tilde{\mathcal{T}}(S)$$

is the classical Teichmüller space.

In general, notice that

$$\mathcal{T}_{g,r} := \mathcal{T}(S) \cap \mathcal{T}_c(S)$$

is the space of complete hyperbolic structures on S with finite area. Via the Uniformization Theorem, $\mathcal{T}_{g,r}$ is isomorphic to the Teichmüller space of *conformal structures* on \hat{S} (i.e. on S that extend to \hat{S}) mod $\text{Diff}^0(\hat{S}, \text{rel } V)$. $\mathcal{T}(S)$ is isomorphic to the Teichmüller space of arbitrary conformal structures on S .

Proposition 3.1. *There is a natural isomorphism*

$$\mathcal{K} : \mathcal{T}(S) \rightarrow \mathcal{T}_c(S).$$

Basically $\mathcal{K}[F]$ coincides with $[\mathcal{K}(F)]$, where $\mathcal{K}(F)$ denotes the interior of the *convex core* $\bar{\mathcal{K}}(F)$ of F . Note that $\mathcal{K}(F)^c = \bar{\mathcal{K}}(F)$. This is a bijection because the convex core determines the whole complete surface.

Proposition 3.2. *There is a natural projection*

$$\beta : \tilde{\mathcal{T}}(S) \rightarrow \mathcal{T}(S)$$

such that $\beta|_{\mathcal{T}(S)} = \text{Id}$.

In fact the holonomy of any $[F] \in \tilde{\mathcal{T}}(S)$ is the conjugacy class of a faithful representation of $\pi_1(S)$ onto a discrete, torsion free subgroup Γ of $\text{PSL}(2, \mathbb{R})$, hence

$\beta([F]) = [\hat{F}]$, $\hat{F} = \mathbb{H}^2/\Gamma$. Finally, we can compound the maps of Propositions 3.1 and 3.2 to define the *convex-core map*

$$\mathcal{K} : \tilde{\mathcal{T}}(S) \rightarrow \mathcal{T}_c(S), \quad \mathcal{K}([F]) = \mathcal{K}([\hat{F}]).$$

In fact we can realize the representatives of the classes involved in such a way that

$$\overline{\mathcal{K}(\hat{F})} \subset F^c \subset \hat{F}$$

since F^c is a closed convex set in \hat{F} homotopically equivalent to S , and $\overline{\mathcal{K}(\hat{F})}$ is the minimal one with these properties. In what follows we will often make the abuse of identifying the classes with their representatives.

Partition by types

Proposition 3.3. *For every complete hyperbolic surface $F \in \mathcal{T}(S)$ there is a partition*

$$V = V_{\mathcal{P}} \cup V_{\mathcal{H}}$$

such that a point p belongs to $V_{\mathcal{P}}$ if and only if the following equivalent properties are satisfied:

- (1) F is of finite area at p (that is F has a cusp at p),
- (2) the holonomy of a circle in S surrounding p is of parabolic type.

On the other hand, p belongs to $V_{\mathcal{H}}$ if and only if the following equivalent properties are satisfied:

- (3) p corresponds to a boundary component of the convex core $\overline{\mathcal{K}(F)}$,
- (4) the holonomy of a circle in S surrounding p is of hyperbolic type.

The partition $V = V_{\mathcal{P}} \cup V_{\mathcal{H}}$, so that $r = r_{\mathcal{P}} + r_{\mathcal{H}}$, is called the *type* $\theta(F)$ of F . More generally, for every $F \in \tilde{\mathcal{T}}(S)$, set $\theta(F) = \theta(\hat{F})$. Any fixed type θ determines the subspace $\tilde{\mathcal{T}}^{\theta}(S)$ of hyperbolic structures that share that type. Varying θ we get the *partition by types* of $\tilde{\mathcal{T}}(S)$.

The fibers of the convex-core map. We want to describe the fibers of the convex-core map

$$\mathcal{K} : \tilde{\mathcal{T}}(S) \rightarrow \mathcal{T}_c(S).$$

Let $h \in \text{Isom}^+(\mathbb{H}^2)$ be of hyperbolic type. Denote by $\gamma = \gamma_h$ its invariant geodesic. Let P be the closed hyperbolic half-plane determined by γ such that the orientation of γ as the boundary of P is *opposite* to the sense of the translation $h|_{\gamma}$.

Definition 3.4. A *crown* is a hyperbolic surface of the form

$$\mathcal{E} = H/h$$

where H is the convex hull in P of an h -invariant closed subset, say $\mathcal{E}_{\infty} \subset \partial\mathbb{H}^2 = S_{\infty}^1$, contained in the frontier at infinity of P .

A crown \mathcal{E} is complete and has geodesic boundary made out of the union of the closed geodesic γ/h and complete open geodesics. $\mathcal{E} \setminus \partial\mathcal{E}$ is homeomorphic to $S^1 \times (0, +\infty)$.

Now, let $F \in \tilde{\mathcal{T}}^\theta(S)$ and $\bar{\mathcal{K}}(\hat{F}) \subset F^c \subset \hat{F}$ be as above. Then F^c is obtained by gluing a (possibly empty) crown at each boundary component C of $\bar{\mathcal{K}}(\hat{F})$, associated to some point $p \in V_{\mathcal{H}}$. This is possible if and only if, for every C we take a crown \mathcal{E} such that h is in the same conjugacy class of the \hat{F} -holonomy of the loop C , endowed with the boundary orientation of $\bar{\mathcal{K}}(\hat{F})$ (in other words, $\text{length}(\gamma/h) = l(C)$ and both orientations of $\bar{\mathcal{K}}(\hat{F})$ and \mathcal{E} are induced by the one of \hat{F}).

Lemma 3.5. *F is of finite area if and only if all crowns are. A crown \mathcal{E} is of finite area if and only if one of the following equivalent conditions is satisfied:*

- (1) \mathcal{E}_∞/h is a finite set.
- (2) \mathcal{E} has finitely many boundary components. For every boundary component l , the hyperbolic distance between each end of l and $\partial\mathcal{E} \setminus l$ is 0.

Furthermore, every crown \mathcal{E} (every $F \in \tilde{\mathcal{T}}(S)$) is the union of exhaustive sequences of increasing sub-crowns $\mathcal{E}_n \subset \mathcal{E}$ (sub-surfaces $F_n \subset F$) of finite area such that $\mathcal{E}_{n,\infty} \subset \mathcal{E}_\infty$.

In fact if \mathcal{E}_∞/h is finite, then the area of \mathcal{E} can be bounded by the sum of the area of a finite set of ideal triangles. If \mathcal{E}_∞ is not a finite set, then \mathcal{E} contains an infinite family of disjoint ideal triangles.

Finally, for every $F \in \mathcal{T}_c(S)$, the fiber $\mathcal{K}^{-1}(F)$ can be identified with the set of all possible patterns of $r_{\mathcal{H}}$ gluable crowns.

Parameters for $\mathcal{T}_c(S)$. The fibers of the convex-core map are in every sense “infinite dimensional”. On the other hand, the base space $\mathcal{T}_c(S)$ is tame and admits nice parameter spaces, that we are going to recall.

Length/twist parameters. This parameterization is based on a fixed *pant decomposition* \mathcal{D} of $\bar{\Sigma}$. It is well known that \mathcal{D} contains $2g+r-2$ pants obtained by cut/opening $\bar{\Sigma}$ at $3g-3+r$ (ordered) disjoint essential simple closed curves z_1, \dots, z_{3g-3+r} in Σ , not isotopic to any boundary component. Each of the r boundary components C_1, \dots, C_r of $\bar{\Sigma}$ is in the boundary of some pant. For every boundary component of a pant P_k , corresponding to some z_j , we also fix the unique “essential” arc ρ in P_k (shown in Figure 1) that has the endpoints on that component, and we furthermore select one among these endpoints, say e .

We use the following notation:

$$\mathbb{R}_+ = \{l \in \mathbb{R} \mid l > 0\}, \quad \bar{\mathbb{R}}_+ = \{l \in \mathbb{R} \mid l \geq 0\}.$$

First consider the simplest case of S having $(g, r) = (0, 3)$. In this case, set

$$\mathcal{T}_c(S) = \mathcal{T}_c(0, 3).$$

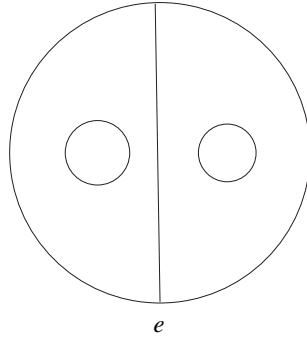


Figure 1. A pant and an arc ρ with endpoint e .

We have just one pant. Let us vary the types. If $r_{\mathcal{H}} = 3$, every hyperbolic structure is determined by the three lengths (l_1, l_2, l_3) of the geodesic boundary components. If $r_{\mathcal{H}} = 2$, the hyperbolic structure is determined by the corresponding two lengths, and it is natural to associate the value 0 to the boundary component that corresponds to the cusp, and so on. Eventually the octant

$$\overline{\mathbb{R}}_+^3 = \{(l_1, l_2, l_3) \mid l_j \geq 0\}$$

is a natural parameter space for the entire space $\mathcal{T}_c(0, 3)$. The canonical stratification by open cells of this closed octant corresponds to the partition by types.

In the general case, let $F \in \mathcal{T}_c(S)$. Then every pant of the topological decomposition \mathcal{D} is associated to a suitable hyperbolic pant $P_i = P_i(F)$ belonging to $\mathcal{T}_c(0, 3)$. Pant geodesic boundary components corresponding to some curve z_j have the same length, so that F^c is obtained by isometrically gluing the hyperbolic pants at the curves z_j . Summing up, F is of the form

$$F = F(l, t),$$

$$(l, t) = (l_{C_1}, \dots, l_{C_r}, l_{z_1}, \dots, l_{z_{3g-3+r}}, t_{z_1}, \dots, t_{z_{3g-3+r}}) \in \overline{\mathbb{R}}_+^r \times \mathbb{R}_+^{3g-3+r} \times \mathbb{R}^{3g-3+r}$$

where l_{C_i} (l_{z_j}) is the length of the geodesic boundary component (the simple closed geodesic) of F^c corresponding to C_i (z_j). The *twist* parameter $t_{z_j} \in \mathbb{R}$ specifies the isometric gluing at z_j as follows. For every hyperbolic pant, the arc ρ is uniquely realized by a geodesic arc orthogonal to this boundary. Then $F(l, 0)$ is the unique hyperbolic structure such that the selected endpoints e of such geometric ρ -arcs match by gluing. A generic $F(l, t)$ is obtained from $F(l, 0)$ by modifying the gluing as follows: if $t_{z_j} > 0$, the two sides at any geodesic line \tilde{z}_j in \mathbb{H}^2 over the closed geodesic z_j of $F(l, 0)$ translate by t_{z_j} along \tilde{z}_j on the *left* to each other. If $t_{z_j} < 0$, they translate on the *right* by $|t_{z_j}|$ (“left” and “right” are well defined and only depend on the orientation of S). Notice that twist parameters are well-defined real numbers since we are considering hyperbolic metrics up to diffeomorphisms isotopic to the identity.

We eventually realize in this way that

$$\overline{\mathbb{R}}_+^r \times \mathbb{R}_+^{3g-3+r} \times \mathbb{R}^{3g-3+r}$$

is a parameter space (depending on the choice of \mathcal{D}) for the space $\mathcal{T}_c(S)$. The product by $\mathbb{R}_+^{3g-3+r} \times \mathbb{R}^{3g-3+r}$ of the natural stratification by open cells of $\overline{\mathbb{R}}_+^r$, corresponds to the partition by types. Every cell has dimension

$$6g - 6 + 2r + r_{\mathcal{H}}$$

according to its type. The top-dimensional cell ($r_{\mathcal{H}} = r$) corresponds to hyperbolic surfaces without cusps. $\mathcal{T}_{g,r}$ is the lowest dimensional one. Cells that share the same $r_{\mathcal{H}}$ are isomorphic as well as the corresponding $\mathcal{T}_c^\theta(S)$. By varying \mathcal{D} we actually get an atlas for $\mathcal{T}_c(S)$ that gives it a *real analytic manifold with corner* structure.

Marked length spectrum. Length and twist parameters are of somewhat different nature; in fact we can deal with *length parameters only*. For every j , consider the “double pant” obtained by gluing the two pants of \mathcal{D} at z_j ; the simple closed curve z'_j obtained by gluing the respective two ρ arcs, and z''_j the curve obtained from z'_j via a Dehn twist along z_j .

Thus we point out a set of essential simple closed curves in S

$$\mathfrak{S}_{\mathcal{D}} = \{C_1, \dots, C_r, z_1, z'_1, z''_1, \dots, z_{3g-3+r}, z'_{3g-3+r}, z''_{3g-3+r}\} \quad (3.1)$$

and for every F we take the length of the corresponding simple closed geodesics. In this way we get an *embedding*

$$\mathcal{T}_c(S) \subset \overline{\mathbb{R}}_+^r \times \mathbb{R}_+^{9g-9+3r}.$$

This is the projection onto a finite set of factors of the *marked length spectrum* injection

$$L: \mathcal{T}_c(S) \rightarrow \overline{\mathbb{R}}_+^r \times \mathbb{R}_+^{\mathfrak{S}}$$

where \mathfrak{S} denotes the set of isotopy classes of essential simple closed curves in S , not isotopic to any boundary component.

For more details about the length/twist parameters and the length spectrum see for instance [37], [19].

Shear parameters. This is based on a fixed *topological ideal triangulation* T of (\hat{S}, V) , and works only if $V \neq \emptyset$. By definition T is a (possibly singular-, multi- and self-adjacency of triangles are allowed) triangulation of \hat{S} such that V coincides with the set of vertices of T . There are $6g - 6 + 3r$ edges $E_1, \dots, E_{6g-6+3r}$. The idea is to consider every triangle of T as a hyperbolic ideal triangle and realize hyperbolic structures F on S by isometrically gluing them at the geodesic edges, according to the pattern of edge-identifications given by T . By the way, T will be converted into a *geometric* ideal triangulation T_F of F . Let us decorate every edge E of T by a real

number $s(E)$ and get

$$s = (s(E_1), \dots, s(E_{6g-6+3r})) \in \mathbb{R}^{6g-6+3r}.$$

These *shear* parameters encode the isometric gluing at each E_j , and are of the same nature as the above twist parameters. Every edge of an ideal triangle has a distinguished point, say e , that is the intersection of the edge with the unique geodesic line emanating from the opposite ideal vertex and which is orthogonal to it. Then set $F = F(0)$ to be the unique hyperbolic structure such that the distinguished points match by gluing. A generic $F = F(s)$ is obtained from $F(0)$ by modifying the gluing according to the left/right moving rule as before. It turns out that all so obtained hyperbolic structures F belong to $\mathcal{T}_c(S)$, and all elements of $\mathcal{T}_c(S)$ arise in this way. For every s and every $p_i \in V$, set

$$s(p_i) = \sum_{E_j \in \text{Star}(p_i)} s(E_j).$$

We realize that

$$l_{C_i}(F(s)) = |s(p_i)|$$

so that, in particular, $p_i \in V_{\mathcal{P}}$ if and only if $s(p_i) = 0$ and this determines the type $\theta = \theta(F(s))$. This also shows that the map

$$\mathcal{S}: \mathbb{R}^{6g-6+3r} \rightarrow \mathcal{T}_c(S)$$

that sends s to $F(s)$, is *not* injective. For every $p_i \in V_{\mathcal{H}}$, define the sign $\epsilon_s(p_i)$ by

$$|s(p_i)| = \epsilon_s(p_i)s(p_i).$$

Then, the generic fiber $\mathcal{S}^{-1}(F)$ consists of $2^{r_{\mathcal{H}}}$ points, that is, \mathcal{S} realizes all the possible *signatures* $V_{\mathcal{H}} \rightarrow \{\pm 1\}$. For the geometric meaning of these signs, see below. For more details about shear parameters, see for instance [24].

The enhanced $\mathcal{T}_c(S)^{\#}$. Let us reflect a length/twist parameter space

$$\bar{\mathbb{R}}_+^r \times \mathbb{R}_+^{3g-3+r} \times \mathbb{R}^{3g-3+r}$$

of $\mathcal{T}_c(S)$ along its boundary components to get

$$\mathbb{R}^r \times \mathbb{R}_+^{3g-3+r} \times \mathbb{R}^{3g-3+r}.$$

This can be considered as a parameter space of the *enhanced Teichmüller space* $\mathcal{T}_c(S)^{\#}$, obtained by decorating each F with a signature

$$\epsilon: V_{\mathcal{H}} \rightarrow \{\pm 1\}.$$

Moreover, we stipulate that the sign ϵ_i associated to i has the meaning of selecting an orientation of the corresponding C_i , by the rule: $\epsilon_i = +1$ if and only if C_i is equipped with the boundary orientation.

To make the notation simpler, it is convenient to extend the signature ϵ to the whole of V by stating that $\epsilon_i = 1$ on $V_{\mathcal{P}}$. In this way an enhanced surface can be written as $(F, \epsilon_1, \dots, \epsilon_r)$ with $\epsilon_i \in \{\pm 1\}$ and $\epsilon_i = 1$ for i corresponding to a cusp of F .

In the same way one can show that the shearing parameters are global coordinates on $\mathcal{T}_c(S)^\#$, namely the map

$$\mathcal{S}^\# : \mathbb{R}^{6g-6+3r} \rightarrow \mathcal{T}_c(S)^\#$$

defined by $\mathcal{S}^\#(s) = (F(s), \text{sign}(s(p_1)), \dots, \text{sign}(s(p_n)))$ is a homeomorphism (see [38] for details).

There is a natural *forgetting projection*

$$\phi^\# : \mathcal{T}_c(S)^\# \rightarrow \mathcal{T}_c(S).$$

We can also define in a coherent way the *enhanced length spectrum*

$$L^\# : \mathcal{T}_c(S)^\# \rightarrow \mathbb{R}^r \times \mathbb{R}_+^{\mathfrak{S}}$$

by setting

$$l_{C_i}^\#(F, \epsilon) = \epsilon_i l_{C_i}(F)$$

on the peripheral loops, and $l_\gamma^\#(F, \epsilon) = l_\gamma(F)$ elsewhere. This is an injection of $\mathcal{T}_c(S)^\#$, and already the projection onto the usual finite set of factors as above is an embedding.

Remark 3.6. For each C_i , the enhanced length is a continuous function on $\mathcal{T}_c^\#(S)$. On the other hand, notice that ϵ_i coincides with the sign of $l_{C_i}^\#$, with the rule that the sign of 0 is 1.

3.3 The space of measured geodesic laminations

Definition 3.7. A *simple* (complete) geodesic in $F \in \tilde{\mathcal{T}}(S)$ is a geodesic which admits an arc length parametrization defined on the whole real line \mathbb{R} that either is injective (and we call its image a *geodesic line* of F), or such that its image is a simple closed geodesic. A *geodesic lamination* \mathcal{L} on F consists of

- (1) a *closed* subset L of F (the *support*);
- (2) a partition of L by simple geodesics (the *leaves*).

The leaves together with the connected components of $F \setminus L$ make a *stratification* of S .

Definition 3.8. Given a geodesic lamination \mathcal{L} on $F \in \tilde{\mathcal{T}}(S)$, a rectifiable arc k in F is *transverse* to the lamination if for every point $p \in k$ there exists a neighbourhood U of p in F such that $U \cap k$ intersects each connected component of $U \cap L$ in at most a point and each connected component of $U \setminus L$ in a connected set. A *transverse measure* μ on \mathcal{L} is the assignment of a Borel measure μ_k on each rectifiable arc k

transverse to \mathcal{L} (this means that μ_k assigns a non-negative *mass* $\mu_k(A)$ to every Borel subset of the arc, in a countably additive way, and $\mu_k(A)$ is finite if A is compact) in such a way that

- (1) the support of μ_k is $k \cap L$;
- (2) if $k' \subset k$, then $\mu_{k'} = \mu_k|_{k'}$;
- (3) if k and k' are homotopic through a family of arcs transverse to \mathcal{L} , then the homotopy sends the measure μ_k to $\mu_{k'}$.

Notice that we allow an arc k hitting the boundary of $F^{\mathbb{C}}$ to have infinite mass, that is, $\mu_k(k) = +\infty$.

Definition 3.9. A *measured geodesic lamination* on F is a pair $\lambda = (\mathcal{L}, \mu)$, where \mathcal{L} is a geodesic lamination and μ is a transverse measure on \mathcal{L} . For every $F \in \tilde{\mathcal{T}}(S)$, denote by $\mathcal{ML}(F)$ the set of measured geodesic laminations on F . Finally, let us define $\mathcal{ML}(S)$ to be the set of pairs (F, λ) , such that $F \in \tilde{\mathcal{T}}(S)$, and $\lambda \in \mathcal{ML}(F)$. We have a natural projection

$$p: \mathcal{ML}(S) \rightarrow \tilde{\mathcal{T}}(S).$$

Definition 3.10. Given $(F, \lambda) \in \mathcal{ML}(S)$, the *simplicial part* \mathcal{L}_S of \mathcal{L} consists of the union of the isolated leaves of \mathcal{L} . Hence \mathcal{L}_S does not depend on the measure μ . A leaf, l , is called *weighted* if there exists a transverse arc k such that $k \cap l$ is an atom of μ_k . The *weighted part* of λ is the union of all weighted leaves. It depends on the measure and it is denoted by $\mathcal{L}_W = \mathcal{L}_W(\mu)$.

Remark 3.11. The word “simplicial” mostly refers to the “dual” geometry of the initial singularity of the spacetimes that we will associate to every (F, λ) , see Section 4.

By Property (3) of the definition of a transverse measure, if l is weighted then for every transverse arc k the intersection of k with l consists of atoms of μ_k whose masses are equal to a positive number A independent of k . We call this number the *weight* of l . Since every compact set $K \subset F$ intersects finitely many weighted leaves with weight bigger than $1/n$, it follows that \mathcal{L}_W is a countable set. As L is the support of μ , then we have the inclusion $\mathcal{L}_S \subset \mathcal{L}_W(\mu)$.

Remark 3.12. There is a slightly different but equivalent definition of $\mathcal{ML}(S)$ that goes as follows. We can consider measured geodesic laminations $\lambda = (\mathcal{L}, \mu)$ of $F^{\mathbb{C}}$ requiring furthermore that

- (1) the boundary components of $F^{\mathbb{C}}$ are leaves of \mathcal{L} ;
- (2) every arc k hitting the boundary of $F^{\mathbb{C}}$ *necessarily* has infinite mass ($\mu_k(k) = +\infty$).

If a boundary component of $F^{\mathbb{C}}$ is isolated in \mathcal{L} we stipulate that it has *weight* $+\infty$. Notice that while a geodesic lamination on $F^{\mathbb{C}}$ can be regarded also as a particular lamination on the associated complete surface \hat{F} , Condition (2) ensures that such a

measured lamination cannot be extended beyond F^c . On the other hand, a lamination on F is not in general a lamination on \hat{F} .

Given any measured geodesic lamination λ of F we get a corresponding measured lamination $\hat{\lambda}$ of F^c by adding the (possibly $+\infty$ -weighted) boundary components to the lamination and keeping the same measure. Given $\hat{\lambda}$ in F^c we get λ in F just by forgetting the boundary leaves. In particular the empty lamination on F corresponds to the lamination on F^c reduced to its boundary components. Clearly this establishes a canonical bijection, hence an equivalent definition of $\mathcal{ML}(S)$. This second definition could sound at present somewhat unmotivated, so in this section we prefer to deal with F instead of F^c . However, we will see in Section 4 that it is the suitable one when dealing with the Lorentzian “materializations” of $\mathcal{ML}(S)$.

Marked measure spectrum. Similarly to the above length spectrum L , for every $F \in \tilde{\mathcal{T}}(S)$, one defines the *marked measure spectrum*

$$I: \mathcal{ML}(F) \rightarrow \overline{\mathbb{R}}_+^r \times \overline{\mathbb{R}}_+^S$$

where for every $\lambda \in \mathcal{ML}(F)$ and for every isotopy class γ of essential simple closed curves on S , $I_\gamma(\lambda)$ is the minimum of the *total variation* $\mu(c)$ of the “ λ -transverse component” of c , c varying among the representatives of s . The first r factors correspond as usual to the curves parallel to the boundary components.

Ray structure. Every $\lambda = (\mathcal{L}, \mu) \in \mathcal{ML}(F)$ determines a ray

$$R_\lambda = \{t\lambda = (\mathcal{L}, t\mu) \mid t \in [0, +\infty)\} \subset \mathcal{ML}(F)$$

where we stipulate that for $t = 0$ we take the empty lamination of F . If $I_\lambda \neq 0$, then $I(R_\lambda) = R_{I_\lambda}$, that is the corresponding ray in $\overline{\mathbb{R}}_+^r \times \overline{\mathbb{R}}_+^S$.

3.4 The subspace $\mathcal{ML}(S)$

We use the following notation:

$$\mathcal{ML}_c(S) = \{(F, \lambda) \in \mathcal{ML}(S) \mid F \in \mathcal{T}_c(S)\}$$

and

$$p_c: \mathcal{ML}_c(S) \rightarrow \mathcal{T}_c(S)$$

is the natural restriction of p with fibers $\mathcal{ML}_c(F)$.

For any $F \in \mathcal{T}_c(S)$, we denote by $\mathcal{ML}_c(F)^0$ the set of laminations on F that do not enter any cusp (namely the closure in F^c of the lamination support is compact). For a fixed type θ , we denote by

$$\mathcal{ML}_c(S)^\theta = \{(F, \lambda) \mid F \in \mathcal{T}_c^\theta(S), \lambda \in \mathcal{ML}_c(F)^0\}$$

and we still denote by p_c the restriction of the projection on every $\mathcal{ML}_c(S)^\theta$.

The spectrum I and the ray structure naturally restrict. In particular, if $\lambda \in \mathcal{ML}_c(F)^0$, and s surrounds a cusp of F , then $I_\lambda(s) = 0$. On the other hand, if s is parallel to a boundary component of F^c , then $I_\lambda(s) = 0$ if and only if the closure in F^c of the support L of the lamination does not intersect that boundary component.

The following proposition summarizes some basic properties of the fibers of p_c .

Proposition 3.13. *Let $\lambda \in \mathcal{ML}_c(F)$. Then:*

- (1) $F \setminus L$ has a finite number of connected components, and each component belongs to some $\tilde{T}(S')$, provided that we drop out the requirement that S' is non-elementary.
- (2) λ is the disjoint union of a finite set of minimal (with respect to inclusion) measured sublaminations. Every minimal sublamination is either compact (and coincides with the closure of any of its half-leaves) or consists of a geodesic line such that each sub half-line either enters a cusp or spirals towards a boundary component of F^c .
- (3) $\mathcal{L}_W = \mathcal{L}_S$.
- (4) Either any cusp or any boundary component has a neighbourhood U such that $\mathcal{L} \cap U = \mathcal{L}_S \cap U$.
- (5) For every arc c in F transverse to λ , $c \cap L$ is a union of isolated points and of a finite union of Cantor sets.

For a proof when $F \in \mathcal{T}_{g,r}$ we refer for instance to the body and the references of [21]. The details for the extension to the whole of $\mathcal{ML}_c(S)$ are given for instance in [29].

Remark 3.14. If the lamination $\hat{\lambda}$ of F^c corresponds to the lamination λ of F as in Remark 3.12, then a leaf spiraling towards a boundary component of F^c as in (2) is no longer a minimal sublamination of $\hat{\lambda}$.

Example 3.15. We refer to the above length/twist or shear parameters for $\mathcal{T}_c(S)$.

(a) Let $F = F(l, t)$. The union of simple closed geodesics of F corresponding to the curves z_j is a geodesic lamination $\mathcal{L} = \mathcal{L}_S$ of F . By giving each z_j an arbitrary real weight $w_j > 0$, we get $\lambda(w) \in \mathcal{ML}_c(F(l, t))^0$.

(b) Let $F = F(s)$. The 1-skeleton of the geometric ideal triangulation T_F (which is made out of geodesic lines) makes a geodesic lamination of F . Every geodesic line is a minimal sublamination. By giving each geodesic line an arbitrary weight $w_j > 0$, we get $\lambda(w) \in \mathcal{ML}_c(F(s))$. For such a $\lambda = \lambda(w)$,

$$I_{C_i}(\lambda) = \sum_{E_j \in \text{Star}(p_i)} w(E_j).$$

Lamination signatures. Let $\lambda \in \mathcal{ML}_c(F)$. Leaves of λ can spiral around a boundary component C_i in the *negative sense* or in *positive sense* with respect to the orientation of C_i . On the other hand two leaves that spiral around C_i must spiral in the same way (otherwise they would meet each other).

This rotation phenomenon is not possible with a cusp, as a spiralling geodesic returns after some time. This means that the only way to accumulate to a cusp is to go straightly on it.

This determines a *signature*

$$\sigma(\lambda): V_{\mathcal{H}} \rightarrow \{\pm 1\}$$

such that $\sigma_i(\lambda) = -1$ if and only if there are leaves of λ spiraling around the corresponding geodesic boundary C_i with a negative sense with respect to the boundary orientation. In other words, $\sigma_i(\lambda)$ is possibly equal to -1 only if $p_i \in V_{\mathcal{H}}$ and $I_{C_i}(\lambda) \neq 0$, $\sigma_i(\lambda) = 1$ otherwise. The signature depends indeed only on the lamination \mathcal{L} , not on the measure.

Remark 3.16. If $\lambda = \lambda(w)$ as in Example 3.15 (b), then σ_λ recovers the signs $\epsilon_s(p_i)$ already defined at the end of Section 3.2.

3.5 Enhanced bundle $\mathcal{ML}(S)^\#$

and measure spectrum Here we address the question to which extent the (restricted) marked measure spectrum determines $\mathcal{ML}_c(S)$. For example, this is known to be the case if we restrict to $\mathcal{ML}_{g,r}^0$, *i.e.* to laminations over $\mathcal{T}_{g,r}$ that do not enter the cusps (see for instance [21]). We want to extend this known result.

We have seen in Proposition 3.13 that a measured geodesic lamination λ on $F \in \mathcal{T}_c(S)$ is the disjoint union of a compact part, say λ_c (that is far away from the geodesic boundary of F^c and does not enter any cusp), with a part, say λ_b , made out of a finite set of weighted geodesic lines l_1, \dots, l_n whose ends leave every compact subset of F . Notice that $\sigma(\lambda) = \sigma(\lambda_b)$.

Let us take such a geodesic line l on $F \in \mathcal{T}_c(S)$. We can select a compact interval J in l such that both components of $l \setminus J$ definitely stay either within a small ε -neighbourhood of some boundary component of F^c , or within some cusp. The interval J can be completed to a simple arc c in \hat{S} with endpoints in V , just by going straight from each endpoint of J to the corresponding puncture. It is easy to see that the homotopy class with fixed endpoints of the arc c obtained in this way does not depend on the choice of J . For simplicity we refer to it as the “homotopy class” of l . We can also give the endpoints of c a sign ± 1 in the very same way we have defined the signature of a lamination on F (recall that the sign is always equal to 1 at cusps). We can prove

Lemma 3.17. *Given any $F \in \mathcal{T}_c(S)$, every homotopy class α of simple arcs on \hat{S} with endpoints on V , and every signature of the endpoints (compatible with the type of F)*

can be realized by a unique geodesic line l of F whose ends leave every compact set of F . Moreover, the members of a finite family of such geodesic lines are pairwise disjoint if and only if the signs agree on every common endpoint and there are disjoint representatives with endpoints on V of the respective homotopy classes. Analogously, a geodesic line does not intersect a compact lamination λ_c , if any representative in its homotopy class is disjoint from λ_c .

By using the lemma, we can prove (see [29]) the following statement.

Proposition 3.18. *Let $\lambda \in \mathcal{ML}_c(F)$. Then the support of λ_b is determined by the homotopy classes of its geodesic lines l_i and the signature of λ . More precisely, given any λ_c , every finite set of homotopy classes of simple weighted arcs on \hat{S} , with signed endpoints in V (provided that the signature is compatible with the type of F), admitting representatives that are pairwise disjoint and do not intersect λ_c , is uniquely realized by a lamination λ_b such that $\lambda = \lambda_b \cup \lambda_c \in \mathcal{ML}_c(F)$.*

Proposition 3.19. *Let $F, F' \in \mathcal{T}_c(S)$. Assume that F^c is without cusps (that is F belongs to the top-dimensional cell of $\mathcal{T}_c(S)$). Then there is a natural map*

$$\iota: \mathcal{ML}_c(F) \rightarrow \mathcal{ML}_c(F')$$

such that for every (isotopy class of) simple closed curve γ on S , we have

$$I_\gamma(\lambda) = I_\gamma(\iota(\lambda)).$$

Proof. Assume first that $\lambda = \lambda_c \in \mathcal{ML}_c(F)$. Then there is a unique $\lambda' = \lambda'_c \in \mathcal{ML}_c(F')$ with the same spectrum. Since we can embed F' in the double surfaces of $(F')^c$, say DF' , which is complete and of finite area, the measure spectrum of λ_c induces a measure spectrum of a unique lamination λ''_c on DF' (by applying the result on the spectrum in the special case recalled at the beginning of this section). Finally we realize that the compact support of λ''_c is contained in F' giving us the required λ'_c . So the map ι can be defined for laminations with compact support.

Given a general lamination $\lambda = \lambda_c \cup \lambda_b$, we can define λ'_c as before, while λ'_b is the unique lamination of F' (according to Proposition 3.18) that shares with λ_b the same homotopy classes, weights and signs at $V_{\mathcal{H}}(F')$ (since F is without cusps, signs are defined for each puncture of S). Notice that λ'_b is disjoint from λ'_c : in fact one can construct an isotopy of S sending the supports of λ_b and λ_c to the supports of λ'_b and λ'_c . Finally set $\iota(\lambda) = \iota(\lambda_b) \cup \iota(\lambda_c)$. \square

Corollary 3.20. *If both F and F' are without cusps, then the map ι is bijective. More generally, if only F is supposed to be without cusps, for every $\lambda' \in \mathcal{ML}_c(F')$, $\iota^{-1}(\lambda')$ consists of 2^k points, where k is the number of cusps of F' entered by λ' .*

In fact, for every F contained in $\mathcal{T}_c(S)$ (not necessarily in the top-dimensional cell), there is a natural action of $(\mathbb{Z}/2\mathbb{Z})^r$ on $\mathcal{ML}_c(F)$ determined as follows. Let

$\rho_i = (0, \dots, 1, \dots, 0)$, $i = 1, \dots, r$, be the i th element of the standard basis of $(\mathbb{Z}/2\mathbb{Z})^r$. Let $\lambda \in \mathcal{ML}_c(F)$. First define the new signature $\rho_i\sigma(\lambda)$ by setting

$$\begin{aligned} \rho_i\sigma(\lambda)(p_j) &= \sigma(\lambda)(p_j) \text{ if } i \neq j; \\ \rho_i\sigma(\lambda)(p_i) &= \sigma(\lambda)(p_i) \text{ if either } p_i \in V_{\mathcal{P}}(F) \text{ or } p_i \in V_{\mathcal{H}}(F) \text{ and } I_{C_i}(\lambda) = 0; \\ \rho_i\sigma(\lambda)(p_i) &= -\sigma(\lambda)(p_i), \text{ otherwise.} \end{aligned}$$

This naturally extends to every $\rho \in (\mathbb{Z}/2\mathbb{Z})^r$, giving the signature $\rho\sigma(\lambda)$. Finally set $\rho(\lambda) = \rho(\lambda_b) \cup \lambda_c$ where (according again to Proposition 3.18) $\rho(\lambda_b)$ is the unique lamination that shares with λ_b the homotopy classes and the weights, while its signature is $\rho\sigma(\lambda)$. Clearly the orbit of λ consists of 2^k points, where k is the number of p_i in $V_{\mathcal{H}}(F)$ such that $I_{C_i}(\lambda) \neq 0$. Finally $\iota^{-1}(\lambda')$ in Corollary 3.20 is just an orbit of such an action. We call the action on $\mathcal{ML}_c(F)$ of the generator ρ_i , the *reflection along C_i* (even if it could be somewhat misleading, as in some case it is just the identity).

If we restrict to the top-dimensional cell of $\mathcal{T}_c(S)$, \mathfrak{p}_c is a bundle and we can use the first statement of the corollary in order to fix a trivialization. The same fact holds for every restriction $\mathfrak{p}_c: \mathcal{ML}_c(S)^\theta \rightarrow \mathcal{T}_c(S)^\theta$, type by type. On the other hand, because of the last statement of the corollary, this is no longer true for the whole \mathfrak{p}_c . In order to overcome such a phenomenon, one can introduce the notion of *enhanced lamination*. An enhanced lamination on $F \in \mathcal{T}_c(S)$, is a couple (λ, η) where $\lambda \in \mathcal{ML}_c(F)$, and $\eta: V \rightarrow \{\pm 1\}$ is a *relaxed signature* such that

$$\begin{aligned} \eta_i &= \sigma_i(\lambda) \text{ if either } p_i \in V_{\mathcal{H}}(F) \text{ or } p_i \in V_{\mathcal{P}}(F) \text{ and } I_{C_i}(\lambda) = 0; \\ \eta_i &\text{ is arbitrary otherwise.} \end{aligned}$$

Notice that there are exactly 2^k relaxed signatures (λ, η) enhancing a given $\lambda \in \mathcal{ML}_c(F)$, where k is the number of cusps entered by λ . Clearly the above action of $(\mathbb{Z}/2\mathbb{Z})^r$ extends on enhanced laminations: $\rho(\lambda, \eta) = (\rho(\lambda), \rho(\eta))$, where $\rho(\eta)$ is uniquely determined by the above requirements and by the fact that $\rho\sigma(\lambda)$ possibly modifies $\sigma(\lambda)$ only on $V_{\mathcal{H}}$. In particular this holds for the generating reflections ρ_i .

We denote by $\mathcal{ML}_c^\#(F)$ the set of such (λ, η) on F . Finally we can define the *enhanced measure spectrum*

$$I^\#: \mathcal{ML}_c^\#(F) \rightarrow \mathbb{R}^r \times \mathbb{R}_+^\mathfrak{S}$$

such that

$$I_\gamma^\#(\lambda, \eta) = I_\gamma(\lambda)$$

for every $\gamma \in \mathfrak{S}$, and

$$I_{C_i}^\#(\lambda, \eta) = \eta_i I_{C_i}(\lambda)$$

for every peripheral loop C_i .

Here is the enhanced version of Proposition 3.19.

Corollary 3.21. *Let $F, F' \in \mathcal{T}_c(S)$. Then there is a natural bijection*

$$\iota^\#: \mathcal{ML}_c(F)^\# \rightarrow \mathcal{ML}_c(F')^\#$$

such that for every (isotopy class of) simple closed curve γ on S , we have

$$\Gamma_\gamma^\#((\lambda, \eta)) = \Gamma_\gamma^\#(t^\#(\lambda, \eta)).$$

Proposition 3.22. (i) *The enhanced spectrum $\mathbb{I}^\#$ realizes an embedding of every $\mathcal{ML}_c(F)^\#$ into $\mathbb{R}^r \times \overline{\mathbb{R}}_+^\mathfrak{S}$. Only the empty lamination goes to 0. The image is homeomorphic to $\mathbb{R}^{6g-6+3r}$. The image of $\mathcal{ML}_c(F)^\#,0$ (that is the set of enhanced laminations that do not enter any cusp) is homeomorphic to $\mathbb{R}^{6g-6+2r+r\mathfrak{K}}$.*

(ii) *For every pant decomposition \mathcal{D} of $\overline{\Sigma}$, consider the set of essential simple curves defined in (3.1):*

$$\mathfrak{S}_{\mathcal{D}} = \{C_1, \dots, C_r, z_1, z'_1, z''_1, \dots, z_{3g-3+r}, z'_{3g-3+r}, z''_{3g-3+r}\}.$$

The projection onto this finite set of factors is already an embedding of $\mathcal{ML}_c(F)^\#$. By varying \mathcal{D} we get an atlas of a PL structure on $\mathcal{ML}_c(F)^\#$ (i.e. on $\mathbb{R}^{6g-6+3r}$). Similar facts hold for the restriction to $\mathcal{ML}_c(F)^\#,0$.

(iii) *Finite laminations are dense in $\mathcal{ML}_c(F)^\#$ ($\mathcal{ML}_c(F)^\#,0$).*

(iv) *For every $F, F' \in \mathcal{T}_c(S)$, there is a canonical identification between the respective sets of finite enhanced measured geodesic laminations, and this extends to a canonical PL isomorphism between $\mathcal{ML}_c(F)^\#$ and $\mathcal{ML}_c(F')^\#$, which respects the ray structures. Similarly for $\mathcal{ML}_c(\cdot)^\#,0$.*

Proof. We will sketch the proof of this proposition. We assume that the result is known when S is compact (see [21], [37]). Thanks to Corollary 3.21 it is enough to deal with F without cusps. Then the double DF of F^c is compact, and we consider on DF the involution w that exchanges the two copies of F . Let us denote by $ML(F)$ the set of w -invariant measured geodesic laminations on DF that do not contain any component of ∂F^c . The idea is to construct a map

$$T: \mathcal{ML}_c(F) \rightarrow ML(F)$$

that is surjective and such that

(1) the fiber over a lamination $\lambda' \in ML(F)$ consists of 2^k laminations of $\mathcal{ML}_c(F)$, where k is the number of boundary components of F^c that intersect the support of λ' ;

(2) for every $\lambda \in \mathcal{ML}_c(F)$, the restrictions to \mathfrak{S} of both the spectrum of $T(\lambda)$ and of λ coincide.

The existence of the map T and the known results in the special cases recalled above will imply the proposition.

The construction of the map T runs as follows. Let $\lambda = \lambda_b \cup \lambda_c \in \mathcal{ML}_c(F)$ be decomposed as above. We define $T(\lambda_c)$ to be the double of λ_c in DF . For each leaf l_i of λ_b , take a “big” segment $J_i \subset l_i$, and complete it to a simple arc l'_i properly embedded in $(F^c, \partial F^c)$, obtained by going straight from each endpoint of J_i to the corresponding boundary component along an orthogonal segment. Clearly the double of l'_i is a simple non-trivial curve in DF , so there is a geodesic representative, say c_i , that is w -invariant and simple. Since for any other l_j of λ_b , $l_i \cap l_j = \emptyset$, the other such

c_j does not meet c_i . Moreover, since $l_i \cap \lambda_c = \emptyset$, the intersection of c_i with $T(\lambda_c)$ is also empty. So we can define

$$T(\lambda) = T(\lambda_c) \cup (c_1, a_1) \cup (c_2, a_2) \cup \dots \cup (c_n, a_n),$$

where a_i is the initial weight of l_i . This map satisfies (2) by construction; moreover, it follows from Corollary 3.20 that (1) holds for every λ' belonging to the image of T . The only point to check is that the map is surjective. The key remark is that for every $\lambda' \in ML(F)$, every leaf l hitting ∂F^c is necessarily closed. As it is w -invariant, then l is orthogonal to ∂F^c , and if l intersects ∂F^c twice, then it is closed. Suppose that l is a geodesic line, so that l meets ∂F exactly once. On the other hand, we know that the closure of l is a minimal sublamination λ'' , such that every leaf is dense in it. Thus if $l' \neq l$ is another leaf in λ'' , then it intersects ∂F^c in a point p . Since l is dense in λ'' , there is a sequence of points in $l \cap \partial F^c$ converging to p and this contradicts the assumption that l intersects ∂F^c once.

Thus a lamination in $ML(F)$ is given by the double of a compact lamination λ_c in F and of a finite number of weighted simple geodesics arcs in F hitting orthogonally ∂F^c . These arcs can be completed to give a family of simple arcs on \hat{S} with endpoints on V . Fix a signature on the endpoints of such arcs. Finally we can apply Proposition 3.18 to these data and we get a suitable $\lambda = \lambda_b \cup \lambda_c \in \mathcal{ML}_c(F)$ such that $T(\lambda) = \lambda'$. □

Finally we can define the map

$$p_c^\# : \mathcal{ML}_c(S)^\# \rightarrow \mathcal{T}_c^\#(S).$$

The total space is defined as the set of pairs

$$((F, \epsilon), (\lambda, \eta))$$

such that

- (1) $(F, \epsilon) = (F, \epsilon_1, \dots, \epsilon_r) \in \mathcal{T}_c(S)^\#$;
- (2) $(\lambda, \eta) = (\lambda, \eta_1, \dots, \eta_r) \in \mathcal{ML}_c(F)^\#$.

Clearly

$$\phi^\# \circ p_c^\# = p \circ \phi_{\mathcal{ML}}^\#$$

where $\phi_{\mathcal{ML}}^\#$ denotes the *forgetting projection* of $\mathcal{ML}_c(S)^\#$ onto $\mathcal{ML}_c(S)$. We are going to see that in fact p_c determines a *bundle* of enhanced laminations, that admits furthermore a natural *trivialization* \mathfrak{t} . It follows from the previous discussion that the image of $I^\#$ does not depend on the choice of F , hence $I^\#(S)$ is well defined. We want to define a natural bijection

$$\mathfrak{t} : \mathcal{T}_c^\#(S) \times I^\#(S) \rightarrow \mathcal{ML}_c(S)^\#.$$

Fix F_0 a base surface in the top-dimensional cell of $\mathcal{T}_c(S)$. For every $\xi \in I^\#(S)$ there is a unique $(\lambda_0(\xi), \eta_0(\xi)) \in \mathcal{ML}_c(F_0)^\#$ that realizes ξ . Moreover, in such a case $\eta_0(\xi) = \sigma(\lambda_0(\xi))$. For every $\epsilon = (\epsilon_1, \dots, \epsilon_r)$, let $\rho_\epsilon(\lambda_0(\xi), \eta_0(\xi)) \in \mathcal{ML}_c(F_0)^\#$

be obtained via the action of $\rho_\epsilon \in (\mathbb{Z}/2\mathbb{Z})^r$, where ρ_ϵ is the composition of some reflections ρ_j , one for each $\epsilon_j = -1$. Recall the bijection

$$i^\# : \mathcal{ML}_c(F_0)^\# \rightarrow \mathcal{ML}_c(F)^\#$$

and finally set

$$t(F, \epsilon, \xi) = (F, \epsilon, (i^\#)^{-1}(\rho_\epsilon(\lambda_0(\xi), \eta_0(\xi)))).$$

It follows from the previous discussion that t is a bijection. We stipulate that it is a homeomorphism, determining by the way a topology on $\mathcal{ML}_c(S)^\#$. Summing up, the map

$$p^\# : \mathcal{ML}_c(S)^\# \rightarrow \mathcal{T}_c(S)^\#$$

can be considered as a *canonically trivialized* fiber bundle having both the base space and the fiber (analytically or PL) isomorphic to $\mathbb{R}^{6g-6+3r}$. Different choices of the base surface F_0 lead to isomorphic trivializations, via isomorphisms that preserve all the structures. These trivializations respect the ray structures. When S is compact this specializes to the trivialized bundle $\mathcal{T}_g \times \mathcal{ML}_g \rightarrow \mathcal{T}_g$ mentioned in the Introduction.

Remark 3.23. The definition of t could appear a bit distressing at first sight. However the geometric meaning is simple. Given a spectrum of positive numbers, this determines the lamination up to choosing the way of spiraling towards the boundary components. If we give a sign to the elements of the spectrum corresponding to the boundary components, this allows to reconstruct the lamination by the following rule: if the sign is positive, the lamination spirals in the positive way, if the sign is negative the lamination spirals in the negative way *with respect to a fixed orientation of the boundary component*. In the non-enhanced set-up, we have stipulated to use the boundary orientation induced by the one of the surface. Since the elements of an enhanced Teichmüller space can be regarded as hyperbolic surfaces equipped with an (arbitrary) orientation on each boundary component, it seems natural to reconstruct the lamination from the spectrum $I^\#$ by means of such boundary component orientations.

This choice is suitable in view of the *earthquake flow* that we are going to define on $\mathcal{T}_c(S)^\#$.

3.6 Grafting, bending, earthquakes

Let $(F, \lambda) \in \mathcal{ML}(S)$. *Grafting* (F, λ) produces a deformation $\text{Gr}_\lambda(F)$ of F in $\mathcal{P}(S)$, the Teichmüller space of *complex projective structures* (i.e. $(S^2, \text{PSL}(2, \mathbb{C}))$ -structures) on S .

3-dimensional *hyperbolic bending* produces the H -hull of $\text{Gr}_\lambda(F)$, that is, in a sense, its “holographic image” in \mathbb{H}^3 .

The *left (right) earthquake* produces (in particular) a new element $\beta_\lambda^L(F)$ ($\beta_\lambda^R(F)$) in $\tilde{\mathcal{T}}(S)$.

We will see in Section 4 how these constructions are *materialized* within the *canonical Wick rotation-rescaling theory* for MGH Einstein spacetimes. For example, grafting is eventually realized by the *level surfaces of the cosmological times*; earthquakes are strictly related to the *Anti de Sitter bending* procedure.

Here we limit ourselves to recalling a few details about earthquakes, purely in terms of hyperbolic geometry.

Features of arbitrary (F, λ) . In such a general case, the leaves of λ possibly enter the crowns of F . If F is of finite area (see Lemma 3.5), basically the conclusions of Proposition 3.13 still hold. The only new fact is that possibly there is a finite number of isolated geodesic lines of λ having at least one end converging to a point of some \mathcal{E}_∞ .

The situation is quite different if F is of infinite area. The set of isolated geodesic lines of λ that are not entirely contained in one crown \mathcal{E} is always finite. On the other hand, (1), (2), (3) and (5) of Proposition 3.13 definitely fail. For example, the support of a lamination λ could contain bands homeomorphic to $[0, 1] \times \mathbb{R}$, such that every $\{t\} \times \mathbb{R}$ maps onto a geodesic line of λ . Both ends of every such line converge to some \mathcal{E}_∞ . We can also construct transverse measures such that L_W is dense in such bands. This also shows that in general \mathcal{L}_S is strictly contained in \mathcal{L}_W .

In general the fibers of the marked measure spectrum map \mathbf{I} are, in any sense, infinite dimensional. For example we have:

Lemma 3.24. $\mathbf{I}^{-1}(0) \subset \mathcal{ML}(F)$ consists of laminations such that the support is entirely contained in the union of crowns.

On the other hand, the image of \mathbf{I} is tame, in fact:

Proposition 3.25. $\mathbf{I}(\mathcal{ML}(F)) = \mathbf{I}(\mathcal{ML}_c(\mathcal{K}(F)))$.

Earthquakes along finite laminations of $\mathcal{ML}_c(F)$. As finite laminations are dense, and arbitrary laminations $\lambda \in \mathcal{ML}_c(F)$ look like finite ones at cusps and boundary components of F^c , it is important (and easy) to understand earthquakes in the finite case.

Example 3.26. Let us consider again Examples 3.15. Let $F(l, t)$ be such that all twist parameters are strictly positive. Then, by definition, $(F(l, t), \lambda(t))$ is obtained from $(F(l, 0), \lambda(t))$ via a *left earthquake (along the measured geodesic lamination $\lambda(t)$ on $F(l, 0)$)*. $(F(l, -t), \lambda(t))$ is obtained from $(F(l, 0), \lambda(t))$ via a *right earthquake (along the measured geodesic lamination $\lambda(t)$ on $F(l, 0)$)*. In the reverse direction, $(F(l, 0), \lambda(t))$ is obtained from $(F(l, t), \lambda(t))$ via a *right earthquake*, and so on. This pattern of earthquakes does preserve the types.

Similarly, let $F(s)$ be such that all shear parameters are strictly positive. Then, by definition $(F(s), \lambda(s))$ is obtained from $(F(0), \lambda(s))$ via a *left earthquake (along the measured geodesic lamination $\lambda(s)$ on $F(0)$)*. $(F(-s), \lambda(s))$ is obtained from

$(F(0), \lambda(s))$ via a *right earthquake* (along the measured geodesic lamination $\lambda(s)$ on $F(0)$). In the reverse direction, $(F(0), \lambda(s))$ is obtained from $(F(s), \lambda(s))$ via a *right earthquake*, and so on. This pattern does not preserve the types, for $F(0) \in \mathcal{T}_{g,r}$, while $F(s)$ is without cusps. Moreover, $\lambda(s)$ has the following special property:

For every boundary component C_i of $F(s)^c$

$$l_{C_i}(F(s)) = I_{C_i}(\lambda(s)).$$

For every $(F, \lambda) \in \mathcal{ML}_c(S)$, λ being finite, the definition of (F', λ') obtained from (F, λ) via a *left (right) earthquake* extends *verbatim* the one of the above examples, so that $(F', \lambda') \in \mathcal{ML}_c(S)$, λ' is also a finite lamination, and (F, λ) is obtained from (F', λ') via the *inverse* right (left) earthquake.

Quake cocycles and general earthquakes. It is convenient to describe earthquakes by lifting everything to the universal coverings. Let it be as usual

$$\bar{\mathcal{K}}(\hat{F}) \subset F^c \subset \hat{F} = \mathbb{H}^2 / \Gamma.$$

Then F^c lifts to a Γ -invariant *straight convex set* H of \mathbb{H}^2 (i.e. H is the closed convex hull of an ideal subset of S_∞^1), and λ lifts to a Γ -invariant measured geodesic lamination on \hat{H} , that, for simplicity, we still denote by λ . If $F \in \mathcal{T}_c$, then $\bar{\mathcal{K}}(\hat{F}) = F^c$.

Lemma 3.27. *Let $(F, \lambda) \in \mathcal{ML}_c(S)$ such that λ is finite. Then there exists a left-quake cocycle*

$$B_\lambda^L: \hat{H} \times \hat{H} \rightarrow \text{PSL}(2, \mathbb{R})$$

such that

- (1) $B_\lambda^L(x, y) \circ B_\lambda^L(y, z) = B_\lambda^L(x, z)$ for every $x, y, z \in \hat{H}$;
- (2) $B_\lambda^L(x, x) = \text{Id}$ for every $x \in \hat{H}$;
- (3) B_λ^L is constant on the strata of the stratification of \hat{H} determined by λ ;
- (4) $B_\lambda^L(\gamma x, \gamma y) = \gamma B_\lambda^L(x, y) \gamma^{-1}$, for every $\gamma \in \Gamma$;
- (5) For every x_0 belonging to a 2-stratum of \hat{H} ,

$$\hat{H} \ni x \mapsto B_\lambda^L(x_0, x)x \in \mathbb{H}^2$$

lifts the left earthquake $\beta_\lambda^L(F)$ to \hat{H} . This cocycle is essentially unique. There exists a similar right-quake cocycle B_λ^R .

The proof is easy and the earthquake is equivalently encoded by its cocycle. For a general (F, λ) we look for (essentially unique) *quake-cocycles* that satisfy all the properties of the previous lemma, with the exception of the last one, and requiring furthermore that

(*) If $\lambda_n \rightarrow \lambda$ on an ε -neighbourhood of the segment $[x, y]$ and $x, y \notin L_W$, then $B_{\lambda_n}(x, y) \rightarrow B_\lambda(x, y)$.

Given such cocycles we can use the map of (5) in the previous lemma as the *general definition of earthquakes*.

For example, if $(F, \lambda) \in \mathcal{ML}_c(S)$ the cocycle can be derived by Lemma 3.27, the density of finite laminations and the fact that we require (*). If (F', λ') results from the left earthquake starting at (F, λ) , then it belongs to $\mathcal{ML}_c(S)$ and (F, λ) is obtained from it via the inverse right earthquake.

In fact, in [36] Epstein–Marden defined these quake-cocycles in general (extending the construction via finite approximations). Strictly speaking they consider only the case of (arbitrary) measured geodesic laminations on \mathbb{H}^2 , but the same arguments hold for laminations on arbitrary straight convex sets H - see also [15] for more details. Hence general left (right) earthquakes

$$(F', \lambda') = \beta^L(F, \lambda)$$

so that

$$(F, \lambda) = \beta^R(F', \lambda')$$

are eventually defined for arbitrary $(F, \lambda) \in \mathcal{ML}(S)$. We will also write $F' = \beta_\lambda^L(F)$, $\lambda' = \beta_\lambda^L(\lambda)$.

Earthquake flows on $\mathcal{ML}_c(S)$. Let $\lambda \in \mathcal{ML}_c(F)$ and consider the ray $(F, t\lambda)$, $t \in [0, +\infty)$. Then, for every $t > 0$, set

$$(F_t, \lambda_t) = (\beta_{t\lambda}^L(F), \frac{1}{t}\beta_{t\lambda}^L(t\lambda)), \quad t \geq 0.$$

This continuously extends at $t = 0$ by

$$(F_0, \lambda_0) = (F, \lambda).$$

We have

$$((F_t)_s, (\lambda_t)_s) = (F_{t+s}, \lambda_{t+s})$$

hence this defines the so called *left-quake flow* on $\mathcal{ML}_c(S)$. In particular this allows to define a sort of “exponential” map

$$\psi^L : \mathcal{ML}_c(F) \rightarrow \mathcal{ML}_c(S)$$

by evaluating the flow at $t = 1$. We do similarly for the *right-quake flow*.

For every t , we denote by $l(t)$ the marked length spectrum of F_t , by $\theta(t)$ its type, by $I(t)$ the marked measure spectrum of λ_t , by $\sigma_t : V \rightarrow \{\pm 1\}$ its signature, and so on. The following lemma describes the behaviour of these objects along the flow.

Lemma 3.28. *The marked measure spectrum is constant for every t , that is,*

$$I_\gamma(t) = I_\gamma(0) \quad \text{for every } \gamma \in \mathfrak{S}.$$

Let $p_i \in V$ and C_i be the curve surrounding it.

If $p_i \in V_{\mathcal{H}}(0)$, then

$$l_{C_i}(t) = |l_{C_i}(0) - t\sigma_i(\lambda)I_{C_i}(0)|$$

and

$$\sigma_i(t) = \text{sign}[l_{C_i} - t\sigma_i(\lambda)I_{C_i}(0)]\sigma_i(0).$$

If $p_i \in V_{\mathcal{P}}(0)$, then

$$l_{C_i}(t) = tI_{C_i}(0)$$

and

$$\sigma_i(t) = -1.$$

As every $\lambda \in \mathcal{ML}_c(F)$ looks finite at cusps and boundary components of F^c , it is enough (and fairly easy) to check the lemma in the finite case, by using also Examples 3.26.

Remark 3.29. If $p_i \in V_{\mathcal{P}}$ and the lamination enters the corresponding cusp, then for $t > 0$ the cusp opens on a geodesic boundary component whose length linearly depends on t with slope equal to $I_{C_i}(0)$. The way of spiraling of λ_t around p_i is always negative (positive for right earthquakes).

Let us consider more carefully the case $p_i \in V_{\mathcal{H}}$. Notice that if λ does not spiral around C_i then the length of C_i is constant. In the other cases let us distinguish two possibilities according to the sense of spiraling of λ .

(1) Case $\sigma_i(0) = -1$. Then for every $t > 0$,

$$\sigma_i(t) = -1, \quad l_{C_i} = l_{C_i}(0) + tI_{C_i}(0).$$

Thus the length of C_i increases affinely with slope $I_{C_i}(0)$ and the lamination continues to spiral in the negative direction.

(2) Case $\sigma_i(0) = 1$. There is a critical time $t_i = l_{C_i}(0)/I_{C_i}(0)$. Before t_i the length of C_i decreases affinely and the lamination spirals in the positive direction. At t_i , C_i becomes a cusp. After t_i , C_i is again a boundary component but the way of spiraling is now negative.

Remark 3.30. From the above lemma a remarkable subset of $\mathcal{ML}_c(F)$ is pointed out, namely:

$$\mathcal{V}_c(F) = \{\lambda \mid I_{C_i}(\lambda) < l_{C_i}(F) \text{ for } i \in V_{\mathcal{H}}\}.$$

Note that this set is *not* preserved by the canonical bijections stated in Proposition 3.22 (iv).

Corollary 3.31. *The restriction of the exponential-like map ψ^L to $\mathcal{V}_c(F) \cap \mathcal{ML}_c(F)^0$ preserves the type and the signatures. The restriction of this map to the whole of $\mathcal{V}_c(F)$ has generic image over the top-dimensional cell of $\mathcal{T}_c(S)$.*

The quake-flow on $\mathcal{ML}_c(S)^\#$. We will define an earthquake flow on $\mathcal{ML}_c(S)^\#$ that satisfies the following properties:

- (1) $\beta_t^\# \circ \beta_s^\# = \beta_{t+s}^\#$.
- (2) Every flow line $\{\beta_t^\#(F, \epsilon, \lambda, \eta) \mid t > 0\}$ is *horizontal* with respect to the trivialization of $\mathcal{ML}_c^\#(S)$. This means that the enhanced lamination is constant along the flow.
- (3) If we include $\mathcal{ML}_c(S)$ into $\mathcal{ML}_c(S)^\#$ by sending (F, λ) to $(F, \epsilon, \lambda, \eta)$ with $\epsilon_i = 1$ for every i and $\eta_i = 1$ for every $i \in V_{\mathcal{P}}$ then $\beta = \phi_{\mathcal{ML}}^\# \circ \beta^\#$ (where $\phi_{\mathcal{ML}}^\#$ is the usual forgetting map).

Remark 3.32. Before giving the actual definition, we describe the qualitative idea. Earthquake paths in $\mathcal{T}_c(S)$ rebound when they reach a cusp. Since $\mathcal{T}_c(S)^\#$ is obtained by reflecting $\mathcal{T}_c(S)$ along its faces, it is natural to lift such paths to horizontal paths on $\mathcal{T}_c(S)^\#$. Instead of rebounding the enhanced lamination after a cusp is obtained by a reflection along a boundary component of the initial lamination. This liftings are unique (up to the choice of an initial signature ϵ) when F does not contain any cusp. When F contains a cusp then there are many possible liftings due to the possible choices of the signature of the cusp after the earthquake. Thus the data (F, ϵ, λ) are not sufficient to determines the lifting. On the other hand the information of a signature of λ around the cusp solves this ambiguity.

Let us come to the actual definition:

$$\beta_t^\#(F, \epsilon, \lambda, \eta) = (\bar{F}, \bar{\eta}, \bar{\lambda}, \bar{\epsilon})$$

where

- (a) similarly to the definition of the map $t, (\bar{F}, \bar{\lambda}) = \beta(F, \rho_\epsilon(\lambda))$;
- (b) $\bar{\eta}_i = \epsilon_i \text{sign}(l_{C_i}(F) + t\eta_i I_{C_i}(\lambda))$;
- (c) $\bar{\epsilon}_i = \eta_i \text{sign}(l_{C_i}(F) + t\eta_i I_{C_i}(\lambda))$.

Property (1) follows from the fact that β is a flow. Point (2) depends on the fact that the spectrum of λ_t is constant and the products $\epsilon_i(t)\eta_i(t)$ are constant. Point (3) is straightforward. The only point to check is that $\beta^\#$ is continuous, as a map $\mathbb{R}_{\geq 0} \times \mathcal{ML}_c^\#(S) \rightarrow \mathcal{ML}_c^\#(S)$. By definition of the topology of $\mathcal{ML}_c^\#(S)$ it is enough to show that for every essential closed curve γ the functions

$$(t, (F, \epsilon, \lambda, \eta)) \mapsto I_\gamma^\#(\beta_t^\#(F, \epsilon, \lambda, \eta)), \quad (t, (F, \epsilon, \lambda, \eta)) \mapsto \mathbb{I}_\gamma^\#(\beta_t^\#(F, \epsilon, \lambda, \eta))$$

are continuous. If γ is not peripheral, then $I_\gamma^\#(\beta_t^\#(F, \epsilon, \lambda, \eta))$ and $\mathbb{I}_\gamma^\#(\beta_t^\#(F, \epsilon, \lambda, \eta))$ depend only on \bar{F} and $\bar{\lambda}$ so the continuity is a consequence of the continuity of β .

If γ is peripheral, then by Lemma 3.28 we have

$$\begin{aligned} I_\gamma^\#(\beta_t^\#(F, \epsilon, \lambda, \eta)) &= I_\gamma^\#(F, \epsilon) - tI_\gamma^\#(F, \epsilon, \lambda, \eta), \\ \mathbb{I}_\gamma^\#(\beta_t^\#(F, \epsilon, \lambda, \eta)) &= \mathbb{I}_\gamma^\#(F, \epsilon). \end{aligned}$$

For every $\xi \in I^\#(S)$ let us consider the map $\mathbb{R}_{\geq 0} \times \mathcal{T}_c^\#(S) \rightarrow \mathcal{T}_c^\#(S)$ that sends (t, F, ϵ) to the projection on $\mathcal{T}_c^\#(S)$ of $\beta_t(F, \epsilon, \xi(F))$ (where $\xi(F)$ is the realization of ξ with respect to the structure given by F). By (2) it is a flow on $\mathcal{T}_c(S)^\#$. We will denote by $\mathcal{E}_\xi^\#$ the homeomorphism of $\mathcal{T}_c(S)^\#$ corresponding to such a flow at time 1 (notice that $\mathcal{E}_\xi \circ \mathcal{E}_\xi = \mathcal{E}_{2\xi}$), it will be called the *enhanced earthquake along ξ* .

Earthquake Theorems

Theorem 3.33 (Earthquake Theorem on $\mathcal{T}_c(S)$). *For every $F_0, F_1 \in \mathcal{T}_c(S)$, let us denote by m the number of points in V that do not correspond to any cusp of F_1 nor of F_2 . Then there exist exactly 2^m left earthquakes such that $F_1 = \beta_\lambda^L(F_0)$. A similar statement holds with respect to right-quakes.*

This is a consequence of the somewhat more precise statement.

Theorem 3.34 (Earthquake Theorem on $\mathcal{T}_c(S)^\#$). *For every $(F_0, \epsilon_0), (F_1, \epsilon_1) \in \mathcal{T}_c(S)^\#$, there is a unique $\xi \in I^\#(S)$ such that $\mathcal{E}_\xi^\#(F_0, \epsilon_0) = (F_1, \epsilon_1)$. Similarly for the right-quakes.*

Given two “signed” surfaces (F_0, σ_0) and (F_1, σ_1) in $\mathcal{T}_c(S)$, where the respective signatures are arbitrary maps $\sigma_j : V \rightarrow \{\pm 1\}$, we say that they are *left-quake compatible* if there exists a left earthquake $(F_1, \lambda_1) = \beta^L(F_0, \lambda_0)$ such that $\sigma_j = \sigma_{\lambda_j}$. The following is an easy corollary of Lemma 3.28 and of Theorem 3.33.

Corollary 3.35. *The signed surfaces (F_0, σ_0) and (F_1, σ_1) are left-quake compatible if and only if for every $i = 1, \dots, r$ the following conditions are satisfied:*

If $l_{C_i}(F_1) < l_{C_i}(F_0)$, then $\sigma_0(i) = 1$.

If $l_{C_i}(F_1) > l_{C_i}(F_0)$, then $\sigma_1(i) = 1$.

Symmetric statements hold with respect to the right-quake compatibility.

In Section 5 we will outline an AdS proof of the Earthquake Theorem (that has been proved in [29]) that generalizes Mess’ proof in the special case of compact S .

$\mathcal{ML}_c(S)$ as tangent bundle of $\mathcal{T}_c(S)$. We have seen above that the bundle

$$p^\# : \mathcal{ML}_c(S)^\# \rightarrow \mathcal{T}_c(S)^\#$$

shares some properties with the *tangent bundle* $T\mathcal{T}_c^\#$ of its base space. We are going to substantiate this fact by means of quake-flows. In fact we have associated to every $\xi \in I^\#(S)$ a flow of $\mathcal{T}_c(S)^\#$, so we can consider the infinitesimal generator of such a flow, that is a vector field on $\mathcal{T}_c(S)^\#$, say X_ξ .

Proposition 3.36. *The map*

$$\Pi : \mathcal{T}_c(S)^\# \times I^\#(S) \rightarrow T\mathcal{T}_c(S)^\#$$

defined by $\Pi(\xi, F) = X_\xi(F)$ is a trivialization of $T\mathcal{T}_c(S)$.

Remark 3.37. The map Π is only a *topological trivialization*. This means that the identifications between tangent spaces arising from Π are not linear.

For a fixed type θ , denote by $I^\#(S)^\theta$ the points corresponding to laminations that do not enter any cusp. It is clear that for a point $F \in \mathcal{T}_c^\theta(S)^\#$ we have that $X_\xi(F)$ is tangent to $T_c^\theta(S)^\#$. So we get that the restriction of Π to $\mathcal{T}_c^\theta(S)^\# \times I^\#(S)^\theta$ is a trivialization of $T\mathcal{T}_c^\theta(S)^\#$.

4 Wick rotation-rescaling theory

We refer to [15]. Let S be a base surface of finite type. Recall from the Introduction and Section 2, that $\mathcal{M}\mathcal{G}\mathcal{H}_\kappa(S)$ denotes the Teichmüller-like space of Einstein maximal globally hyperbolic spacetimes of constant curvature $\kappa = 0, \pm 1$, that contain a *complete* Cauchy surface homeomorphic to S .

Denote by $\mathcal{P}(S)$ the Teichmüller-like space of *complex projective* structures (that is, $(S^2, \text{PSL}(2, \mathbb{C}))$ -structures) on S . Here S^2 is the Riemann sphere, identified with $S_\infty^2 = \partial\mathbb{H}^3$, and $\text{PSL}(2, \mathbb{C}) \cong \text{Isom}^+(\mathbb{H}^3)$ (see also [34]).

The aim of this section is to illustrate the following pattern of statements (given here in a somewhat informal way):

Classifications. For every surface S of finite type, and every $\kappa = 0, \pm 1$, there are geometrically defined “*materialization*” maps

$$m_{\mathcal{P}} : \mathcal{M}\mathcal{L}(S) \rightarrow \mathcal{P}(S),$$

$$m_\kappa : \mathcal{M}\mathcal{L}(S) \rightarrow \mathcal{M}\mathcal{G}\mathcal{H}_\kappa(S)$$

that actually make $\mathcal{M}\mathcal{L}(S)$ a *universal parameter space*.

Canonical correlations. For every $(F, \lambda) \in \mathcal{M}\mathcal{L}(S)$, there are geometrical correlations between the spacetimes $m_\kappa(F, \lambda)$ or between them and the projective surface $m_{\mathcal{P}}(F, \lambda)$. Such correlations are either realized by means of canonical *rescalings* or *Wick rotations* directed by the respective *cosmological times*, with *universal rescaling functions*.

Let us explain first some terms entering the last statement.

Definition 4.1. Let (M, h) be any spacetime and X be a nowhere vanishing h -timelike and future directed vector field on M . Let $\alpha, \beta : M \rightarrow \mathbb{R}_{>0}$ be positive functions.

We say that the Riemannian manifold (M, g) is obtained from (M, h) via the *Wick rotation* directed by X , with *vertical* (resp. *horizontal*) rescaling function β (resp. α), if for every $y \in M$, the g - and h -orthogonal spaces to $X(y)$ coincide (denoted by

$\langle X(y) \rangle^\perp$), and

$$g(X(y), X(y)) = -\beta(y)h(X(y), X(y)) \quad \text{and} \quad g|_{\langle X(y) \rangle^\perp} = \alpha(y)h|_{\langle X(y) \rangle^\perp}.$$

Similarly, the spacetime (M, h') is obtained from (M, h) via the *rescaling* directed by X , with *vertical* (resp. *horizontal*) rescaling function β (resp. α), if for every $y \in M$, the h' - and h -orthogonal spaces to $X(y)$ coincide, and

$$h'(X(y), X(y)) = \beta(y)h(X(y), X(y)) \quad \text{and} \quad h'|_{\langle X(y) \rangle^\perp} = \alpha(y)h|_{\langle X(y) \rangle^\perp}.$$

4.1 Cosmological time

We refer to [3] for a general treatment of this matter. Here we limit ourselves to recalling the main features of this notion. Let (M, h) be any spacetime. The *cosmological function*

$$\tau: M \rightarrow (0, +\infty]$$

is defined as follows. Let $C^-(q)$ be the set of past-directed causal curves in M that start at $q \in M$. Then

$$\tau(q) = \sup\{L(c) \mid c \in C^-(q)\},$$

where $L(c)$ denotes the Lorentzian length of c . Roughly speaking, this gives the (possibly infinite) *proper time that every event $q \in M$ has been in existence in M* . The function τ is said *regular* if it is finite valued for every $q \in M$, and $\tau \rightarrow 0$ along every past-directed inextendible causal curve. In such a case it turns out that τ is a continuous *global time* on M , called its *cosmological time*. This cosmological time (if it exists) represents an intrinsic feature of the spacetime. Having cosmological time has strong consequences for the structure of M , and τ itself has stronger properties (it is locally Lipschitz and twice differentiable almost everywhere). In particular: M is globally hyperbolic; for every $q \in M$, there exists a future-directed timelike unit speed geodesic ray ending at q , whose length equals $\tau(q)$. Up to a suitable past-asymptotic equivalence, these rays form the *initial singularity* of M . In a sense τ gives the Lorentzian distance of every event from the initial singularity.

4.2 Grafting and Lorentzian grafting

Before describing in some formal way how to get parameterizations of $\mathcal{MG}_\kappa(S)$ and $\mathcal{P}(S)$ in terms of $\mathcal{ML}(S)$, we will explain how to associate to a pair $(F, \lambda) \in \mathcal{ML}(S)$ a projective structure on S and a spacetime of constant curvature κ , in some simple cases (see also [34]).

First consider the case where S is closed and λ is empty. Given a hyperbolic structure $F = (S, h)$ on S , the projective structure associated to it, that, with a little abuse, we will denote simply by F , is the structure whose developing map coincides (up to post-composition with $g \in \text{PSL}(2, \mathbb{C})$) with the isometric developing map

of F . Structures obtained in this way are called Fuchsian and are characterized by the following requirements:

- (i) the developing map is injective;
- (ii) the holonomy is conjugate to a representation into $\text{PSL}(2, \mathbb{R})$.

For the Lorentzian side, define $m_\kappa(F)$ to be the spacetime $(S \times I, g_\kappa)$ where I is the interval $(0, +\infty)$ for $\kappa \geq 0$ and $I = (0, \pi/2)$ for $\kappa = -1$ and $g_\kappa = g_\kappa(F)$ is so defined

$$g_\kappa = \begin{cases} -dt^2 + t^2h & \text{if } \kappa = 0, \\ -dt^2 + \text{sh}^2(t)h & \text{if } \kappa = 1, \\ -dt^2 + \sin^2(t)h & \text{if } \kappa = -1. \end{cases} \tag{4.1}$$

The fact that g_κ has constant curvature κ is just a local computation independent of the compactness of F . Thus one does the computation assuming $F = \mathbb{H}^2$. For instance, for $\kappa = 0$, one embeds \mathbb{H}^2 in the Minkowski space \mathbb{X}_0 and takes the normal evolution of \mathbb{H}^2 (that is a map $\mathbb{H}^2 \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{X}_0$ sending (x, t) to tx). The pull-back of the Minkowski metric takes the form (4.1).

Remark 4.2. Strictly speaking $m_{-1}(F)$ is not maximal. In fact the metric g_κ can be defined as well on the interval $(0, \pi)$. On the other hand, for some reason that will appear clear it is better to define $m_{-1}(F)$ in this way and then to take its maximal extension.

Now suppose S is closed and λ is a weighted curve (c, a) . The projective surface $m_{\mathcal{P}}(F, \lambda)$ is the grafting of F along λ , that we sometimes denote by $\text{Gr}_\lambda(F)$. We cut F along c and graft a projective annulus $A = c \times [0, a]$ whose developing map can be explicitly described in the following way. We can choose a developing map $\text{dev}: \tilde{F} \rightarrow \mathbb{H}^2 = \{z \in \mathbb{C} \mid \text{Im}z > 0\}$ such that c lifts to a geodesic \tilde{c} with endpoints at 0 and ∞ . The developing map of A is given by

$$\tilde{c} \times [0, a] \ni (x, t) \rightarrow \text{dev}(x)e^{it} \in \mathbb{C}P^1 = S^2.$$

The fact that A can be grafted on F is based on the fact that the developing map of each component of ∂A is conjugate by some element of $\text{PSL}(2, \mathbb{C})$ to the developing map of c . Notice that A carries a natural Euclidean metric. The length of each boundary component of A is equal to the length of c whereas the width of A is equal to a . Thus we can consider on $\text{Gr}_\lambda(F)$ the metric that is hyperbolic on $F \setminus c$ and Euclidean on A . Such a metric is C^1 and compatible with the conformal structure underlying the projective structure of $\text{Gr}_\lambda(F)$. We call it the Thurston metric of $\text{Gr}_\lambda(F)$. In what follows we often indicate by $\text{Gr}_\lambda(F)$ both the projective structure and the metric structure on S .

Remark 4.3. Thurston’s distance is defined on every projective structure on S and is a metric compatible with the conformal class of the projective surface. The interesting point shown by Thurston is that Thurston’s metric determines the projective structure.

This means that a map between projective surfaces is a projective equivalence if and only if it is an isometry with respect to the corresponding Thurston's distances.

Remark 4.4. If a is *small*, then the holonomy group of $\text{Gr}_\lambda(F)$, say Γ , is quasi-Fuchsian and the developing map is injective with image a component of the discontinuity domain. Thus, $\text{Gr}_\lambda(F)$ can be regarded as an asymptotic end of the quasi-Fuchsian manifold \mathbb{H}^3/Γ . In fact, the boundary component of the convex core facing $\text{Gr}_\lambda(F)$ is isometric to F bent along c with bending angle a , and the annulus A coincides with the set of points in $\text{Gr}_\lambda(F)$ that are sent by the retraction on the convex core to the bending line.

Moreover, let us consider the connected component of the complement of the convex core in \mathbb{H}^3/Γ close to F . Then the distance d from the convex core is a C^1 -function on it whose level surfaces are isometric to $\text{ch } d \cdot \text{Gr}_{(\text{tgh } d)\lambda}(F)$ (if X is a metric space, $\lambda \cdot X$ denotes the metric space obtained by multiplying the distance function by λ).

Thurston generalized this idea and showed how to associate to each projective structure on S a non-complete hyperbolic structure on $S \times (0, 1)$, called the H -hull such that

- (i) its completion is $S \times [0, 1)$ and $S \times \{0\}$ is a locally convex bent surface F along a lamination λ ;
- (ii) the asymptotic end $S \times \{1\}$ carries the original projective surface that in turns coincides with $\text{Gr}_\lambda(F)$.

Moreover the distance d from $S \times \{0\}$ is a C^1 -function and level surfaces are isometric to

$$\text{ch } d \cdot \text{Gr}_{\text{tgh } d\lambda}(F). \quad (4.2)$$

Clearly in the quasi-Fuchsian case the H -hull is simply the end of the corresponding quasi-Fuchsian manifold facing the projective surface.

Consider now the Lorentzian case.

To construct $\mathfrak{m}_\kappa(F, \lambda)$ we will deform the structure on $\mathfrak{m}_\kappa(F)$ by means of a construction that is reminiscent of the grafting procedure, so we call it the *Lorentzian grafting*.

With a little abuse let us denote by c the geodesic representative of c with respect to the hyperbolic structure F . Then one shows that the timelike surface $c \times I$ is totally geodesic in $\mathfrak{m}_\kappa(F)$ (it is still a local computation – for instance, in the flat case it is a direct consequence of the fact that geodesics of \mathbb{H}^2 are intersections of \mathbb{H}^2 with linear timelike planes of Minkowski space). Then one cuts $\mathfrak{m}_\kappa(F)$ along $c \times I$ and *grafts* a piece, say $\mathfrak{m}_\kappa(A)$ such that

- (i) topologically $\mathfrak{m}_\kappa(A) = (c \times [0, a]) \times I$, that is, the product of the annulus $A = c \times [0, a]$ by the time interval I ;
- (ii) $\mathfrak{m}_\kappa(A)$ is equipped with a Lorentzian metric of constant curvature κ ;

- (iii) the restriction of the metric on each slice $A \times \{t\}$ is a Euclidean annulus, whose width depends only on a and on t and whose boundary length is equal to the length of $c \times \{t\} \subset m_\kappa(F)$;
- (iv) the boundary of $m_\kappa(A)$ (that is $\partial A \times I = c \times I \times \{0, a\}$) is totally geodesic and each component is isometric to $c \times I$.

For instance in the flat case $m_0(A)$ is just $(c \times I) \times [0, a]$ with the product metric (which is flat since it is the product of two flat metrics). For the other curvatures, the expression of the metric on $m_0(A)$ takes the more complicated form given by

$$\begin{cases} -dt^2 + (\operatorname{ch}^2(t)dr^2 + \operatorname{sh}^2(t)d\theta^2) & \text{for } \kappa = 1, \\ -dt^2 + (\cos^2(t)dr^2 + \sin^2(t)d\theta^2) & \text{for } \kappa = -1 \end{cases} \quad (4.3)$$

where θ is an arc parameter on c and r is the variable on $[0, a]$. Notice that the width of $A \times \{t\}$ is independent of t only in the flat case.

Remark 4.5. The piece $m_{-1}(A)$ is well defined only for $t \in (0, \pi/2)$ and this explains the definition of $m_{-1}(F)$. In general the spacetime obtained for $\kappa = -1$ is never maximal, so more correctly $m_{-1}(F)$ will denote the maximal extension of the spacetime we have defined. In the next sections we will explain the reason of this asymmetry and also how the spacetime we have defined is uniquely determined by its maximal extension.

Remark 4.6. A way to define $m_\kappa(F, \lambda)$ for a generic λ is by means of an approximation argument. We take a sequence of simple weighted curves $\lambda_n = (c_n, a_n)$ converging to λ and define $m_\kappa(F, \lambda) = \lim m_\kappa(F, \lambda_n)$. Clearly the existence of this limit has to be checked. To this aim it is better to work in the framework of (G, X) -structures and study the behaviour of the developing maps of $m_\kappa(F, \lambda)$. This will be the theme of the next sections.

Notice that the construction of $m_\kappa(F, \lambda)$ gives, as a by-product, a natural foliation of spacetime by spacelike surfaces homeomorphic to S . In fact, in both $m_\kappa(F)$ and $m_\kappa(A)$ we have pointed out a time-function t to express the metric in some explicit way. These functions glue to a time-function on $m_\kappa(F, \lambda)$. Notice however that the function we get in $m_\kappa(F, \lambda)$ is not smooth: its level surfaces are made out of hyperbolic pieces and Euclidean annuli. In fact they are reminiscent of the usual grafted surfaces.

Let us consider the flat case. In such a case the t level surface corresponding to some value t_0 is obtained by multiplying the hyperbolic metric on F by the factor t_0^2 (that is, by multiplying the hyperbolic distance by the factor t_0), cutting along c and gluing a Euclidean annulus of width a . This is the same as grafting an annulus of width a/t_0 on F and then multiplying the grafted distance by the factor t_0 .

More generally one can check explicitly that for a weighted multi-curve $\lambda = (c, a)$ the surface $t^{-1}(t_0) \subset \mathfrak{m}_\kappa(F, \lambda)$ is metrically equal to

$$\begin{aligned} t_0 \cdot \text{Gr}_{\lambda/t_0}(F) & \quad \text{if } \kappa = 0, \\ \text{sh } t_0 \cdot \text{Gr}_{\lambda/\text{tgh } t_0}(F) & \quad \text{if } \kappa = 1, \\ \sin t_0 \cdot \text{Gr}_{\lambda/\tan t_0}(F) & \quad \text{if } \kappa = -1. \end{aligned} \tag{4.4}$$

The point that makes this remark interesting is that the function t is the cosmological time of $\mathfrak{m}_\kappa(F, \lambda)$, so it is somehow independent of the parameterization and the same formulae to express the level surface work for every (F, λ) . This remark motivates the idea of finding a canonical rescaling directed by the gradient of the cosmological time transforming $\mathfrak{m}_0(F, \lambda)$ into $\mathfrak{m}_{\pm 1}(F, \lambda)$ and a Wick rotation transforming $\mathfrak{m}_0(F, \lambda)$ into the H -hull of $\text{Gr}_\lambda(F)$.

Remark 4.7. Consider the case where S is a surface of finite type. For $F \in \mathcal{T}_c(S)$ we could try to define $\mathfrak{m}_\kappa(F)$ as in the closed case. Notice however that the slice $S \times \{t\}$ is in general not complete. In fact such a spacetime has a natural totally geodesic timelike boundary that is homeomorphic to $\partial F^c \times I$. A way to get a complete level surface is then for each boundary component c of F to glue a piece $\mathfrak{m}_\kappa(\Delta)$ where $\Delta = c \times [0, +\infty)$ is an annulus with infinite width and $\mathfrak{m}_\kappa(\Delta) = \Delta \times I$ with a metric given in (4.3). Notice that the definition of $\mathfrak{m}_\kappa(F)$ is then consistent with the previous case provided that we allow a boundary component of F to carry infinite weight.

In fact, one can show that to define $\mathfrak{m}_\kappa(F, \lambda)$ it is necessary to glue a cylindrical end for each boundary component of F that is not close to the lamination. On the other hand, if λ contains a leaf l spiraling around a boundary curve, it is clear that it is possible to define the analogue of $\mathfrak{m}_\kappa(A)$ for this leaf (that now is the product of an infinite band of width equal to the weight of l and the time-interval I) and apply the grafting procedure. Notice that if l spirals around a boundary component c , the corresponding end on the slice $S \times \{t\}$ in $\mathfrak{m}_\kappa(F, \lambda)$ appears complete (in fact a path entering the ends meets the band infinitely many times so its length cannot be bounded).

From this discussion it appears clearly that in this context it is more convenient to use the notion of geodesic lamination on a surface given in Remark 3.12. That is, we require that the boundary components of F are contained in the lamination and that paths arriving on the boundary have infinite total mass. In particular, for each boundary component either a leaf spirals around it or it carries an infinite weight. With this definition the 0 lamination on F is obtained by putting the weight $+\infty$ on each boundary component.

4.3 Wick rotation-rescaling set-up

Let us go back to the statement concerning the canonical correlations. We will see that every spacetime $\mathfrak{m}_\kappa(F, \lambda)$ has (rather tame) cosmological time, so that the geometry of the initial singularity will quite naturally arise. The above mentioned Wick rotations and rescalings (possibly only defined on suitable “slabs” of the spacetimes) will be

directed by the *gradient of the cosmological times*. The rescaling functions will be *universal* in the sense that their values only depend on the cosmological time values: for every y in the domain of definition, $\beta(y) = \beta(\tau(y))$, $\alpha(y) = \alpha(\tau(y))$. We stress that they do not depend on (F, λ) .

We are going to outline the *linked* arguments establishing both the constructions of the maps m_* , the geometric correlations and the fact that the materialization maps induce bijections.

Let $(F, \lambda) \in \mathcal{ML}(S)$. With the notation of Section 3, we have

$$F \subset F^c \subset \hat{F} = \mathbb{H}^2 / \Gamma.$$

(F^c, λ) lifts to a Γ -invariant couple $(H, \tilde{\lambda})$ where H is a straight convex set in \mathbb{H}^2 equipped with the measured geodesic lamination $\tilde{\lambda}$. The universal covering map

$$\mathbb{H}^2 \rightarrow \hat{F}$$

restricts to the universal covering maps $H \rightarrow F^c$, $\mathring{H} \rightarrow F$, where \mathring{H} is the interior of H . To simplify the notation we make the abuse of always writing λ instead of either $\tilde{\lambda}$, (F, λ) or $(\mathring{H}, \tilde{\lambda})$, that is, we will understand F or \mathring{H} .

The projective surface

$$S_{\mathcal{P}}^\lambda = m_{\mathcal{P}}(\lambda)$$

will be given in terms of a pair $(d_{\mathcal{P}}^\lambda, h_{\mathcal{P}}^\lambda)$ of compatible developing map

$$d_{\mathcal{P}}^\lambda : \tilde{S} \rightarrow S^2$$

and holonomy representation

$$h_{\mathcal{P}}^\lambda : \pi_1(S) \rightarrow \text{PSL}(2, \mathbb{C}).$$

We denote by

$$p_{\mathcal{P}}^\lambda : \tilde{S}_{\mathcal{P}}^\lambda \rightarrow S_{\mathcal{P}}^\lambda$$

the corresponding projective universal covering.

Similarly, every spacetime

$$Y_\kappa^\lambda = m_\kappa(\lambda)$$

will be specified by a compatible pair $(d_\kappa^\lambda, h_\kappa^\lambda)$,

$$d_\kappa^\lambda : \tilde{S} \times \mathbb{R} \rightarrow \mathbb{X}_\kappa,$$

$$h_\kappa^\lambda : \pi_1(S) \rightarrow \text{Isom}^+(\mathbb{X}_\kappa).$$

We denote

$$p_\kappa^\lambda : \mathcal{U}_\kappa^\lambda \rightarrow Y_\kappa^\lambda$$

the corresponding Lorentzian universal covering.

For simplicity, we will often identify S with F , \tilde{S} with \mathring{H} , $\pi_1(S)$ with Γ , and so on.

For every F as above, denote by λ_0 the measured geodesic lamination consisting of the $+\infty$ weighted boundary components of F^c . Recall that λ_0 is the initial endpoint of any ray in $\mathcal{ML}(F)$. We will describe explicitly the corresponding surface $S_{\mathcal{P}}^{0_\lambda}$ and

spacetimes $Y_\kappa^0, \mathcal{U}_\kappa^0$. Every $\lambda \in \mathcal{ML}(F)$ somehow encodes the instructions in order to *deform* (d_*^0, h_*^0) towards $(d_*^\lambda, h_*^\lambda)$ as it has been made explicit in the case of finite laminations.

4.4 Flat spacetimes classification

Take the hyperboloid model $\mathbb{H}^2 \subset \mathbb{X}_0$ of the hyperbolic plane. The *chronological future* of 0 in \mathbb{X}_0 is the cone $I^+(0) = \{-x_0^2 + x_1^2 + x_2^2 < 0 \mid x_2 > 0\}$ from 0 over \mathbb{H}^2 . $I^+(0)$ has cosmological time $\tau = (x_0^2 - (x_1^2 + x_2^2))^{1/2}$, so that $\mathbb{H}^2 = \{\tau = 1\}$ and 0 is the initial singularity. The future $I^+(r) = \{-x_0^2 - x_2^2 < 0 \mid x_2 > 0\}$ of the spacelike geodesic $r = \{x_0 = x_2 = 0\}$ has cosmological time $\tau = (x_0^2 - x_2^2)^{1/2}$; r is the initial singularity.

Construction of \mathcal{U}_0^0 . The cone C_0H from 0 over $H \subset \mathbb{H}^2$ is contained in $I^+(0)$. The boundary of C_0H is made by the cone over the boundary of H . Each component of ∂C_0H , corresponding to a geodesic line $\gamma \subset \partial H$, is the intersection with $I^+(0)$ of a hyperplane P_γ , orthogonal to a determined unitary spacelike vector v_γ , that points out of C_0H . The developing map d_0^0 is an *embedding* onto the convex domain \mathcal{U}_0^0 of \mathbb{X}_0 made by the union of C_0H with the future of all the rays of the form $\{tv_\gamma \mid t \geq 0\}$.

A convenient description of the domain \mathcal{U}_0^0 is as the intersection of half-spaces. In fact, for each $x \in H$, let x^\perp denote the spacelike plane orthogonal to x . Then $I^+(x^\perp)$ is a half-space and

$$\mathcal{U}_0^0 = \bigcap_{x \in H} I^+(x^\perp).$$

This shows that \mathcal{U}_0^0 is convex and future complete.

Up to isometry of \mathbb{X}_0 , the local model for \mathcal{U}_0^0 at each component of $\partial C_0(H)$, is the future $I^+(r_+)$ of the ray $\{x_1 \geq 0\} \subset r$, that is,

$$I^+(r_+) = (I^+(0) \cap \{x_1 \leq 0\}) \cup (I^+(r) \cap \{x_1 \geq 0\}).$$

The above cosmological times match at the intersection, producing the cosmological time of the union, that turns out to be a C^1 -function. The ray r_+ is the initial singularity. Then the cosmological time of \mathcal{U}_0^0 coincides with the one of $I^+(0)$ on C_0H . The initial singularity is the spacelike tree with one vertex at 0 and rays $tv_\gamma, t \geq 0$, emanating from the origin. The action of $\pi_1(S)$ on H naturally extends to the whole of \mathcal{U}_0^0 , giving the holonomy h_0^0 .

Construction of \mathcal{U}_0^λ . Let us consider now an arbitrary lamination $\lambda = (L, \mu) \in \mathcal{ML}(F)$. The developing map d_0^λ will always be an *embedding* onto a convex domain \mathcal{U}_0^λ in \mathbb{X}_0 , obtained as follows. Fix a basepoint $x_0 \in \mathring{H}$ not belonging to the weighted part L_W of λ . For every $x \in \mathring{H} \setminus L_W$ choose an arc c transverse to λ with endpoints x_0 and x . For $t \in c \cap L$, let $v(t) \in \mathbb{R}^3$ denote the unitary spacelike vector tangent to

\mathbb{H}^2 at t , orthogonal to the leaf through t and pointing towards x . For $t \in c \setminus L$, let us set $v(t) = 0$. In this way we define a function

$$v: c \rightarrow \mathbb{R}^3$$

whose restriction on the support of μ is continuous. We can define

$$s(x) = \int_c v(t) d\mu(t).$$

As the homological intersection of a closed transverse path with each leaf is 0, using that μ is invariant along \mathcal{L} preserving homotopies, it is not hard to see that s does not depend on the path c . Moreover, it is constant on every stratum of the stratification determined by λ , and it is a continuous function on $H \setminus L_W$.

The domain \mathcal{U}_0^λ can be defined in the following way:

$$\mathcal{U}_0^\lambda = \bigcap_{x \in H \setminus L_W} I^+(s(x) + x^\perp).$$

Note that this definition is compatible with the one already given for \mathcal{U}_0^0 .

The holonomy of Y_0^λ can be defined in this way:

$$h_0^\lambda(\gamma) = h_0^0(\gamma) + \tau(\gamma)$$

where $h_0^0: \pi_1(S) \rightarrow \text{SO}(2, 1)$ is the hyperbolic holonomy of F and $\tau(\gamma)$ is the translation by the vector $s(\gamma x_0)$. Since the lamination λ is h -invariant (being the pull-back of a lamination on F) the domain \mathcal{U}_0^λ turns out to be h_0^λ -invariant and Y_0^λ is the quotient of \mathcal{U}_0^λ by this action.

Let us summarize the main properties of these constructions (see [15], [7] for all details).

Theorem 4.8. (1) \mathcal{U}_0^λ coincides with the intersection of the future of its null support planes. In particular it is future complete.

(2) \mathcal{U}_0^λ has C^1 cosmological time T_0^λ with range $(0, +\infty)$. Every level surface $\mathcal{U}_0^\lambda(a) = (T_0^\lambda)^{-1}(a)$ is a complete Cauchy surface. For every $y \in \mathcal{U}_0^\lambda$, there is a unique past timelike geodesic segment emanating from y that realizes $T_0^\lambda(y)$. The union of the past endpoints of such segments makes the initial singularity Σ_0^λ . This is a spacelike \mathbb{R} -tree injectively immersed in \mathbb{X}_0 .

(3) The action of $\pi_1(S)$ on H induces a natural flat spacetime holonomy action on $\mathcal{U}_0^\lambda \cup \Sigma_0^\lambda$. The cosmological time descends to the quotient spacetime Y_0^λ .

It is convenient to give a general definition of a convex subset of \mathbb{X}_0 satisfying statement (1) in this theorem.

Definition 4.9. A regular domain is an open convex subset of \mathbb{X}_0 that coincides with the intersection of the future of its null support planes and admits at least two non-parallel null support planes.

Hence we have a well-defined map

$$m_0: \mathcal{ML}(S) \rightarrow \mathcal{MG}\mathcal{H}_0(S).$$

The spacetimes \mathcal{U}_0^λ (and \mathcal{Y}_0^λ) are particularly simple to figure out when λ is a *finite* lamination. In such a case, the local model consists of the future, say \mathcal{U}_0 , of a segment $I = [0, \alpha_0 v_0]$, where v_0 is a unitary spacelike vector and $0 < \alpha_0 < \pi$. Here a local model means that there is a neighbourhood of each point $p \in \mathcal{U}_0^\lambda$ that embeds in \mathcal{U}_0 via an isometry that preserves the cosmological time.

The cosmological time on \mathcal{U}_0 is realized by geodesics with starting point on $[0, \alpha_0 v]$, so there is a natural projection say $r: \mathcal{U}_0 \rightarrow [0, \alpha_0 v]$ sending p to the point on the segment that realizes the cosmological time.

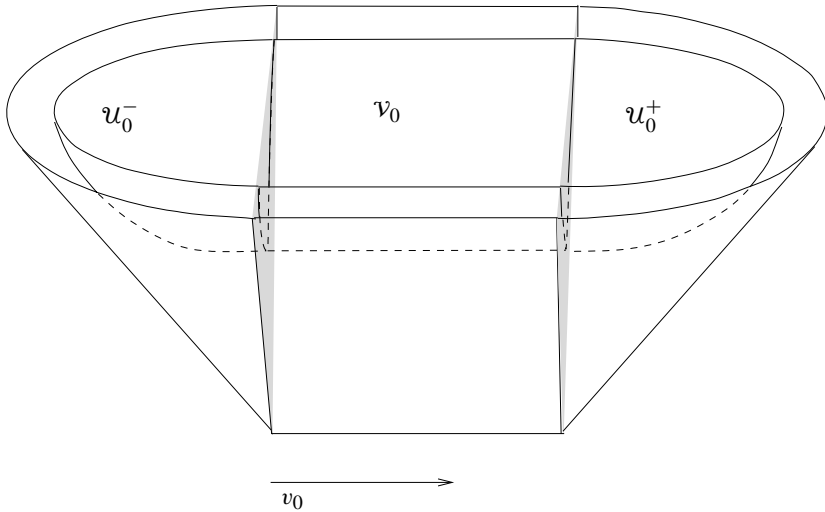


Figure 2. The domain \mathcal{U}_0 , its decomposition, and a level surface.

We have a decomposition of \mathcal{U}_0 in three pieces $\mathcal{U}_0^-, \mathcal{U}_0^+, \mathcal{V}$ defined in the following way:

$$\begin{aligned} \mathcal{U}_0^- &= r^{-1}(0); \\ \mathcal{V} &= r^{-1}(0, \alpha_0 v_0); \\ \mathcal{U}_0^+ &= r^{-1}(\alpha_0 v_0). \end{aligned}$$

We denote by $\mathcal{U}_0^+(a), \mathcal{U}_0^-(a), \mathcal{V}(a)$ the intersections of the corresponding domains with the surface $\mathcal{U}_0(a)$. The surfaces $\mathcal{U}_0^\pm(a)$ are hyperbolic of constant curvature $-1/a^2$. On the other hand, the parametrization of $\mathcal{V}(a)$ given by

$$(0, \alpha_0) \times I_0 \ni (t, y) \mapsto ay + tv_0 \in \mathcal{V}(a)$$

produces two orthogonal geodesic foliations on \mathcal{V} . The parametrization restricted to horizontal leaves is an isometry, whereas on the vertical leaves it acts as a rescaling of factor a . Thus $\mathcal{V}(a)$ is a Euclidean band of width α_0 . Note that by formally setting $\alpha_0 = +\infty$, and removing \mathcal{U}_0^+ , we recover the above local model for \mathcal{U}_0^0 at each component of $\partial C_0(H)$.

The initial singularity of a flat spacetime corresponding to a finite lamination is a *simplicial* metric tree. On the other hand, in [15] we prove also a suitable *continuous dependence* of \mathcal{U}_0^λ on λ . By using the density of finite laminations, this implies that spacetimes corresponding to finite laminations provide us with good approximations of arbitrary ones.

Asymptotic states. In general, the cosmological time level surface $\mathcal{U}_0^\lambda(1)$ ($Y_0^\lambda(1)$) is a C^1 spacelike surface; with the induced Riemannian metric it realizes the *grafting* of \mathring{H} (the hyperbolic surface F) at the measured geodesic lamination λ . By taking the rescaled level surface $(1/a)Y_0^\lambda(a)$, we get a 1-parameter family of grafting of F . More precisely, we get that $(1/a)Y_0^\lambda(a)$ is obtained by grafting F along λ/a .

When $a \rightarrow +\infty$ the geometries of $(1/a)Y_0^\lambda(a)$ converge to F . The geometry of the initial singularity Σ_0^λ of \mathcal{U}_0^λ , together with the isometric action of Γ on it, is “dual” to the geometry of the measured lamination λ , and can be recovered by means of the asymptotic behaviour of the level surfaces $\mathcal{U}_0^\lambda(a)$ (equipped with the respective isometric actions of Γ on them), when $a \rightarrow 0$.

The inverse map of \mathfrak{m}_0 . The image of \mathfrak{m}_0 consists of spacetimes whose universal covering is a regular domain that is future complete. On the other hand, general results due to Barbot [7] on flat spacetimes, applied in our finite type situation, imply that, *possibly reversing the time orientation*, every spacetime Y in $\mathcal{M}\mathcal{G}\mathcal{H}_0(S)$ is future complete, and its universal covering is a regular domain $\mathcal{U} \neq I^+(r)$. So it is natural to consider the quotient $\mathcal{M}\mathcal{G}\mathcal{H}_0(S)/\pm$, up to time orientation reversing. We are going to outline the steps leading to the inverse map of \mathfrak{m}_0 , defined on it. First one shows that every regular domain \mathcal{U} has cosmological time T that satisfies point (2) of Theorem 4.8. We consider the level surface $\mathcal{U}(1)$. We have a natural continuous *retraction*

$$r: \mathcal{U}(1) \rightarrow \Sigma_{\mathcal{U}}$$

onto the initial singularity. Moreover, the gradient of T is a unitary vector field, hence it induces the *Gauss map*

$$N: \mathcal{U}(1) \rightarrow \mathbb{H}^2.$$

The closure $H_{\mathcal{U}}$ of the image of N in \mathbb{H}^2 is a straight convex set. If $\mathcal{U} \rightarrow Y$ is a universal covering of $Y \in \mathcal{M}\mathcal{G}\mathcal{H}_0(S)$, the action of $\pi_1(S)$ extends to $H_{\mathcal{U}}$, and makes it a universal covering of $F_{\mathcal{U}}^c$, for some $F_{\mathcal{U}} \in \tilde{\mathcal{T}}(S)$. We take the partition of $\mathcal{U}(1)$ given by the closed sets $r^{-1}(y)$, $y \in \Sigma_{\mathcal{U}}$. Via the retraction, we can pullback to this partition the metric structure of $\Sigma_{\mathcal{U}}$, and (in a suitable sense) we can project everything onto $H_{\mathcal{U}}$, by means of the Gauss map. More precisely, if $r^{-1}(y)$ is 1-dimensional, then it is a geodesic line, so that the union of such lines makes a lamination in $\mathcal{U}(1)$. We can

define on it a transverse measure such that the mass of any transverse path is given by the integral of the Lorentzian norm of the derivative of r . A measured geodesic lamination $\lambda_{\mathcal{U}}$ on $H_{\mathcal{U}}$ is obtained via the push-forward by N of this lamination on $\mathcal{U}(1)$. This descends to a lamination $\lambda_{\mathcal{U}}$ on F . So we eventually get $m_0^{-1}(Y) = (F_{\mathcal{U}}, \lambda_{\mathcal{U}})$. This achieves our classification of *flat* MGH spacetimes of finite type.

4.5 Wick rotation: flat Lorentzian vs. hyperbolic geometry

Although we adopt a slightly different definition of the measured geodesic laminations involved, the bijective map

$$m_{\mathcal{P}}: \mathcal{ML}(S) \rightarrow \mathcal{P}(S)$$

is due to Kulkarni–Pinkall [50] and extends one due to Thurston for compact S . This is unfolded in terms of a 3-dimensional hyperbolic construction. We are going to describe it, by performing at the same time the canonical Wick rotation establishing the correlation between flat spacetimes $Y \in \mathcal{MGH}_0(S)$ and suitable hyperbolic 3-manifolds.

For every $Y \in \mathcal{MGH}_0(S)$ ($Y = Y_0^\lambda$), with universal covering $\mathcal{U} \rightarrow Y$ and cosmological time T , we construct a local C^1 -diffeomorphism

$$d_{\mathbb{H}}: \mathcal{U}(> 1) \rightarrow \mathbb{H}^3$$

and a compatible holonomy

$$h_{\mathbb{H}}: \pi_1(S) \rightarrow \mathrm{PSL}(2, \mathbb{C})$$

realizing a (non complete) hyperbolic structure $M = M_Y$ on $Y(> 1)$. This verifies the following properties:

(1) The hyperbolic metric of M is obtained by the Wick rotation of the flat Lorentzian metric on $Y(> 1)$, directed by the gradient of T , with universal rescaling functions

$$\alpha = \frac{1}{T^2 - 1} \quad \text{and} \quad \beta = \frac{1}{(T^2 - 1)^2}.$$

(2) Recall that the closure H of the Gauss map image is the straight convex set realizing the future asymptotic geometry of \mathcal{U} . Then the map $d_{\mathbb{H}}$ extends (in an $h_{\mathbb{H}}$ equivariant way) to

$$d_{\mathbb{H}}: \mathcal{U}(\geq 1) \cup H \rightarrow \mathbb{H}^3$$

such that the following is valid.

(a) The restriction of $d_{\mathbb{H}}$ to $\mathcal{U}(> 1) \cup H$ corresponds to the completion of the manifold M . The restriction to \mathring{H} is a locally isometric *pleated immersion* in \mathbb{H}^3 , having the measured geodesic lamination λ as *bending locus*. This gives the so-called *hyperbolic boundary* of M . The level surfaces $\mathcal{U}(a)$, $a > 1$, correspond via $d_{\mathbb{H}}$ to level surfaces of the distance function Δ on M from its hyperbolic boundary, so that the

inverse Wick rotation is directed by the gradient of Δ . More precisely, the following formula holds:

$$\Delta = \operatorname{arctgh}(1/T).$$

(b) The restriction $d_{\mathbb{H}}|_{\mathcal{U}(1)}$ actually coincides with the developing map of the complex projective structure $S_{\mathcal{P}}^\lambda$

$$d_{\mathcal{P}} : \mathcal{U}(1) \rightarrow S^2$$

so that

$$h_{\mathcal{P}} = h_{\mathbb{H}}.$$

The spacelike metric of $\mathcal{U}(1)$ ($Y(1)$) coincides with the *Thurston metric* (see Remarks 4.3, 4.4) of this projective surface, and its *canonical stratification* coincides with the stratification induced by the retraction r of $\mathcal{U}(1)$ onto the initial singularity. This gives the so called *asymptotic complex projective boundary* of M . In fact M turns to be the H -hull of $Y(1)$.

Remark 4.10. We give here a more precise description of the Thurston metric specifying what the canonical stratification is.

Let us take a complex projective structure on our surface S and consider a developing map

$$D : \tilde{S} \rightarrow S^2.$$

Pulling back the standard unit-sphere metric of S^2 on \tilde{S} is *not* a well-defined operation, as it depends on the choice of the developing map. Nevertheless, by the compactness of S^2 , the completion \bar{S} of \tilde{S} with respect to such a metric is well defined. It turns out that in our finite-type situation, $\bar{S} \setminus \tilde{S}$ contains at least 2 points (we say that it is of *hyperbolic* type). A *round disk* in \tilde{S} is a set Δ such that $D|_{\Delta}$ is injective and the image of Δ is a round disk in S^2 (this notion is well defined because $\operatorname{PSL}(2, \mathbb{C})$ sends round disks onto round disks). Given a *maximal* disk Δ (with respect to the inclusion), we can consider its closure $\bar{\Delta}$ in \bar{S} .

The closed disk $\bar{\Delta}$ is sent by D to the closed disk $\overline{D(\Delta)}$. In particular, if g_{Δ} denotes the pull-back on Δ of the standard *hyperbolic* metric on $D(\Delta)$, we can consider the boundary of Δ in \hat{S} as its ideal boundary. Since Δ is maximal, $\bar{\Delta}$ is not contained in \tilde{S} . So, if Λ_{Δ} denotes the set of points in $\bar{\Delta} \setminus \tilde{S}$, let $\hat{\Delta}$ be the convex hull in (Δ, g_{Δ}) of Λ_{Δ} (by maximality Λ_{Δ} contains at least two points). In [50] it is proved that for every point $p \in \tilde{S}$, there exists a unique maximal disk Δ containing p such that $p \in \hat{\Delta}$. So, $\{\hat{\Delta} \mid \Delta \text{ is a maximal disk}\}$ is a partition of \tilde{S} . We call it the *canonical stratification* of \tilde{S} . Clearly the stratification is invariant under the action of $\pi_1(S)$.

Let g be the Riemannian metric on \tilde{S} that coincides at p with the metric g_{Δ} , where Δ is the maximal disk such that $p \in \hat{\Delta}$. It is a conformal metric, in the sense that it makes D a conformal map. It is $C^{1,1}$ and is invariant under the action of $\pi_1(S)$. So, it induces a metric on \tilde{S} . We call it the *Thurston metric* on \tilde{S} .

Finally let us recall the construction of the H -hull of S . For $p \in \tilde{S}$, let $\Delta(p)$ be the maximal disk such that $p \in \hat{\Delta}$. The image of Δ via dev is a round disk in S^2 , so

its boundary is the trace of a hyperbolic plane $P(p)$ in \mathbb{H}^3 . Let c_p be the geodesic half-line with an endpoint at $\text{dev}(p)$ and an endpoint on $P(p)$ and orthogonal to $P(p)$. Then the developing map of the H -hull of S is the map

$$\tilde{S} \times (0, +\infty) \ni (p, t) \mapsto c_p(t) \in \mathbb{H}^3.$$

Notice that if S is quasi-Fuchsian, the H -hull is simply the end of the corresponding quasi-Fuchsian manifold facing S .

About the rescaling function. Before proving the theorem we want to give some heuristic motivation for the formulae of Wick rotation. The point is that we want to construct a Wick rotation transforming Y_0^λ (or some slab) into the H -hull, say M , of $\text{Gr}_\lambda(F)$, in such a way that the cosmological time level surfaces are sent to level surfaces of the distance function, Δ , from the hyperbolic boundary and rescaling functions are constant on level surfaces. Now suppose that such a Wick rotation exists. Let $\Delta(T)$ be in such a way that the Wick rotation transforms $Y(T)$ into $M(\Delta(T))$, and let $\alpha(T)$ and $\beta(T)$ be the horizontal and vertical rescaling functions.

By Formulae (4.2) and (4.4) we have

$$(\alpha(T))^{1/2} T \text{Gr}_{\lambda/T}(F) = \text{ch } \Delta(T) \text{Gr}_{\text{tgh } \Delta(T)\lambda}$$

Since $\text{Gr}_{t\lambda}(F)$ is conformally equivalent to $\text{Gr}_{s\lambda}(F)$ if and only if $s = t$, we deduce that

$$T = 1/\text{tgh } (\Delta(T)),$$

that is, $\Delta(T) = \text{arctgh } 1/T$. Moreover, we have

$$\alpha(T) = \text{ch}^2(\Delta(T))/T^2 = 1/(T^2 - 1).$$

Finally, let X denote the gradient of T with respect to the flat metric and Y denote the gradient of Δ with respect to the hyperbolic metric. We have $X = -\beta^{1/2}Y$. On the other hand, $\langle X, Y \rangle_{\text{Hyp}} = d\Delta(X) = \Delta'(T)dT(X) = -1/(T^2 - 1)$. Thus $\beta(T) = 1/(T^2 - 1)^2$.

Summing up if some Wick rotation exists satisfying the required properties, then necessarily $\alpha = 1/(T^2 - 1)$ and $\beta = 1/(T^2 - 1)^2$.

Bending cocycle. A key step in the construction is the *bending* of \mathring{H} in \mathbb{H}^3 along a measured geodesic lamination λ . We mostly refer to the Epstein–Marden paper [36] where this hyperbolic bending has been carefully studied (in the case of $\mathring{H} = \mathbb{H}^2$; however the constructions extend straightforwardly to the general case). In fact, in [36] one considers *quake-bend* maps, more generally associated to *complex valued* transverse measures on a lamination \mathcal{L} . Bending maps correspond to *imaginary valued* measures. So, given a measured geodesic lamination $\lambda = (\mathcal{L}, \mu) \in \mathcal{ML}(F)$, we take $i\mu$ in order to get the corresponding bending map.

The bending cocycles. We fix once and for all an embedding of \mathbb{H}^2 into \mathbb{H}^3 as a totally geodesic hyperbolic plane.

Given λ on H as usual, we first define the associated *bending cocycle* (recall a similar notion already introduced in Section 3 relatively to earthquakes). This is a map

$$B_\lambda: \mathring{H} \times \mathring{H} \rightarrow \text{PSL}(2, \mathbb{C})$$

which satisfies the following properties:

- (1) $B_\lambda(x, y) \circ B_\lambda(y, z) = B_\lambda(x, z)$ for every $x, y, z \in \mathring{H}$;
- (2) $B_\lambda(x, x) = \text{Id}$ for every $x \in \mathring{H}$;
- (3) B_λ is constant on the strata of the stratification of \mathring{H} determined by λ ;
- (4) If $\lambda_n \rightarrow \lambda$ on an ε -neighbourhood of the segment $[x, y]$ and $x, y \notin L_W$, then $B_{\lambda_n}(x, y) \rightarrow B_\lambda(x, y)$.

If λ is finite, then there is an easy description of B_λ . If l is an oriented geodesic of \mathbb{H}^3 , let $X_l \in \mathfrak{sl}(2, \mathbb{C})$ denote the infinitesimal generator of the positive rotation around l such that $\exp(2\pi X_l) = \text{Id}$ (since l is oriented the notion of *positive* rotation is well defined). Now take $x, y \in \mathring{H}$. If they lie in the same leaf of λ , then put $B_\lambda(x, y) = \text{Id}$. If both x and y do not lie in the support of λ , then let l_1, \dots, l_s be the geodesics of λ meeting the segment $[x, y]$ and a_1, \dots, a_s be the respective weights. Let us consider the orientation on l_i induced by the half plane bounded by l_i containing x and non-containing y . Then, put

$$B_\lambda(x, y) = \exp(a_1 X_{l_1}) \circ \exp(a_2 X_{l_2}) \circ \dots \circ \exp(a_s X_{l_s}),$$

where we have identified \mathbb{H}^2 with a geodesic plane of \mathbb{H}^3 . If x lies in l_1 , use the same construction, but replace a_1 by $a_1/2$; if y lies in l_s replace a_s by $a_s/2$.

The bending cocycle is not continuous on the whole definition set. However, there is a natural continuous “pull-back” of it to a cocycle defined on the flat spacetime $\mathcal{U} = \mathcal{U}_0^\lambda$

$$\hat{B}_\lambda: \mathcal{U}(1) \times \mathcal{U}(1) \rightarrow \text{PSL}(2, \mathbb{C})$$

such that

$$\hat{B}_\lambda(p, q) = B_\lambda(N(p), N(q))$$

for p, q such that $N(p)$ and $N(q)$ do not lie on L_W .

This map is locally Lipschitz (with respect to the Euclidean distance on \mathcal{U}). Moreover, for every compact set K of \mathcal{U} , the Lipschitz constant on $K \times K$ depends only on $N(K)$, on the diameter of the retraction of K on $\mathcal{U}(1)$ (via $r(1, \cdot)$) and on the maximum M and minimum m of T on K .

The bending map. Fix a basepoint x_0 of \mathring{H} (x_0 is supposed not to be in L_W). The *bending map* of \mathring{H} along λ is

$$F = F_\lambda: \mathring{H} \ni x \mapsto B(x_0, x) \cdot x \in \mathbb{H}^3.$$

The map F satisfies the following properties:

- (1) it does not depend on x_0 up to post-composition by elements of $\text{PSL}(2, \mathbb{C})$;

- (2) it is a 1-Lipschitz map;
- (3) if $\lambda_n \rightarrow \lambda$ then $F_{\lambda_n} \rightarrow F_\lambda$ with respect to the compact open topology.

The Wick rotation. We are ready to construct the local C^1 -diffeomorphism

$$d_{\mathbb{H}}: \mathcal{U}(> 1) \rightarrow \mathbb{H}^3$$

with the properties outlined at the beginning of this section.

Recall the continuous cocycle $\hat{B} = \hat{B}_\lambda$ defined above on the whole of $\mathcal{U} \times \mathcal{U}$. Since both \mathbb{H}^3 and $\mathbb{H}^2 \subset \mathbb{H}^3$ are oriented, the normal bundle is oriented too. Let v denote the normal vector field on \mathbb{H}^2 that is positively oriented with respect to the orientation of the normal bundle. Let us take $p_0 \in N^{-1}(x_0)$ and for $p \in \mathcal{U}(> 1)$ consider the geodesic ray c_p of \mathbb{H}^3 starting at $F(N(p))$ with speed vector equal to $w(p) = \hat{B}(p_0, p)_*(v(N(p)))$. Thus $d_{\mathbb{H}}$ is defined in the following way:

$$d_{\mathbb{H}}(p) = c_p(\operatorname{arctgh}(1/T(p))) = \exp_{F(N(p))} \left(\operatorname{arctgh} \left(\frac{1}{T(p)} \right) w(p) \right).$$

As usual, we make everything explicit on the local models of \mathcal{U}_0^0 and of flat spacetimes associated to finite laminations.

Local model of the Wick rotation for finite laminations. Consider as above the future \mathcal{U}_0 of a spacelike segment $[0, v_0]$ in \mathbb{X}_0 (adopting the same notation). We introduce suitable $C^{1,1}$ coordinates on \mathcal{U}_0 . For any $a > 0$, denote by l_a the boundary of $\mathcal{U}_0^-(a)$ and by d_a the intrinsic distance of $\mathcal{U}_0(a)$. Let l_0 be the geodesic in \mathbb{H}^2 where each l_a is sent by the Gauss map N . Fix a point z_0 on $l_0 \subset \mathbb{H}^2$ and denote by $\hat{z}_a \in l_a$ the point such that $N(\hat{z}_a) = z_0$.

For every $x \in \mathcal{U}_0(a)$, there is a unique point $\pi(x) \in l_a$ such that $d_a(x, l_a) = d_a(x, \pi(x))$. Then, we consider coordinates T, ζ, u , where T is again the cosmological time, and ζ, u are defined in the following way

$$\begin{aligned} \zeta(x) &= \epsilon(x) d_{T(x)}(x, l_{T(x)}) / T(x), \\ u(x) &= \epsilon'(x) d_{T(x)}(\pi(x), \hat{z}_{T(x)}) / T(x) \end{aligned}$$

where $\epsilon(x)$ (resp. $\epsilon'(x)$) is -1 if $x \in \mathcal{U}_0^-$ (resp. $\pi(x)$ is on the left of $\hat{z}_{T(x)}$) and is 1 otherwise.

Choose coordinates (y_0, y_1, y_2) of Minkowski space such that $v_0 = (0, 0, 1)$ and $z_0 = (1, 0, 0)$. Thus, the parametrization induced by T, ζ, u is

$$(T, u, \zeta) \mapsto \begin{cases} T \cdot (\operatorname{ch} uch \zeta, \operatorname{sh} uch \zeta, \operatorname{sh} \zeta) & \text{if } \zeta < 0, \\ T \cdot (\operatorname{ch} u, \operatorname{sh} u, \zeta) & \text{if } \zeta \in [0, \alpha_0/T], \\ T \cdot (\operatorname{ch} uch \zeta', \operatorname{sh} uch \zeta', \operatorname{sh} \zeta' + \alpha_0/T) & \text{otherwise} \end{cases}$$

where $\zeta' = \zeta - \alpha_0/T$.

With respect to these coordinates the metric takes the following form:

$$h_0(T, \zeta, u) = \begin{cases} -dT^2 + T^2(d\zeta^2 + \operatorname{ch}^2 \zeta du^2) & \text{if } \zeta < 0, \\ -dT^2 + T^2(d\zeta^2 + du^2) & \text{if } \zeta \in [0, \alpha_0/T], \\ -dT^2 + T^2(d\zeta^2 + \operatorname{ch}^2(\zeta') du^2) & \text{otherwise.} \end{cases}$$

Notice that the gradient of T is just the coordinate field $\frac{\partial}{\partial T}$.

The Gauss map takes the form

$$N(T, \zeta, u) = \begin{cases} (\operatorname{ch} u \operatorname{ch} \zeta, \operatorname{sh} u \operatorname{ch} \zeta, \operatorname{sh} \zeta) & \text{if } \zeta < 0, \\ (\operatorname{ch} u, \operatorname{sh} u, 0) & \text{if } \zeta \in [0, \alpha_0/T], \\ (\operatorname{ch} u \operatorname{ch} \zeta', \operatorname{sh} u \operatorname{ch} \zeta', \operatorname{sh} \zeta') & \text{otherwise,} \end{cases}$$

and the bending cocycle $\hat{B}_0(p_0, (T, \zeta, u))$ is the rotation around l_0 of angle equal to 0 if $\zeta < 0$, ζ if $\zeta \in [0, \alpha_0/T]$, α_0/T otherwise.

Let \mathbb{H}^3 be identified with the set of timelike unit vectors in the $(3 + 1)$ -Minkowski space \mathbb{M}^4 . We can choose affine coordinates on \mathbb{M}^4 in such a way that the inclusion $\mathbb{H}^3 \subset \mathbb{H}^4$ is induced by the inclusion $\mathbb{X}_0 \rightarrow \mathbb{M}^4$ given by $(x_0, x_1, x_2) \mapsto (x_0, x_1, x_2, 0)$. Thus the general rotation around l_0 of angle α is represented by the linear transformation T_α , such that

$$\begin{aligned} T_\alpha(e_0) &= e_0, & T_\alpha(e_1) &= e_1, \\ T(e_2) &= \cos \alpha e_2 + \sin \alpha e_3, & T_\alpha(e_3) &= -\sin \alpha e_2 + \cos \alpha e_3 \end{aligned}$$

where (e_0, e_1, e_2, e_3) is the canonical basis of \mathbb{R}^4 . Thus, we can write in local coordinates $d_{\mathbb{H}} = D_0$

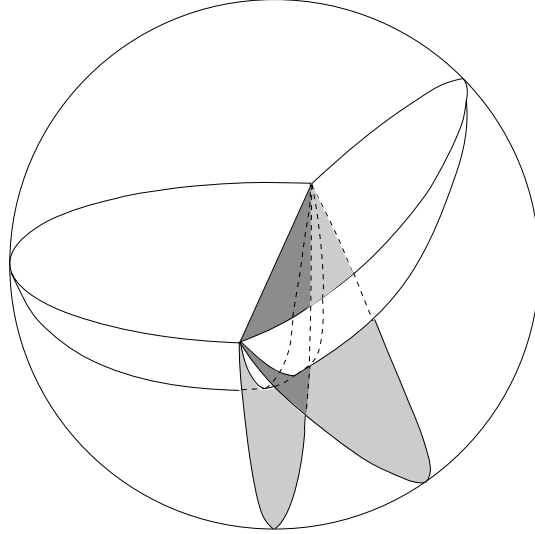
$$D_0(T, u, \zeta) \mapsto \begin{cases} \operatorname{ch} \delta (\operatorname{ch} \zeta \operatorname{ch} u, \operatorname{ch} \zeta \operatorname{sh} u, \operatorname{sh} \zeta, 0) + \operatorname{sh} \delta (0, 0, 0, 1) & \text{if } \zeta \leq 0, \\ \operatorname{ch} \delta (\operatorname{ch} u, \operatorname{sh} u, 0, 0) & \\ \quad + \operatorname{sh} \delta (0, 0, -\sin \frac{\zeta}{\operatorname{tgh} \delta}, \cos \frac{\zeta}{\operatorname{tgh} \delta}) & \text{if } \zeta \in [0, \alpha_0/T], \\ \operatorname{ch} \delta (\operatorname{ch} \zeta' \operatorname{ch} u, \operatorname{ch} \zeta' \operatorname{sh} u, \operatorname{sh} \zeta' \cos \alpha_0, \operatorname{sh} \zeta' \sin \alpha_0) & \\ \quad + \operatorname{sh} \delta (0, 0, -\sin \alpha_0, \cos \alpha_0) & \text{otherwise} \end{cases}$$

where $\delta = \operatorname{arctgh}(1/T)$ and $\zeta' = \eta - \alpha_0/T$. This map is clearly smooth for $\zeta \neq 0, \alpha_0/T$. Since the derivatives of D_0 with respect to the coordinate fields glue along $\zeta = 0$ and $\zeta = \alpha_0 T$, the map D_0 is C^1 . It is not hard to see that the derivatives are locally Lipschitz. One can check by direct computation that $D_0^*(g)$ is obtained by the canonical Wick rotation. The same formulae hold on \mathcal{U}_0^0 , provided we replace $\mathcal{U}_0^+ \cup \mathcal{V}$ by $r^{-1}(\mathbb{R}_+ \cdot v_0)$, the inverse image of the open ray.

The holonomy $h_{\mathbb{H}}$. Recall that $(F, \lambda) \in \mathcal{ML}(S)$, $F \subset F^c \subset \hat{F} = \mathbb{H}^2/\Gamma$, $F = \mathring{H}/\Gamma$. Then (see [36]) the bending cocycle satisfies

$$B_\lambda(\gamma x, \gamma y) = \gamma \circ B(x, y) \circ \gamma^{-1}$$

for every $\gamma \in \Gamma$.

Figure 3. The image \mathcal{E}_0 of D_0 and its decomposition.

Consider a bending map

$$F_\lambda: \overset{\circ}{H} \rightarrow \mathbb{H}^3.$$

For $\gamma \in \Gamma$ let us define

$$h_{\mathbb{H}}(\gamma) = B_\lambda(x_0, \gamma x_0) \circ \gamma \in \text{PSL}(2, \mathbb{C}).$$

Clearly F_λ is $h_{\mathbb{H}}$ -equivariant. We eventually get that the Wick rotation descends on the quotient spacetime $Y = Y_0^\lambda$. This gives the required hyperbolic structure M on $Y (> 1)$, having as asymptotic boundary the projective surface $S_{\mathcal{P}}^\lambda$.

4.6 Flat vs. de Sitter Lorentzian geometry

In order to classify MGH de Sitter spacetimes of finite type in terms of complex projective structures, we refer to [60] where the case of *compact* Cauchy surfaces was treated. In fact we can check that all constructions work as well by simply letting the Cauchy surface be complete of finite type. Let us summarize the main steps of this classification:

(1) We associate with every complex projective structure on a surface of finite type S a so-called *standard* spacetime belonging to $\mathcal{MGH}_1(S)$. It turns out that it is future complete. By composing with the parametrization $m_{\mathcal{P}}: \mathcal{ML}(S) \rightarrow \mathcal{P}(S)$, we eventually construct the *injective* map $m_1: \mathcal{ML}(S) \rightarrow \mathcal{MGH}_1(S)$.

(2) We show that, possibly inverting the time orientation, every spacetime in $\mathcal{MGH}_1(S)$ is standard, that is $m_1: \mathcal{ML}(S) \rightarrow \mathcal{MGH}_1(S)/\pm$ is a bijection (with the same meaning of \pm as for m_0).

We recall the construction of these standard spacetimes. Given a projective structure on S with developing map

$$d: \tilde{S} \rightarrow S^2,$$

we perform a construction which is dual to the one made for the H -hulls. Recall the canonical stratification of \tilde{S} described in Remark 4.10. For every $p \in \tilde{S}$ let $\Delta(p)$ denote the stratum passing through p and $\Delta^*(p)$ be the maximal ball containing $\Delta(p)$. Now $d(\Delta^*(p))$ is a ball in S^2 which determines a hyperbolic plane in \mathbb{H}^3 . Let $\rho(p)$ denote the point in \mathbb{X}_1 corresponding to this plane: the map $\rho: \tilde{S} \rightarrow \mathbb{X}_1$ turns out to be continuous. There exists a unique timelike geodesic c_p in \mathbb{X}_1 joining $\rho(p)$ to $d(p)$ so we can define the map

$$\hat{d}: \Delta(p) \times (0, +\infty) \ni (p, t) \mapsto c_p(t) \in \mathbb{X}_1.$$

This map is a developing map for the required standard de Sitter spacetime. A compatible holonomy follows by a natural equivariant version of the construction.

Assume now that the projective structure is encoded by $(F, \lambda) \in \mathcal{ML}(S)$, via $m_{\mathcal{P}}$. We eventually realize that the construction of m_1 can be obtained via a canonical rescaling performed on $Y_\lambda^0 (< 1)$. More precisely, we realize \hat{d} as a C^1 developing map

$$d_1^\lambda: \mathcal{U}_\lambda(< 1) \rightarrow \mathbb{X}_1$$

obtained as a sort of semi-analytic continuation of the hyperbolic developing map $d_{\mathbb{H}}$ constructed in the previous section, and we have:

Theorem 4.11. *The spacetime $\mathcal{U}_\lambda^1(Y_\lambda^1)$, obtained from $\mathcal{U}_\lambda^0(< 1)$ ($Y_\lambda^0(< 1)$) via the rescaling directed by the gradient of its cosmological time T and with rescaling functions*

$$\alpha = \frac{1}{1 - T^2} \quad \text{and} \quad \beta = \frac{1}{(1 - T^2)^2}$$

is the standard de Sitter spacetime corresponding to the projective structure on $\mathcal{U}_\lambda^0(1)$ ($Y_\lambda^0(1)$) produced by the Wick rotation.

The construction of d_1^λ is very simple. We regard both \mathbb{H}^3 and \mathbb{X}_1 as open sets of the real projective space (Klein models), separated by the quadric S^2 . If s is a geodesic integral line of the gradient of the cosmological time, $s_{>1} = s \cap \mathcal{U}_\lambda^0(> 1)$ is sent by $d_{\mathbb{H}}$ onto a geodesic ray of \mathbb{H}^3 . We define d_1^λ on $s_{<1}$ in such a way that it parameterizes the timelike geodesic ray in \mathbb{X}_1 contained in the projective line (in the Klein model) determined by $d_{\mathbb{H}}(s_{>1})$.

The proof, as well as the explicit computation for our favourite local models (evoked in the figure) are similar to the ones made for the Wick rotation, so we omit them.

An essential step in proving that m_1 is a bijection consists in

Proposition 4.12. (1) Every $Y \in \mathcal{MGH}_1(S)$ has C^1 cosmological time, and every level surface is a complete Cauchy surface.

(2) If $Y = Y_\lambda^1$ with universal covering \mathcal{U}_λ^1 , then the cosmological time of \mathcal{U}_λ^1 is

$$\tau = \operatorname{arctgh}(T),$$

T being the cosmological time of $\mathcal{U}_\lambda^0 (< 1)$. Hence the inverse rescaling is directed by the gradient of τ and has universal rescaling functions.

(3) Let Σ^0 be the initial singularity of \mathcal{U}_λ^0 . Then the map d_λ^1 extends to a continuous map

$$\mathcal{U}_\lambda^0 (\leq 1) \cup \Sigma^0 \rightarrow \mathbb{X}_1 \cup S^2.$$

Moreover, its restriction to $\mathcal{U}_\lambda^0(1)$ coincides with $d_{\mathbb{H}}$ and the restriction to Σ^0 is an (equivariant) isometry onto the initial singularity Σ^1 of \mathcal{U}_λ^1 .

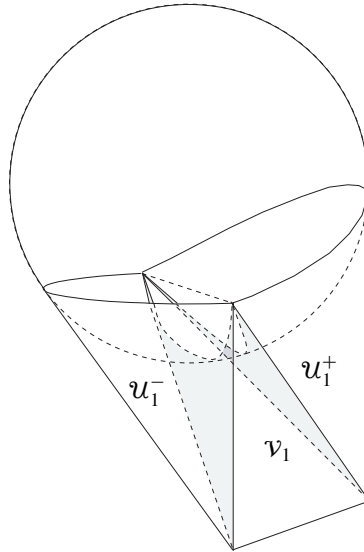


Figure 4. A standard de Sitter spacetime – local model.

Note that, in contrast with the flat Lorentzian case, these de Sitter developing maps as well as the dual hyperbolic ones are in general *not* injective.

4.7 Flat vs. Anti de Sitter Lorentzian geometry

We are going to outline first a few features of the spacetimes in $\mathcal{MGH}_{-1}(S)$. Recall the content of Section 2.2; in particular the duality between points x of \mathbb{X}_{-1} and spacelike planes $P(x)$ of \mathbb{X}_{-1} , or between spacelike lines, $l \mapsto l^*$. Recall also that the boundary

$\partial\mathbb{X}_{-1}$ has a natural causal structure, so that the notion of a *nowhere timelike simple closed curve* embedded in $\partial\mathbb{X}_{-1}$ makes sense.

Standard AdS spacetimes. Given such a curve $C \subset \mathbb{X}_{-1}$, assume furthermore that C is a *meridian* of $\partial\mathbb{X}_{-1}$ with respect to \mathbb{X}_{-1} (that means C is homologous to the difference of a left and right future-oriented leaves). We call C an *admissible achronal curve*. Then its *Cauchy development* is defined as

$$\mathcal{Y}(C) = \{p \in \mathbb{X}_{-1} \mid \partial P(p) \cap C = \emptyset\}$$

and the so obtained spacetime is called a (simply connected) *standard AdS spacetime*. C is called the *curve at infinity* of $\mathcal{Y}(C)$. An AdS spacetime is said to be standard if its universal covering is standard.

The convex core. There exists a spacelike plane P not intersecting $\mathcal{Y}(C)$ (see [55]). In the Klein model, we can cut \mathbb{P}^3 along the projective plane \hat{P} containing P and we have that $\mathcal{Y}(C)$ is contained in $\mathbb{R}^3 = \mathbb{P}^3 \setminus \hat{P}$.

Since C is nowhere timelike, the intersection of C with leaves of each foliation of $\partial\mathbb{X}_{-1}$ is either everywhere positive or negative. Since we assume C to be a meridian, for every point $p \in C$ the plane $P(p)$ tangent to $\partial\mathbb{X}_{-1}$ at p (that cuts \mathbb{X}_{-1} at a null totally geodesic plane) does not separate C : the intersection of $P(p)$ with $\partial\mathbb{X}_{-1}$ is the union of the left and right leaves through p , thus $P(p)$ intersects C only in p .

It follows that the convex hull $\mathcal{K}(C)$ of C in \mathbb{R}^3 is actually contained in \mathbb{X}_{-1} . We realize that $\mathcal{K}(C)$ does not depend on the choice of \hat{P} , and it is called the *convex core* of $\mathcal{Y}(C)$.

Support planes of $\mathcal{K}(C)$ are non-timelike and the closure $\overline{\mathcal{Y}(C)}$ of $\mathcal{Y}(C)$ in \mathbb{X}_{-1} coincides with the set of points dual to spacelike support planes of $\mathcal{K}(C)$ whereas the set of points dual to null support planes of $\mathcal{K}(C)$ coincides with C . $\overline{\mathcal{Y}(C)}$ is convex and the closure of $\mathcal{Y}(C)$ in $\overline{\mathbb{X}_{-1}}$ is $\overline{\mathcal{Y}(C)} \cup C$. It follows that $\mathcal{K}(C) \subset \overline{\mathcal{Y}(C)}$. A point $p \in \partial\mathcal{K}(C)$ lies in $\mathcal{Y}(C)$ if and only if it is touched only by spacelike support planes.

Being the boundary of a convex set in \mathbb{R}^3 , $\partial\mathcal{K}(C) \cup C$ is homeomorphic to a sphere. In particular, $\partial\mathcal{K}(C)$ (that is the boundary of $\mathcal{K}(C)$ in \mathbb{X}_{-1}) is obtained by removing a circle from a sphere, so it is the union of two disks. These components will be called *the past and the future boundary* of $\mathcal{K}(C)$ (with respect to time orientation), and denoted by $\partial_-\mathcal{K}(C)$ and $\partial_+\mathcal{K}(C)$ respectively. Given any inextensible timelike ray contained in $\mathcal{K}(C)$, its future endpoint lies on the future boundary, and the past endpoint lies on the past boundary.

The intersection $\partial_+\mathcal{K}(C) \cap \mathcal{Y}(C)$ is obtained by removing from $\partial_+\mathcal{K}(C)$ the set of points that admit a null support plane. Now suppose that a null support plane P passes through $x \in \partial_+\mathcal{K}(C)$. Then $P \cap \mathcal{K}(C)$ is a triangle with a vertex at some point $p \in \partial\mathbb{X}_{-1}$, two ideal edges (that are segments on the leaves of the double foliation of $\partial\mathbb{X}_{-1}$) and a complete spacelike geodesic of $\mathcal{K}(C)$. It follows that the set $\partial_+\mathcal{K}(C) \cap \mathcal{Y}(C)$ is obtained by removing from $\partial_+\mathcal{K}(C)$ (at most) countably many ideal triangles, so it is homeomorphic to an open disk. The only case where

$\partial_+ \mathcal{K}(C) \cap \mathcal{Y}(C)$ is empty is when the curve C is obtained by joining the endpoints of a spacelike geodesic l with the endpoints of its dual geodesic l^* . In that case $\mathcal{Y}(C) = \mathcal{K}(C)$, and we call it the *degenerate standard spacetime*. So, from now on, we incorporate in the definition of *standard AdS spacetime* that it is not degenerate. Moreover, since we will be mainly interested in $\partial_+ \mathcal{K}(C) \cap \mathcal{Y}(C)$, from now on we will use $\partial_+ \mathcal{K}(C)$ just to denote that set.

Proposition 4.13. $\partial_+ \mathcal{K}(C)$ is locally C^0 -isometric to \mathbb{H}^2 .

Remark 4.14. If $\partial_+ \mathcal{K}$ is complete then it is isometric to \mathbb{H}^2 . In general $\partial_+ \mathcal{K}$ is not complete, not even in the special case where C is the graph of a homeomorphism of S^1 onto itself. Moreover, it can be not complete even when there are no null triangles on the boundary.

The past part of a standard spacetime. The *past part* $\mathcal{P} = \mathcal{P}(C)$ of a standard AdS spacetime $\mathcal{Y}(C)$ is the past in $\mathcal{Y}(C)$ of the future boundary $\partial_+ \mathcal{K}$ of its convex core. The complement of $\partial_+ \mathcal{K}$ in the frontier of $\mathcal{P}(C)$ in \mathbb{X}_{-1} is called the *past boundary* of $\mathcal{Y}(C)$, denoted by $\partial_- \mathcal{P}$.

Proposition 4.15. Let \mathcal{P} be the past part of some $\mathcal{Y}(C)$. Then \mathcal{P} has cosmological time τ that takes values in $(0, \pi/2)$. For every point $p \in \mathcal{P}$ there exist only one point $\rho_-(p) \in \partial_- \mathcal{P}$, and only one point $\rho_+(p) \in \partial_+ \mathcal{K}$ such that the following holds:

- (1) p is on the timelike segment joining $\rho_-(p)$ to $\rho_+(p)$.
- (2) $\tau(p)$ is equal to the length of the segment $[\rho_-(p), p]$.
- (3) The length of $[\rho_-(p), \rho_+(p)]$ is $\pi/2$.
- (4) $P(\rho_-(p))$ is a support plane for \mathcal{P} passing through $\rho_+(p)$ and $P(\rho_+(p))$ is a support plane for \mathcal{P} passing through $\rho_-(p)$.
- (5) The map $p \mapsto \rho_-(p)$ is continuous. The function τ is C^1 and its gradient at p is the unit timelike tangent vector $\text{grad } \tau(p)$ such that

$$\exp_p(\tau(p)\text{grad } \tau(p)) = \rho_-(p).$$

Summing up, given the past part \mathcal{P} of a standard AdS spacetime $\mathcal{Y}(C)$, we can construct

- the cosmological time $\tau : \mathcal{P} \rightarrow (0, \pi/2)$;
- the future retraction $\rho_+ : \mathcal{P} \rightarrow \partial_+ \mathcal{K}$;
- the past retraction $\rho_- : \mathcal{P} \rightarrow \partial_- \mathcal{P}$.

Corollary 4.16. (1) Given r in the past boundary of \mathcal{Y} , $\rho_-^{-1}(r)$ is the set of points p such that the ray starting from r towards p meets at time $\pi/2$ the future boundary of \mathcal{K} .

(2) The image of ρ_- is the set of points of $\partial_- \mathcal{P}$ whose dual plane meets C at least in two points.

(3) The image of ρ_+ is the whole $\partial_+ \mathcal{K}$.

The image of the past retraction is called the *initial singularity* of $\mathcal{Y}(C)$.

For every surface of finite type S , $\text{Stand}_{-1}(S)$ denotes the Teichmüller-like space of standard AdS spacetimes admitting a Cauchy surface homeomorphic to S . The following is a fundamental step towards the classification.

Theorem 4.17. $\text{Stand}_{-1}(S) = \mathcal{MGH}_{-1}(S)$.

Note that a consequence of the inclusion \supset stated in this theorem is that, similarly to the flat case, the developing maps of finite type MGH AdS spacetimes are *embeddings* onto convex domains.

Both inclusions of the theorem are not trivial. The fact that every spacetime in $\mathcal{MGH}_{-1}(S)$ is standard follows from the following more general result (Section 7 of [55]). The fact that every standard spacetime is $\mathcal{MGH}_{-1}(S)$ is a consequence of the fact that $\mathcal{P}(a)$ is a complete Cauchy surface (see Proposition 4.19).

Proposition 4.18. *Let Y be an Anti de Sitter simply connected spacetime, and $F \subset Y$ be a complete Cauchy surface. Then the developing map $Y \rightarrow \mathbb{X}_{-1}$ is an embedding onto a convex subset of \mathbb{X}_{-1} .*

The closure of F in \mathbb{X}_{-1} is a closed disk and its boundary ∂F is a nowhere timelike curve of $\partial \mathbb{X}_{-1}$.

If Y is the maximal globally hyperbolic Anti de Sitter spacetime containing F then $Y = \mathcal{Y}(\partial F)$. The curve ∂F determines Y , namely $p \in Y$ if and only if the dual plane $P(p)$ does not meet ∂F .

Conversely ∂F is determined by Y . In fact ∂F is the set of accumulation points of Y on $\partial \mathbb{X}_{-1}$. If F' is another complete spacelike Cauchy surface of Y then $\partial F' = \partial F$.

The main step in order to prove the opposite inclusion is the following proposition (recently achieved also by Barbot [8] with a different approach with respect to [15]), that also holds for arbitrary standard spacetimes.

Proposition 4.19. *If \mathcal{P} is the past part of $\mathcal{Y}(C)$, then every level surface $\mathcal{P}(a)$ of the cosmological time is complete.*

Corollary 4.20. *Every level surface $\mathcal{P}(a)$ of the past part \mathcal{P} of a standard AdS spacetime $\mathcal{Y}(C)$ is a complete Cauchy surface of $\mathcal{Y}(C)$, and the latter is the maximal globally hyperbolic AdS spacetime that extends \mathcal{P} .*

Remark 4.21. The function τ extends to the cosmological time of $\mathcal{Y}(C)$, that takes values on some interval $(0, a_0(C))$, for some $a_0(C)$ satisfying $\pi/2 < a_0(C) < \pi$. Notice however that τ is C^1 only on the past part.

The map m_{-1} . Let $(F, \lambda) \in \mathcal{ML}(S)$, $F \subset F^c \subset \hat{F} = \mathbb{H}^2/\Gamma(F^c, \lambda)$, with universal coverings $\mathbb{H}^2 \rightarrow \hat{F}$, $H \rightarrow F^c$, and $\mathring{H} \rightarrow F$ respectively, as usual. Fix an embedding of \mathbb{H}^2 in \mathbb{X}_{-1} as a spacelike plane. The key ingredient to construct m_{-1} is the AdS version of the *bending of $H \subset \mathbb{H}^2$ along the lamination λ* (see below). This produces a *convex embedding* $\varphi_\lambda: \mathring{H} \rightarrow \mathbb{X}_{-1}$. Recall that to construct the H -hull (via the Wick rotation) we used the bending map $f_\lambda: \mathring{H} \rightarrow \mathbb{H}^3$, that is a local convex embedding, and then we followed the geodesic rays normal to $f_\lambda(\mathring{H})$, in the *non-convex* side bounded by $f_\lambda(\mathring{H})$. Eventually the developing map $d_{\mathbb{H}}$ has been obtained by requiring that the integral lines of the cosmological times are sent to the integral lines of the normal flow. Likewise, in the present situation, we construct a C^1 developing map

$$d_{-1}^\lambda: \mathcal{U}_\lambda^0 \rightarrow \mathbb{X}_{-1}$$

by requiring that the integral lines of the cosmological time of \mathcal{U}_λ^0 are sent to the integral lines of the normal flow. An important difference with respect to the hyperbolic case is that the normal flow is followed now in the *convex* side bounded by $\varphi_\lambda(\mathring{H})$ (otherwise singularities would be reached). It turns out that the image of d_{-1}^λ is the past part of a standard AdS spacetime, that plays here the role of a sort of AdS-hull. More precisely, we have:

Theorem 4.22. (1) d_{-1}^λ is an embedding onto the past part \mathcal{P}^λ of a determined $\mathcal{U}_{-1}^\lambda = \mathcal{Y}(C^\lambda)$, which is the universal covering of $Y_{-1}^\lambda \in \mathcal{MGH}_{-1}(S)$. The image of the AdS bending map φ_λ coincides with $\partial_+ \mathcal{K}(C^\lambda)$. The map d_{-1}^λ continuously extends to an isometry between the respective initial singularities.

(2) \mathcal{U}_{-1}^λ is produced by the rescaling of \mathcal{U}_0^λ , directed by the gradient of the cosmological time T , with universal rescaling functions

$$\alpha = \frac{1}{1 + T^2} \quad \text{and} \quad \beta = \frac{1}{(1 + T^2)^2}.$$

(3) The cosmological time τ on \mathcal{P}^λ is given by

$$T(p) = \tan \tau(p).$$

In this way, we construct an *injective* map

$$m_{-1}: \mathcal{ML}(S) \rightarrow \mathcal{MGH}_{-1}(S).$$

The following general proposition (specialized to $\text{Stand}_{-1}(S)$) implies that m_{-1} is in fact a *bijection*.

Proposition 4.23. For every standard AdS domain $\mathcal{Y} = \mathcal{Y}(C)$, the rescaling of its past part \mathcal{P} , directed by the gradient of the cosmological time τ , with universal rescaling functions

$$\alpha = \frac{1}{\cos^2 \tau} \quad \text{and} \quad \beta = \frac{1}{\cos^4 \tau}$$

produces a regular domain, whose cosmological time is given by the formula

$$T = \tan \tau.$$

It is not too hard to see, by means of local considerations, that such a rescaling produces a flat spacetime. Showing that it is a regular domain is actually more demanding. This is equivalent to showing that the future boundary of the convex core is isometric to a straight convex set pleated at a measured lamination. The key point is the fact that level surfaces of τ are complete (Proposition 4.20).

On AdS bending. We are going to outline more precisely the construction of d_{-1}^λ . The AdS bending runs similarly to the hyperbolic one, but with some remarkable differences (that are eventually responsible, for example, that the AdS developing maps are embeddings, in contrast with the hyperbolic ones). We also stress that orientations play a subtle role in the AdS bending procedure. The basic difference arises from the different behaviour of the “angles” between hyperbolic planes (that is, spacelike planes) and of “rotations” around spacelike geodesics in \mathbb{X}_{-1} , with respect to \mathbb{H}^3 . In fact, given two spacelike planes P_1, P_2 meeting each other along a geodesic l , the dual points $x_i = x(P_i)$ lie on the geodesic l^* dual to l . Then, we define the *angle between P_1 and P_2* as the distance between x_1 and x_2 along l^* . Notice the following:

Fix P_1 , then by varying P_2 , the angles between them are well-defined numbers that span the whole interval $(0, +\infty)$.

Define a *rotation around a spacelike geodesic l* simply to be an isometry of \mathbb{X}_{-1} which pointwise fixes l . We have

Lemma 4.24. *Rotations around a geodesic l act freely and transitively on the dual geodesic l^* . Such an action induces an isomorphism between the set of rotations around l and the set of translations of l^* .*

By duality, rotations around l act freely and transitively on the set of spacelike planes containing l . Given two spacelike planes P_1, P_2 such that $l \subset P_i$, then there exists a unique rotation $T_{1,2}$ around l such that $T_{1,2}(P_1) = P_2$.

Lemma 4.25. *An isometry of \mathbb{X}_{-1} is a rotation around a geodesic if and only if it is represented by a pair (α, β) such that α and β are isometries of \mathbb{H}^2 of hyperbolic type with the same translation length.*

Given two spacelike planes P_1, P_2 meeting along a geodesic l , let (α, β) be the rotation taking P_1 to P_2 . Then the translation length τ of α coincides with the angle between P_1 and P_2 .

There is a natural definition of positive rotation around an oriented spacelike geodesic l (depending only on the orientations of l and \mathbb{X}_{-1}). Thus, an orientation on the dual line l^* is induced by requiring that positive rotations act by positive translations on l^* . In particular, if we take an oriented geodesic l in $P(\text{Id})$, and

denote by $X \in \mathfrak{sl}(2, \mathbb{R})$ the infinitesimal generator of positive translations along l then it is not difficult to show that the positive rotations around l are of the form $(\exp(-tX), \exp(tX))$ for $t > 0$. Actually, by looking at the action on the boundary we deduce that both the maps $(\exp(-tX), \text{Id})$ and $(\text{Id}, \exp(tX))$ rotate planes through l in the positive direction (see Figure 5).

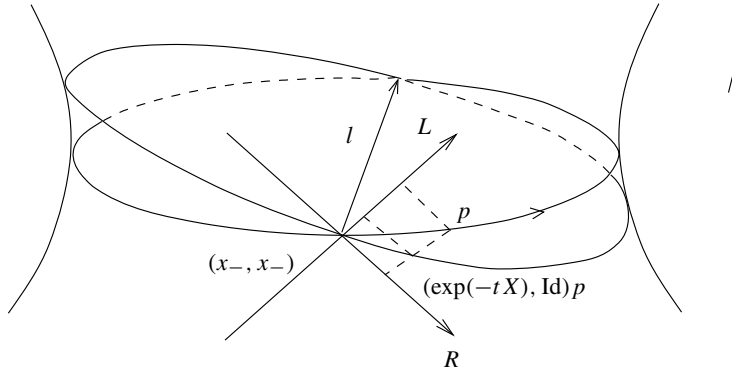


Figure 5. $(\exp(-tX), \text{Id})$ rotates planes around l in the positive sense.

Given λ on H as usual, we construct now an *AdS bending cocycle*

$$B^\lambda = (B_-^\lambda, B_+^\lambda): \mathring{H} \times \mathring{H} \rightarrow \text{PSL}(2, \mathbb{R}) \times \text{PSL}(2, \mathbb{R})$$

which formally satisfies properties similar to these of the quake cocycles of Section 3.6, or the above hyperbolic bending cocycle. In fact, B_-^λ and B_+^λ are exactly the Epstein–Marden cocycles (like the quake cocycles), corresponding to the *real-valued* measured laminations $-\lambda$ and λ . Here $-\lambda = (L, -\mu)$, that is, we take the *negative-valued* measure $-\mu$. Although this is no longer a measured lamination in the ordinary sense, the construction of [36] does apply. Besides the usual cocycle properties, B^λ also verifies that if x, y lie in different strata then $B_+^\lambda(x, y)$ (resp. $B_-^\lambda(x, y)$) is a non-trivial hyperbolic transformation whose axis separates the stratum through x and the stratum through y . Moreover the translation length is bigger than the total mass of $[x, y]$.

All this is very simple in the usual local model for finite laminations. In fact, take a finite measured geodesic lamination λ of \mathbb{H}^2 . Take a pair of points $x, y \in \mathbb{H}^2$ and enumerate the geodesics in λ that cut the segment $[x, y]$ in the natural way l_1, \dots, l_n . Moreover, we can orient l_i as the boundary of the half-plane containing x . With a little abuse, denote by l_i also the geodesic in $P(\text{Id})$ corresponding to l_i , then let $B^\lambda(x, y)$ be the isometry of \mathbb{X}_{-1} obtained by composition of positive rotations around l_i of angle a_i equal to the weight of l_i . In particular, if X_i denotes the unit positive generator of the hyperbolic transformations with axis equal to l_i , then we have

$$B^\lambda(x, y) = (B_-^\lambda(x, y), B_+^\lambda(x, y)) \in \text{PSL}(2, \mathbb{R}) \times \text{PSL}(2, \mathbb{R})$$

where

$$B_-^\lambda(x, y) = \exp(-a_1 X_1/2) \circ \exp(-a_2 X_2/2) \circ \cdots \circ \exp(-a_n X_n/2)$$

and

$$B_+^\lambda(x, y) = \exp(a_1 X_1/2) \circ \exp(a_2 X_2/2) \circ \cdots \circ \exp(a_n X_n/2)$$

with the following possible modifications: a_1 is replaced by $a_1/2$ when x lies on l_1 and a_n is replaced by $a_n/2$ when y lies on l_1 . The factor $1/2$ in the definition of β_\pm arises because the translation length of $\exp tX$ is $2t$.

By means of the bending cocycle, we construct an *AdS bending map*: take a basepoint x_0 in \mathring{H} and set

$$\varphi_\lambda : \mathring{H} \ni x \mapsto B^\lambda(x_0, x)x.$$

Proposition 4.26. *The bending map φ_λ is an isometric C^0 embedding of \mathring{H} onto an achronal set of \mathbb{X}_{-1} .*

Let $\mathcal{U} = \mathcal{U}_\lambda^0$ be the flat spacetime encoded by λ . Just as in the hyperbolic case, we “pull-back” the bending cocycle B^λ to a *continuous* bending cocycle

$$\hat{B}^\lambda : \mathcal{U} \times \mathcal{U} \rightarrow \mathrm{PSL}(2, \mathbb{R}) \times \mathrm{PSL}(2, \mathbb{R}).$$

In fact, we get a natural extension such that the following holds true:

(1) For every $p, q \in \mathcal{U}$ such that $N(p)$ and $N(q)$ do not lie on the weighted part of the lamination, we have

$$\hat{B}^\lambda(p, q) = B^\lambda(N(p), N(q)).$$

(2) \hat{B}^λ on the whole of \mathcal{U} is constant along the integral geodesics of the gradient of the cosmological time T .

(3) The extension is locally Lipschitzian (with respect to the Euclidean distance on \mathcal{U}), and the Lipschitz constant on $K \times K$ (K being any compact set in \mathcal{U}) depends only on the image of the Gauss map $N(K)$, the maximum of the total masses of geodesic paths of H joining points in $N(K)$, and the maximum and the minimum of the cosmological time T on K .

Finally, we can define our developing map $d_{-1}^\lambda : \mathcal{U}_0^\lambda \rightarrow \mathbb{X}_{-1}$. For every $p \in \mathcal{U}_\lambda^0$, we define $x_-(p)$ as the dual point of the plane $\hat{B}^\lambda(p_0, p)(P(\mathrm{Id}))$ (that is, $x_-(p) = \hat{B}^\lambda(p_0, p)(\mathrm{Id})$), and $x_+(p) = \hat{B}^\lambda(p_0, p)(N(p))$. Take representatives $\hat{x}_-(p)$ and $\hat{x}_+(p)$ in $\mathrm{SL}(2, \mathbb{R})$ such that the geodesic segment between $\hat{x}_-(p)$ and $\hat{x}_+(p)$ is future directed. Finally set

$$d_{-1}^\lambda(p) = [\cos \tau(p)\hat{x}_-(p) + \sin \tau(p)\hat{x}_+(p)]$$

where $\tau(p) = \arctan T(p)$.

As usual, we end with a few explicit computations for our favourite local model, that is when λ is a single weighted geodesic.

Let us set $\lambda_0 = (l_0, a_0)$ and choose a basepoint $p_0 \in \mathbb{H}^2 - l_0$. The surface $P = \varphi_\lambda(\mathbb{H}^2)$ is simply the union of two half-planes P_- and P_+ meeting along a

geodesic (that, with a little abuse of notation, is denoted by l_0). We can suppose that p_0 is in P_- , and l_0 is oriented as the boundary of P_- . If v_{\pm} denote the dual points of the planes containing P_{\pm} we have $v_- = \text{Id}$ and $v_+ = \exp(-a_0 X_0)$, X_0 being the standard generator of translations along l_0 . The vector X_0 is tangent to $P(\text{Id})$ along l_0 , orthogonal to it, and points towards p_0 .

By definition, the image, say \mathcal{P} , of $\Delta_0 = d_{-1}^{\lambda_0}$ is the union of three pieces: the cone with vertex at v_- and basis P_- , say \mathcal{P}_- , the cone with vertex at v_+ and basis P_+ , say \mathcal{P}_+ , and the join of the geodesic l_0 and the segment $[v_-, v_+]$, say \mathcal{Q} .

Fix a point in l_0 , say p_0 , and denote by v_0 the unit positively oriented tangent vector of l_0 at p_0 (that we will identify with a matrix in $M(2, \mathbb{R})$). Consider the coordinates on \mathcal{U}_0 , say (T, u, ζ) introduced in Section 4.5. With respect to these coordinates we have

$$\Delta_0(T, u, \zeta)$$

$$= \begin{cases} \sin \tau (\text{ch } \zeta (\text{ch } u \hat{p}_0 + \text{sh } u v_0) - \text{sh } \zeta X_0) + \cos \tau \hat{v}_- & \text{if } \zeta < 0, \\ \sin \tau (\text{ch } u \hat{p}_0 + \text{sh } u v_0) + \cos \tau \exp(-\zeta \tan \tau X_0) & \text{if } \zeta \in [0, a_0/T], \\ \sin \tau (\text{ch } \zeta' (\text{ch } u \hat{p}_0 + \text{sh } u v_0) - \text{sh } \zeta' X_0) + \cos \tau \hat{v}_+ & \text{otherwise} \end{cases}$$

where $\zeta' = \zeta - a_0/T$, $\tau = \arctan T$ and $\hat{p}_0, \hat{v}_+, \hat{v}_- \in \text{SL}(2, \mathbb{R})$ are chosen as above.

Clearly Δ_0 is C^∞ for $\zeta \neq 0, a_0/T$. A direct computation shows that the derivatives along the coordinate fields glue on $\zeta = 0$ and $\zeta = a_0/T$ and this proves that Δ_0 is C^1 .

By a direct computation we have

$$\Delta_0^*(\eta) = \begin{cases} -d\tau^2 + \sin^2 \tau (d\zeta^2 + \text{ch}^2 \zeta du^2) & \text{if } \zeta < 0, \\ -d\tau^2 + \sin^2 \tau (d\zeta^2 + du^2) & \text{if } \zeta \in [0, a_0/T], \\ -d\tau^2 + \sin^2 \tau (d\zeta^2 + \text{ch}^2 \zeta' du^2) & \text{otherwise.} \end{cases}$$

Since $d\tau^2 = \frac{dT^2}{(1+T^2)^2}$ and $\sin^2 \tau = \frac{T^2}{1+T^2}$, we finally see that Δ_0 is obtained by a rescaling directed by the gradient of T with the right rescaling functions.

Compatible holonomy. We obtain the holonomy representation of $Y_{-1}^\lambda \in \mathcal{M}\mathcal{G}\mathcal{L}(S)$, $h_{-1}^\lambda: \pi_1(S) \rightarrow \text{PSL}(2, \mathbb{R}) \times \text{PSL}(2, \mathbb{R})$, compatible with d_{-1}^λ as follows. If $x_0 \in \mathring{H}$ is the usual fixed basepoint of the construction, then for every $\gamma \in \pi_1(S) = \Gamma$ we have

$$h_{-1}^\lambda(\gamma) = B^\lambda(x_0, \gamma x_0) \circ (\gamma, \gamma).$$

Remark 4.27. It follows from the previous discussion that the spacetimes in $\mathcal{M}\mathcal{G}\mathcal{H}_{-1}(S)$ have a few analogies with the hyperbolic 3-manifolds arising as H -hulls of *quasi-Fuchsian* projective surfaces belonging to $\mathcal{P}(S)$. For instance, the curves at infinity $C \subset \partial\mathbb{X}_{-1}$ of $\mathcal{Y}(C)$ play a similar role than the Jordan curves that bound the universal coverings embedded in $\partial\mathbb{H}^3 = S^2$ of quasi-Fuchsian surfaces. However, there are important differences that make the AdS behaviour much more “tame”. For example such Jordan curves are in general rather wild, while the curves C are Lipschitz. Moreover, taking for example S compact, for every $(F, \lambda) \in \mathcal{M}\mathcal{L}(S)$, along

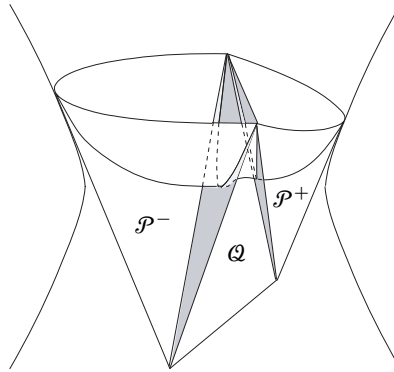


Figure 6. The domain \mathcal{P} with its decomposition. Also the surface $\mathcal{P}(a)$ is shown.

the ray $(F, t\lambda)$ there is a critical value $t_0 > 0$ such that $S^{t\lambda} \in \mathcal{P}(S)$ is quasi-Fuchsian only for $t < t_0$. On the other hand, the description of $Y^{t\lambda}$ is qualitatively the same for every $t > 0$; in particular all AdS developing maps are embeddings.

5 Causal AdS spacetimes, earthquakes and black holes

Beyond the classification achieved in the previous section, the AdS case displays a rich phenomenology that we are going to point out.

5.1 On holonomy information

Let us recall first the following results of [55], in the case of compact S .

Theorem 5.1. *If S is compact, and $Y \in \mathcal{M}\mathcal{G}\mathcal{H}_{-1}(S)$, then, seeing $\text{Isom}(\mathbb{X}_{-1})$ as $\text{PSL}(2, \mathbb{R}) \times \text{PSL}(2, \mathbb{R})$:*

- (a) *The holonomy $h = (h_L, h_R)$ of Y is made by a pair of Fuchsian representations of $\pi_1(S)$, and every such pair arises in this way (by varying Y).*
- (b) *Y is completely determined by its holonomy $h = (h_L, h_R)$. In fact $Y = \mathcal{Y}(C)$, where C is the graph in $S_\infty^1 \times S_\infty^1 = \partial\mathbb{X}_{-1}$ of the unique orientation preserving homeomorphism that conjugates the action of h_L on $S_\infty^1 = \partial\mathbb{H}^2$ with that of h_R . This curve C is the unique h -invariant curve on $\partial\mathbb{X}_{-1}$.*

In [8], [9] we can find the following generalization of point (a). Here we use the notation of Section 3.

Proposition 5.2. *Let $Y \in \mathcal{MGH}_{-1}(S)$ be of finite type, with holonomy representation*

$$h = (h_L, h_R): \pi_1(Y) \rightarrow \mathrm{PSL}(2, \mathbb{R}) \times \mathrm{PSL}(2, \mathbb{R}).$$

Then both h_L and h_R are holonomy representations of hyperbolic structures belonging to $\mathcal{T}(S)$. Conversely given a pair of representations $h = (h_L, h_R)$ corresponding to elements of $\mathcal{T}(S)$, then there exists a spacetime $Y \in \mathcal{MGH}_{-1}(S)$ whose holonomy is h .

Concerning point (b), the following partial generalization holds. Here we adopt the notation of Corollary 3.31.

Proposition 5.3. *$Y \in \mathcal{MGH}_{-1}(S)$ is completely determined by its holonomy provided that $Y = Y_{-1}^\lambda$ for some $\lambda \in \mathcal{V}_c(F) \cap \mathcal{ML}_c(F)^0$.*

This is essentially a consequence of the proof of the Earthquake Theorem considered below. On the other hand, non-equivalent spacetimes in $\mathcal{MGH}_{-1}(S)$ can actually share the same holonomy. This is the theme of the following construction.

5.2 Canonical causal AdS spacetimes with prescribed holonomy

We mostly refer to [8], [9]. Let us fix a representation $h = (h_L, h_R)$ of $\pi_1(S, p_0)$ as in Theorem 5.2. We stress that the representation is fixed, and not only its conjugacy class. For this reason, we have also fixed a basepoint $p_0 \in S$. We consider the domain

$$\tilde{\Omega}(h)$$

of points $x \in \mathbb{X}_{-1}$ such that, for every $\gamma \in \pi_1(S, p_0)$, x and $h(\gamma)(x)$ are not causally related. We can prove:

Proposition 5.4. *$\tilde{\Omega}(h)$ is simply connected and h -invariant; the action of $\pi_1(S, p_0)$ on it is free and properly discontinuous. The quotient, say $\Omega(h)$, is a causal AdS spacetime homeomorphic to $S \times \mathbb{R}$.*

Let us now consider the set

$$\mathcal{MGH}(h)$$

of all AdS MGH spacetimes Y homeomorphic to $S \times \mathbb{R}$ of the form $Y = \mathcal{Y}(C)/h$ such that the nowhere timelike curve at infinity C is h -invariant. Note again that each such maximal globally hyperbolic spacetime is fixed and *not* considered up to Teichmüller-like equivalence. However, the following is not hard to see:

Remark 5.5. *If $h' = ghg^{-1}$ is conjugate to h , then $\tilde{\Omega}(h') = g\tilde{\Omega}(h)$, and $\mathcal{MGH}(h') = g\mathcal{MGH}(h)$, so that they have the same image in $\mathcal{MGH}_{-1}(S)$. By fixing a representative h in any conjugation class, we get in this way an open partition of $\mathcal{MGH}_{-1}(S)$.*

Lemma 5.6. *The natural map $\mathcal{M}\mathcal{G}\mathcal{H}(h) \rightarrow \mathcal{M}\mathcal{G}\mathcal{H}_{-1}(S)$ is injective. By fixing a representative h in any conjugacy class we get a partition of $\mathcal{M}\mathcal{G}\mathcal{H}_{-1}(S)$.*

In fact we can prove

Proposition 5.7. *$\Omega(h)$ is the union of the $Y \in \mathcal{M}\mathcal{G}\mathcal{H}(h)$, and $\tilde{\Omega}(h)$ is the union of the h -invariant $\mathcal{Y}(C)$'s.*

See below for a description of the curves C arising in this way.

$\partial_\infty \tilde{\Omega}(h)$ and the limit set. For a generic representation h , $\mathcal{M}\mathcal{G}\mathcal{H}(h)$ contains more than one element. By Proposition 5.7 it follows that in general $\Omega(h)$ is not globally hyperbolic. In a sense, it is just the *maximal causal extension* of every globally hyperbolic $Y \in \mathcal{M}\mathcal{G}\mathcal{H}(h)$. The reason is that the boundary at infinity

$$\partial_\infty \tilde{\Omega}(h)$$

of $\tilde{\Omega}(h)$ in the boundary of \mathbb{X}_{-1} can have non-empty interior. Such a closure can be explicitly described by means of the holonomy of the peripheral loops. Let $\gamma \in \pi_1(S, p_0)$ freely homotopic to a loop surrounding a point in V . If $h_L(\gamma)$ (resp. $h_R(\gamma)$) is of hyperbolic type, we set $I_L(\gamma)$ (resp. $I_R(\gamma)$) to be the interval of S^1 whose endpoints are the fixed points of $h_L(\gamma)$ (resp. $h_R(\gamma)$) and that does not meet the limit set of h_L (resp. h_R). If $h_L(\gamma)$ is parabolic, then let $I_L(\gamma)$ be the fixed point of $h_L(\gamma)$. Similarly for $I_R(\gamma)$. Then the ‘‘rectangle’’ $R(\gamma) = I_L(\gamma) \times I_R(\gamma)$ is contained in the closure of $\tilde{\Omega}(h)$, and in fact $\partial_\infty \tilde{\Omega}(h)$ is the closure of $\bigcup_\gamma R(\gamma)$. The closure of the complement of the union of these rectangles $R(\gamma)$ in $\partial_\infty \tilde{\Omega}(h)$ can be regarded as a *limit set*

$$\Lambda = \Lambda(h)$$

in the sense that it is contained in the closure of the orbit of any point $x \in \overline{\mathbb{X}_{-1}}$.

A rectangle is *non-degenerate* if both $h_L(\gamma)$, $h_R(\gamma)$ are of hyperbolic type.

Lemma 5.8. *If some rectangle $R(\gamma)$ is non-degenerate, then the interior of $\partial_\infty \tilde{\Omega}(h)$ is not empty. $\Omega(h)$ is not globally hyperbolic if and only if there is some non-degenerate rectangles.*

It is possible to find points p, q close to $R(\gamma)$ such that $I^+(p) \cap I^-(q)$ is not pre-compact in \mathbb{X}_{-1} and this contradicts global hyperbolicity.

The asymptotic regions. Notice that a non degenerate rectangle $R(\gamma)$ has exactly two vertices that are the endpoints of a spacelike geodesic l_γ that is invariant under $(h_L(\gamma), h_R(\gamma))$. The boundary lines of $R(\gamma)$ together with l_γ span a surface $\tilde{H}(\gamma)$ embedded in $\tilde{\Omega}(h)$ made by two null triangles intersecting at l_γ . This surface divides $\tilde{\Omega}(h)$ in two components. The component whose closure in $\partial_\infty \tilde{\Omega}(h)$ is $R(\gamma)$ is called an *asymptotic region* of $\tilde{\Omega}(h)$, denoted by $\tilde{A}(\gamma)$, and $\tilde{H}(\gamma)$ is its *horizon*.

$(A(\gamma), H(\gamma))$ is invariant under $(h_L(\gamma), h_R(\gamma))$ and the quotient embeds in $\Omega(h)$ giving us an asymptotic region $A(\gamma)$ with horizon $H(\gamma)$. The latter is the union of two null annuli along a spacelike closed geodesic. The length of this spacelike geodesic is called the *size* of the horizon, whereas the *momentum* is the twist factor for the parallel transport along it. If l_L, l_R are the translation lengths of $h_L(\gamma)$ and $h_R(\gamma)$, the size is simply

$$s = (l_L + l_R)/2,$$

whereas the momentum is

$$m = (l_L - l_R)/2.$$

$\Omega(h)$ has exactly k asymptotic regions, where k is the number of points $p \in V$ such that the surrounding circle is of hyperbolic type for both h_L and h_R .

More about $\mathcal{Y}(C) \subset \tilde{\Omega}(h)$. Clearly the h -invariant curve at infinity C is contained in $\partial_\infty \tilde{\Omega}(h)$. On the other hand, every nowhere timelike meridian of $\partial \mathbb{X}_{-1}$ contained in $\partial_\infty \tilde{\Omega}(h)$ is determined by drawing in each non-degenerate rectangle $R(\gamma)$ an arc l_γ joining the vertices that are the endpoints of the spacelike geodesic of the corresponding horizon. In the degenerate case the segment l_γ coincides with $R(\gamma)$. The closure C of the union of these l_γ 's is a nowhere timelike meridian. Moreover, if the segments are chosen in an h -invariant way (that is, $l_{\alpha\gamma\alpha^{-1}} = h(\alpha)l_\gamma$), then C is the curve at infinity of some $\mathcal{Y}(C) \subset \tilde{\Omega}(h)$.

5.3 AdS bending and Earthquake Theorems

By extending the arguments given in [55] in the case of compact S , we have for a general S of finite type:

Proposition 5.9. *Let $Y \in \mathcal{M}\mathcal{G}\mathcal{H}_{-1}(S)$ be encoded by $(F, \lambda) \in \mathcal{ML}(S)$, and let $h = (h_L, h_R)$ be its holonomy. Then h_L (resp. h_R) is the holonomy of the surface $F_L = \beta_\lambda^L(F)$ (resp. $F_R = \beta_\lambda^R(F)$), that is, the surface in $\tilde{\mathcal{T}}(S)$ obtained by the left (right) earthquake on F along λ .*

We stress that Proposition 5.9, together with Theorem 5.1, actually gives an AdS proof of the Earthquake Theorem 3.33 when S is compact. For, given F^0, F^1 two hyperbolic structures on a compact surface S , there exists a unique spacetime $Y_{-1}^\lambda \in \mathcal{M}\mathcal{G}\mathcal{H}_{-1}(S)$ whose holonomy is $h = (h^0, h^1)$, where h^j is the hyperbolic holonomy of F^j . Then the left earthquake along 2λ transforms F^0 into F^1 .

Now let us consider the general case (S non necessarily compact), we consider the subset

$$\mathcal{M}\mathcal{G}\mathcal{H}_c(h)$$

of $\mathcal{M}\mathcal{G}\mathcal{H}(h)$ consisting of the spacetimes Y that satisfy the further condition of being encoded by pairs $(F, \lambda) \in \mathcal{ML}_c(S)$. In order to get such an AdS proof of the full

Earthquake Theorem 3.33, we need to characterize the spacetimes $Y = \mathcal{Y}(C)/h \in \mathcal{M}\mathcal{G}\mathcal{H}_c(h)$ in terms of the curve at infinity C . Consider again the general description of an h -invariant meridian C given above. A case of particular interest is when the segments l_γ are chosen on the boundary of $R(\gamma)$. Meridians C obtained in this way are called *extremal*. Notice that for each asymptotic region there are only two ways to choose such an arc: an upper extremal arc and a lower extremal arc. Thus, there are exactly 2^k h -invariant extremal arcs where k is defined as above. This holds also when $k = 0$; in such a case $\Omega(h) = \mathcal{Y}(C)/h$ is globally hyperbolic, and C is its extremal meridian. Finally we note the following nice geometric characterization (see [29]).

Proposition 5.10. *$\mathcal{Y}(C)$ is the universal covering of some $Y \in \mathcal{M}\mathcal{G}\mathcal{H}_c(h)$ if and only if C is an h -invariant extremal meridian.*

Corollary 5.11. *$\Omega(h)$ is globally hyperbolic if and only if it belongs to $\mathcal{M}\mathcal{G}\mathcal{H}_c(h)$ and is encoded by (F, λ) such that $F \in \mathcal{T}_{g,r}$ and the lamination does not enter the cusps.*

We can now sketch the proof of Theorem 3.33. Let F^0, F^1 be the interior of the convex cores of $\mathbb{H}^2/h^0, \mathbb{H}^2/h^1$ respectively, that are both homeomorphic to S . Set $h = (h^0, h^1)$ and take $\Omega(h)$. Let us apply Proposition 5.9 to every $Y \in \mathcal{M}\mathcal{G}\mathcal{H}_c(h)$, encoded by some $(F, \lambda) \in \mathcal{M}\mathcal{L}_c(S)$. As the convex cores are uniquely determined by the holonomy, and $\mathcal{T}_c(S)$ is closed under earthquakes, it follows that $F^0 = \beta_\lambda^L(F)$, $F^1 = \beta_\lambda^R(F)$, so that $F^1 = \beta_{2\lambda}^L(F^0)$. The determined lack of uniqueness in Theorem 3.33, the “enhanced” version 3.34, as well as Corollary 3.35 are now rather easy consequences of Proposition 5.10, Lemma 3.28, and the definition of the (enhanced) quake-flow.

5.4 Convex core of $\Omega(h)$ and black holes

We denote by

$$\Omega_c(h) \subset \Omega(h)$$

the union of the spacetimes belonging to $\mathcal{M}\mathcal{G}\mathcal{H}_c(h)$. We do similarly for $\tilde{\Omega}_c(h) \subset \tilde{\Omega}(h)$. It follows from Proposition 5.10 that the connected components of $\Omega(h) \setminus \Omega_c(h)$ coincide with the asymptotic regions defined above. Similarly for $\tilde{\Omega}(h) \setminus \tilde{\Omega}_c(h)$.

The limit set Λ is contained in the closure of $\tilde{\Omega}_c(h)$ which is the union of a finite number of globally hyperbolic spacetimes. Hence there is a spacelike plane P that does not intersect $\tilde{\Omega}_c(h)$, so that we can take the convex hull

$$\tilde{\mathcal{K}}(h)$$

of Λ in $\mathbb{R}^3 = \mathbb{P}^3 \setminus \hat{P}$, where \hat{P} is the projective plane containing P . It turns out that $\tilde{\mathcal{K}}(h)$ is contained in the closure of $\tilde{\Omega}_c(h)$, it is h -invariant and does not depend on the choice of P . $\tilde{\mathcal{K}}(h)$ is called the *convex core* of $\tilde{\Omega}(h)$, and its quotient

$$\mathcal{K}(h)$$

is the convex core of $\Omega(h)$. We can see that $\tilde{\Omega}(h)$ coincides with the set of points in \mathbb{X}_{-1} whose dual plane does not intersect $\tilde{\mathcal{K}}(h)$, and that every plane dual to some point of $\tilde{\mathcal{K}}(h)$ does not intersect $\tilde{\Omega}(h)$. The boundary of $\tilde{\mathcal{K}}(h)$ contains the spacelike geodesics of the horizons of $\tilde{\Omega}_c(h)$. Such a geodesic disconnects $\partial\tilde{\mathcal{K}}(h)$ into two h -invariant pleated surfaces whose quotients are homeomorphic to S . One of them, say $\partial_+\tilde{\mathcal{K}}(h)$, is in the future of the other one, say $\partial_-\tilde{\mathcal{K}}(h)$, and they are called the *future* and the *past boundary* of $\tilde{\mathcal{K}}(h)$ respectively. It turns out that $\partial_+\tilde{\mathcal{K}}(h)$ is obtained via the AdS bending of $(F_+, \lambda_+) \in \mathcal{ML}_c(S)$ (according to Section 4) so that it is the future boundary of the past part of a specific $\mathcal{Y}(C_+)/h \in \mathcal{MGH}_c(h)$. In fact:

The extremal h -invariant meridian C_+ is obtained by taking the lower extremal arc in each rectangle.

Similarly $\partial_-\tilde{\mathcal{K}}(h)$ is the *past* boundary of the *future* part of a specific spacetime $\mathcal{Y}(C_-)/h \in \mathcal{MGH}_c(h)$ (whose future boundary of the past part is obtained by bending a certain $(F_-, \lambda_-) \in \mathcal{ML}_c(S)$). The corresponding extremal h -invariant meridian C_- is obtained by taking the *upper* extremal arc in each rectangle. This makes sense also when $\Omega(h)$ is globally hyperbolic; in such a case $C_- = C_+$.

Assume now that $\Omega(h)$ is not globally hyperbolic. For each boundary component c_i of F_+ , l_{c_i} is the size of the corresponding horizon, whereas $I_{c_i}(\lambda_+)$ is the corresponding momentum. It follows that λ_+ belongs to the closure of $\mathcal{V}_c(F_+)$ (recall Corollary 3.31). In fact this property uniquely characterizes Y_+ within $\mathcal{MGH}_c(h)$. In particular this selects a *privileged* one among the earthquakes of Theorem 3.33. For (F_-, λ_-) we have the somehow *opposite* behaviour, that is, for every boundary component $I_{c_i}(\lambda_-) > l_{c_i}$.

Set

$$\tilde{B}(h) = \mathcal{Y}(C_-), \quad B(h) = \tilde{B}(h)/h, \quad \tilde{W}(h) = \mathcal{Y}(C_+), \quad W(h) = \tilde{W}(h)/h.$$

Denote by $KB(h)$, $KW(h)$ the respective convex cores as MGH spacetimes. We have

Proposition 5.12. (1) $\mathcal{K}(h) = KB(h) \cap KW(h)$.

$$(2) \tilde{\Omega}_c(h) = \tilde{B}(h) \cup \tilde{W}(h), \quad \Omega_c(h) = B(h) \cup W(h).$$

In Physics literature the special globally hyperbolic spacetime $B(h)$ ($W(h)$) is known as the *multi black hole* (*multi white hole*) contained in the causal spacetime $\Omega(h)$. The attribute “multi” mostly refer to the fact that it has a “multi” horizon. $B(h)$ looks like an honest black hole in the sense that every future inextensible causal curve emanating from any event in $B(h)$ never leaves $B(h)$ and eventually reaches the final singularity Σ_+ of $B(h)$ in finite time. In particular, lightlike rays emanating from $B(h)$ do not reach $\partial_\infty\Omega(h)$. So the final singularity Σ_+ is an actual singularity for the spacetime $\Omega(h)$ itself, as it reflects its future timelike geodesic incompleteness (at the initial singularity of $B(h)$ that is contained in its interior, $\Omega(h)$ is perfectly non singular). The initial singularity Σ_- of the white hole $W(h)$ plays a similar role with respect to the past. However, Σ_+ is “censored” by the multihorizon of $B(h)$, while Σ_- is a “naked” singularity. In Figure 7 we see a schematic picture of $\Omega(h)$ with its convex core and its black hole.

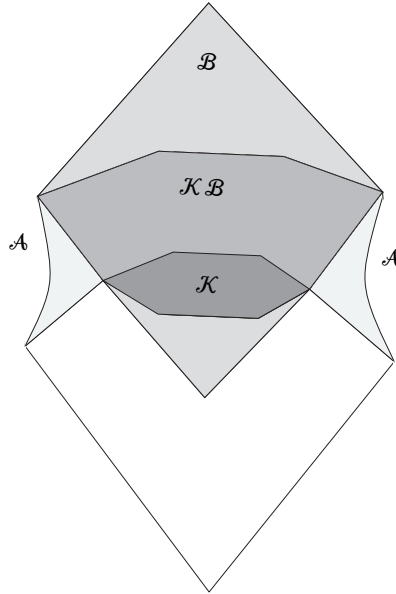


Figure 7. A schematic picture of Ω with its convex core \mathcal{K} and asymptotic regions \mathcal{A} . It contains the black hole \mathcal{B} with its convex core \mathcal{KB} containing \mathcal{K} .

Asymptotic regions and BTZ black holes. Every asymptotic region $R = R(\gamma)$ of $\Omega(h)$ has by itself a natural extension to a maximal *causal* AdS spacetime $\mathcal{B} = \mathcal{B}(\gamma)$, homeomorphic to $(S^1 \times \mathbb{R}) \times \mathbb{R}$. \mathcal{B} contains a maximal globally hyperbolic spacetime \mathcal{BH} , with a complete Cauchy surface homeomorphic to the annulus $S^1 \times \mathbb{R}$, which is known as the *BTZ black hole* contained in \mathcal{B} (see [6], [30]). \mathcal{B} has been particularly studied because it supports *Kerr-like metrics* with several qualitative analogies with the classical rotating black hole solutions of $(3 + 1)$ gravity. Let us briefly recall this matter. It is convenient to lift $\mathbb{X}_{-1} = \text{PSL}(2, \mathbb{R})$ to $\hat{\mathbb{X}}_{-1} = \text{SL}(2, \mathbb{R})$ so that it is given by the matrices of the form

$$X = \begin{pmatrix} T_1 + X_1 & T_2 + X_2 \\ -T_2 + X_2 & T_1 - X_1 \end{pmatrix}$$

such that $\det(X) = 1$, $0 < T_1^2 - X_1^2 < 1$, X_1, T_1 have a definite sign. We also fix a suitable $\text{SL}(2, \mathbb{C})$ -lifting of the isometry $(h_L(\gamma), h_R(\gamma))$ corresponding as above to the given asymptotic region. Let us assume for simplicity that it is of the form

$$\left(\begin{pmatrix} \exp(r_+ - r_-) & 0 \\ 0 & \exp(r_- - r_+) \end{pmatrix}, \begin{pmatrix} \exp((r_+ + r_-)) & 0 \\ 0 & \exp(-(r_+ + r_-)) \end{pmatrix} \right)$$

and that $r_+ > r_- \geq 0$. This isometry generates a group \mathcal{G} that acts on the whole of $\hat{\mathbb{X}}_{-1}$, with a constant vector field ξ as infinitesimal generator, and we have $q(\xi) =$

$(T_2^2 - X_2^2)r_+ + (T_1^2 - X_1^2)r_-$ (q is defined in Section 2). Roughly speaking, $\tilde{\mathcal{B}}$ is the maximal region of \mathbb{X}_{-1} such that:

- (1) $q(\xi) > 0$ on $\tilde{\mathcal{B}}$, so that we can take the function $r = q(\xi)^{1/2} > 0$;
- (2) $\{r_+ > r > r_-\} \subset \tilde{\mathcal{B}}$;

(3) $\tilde{\mathcal{B}}$ is \mathcal{G} -invariant, the group acts nicely and the quotient \mathcal{B} is a causal spacetime homeomorphic to $(S^1 \times \mathbb{R}) \times \mathbb{R}$.

$\tilde{\mathcal{B}}$ admits a \mathcal{G} -invariant “tiling” by regions of three types I, II, III contained in $\{r > r_+\}$, $\{r_+ > r > r_-\}$, $\{r_- > r\}$ respectively. Each region is bounded by suitable null horizons at which $r = r_{\pm}$. We can see that our asymptotic regions \tilde{A} are of type III.

By “joining” (the lifting of) the spacelike line l_γ with the two liftings of the dual line l_γ^* respectively, we get two “tetrahedra” say $\tilde{\mathcal{B}}\mathcal{H}$ and $\tilde{\mathcal{W}}\mathcal{H}$ embedded in $\tilde{\mathcal{B}}$ intersecting at l_γ . These are the two regions of type II that form the whole of $\{r_+ > r > r_-\}$. One projects onto the BTZ black hole $\mathcal{B}\mathcal{H}$, the other one covers the white hole embedded in \mathcal{B} , say $\mathcal{W}\mathcal{H}$. Note that both $\mathcal{B}\mathcal{H}$ and $\mathcal{W}\mathcal{H}$ are instances of “degenerate” globally hyperbolic spacetimes in the sense of Section 4.7.

For suitable coordinates (v, r, ϕ) on \mathcal{B} , where (r, ϕ) look like polar coordinates on the v -level surfaces, the Kerr-like metric is of the form

$$ds^2 = (M - r^2)dv^2 + f^{-1}dr^2 + r^2d\phi^2 - Jdv d\phi$$

where

$$M = r_+^2 + r_-^2, \quad J = 2r_+r_-, \quad M \geq J, \quad f = -M + r^2 + \frac{J^2}{4r^2}$$

and they are related to the previously defined “size” and “momentum” by

$$M + J = s^2, \quad M - J = m^2.$$

Each region of \mathcal{B} supports this metric, the null horizons of the regions being just “coordinate singularities”.

BTZ black holes naturally arise in the framework of Wick rotation-rescaling theory for the elementary surfaces of finite type, that is having Abelian fundamental group: $S = S^1 \times \mathbb{R}$ and $S = S^1 \times S^1$. This displays an interesting role of quadratic differentials instead of geodesic laminations. See Chapter 7 of [15] for more details.

5.5 (Broken) T -symmetry

Let $Y \in \mathcal{M}\mathcal{G}\mathcal{H}_{-1}(S)$. By reversing time orientation we get another spacetime $Y^* \in \mathcal{M}\mathcal{G}\mathcal{H}_{-1}$. This involution is called T -symmetry as the involution induced on $\mathcal{M}\mathcal{L}(S)$ via the map m_{-1} . If $\mathcal{Y}(C)$ is the universal covering of Y , then the universal covering of Y^* is $\mathcal{Y}^* = \mathcal{Y}(C^*)$, where C^* is the image of the curve C under the involution of $\partial\mathbb{X}_{-1} = S_\infty^1 \times S_\infty^1$

$$(x, y) \mapsto (y, x).$$

Moreover, the holonomy h^* of Y^* is obtained by exchanging the components of the holonomy h of Y

$$h = (h_-, h_+) \longleftrightarrow h^* = (h_+, h_-).$$

If $B(h)$ is the black hole of $\Omega(h)$, then $B(h)^* = W(h^*)$. The opposite behaviour “ $I_c(\lambda_+) \leq l_c$ vs $I_c(\lambda_-) > l_c$ ” at the future boundary of the respective past parts (see above), can be considered as the basic feature of “broken T -symmetry”. A particular instance is when $B(h)$ is encoded by (F, λ) such that $F \in \mathcal{T}_{g,r} = \mathcal{T}(S) \cap \mathcal{T}_c(S)$ (the smallest stratum of $\mathcal{T}_c(S)$), and λ enters the cusps (in Figure 8 we show an example of $B(h)$ where F has $g = 0, r = 3$, and the lamination is like in Example 3.15 with respect to a standard ideal triangulation of F by two triangles). In this case F^* belongs indeed to a higher dimensional cell of $\mathcal{T}_c(S)$ and the white hole $W(h^*)$ has the property that $I_{\lambda^*}(c) = l(c)$ at every boundary curve, and every asymptotic region has null momentum.

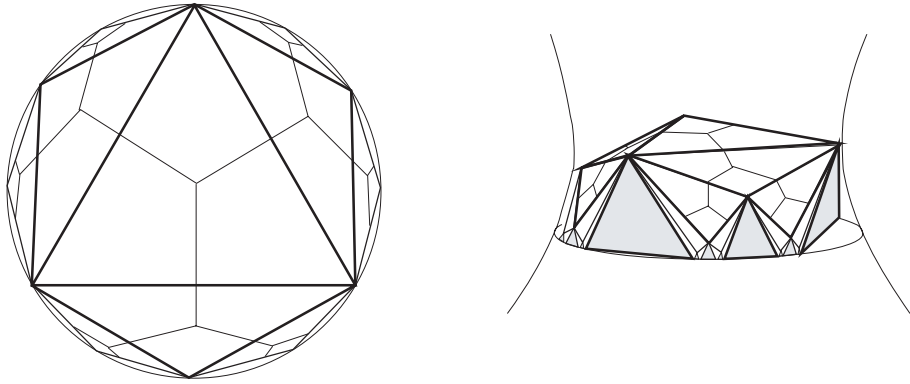


Figure 8. The convex core of a black hole $B(h)$. On the left the lamination with its dual spine. On the right the bending of \mathbb{H}^2 along λ in \mathbb{X}_{-1} . Grey regions are lightlike components of the past boundary of \mathcal{K}_λ .

6 Including particles

In 3-dimensional gravity massive point particles can be modeled as cone singularities along timelike lines. In particular, the rest mass m of a particle is related to the curvature k concentrated along its timelike geodesic “world line” by

$$k = 2\pi m, \quad k = 2\pi - \alpha$$

where α is the cone angle. If we require that the mass is positive, then it is bounded by $0 \leq m \leq 1$, while $2\pi \geq \alpha \geq 0$. However, there is no real geometric reason to exclude cone angles bigger than 2π .

It is a natural question whether Wick rotation-rescaling theory does apply also on cone spacetimes. In such a perspective, it is quite natural to extend the space $\mathcal{T}_c(S)$ defined in Section 3, by extending the notion of “type” to $\theta = V_{\mathcal{H}} \cup V_C$, and allowing hyperbolic structures F on S whose completion F^c is compact, with geodesic boundary components corresponding to the points of $V_{\mathcal{H}}$ and possibly *conical singularities* at the points of V_C . In particular we allow also holonomy of *elliptic type* at the circles surrounding these points. The parabolic holonomies correspond now to cone angles equal to 0, hence to particles of extremal mass. In order to preserve the conical structure, we consider measured geodesic laminations on such cone surfaces F that have *compact* support L in F , that is whose closure in F^c does not intersect the singularities.

For the sake of simplicity (and following [28] to which we will refer for most results stated in this section), from now on we will consider the particular case where $V_{\mathcal{H}} = \emptyset$. Even in this simplest case, a complete answer to the above question is unknown. Only a few partial results are known, mostly concerning the case of “small” cone angles ($< \pi$), or equivalently the case of particles with “big” masses.

Let $\hat{S}, S = \hat{S} \setminus V, V = \{p_1, \dots, p_r\}$ be as before. Let g be the genus of \hat{S} . Fix an r -tuple of angles $\mathcal{A} = (\alpha_1, \dots, \alpha_r)$, such that the “Gauss–Bonnet inequality”

$$\sum_j \left(1 - \frac{\alpha_j}{2\pi}\right) > 2 - 2g$$

holds; notice that we are not requiring here that the cone angles are smaller than 2π .

We denote by

$$\mathcal{T}_c(S, \mathcal{A})$$

the Teichmüller space of hyperbolic structures F on S whose completion F^c has conical singularities at p_1, \dots, p_r , of cone angles $\alpha_1, \dots, \alpha_r$.

By a general result of Troyanov [65], we have

Proposition 6.1. *The natural map*

$$\mathcal{T}_c(S, \mathcal{A}) \rightarrow \mathcal{T}_{g,r}$$

that associates to every $F \in \mathcal{T}_c(S, \mathcal{A})$ the unique complete hyperbolic structure of finite area \tilde{F} on S in the same conformal class of F , is a bijection.

This means in particular that $\mathcal{T}_c(S, \mathcal{A})$ is not empty. If $\alpha_j = 0$ for every j , then $\tilde{F} = F$ and $\mathcal{T}_c(S, 0)$ just coincides with $\mathcal{T}_{g,r}$.

For every $F \in \mathcal{T}_c(S, \mathcal{A})$, we denote by $\mathcal{ML}_c(F, \mathcal{A})$ the space of measured geodesic laminations on F with compact support. The space of all such (F, λ) 's is denoted by $\mathcal{ML}_c(S, \mathcal{A})$. When $\mathcal{A} = 0$, then $\mathcal{ML}_c(F, 0)$ just coincides with $\mathcal{ML}_c(\tilde{F})^0$ (defined in Section 3). More generally we have:

Proposition 6.2. *Assume that for every $j, \alpha_j < \pi$. Then there is a natural identification between $\mathcal{ML}_c(F, \mathcal{A})$ and $\mathcal{ML}_c(\tilde{F})^0$.*

We give a brief sketch of the proof of this proposition, referring to [28] for details. Since $\alpha_i < \pi$, for small $\epsilon > 0$, any non-peripheral loop on F admits a geodesic representative whose distance from V is at least ϵ .

Denote by Σ_ϵ the complement of a regular neighborhood of V in F^ϵ of ray ϵ . Given $\tilde{\lambda} \in \mathcal{ML}_\epsilon(\tilde{F})^0$, its support is contained in Σ_ϵ , for ϵ sufficiently small. Since Σ_ϵ is compact, it follows that the leaves of $\tilde{\lambda}$ are quasi-geodesic in F . So they can be stretched to become geodesic with respect to F . The union of all these leaves makes a geodesic lamination λ on F . A train-track carrying $\tilde{\lambda}$ carries also λ so the latter can be equipped with a transverse measure corresponding to the measure on $\tilde{\lambda}$.

Remark 6.3. If some cone angle α_i is bigger than π , then Proposition 6.2 fails. In fact it is not difficult to construct a surface with cone angles bigger than π and a loop c whose geodesic representative passes through the singular point.

6.1 Maximal globally hyperbolic spacetimes with particles

Since the causal structure of a spacetime with timelike geodesic world lines of conical singularities extends also on the singular locus, we can extend as well the notion of *Cauchy surfaces*. These turn out to be spacelike with conical singularities. Such a cone spacetime is said to be *globally hyperbolic* if it contains a Cauchy surface. Similarly to the smooth case, we can restrict our study to *maximal globally hyperbolic* ones. More precisely we set

$$\mathcal{MGH}_\kappa(S, \mathcal{A})$$

the Teichmüller-like space of cone spacetime structures h on $\hat{S} \times \mathbb{R}$, of constant curvature $\kappa \in \{-1, 0, 1\}$, such that

- h is non singular on $S \times \mathbb{R}$;
- h has a timelike geodesic line of conical singularity of angle α_i at each $\{p_i\} \times \mathbb{R}$;
- h is maximal globally hyperbolic and has a Cauchy surface orthogonal to the singular set;
- these structures are considered up to isotopies of $\hat{S} \times \mathbb{R}$ preserving $V \times \mathbb{R}$.

By easily adapting the constructions of Section 4 we get, for every $\kappa = 0, \pm 1$, the map

$$m_\kappa : \mathcal{ML}_\epsilon(S, \mathcal{A}) \rightarrow \mathcal{MGH}_\kappa(S, \mathcal{A}).$$

The main differences are that even for $\kappa = 0, -1$ the developing maps are no longer embeddings; moreover, the asymptotic complex projective structures produced by the Wick rotations have also conical singularities. As in the smooth case, the maps m_κ are injective. For the so obtained spacetimes have cosmological time, and one can recover the corresponding data (F, λ) by looking at level surfaces of cosmological time. Moreover, by construction, canonical Wick rotations and rescalings, directed by the gradient of cosmological times, with the usual universal rescaling functions, apply to the spacetimes belonging to the images of the maps m_κ .

On the other hand, the question of having an *intrinsic characterization of the images* $\text{Im}(\mathfrak{m}_\kappa)$ is largely open. In particular one asks to determine, for every κ , the angle assignments \mathcal{A} such that \mathfrak{m}_κ gives a parametrization of the whole of $\mathcal{MGH}_\kappa(S, \mathcal{A})$ (possibly inverting the time orientation). We have (see [28]):

Proposition 6.4. *If all the cone angles are less than π , then the spacetimes belonging to $\text{Im}(\mathfrak{m}_{-1})$ are precisely those admitting a convex Cauchy surface orthogonal to the singular locus.*

In fact, under such a “big masses” hypothesis, being in the image of \mathfrak{m}_{-1} turns out to be equivalent to admitting a *convex core*, that is a minimal convex subset. The convex core is homeomorphic to $\hat{S} \times \mathbb{R}$ and its boundary is the union of two $C^{0,1}$ -spacelike, intrinsically hyperbolic bent cone surfaces $\partial_+ \mathcal{K}(Y)$ and $\partial_- \mathcal{K}(Y)$, orthogonal to the singular locus. Just like the non-singular case, Y is encoded by (F, λ) if and only if the future boundary of its convex core is obtained by bending F along λ ; similarly for the past boundary, via T -symmetry.

One would expect that for big masses, the map \mathfrak{m}_{-1} actually is a bijection, that is, a convex Cauchy surface should always exist.

If some cone angle is bigger or equal than π , it is known that in general the maps \mathfrak{m}_κ are not onto, even if all masses are positive. For example in [17], by applying a so called “patchwork” construction, one produces flat MGH cone spacetimes with positive masses and with some cone angles equal to π , that do not belong to the image of \mathfrak{m}_0 . In fact it is remarkable that these spacetimes have nevertheless cosmological time whose level surfaces are orthogonal to the singular locus, and are *flat* instead of hyperbolic at the singular points of cone angle π . The canonical rescalings apply to them so that we finally also get spacetimes that do not belong to the images of $\mathfrak{m}_{\pm 1}$.

6.2 Earthquakes on hyperbolic cone surfaces

As every lamination $\lambda \in \mathcal{ML}_c(F, \mathcal{A})$ avoids the conical points, the notion of earthquake along such a lamination is defined as well. Similarly to the non-singular case, we have (see [28])

Theorem 6.5. *If Y has big masses, belongs to $\text{Im}(\mathfrak{m}_{-1})$, and is encoded by $(F, \lambda) \in \mathcal{ML}_c(S, \mathcal{A})$, then the left (resp. right) earthquake on F along λ produces surfaces $\beta_L(F, \lambda)$ (resp. $\beta_R(F, \lambda)$) $\in \mathcal{T}_c(S, \mathcal{A})$ whose holonomy coincides with the right (resp. left) holonomy of Y .*

Under the big masses hypothesis, let us consider the map

$$\mu: \text{Im}(\mathfrak{m}_{-1}) \rightarrow \mathcal{T}_c(S, \mathcal{A}) \times \mathcal{T}_c(S, \mathcal{A})$$

that associates to every Y the points obtained by left and right earthquake on (F, λ) respectively, as above. We have ([28])

Theorem 6.6. *The following equivalent facts hold.*

- (1) *Given $F, F' \in \mathcal{T}_c(S, \mathcal{A})$ there exists a unique $\lambda \in \mathcal{ML}_c(F, \mathcal{A})$ such that $\beta_L(F, \lambda) = F'$.*
- (2) *The map μ is bijective.*

Notice that the first statement is in purely hyperbolic terms. The equivalence between the two statements follows from Theorem 6.5. This equivalence between the hyperbolic and Lorentzian formulations plays a subtle role in the proof of Theorem 6.6. In fact by means of the hyperbolic formulation the map μ is proved to be locally injective, whereas Lorentzian geometry is used to prove that it is a proper map.

Finally we mention that in Chapter 7 of [15], we have described a quite different family of spacetimes with cone angles $\geq \pi$ (i.e. possibly with negative masses) that are governed by quadratic differentials rather than by measured geodesic laminations, and such that Wick rotation-rescaling machinery does apply to them.

References

- [1] W. Abikoff, *The real analytic theory of Teichmüller space*. Lecture Notes in Math. 820, Springer-Verlag, Berlin 1980.
- [2] L. Andersson, T. Barbot, R. Benedetti, F. Bonsante, W. M. Goldman, F. Labourie, K. P. Scannell and J. M. Schleker, Notes on a paper of Mess. *Geom. Dedicata* 126 (1) (2007), 47–70.
- [3] L. Andersson, G. J. Galloway and R. Howards, The cosmological time function. *Class. Quantum Grav.* 15 (1998), 309–322. [567](#)
- [4] L. Andersson, V. Moncrief and A. J. Tromba, On the global evolution problem in 2 + 1 gravity. *J. Geom. Phys.* 23 (3–4) (1997), 191–205. [535](#)
- [5] B. Apanasov, The geometry of Nielsen’s hull for a Kleinian group in space of quasi-conformal mappings. *Ann. Glob. Anal. Geom.* 6 (1998), 207–230.
- [6] M. Banados, M. Henneaux, C. Teitelboim and J. Zanelli, Geometry of the 2 + 1 black hole. *Phys. Rev. D* 48 (1993), 1506–1525. [600](#)
- [7] T. Barbot, Globally hyperbolic flat spacetimes. *J. Geom. Phys.* 53 (2005), 123–165. [535](#), [574](#), [576](#)
- [8] T. Barbot, Causal properties of AdS-isometry groups I: Causal actions and limit sets. *Adv. Theor. Math. Phys.* 12 (2008), 1–66. [535](#), [588](#), [594](#), [595](#)
- [9] T. Barbot, Causal properties of AdS-isometry groups II: BTZ multi black-holes. *Adv. Theor. Math. Phys.* 12 (2008), 1209–1257. [594](#), [595](#)
- [10] T. Barbot, G. Zeghib and F. Beguin, Constant mean curvature foliations of globally hyperbolic spacetimes locally modelled on AdS₃. *Geom. Dedicata* 126 (2007), 71–129. [535](#)
- [11] S. Baseilhac and R. Benedetti, Quantum hyperbolic invariants of 3-manifolds with PSL(2, \mathbb{C})-characters. *Topology* 43 (6) (2004), 1373–1423. [535](#)

- [12] S. Baseilhac and R. Benedetti, Classical and quantum dilogarithmic invariants of flat $\mathrm{PSL}(2, \mathbb{C})$ -bundles over 3-manifolds. *Geom. Topol.* 9 (2005), 493–569. [535](#)
- [13] S. Baseilhac and R. Benedetti, Quantum hyperbolic geometry. *Alg. Geom. Topology* 7 (2007), 845–917. [535](#)
- [14] J. T. Beem, P. E. Ehrlich and K. L. Easley, *Global Lorentzian Geometry*. 2nd ed., Monogr. Textbooks Pure Appl. Math. 202, Marcel Dekker, New York 1996. [538](#)
- [15] R. Benedetti and F. Bonsante, Canonical Wick rotations in 3-dimensional gravity. *Mem. Amer. Math. Soc.* 198, Amer. Math. Soc., Providence, RI, 2009. [534](#), [535](#), [562](#), [566](#), [574](#), [576](#), [588](#), [601](#), [606](#)
- [16] R. Benedetti and E. Guadagnini, Cosmological time in $(2 + 1)$ -gravity. *Nuclear Phys. B* 613 (2001), 330–352.
- [17] R. Benedetti and E. Guadagnini, Geometric cone surfaces and $(2 + 1)$ -gravity coupled to particles. *Nuclear Phys. B* 588 (1–2) (2000), 436–450. [535](#), [605](#)
- [18] R. Benedetti and E. Guadagnini, Classical Teichmüller theory and $(2 + 1)$ gravity. *Phys. Lett. B* 441 (1998), 60–68.
- [19] R. Benedetti and C. Petronio, *Lectures on Hyperbolic Geometry*. Universitext, Springer-Verlag, Berlin 1992. [548](#)
- [20] L. Bers, Uniformization by Beltrami equations. *Comm. Pure Appl. Math.* 14 (1961), 215–228.
- [21] F. Bonahon, Geodesic laminations on surfaces. In *Laminations and foliations in dynamics, geometry and topology* (Stony Brook, NY, 1998), Contemp. Math. 269, Amer. Math. Soc., Providence, RI, 2001, 1–37. [553](#), [554](#), [557](#)
- [22] F. Bonahon, Kleinian groups which are almost Fuchsian. *J. Reine Angew. Math.* 587 (2005), 1–15.
- [23] F. Bonahon, The geometry of Teichmüller space via geodesic currents. *Invent. Math.* 92 (1988), 139–162.
- [24] F. Bonahon, Shearing hyperbolic surfaces, bending pleated surfaces and Thurston’s symplectic form. *Ann. Fac. Sci. Toulouse Math.* 5 (1996), 233–297. [549](#)
- [25] F. Bonahon and J.-P. Otal, Laminations mesurées de plissage des variétés hyperboliques de dimension 3, *Ann. of Math.* 160 (2004), 1013–1055.
- [26] F. Bonsante, PhD Thesis. Scuola Normale Superiore, Pisa, 2005.
- [27] F. Bonsante, Flat Spacetimes with Compact Hyperbolic Cauchy Surfaces. *J. Differential Geom.* 69 (2005), 441–521.
- [28] F. Bonsante and J. M. Schlenker, AdS manifolds with particles and earthquakes on singular surfaces. *Geom. Funct. Anal.*, to appear; Preprint, arXiv:math.GT/0609116. [535](#), [603](#), [604](#), [605](#)
- [29] F. Bonsante, K. Krasnov and J. M. Schlenker, Multi black holes and earthquakes on Riemann surfaces with boundaries. Preprint, arXiv: math/0610429. [535](#), [553](#), [555](#), [565](#), [598](#)
- [30] S. Carlip, The $(2 + 1)$ -dimensional black hole. *Classical Quantum Gravity* 12 (12) (1995), 2853–2879. [600](#)

- [31] S. Carlip, *Quantum gravity in 2 + 1 dimensions*. Cambridge Monogr. Math. Phys., Cambridge University Press, Cambridge 1998.
- [32] Y. Choquet-Bruhat and R. Geroch, Global aspects of the Cauchy problem in general relativity. *Comm. Math. Phys.* 14 (1969), 329–335. [540](#)
- [33] S. Deser, R. Jackiw and G. 't Hooft, Three-dimensional Einstein gravity: dynamics of flat space. *Ann. Phys.* 152 (1) (1984), 220–235.
- [34] D. Dumas, Complex projective structures. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 455–508. [566](#), [567](#)
- [35] N. Dunfield, Cyclic surgery, degree of maps of character curves, and volume rigidity for hyperbolic manifolds, *Invent. Math.* 136 (1999), 623–657.
- [36] D. B. A. Epstein and A. Marden, Convex hulls in hyperbolic space, a theorem of Sullivan and measured pleated surfaces. In *Fundamentals of hyperbolic geometry: selected expositions*, London Math. Soc. Lectures Note Ser. 111, Cambridge University Press, Cambridge 1987, 113–253. [562](#), [579](#), [582](#), [591](#)
- [37] A. Fathi, F. Laudenbach and V. Poénaru (eds.), *Travaux de Thurston sur les surfaces*. *Astérisque* 66–67 (1979). [548](#), [557](#)
- [38] V. V. Fock and A. B. Goncharov, Dual Teichmüller and lamination spaces, In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 647–684. [550](#)
- [39] G. W. Gibbons, Real tunnelling geometries. In *Topology of the Universe Conference* (Cleveland, OH, 1997). *Classical Quantum Gravity* 15 (9) (1998), 2605–2612.
- [40] G. W. Gibbons and J.B. Hartle, Real tunneling geometries and the large-scale topology of the universe, *Phys. Rev. D* (3) 42 (8) (1990), 2458–2468.
- [41] W. M. Goldman, Projective structures with Fuchsian holonomy. *J. Differential Geom.* 25 (1987) 297–326.
- [42] W. M. Goldman, The Margulis invariant of isometric actions of Minkowski $(2 + 1)$ -space. In *Rigidity in Dynamics and Geometry* (Cambridge 2000), Springer-Verlag, Berlin 2002, 187–201.
- [43] W. M. Goldman and G. A. Margulis, Flat Lorentz 3-manifolds and cocompact Fuchsian groups. *Contemp. Math.* 262 (2000), 135–145.
- [44] S. Hawking and G. Ellis, *The large scale structure of space-time*. Cambridge University Press, Cambridge 1978. [538](#)
- [45] M. Kapovich, *Hyperbolic manifolds and discrete groups*. Progr. Math. 183, Birkhäuser, Basel 2001.
- [46] R. M. Kashaev, A link invariant from quantum dilogarithm. *Modern Phys. Lett. A* 10 (1995) 1409–1418
- [47] R. M. Kashaev, The hyperbolic volume of knots from the quantum dilogarithm. *Lett. Math. Phys.* 39 (1997), 269–275
- [48] S. P. Kerckhoff, The asymptotic geometry of Teichmüller space. *Topology* 19 (1) (1980), 23–41.
- [49] K. Krasnov and J. M. Schlenker, Minimal surfaces and particles in 3-manifolds. *Geom. Dedicata* 126 (2007), 187–254. [535](#)

- [50] R.S. Kulkarni and U. Pinkall, A canonical metric for Möbius structures and its applications. *Math. Z.* 216 (1) (1994), 89–129. [577](#), [578](#)
- [51] G. A. Margulis, Complete affine locally flat manifolds with free fundamental group. *J. Soviet Math.* 134 (1) (1987), 129–134.
- [52] H. Masur and J. Smillie *Quadratic differentials with prescribed singularities and pseudo-Anosov diffeomorphisms*, *Comment. Math. Helv.* 68 (2) (1993), 289–307.
- [53] H.-J. Matschull, The phase space structure of multi-particle models in $2 + 1$ gravity. *Classical Quantum Gravity* 18 (17) (2001), 3497–3560.
- [54] C. McMullen, Complex earthquakes and Teichmüller theory. *J. Amer. Math. Soc.* 11 (1998), 283–320.
- [55] G. Mess, Lorentz Spacetimes of Constant Curvature. *Geom. Dedicata* 126 (2007), 3–45. [534](#), [535](#), [536](#), [586](#), [588](#), [594](#), [597](#)
- [56] V. Moncrief, Reduction of the Einstein equations in $2 + 1$ dimensions to a Hamiltonian system over Teichmüller space. *J. Math. Phys.* 30 (12) (1989), 2907–2914. [535](#)
- [57] J.-P. Otal, *The hyperbolization theorem for fibered 3-manifolds*. SMF/AMS Texts Monogr. 7, Amer. Math. Soc., Providence, RI; Soc. Math. France, Paris 2001.
- [58] R. C. Penner and J. L. Harer, *Combinatorics of train tracks*. Ann. of Math. Stud. 125, Princeton University Press, Princeton, N.J., 1992.
- [59] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, Grad. Texts in Math., Springer-Verlag, New York 1994.
- [60] K. P. Scannell, Flat conformal structures and classification of de Sitter manifolds. *Comm. Anal. Geom.* 7 (1999), 283–320. [534](#), [583](#)
- [61] G. 't Hooft, Causality in $(2 + 1)$ -dimensional gravity. *Classical Quantum Gravity* 9 (5) (1992), 1335–1348.
- [62] G. 't Hooft, Nonperturbative 2 particle scattering amplitudes in $2 + 1$ -dimensional quantum gravity. *Comm. Math. Phys.* 117 (4) (1988), 685–700.
- [63] W. P. Thurston, Geometry and topology of 3-manifolds, Electronic Version 1.0 October 1997, <http://www.msri.org/gt3m/>.
- [64] W. P. Thurston, Earthquakes in two-dimensional hyperbolic geometry. In *Analytical and Geometric Aspects of Hyperbolic Space*, D. B. A. Epstein, ed., London Math. Soc. Lecture Notes Ser. 111, Cambridge University Press, London 1987.
- [65] M. Troyanov, Prescribing curvature on compact surfaces with cone singularities. *Trans. Amer. Math. Soc.* 324 (1991), 793–821. [603](#)
- [66] E. Witten, $2 + 1$ -dimensional gravity as an exactly soluble system. *Nuclear Phys. B* 311 (1) (1988/89), 46–78.
- [67] E. Witten, Topology-changing amplitudes in $(2 + 1)$ -dimensional gravity. *Nuclear Phys. B* 323 (1) (1989), 113–140.

Trace coordinates on Fricke spaces of some simple hyperbolic surfaces

William M. Goldman

Contents

1	Introduction	612
2	Traces in $SL(2, \mathbb{C})$	617
	2.1 Cyclic groups	618
	2.2 Two-generator groups	619
	2.3 Injectivity of the character map: the general case	626
3	Coxeter triangle groups in hyperbolic 3-space	629
	3.1 Lifting representations to $SL(2, \mathbb{C})$	629
	3.2 The three-holed sphere	630
	3.3 Orthogonal reflection groups	639
	3.4 Real characters and real forms	643
4	Hyperbolic structures on surfaces of $\chi = -1$	645
	4.1 Fricke spaces	647
	4.2 Two-dimensional hyperbolic geometry	648
	4.3 The three-holed sphere	651
	4.4 The one-holed torus	654
	4.5 Fenchel–Nielsen coordinates	657
	4.6 The two-holed cross-surface	660
	4.7 The one-holed Klein bottle	661
5	Three-generator groups and beyond	662
	5.1 The $SL(2, \mathbb{C})$ -character ring of \mathbb{F}_3	662
	5.2 The four-holed sphere	668
	5.3 The two-holed torus	674
	5.4 Orientable double covering spaces	675
	5.5 The two-holed cross-surface	677
	5.6 The one-holed Klein bottle	678
	5.7 Free groups of rank ≥ 3	680
	References	680

1 Introduction

The work of Fricke–Klein [21] develops the deformation theory of hyperbolic structures on a surface Σ in terms of the space of representations of its fundamental group $\pi = \pi_1(\Sigma)$ in $\mathrm{SL}(2, \mathbb{C})$. This leads to an algebraic structure on the deformation spaces. Here we expound this theory from a modern viewpoint.

We emphasize the close relationship between algebra and geometry. In particular algebraic properties of 2×2 matrices are applied to hyperbolic geometry in low dimensions. Our main object of interest is the *deformation space* of hyperbolic structures on a fixed compact surface-with-boundary Σ . The points of this deformation space correspond to equivalence classes of *marked* hyperbolic structures on $\mathrm{int}(\Sigma)$ where the ends are either cusps (complete ends of finite area) or are collar neighborhoods of closed geodesics. Such deformation spaces have been named *Fricke spaces* by Bers–Gardiner [3]. When Σ is closed, then the uniformization theorem identifies hyperbolic structures with conformal structures and the Fricke space is commonly identified with the *Teichmüller space* of marked *conformal* structures on Σ .

Hyperbolic structures are a special case of *locally homogeneous geometric structures* modelled on a homogeneous space of a Lie group G . These structures were first systematically defined by Ehresmann [15], and they determine representations of the fundamental group $\pi_1(\Sigma)$ in G . Equivalence classes of structures determine *equivalence classes* of representations, and the first part of this chapter deals with the algebraic problem of determining the moduli space of equivalence classes of pairs of unimodular 2×2 matrices.

Our starting point is the following well-known yet fundamental fact when π is a free group \mathbb{F}_2 of rank two. This fact may be found in the book of Fricke and Klein [21] and the even earlier paper of Vogt [76]. Perhaps much was known at the time about invariants of 2×2 matrices among the early practitioners of what has since become known as “classical invariant theory”. Now this algebraic work is contained in the powerful general theory developed by Procesi [67] and others, which in a sense completes the work begun in the 19th century.

Procesi’s theorem implies that the ring of invariants on the space of representations $\pi \xrightarrow{\rho} \mathrm{SL}(2, \mathbb{C})$ is generated by *characters*

$$\rho \mapsto \mathrm{tr}(\rho(\gamma)),$$

where $\gamma \in \pi$, and hence we call this ring the *character ring*. We begin by proving the elementary fact that character ring \mathfrak{R}_1 of a cyclic group is the polynomial ring $\mathbb{C}[\mathrm{tr}]$ generated by the *trace function* $\mathrm{SL}(2, \mathbb{C}) \xrightarrow{\mathrm{tr}} \mathbb{C}$. From this we proceed to the basic fact, that the character ring \mathfrak{R}_2 of the rank two free group \mathbb{F}_2 is a polynomial ring on *three variables*:

Theorem A (Vogt [76], Fricke [20]). *Let $\mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C}) \xrightarrow{f} \mathbb{C}$ be a regular function which is invariant under the diagonal action of $\mathrm{SL}(2, \mathbb{C})$ by conjugation.*

There exists a polynomial function $F(x, y, z) \in \mathbb{C}[x, y, z]$ such that

$$f(\xi, \eta) = F(\text{tr}(\xi), \text{tr}(\eta), \text{tr}(\xi\eta)).$$

Furthermore, for all $(x, y, z) \in \mathbb{C}^3$, there exists $(\xi, \eta) \in \text{SL}(2, \mathbb{C}) \times \text{SL}(2, \mathbb{C})$ such that

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \text{tr}(\xi) \\ \text{tr}(\eta) \\ \text{tr}(\xi\eta) \end{bmatrix}.$$

Conversely, if $x^2 + y^2 + z^2 - xyz \neq 4$ and $(\xi, \eta), (\xi', \eta') \in \text{SL}(2, \mathbb{C}) \times \text{SL}(2, \mathbb{C})$ satisfy

$$\begin{bmatrix} \text{tr}(\xi) \\ \text{tr}(\eta) \\ \text{tr}(\xi\eta) \end{bmatrix} = \begin{bmatrix} \text{tr}(\xi') \\ \text{tr}(\eta') \\ \text{tr}(\xi'\eta') \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

then $(\xi', \eta') = g.(\xi, \eta)$ for some $g \in G$.

Algebrao-geometrically, Theorem A asserts that the $\text{SL}(2, \mathbb{C})$ -character variety V_2 of a free group of rank two equals \mathbb{C}^3 . This will be our basic algebraic tool for describing moduli spaces of structures on the surface Σ and their automorphisms arising from transformations of Σ .

The condition $x^2 + y^2 + z^2 - xyz \neq 4$ also means that the matrix group $\langle \xi, \eta \rangle$ acts *irreducibly* on \mathbb{C}^2 . That is, $\langle \xi, \eta \rangle$ preserves no proper nonzero linear subspace of \mathbb{C}^2 . The condition that ξ, η generate an irreducible representation is crucial in several alternate descriptions of $\text{SL}(2, \mathbb{C})$ -representations of \mathbb{F}_2 . In particular, it is equivalent to the condition that the $\text{PGL}(2, \mathbb{C})$ -orbit is closed in $\text{Hom}(\mathbb{F}_2, \text{SL}(2, \mathbb{C}))$. This condition is in turn equivalent to the orbit being *stable* in the sense of Geometric Invariant Theory.

A more geometric description involves the action of the subgroup $\langle \xi, \eta \rangle \subset \text{SL}(2, \mathbb{C})$ on hyperbolic 3-space \mathbb{H}^3 . The group $\text{PSL}(2, \mathbb{C})$ acts by orientation-preserving isometries of \mathbb{H}^3 . An *involution*, that is, an element $g \in \text{PSL}(2, \mathbb{C})$ having order two, is reflection in a unique geodesic $\text{Fix}(g) \subset \mathbb{H}^3$. Denote the space of such involutions by Inv . Denote the identity by Id .

Theorem B (Coxeter extension). *Suppose that $\xi, \eta \in \text{SL}(2, \mathbb{C})$ generate an irreducible representation and let $\zeta = \eta^{-1}\xi^{-1}$ so that*

$$\xi\eta\zeta = \text{Id}.$$

Then there exists a unique triple of involutions

$$\iota_{\xi\eta}, \iota_{\eta\zeta}, \iota_{\zeta\xi} \in \text{Inv}$$

such that the corresponding elements $\mathbb{P}(\xi), \mathbb{P}(\eta), \mathbb{P}(\zeta) \in \mathrm{PSL}(2, \mathbb{C})$ satisfy

$$\begin{aligned}\mathbb{P}(\xi) &= \iota_{\zeta\xi}\iota_{\xi\eta}, \\ \mathbb{P}(\eta) &= \iota_{\xi\eta}\iota_{\eta\zeta}, \\ \mathbb{P}(\zeta) &= \iota_{\eta\zeta}\iota_{\zeta\xi}.\end{aligned}$$

From Theorem A follows the identification of the Fricke space of the three-holed sphere in terms of trace coordinates as $(-\infty, -2]^3$. The three trace parameters correspond to the three boundary components of Σ . From Theorem B follows the identification of the Fricke space of the three-holed sphere with the space of (mildly degenerate) right-angled hexagons in the hyperbolic plane \mathbb{H}^2 . (Right-angled hexagons are allowed to degenerate when some of the alternate edges covering boundary components degenerate to ideal points.)

The condition $x^2 + y^2 + z^2 - xyz \neq 4$ means that (ξ, η) defines an irreducible representation on \mathbb{C}^2 . This is equivalent to the condition that

$$\mathrm{tr}[\xi, \eta] \neq 2.$$

Thus the commutator trace plays an important role, partially because the fundamental group of the one-holed torus admits free generators X, Y such that the boundary component corresponds to $[X, Y]$. In particular trace coordinates identify the Fricke space of the one-holed torus with

$$\{(x, y, z) \in (2, \infty) \mid x^2 + y^2 + z^2 - xyz \leq 0\},$$

where the boundary trace equals

$$\mathrm{tr}[\xi, \eta] = x^2 + y^2 + z^2 - xyz \leq -2.$$

The trace coordinates are related to Fenchel–Nielsen coordinates. Similar descriptions of the Fricke spaces of the two-holed cross-surface (projective plane) and the one-holed Klein bottle are also given.

The character variety of \mathbb{F}_3 is more complicated. Let X_1, X_2, X_3 be free generators. The traces of the words

$$X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3, X_1X_2X_3, X_1X_3X_2$$

generate the $\mathrm{SL}(2, \mathbb{C})$ -character ring of \mathbb{F}_3 . We denote these functions by

$$x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_{123}, x_{132}$$

respectively. However, the character ring is not a polynomial ring on these generators, due to the trace identities expressing the *triple traces* x_{123} and x_{132} as the roots of a monic quadratic polynomial whose coefficients are polynomials in the *single traces* x_i

and *double traces* x_{ij} :

$$\begin{aligned}x_{123} + x_{132} &= x_{12}x_3 + x_{13}x_2 + x_{23}x_1 - x_1x_2x_3, \\x_{123} x_{132} &= (x_1^2 + x_2^2 + x_3^2) + (x_{12}^2 + x_{23}^2 + x_{13}^2) \\&\quad - (x_1x_2x_{12} + x_2x_3x_{23} + x_3x_1x_{13}) + x_{12}x_{23}x_{13} - 4.\end{aligned}$$

Furthermore the character variety is a hypersurface in \mathbb{C}^7 which is a double branched covering of \mathbb{C}^6 . In particular its coordinate ring, the *character ring*, is the quotient

$$\mathfrak{R}_3 := \mathbb{C}[x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_{123}]/\mathfrak{J}$$

by the principal ideal \mathfrak{J} generated by the polynomial

$$\begin{aligned}\Phi(x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_{123}) &:= x_1x_2x_3x_{123} + x_{12}x_{13}x_{23} \\&\quad - x_1x_2x_{12} - x_1x_3x_{13} - x_2x_3x_{23} \\&\quad - x_1x_{23}x_{123} - x_2x_{13}x_{123} - x_3x_{12}x_{123} \\&\quad + x_1^2 + x_2^2 + x_3^2 + x_{12}^2 + x_{13}^2 + x_{23}^2 + x_{123}^2 - 4.\end{aligned}$$

We use this description to discuss the Fricke spaces of the four-holed sphere $\Sigma_{0,4}$ and the two-holed torus $\Sigma_{1,2}$. In these cases, the generators X_i and their products correspond to curves on the surface, and we pay special attention to the elements corresponding to the boundary $\partial\Sigma$.

In particular we describe the homomorphisms on character rings induced by the orientable double coverings of the two-holed cross-cap $C_{0,2}$,

$$\Sigma_{0,4} \longrightarrow C_{0,2},$$

and the one-holed Klein bottle $C_{1,1}$,

$$\Sigma_{1,2} \longrightarrow C_{1,1},$$

respectively.

Finally we end with the important observation (see Vogt [76]) that the $\mathrm{SL}(2, \mathbb{C})$ -character ring \mathfrak{R}_n of a free group \mathbb{F}_n where $n \geq 4$, is generated by traces of words of length ≤ 3 .

This chapter began as an effort [31], to provide a self-contained exposition of Theorem A. Later it grew to include several results on hyperbolic geometry, which were used, for example in [30] but with neither adequate proofs nor references to the literature. In this version, we have tried to give a leisurely and elementary description of basic results on moduli of hyperbolic structures using trace coordinates. In Chapter 17 of this volume, written by Feng Luo [18], there is another point of view on the trace coordinates, following the Grothendieck reconstruction principle.

Acknowledgements. I am grateful to Hyman Bass, Steve Boyer, Richard Brown, Serge Cantat, Virginie Charette, Daryl Cooper, John Conway, Marc Culler, Todd Drumm, Art Dupre, Elisha Falbel, Carlos Florentino, Charlie Frohman, Alexander Gamburd, Jane Gilman, Ryan Hoban, Misha Kapovich, Linda Keen, François Labourie, Albert

Marden, John Millson, Maryam Mirzakhani, Greg McShane, Walter Neumann, John Parker, Julien Paupert, Bob Penner, Peter Sarnak, Caroline Series, Peter Shalen, Adam Sikora, George Stantchev, Ser-Peow Tan, Domingo Toledo, Richard Wentworth, Anna Wienhard, Scott Wolpert and Eugene Xia for their interest in this manuscript, and for providing valuable suggestions. In particular I wish to thank Sean Lawton, Elisha Peterson, and Ying Zhang for carefully reading parts of this manuscript and pointing out several technical mistakes. I would like to thank Guillaume Th  ret for help with the illustrations. I also wish to express my gratitude for the hospitality of the Mathematical Sciences Research Institute in Fall 2007, and the Institute for Advanced Study and Princeton University in Spring 2008, where this work was completed. Partial support by National Science Foundation grant DMS070781, a Semester Research Award (Fall 2005) from the General Research Board of the University of Maryland, and the Oswald Veblen Fund at the Institute for Advanced Study are gratefully acknowledged.

Notation and terminology. We mainly work over the field \mathbb{C} of complex numbers and its subfield \mathbb{R} of real numbers. Denote the ring of rational integers by \mathbb{Z} . We denote projectivization by \mathbb{P} , so that if V is a \mathbb{C} -vector space (respectively an \mathbb{R} -vector space), then $\mathbb{P}(V)$ denotes the set of all complex (respectively real) lines in V . Similarly if $V \xrightarrow{\xi} W$ is a linear transformation between vector spaces V, W , denote the corresponding projective transformation by $\mathbb{P}(\xi)$, wherever it is defined. For example the *complex projective line* $\mathbb{C}\mathbb{P}^1 = \mathbb{P}(\mathbb{C}^2)$. The noncommutative field of Hamilton quaternions is denoted \mathbb{H} . The set of positive real numbers is denoted \mathbb{R}_+ .

Denote the algebra of 2×2 matrices over \mathbb{C} by $M_2(\mathbb{C})$.

The trace and determinant functions are denoted tr and \det respectively. Denote the transpose of a matrix A by A^\top .

Let \mathbb{k} be a field (either \mathbb{R} or \mathbb{C}). Denote the multiplicative group of \mathbb{k} (the group of nonzero elements) by \mathbb{k}^* .

Let $n > 0$ be an integer. The *general linear group* is denoted $GL(n, \mathbb{k})$; for example $GL(2, \mathbb{C})$ is the group of all invertible 2×2 complex matrices. We also denote the group of scalar matrices

$$\mathbb{k}^*\text{Id} \subset GL(n, \mathbb{k})$$

by \mathbb{k}^* . The *special linear group* consists of all matrices in $GL(n, \mathbb{k})$ having determinant one, and is denoted $SL(n, \mathbb{k})$. The *projective linear groups* $PGL(n, \mathbb{k})$ (and respectively $PSL(n, \mathbb{k})$) are the quotients of $GL(n, \mathbb{k})$ (respectively $SL(n, \mathbb{k})$) by the central subgroup $\{\lambda\text{Id} \mid \lambda \in \mathbb{k}^*\}$ of scalar matrices, which we also denote \mathbb{k}^* .

If A, B are matrices, then their multiplicative commutator is denoted $[A, B] := ABA^{-1}B^{-1}$ and their additive commutator (their *Lie product*) is denoted $\text{Lie}(A, B) := AB - BA$.

If A is a transformation, denote its set of fixed points by $\text{Fix}(A)$. Denote the relation of conjugacy in a group by \sim . Denote free product of two groups A, B by $A * B$. If a_1, \dots, a_n are elements of a group, then $\langle a_1, \dots, a_n \rangle$ denotes the subgroup generated by a_1, \dots, a_n . The presentation of a group with generators g_1, \dots, g_m and relations

$r_1(g_1, \dots, g_m), \dots, r_n(g_1, \dots, g_m)$ is denoted

$$\langle g_1, \dots, g_m \mid r_1, \dots, r_n \rangle.$$

Denote the free group of rank n by \mathbb{F}_n . Denote the symmetric group on n letters by \mathfrak{S}_n .

Denote the (real) hyperbolic n -space by \mathbb{H}^n .

We briefly summarize the topology of surfaces.

A compact surface with n boundary components will be called n -holed. If M is a closed surface, then the complement in M of n open discs will be called an “ n -holed M .” For example a one-holed sphere is a disc and a two-holed sphere is an annulus.

We adopt the following notation for topological types of connected compact surfaces, beginning with orientable surfaces. $\Sigma_{g,n}$ denotes the n -holed (orientable) surface of genus g . Thus $\Sigma_{0,0}$ is a sphere, $\Sigma_{1,0}$ is a torus, $\Sigma_{0,1}$ is a disc and $\Sigma_{0,2}$ is an annulus.

The connected sum operation $\#$ satisfies

$$\Sigma_{g_1, n_1} \# \Sigma_{g_2, n_2} \approx \Sigma_{g_1+g_2, n_1+n_2}.$$

Other basic facts about orientable surfaces involve the Euler characteristic and the fundamental group:

$$\chi(\Sigma_{g,n}) = 2 - 2g - n$$

and if $n > 0$, the fundamental group $\pi_1(\Sigma_{g,n})$ is free of rank $2g + n - 1$.

For non-orientable surfaces, our starting point is the topological surface $C_{0,0}$ homeomorphic to the real projective plane, which J. H. Conway has proposed calling a *cross-surface*. We denote the n -holed $k + 1$ -fold connected sum of cross-surfaces by $C_{k,n}$. Thus the Möbius band is represented by $C_{0,1}$ and the Klein bottle by

$$C_{1,0} \approx C_{0,0} \# C_{0,0}.$$

The operation of connected sum satisfies

$$\Sigma_{g,n_1} \# C_{k,n_2} \approx C_{2g+k, n_1+n_2}, \quad \text{and} \quad C_{k_1, n_1} \# C_{k_2, n_2} \approx C_{k_1+k_2+1, n_1+n_2}.$$

The Euler characteristic and the fundamental group satisfy

$$\chi(C_{k,n}) = 1 - n - k$$

and $\pi_1(C_{k,n})$ is free of rank $n + k$ if $n > 0$.

The orientable double covering space of $C_{g,n}$ is $\Sigma_{g,2n}$.

2 Traces in $\mathrm{SL}(2, \mathbb{C})$

The purpose of this section is an elementary and relatively self-contained proof of Theorem A. This basic result explicitly describes the $\mathrm{SL}(2, \mathbb{C})$ -character variety of a rank-two free group as the affine space \mathbb{C}^3 , parametrized by the traces of the free

generators X, Y and the trace of their product XY . Apparently due to Vogt [76], it is also in the work of Fricke [20] and Fricke–Klein [21].

We motivate the discussion by starting with the simpler case of conjugacy classes of single elements, that is *cyclic groups* (free groups of *rank one*). In this case the $\mathrm{SL}(2, \mathbb{C})$ -character variety V_1 is the *affine line* \mathbb{C}^1 , parametrized by the trace.

2.1 Cyclic groups

Theorem 2.1.1. *Let $\mathrm{SL}(2, \mathbb{C}) \xrightarrow{f} \mathbb{C}$ be a polynomial function invariant under inner automorphisms of $\mathrm{SL}(2, \mathbb{C})$. Then there exists a polynomial $F(t) \in \mathbb{C}[t]$ such that $f(g) = F(\mathrm{tr}(g))$. Conversely, if $g, g' \in \mathrm{SL}(2, \mathbb{C})$ satisfy*

$$\mathrm{tr}(g) = \mathrm{tr}(g') \neq \pm 2,$$

then $g' = hgh^{-1}$ for some $h \in \mathrm{SL}(2, \mathbb{C})$.

Proof. Suppose f is an invariant function. For $t \in \mathbb{C}$, define

$$\xi_t := \begin{bmatrix} t & -1 \\ 1 & 0 \end{bmatrix}$$

and define $F(t)$ by

$$F(t) = f(\xi_t).$$

Suppose that $t \neq \pm 2$ and $\mathrm{tr}(g) = t$. Then g and ξ_t each have distinct eigenvalues

$$\lambda_{\pm} = \frac{1}{2}(t \pm (t^2 - 4)^{1/2})$$

and $hgh^{-1} = \xi_t$ for some $h \in \mathrm{SL}(2, \mathbb{C})$. Thus

$$f(g) = f(h^{-1}\xi_t h) = f(\xi_t) = F(t)$$

as desired. If $t = \pm 2$, then by taking Jordan normal form, either $g = \pm \mathrm{Id}$ or g is conjugate to ξ_t . In the latter case, $f(g) = F(t)$ follows from invariance. Otherwise g lies in the closure of the $\mathrm{SL}(2, \mathbb{C})$ -orbit of ξ_t and $f(g) = f(\xi_t) = F(t)$ follows by continuity of f .

The converse direction follows from Jordan normal form as already used above. \square

The map

$$\mathrm{SL}(2, \mathbb{C}) \xrightarrow{\mathrm{tr}} \mathbb{C}$$

is a *categorical* quotient map in the sense of algebraic geometry, although it fails to be a quotient map in the usual sense. The discrepancy occurs at the critical level sets $\mathrm{tr}^{-1}(\pm 2)$. The critical values of tr are ± 2 , and the restriction of tr to the regular set

$$\mathrm{tr}^{-1}(\mathbb{C} \setminus \{\pm 2\})$$

is a quotient map (indeed a holomorphic submersion). The critical level set $\text{tr}^{-1}(2)$ consists of all unipotent matrices, and these are conjugate to the one-parameter subgroup

$$\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix},$$

where $t \in \mathbb{C}$. For $t \neq 0$, these matrices comprise a single orbit. This orbit does *not* contain the identity matrix Id (where $t = 0$), although its closure does. Any regular function cannot separate a non-identity unipotent matrix from Id . Thus $\text{tr}^{-1}(2)$ contains two orbits: the non-identity unipotent matrices, and the identity matrix Id . Similar remarks apply to the other critical level set $\text{tr}^{-1}(-2) = -\text{tr}^{-1}(2)$.

For example,

$$\begin{aligned} \text{SL}(2, \mathbb{C}) &\longrightarrow \mathbb{C}, \\ \xi &\longmapsto \text{tr}(\xi^2) \end{aligned}$$

is an invariant function and can be expressed in terms of $\text{tr}(\xi)$ by

$$\text{tr}(\xi^2) = \text{tr}(\xi)^2 - 2 \tag{2.1}$$

which follows from the Cayley–Hamilton theorem (see (2.3) below) by taking traces.

2.2 Two-generator groups

We begin by recording the first (trivial) normalization for computing traces:

$$\text{tr}(\text{Id}) = 2. \tag{2.2}$$

This will be the first of three properties of the trace function which enables the computation of traces of arbitrary words in elements of $\text{SL}(2, \mathbb{C})$.

The Cayley–Hamilton theorem. If ξ is a 2×2 -matrix, then

$$\xi^2 - \text{tr}(\xi)\xi + \det(\xi)\text{Id} = 0. \tag{2.3}$$

Suppose $\xi, \eta \in \text{SL}(2, \mathbb{C})$. Multiplying (2.3) by ξ^{-1} and rearranging yields

$$\xi + \xi^{-1} = \text{tr}(\xi)\text{Id} \tag{2.4}$$

from which follows (using (2.2)) that

$$\text{tr}(\xi) = \text{tr}(\xi^{-1}). \tag{2.5}$$

Multiplying (2.4) by η and taking traces, we obtain (switching ξ and η):

Theorem 2.2.1 (The Basic Identity). *Let $\xi, \eta \in \text{SL}(2, \mathbb{C})$. Then*

$$\text{tr}(\xi\eta) + \text{tr}(\xi\eta^{-1}) = \text{tr}(\xi)\text{tr}(\eta). \tag{2.6}$$

As we shall see, the three identities (2.2), (2.4) and (2.5) apply to compute the trace of any word $w(\xi, \eta)$ for $\xi, \eta \in \mathrm{SL}(2, \mathbb{C})$.

Traces of reduced words: an algorithm. Here is an important special case of Theorem A. Namely, let $w(X, Y) \in \pi$ be a reduced word. Then

$$\begin{aligned} \mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C}) &\longrightarrow \mathbb{C}, \\ (\xi, \eta) &\longmapsto \mathrm{tr}(w(\xi, \eta)) \end{aligned}$$

is an $\mathrm{SL}(2, \mathbb{C})$ -invariant function on $\mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C})$. Theorem A guarantees a polynomial

$$f_w(x, y, z) \in \mathbb{C}[x, y, z]$$

such that

$$\mathrm{tr}(w(\xi, \eta)) = f_w(\mathrm{tr}(\xi), \mathrm{tr}(\eta), \mathrm{tr}(\xi\eta)) \quad (2.7)$$

for all $\xi, \eta \in \mathrm{SL}(2, \mathbb{C})$. We describe an algorithm for computing $f_w(x, y, z)$. For notational convenience we write

$$\mathrm{tr}(w(\xi, \eta)) := f_{w(X, Y)}(x, y, z).$$

For example,

$$\begin{aligned} \mathrm{tr}(\mathrm{Id}) &= 2, \\ \mathrm{tr}(\xi^{-1}) &= \mathrm{tr}(\xi) = x, \\ \mathrm{tr}(\eta^{-1}) &= \mathrm{tr}(\eta) = y, \end{aligned}$$

verifying assertion (2.7) for words w of length $\ell(w) \leq 1$. For symmetry, we write $Z = Y^{-1}X^{-1}$, so that X, Y, Z satisfy the relation

$$XYZ = \mathrm{Id}.$$

(For a geometric interpretation of this presentation in terms of the three-holed sphere $\Sigma_{0,3}$, compare §3.2.) Write $\zeta = (\xi\eta)^{-1}$ so that $\xi\eta\zeta = \mathrm{Id}$. Then

$$\mathrm{tr}(\xi\eta) = \mathrm{tr}(\eta\xi) = \mathrm{tr}(\xi^{-1}\eta^{-1}) = \mathrm{tr}(\eta^{-1}\xi^{-1}) = \mathrm{tr}(\zeta) = \mathrm{tr}(\zeta^{-1}) = z.$$

The reduced words of length two are

$$\begin{aligned} X^2, \quad Y^2, \quad XY, \quad XY^{-1}, \quad YX, \quad YX^{-1}, \\ X^{-2}, \quad Y^{-2}, \quad X^{-1}Y^{-1}, \quad X^{-1}Y, \quad Y^{-1}X^{-1}, \quad Y^{-1}X. \end{aligned}$$

As mentioned above, the trace of a square (2.1) follows immediately by taking the trace of (2.3). Thus

$$\begin{aligned} \mathrm{tr}(\xi^2) &= x^2 - 2, \\ \mathrm{tr}(\eta^2) &= y^2 - 2, \\ \mathrm{tr}((\xi\eta)^2) &= z^2 - 2. \end{aligned}$$

Further applications of the trace identities imply

$$\begin{aligned}\mathrm{tr}(\xi\eta^{-1}) &= xy - z, \\ \mathrm{tr}(\eta(\xi\eta)) &= \mathrm{tr}(\eta\xi^{-1}) = yz - x, \\ \mathrm{tr}((\xi\eta)^{-1}\xi^{-1}) &= \mathrm{tr}(\zeta\xi^{-1}) = zx - y.\end{aligned}$$

For example, taking $w(X, Y) = XY^{-1}$,

$$\mathrm{tr}(\xi\eta^{-1}) = \mathrm{tr}(\xi)\mathrm{tr}(\eta) - \mathrm{tr}(\xi\eta).$$

Furthermore,

$$\begin{aligned}\mathrm{tr}(\xi\eta\xi^{-1}\eta) &= \mathrm{tr}(\xi\eta)\mathrm{tr}(\xi^{-1}\eta) - \mathrm{tr}(\xi^2) \\ &= z(xy - z) - (x^2 - 2) \\ &= 2 - x^2 - z^2 + xyz.\end{aligned}\tag{2.8}$$

An extremely important example is the *commutator word*

$$k(X, Y) := XYX^{-1}Y^{-1}.$$

Computation of its trace polynomial $\kappa = f_k$ follows easily from applying (2.6) to (2.8):

$$\begin{aligned}\mathrm{tr}(\xi\eta\xi^{-1}\eta^{-1}) &= \mathrm{tr}(\xi\eta\xi^{-1})\mathrm{tr}(\eta) - \mathrm{tr}(\xi\eta\xi^{-1}\eta) \\ &= y^2 - (2 - x^2 - z^2 + xyz) \\ &= x^2 + y^2 + z^2 - xyz - 2\end{aligned}$$

whence

$$\kappa(x, y, z) = f_k(x, y, z) = x^2 + y^2 + z^2 - xyz - 2.\tag{2.9}$$

Assume inductively that for all reduced words $w(X, Y) \in \pi$ with $\ell(w) < m$, there exists a polynomial $f_w(x, y, z) = \mathrm{tr}(w(\xi, \eta))$ satisfying (2.7). Suppose that $u(X, Y) \in \mathbb{F}_2$ is a reduced word of length $\ell(u) = m$.

The explicit calculations above begin the induction for $m \leq 2$. Thus we assume $m > 2$.

Furthermore, we can assume that u is *cyclically reduced*, that is the initial symbol of u is not inverse to the terminal symbol of u . For otherwise

$$u(X, Y) = Su'(X, Y)S^{-1},$$

where S is one of the four symbols

$$X, \quad Y, \quad X^{-1}, \quad Y^{-1}$$

and $\ell(u') = m - 2$. Then $u(X, Y)$ and $u'(X, Y)$ are conjugate and

$$\mathrm{tr}(u(X, Y)) = \mathrm{tr}(u'(X, Y)).$$

If $m > 2$, and u is cyclically reduced, then $u(X, Y)$ has a repeated letter, which we may assume to equal X . That is, we may write, after conjugating by a subword,

$$u(X, Y) = u_1(X, Y)u_2(X, Y),$$

where u_1 and u_2 are reduced words each ending in $X^{\pm 1}$. Furthermore we may assume that

$$\ell(u_1) + \ell(u_2) = \ell(u) = m,$$

so that $\ell(u_1) < m$ and $\ell(u_2) < m$. Suppose first that u_1 and u_2 both end in X . Then

$$u(X, Y) = (u_1(X, Y)X^{-1})X(u_2(X, Y)X^{-1})X$$

and each of

$$u_1(X, Y)X^{-1}, \quad u_2(X, Y)X^{-1}$$

has a terminal XX^{-1} , which we cancel to obtain the corresponding reduced words $u'_1(X, Y)$, $u'_2(X, Y)$ respectively with

$$\ell(u'_i), \quad \ell(u_i)$$

for $i = 1, 2$, and

$$u(X, Y) = u_1(X, Y)u_2(X, Y) = u'_1(X, Y)Xu'_2(X, Y)X$$

in \mathbb{F}_2 . Then

$$(u_1(X, Y)X^{-1})(u_2(X, Y)X^{-1})^{-1} = u'_1(X, Y)u'_2(X, Y)^{-1}$$

is represented by a reduced word $u_3(X, Y)$ satisfying $\ell(u_3) < m$. By the induction hypothesis, there exist polynomials

$$f_{u_1(X, Y)}, f_{u_2(X, Y)}, f_{u_3(X, Y)} \in \mathbb{C}[x, y, z]$$

such that, for all $\xi, \eta \in \mathrm{SL}(2, \mathbb{C})$, $i = 1, 2, 3$,

$$\mathrm{tr}(u_i(\xi, \eta)) = f_{u_i(X, Y)}(\mathrm{tr}(\xi), \mathrm{tr}(\eta), \mathrm{tr}(\xi\eta)).$$

By (2.6),

$$f_u = f_{u_1}f_{u_2} - f_{u_3}$$

is a polynomial in $\mathbb{C}[x, y, z]$. The cases when u_1 and u_2 both end in the symbols X^{-1} , Y , Y^{-1} are completely analogous. Since there are only four symbols, the only cyclically reduced words without repeated symbols are commutators of the symbols, for example $XYX^{-1}Y^{-1}$. Repeated applications of the trace identities evaluate this trace polynomial as $\kappa(x, y, z)$ defined in (2.9). The other commutators of distinct symbols also have trace $\kappa(x, y, z)$ by identical arguments.

Surjectivity of characters of pairs: a normal form. We first show that

$$\begin{aligned} \tau : \mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C}) &\longrightarrow \mathbb{C}^3, \\ (\xi, \eta) &\longmapsto \begin{bmatrix} \mathrm{tr}(\xi) \\ \mathrm{tr}(\eta) \\ \mathrm{tr}(\xi\eta) \end{bmatrix} \end{aligned}$$

is surjective. Let $(x, y, z) \in \mathbb{C}^3$. Choose $\mathfrak{z} \in \mathbb{C}$ so that

$$\mathfrak{z} + \mathfrak{z}^{-1} = z,$$

that is, $\mathfrak{z} = \frac{1}{2}(z \pm \sqrt{z^2 - 4})$. Let

$$\xi_x = \begin{bmatrix} x & -1 \\ 1 & 0 \end{bmatrix}, \quad \eta_{(y, \mathfrak{z})} = \begin{bmatrix} 0 & \mathfrak{z}^{-1} \\ -\mathfrak{z} & y \end{bmatrix}. \quad (2.10)$$

Then $\tau(\xi_x, \eta_{(y, \mathfrak{z})}) = (x, y, z)$.

Next we show that every $\mathrm{SL}(2, \mathbb{C})$ -invariant regular function

$$\mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C}) \xrightarrow{f} \mathbb{C}$$

factors through τ . To this end we need the following elementary lemma on symmetric functions:

Lemma 2.2.2. *Let R be an integral domain where 2 is invertible, and let $R' = R[\mathfrak{z}, \mathfrak{z}^{-1}]$ be the ring of Laurent polynomials over R . Let $R' \xrightarrow{\sigma} R'$ be the involution which fixes R and interchanges \mathfrak{z} and \mathfrak{z}^{-1} . Then the subring of σ -invariants is the polynomial ring $R[\mathfrak{z} + \mathfrak{z}^{-1}]$.*

Proof. Let $F(\mathfrak{z}, \mathfrak{z}^{-1}) \in R[\mathfrak{z}, \mathfrak{z}^{-1}]$ be a σ -invariant Laurent polynomial. Begin by rewriting R' as the quotient of the polynomial ring $R[x, y]$ by the ideal generated by $xy - 1$. Then σ is induced by the involution $\tilde{\sigma}$ of $R[x, y]$ interchanging x and y . Let $f(x, y) \in R[x, y]$ be a polynomial whose image in R' is F . Then there exists a polynomial $g(x, y)$ such that

$$f(x, y) - f(y, x) = g(x, y)(xy - 1).$$

Clearly $g(x, y) = -g(y, x)$. Let

$$\tilde{f}(x, y) = f(x, y) - \frac{1}{2} g(x, y)(xy - 1)$$

so that $\tilde{f}(x, y) = \tilde{f}(y, x)$. By the theorem on elementary symmetric functions,

$$\tilde{f}(x, y) = h(x + y, xy)$$

for some polynomial $h(u, v)$. Therefore $F(\mathfrak{z}, \mathfrak{z}^{-1}) = h(\mathfrak{z} + \mathfrak{z}^{-1}, 1)$ as desired. \square

By definition $f(\xi, \eta)$ is a polynomial in the matrix entries of ξ and η ; regard two polynomials differing by elements in the ideal generated by $\det(\xi) - 1$ and $\det(\eta) - 1$

as equal. Thus $f(\xi_x, \eta_{(y, \mathfrak{z})})$ equals a function $g(x, y, \mathfrak{z})$ which is a polynomial in $x, y \in \mathbb{C}$ and a Laurent polynomial in $\mathfrak{z} \in \mathbb{C}^*$ where ξ_x and $\eta_{(y, \mathfrak{z})}$ were defined in (2.10).

Lemma 2.2.3. *Let $\xi, \eta \in \mathrm{SL}(2, \mathbb{C})$ such that $\kappa(\tau(\xi, \eta)) \neq 2$. Then there exists $h \in \mathrm{SL}(2, \mathbb{C})$ such that*

$$h \cdot (\xi, \eta) = (\xi^{-1}, \eta^{-1}).$$

Proof. Let $(x, y, z) = \tau(\xi, \eta)$. By the commutator trace formula (2.9),

$$\mathrm{tr}[\xi, \eta] = \kappa(x, y, z),$$

where $[\xi, \eta] = \xi\eta\xi^{-1}\eta^{-1}$.

Let $L = \xi\eta - \eta\xi$. (Compare §4 of Jørgensen [43] or Fenchel [16].) Then

$$\mathrm{tr}(L) = \mathrm{tr}(\xi\eta) - \mathrm{tr}(\eta\xi) = 0.$$

Furthermore for any 2×2 matrix M , the characteristic polynomial

$$\lambda_M(t) := \det(t \mathrm{Id} - M) = t^2 - \mathrm{tr}(M)t + \det(M).$$

Thus

$$\begin{aligned} \det(L) &= \det([\xi, \eta] - \mathrm{Id})\det(\eta\xi) \\ &= \det([\xi, \eta] - \mathrm{Id}) \\ &= -\lambda_{[\xi, \eta]}(1) \\ &= -2 + \mathrm{tr}[\xi, \eta] \\ &= -2 + \kappa(x, y, z) \neq 0. \end{aligned}$$

Choose $\mu \in \mathbb{C}^*$ such that $\mu^2 \det(L) = 1$ and let $h = \mu L \in \mathrm{SL}(2, \mathbb{C})$.

Since $\mathrm{tr}(h) = 0$ and $\det(h) = 1$, the Cayley–Hamilton Theorem

$$\lambda_M(M) = 0$$

implies that $h^2 = -\mathrm{Id}$. Similarly

$$\begin{aligned} \det(h\xi) &= \det(h) = 1, \\ \det(h\eta) &= \det(h) = 1, \end{aligned}$$

and

$$\mathrm{tr}(h\xi) = \mu(\mathrm{tr}((\xi\eta)\xi) - \mathrm{tr}((\eta\xi)\xi)) = \mu(\mathrm{tr}(\xi(\eta\xi)) - \mathrm{tr}((\eta\xi)\xi)) = 0$$

and

$$\mathrm{tr}(h\eta) = \mu(\mathrm{tr}((\xi\eta)\eta) - \mathrm{tr}((\eta\xi)\eta)) = \mu(\mathrm{tr}((\xi\eta)\eta) - \mathrm{tr}(\eta(\xi\eta))) = 0,$$

so $(h\xi)^2 = (h\eta)^2 = -\mathrm{Id}$. Thus

$$h\xi h^{-1}\xi = -h\xi h\xi = \mathrm{Id}$$

whence $h\xi h^{-1} = \xi^{-1}$. Similarly $h\eta h^{-1} = \eta^{-1}$, concluding the proof of Lemma 2.2.3. \square

Apply Lemma 2.2.3 to $\xi = \xi_x$ and $\eta = \eta_{(y, \mathfrak{z})}$ as above to obtain h such that conjugation by h maps

$$\xi \mapsto \xi^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & x \end{bmatrix}$$

and

$$\eta \mapsto \eta^{-1} = \begin{bmatrix} y & -1/\mathfrak{z} \\ \mathfrak{z} & 0 \end{bmatrix}.$$

If

$$u = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

then

$$uh\xi(uh)^{-1} = u\xi^{-1}u^{-1} = \begin{bmatrix} x & -1 \\ 1 & 0 \end{bmatrix} = \xi$$

and

$$uh\eta(uh)^{-1} = u\eta^{-1}u^{-1} = \begin{bmatrix} 0 & \mathfrak{z} \\ -1/\mathfrak{z} & y \end{bmatrix} = \eta_{(y, \mathfrak{z}^{-1})}.$$

Thus

$$g(x, y, \mathfrak{z}) = f(\xi, \eta) = f(uh\xi(uh)^{-1}, uh\eta(uh)^{-1}) = g(x, y, \mathfrak{z}^{-1}).$$

Lemma 2.2.2 implies that

$$g(x, y, \mathfrak{z}) = F(x, y, \mathfrak{z} + 1/\mathfrak{z}) \tag{2.11}$$

for some polynomial $F(x, y, z) \in \mathbb{C}[x, y, z]$, whenever $\kappa(x, y, \mathfrak{z} + 1/\mathfrak{z}) \neq 2$. Since this condition defines a nonempty Zariski-dense open set, (2.11) holds on all of $\mathbb{C}^2 \times \mathbb{C}^*$ and

$$f(\xi, \eta) = F(\text{tr}(\xi), \text{tr}(\eta), \text{tr}(\xi\eta))$$

as claimed.

Injectivity of $\text{SL}(2, \mathbb{C})$ -characters of pairs. Finally we show that if $(\xi, \eta), (\xi', \eta') \in H$ satisfy

$$\begin{bmatrix} \text{tr}(\xi) \\ \text{tr}(\eta) \\ \text{tr}(\xi\eta) \end{bmatrix} = \begin{bmatrix} \text{tr}(\xi') \\ \text{tr}(\eta') \\ \text{tr}(\xi'\eta') \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \tag{2.12}$$

and $\kappa(x, y, z) \neq 2$, then (ξ, η) and (ξ', η') are $\text{SL}(2, \mathbb{C})$ -equivalent. By §2.2, the triple

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \text{tr}(\xi) \\ \text{tr}(\eta) \\ \text{tr}(\xi\eta) \end{bmatrix}$$

determines the character function

$$\begin{aligned}\pi &\longrightarrow \mathbb{C}, \\ w(X, Y) &\longmapsto \operatorname{tr}(w(\xi, \eta)) = f_w(x, y, z).\end{aligned}$$

Let ρ and ρ' denote the representations $\pi \rightarrow \operatorname{SL}(2, \mathbb{C})$ taking X, Y to ξ, η and ξ', η' respectively and let χ, χ' denote their respective characters. Then our hypothesis (2.12) implies that $\chi = \chi'$.

2.3 Injectivity of the character map: the general case

The conjugacy of representations (one of which is irreducible) having the same character follows from a general argument using the Burnside theorem. I am grateful to Hyman Bass [1] for explaining this to me.

Suppose ρ and ρ' are irreducible representations on \mathbb{C}^2 . Burnside's Theorem (see Lang [49], p. 445) implies the corresponding representations (also denoted ρ, ρ' respectively) of the group algebra $\mathbb{C}\pi$ into $M_2(\mathbb{C})$ are surjective. Since the trace form

$$\begin{aligned}M_2(\mathbb{C}) \times M_2(\mathbb{C}) &\longrightarrow \mathbb{C}, \\ (A, B) &\longmapsto \operatorname{tr}(AB)\end{aligned}$$

is nondegenerate, the kernel K of $\mathbb{C}\pi \xrightarrow{\rho} M_2(\mathbb{C})$ consists of all

$$\sum_{\alpha \in \pi} a_\alpha \alpha \in \mathbb{C}\pi$$

such that

$$0 = \operatorname{tr}\left(\left(\sum_{\alpha \in \pi} a_\alpha \rho(\alpha)\right)\rho(\beta)\right) = \sum_{\alpha \in \pi} a_\alpha \operatorname{tr}(\rho(\alpha\beta)) = \sum_{\alpha \in \pi} a_\alpha \chi(\alpha\beta)$$

for all $\beta \in \pi$. Thus the kernels of both representations of $\mathbb{C}\pi$ are equal, and ρ and ρ' respectively induce algebra isomorphisms

$$\mathbb{C}\pi/K \longrightarrow M_2(\mathbb{C}),$$

denoted $\tilde{\rho}, \tilde{\rho}'$.

The composition $\tilde{\rho}' \circ \tilde{\rho}^{-1}$ is an automorphism of the algebra $M_2(\mathbb{C})$, which must be induced by conjugation by $g \in \operatorname{GL}(2, \mathbb{C})$. (See, for example, Corollary 9.122, p. 734 of Rotman [69].) In particular $\rho'(\gamma) = g\rho(\gamma)g^{-1}$ as desired.

Irreducibility. The theory is significantly different for reducible representations. Representations

$$\rho_1, \rho_2 \in \operatorname{Hom}(\pi, \operatorname{SL}(2, \mathbb{C}))$$

are *equivalent* if and only if they define the same point in the character variety, that is, for all regular functions f in the character ring,

$$f(\rho_1) = f(\rho_2).$$

If both are irreducible, then ρ_1 and ρ_2 are conjugate. Closely related is the fact that the conjugacy class of an irreducible representation is closed. Here are several equivalent conditions for irreducibility of two-generator subgroups of $\mathrm{SL}(2, \mathbb{C})$:

Proposition 2.3.1. *Let $\xi, \eta \in \mathrm{SL}(2, \mathbb{C})$. The following are equivalent:*

- (1) ξ, η generate an irreducible representation on \mathbb{C}^2 .
- (2) $\mathrm{tr}(\xi\eta\xi^{-1}\eta^{-1}) \neq 2$.
- (3) $\det(\xi\eta - \eta\xi) \neq 0$.
- (4) The pair (ξ, η) is not $\mathrm{SL}(2, \mathbb{C})$ -conjugate to a representation by upper-triangular matrices

$$\begin{bmatrix} a & b \\ 0 & a^{-1} \end{bmatrix},$$

where $a \in \mathbb{C}^*, b \in \mathbb{C}$.

- (5) Either the group $\langle \xi, \eta \rangle$ is not solvable, or there exists a decomposition

$$\mathbb{C}^2 = L_1 \oplus L_2$$

into an invariant pair of lines L_i such that one of ξ, η interchanges L_1 and L_2 .

- (6) $\{\mathrm{Id}, \xi, \eta, \xi\eta\}$ is a basis for $M_2(\mathbb{C})$.

In the next section we will find a further condition (Theorem 3.2.2) involving extending the representation to a representation of the free product $\mathbb{Z}/2 * \mathbb{Z}/2 * \mathbb{Z}/2$.

Proof. The equivalence (1) \iff (2) is due to Culler–Shalen [13]. For completeness we give the proof here.

To prove (2) \implies (1), suppose that ρ is reducible. If ξ, η generate a representation with an invariant subspace of \mathbb{C}^2 of dimension one, this representation is conjugate to one in which ξ and η are upper-triangular. Denoting their diagonal entries by a, a^{-1} and b, b^{-1} respectively, the diagonal entries of $\xi\eta$ are $ab, a^{-1}b^{-1}$. Thus

$$\begin{aligned} x &= a + a^{-1}, \\ y &= b + b^{-1}, \\ z &= ab + a^{-1}b^{-1}. \end{aligned}$$

By direct computation, $\kappa(x, y, z) = 2$.

To prove (1) \implies (2), suppose that $\kappa(x, y, z) = 2$. Let $\mathfrak{A} \subset M_2(\mathbb{C})$ denote the linear span of $\mathrm{Id}, \xi, \eta, \xi\eta$. Identities derived from the Cayley–Hamilton theorem (2.3) such as (2.4) imply that \mathfrak{A} is a subalgebra of $M_2(\mathbb{C})$. For example, ξ^2 equals the linear combination

$$\xi^2 = -\mathrm{Id} + x\xi \tag{2.13}$$

and

$$\eta\xi = (z - xy)\text{Id} + y\xi + x\eta - \xi\eta. \quad (2.14)$$

The latter identity follows by writing

$$\xi^{-1}\eta + \eta^{-1}\xi = \text{tr}(\xi^{-1}\eta)\text{Id} = (xy - z)\text{Id}$$

and summing

$$\begin{aligned} x\eta &= \xi\eta + \xi^{-1}\eta, \\ y\xi &= \eta\xi + \eta^{-1}\xi \end{aligned}$$

to obtain

$$\xi\eta + \eta\xi = (z - xy)\text{Id} + x\eta + y\xi$$

as desired.

In the basis of $M_2(\mathbb{C})$ by elementary matrices, the linear map

$$\begin{aligned} \mathbb{C}^4 &\longrightarrow M_2(\mathbb{C}), \\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} &\longmapsto x_1\text{Id} + x_2\xi + x_3\eta + x_4\xi\eta \end{aligned}$$

has determinant $2 - \kappa(x, y, z) = 0$ and is not surjective. Thus \mathfrak{A} is a proper subalgebra of $M_2(\mathbb{C})$ and the representation is reducible, as desired.

(2) \iff (3) follows from the suggestive formula, valid for $\xi, \eta \in \text{SL}(2, \mathbb{C})$,

$$\text{tr}(\xi\eta\xi^{-1}\eta^{-1}) + \det(\xi\eta - \eta\xi) = 2, \quad (2.15)$$

whose proof is left as an exercise.

The equivalence (1) \iff (4) is essentially the definition of reducibility. If $L \subset \mathbb{C}^2$ is an invariant subspace, then conjugating by a linear automorphism which maps L to the first coordinate line $\mathbb{C} \times \{0\}$ makes the representation upper triangular.

(4) \iff (5) follows from the classification of solvable subgroups of $\text{SL}(2, \mathbb{C})$: a solvable subgroup is either conjugate to a group of upper-triangular matrices, or is conjugate to a *dihedral representation*, where one of ξ, η is a diagonal matrix and the other is the *involution*

$$i \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(where the coefficient i is required for unimodularity). A dihedral representation is one which interchanges an invariant pair of lines although the lines themselves are not invariant. For a description of these representations in terms of hyperbolic geometry, see §3.2.

(1) \iff (6) follows from the Burnside lemma, and identities such as (2.13) and (2.14) to express products of $\text{Id}, \xi, \eta, \xi\eta$ with the generators ξ, η as linear combinations of $\text{Id}, \xi, \eta, \xi\eta$. \square

3 Coxeter triangle groups in hyperbolic 3-space

An alternate geometric approach to the algebraic parametrization using traces involves right-angled hexagons in \mathbb{H}^3 . Specifically, a marked two-generator group corresponds to an ordered triple of lines in \mathbb{H}^2 , no two of which are asymptotic. This triple completes to a right-angled hexagon by including the three common orthogonal lines. We use this geometric construction to identify, in terms of traces, which representations correspond to geometric structures on surfaces. However, since the trace is only defined on $\mathrm{SL}(2, \mathbb{C})$, and not on $\mathrm{PSL}(2, \mathbb{C})$, we must first discuss the conditions which ensure that a representation into $\mathrm{PSL}(2, \mathbb{C})$ lifts to $\mathrm{SL}(2, \mathbb{C})$.

3.1 Lifting representations to $\mathrm{SL}(2, \mathbb{C})$

The group of orientation-preserving isometries of \mathbb{H}^3 identifies with $\mathrm{PSL}(2, \mathbb{C})$, which is doubly covered by $\mathrm{SL}(2, \mathbb{C})$. In general, a representation $\Gamma \rightarrow \mathrm{PSL}(2, \mathbb{C})$ may or may not lift to a representation to $\mathrm{SL}(2, \mathbb{C})$. Clearly if Γ is a free group, every representation lifts, since lifting each generator suffices to define a lifted representation. In general the obstruction to lifting a representation $\Gamma \rightarrow \mathrm{PSL}(2, \mathbb{C})$ is a cohomology class $\mathfrak{o} \in H^2(\Gamma, \mathbb{Z}/2)$. Furthermore there exists a central $\mathbb{Z}/2$ -extension $\hat{\Gamma} \rightarrow \Gamma$ (corresponding to \mathfrak{o}) and a lifted representation $\hat{\Gamma}$ such that

$$\begin{array}{ccc} \hat{\Gamma} & \longrightarrow & \mathrm{SL}(2, \mathbb{C}) \\ \downarrow & & \downarrow \\ \Gamma & \longrightarrow & \mathrm{PSL}(2, \mathbb{C}) \end{array}$$

commutes. This lift is *not unique*; the various lifts differ by multiplication by homomorphisms

$$\Gamma \longrightarrow \{\pm \mathrm{Id}\} = \text{center}(\mathrm{SL}(2, \mathbb{C}))$$

which comprise the group

$$\mathrm{Hom}(\pi_1(\Sigma), \{\pm \mathrm{Id}\}) \cong H^1(\Sigma; \mathbb{Z}/2).$$

The cohomology class in $H^2(\Gamma, \mathbb{Z}/2)$ may be understood in terms of *Hopf's formula* for the second homology of a group. (See, for example, Brown [6].) Consider a presentation $\Gamma = F/R$ where F is a finitely generated free group and $R \triangleleft F$ is a normal subgroup. A set $\{f_1, \dots, f_N\}$ of free generators for F corresponds to the generators of Γ and R corresponds to the *relations* among these generators. Then Hopf's formula identifies $H_2(\Gamma)$ with the quotient group

$$([F, F] \cap R)/[F, R],$$

where $[F, F] \triangleleft F$ is the commutator subgroup and $[F, R]$ is the (normal) subgroup of F generated by commutators $[f, r]$ where $f \in F$ and $r \in R$. Intuitively, $H_2(\Gamma)$ is generated by relations which are products of simple commutators $[a_1, b_1] \dots [a_g, b_g]$,

where $a_i, b_i \in F$ are words in f_1, \dots, f_N . Such *commutator relations* correspond to maps of a closed orientable surface Σ_g into the classifying space $B\Gamma$ of Γ . If $\Gamma \xrightarrow{\rho} G$ is a homomorphism into G and $\tilde{G} \rightarrow G$ is a central extension (such as a covering group of a Lie group), then the obstruction is calculated for each commutator relation

$$w = [a_1, b_1] \dots [a_g, b_g] \in [F, F] \cap R$$

corresponding to a 2-cycle z , as follows. (Here each $a_i, b_i \in F$ is a word in the free generators f_1, \dots, f_N .) Lift each generator $\rho(f_i)$ to $\tilde{\rho}(f_i) \in \tilde{G}$ and evaluate the word $w(f_1, \dots, f_N)$ on the lifts $\tilde{\rho}(f_i)$ to obtain an element in the kernel K of $\tilde{G} \rightarrow G$ (since $w \in R$). Furthermore since $w \in [F, F]$ and two lifts differ by an element of $K \subset \text{center}(\tilde{G})$, this element is independent of the chosen lift $\tilde{\rho}$. This procedure defines an element of

$$H^2(\Gamma, K) \cong \text{Hom}\left(\frac{[F, F] \cap R}{[F, R]}, K\right)$$

which evidently vanishes if and only if ρ lifts. (Compare Milnor [59]. For more discussion of lifting homomorphisms to $\text{SL}(2, \mathbb{C})$, compare Culler [12], Kra [48], Goldman [27] or Patterson [65]. According to Patterson [65], the first result of this type, due to H. Petersson [66], is that a Fuchsian subgroup of $\text{PSL}(2, \mathbb{R})$ lifts to $\text{SL}(2, \mathbb{R})$ if and only if it has no elements of order two.)

A representation $\Gamma \rightarrow \text{PSL}(2, \mathbb{C})$ is *irreducible* if one (and hence every) lift $\hat{\Gamma} \rightarrow \text{SL}(2, \mathbb{C})$ is irreducible.

3.2 The three-holed sphere

The basic building block for hyperbolic surfaces is the three-holed sphere $\Sigma_{0,3}$.

Geometric version of Theorem A. Theorem A has a suggestive interpretation in terms of the three-holed sphere $\Sigma_{0,3}$, or “pair-of-pants.” Namely, the fundamental group

$$\pi_1(\Sigma_{0,3}) \cong \mathbb{F}_2$$

admits the *redundant geometric presentation*

$$\pi = \pi_1(\Sigma_{0,3}) = \langle X, Y, Z \mid XYZ = 1 \rangle,$$

where X, Y, Z correspond to the three components of $\partial\Sigma_{0,3}$. Denoting the corresponding trace functions by lower case, for example

$$\begin{aligned} \text{Hom}(\pi, G) &\xrightarrow{x} \mathbb{C}, \\ \rho &\longmapsto \text{tr}(\rho(X)), \end{aligned}$$

Theorem A asserts that the $\text{SL}(2, \mathbb{C})$ -character ring of π is the polynomial ring $\mathbb{C}[x, y, z]$.

Theorem 3.2.1. *The equivalence class of a flat $SL(2, \mathbb{C})$ -bundle over $\Sigma_{0,3}$ with irreducible holonomy is determined by the equivalence classes of its restrictions to the three components of $\partial\Sigma_{0,3}$. Furthermore any triple of isomorphism classes of flat $SL(2, \mathbb{C})$ -bundles over $\partial\Sigma_{0,3}$ whose holonomy traces satisfy*

$$x^2 + y^2 + z^2 - xyz \neq 4$$

extends to a flat $SL(2, \mathbb{C})$ -bundle over $\Sigma_{0,3}$.

The hexagon orbifold. Every irreducible representation ρ corresponds to a geometric object in H^3 , a *triple of geodesics*. Any two of these geodesics admits a unique common perpendicular geodesic. These perpendiculars cut off a hexagon bounded by geodesic segments, with all six angles right angles. Such a *right hexagon* in H^3 is an alternate geometric object corresponding to ρ .

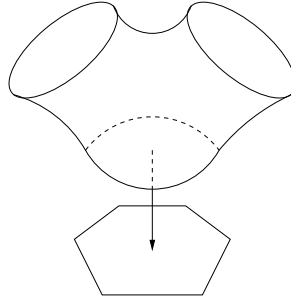


Figure 1. The three-holed sphere double covers a hexagon orbifold.

The surface $\Sigma_{0,3}$ admits an *orientation-reversing involution*

$$\Sigma_{0,3} \xrightarrow{\iota_{\text{Hex}}} \Sigma_{0,3}$$

whose restriction to each boundary component is a reflection. The quotient Hex by this involution is a disc, combinatorially equivalent to a hexagon. The three boundary components map to three intervals $\partial_i(\text{Hex})$, for $i = 1, 2, 3$, in the boundary ∂Hex . The other three edges in ∂Hex correspond to the three arcs comprising the fixed point set $\text{Fix}(\iota_{\text{Hex}})$. The orbifold structure on Hex is defined by mirrors on these three arcs on ∂Hex . The quotient map

$$\Sigma_{0,3} \xrightarrow{\Pi_{\text{Hex}}} \text{Hex}$$

is an orbifold covering-space, representing $\Sigma_{0,3}$ as the orientable double covering of the orbifold Hex . The orbifold fundamental group is

$$\begin{aligned} \hat{\pi} &:= \pi_1(\text{Hex}) = \langle \iota_{YZ}, \iota_{ZX}, \iota_{XY} \mid \iota_{YZ}^2 = \iota_{ZX}^2 = \iota_{XY}^2 = 1 \rangle \\ &\cong \mathbb{Z}/2 * \mathbb{Z}/2 * \mathbb{Z}/2. \end{aligned}$$

The covering-space $\Sigma_{0,3} \xrightarrow{\Pi_{\text{Hex}}} \text{Hex}$ induces the embedding of fundamental groups:

$$\begin{aligned} \pi_1(\Sigma_{0,3}) &\xrightarrow{(\Pi_{\text{Hex}})_*} \pi_1(\text{Hex}); \\ X &\longmapsto \iota_{ZX}\iota_{XY}, \\ Y &\longmapsto \iota_{XY}\iota_{YZ}, \\ Z &\longmapsto \iota_{YZ}\iota_{ZX}. \end{aligned}$$

Theorem 3.2.2. *Let $\pi \xrightarrow{\rho} \text{PGL}(2, \mathbb{C})$ be an irreducible representation. Then there exists a unique representation $\hat{\pi} \xrightarrow{\hat{\rho}} \text{PGL}(2, \mathbb{C})$ such that $\rho = \hat{\rho} \circ (\Pi_{\text{Hex}})_*$.*

Every element of order two in $\text{PGL}(2, \mathbb{C})$ is reflection about some geodesic. Therefore a representation $\hat{\rho}$ corresponds exactly to an ordered triple of geodesics in \mathbb{H}^3 . Denote this ordered triple of geodesics in \mathbb{H}^3 corresponding to ρ by ι^ρ .

Corollary 3.2.3. *Irreducible representations $\pi \xrightarrow{\rho} \text{PGL}(2, \mathbb{C})$ correspond to triples ι^ρ of geodesics in \mathbb{H}^3 , which share neither a common endpoint nor a common orthogonal geodesic.*

The proofs of Theorem 3.2.2 and Corollary 3.2.3 occupy the remainder of this section.

Involutions in $\text{PGL}(2, \mathbb{C})$. We are particularly interested in projective transformations of \mathbb{CP}^1 of order two, which we call *involutions*. Such an involution is given by a matrix $\xi \in \text{GL}(2, \mathbb{C})$ such that ξ^2 does act identically on \mathbb{CP}^1 but ξ does *not* act identically on \mathbb{CP}^1 . Thus ξ is a matrix whose square is a scalar matrix but ξ itself is not scalar. Since $\det(\xi) \neq 0$, replacing ξ by

$$\det(\xi)^{-1/2}\xi$$

– for either choice of $\det(\xi)^{-1/2}$ – ensures that $\det(\xi) = 1$. Then the scalar matrix $\xi^2 = \pm \text{Id}$. If $\xi^2 = \text{Id}$, then $\det(\xi) = 1$ implies $\xi = -\text{Id}$, a contradiction. Hence $\xi^2 = -\text{Id}$, and ξ must have distinct reciprocal eigenvalues $\pm i$. Thus ξ is conjugate to

$$\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

The corresponding projective transformation $\mathbb{P}(\xi)$ has two fixed points. The orbit of any point not in $\text{Fix}(\mathbb{P}(\xi))$ has cardinality two.

Proposition 3.2.4. *Let $\xi \in M_2(\mathbb{C})$. The following conditions are equivalent:*

- $\mathbb{P}(\xi) \in \text{Inv}$;
- ξ is conjugate to $\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$;
- $\det(\xi) = 1$ and $\text{tr}(\xi) = 0$;

- $\xi^2 = -\text{Id}$ and $\xi \neq \pm i \text{Id}$;
- $\xi^2 = -\text{Id}$ and ξ is not a scalar matrix.

The proof is left as an exercise. Denote the collection of such matrices by

$$\begin{aligned} \widetilde{\text{Inv}} &:= \text{SL}(2, \mathbb{C}) \cap \mathfrak{sl}(2, \mathbb{C}) \\ &= \{\xi \in \text{M}_2(\mathbb{C}) \mid \det(\xi) = 1, \text{tr}(\xi) = 0\}. \end{aligned}$$

Notice that $\widetilde{\text{Inv}}$ is invariant under $\pm \text{Id}$, and the quotient

$$\text{Inv} := \widetilde{\text{Inv}} / \{\pm \text{Id}\} \subset \text{PGL}(2, \mathbb{C})$$

consists of all projective involutions of \mathbb{CP}^1 . It naturally identifies with the collection of unordered pairs of distinct points in \mathbb{CP}^1 , that is, the quotient

$$(\mathbb{CP}^1 \times \mathbb{CP}^1 \setminus \Delta_{\mathbb{CP}^1}) / \mathfrak{S}_2$$

of the complement in $\mathbb{CP}^1 \times \mathbb{CP}^1$ of the diagonal

$$\Delta_{\mathbb{CP}^1} \subset \mathbb{CP}^1 \times \mathbb{CP}^1$$

by the symmetric group \mathfrak{S}_2 . In §3.2, we interpret $\widetilde{\text{Inv}}$ as the space of *oriented* geodesics in hyperbolic 3-space \mathbb{H}^3 .

Involutions and the complex projective line. Denote by $\overline{\text{Inv}}$ the closure of Inv in the projective space $\mathbb{P}(\mathfrak{sl}(2, \mathbb{C}))$. The complement $\overline{\text{Inv}} \setminus \text{Inv}$ corresponds to \mathbb{CP}^1 embedded as the diagonal $\Delta_{\mathbb{CP}^1}$ in the above description. For example the elements of $\overline{\text{Inv}}$ corresponding to $0, \infty \in \mathbb{CP}^1$ are the respective lines

$$\begin{bmatrix} 0 & * \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ * & 0 \end{bmatrix} \subset \text{M}_2(\mathbb{C}).$$

The closure corresponds to the full quotient space

$$(\mathbb{CP}^1 \times \mathbb{CP}^1) / \mathfrak{S}_2.$$

An element $\xi \in \text{PGL}(2, \mathbb{C}) \setminus \{\text{Id}\}$ stabilizes a unique element $\iota_\xi \in \overline{\text{Inv}}$. If ξ is semisimple ($\#\text{Fix}(\xi) = 2$), then ι_ξ is the unique involution with the same fixed points. Otherwise ξ is parabolic ($\#\text{Fix}(\xi) = 1$), and ι_ξ corresponds to the line

$$\text{Fix}(\text{Ad}(\xi)) = \text{Ker}(\text{Id} - \text{Ad}(\xi)) \subset \mathfrak{sl}(2, \mathbb{C}),$$

the Lie algebra centralizer of ξ in $\mathfrak{sl}(2, \mathbb{C})$. Further discussion of semisimple elements in $\text{SL}(2, \mathbb{C})$ and $\text{PSL}(2, \mathbb{C})$ is given in §3.2.

Here is an elegant matrix representation. If $\xi \in \text{SL}(2, \mathbb{C})$ is semisimple and $\neq \text{Id}$, then the two lifts of $\iota_\xi \in \text{Inv}$ to $\widetilde{\text{Inv}} \subset \text{SL}(2, \mathbb{C})$ differ by $\pm \text{Id}$. Since ξ is semisimple, its *traceless projection*

$$\xi' := \xi - \frac{1}{2} \text{tr}(\xi) \text{Id}$$

satisfies

- $\text{tr}(\xi') = 0$;
- ξ' commutes with ξ ;
- $\det(\xi') \neq 0$ (semisimplicity).

Choose $\delta \in \mathbb{C}^*$ such that

$$\delta^2 = \det(\xi') = \frac{4 - \text{tr}(\xi)^2}{4}.$$

Then $\delta^{-1}\xi' \in \widetilde{\text{Inv}}$ and represents the involution ι_ξ centralizing ξ :

$$\widetilde{\iota}_\xi = \pm \frac{2}{\sqrt{4 - \text{tr}(\xi)^2}} \left(\xi - \frac{\text{tr}(\xi)}{2} \text{Id} \right). \quad (3.1)$$

This formula will be used later in (4.3).

3-dimensional hyperbolic geometry. The group $\text{GL}(2, \mathbb{C})$ acts by orientation-preserving isometries on *hyperbolic 3-space* \mathbb{H}^3 . The kernel of the action equals the center of $\text{GL}(2, \mathbb{C})$, the group \mathbb{C}^* of nonzero scalar matrices. The quotient

$$\text{PGL}(2, \mathbb{C}) := \text{GL}(2, \mathbb{C})/\mathbb{C}^*$$

acts effectively on \mathbb{H}^3 . The restriction of the quotient homomorphism

$$\text{GL}(2, \mathbb{C}) \longrightarrow \text{PGL}(2, \mathbb{C})$$

to $\text{SL}(2, \mathbb{C}) \subset \text{GL}(2, \mathbb{C})$ defines an isomorphism

$$\text{PSL}(2, \mathbb{C}) \xrightarrow{\cong} \text{PGL}(2, \mathbb{C}).$$

The projective line \mathbb{CP}^1 identifies naturally with the ideal boundary $\partial\mathbb{H}^3$. The center of $\text{SL}(2, \mathbb{C})$ consists of $\pm\text{Id}$, which is the kernel of the actions on \mathbb{H}^3 and \mathbb{CP}^1 . The only element of order two in $\text{GL}(2, \mathbb{C})$ is $-\text{Id}$, and an element of even order $2k$ in $\text{PGL}(2, \mathbb{C})$ corresponds to an element of order $4k$ in $\text{GL}(2, \mathbb{C})$. Elements of odd order $2k + 1$ in $\text{PGL}(2, \mathbb{C})$ have two lifts to $\text{SL}(2, \mathbb{C})$, one of order $2k + 1$ and the other of order $2(2k + 1)$.

We use the *upper-half-space model* of \mathbb{H}^3 as follows. The algebra \mathbb{H} of *Hamilton quaternions* is the \mathbb{R} -algebra generated by $1, i, j$ subject to the relations

$$i^2 = j^2 = -1, \quad ij + ji = 0.$$

\mathbb{H} contains the smaller subalgebra \mathbb{C} having basis $\{1, i\}$. Define

$$\mathbb{H}^3 := \{z + uj \in \mathbb{H} \mid z \in \mathbb{C}, u \in \mathbb{R}, u > 0\},$$

where

$$\xi = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \text{GL}(2, \mathbb{C})$$

acts by

$$z + uj \longmapsto (a(z + uj) + b)(c(z + uj) + d)^{-1}.$$

$\mathrm{PSL}(2, \mathbb{C})$ is the group of orientation-preserving *isometries* of \mathbb{H}^3 with respect to the *Poincaré metric*

$$u^{-2}(|dz|^2 + du^2)$$

of constant curvature -1 . The restriction of $\mathrm{GL}(2, \mathbb{C})$ to $\partial\mathbb{H}^3$ identifies with the usual projective action of $\mathrm{PGL}(2, \mathbb{C})$ on

$$\partial\mathbb{H}^3 := \mathbb{CP}^1 = \mathbb{C} \cup \{\infty\},$$

the space of complex lines (1-dimensional linear subspaces) in \mathbb{C}^2 .

Oriented geodesics in \mathbb{H}^3 correspond to ordered pairs of distinct points in \mathbb{CP}^1 , via their endpoints. Unoriented geodesics correspond to unordered pairs. For example geodesics with an endpoint at ∞ are represented by vertical rays $z + \mathbb{R}_+j$, where $z \in \mathbb{C}$ is the other endpoint. The unit-speed parametrization of this geodesic is

$$\begin{aligned} \mathbb{R} &\longrightarrow \mathbb{H}^3, \\ t &\longmapsto z + e^t j. \end{aligned}$$

Distinct $z_1, z_2 \in \mathbb{C}$ span a geodesic in \mathbb{H}^3 whose unit-speed parametrization is

$$\begin{aligned} \mathbb{R} &\longrightarrow \mathbb{H}^3, \\ t &\longmapsto \frac{z_1 + z_2}{2} + \frac{z_2 - z_1}{2} (\tanh(t) + \operatorname{sech}(t)j). \end{aligned}$$

A geodesic $l \subset \mathbb{H}^3$ corresponds uniquely to the involution $\iota = \iota_l \in \mathrm{PSL}(2, \mathbb{C})$ for which

$$l = \operatorname{Fix}(\iota).$$

For example, if $z_1, z_2 \in \mathbb{C}$, the involution in $\mathrm{PGL}(2, \mathbb{C})$ fixing z_1, z_2 is given by the pair of matrices

$$\pm \frac{i}{z_1 - z_2} \begin{bmatrix} z_1 + z_2 & -2z_1 z_2 \\ 2 & -(z_1 + z_2) \end{bmatrix} \in \mathrm{SL}(2, \mathbb{C}). \quad (3.2)$$

If $z_2 = \infty$, the corresponding matrices are

$$\pm i \begin{bmatrix} -1 & 2z_1 \\ 0 & 1 \end{bmatrix} \in \mathrm{SL}(2, \mathbb{C}). \quad (3.3)$$

Compare Fenchel [16] for more details.

Let $\xi \in \mathrm{SL}(2, \mathbb{C})$ be *non-central*: $\xi \neq \pm \operatorname{Id}$. Then the following conditions are equivalent:

- ξ has two distinct eigenvalues;
- $\operatorname{tr}(\xi) \neq \pm 2$;
- the corresponding collineation of \mathbb{CP}^1 has two fixed points;
- the corresponding orientation-preserving isometry of \mathbb{H}^3 leaves invariant a unique geodesic ℓ_ξ , each of whose endpoints is fixed;
- a unique involution ι_ξ centralizes ξ .

(In the standard terminology, the corresponding isometry of \mathbb{H}^3 is either *elliptic* or *loxodromic*.)

We shall say that ξ is *semisimple*. Otherwise ξ is *parabolic*: it has a repeated eigenvalue (necessarily ± 1 , because $\det(\xi) = 1$), and fixes a *unique* point on $\mathbb{C}\mathbb{P}^1$.

Suppose $\xi \in \text{SL}(2, \mathbb{C})$ and the corresponding isometry $\mathbb{P}(\xi) \in \text{PSL}(2, \mathbb{C})$ leaves invariant a geodesic $l \subset \mathbb{H}^3$. Then the restriction $\mathbb{P}(\xi)|_l$ is an isometry of $l \approx \mathbb{R}$. Any isometry of \mathbb{R} is either a translation of \mathbb{R} , a reflection in a point of \mathbb{R} , or the identity. We distinguish these three cases as follows. For concreteness choose coordinates so that l is represented by the imaginary axis $\mathbb{R}+j \subset \mathbb{H}^3$ in the upper-half-space model. The endpoints of l are $0, \infty$:

- $\mathbb{P}(\xi)|_l$ acts by translation. Then $\mathbb{P}(\xi)$ is *loxodromic*, represented by

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}, \tag{3.4}$$

where $\lambda \in \mathbb{C}^*$ is a nonzero complex number and $|\lambda| \neq 1$. The fixed point set is

$$\text{Fix}(\mathbb{P}(\xi)) = \{0, \infty\}.$$

The restriction of $\mathbb{P}(\xi)$ to l is translation along l by distance $2 \log |\lambda|$, in the direction from 0 (its repellor) to ∞ (its attractor) if $|\lambda| > 1$. (If $|\lambda| < 1$, then 0 is the attractor and ∞ is the repellor.)

- $\mathbb{P}(\xi)|_l$ acts identically. Now $\mathbb{P}(\xi)$ is *elliptic* and is represented by the diagonal matrix (3.4), except now $|\lambda| = 1$. If $\lambda = e^{i\theta}$, then $\mathbb{P}(\xi)$ represents a rotation through angle 2θ about l . In particular if $\lambda = \pm i$, then $\mathbb{P}(\xi)$ is the involution fixing l . Although $\mathbb{P}(\xi)$ has order two in $\text{PGL}(2, \mathbb{C})$, its matrix representatives in $\text{SL}(2, \mathbb{C})$ each have order 4. (Compare Proposition 3.2.4.)
- $\mathbb{P}(\xi)|_l$ acts by reflection. In this case $\mathbb{P}(\xi)$ interchanges the two endpoints $0, \infty$ and is necessarily of order two. Its restriction $\mathbb{P}(\xi)|_l$ to l fixes the point $p = \text{Fix}(\mathbb{P}(\xi)) \cap l$, and is reflection in p . The corresponding matrix is

$$\begin{bmatrix} 0 & -\lambda \\ \lambda^{-1} & 0 \end{bmatrix},$$

where $\lambda \in \mathbb{C}^*$ and $p = |\lambda|j$ is the fixed point of $\mathbb{P}(\xi)|_l$. Necessarily $\mathbb{P}(\xi) \in \text{Inv}$ and

$$\text{Fix}(\mathbb{P}(\xi)) = \{\pm i\lambda\}.$$

Dihedral representations. The following lemma is crucial in the proof of Theorem 3.2.2.

Lemma 3.2.5. *Suppose that $\xi \in \text{SL}(2, \mathbb{C}) \setminus \{\pm \text{Id}\}$ and $\iota \in \text{Inv}$.*

- (1) *Suppose that $\#\text{Fix}(\mathbb{P}(\xi)) = 2$. Let $\ell_\xi \subset \mathbb{H}^3$ denote the unique ξ -invariant geodesic (the geodesic with endpoints $\text{Fix}(\mathbb{P}(\xi))$). Then*

$$\iota\xi\iota = \xi^{-1} \tag{3.5}$$

if and only if ι preserves ℓ_ξ and its restriction acts by reflection. In that case ι interchanges the two elements of $\text{Fix}(\mathbb{P}(\xi))$.

(2) Suppose that $\#\text{Fix}(\mathbb{P}(\xi)) = 1$. Then $\iota\xi\iota = \xi^{-1}$ if and only if $\text{Fix}(\mathbb{P}(\xi)) \subset \text{Fix}(\iota)$.

Proof. Consider first the case that ξ is semisimple, that is, when $\#\text{Fix}(\mathbb{P}(\xi)) = 2$. Let $\ell_\xi \subset \mathbb{H}^3$ be the ξ -invariant geodesic with endpoints $\text{Fix}(\mathbb{P}(\xi))$. Then ι interchanges the two elements of $\text{Fix}(\mathbb{P}(\xi))$. In terms of the linear representation, ξ preserves a decomposition into eigenspaces

$$\mathbb{C}^2 = L_1 \oplus L_2,$$

where each line $L_i \subset \mathbb{C}^2$ corresponds to a fixed point in $\mathbb{C}\mathbb{P}^1$, and ι interchanges L_1 and L_2 .

When $\#\text{Fix}(\mathbb{P}(\xi)) = 1$, the corresponding matrix has a unique eigenspace, which we take to be the first coordinate line. Then ξ is represented by the upper-triangular matrix

$$\pm \begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix},$$

and ι is also represented by an upper-triangular matrix of the form

$$\pm \begin{bmatrix} i & w \\ 0 & -i \end{bmatrix}.$$

Rewrite (3.5) as

$$(\iota\xi)^2 = \text{Id},$$

so that $\iota\xi \in \text{Inv}$. Thus ξ factors as the product of two involutions

$$\xi = \iota(\iota\xi).$$

Conversely if $\iota, \iota' \in \text{Inv}$, then the product $\xi := \iota'$ satisfies (3.5). □

Geometric interpretation of the Lie product. These ideas provide an elegant formula for the common orthogonal of the invariant axes of elements of $\text{PSL}(2, \mathbb{C})$. Suppose $\xi, \eta \in \text{SL}(2, \mathbb{C})$. Then the *Lie product*

$$\text{Lie}(\xi, \eta) := \xi\eta - \eta\xi \tag{3.6}$$

has trace zero, and vanishes if and only if ξ, η commute. Furthermore (2.15) and Proposition 2.3.1, (1) \iff (2) imply that $\text{Lie}(\xi, \eta)$ is invertible if and only if $\langle \xi, \eta \rangle$ acts irreducibly (as defined in §3). Suppose $\langle \xi, \eta \rangle$ acts irreducibly, so that $\text{Lie}(\xi, \eta)$ defines an element $\lambda \in \text{PSL}(2, \mathbb{C})$. Since $\text{tr}(\text{Lie}(\xi, \eta)) = 0$, the isometry λ has order two, that is, lies in Inv .

Now

$$\begin{aligned}\mathrm{tr}(\xi \mathrm{Lie}(\xi, \eta)) &= \mathrm{tr}(\xi(\xi\eta)) - \mathrm{tr}(\xi(\eta\xi)) \\ &= \mathrm{tr}(\xi(\xi\eta)) - \mathrm{tr}((\xi\eta)\xi) \\ &= 0\end{aligned}$$

which implies that $\xi\lambda$ also has order two, that is, $\lambda\xi\lambda = \xi^{-1}$. Lemma 3.2.5 implies that λ acts by reflection on the invariant axis ℓ_ξ . Similarly λ acts by reflection on the invariant axis ℓ_η . Hence the fixed axis ℓ_λ is orthogonal to both ℓ_ξ and ℓ_η :

Proposition 3.2.6. *If $\xi, \eta \in \mathrm{GL}(2, \mathbb{C})$, then the Lie product $\mathrm{Lie}(\xi, \eta)$ represents the common orthogonal geodesic $\perp (\ell_{\mathbb{P}(\xi)}, \ell_{\mathbb{P}(\eta)})$ to the invariant axes $\ell_{\mathbb{P}(\xi)}, \ell_{\mathbb{P}(\eta)}$ of $\mathbb{P}(\xi)$ and $\mathbb{P}(\eta)$ respectively.*

Compare Marden [55] and the references given there.

Geometric proof of Theorem 3.2.2

Proof of Theorem 3.2.2. Abusing notation, write X, Y, Z for

$$\rho(X), \rho(Y), \rho(Z) \in \mathrm{PGL}(2, \mathbb{C})$$

respectively. We seek respective involutions

$$\rho(\iota_{XY}), \quad \rho(\iota_{YZ}), \quad \rho(\iota_{ZX}),$$

which we respectively denote $\rho_{XY}, \rho_{YZ}, \rho_{ZX}$. These involutions will be the ones fixing the respective pairs. For example we take ρ_{XY} to be the involution fixing $\perp (l_X, l_Y)$, and similarly for ρ_{YZ} and ρ_{ZX} .

Suppose that $\langle X, Y \rangle \subset \mathrm{SL}(2, \mathbb{C})$ acts irreducibly on \mathbb{C}^2 . Write $Z = Y^{-1}X^{-1}$ so that

$$XYZ = \mathrm{Id}.$$

Since $\langle X, Y \rangle$ acts irreducibly, none of X, Y, Z act identically.

Let $\rho_{XY} \in \widetilde{\mathrm{Inv}}$ be the unique involution such that

$$\begin{aligned}\rho_{XY}X\rho_{XY} &= X^{-1}, \\ \rho_{XY}Y\rho_{XY} &= Y^{-1}\end{aligned}\tag{3.7}$$

respectively. (By Proposition 3.2.6, it is represented by the Lie product $\mathrm{Lie}(X, Y)$.) The involution ρ_{XY} will be specified by its fixed line $l_{XY} = \mathrm{Fix}(\rho_{XY})$, which is defined as follows: If both X, Y are semisimple, then Lemma 3.2.5 implies ρ_{XY} is the involution fixing the unique common orthogonal geodesic to the invariant axes of X or Y . If both X, Y are parabolic, then ρ_{XY} is the involution in the geodesic bounded by the fixed points of X, Y . Finally consider the case when one element is semisimple and the other element is parabolic. Then l_{XY} is the unique geodesic, for which one endpoint is the fixed point of the parabolic element, and which is orthogonal to the invariant axis of the semisimple element.

Similarly define lines l_{YZ}, l_{ZX} with respective involutions $\rho_{YZ}, \rho_{ZX} \in \text{Inv}$. The triple $(\rho_{XY}, \rho_{YZ}, \rho_{ZX})$ defines the homomorphism $\hat{\rho}$ of Theorem 3.2.2.

Claim. $X = \rho_{ZX}\rho_{XY}$. To this end we show $X\rho_{XY}$ equals ρ_{ZX} . First, $X\rho_{XY}$ fixes $\text{Fix}(X)$, since both X and ρ_{XY} fix $\text{Fix}(X)$. By (3.7),

$$\rho_{XY}XY\rho_{XY} = X^{-1}Y^{-1} = X^{-1}(XY)^{-1}X.$$

Equivalently,

$$\rho_{XY}Z^{-1}\rho_{XY} = X^{-1}Y^{-1} = X^{-1}ZX,$$

which implies

$$(X\rho_{XY})Z^{-1}(X\rho_{XY})^{-1} = Z.$$

Now Lemma 3.2.5 (1) implies that $X\rho_{XY}$ preserves $\text{Fix}(Z)$ and its restriction to the corresponding line is a reflection. Thus $X\rho_{XY}$ is itself an involution ℓ_Z with $\text{Fix}(X\rho_{XY})$ orthogonal to the axis of Z .

Since $X\rho_{XY}$ fixes $\text{Fix}(X)$, it follows that $X\rho_{XY} = \rho_{ZX}$ as claimed. Similarly $Y\rho_{YZ} = \rho_{XY}$ and $Z\rho_{ZX} = \rho_{YZ}$, completing the proof of Theorem 3.2.2. \square

3.3 Orthogonal reflection groups

An algebraic proof of Theorem 3.2.2 involves three-dimensional inner product spaces and is described in Goldman [27]. This proof exploits the isomorphism $\text{PSL}(2, \mathbb{C}) \rightarrow \text{SO}(3, \mathbb{C})$.

The 3-dimensional orthogonal representation of $\text{PSL}(2, \mathbb{C})$. Let $W = \mathbb{C}^2$ with a nondegenerate symplectic form ω . The *symmetric square* $\text{Sym}^2(W)$ is a 3-dimensional vector space based on monomials $e \cdot e, e \cdot f, f \cdot f$, where e, f is a basis of W , and $x \cdot y$ denotes the *symmetric product* of x, y (the image of the tensor product $x \otimes y$ under symmetrization). $\text{Sym}^2(W)$ inherits a symmetric inner product defined by:

$$(u_1 \cdot u_2, v_1 \cdot v_2) \mapsto \frac{1}{2}(\omega(u_1, v_1)\omega(u_2, v_2) + \omega(u_1, v_2)\omega(u_2, v_1)).$$

If $e, f \in W$ is a symplectic basis for W , the corresponding inner product for $\text{Sym}^2(W)$ has matrix

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & -1/2 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

with respect to the above basis of $\text{Sym}^2(W)$. In particular the inner product is nondegenerate. Every $\xi \in \text{SL}(2, \mathbb{C})$ induces an isometry of $\text{Sym}^2(W)$ with respect to this inner product. This correspondence defines a local isomorphism

$$\text{SL}(2, \mathbb{C}) \xrightarrow{\text{Sym}^2} \text{SO}(3, \mathbb{C})$$

with kernel $\{\pm \text{Id}\}$ and a resulting *isomorphism* $\text{PSL}(2, \mathbb{C}) \rightarrow \text{SO}(3, \mathbb{C})$.

If $\xi \in \mathrm{SL}(2, \mathbb{C})$, then

$$\mathrm{tr}(\mathrm{Sym}^2(\xi)) = \mathrm{tr}(\xi)^2 - 1. \quad (3.8)$$

For example, the diagonal matrix

$$\xi = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$$

induces the diagonal matrix

$$\mathrm{Sym}^2(\xi) = \begin{bmatrix} \lambda^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda^{-2} \end{bmatrix}$$

and

$$\mathrm{tr}(\mathrm{Sym}^2(\xi)) = \lambda^2 + 1 + \lambda^{-2} = (\lambda + \lambda^{-1})^2 - 1.$$

Alternatively, this is the adjoint representation of $\mathrm{SL}(2, \mathbb{C})$ on its Lie algebra $\mathfrak{sl}(2) \cong \mathrm{Sym}^2(W)$. Here the standard basis of \mathbb{C} is

$$e = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad f = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and the monomials correspond to

$$e \cdot e = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad e \cdot f = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad f \cdot f = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

The inner product corresponds to the *trace form*

$$(X, Y) \mapsto \frac{1}{2} \mathrm{tr}(XY)$$

which is 1/8 the Killing form on $\mathfrak{sl}(2, \mathbb{C})$.

3-dimensional inner product spaces. Let e_1, e_2, e_3 denote the standard basis of \mathbb{C}^3 . A 3×3 symmetric matrix B determines an inner product \mathbb{B} on \mathbb{C}^3 by the usual rule:

$$(v, w) \xrightarrow{\mathbb{B}} v^\top B w.$$

We suppose that \mathbb{B} is nonzero on e_1, e_2, e_3 ; in fact, let's normalize \mathbb{B} so that its basic values are 1:

$$\mathbb{B}(e_i, e_i) = 1 \quad \text{for } i = 1, 2, 3.$$

In other words, the diagonal entries satisfy $B_{11} = B_{22} = B_{33} = 1$.

Let $R_i = R_i^{(B)}$ denote the orthogonal reflection in e_i defined by \mathbb{B} :

$$v \xrightarrow{R_i} v - 2\mathbb{B}(v, e_i)e_i,$$

with corresponding matrix

$$R_i := \mathrm{Id} - 2e_i(e_i)^\top B.$$

$(e_i(e_i)^\top B$ is the 3×3 matrix with the same i -th row as B and the other two rows zero.) Since

$$\text{Id} = \mathbb{B}(e_i, e_i) = (e_i)^\top B e_i \quad (\text{matrix multiplication}),$$

$$\begin{aligned} R_i^\top B R_i - B &= (\text{Id} - 2B e_i (e_i)^\top) B (\text{Id} - 2e_i (e_i)^\top B) - B \\ &= -2B e_i (e_i)^\top B - 2B e_i (e_i)^\top B + 4B e_i (e_i)^\top B e_i (e_i)^\top B \\ &= -2B e_i (e_i)^\top B - 2B e_i (e_i)^\top B + 4B e_i (e_i)^\top B \\ &= 0, \end{aligned}$$

so R_i is orthogonal with respect to \mathbb{B} .

Thus the matrix B determines a triple of involutions $R_1^{(B)}, R_2^{(B)}, R_3^{(B)}$ in the orthogonal group of \mathbb{B} :

$$\text{O}(\mathbb{C}^3, \mathbb{B}) := \{\xi \in \text{GL}(2, \mathbb{C}) \mid \xi^\top B \xi = B\}.$$

In other words, B defines a representation $\hat{\rho} := \hat{\rho}^{(B)}$ of the free product

$$\hat{\pi} := \mathbb{Z}/2 * \mathbb{Z}/2 * \mathbb{Z}/2$$

in $\text{O}(\mathbb{C}^3, \mathbb{B})$, taking the free generators $\iota_{XY}, \iota_{YZ}, \iota_{ZX}$ of $\hat{\pi}$ into $R_1^{(B)}, R_2^{(B)}, R_3^{(B)}$ respectively. The restriction $\rho := \rho^{(B)}$ of $\hat{\rho}^{(B)}$ to the index-two subgroup

$$\mathbb{Z} * \mathbb{Z} \cong \pi \subset \hat{\pi}$$

(compare §3.2) assumes values in the subgroup

$$\text{SO}(\mathbb{C}^3, \mathbb{B}) := \text{SL}(3, \mathbb{C}) \cap \text{O}(\mathbb{C}^3, \mathbb{B}).$$

When \mathbb{B} is nondegenerate, then $\text{SO}(\mathbb{C}^3, \mathbb{B}) \cong \text{SO}(3, \mathbb{C})$ (a specific isomorphism corresponds to an orthonormal basis for \mathbb{B}). There are exactly four lifts $\tilde{\rho}$ of ρ to the double covering-space

$$\text{SL}(2, \mathbb{C}) \longrightarrow \text{SO}(3, \mathbb{C}) \xrightarrow{\cong} \text{SO}(\mathbb{C}^3, \mathbb{B}).$$

To see this, for each generator X, Y, Z of π , its ρ -image has exactly two lifts, differing by $\pm \text{Id}$. Lifting the generators to $\widetilde{\rho(X)}, \widetilde{\rho(Y)}, \widetilde{\rho(Z)}$ respectively, exactly half of the eight choices satisfy

$$\widetilde{\rho(X)} \widetilde{\rho(Y)} \widetilde{\rho(Z)} = \text{Id} \tag{3.9}$$

(as desired), and for the other four choices the product equals $-\text{Id}$.

Choose one of the four lifts satisfying (3.9), and denote it $\tilde{\rho}$. If $i \neq j$, the trace of $R_i R_j \in \text{SL}(3, \mathbb{C})$ equals $4(B_{ij})^2 - 1$. For example, take $i = 1, j = 2$:

$$\begin{aligned} \widetilde{\rho}(Z) = R_1 R_2 &= \begin{bmatrix} -1 & -2B_{12} & -2B_{13} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2B_{12} & -1 & -2B_{23} \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 4(B_{12})^2 - 1 & -2B_{12} & 4B_{23}B_{12} - 2B_{13} \\ -2B_{12} & -1 & -2B_{23} \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

has trace $4(B_{12})^2 - 1$.

This calculation gives another proof of surjectivity in Theorem A (as in [27]). For $(x, y, z) \in \mathbb{C}^3$, the matrix

$$B = \begin{bmatrix} 1 & z/2 & y/2 \\ z/2 & 1 & x/2 \\ y/2 & x/2 & 1 \end{bmatrix} \tag{3.10}$$

defines a bilinear form \mathbb{B} and a representation $\rho^{(B)}$ as above.

The corresponding $\text{SL}(2, \mathbb{C})$ -traces of the $\tilde{\rho}$ -images of the generators X, Y, Z of π satisfy

$$\begin{aligned} \text{tr}(\tilde{\rho}(X)) &= \pm 2B_{23}, \\ \text{tr}(\tilde{\rho}(Y)) &= \pm 2B_{13}, \\ \text{tr}(\tilde{\rho}(Z)) &= \pm 2B_{12}, \end{aligned}$$

because (using (3.8))

$$\begin{aligned} \text{tr}(\tilde{\rho}(X))^2 &= 1 + \text{tr}(\text{Sym}^2(\tilde{\rho}(X))) = 1 + \text{tr}(R_2 R_3) = 4(B_{23})^2, \\ \text{tr}(\tilde{\rho}(Y))^2 &= 1 + \text{tr}(\text{Sym}^2(\tilde{\rho}(Y))) = 1 + \text{tr}(R_3 R_1) = 4(B_{31})^2, \\ \text{tr}(\tilde{\rho}(Z))^2 &= 1 + \text{tr}(\text{Sym}^2(\tilde{\rho}(Z))) = 1 + \text{tr}(R_1 R_2) = 4(B_{12})^2. \end{aligned}$$

Now adjust the lifts as above to arrange a representation $\pi \xrightarrow{\tilde{\rho}} \text{SL}(2, \mathbb{C})$ with

$$\begin{aligned} \text{tr}(\tilde{\rho}(X)) &= x, \\ \text{tr}(\tilde{\rho}(Y)) &= y, \\ \text{tr}(\tilde{\rho}(Z)) &= z \end{aligned}$$

as desired.

When $\kappa(x, y, z) = 2$, the matrix B is singular and we obtain *reducible* representations. There are two cases, depending on whether $\text{rank}(B) = 2$ or $\text{rank}(B) = 1$. (Since $B \neq 0$, its rank cannot be zero.)

3.4 Real characters and real forms

A *real character* $(x, y, z) \in \mathbb{R}^3$ corresponds to a representation of a rank-two free group in one of the two *real forms* $SU(2)$, $SL(2, \mathbb{R})$ of $SL(2, \mathbb{C})$. This was first stated and proved in general by Morgan–Shalen [61]. Geometrically, $SU(2)$ -representations are those which fix a point in \mathbb{H}^3 , and $SL(2, \mathbb{R})$ -representations are those which preserve a plane $\mathbb{H}^2 \subset \mathbb{H}^3$ as well as an orientation on the plane.

Theorem 3.4.1. *Let $(x, y, z) \in \mathbb{R}^3$ and*

$$\kappa(x, y, z) := x^2 + y^2 + z^2 - xyz - 2.$$

Let $\pi \xrightarrow{\rho} SL(2, \mathbb{C})$ be a representation with character (x, y, z) . Suppose first that $\kappa(x, y, z) \neq 2$.

- *If $-2 \leq x, y, z \leq 2$ and $\kappa(x, y, z) < 2$, then $\rho(\pi)$ fixes a unique point in \mathbb{H}^3 and is conjugate to an $SU(2)$ -representation.*
- *Otherwise $\rho(\pi)$ preserves a unique plane in \mathbb{H}^3 and its restriction to that plane preserves orientation.*

If $\kappa(x, y, z) = 2$, then ρ is reducible and one of the following must occur:

- *$\rho(\pi)$ acts identically on \mathbb{H}^3 , in which case $\rho(\pi) \subset \{\pm \text{Id}\}$ is a central representation.*
- *$\rho(\pi)$ fixes a line in \mathbb{H}^3 , in which case $-2 \leq x, y, z \leq 2$ and ρ is conjugate to a representation taking values in $SO(2) = SU(2) \cap SL(2, \mathbb{R})$.*
- *$\rho(\pi)$ acts by transvections along a unique line in \mathbb{H}^3 , in which case*

$$x, y, z \in \mathbb{R} \setminus (-2, 2).$$

Then ρ is conjugate to a representation taking values in $SO(1, 1) \subset SL(2, \mathbb{R})$.

- *$\rho(\pi)$ fixes a unique point on $\partial_\infty \mathbb{H}^3$.*

Recall that $SO(1, 1)$ is isomorphic to the multiplicative group \mathbb{R}^* of nonzero real numbers, and is conjugate to the subgroup of $SL(2, \mathbb{R})$ consisting of diagonal matrices.

Corollary 3.2.3 associates to a generic representation ρ an ordered triple ι^ρ of geodesics in \mathbb{H}^3 . When ρ is irreducible, the corresponding cases for ι^ρ are the following:

- *If ρ fixes a unique point $p \in \mathbb{H}^3$, then the three lines are distinct and intersect in p . Conversely if the three lines are concurrent, then ρ is conjugate to an $SU(2)$ -representation.*
- *If ρ preserves a unique plane P , then the three lines are distinct. There are two cases:*
 - *The three lines are orthogonal to P .*
 - *The three lines lie in P .*

The first case, when the lines are orthogonal to P , occurs when $\kappa(x, y, z) < 2$. In this case the corresponding involutions preserve orientation on P . The second case, when the lines lie in P , occurs when $\kappa(x, y, z) > 2$. In that case the involutions restrict to reflections in geodesics in P which reverse orientation.

Real symmetric 3×3 matrices. We deduce these facts from the classification given in Theorem A. First assume that (x, y, z) is an *irreducible* character, i.e., $\kappa(x, y, z) \neq 2$. Theorem A implies that (x, y, z) is the character of the representation ρ given by (2.10), and all such characters are $\text{PGL}(2, \mathbb{C})$ -conjugate.

The matrix B defining the bilinear form \mathbb{B} in (3.10) satisfies

$$4 \det(B) = 2 - \kappa(x, y, z)$$

so \mathbb{B} is degenerate if and only if $\kappa(x, y, z) = 2$, in which case ρ is reducible.

Suppose that \mathbb{B} is nondegenerate, so that either $\kappa(x, y, z) > 2$ or $\kappa(x, y, z) < 2$. Suppose first that $\kappa(x, y, z) > 2$. Then $\det(B) < 0$. Since the diagonal entries of B are positive, \mathbb{B} is indefinite of signature $(2, 1)$. In particular, it cannot be negative definite. In this case the triple of lines corresponding to ρ are all coplanar. The corresponding involutions reverse orientation on P and act by reflections of P in the three geodesics respectively.

When $\kappa(x, y, z) < 2$, there are two cases: either \mathbb{B} is positive definite (signature $(3, 0)$) or indefinite (signature $(1, 2)$). The restriction of \mathbb{B} to the coordinate plane spanned by e_i and e_j is given by the 2×2 symmetric matrix

$$\begin{bmatrix} 1 & B_{ij} \\ B_{ij} & 1 \end{bmatrix}$$

which is positive definite if and only if $-1 < B_{ij} < 1$. Thus \mathbb{B} is positive definite if and only if $-2 < x, y, z < 2$.

Otherwise \mathbb{B} is indefinite and ρ corresponds to a representation in $\text{SO}(1, 2)$. In this case the triple of lines in \mathbb{H}^3 are all orthogonal to the invariant plane P in \mathbb{H}^3 . The corresponding three involutions preserve orientation on P and act by symmetries about points in P .

The two-dimensional normal form. Another approach to finding a representation with given traces involves a direct computation with the explicit normal form (2.10) as follows. Let (x, y, z) be as above. First solve $z = 2 \cos(\theta)$ to obtain representative matrices:

$$\xi_x := \begin{bmatrix} x & -1 \\ 1 & 0 \end{bmatrix}, \quad \eta_{y,\theta} := \begin{bmatrix} 0 & e^{-i\theta} \\ -e^{i\theta} & y \end{bmatrix},$$

with a slight change of notation from (2.10).

A Hermitian form on \mathbb{C}^2 is given by a Hermitian 2×2 -matrix H . A complex 2×2 matrix H is Hermitian if and only if $H = \bar{H}^\top$. The corresponding Hermitian form

on \mathbb{C}^2 is:

$$(u, v) \mapsto \bar{v}^\top H u,$$

where $u, v \in \mathbb{C}^2$. A linear transformation $\mathbb{C}^2 \xrightarrow{\xi} \mathbb{C}^2$ preserves H if and only if $\bar{\xi}^\top H \xi = H$.

The ξ_x -invariant Hermitian forms comprise the real vector space with basis

$$\begin{bmatrix} 2 & x \\ x & 2 \end{bmatrix}, \quad \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$$

and the $\eta_{y,\theta}$ -invariant Hermitian forms comprise the real vector space with basis

$$\begin{bmatrix} 2 & -y \sec(\theta) \\ -y \sec(\theta) & 2 \end{bmatrix}, \quad \begin{bmatrix} 0 & i - y \tan(\theta) \\ -i - y \tan(\theta) & 0 \end{bmatrix}.$$

The ρ -invariant Hermitian forms thus comprise the intersection of these two vector spaces, the vector space with basis

$$H = \begin{bmatrix} 2 \sin(\theta) & x \sin(\theta) - i(y + xz/2) \\ x \sin(\theta) + i(y + xz/2) & 2 \sin(\theta) \end{bmatrix}.$$

This Hermitian matrix is definite since its determinant is positive:

$$\begin{aligned} \det(H) &= 4 \sin^2(\theta) - x^2 \sin^2(\theta) - (y - xz/2)^2 \\ &= 2 - \kappa(x, y, z) > 0 \end{aligned}$$

(since $\sin^2(\theta) = 1 - (z/2)^2$ and $\kappa(x, y, z) < 2$).

4 Hyperbolic structures on surfaces of $\chi = -1$

We apply this theory to compute, in trace coordinates, the deformation spaces of hyperbolic structures on compact connected surfaces Σ with $\chi(\Sigma) = -1$. Equivalently, such surfaces are characterized by the condition that $\pi_1(\Sigma)$ is a free group of rank two. There are four possibilities:

- Σ is homeomorphic to a three-holed sphere (a “pair-of-pants” or “trinion”) $\Sigma_{0,3}$;
- Σ is homeomorphic to a one-holed torus $\Sigma_{1,1}$;
- Σ is homeomorphic to a one-holed Klein bottle $C_{1,1}$;
- Σ is homeomorphic to a two-holed projective plane $C_{0,2}$.

Each of these surfaces can be realized as a ribbon graph with three bands connecting two 2-cells. The number of boundary components and the orientability can be read off from the parities of the number of twists in the three bands. Compare Figures 2–5.

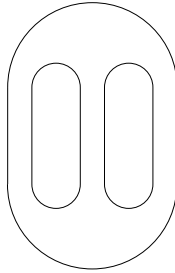


Figure 2. A ribbon graph for a three-holed sphere.

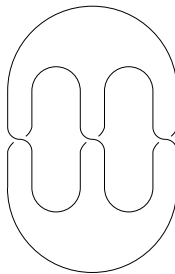


Figure 3. A ribbon graph for a one-holed torus.

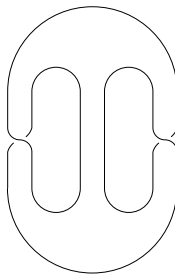


Figure 4. A ribbon graph for a two-holed cross-surface.

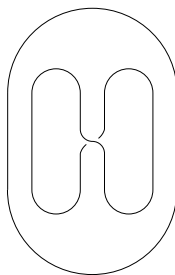


Figure 5. A ribbon graph for a one-holed Klein bottle.

4.1 Fricke spaces

The *Fricke space* $\mathfrak{F}(\Sigma)$ is the space of isotopy classes of marked hyperbolic structures on Σ with $\partial\Sigma$ geodesic. The group of isometries of \mathbb{H}^2 equals $\mathrm{PGL}(2, \mathbb{R})$, which embeds in $\mathrm{PSL}(2, \mathbb{C})$. Its identity component $\mathrm{PSL}(2, \mathbb{R})$ consists of the isometries of \mathbb{H}^2 which *preserve an orientation* on \mathbb{H}^2 . The *holonomy map* embeds $\mathfrak{F}(\Sigma)$ in the deformation space

$$\mathrm{Hom}(\pi_1(\Sigma), \mathrm{PGL}(2, \mathbb{R})) // \mathrm{PGL}(2, \mathbb{R}).$$

Since $\partial\Sigma \neq \emptyset$, $\pi_1(\Sigma)$ is a free group, and the problem of lifting a representation of $\pi_1(\Sigma)$ to $\mathrm{GL}(2, \mathbb{R})$ is unobstructed.

The various lifts are permuted by the group $H^1(\Sigma; \mathbb{Z}/2)$, which is isomorphic to $\mathbb{Z}/2 \oplus \mathbb{Z}/2$ when $\chi(\Sigma) = -1$. In terms of trace coordinates on the \mathbb{R} -locus of the character variety this action is given by:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \begin{bmatrix} x \\ -y \\ -z \end{bmatrix}, \quad \begin{bmatrix} -x \\ y \\ -z \end{bmatrix}, \quad \begin{bmatrix} -x \\ -y \\ z \end{bmatrix}.$$

Theorem 4.1.1. *Using trace coordinates of the boundary, the Fricke space of the three-holed sphere $\Sigma_{0,3}$ identifies with the quotient of the four octants*

$$\begin{aligned} &(-\infty, -2] \times (-\infty, -2] \times (-\infty, -2] \sqcup [(-\infty, -2] \times [2, \infty) \times [2, \infty) \\ &\qquad \qquad \qquad \sqcup [2, \infty) \times [2, \infty) \times (-\infty, -2] \\ &\qquad \qquad \qquad \sqcup [2, \infty) \times (-\infty, -2] \times [2, \infty) \subset \mathbb{R}^3 \end{aligned}$$

by $H^1(\Sigma, \mathbb{Z}/2)$. The octant

$$(-\infty, -2] \times (-\infty, -2] \times (-\infty, -2]$$

defines a slice for the $H^1(\Sigma, \mathbb{Z}/2)$ -action.

The proof will be given in §4.3.

Theorem 4.1.2. *The Fricke space of the one-holed torus $\Sigma_{1,1}$ identifies with the quotient of*

$$\kappa^{-1}((-\infty, -2]) = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 - xyz \leq 0\}$$

by $H^1(\Sigma, \mathbb{Z}/2)$. The region

$$\{(x, y, z) \in (2, \infty)^3 \mid x^2 + y^2 + z^2 - xyz \leq 0\}$$

is a connected component of $\kappa^{-1}((-\infty, -2])$, defines a slice for the $H^1(\Sigma, \mathbb{Z}/2)$ -action, and hence identifies with the Fricke space of Σ .

The proof will be given in §4.4.

4.2 Two-dimensional hyperbolic geometry

We take for our model of the *hyperbolic plane* \mathbb{H}^2 the subset of \mathbb{H}^3 comprising quaternions $z + uj$, where $u > 0$ and $z \in \mathbb{R}$. A matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{GL}(2, \mathbb{C})$$

determines a projective transformation of $\mathbb{C}\mathbb{P}^1 = \partial\mathbb{H}^3$ which extends to an orientation-preserving isometry of \mathbb{H}^3 . This isometry preserves \mathbb{H}^2 if and only if it is a scalar multiple of a real matrix. As usual, normalize $A \in \mathrm{GL}(2, \mathbb{C})$ by dividing by a square root $\sqrt{\det(A)} \in \mathbb{C}^*$. An element of

$$\mathrm{SL}(2, \mathbb{C}) \cap \mathbb{C}^* \mathrm{GL}(2, \mathbb{R})$$

is either:

- a real matrix of determinant 1, or
- a purely imaginary matrix iA' where $A' \in \mathrm{GL}(2, \mathbb{R})$ satisfies $\det(A') = 1$.

In the first case the corresponding orientation-preserving isometry of \mathbb{H}^3 preserves orientation on \mathbb{H}^2 , and in the second case its restriction *reverses orientation* on \mathbb{H}^2 . The traces of their representative matrices in $\mathrm{SL}(2, \mathbb{C})$ distinguish these cases. We emphasize that these representatives are only determined up to ± 1 . Suppose that $A \in \mathrm{PSL}(2, \mathbb{C})$ preserves a plane $P \subset \mathbb{H}^3$ and let $\tilde{A} \in \mathrm{SL}(2, \mathbb{C})$ be a lift of A . Then:

- The restriction of A to P preserves orientation if and only if $\mathrm{tr}(\tilde{A}) \in \mathbb{R}$.
- The restriction of A to P reverses orientation if and only if $\mathrm{tr}(\tilde{A}) \in i\mathbb{R}$.

Observe that an element $\tilde{A} \in \mathrm{SL}(2, \mathbb{C})$ with $\mathbb{P}(\tilde{A}) = A$ whose trace is both real and purely imaginary – that is, equals zero – is an involution in a line $\ell := \mathrm{Fix}(A)$. The A -invariant planes fall into two types: those which contain ℓ , upon which A reverses orientation, and those which are orthogonal to ℓ , upon which A preserves orientation.

The hyperbolic plane and involutions of \mathbb{H}^3 . The proof of Theorem 4.1.1 requires an algebraic representation of half-planes in \mathbb{H}^2 . Given an orientation on \mathbb{H}^2 , and an oriented geodesic $\ell \subset \mathbb{H}^2$, there is a well-determined half-plane bounded by ℓ , defined as follows. Let $x \in \ell$. Choose the unit vector v_ℓ tangent to ℓ at x determined by the orientation of ℓ . The choice of half-plane $\mathfrak{H} \subset \mathbb{H}^2 \setminus \ell$ is determined by the normal vector ν to ℓ at x pointing outward from \mathfrak{H} . Choose \mathfrak{H} so that the basis $\{v_\ell, \nu\} \subset T_x\mathbb{H}^2$ is positively oriented.

The points of \mathbb{H}^2 identify with geodesics in \mathbb{H}^3 which are orthogonal to \mathbb{H}^2 ; the endpoints of such geodesics are complex-conjugate elements of $\mathbb{C}\mathbb{P}^1 \setminus \mathbb{R}\mathbb{P}^1$. An involution which interchanges these endpoints is given by matrices $\pm I_{x+ju}$, where the 2×2 real matrix

$$I_{x+ju} := \frac{1}{u} \begin{bmatrix} x & -(x^2 + u^2) \\ 1 & -x \end{bmatrix} \quad (4.1)$$

has determinant 1 and trace 0. (Apply (3.2), taking $z_1 = x + iu$ and $z_2 = x - iu$.) The space of such matrices has two components, depending on the signs of the off-diagonal elements. A matrix

$$A \in \mathrm{SL}(2, \mathbb{R}) \cap \mathfrak{sl}(2, \mathbb{R})$$

equals I_{x+ju} , for some $x + ju \in \mathbb{H}^2$ if and only if $A_{12} < 0 < A_{21}$, in which case

$$u = (A_{21})^{-1}, \quad x = (A_{21})^{-1}A_{11}.$$

The above inequalities determine one of the two sheets of the two-sheeted hyperboloid in $\mathfrak{sl}(2, \mathbb{R})$ defined by the unimodularity condition $\det(A) = 1$.

This gives a convenient form of the Klein hyperboloid model for \mathbb{H}^2 , as the quadric in $\mathfrak{sl}(2, \mathbb{R}) \cong \mathbb{R}^3$ defined by

$$1 = \det(A) = -\frac{1}{2}\mathrm{tr}(A^2).$$

Next we represent oriented geodesics by involutions. A geodesic in \mathbb{H}^2 determines an involution by (3.2) and (3.3), where z_1, z_2 are distinct points in $\mathbb{R}\mathbb{P}^1$. Such a matrix is purely imaginary, has trace zero and determinant one. Multiplying by i , we obtain an element A of $\mathfrak{sl}(2, \mathbb{R})$ which has determinant -1 . Such a matrix has well-defined 1-dimensional eigenspaces with eigenvalues $\pm i$. These eigenspaces determine the respective fixed points in $\mathbb{R}\mathbb{P}^1$. Replacing A by $-A$ interchanges the $\pm i$ -eigenspaces. In this way, we identify the set of oriented geodesics in \mathbb{H}^2 with

$$\{A \in \mathfrak{sl}(2, \mathbb{R}) \mid \det(A) = -1\}.$$

This is just the usual hyperboloid model. Think of $\mathfrak{sl}(2, \mathbb{R})$ as a 3-dimensional real inner product space under the inner product

$$\langle A, B \rangle := \frac{1}{2}\mathrm{tr}(AB).$$

The corresponding quadratic form relates to the *determinant* by

$$\langle A, A \rangle = \frac{1}{2}\mathrm{tr}(A^2) = -\det(A).$$

This quadratic form is readily seen to have signature (2, 1) since its value on

$$\begin{bmatrix} a & b \\ c & -a \end{bmatrix}$$

equals $a^2 + bc$. Then \mathbb{H}^2 corresponds to one component of the two-sheeted hyperboloid (say the one with $b < 0 < c$)

$$\{v \in \mathfrak{sl}(2, \mathbb{R}) \mid \langle v, v \rangle = -1\}$$

and the space of oriented geodesics corresponds to the *de Sitter space*

$$d\mathbb{S}_1^2 := \{v \in \mathfrak{sl}(2, \mathbb{R}) \mid \langle v, v \rangle = 1\}.$$

A vector $v \in \text{dS}_1^2$ determines a half-plane $\mathcal{H}(v)$ by:

$$\mathcal{H}(v) := \{w \in \mathbb{H}^2 \mid \langle w, v \rangle \geq 0\}. \quad (4.2)$$

In particular, $\mathcal{H}(-v)$ is the half-plane complementary to $\mathcal{H}(v)$.

For example, the half-plane corresponding to

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

consists of all $x + uj$ where $x \geq 0$, as can easily be verified using (4.1).

The main criterion for disjointness of half-planes is the following lemma, whose proof is an elementary exercise and left to the reader. (Recall that two geodesics in \mathbb{H}^2 are *ultraparallel* if and only if they admit a common orthogonal geodesic; equivalently distances between their respective points have a positive lower bound.)

Lemma 4.2.1. *Let $v_1, v_2 \in \text{dS}_1^2$ determine geodesics $\ell_1, \ell_2 \subset \mathbb{H}^2$ and half-planes $\mathcal{H}_i := \mathcal{H}(v_i)$ with $\partial\mathcal{H}_i = \ell_i$. The following conditions are equivalent:*

- $|\langle v_1, v_2 \rangle| > 1$.
- *The invariant geodesics ℓ_1 and ℓ_2 are ultraparallel.*

In this case, the following two further conditions are equivalent:

- $\langle v_1, v_2 \rangle > 1$.
- *Either $\mathcal{H}_1 \subset \mathcal{H}_2$ or $\mathcal{H}_2 \subset \mathcal{H}_1$.*

Contrariwise, the following two conditions are equivalent:

- $\langle v_1, v_2 \rangle < -1$.
- *Either \mathcal{H}_1 and \mathcal{H}_2 are disjoint or their complements are disjoint.*

Hyperbolic isometries. An element $A \in \text{SL}(2, \mathbb{R})$ is *hyperbolic* if it satisfies any of the following equivalent conditions:

- $\text{tr}(A) > 2$ or $\text{tr}(A) < -2$.
- A has distinct real eigenvalues.
- The isometry of \mathbb{H}^2 defined by A has exactly two fixed points on $\partial\mathbb{H}^2$.
- The isometry of \mathbb{H}^2 defined by A leaves invariant a (necessarily unique) geodesic ℓ_A , upon which it acts by a nontrivial translation.

A geodesic ℓ in \mathbb{H}^2 is specified by its *reflection* ρ_ℓ , an isometry of \mathbb{H}^2 whose fixed point set equals ℓ . If $v \in \text{dS}_1^2$ is a vector corresponding to ℓ , then ρ_ℓ is the restriction to \mathbb{H}^2 of the orthogonal reflection in $\text{SO}(2, 1)$ fixing v :

$$u \xrightarrow{\rho_v} -u + 2\langle u, v \rangle v.$$

(3.1) implies the following useful formula for the invariant axis of a hyperbolic element:

Lemma 4.2.2. *Let A be hyperbolic. Then*

$$\hat{A} := \frac{2A - \text{tr}(A)\text{Id}}{\sqrt{\text{tr}(A)^2 - 4}} \in \widetilde{\text{Inv}} \cap \text{SL}(2, \mathbb{R}) \tag{4.3}$$

defines the reflection in the invariant axis of A .

Notice that

$$\widehat{-A} = \widehat{A^{-1}} = -\hat{A},$$

so that A and A^{-1} determine complementary half-planes.

For example

$$A = \begin{bmatrix} e^{l/2} & 0 \\ 0 & e^{-l/2} \end{bmatrix}$$

represents translation along a geodesic (the imaginary axis in \mathbb{H}^2) by distance $l > 0$ from 0 to ∞ . The corresponding reflection is

$$\hat{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

which determines the half-plane

$$\mathcal{H}(\hat{A}) = \{x + uj \in \mathbb{H}^2 \mid x \geq 0, u > 0\}$$

as above.

4.3 The three-holed sphere

We now show that a representation corresponding to a character $(x, y, z) \in \mathbb{R}^3$ satisfying $x, y, z < -2$ is the holonomy representation of a hyperbolic structure on a three-holed sphere. We find matrices X, Y, Z of the desired type and compute the corresponding reflections $\hat{X}, \hat{Y}, \hat{Z}$. Then we show that the corresponding half-planes are all disjoint (after possibly replacing $\hat{X}, \hat{Y}, \hat{Z}$ by their negatives). From this we construct a developing map for a hyperbolic structure on Σ . For details on geometric structures on manifolds and their developing maps, see Goldman [28], [30], [32] or Thurston [75].

Lemma 4.3.1. *Suppose $X, Y, Z \in \text{SL}(2, \mathbb{C})$ satisfy $XYZ = \text{Id}$ and have real traces $x, y, z < -2$ respectively. Then the inner products*

$$\langle \hat{X}, \hat{Y} \rangle, \quad \langle \hat{Y}, \hat{Z} \rangle, \quad \langle \hat{Z}, \hat{X} \rangle$$

are all < -1 .

Proof. The proof breaks into a series of calculations. By symmetry it suffices to prove $\langle \hat{X}, \hat{Y} \rangle < -1$. By the definition (4.3)

$$\begin{aligned} \langle \hat{X}, \hat{Y} \rangle &= \frac{1}{2} \operatorname{tr} \left(\frac{2X - x \operatorname{Id}}{\sqrt{x^2 - 4}} \frac{2Y - y \operatorname{Id}}{\sqrt{y^2 - 4}} \right) \\ &= \frac{\operatorname{tr}(4XY - 2xY - 2yX + xy \operatorname{Id})}{2\sqrt{(x^2 - 4)(y^2 - 4)}} \\ &= \frac{2z - xy}{\sqrt{(x^2 - 4)(y^2 - 4)}} \end{aligned} \tag{4.4}$$

since $\operatorname{tr}(XY) = z$, $\operatorname{tr}(Y) = y$, $\operatorname{tr}(X) = x$ and $\operatorname{tr}(\operatorname{Id}) = 2$. Because

$$x, y, z < -2 \implies 2z - xy < 0,$$

the calculation above implies

$$\langle \hat{X}, \hat{Y} \rangle < 0. \tag{4.5}$$

Now

$$x^2 + y^2 + z^2 - xyz - 4 > 4 + 4 + 4 - 8 - 4 = 0$$

implies that

$$(2z - xy)^2 - (x^2 - 4)(y^2 - 4) = 4(x^2 + y^2 + z^2 - xyz - 4) > 0$$

and

$$\left(\frac{2z - xy}{\sqrt{(x^2 - 4)(y^2 - 4)}} \right)^2 > 1. \tag{4.6}$$

Thus (4.4) and (4.6) imply

$$\langle \hat{X}, \hat{Y} \rangle^2 > 1$$

whence (4.5) implies

$$\langle \hat{X}, \hat{Y} \rangle < -1$$

as claimed. □

Conclusion of proof of Theorem 4.1.1. Thus the half-planes $\mathcal{H}_{\hat{X}}, \mathcal{H}_{\hat{Y}}, \mathcal{H}_{\hat{Z}}$ are either all disjoint or their complements are all disjoint. Replacing $\hat{X}, \hat{Y}, \hat{Z}$ by their negatives if necessary, assume that the complements to $\mathcal{H}_{\hat{X}}, \mathcal{H}_{\hat{Y}}, \mathcal{H}_{\hat{Z}}$ are pairwise disjoint.

The intersection

$$\Delta_\infty := \mathcal{H}_{\hat{X}} \cap \mathcal{H}_{\hat{Y}} \cap \mathcal{H}_{\hat{Z}}$$

is bounded by the three geodesics

$$\ell_X = \partial \mathcal{H}_{\hat{X}}, \quad \ell_Y = \partial \mathcal{H}_{\hat{Y}}, \quad \ell_Z = \partial \mathcal{H}_{\hat{Z}}$$

and three segments of ∂H^2 . When some of $\rho(X), \rho(Y), \rho(Z)$ are parabolic, then these segments degenerate into ideal points. If a, b are lines or ideal points, denote their

common orthogonal segment by $\perp(a, b)$. Define:

$$\begin{aligned}\sigma_{XY} &:= \perp(\ell_X, \ell_Y), \\ \sigma_{YZ} &:= \perp(\ell_Y, \ell_Z), \\ \sigma_{ZX} &:= \perp(\ell_Z, \ell_X).\end{aligned}$$

Let $\text{Hex}_\rho \subset \Delta_\infty$ denote the right hexagon bounded by $\sigma_{XY}, \sigma_{YZ}, \sigma_{ZX}$ and segments of ℓ_X, ℓ_Y, ℓ_Z as in Figure 6. Map the abstract hexagon Hex of §3.2 to Hex_ρ so that

$$\begin{aligned}\partial_1(\text{Hex}) &\mapsto \ell_X, \\ \partial_2(\text{Hex}) &\mapsto \ell_Y, \\ \partial_3(\text{Hex}) &\mapsto \ell_Z\end{aligned}$$

and the other three edges of ∂Hex map homeomorphically to $\sigma_{XY}, \sigma_{YZ}, \sigma_{ZX}$ respectively. This mapping embeds Hex into \mathbb{H}^2 .

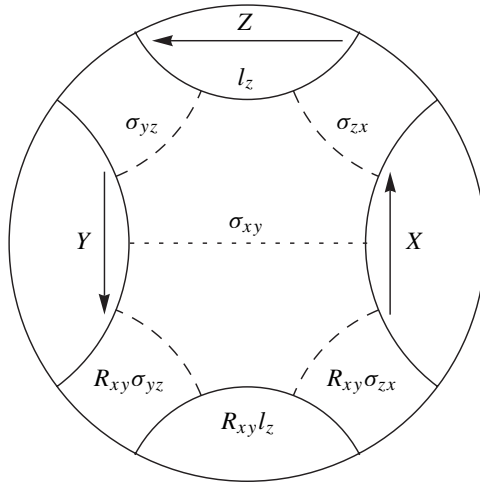


Figure 6. Fundamental domain for hyperbolic structure on $\Sigma_{0,3}$.

A fundamental domain for the action of

$$\pi_1(\Sigma) = \langle X, Y, Z \mid XYZ = 1 \rangle$$

on the universal covering surface

$$\tilde{\Sigma} := (\text{Hex} \times \hat{\pi}) / \sim$$

is the union

$$\Delta := \text{Hex} \cup \iota_{XY}(\text{Hex}).$$

We shall extend the embedding $\text{Hex} \hookrightarrow \mathbb{H}^2$ to a local diffeomorphism (a *developing map*)

$$\tilde{\Sigma} \longrightarrow \mathbb{H}^2$$

which is $\pi_1(\Sigma)$ -equivariant with respect to the action ρ on \mathbb{H}^2 . Pull back the hyperbolic structure from \mathbb{H}^2 to obtain a $\pi_1(\Sigma)$ -invariant hyperbolic structure on Σ , as desired.

Extend $\text{Hex} \hookrightarrow \mathbb{H}^2$ to $\Delta \rightarrow \mathbb{H}^2$ as follows. Map the reflected image of Hex to $\iota_{XY}\text{Hex}_\rho$. Then

$$X = \iota_{ZX}\iota_{XY}$$

identifies the two sides of Δ corresponding to σ_{ZX} and $\iota_{XY}\sigma_{ZX}$, and

$$Y = \iota_{XY}\iota_{YZ}$$

identifies the two sides of Δ corresponding to $\iota_{XY}\sigma_{YZ}$ and σ_{YZ} . (Compare Figure 6.) This defines a hyperbolic structure on Σ with geodesic boundary developing to ℓ_X , ℓ_Y and ℓ_Z .

This completes the proof that every character $(x, y, z) \in (-\infty, 2]^3$ is the holonomy of a hyperbolic structure on $\Sigma_{0,3}$. \square

4.4 The one-holed torus

Now consider the case $\Sigma \approx \Sigma_{1,1}$. Present $\pi = \pi_1(\Sigma)$ as freely generated by X, Y corresponding to simple closed curves which intersect transversely in one point. Then the boundary $\partial\Sigma$ corresponds to the commutator $K = [X, Y]$ and we obtain the presentation

$$\pi = \langle X, Y, Z, K \mid XYZ = \text{Id}, \quad XY = K Y X \rangle$$

The corresponding trace functions are

$$\begin{aligned} x([\rho]) &:= \text{tr}(\rho(X)), \\ y([\rho]) &:= \text{tr}(\rho(Y)), \\ z([\rho]) &:= \text{tr}(\rho(Z)), \\ k([\rho]) &:= \text{tr}(\rho(K)) \\ &= \kappa(x, y, z) = x^2 + y^2 + z^2 - xyz - 2, \end{aligned}$$

which we denote by x, y, z, k (without reference to ρ) when the context is clear.

The goal of this section is to prove Theorem 4.1.2.

Lemma 4.4.1. *Suppose $(x, y, z) \in \mathbb{R}^3$ satisfies $x^2 + y^2 + z^2 - xyz < 4$. Then either*

- $(x, y, z) \in [-2, 2]^3$, or
- $|x|, |y|, |z| > 2$.

In the first case (x, y, z) is the character of an $\text{SU}(2)$ -representation as in Theorem 3.4.1.

Proof. Rewriting the hypothesis as

$$(x^2 - 4)(y^2 - 4) > (2z - xy)^2,$$

it follows that $(x^2 - 4)(y^2 - 4) > 0$. By symmetry $(y^2 - 4)(z^2 - 4) > 0$ and $(z^2 - 4)(x^2 - 4) > 0$ as well. Thus none of $|x|, |y|, |z|$ equal 2, and either

$$x^2 - 4, \quad y^2 - 4, \quad z^2 - 4$$

are all positive or all negative. If

$$|x|, |y|, |z| < 2$$

then (x, y, z) is an $SU(2)$ -character as in Theorem 3.4.1; otherwise

$$|x|, |y|, |z| > 2$$

as desired. This completes the proof of Lemma 4.4.1. □

Denote by $X \subset \Sigma$ a simple closed curve corresponding to the generator $X \in \pi_1(\Sigma)$. The surface-with-boundary $\Sigma' := \Sigma|X$ obtained by *splitting* Σ along X is homeomorphic to a three-holed sphere. Denoting the quotient map by $\Sigma' \xrightarrow{\Pi} \Sigma$, the three components of $\partial\Sigma'$ are the connected preimage $\partial' := \Pi^{-1}(\partial\Sigma)$ and the two components X_{\pm} of the preimage $\Pi^{-1}(X)$. Choose arcs from the basepoint to X_+ and represent the boundary generators of $\pi_1(\Sigma')$ by the elements ∂', X_+, X_- subject to the relation $X_-X_+\partial' = \text{Id}$. The quotient map induces a monomorphism

$$\begin{aligned} \pi_1(\Sigma') &\xrightarrow{\Pi_*} \pi_1(\Sigma), \\ X_+ &\longmapsto X, \\ X_- &\longmapsto YX^{-1}Y^{-1}, \\ \partial' &\longmapsto \partial = X^{-1}YXY^{-1}; \end{aligned}$$

compare Figure 7.

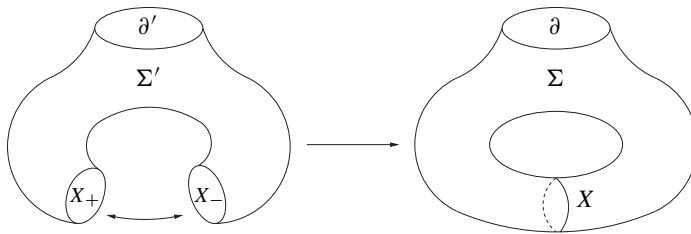


Figure 7. The one-holed torus as an identification space. The identification map Y conjugates X to a boundary element of $\pi_1(\Sigma')$, but with the opposite orientation.

Lemma 4.4.2. *The composition $\rho \circ \Pi_*$ is the holonomy representation of a hyperbolic structure on $\Sigma' \approx \Sigma_{0,3}$.*

Proof. By Theorem 4.1.1, it suffices to show that the images of the boundary generators $\partial', X_-, X_+ \in \pi_1(\Sigma')$ under $\rho \circ \Pi_*$ have trace ≤ -2 . By Lemma 4.4.1,

$$\text{tr}(\Pi_* \circ \rho(X_-)) = \text{tr}(\Pi_* \circ \rho(X_+)) = x$$

is either > 2 or < -2 . In the former case, replace $\rho(X)$ by $-\rho(X)$ to assume that $x < -2$. Now by assumption

$$\text{tr}(\Pi_* \circ \rho(\partial')) = \text{tr}(\rho(K)) = k \leq -2$$

so that all three boundary generators of $\pi_1(\Sigma')$ have trace ≤ -2 , as desired. \square

Conclusion of proof of Theorem 4.1.2. Thus we obtain a hyperbolic structure on Σ' with geodesic boundary. Two of the boundary components correspond to X_- and X_+ . Choosing a developing map with holonomy $\Pi_* \circ \rho$, the isometry $\rho(Y)$ realizes the identification of X_- with X_+ for which Π is the quotient map. The resulting quotient space is homeomorphic to Σ and inherits a hyperbolic structure from the one on Σ' and the identification. Therefore ρ is the holonomy representation of a hyperbolic structure on Σ . This concludes the proof of Theorem 4.1.2. \square

Compare Goldman [30] for a different proof.

The algebraic methods discussed here easily imply several other qualitative geometric facts:

Proposition 4.4.3. *Suppose that $x, y, z \in \mathbb{R}$ satisfy $\kappa(x, y, z) \leq -2$ and $(x, y, z) \neq (0, 0, 0)$. Then (x, y, z) is the character of a representation $\pi \xrightarrow{\rho} \text{SL}(2, \mathbb{R})$ and $\rho(X)$ and $\rho(Y)$ are hyperbolic elements whose axes cross.*

Proof. Since $\kappa(x, y, z) \leq -2 < 2$, Lemma 4.4.1 applies. The representation ρ with character (x, y, z) is conjugate to either an $\text{SU}(2)$ -representation or an $\text{SL}(2, \mathbb{R})$ -representation. Since

$$\text{tr}(\rho([X, Y])) \leq -2,$$

the only possibility for an $\text{SU}(2)$ -representation occurs if $\kappa(x, y, z) = -2$. Then ρ is conjugate to the quaternion representation

$$\rho(X) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \rho(Y) = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix},$$

and $(x, y, z) = (0, 0, 0)$, a contradiction. (Compare §2.6 of Goldman [30].) Thus (x, y, z) corresponds to an $\text{SL}(2, \mathbb{R})$ -representation ρ . Lemma 4.4.1 implies that $\rho(X)$ and $\rho(Y)$ are both hyperbolic. Proposition 3.2.6 implies that the involution fixing the common orthogonal $\perp(\ell_{\rho(X)}, \ell_{\rho(Y)})$ of their respective invariant axes is given by the Lie product $\text{Lie}(\rho(X), \rho(Y))$. In particular their axes cross if and only if $\perp(\ell_{\rho(X)}, \ell_{\rho(Y)})$ is orthogonal to the real plane $\mathbb{H}^2 \subset \mathbb{H}^3$, that is, if $\text{Lie}(\rho(X), \rho(Y))$ defines an orientation-preserving involution of \mathbb{H}^2 . This occurs precisely when the matrix $\text{Lie}(\rho(X), \rho(Y))$ has positive determinant. By (2.15), the Lie product

$\text{Lie}(\rho(X), \rho(Y))$ has positive determinant if and only if the commutator trace inequality

$$\text{tr}([\rho(X), \rho(Y)]) < 2$$

holds, as assumed. The proof of Proposition 4.4.3 is complete. \square

4.5 Fenchel–Nielsen coordinates

In an influential manuscript written in the early 20th century but only recently published, Fenchel and Nielsen [17] gave geometric coordinates for Fricke space. We briefly relate these coordinates to trace coordinates for the surfaces of Euler characteristic -1 .

Pants decompositions. Decompose Σ into a union of three-holed spheres (“pants”)

$$P_1, \dots, P_l$$

along a system of N disjoint simple curves

$$\gamma_1, \dots, \gamma_N \subset \text{int}(\Sigma).$$

Let $\partial_1, \dots, \partial_n$ denote components of $\partial\Sigma$. For a given marked hyperbolic structure on Σ , the curves γ_i, ∂_j may be taken to be simple closed geodesics. Theorem A implies that the isometry type of the complementary subsurfaces P_k are determined by the lengths of the three closed geodesics representing ∂P_k . The resulting map

$$\begin{aligned} \mathfrak{F}(\Sigma) &\xrightarrow{F} (\mathbb{R}_+)^N \times (\mathbb{R}_{\geq 0})^n, \\ \langle M \rangle &\longmapsto \ell_M(\gamma_i) \times \ell_M(\partial_j) \end{aligned}$$

which associates to a hyperbolic surface M the lengths of the geodesics γ_i, ∂_j is onto. Its fibers correspond to the various ways in which the subsurfaces P_k are identified along interior curves γ_i .

Choose a section σ of the map F as follows. Each interior curve γ bounds two subsurfaces, which we denote P' and P'' . The corresponding boundary curves are denoted $\gamma' \subset P'$ and $\gamma'' \subset P''$ respectively. The *twist parameter* $\tau_i \in \mathbb{R}$ represents the displacement between points on the marked surfaces P, P' corresponding to the section σ . This realizes the Fenchel–Nielsen map F as a principal \mathbb{R}^N -bundle over $\mathfrak{F}(\Sigma)$. Wolpert [80], [81], [82] shows that, when Σ is closed and orientable, the Fenchel–Nielsen coordinates on $\mathfrak{F}(\Sigma)$ are *canonical* or *Darboux* coordinates for the symplectic structure arising from the Weil–Petersson Kähler form on Teichmüller space, and indeed F is a moment map for a completely integrable system on $\mathfrak{F}(\Sigma)$.

In the orientable case, $\Sigma \approx \Sigma_{g,n}$, then

$$N = 3(g - 1) + n. \tag{4.7}$$

Since $\chi(\Sigma) = 2 - 2g + n$ and each P_k has Euler characteristic -1 , the number l of subsurfaces P_k equals

$$l = -\chi(\Sigma) = 2 + 2g + n.$$

Consider the set S of pairs (α, C) , where α is one of the $N + n$ curves ∂_i, γ_j and $C \subset P_k$ is a collar neighborhood of α inside P_k for some k . Each C lies in exactly one P_k and each subsurface P_k contains exactly three pairs (α, C) , the cardinality of S equals $3l$. Furthermore the number of collars C equals $2N + n$, since each γ_i is two-sided in Σ and each ∂_j is one-sided. Computing the cardinality of S in two ways

$$2N + n = 3l = 3(2g + 2 + n)$$

implies (4.7).

The nonorientable case, say $\Sigma \approx C_{k,n}$, reduces to the orientable case by cutting along a disjoint family of simple loops: k of them reverse orientation and $N = k + 2 - n$ preserve orientation. This follows easily from the classification of surfaces: $C_{k,n}$ can be obtained from the planar surface $\Sigma_{0,k+n}$ by attaching k cross-caps (copies of $C_{0,k}$) to k of the components of $\Sigma_{0,k+n}$. In the nonorientable surface $\Sigma \approx C_{k,n}$ are k disjoint orientation-reversing simple loops s_1, \dots, s_k so that the surface Σ' obtained by splitting Σ along s_1, \dots, s_k identifies to Σ . Denote the resulting quotient map by

$$\Sigma' \approx \Sigma_{0,n+k} \xrightarrow{\phi} \Sigma \approx C_{k,n}.$$

Let $s'_i \subset \Sigma'$ denote the preimage $\phi^{-1}(s_i)$. Given a hyperbolic structure on Σ' , there is a unique way of extending this hyperbolic structure to Σ as follows. As usual, assume that each $s'_i \subset \Sigma'$ is a closed geodesic. Choose $\epsilon > 0$ sufficiently small so that all the ϵ -collars $N_\epsilon(s'_i)$ of s'_i are disjoint. Denote the complement of these collars by

$$\Sigma'' := \Sigma' \setminus \bigcup_{i=1}^k N_\epsilon(s'_i).$$

Represent the geodesic s'_i as the quotient $\tilde{s}/\langle \xi \rangle$, where $\xi \in \text{PSL}(2, \mathbb{R})$ is hyperbolic and $\tilde{s} \subset \mathbb{H}^2$ is the ξ -invariant geodesic. That is, ξ is a transvection along the geodesic \tilde{s} . Let $\sqrt{\xi}$ denote the unique glide-reflection whose square is ξ ; it is the composition of reflection in \tilde{s} with the transvection of displacement $\ell(\xi)/2$ where $\ell(\xi)$ is the displacement of ξ . If a matrix representative of ξ has trace $x > 2$, then $x = 2 \cosh(\ell(\xi)/2)$ and a matrix representing the glide-reflection $\sqrt{\xi}$ equals

$$\frac{1}{\sqrt{x - 2}}(\xi - \text{Id}).$$

Let $N_\epsilon(\tilde{s}) \subset \mathbb{H}^2$ be the tubular neighborhood of width ϵ about \tilde{s} . The quotient $N_\epsilon(\tilde{s})/\langle \sqrt{\xi} \rangle$ is a cross-cap bounded by a hypercycle (equidistant curve). The union

$$\Sigma'' \cup N_\epsilon(\tilde{s})/\langle \sqrt{\xi} \rangle$$

is the desired hyperbolic structure on Σ .

Fenchel–Nielsen coordinates on $\Sigma_{1,1}$. We relate the Fricke trace coordinates to Fenchel–Nielsen coordinates as follows. We suppose that the boundary $\partial\Sigma$ has length $b \geq 0$, the case $b = 0$ corresponding to the complete finite-area structure (where the holonomy around $\partial\Sigma$ is parabolic).

Suppose that $X \subset \Sigma$ has length $l > 0$, and has holonomy represented by

$$\tilde{\rho}(X) := \begin{bmatrix} e^{l/2} & 0 \\ 0 & e^{-l/2} \end{bmatrix},$$

where $x = 2 \cosh(l/2)$. Then

$$\text{Fix}(\rho(X)) = \{0, \infty\}.$$

A fundamental domain for the cyclic group $\langle \rho(X) \rangle$ is bounded by the geodesics with endpoints $\pm e^{-l/2}$ and $\pm e^{l/2}$ respectively.

Normalize the twist parameter τ so that $\tau = 0$ corresponds to the case that the invariant axes $\ell_{\rho(X)}$, $\ell_{\rho(Y)}$ are orthogonal. In that case take $\text{Fix}(\rho(Y)) = \pm 1$ and define

$$\tilde{\rho}_0(Y) := \begin{bmatrix} \cosh(\mu/2) & \sinh(\mu/2) \\ \sinh(\mu/2) & \cosh(\mu/2) \end{bmatrix},$$

where $y = 2 \cosh(\mu/2)$.

A fundamental domain for the cyclic group $\langle \rho(Y) \rangle$ is bounded by the geodesics with endpoints $-e^{\pm\mu/2}$ and $e^{\pm\mu/2}$ respectively. For this representation,

$$z = \text{tr}(\tilde{\rho}(X)\tilde{\rho}_0(Y)) = xy/2 = 2 \cosh(l/2) \cosh(\mu/2).$$

Let $\tau \in \mathbb{R}$ be the twist parameter for Fenchel–Nielsen flow. (Compare Wolpert [80].) The orbit of the Fenchel–Nielsen twist deformation is defined by the representation

$$\rho(Y) := \rho_0(Y) \exp\left((\tau/2) \widehat{\tilde{\rho}(X)}\right),$$

where

$$\widehat{\tilde{\rho}(X)} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

defines the one-parameter subgroup

$$\exp\left((\tau/2) \widehat{\tilde{\rho}(X)}\right) = \begin{bmatrix} e^{\tau/2} & 0 \\ 0 & e^{-\tau/2} \end{bmatrix}.$$

Now $x = 2 \cosh(l/2)$ is constant but

$$\begin{aligned} y = \text{tr}(\tilde{\rho}(Y)) &= \text{tr}\left(\tilde{\rho}_0(Y) \begin{bmatrix} e^{\tau/2} & 0 \\ 0 & e^{-\tau/2} \end{bmatrix}\right) \\ &= 2 \cosh(l/2) \cosh(\tau/2). \end{aligned}$$

Similarly,

$$z = \text{tr}(\tilde{\rho}(X)\tilde{\rho}(Y)) = \text{tr}\left(\begin{bmatrix} e^{l/2} & 0 \\ 0 & e^{-l/2} \end{bmatrix} \tilde{\rho}_0(Y) \begin{bmatrix} e^{\tau/2} & 0 \\ 0 & e^{-\tau/2} \end{bmatrix}\right) \\ = 2 \cosh(\mu/2) \cosh((l + \tau)/2).$$

Now the commutator trace $\text{tr}(\tilde{\rho}([X, Y]))$ equals

$$-2 \cosh(b/2) = \kappa(x, y, z) = 2 - \sinh^2(l/2) \sinh^2(\mu/2)$$

whence

$$\cosh^2(\mu/2) = 1 - 4 \text{csch}^2(l/2) \sinh^2(b/4).$$

Therefore the Fricke trace coordinates are expressed in terms of Fenchel–Nielsen coordinates by

$$x = 2 \cosh(l/2), \\ y = 2\sqrt{1 - 4 \text{csch}^2(l/2) \sinh^2(b/4)} \cosh(\tau/2), \\ z = 2\sqrt{1 - 4 \text{csch}^2(l/2) \sinh^2(b/4)} \cosh((\tau + l)/2).$$

4.6 The two-holed cross-surface

Following John H. Conway’s suggestion, we call a surface homeomorphic to a real projective plane a *cross-surface*. Suppose that $\Sigma = C_{0,2}$ is a two-holed cross-surface (Figure 4). Then $\pi_1(\Sigma)$ is freely generated by two orientation-reversing simple loops P, Q on the interior. These loops correspond to the two 1-handles in Figure 4. The two boundary components ∂_{\pm} of Σ correspond to elements

$$R := Q^{-1}P^{-1}, \quad R' := QR^{-1},$$

obtaining a redundant geometric presentation of $\pi_1(\Sigma)$:

$$\pi = \langle P, Q, R, R' \mid PQR = PQR^{-1}R' = \text{Id} \rangle.$$

The characters of the generators of this presentation define a presentation of the character ring

$$\mathbb{C}[f_P, f_Q, f_R, f_{R'}] / (f_R + f_{R'} - f_P f_Q)$$

the relation being (2.6). Of course p, q, r (respectively p, q, r') are free generators for the character ring (a polynomial ring in three variables).

The Fricke space of Σ was computed by Stantchev [72]; compare also the forthcoming paper by Goldman–McShane–Stantchev–Tan [33]. For a given hyperbolic structure on Σ , the holonomy transformations $\rho(P)$ and $\rho(Q)$ reverse orientation, their traces are purely imaginary, and the traces of $\rho(R)$ and $\rho(R')$ are real. For this

reason we write

$$\begin{aligned} ip &= f_P = \operatorname{tr}(\rho(P)) \in i\mathbb{R}, \\ iq &= f_Q = \operatorname{tr}(\rho(Q)) \in i\mathbb{R}, \\ r &= f_R = \operatorname{tr}(\rho(R)) \in \mathbb{R}, \\ r' &= f_{R'} = \operatorname{tr}(\rho(R')) \in \mathbb{R}, \end{aligned}$$

where $p, q, r, r' \in \mathbb{R}$ and

$$r' := r + pq \in \mathbb{R}.$$

By an analysis similar to that of $\Sigma_{0,3}$ and $\Sigma_{1,1}$, the Fricke space of Σ identifies with

$$\{(p, q, r) \in \mathbb{R}^3 \mid r \leq -2, pq + r \geq 2\}.$$

Compare [72], [35], [33] for further details.

4.7 The one-holed Klein bottle

Now suppose Σ is a one-holed Klein bottle. (Compare Figure 5.) Once again we choose free generators P, Q for π corresponding to the two 1-handles in Figure 5 which reverse orientation. The boundary component D corresponds to P^2Q^2 and, writing $R = (PQ)^{-1}$, we obtain a redundant geometric presentation

$$\pi = \langle P, Q, R, D \mid PQR = \operatorname{Id}, D = P^2Q^2 \rangle$$

and the character ring has presentation

$$\mathbb{C}[f_P, f_Q, f_R, f_D] / (f_D - (f_P f_Q f_R - f_P^2 - f_Q^2 + 4)).$$

Since P, Q reverse orientation on \mathbb{H}^2 , the functions f_P, f_Q are purely imaginary and f_R, f_D are real. Thus we write

$$\begin{aligned} ip &= f_P = \operatorname{tr}(\rho(P)) \in i\mathbb{R}, \\ iq &= f_Q = \operatorname{tr}(\rho(Q)) \in i\mathbb{R}, \\ r &= f_R = \operatorname{tr}(\rho(R)) \in \mathbb{R}, \\ d &= f_D = \operatorname{tr}(\rho(D)) \in \mathbb{R}, \end{aligned}$$

and the Fricke space of Σ identifies with

$$\{(p, q, r) \in \mathbb{R}^3 \mid p^2 + q^2 - pqr \geq 0\}.$$

See Stantchev [72] and [33] for details.

5 Three-generator groups and beyond

Let Σ be a compact connected surface-with-boundary. Suppose $\partial\Sigma \neq \emptyset$. Then the fundamental group $\pi_1(\Sigma)$ is free of rank 3 if and only if the Euler characteristic $\chi(\Sigma) = -2$. Such a surface is homeomorphic to one of the four topological types:

- a four-holed sphere $\Sigma_{0,4}$;
- a two-holed torus $\Sigma_{1,2}$;
- a three-holed cross-surface (projective planes) $C_{0,3}$;
- a two-holed Klein bottle $C_{1,2}$.

In this section we only consider the orientable topological types, namely $\Sigma_{0,4}$ and $\Sigma_{1,2}$. We relate their character varieties to those of nonorientable surfaces $C_{0,2}$ and $C_{1,1}$, discussed in §4.6 and §4.7.

5.1 The $SL(2, \mathbb{C})$ -character ring of \mathbb{F}_3

Representations ρ of the free group $\langle X_1, X_2, X_3 \rangle$ of rank three correspond to arbitrary triples

$$(\rho(X_1), \rho(X_2), \rho(X_3)) \in SL(2, \mathbb{C})^3.$$

As before we consider the quotient space (in the sense of Geometric Invariant Theory) under the action of $SL(2, \mathbb{C})$ by inner automorphisms, the *character variety*. Its coordinate ring is by definition the subring of invariants (the *character ring*)

$$\mathbb{C}[SL(2, \mathbb{C})^3]^{PSL(2, \mathbb{C})} \subset \mathbb{C}[SL(2, \mathbb{C})^3]$$

of the induced *effective* $PSL(2, \mathbb{C})$ -action on the ring of coordinate ring $\mathbb{C}[SL(2, \mathbb{C})^3]$.

We saw in §2 that for a free group of rank two, the character variety is an affine space and the character ring is a polynomial ring. The situation in rank three is more complicated. The character variety V_3 is a six-dimensional hypersurface in \mathbb{C}^7 , which admits a branched double covering onto the six-dimensional affine space \mathbb{C}^6 .

Explicitly, the character ring \mathfrak{R}_3 is generated by eight trace functions

$$t_1, t_2, t_3, t_{12}, t_{23}, t_{13}, t_{123}, t_{132}$$

defined by

$$t_i(\rho) := \text{tr}(\rho(X_i)), \quad t_{ij}(\rho) := \text{tr}(\rho(X_i X_j)) \quad \text{and} \quad t_{ijk}(\rho) := \text{tr}(\rho(X_i X_j X_k))$$

subject to two relations expressing the sum and product of traces of the length 3 monomials in terms of traces of monomials of length 1 and 2:

$$t_{123} + t_{132} = t_{12}t_3 + t_{13}t_2 + t_{23}t_1 - t_1 t_2 t_3, \tag{5.1}$$

$$t_{123} t_{132} = (t_1^2 + t_2^2 + t_3^2) + (t_{12}^2 + t_{23}^2 + t_{13}^2) - (t_1 t_2 t_{12} + t_2 t_3 t_{23} + t_3 t_1 t_{13}) + t_{12} t_{23} t_{13} - 4. \tag{5.2}$$

We call (5.1) the *Sum Relation* and (5.2) the *Product Relation* respectively. They imply that the *triple traces* t_{123} and t_{132} are the respective roots λ of the irreducible monic quadratic equation

$$\lambda^2 - f_\Sigma \lambda + f_\Pi = 0,$$

where the coefficients

$$f_\Sigma := t_{12}t_3 + t_{23}t_1 + t_{13}t_2 - t_1t_2t_3$$

and

$$\begin{aligned} f_\Pi := & (t_1^2 + t_2^2 + t_3^2) \\ & + (t_{12}^2 + t_{23}^2 + t_{13}^2) \\ & - (t_1t_2t_{12} + t_2t_3t_{23} + t_3t_1t_{13}) \\ & + t_{12}t_{23}t_{13} - 4 \end{aligned}$$

are the polynomials appearing in the right-hand sides of (5.1) and (5.2) respectively.

V_3 is a hypersurface in \mathbb{C}^7 . Eliminating t_{132} in (5.1) as

$$t_{132} = t_{12}t_3 + t_{13}t_2 + t_{23}t_1 - t_1t_2t_3 - t_{123}, \tag{5.3}$$

realizes V_3 as the hypersurface in \mathbb{C}^7 consisting of all

$$(t_1, t_2, t_3, t_{12}, t_{23}, t_{13}, t_{123}) \in \mathbb{C}^7$$

satisfying

$$\begin{aligned} & t_{123}(t_{12}t_3 + t_{13}t_2 + t_{23}t_1 - t_1t_2t_3 - t_{123}) \\ & = (t_1^2 + t_2^2 + t_3^2) + (t_{12}^2 + t_{23}^2 + t_{13}^2) - (t_1t_2t_{12} + t_2t_3t_{23} + t_3t_1t_{13}) + t_{12}t_{23}t_{13} - 4. \end{aligned}$$

V_3 double covers \mathbb{C}^6 . The double covering of V_3 over \mathbb{C}^6 arises from the composition

$$V_3 \hookrightarrow \mathbb{C}^8 \xrightarrow{\Pi} \mathbb{C}^6, \tag{5.4}$$

where

$$\mathbb{C}^8 \xrightarrow{\Pi} \mathbb{C}^6,$$

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_{12} \\ t_{23} \\ t_{13} \\ t_{123} \\ t_{132} \end{bmatrix} \longmapsto \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_{12} \\ t_{23} \\ t_{13} \end{bmatrix}$$

is the coordinate projection.

Proposition 5.1.1. *The composition (5.4)*

$$V \xrightarrow{t} \mathbb{C}^6,$$

$$[\rho] \mapsto \begin{bmatrix} t_1(\rho) \\ t_2(\rho) \\ t_3(\rho) \\ t_{12}(\rho) \\ t_{23}(\rho) \\ t_{13}(\rho) \end{bmatrix}$$

is onto. Furthermore it is a double covering branched along the discriminant hypersurface in \mathbb{C}^6 defined by

$$(t_{12}t_3 + t_{13}t_2 + t_{23}t_1 - t_1t_2t_3)^2 = 4((t_1^2 + t_2^2 + t_3^2) + (t_{12}^2 + t_{23}^2 + t_{13}^2) - (t_1t_2t_{12} + t_2t_3t_{23} + t_3t_1t_{13}) + t_{12}t_{23}t_{13} - 4).$$

The goal of this section is to prove the identities (5.1), (5.2) and Proposition 5.1.1.

Proof of the Sum Relation. To prove these identities, we temporarily introduce the following notation. Let $\xi = \rho(X_1)$, $\eta = \rho(X_2)$, $\zeta = \rho(X_3)$, so that (5.1) becomes

$$\begin{aligned} \operatorname{tr}(\xi\eta\zeta) + \operatorname{tr}(\xi\zeta\eta) &= \operatorname{tr}(\xi\eta)\operatorname{tr}(\zeta) - \operatorname{tr}(\zeta)\operatorname{tr}(\xi)\operatorname{tr}(\eta) \\ &\quad + \operatorname{tr}(\zeta\xi)\operatorname{tr}(\eta) + \operatorname{tr}(\eta\zeta)\operatorname{tr}(\xi). \end{aligned} \quad (5.5)$$

To prove (5.5), apply the Basic Identity (2.6) three times:

$$\operatorname{tr}(\xi\eta\zeta) + \operatorname{tr}(\xi\eta\zeta^{-1}) = \operatorname{tr}(\xi\eta)\operatorname{tr}(\zeta), \quad (5.6)$$

$$\begin{aligned} \operatorname{tr}(\zeta^{-1}\xi\eta) + \operatorname{tr}(\zeta^{-1}\xi\eta^{-1}) &= \operatorname{tr}(\zeta^{-1}\xi)\operatorname{tr}(\eta) \\ &= (\operatorname{tr}(\zeta)\operatorname{tr}(\xi) - \operatorname{tr}(\zeta\xi))\operatorname{tr}(\eta), \end{aligned} \quad (5.7)$$

$$\begin{aligned} \operatorname{tr}(\eta^{-1}\zeta^{-1}\xi) + \operatorname{tr}(\eta^{-1}\zeta^{-1}\xi^{-1}) &= \operatorname{tr}(\eta^{-1}\zeta^{-1})\operatorname{tr}(\xi) \\ &= \operatorname{tr}(\eta\zeta)\operatorname{tr}(\xi). \end{aligned} \quad (5.8)$$

Now add (5.6), subtract (5.7) and add (5.8) to obtain

$$\begin{aligned} (\operatorname{tr}(\xi\eta\zeta) + \operatorname{tr}(\xi\eta\zeta^{-1})) - (\operatorname{tr}(\zeta^{-1}\xi\eta) + \operatorname{tr}(\zeta^{-1}\xi\eta^{-1})) \\ + (\operatorname{tr}(\eta^{-1}\zeta^{-1}\xi) + \operatorname{tr}(\xi\zeta\eta)) \\ = \operatorname{tr}(\xi\eta)\operatorname{tr}(\zeta) - (\operatorname{tr}(\zeta)\operatorname{tr}(\xi) - \operatorname{tr}(\zeta\xi))\operatorname{tr}(\eta) + \operatorname{tr}(\eta\zeta)\operatorname{tr}(\xi). \end{aligned} \quad (5.9)$$

The right hand side of (5.9) is the right-hand side of (5.5). The left-hand side of (5.9) equals

$$\begin{aligned} \operatorname{tr}(\xi\eta\zeta) + (\operatorname{tr}(\xi\eta\zeta^{-1}) - \operatorname{tr}(\zeta^{-1}\xi\eta)) + (-\operatorname{tr}(\zeta^{-1}\xi\eta^{-1}) + \operatorname{tr}(\eta^{-1}\zeta^{-1}\xi)) + \operatorname{tr}(\xi\zeta\eta) \\ = \operatorname{tr}(\xi\eta\zeta) + \operatorname{tr}(\xi\zeta\eta), \end{aligned}$$

the left-hand side of (5.5), from which (5.5) follows.

Proof of the Product Relation. We derive this formula in several steps. Directly applying the Basic Identity (2.6):

$$\begin{aligned}
 \operatorname{tr}(\zeta \xi \zeta \eta) &= \operatorname{tr}(\zeta \xi) \operatorname{tr}(\zeta \eta) - \operatorname{tr}(\xi \eta^{-1}) \\
 &= \operatorname{tr}(\zeta \xi) \operatorname{tr}(\zeta \eta) - (\operatorname{tr}(\xi) \operatorname{tr}(\eta) - \operatorname{tr}(\xi \eta)) \\
 &= \operatorname{tr}(\zeta \xi) \operatorname{tr}(\zeta \eta) - \operatorname{tr}(\xi) \operatorname{tr}(\eta) + \operatorname{tr}(\xi \eta).
 \end{aligned} \tag{5.10}$$

Apply a calculation similar to (2.8) to ξ, ζ^{-1} :

$$\operatorname{tr}(\xi \zeta^{-1} \xi^{-1} \zeta^{-1}) = \operatorname{tr}(\xi) \operatorname{tr}(\zeta) \operatorname{tr}(\zeta \xi) - \operatorname{tr}(\zeta \xi)^2 - \operatorname{tr}(\xi)^2 + 2, \tag{5.11}$$

$$\begin{aligned}
 \operatorname{tr}(\xi \eta \zeta \xi \zeta \eta) &= \operatorname{tr}(\xi \eta) \operatorname{tr}(\zeta \xi \zeta \eta) - \operatorname{tr}(\xi \zeta^{-1} \xi^{-1} \zeta^{-1}) \\
 &= \operatorname{tr}(\xi \eta) (\operatorname{tr}(\zeta \xi) \operatorname{tr}(\zeta \eta) - \operatorname{tr}(\xi) \operatorname{tr}(\eta) + \operatorname{tr}(\xi \eta)) \\
 &\quad - (\operatorname{tr}(\xi) \operatorname{tr}(\zeta) \operatorname{tr}(\zeta \xi) - \operatorname{tr}(\zeta \xi)^2 - \operatorname{tr}(\xi)^2 + 2) \\
 &\quad \text{(by (5.10) and (5.11))} \\
 &= \operatorname{tr}(\xi \eta) \operatorname{tr}(\zeta \xi) \operatorname{tr}(\eta \zeta) \\
 &\quad - \operatorname{tr}(\xi) \operatorname{tr}(\eta) \operatorname{tr}(\xi \eta) - \operatorname{tr}(\zeta) \operatorname{tr}(\xi) \operatorname{tr}(\zeta \xi) \\
 &\quad + \operatorname{tr}(\xi \eta)^2 + \operatorname{tr}(\xi)^2 - 2.
 \end{aligned} \tag{5.12}$$

Finally, applying (5.12) and the Commutator Identity (2.9) to η, ζ :

$$\begin{aligned}
 \operatorname{tr}(\xi \eta \zeta) \operatorname{tr}(\xi \zeta \eta) &= \operatorname{tr}(\xi \eta \zeta \xi \zeta \eta) + \operatorname{tr}(\eta \zeta \eta^{-1} \zeta^{-1}) \\
 &= (\operatorname{tr}(\xi \eta) \operatorname{tr}(\zeta \xi) \operatorname{tr}(\eta \zeta) - \operatorname{tr}(\xi) \operatorname{tr}(\eta) \operatorname{tr}(\xi \eta) \\
 &\quad - \operatorname{tr}(\zeta) \operatorname{tr}(\xi) \operatorname{tr}(\zeta \xi) + \operatorname{tr}(\xi \eta)^2 + \operatorname{tr}(\zeta \xi)^2 + \operatorname{tr}(\xi)^2 - 2) \\
 &\quad + (\operatorname{tr}(\eta)^2 + \operatorname{tr}(\zeta)^2 + \operatorname{tr}(\eta \zeta)^2 - \operatorname{tr}(\eta) \operatorname{tr}(\zeta) \operatorname{tr}(\eta \zeta) - 2)
 \end{aligned}$$

which yields (5.2).

Proof that \mathbf{t} is onto. Now we prove Proposition 5.1.1. For a more general treatment see Florentino [19].

Theorem A guarantees $\xi_1, \xi_2 \in \operatorname{SL}(2, \mathbb{C})$ such that

$$\begin{aligned}
 \operatorname{tr}(\xi_1) &= t_1, \\
 \operatorname{tr}(\xi_2) &= t_2, \\
 \operatorname{tr}(\xi_1 \xi_2) &= t_{12}.
 \end{aligned} \tag{5.13}$$

We seek $\xi_3 \in \operatorname{SL}(2, \mathbb{C})$ such that

$$\begin{aligned}
 \operatorname{tr}(\xi_3) &= t_3, \\
 \operatorname{tr}(\xi_2 \xi_3) &= t_{23}, \\
 \operatorname{tr}(\xi_1 \xi_3) &= t_{13}.
 \end{aligned} \tag{5.14}$$

To this end, consider the affine subspace \mathcal{W} of $M_2(\mathbb{C})$ consisting of matrices ω satisfying

$$\begin{aligned}\operatorname{tr}(\omega) &= t_3, \\ \operatorname{tr}(\xi_2\omega) &= t_{23}, \\ \operatorname{tr}(\xi_1\omega) &= t_{13}.\end{aligned}\tag{5.15}$$

Since the bilinear pairing

$$\begin{aligned}M_2(\mathbb{C}) \times M_2(\mathbb{C}) &\longrightarrow \mathbb{C}, \\ (\xi, \eta) &\longmapsto \operatorname{tr}(\xi\eta)\end{aligned}$$

is nondegenerate, each of the three equations in (5.15) describes an affine hyperplane in $M_2(\mathbb{C})$. We first suppose that (t_1, t_2, t_{12}) describes an irreducible character, that is $\kappa(t_1, t_2, t_{12}) \neq 2$:

$$4 - t_1^2 - t_2^2 - t_{12}^2 + t_1 t_2 t_{12} \neq 0.\tag{5.16}$$

Our goal will be to find an element $\xi_3 \in \mathcal{W}$ such that $\det(\xi_3) = 1$.

Lemma 5.1.2. *There exist $\xi_1, \xi_2 \in \operatorname{SL}(2, \mathbb{C})$ satisfying (5.13) such that \mathcal{W} is a (nonempty) affine line.*

Proof. Since $\kappa(t_1, t_2, t_{12}) \neq 2$, Proposition 2.3.1, (2) implies that the pair ξ_1, ξ_2 generates an irreducible representation.

We claim that $\{\operatorname{Id}, \xi_1, \xi_2\}$ is a linearly independent subset of the 4-dimensional vector space $M_2(\mathbb{C})$. Otherwise the nonzero element ξ_1 is a linear combination of ξ_2 and Id . Let $v \neq 0$ be an eigenvector of ξ_2 . Then the line (v) spanned by v is invariant under ξ_1 as well, and hence under the group generated by ξ_1 and ξ_2 . This contradicts irreducibility of the representation generated by ξ_1 and ξ_2 .

Since $\{\operatorname{Id}, \xi_1, \xi_2\}$ is linearly independent, the three linear conditions of (5.15) are independent. Hence $\mathcal{W} \subset M_2(\mathbb{C})$ is an affine line. \square

Let $\omega_0, \omega_1 \in \mathcal{W}$ be distinct elements in this line. Then the function

$$\begin{aligned}\mathbb{C} &\longrightarrow \mathbb{C}, \\ s &\longmapsto \det(s\omega_1 + (1-s)\omega_0)\end{aligned}$$

is a polynomial of degree ≤ 2 , and is thus onto unless it is constant.

We shall show that this map is onto, and therefore $\mathcal{W} \cap \operatorname{SL}(2, \mathbb{C}) \neq \emptyset$. The desired matrix ξ_3 will be an element of $\mathcal{W} \cap \operatorname{SL}(2, \mathbb{C})$.

Lemma 5.1.3. *Let $\omega_0, \omega_1 \in M_2(\mathbb{C})$. Then*

$$\begin{aligned}\det(s\omega_1 + (1-s)\omega_0) &= \det(\omega_0 + s(\omega_1 - \omega_0)) \\ &= \det(\omega_0) + s(\operatorname{tr}(\omega_0)\operatorname{tr}(\omega_1 - \omega_0) - \operatorname{tr}(\omega_0(\omega_1 - \omega_0))) \\ &\quad + s^2\det(\omega_1 - \omega_0).\end{aligned}$$

Proof. Clearly

$$\mathrm{tr}(\omega_0 + s(\omega_1 - \omega_0)) = \mathrm{tr}(\omega_0) + s\mathrm{tr}(\omega_1 - \omega_0). \quad (5.17)$$

Now

$$\det(\omega) = \frac{\mathrm{tr}(\omega)^2 - \mathrm{tr}(\omega^2)}{2} \quad (5.18)$$

whenever $\omega \in M_2(\mathbb{C})$. Now apply (5.18) to (5.17) taking

$$\omega = \omega_0 + s(\omega_1 - \omega_0). \quad \square$$

Thus the restriction $\det|_{\mathcal{W}}$ is constant only if $\det(\omega_1 - \omega_0) = 0$.

Choose a solution ξ of $\xi + \xi^{-1} = t_{12}$. Work in the slice

$$\xi_1 = \begin{bmatrix} t_1 & -1 \\ 1 & 0 \end{bmatrix}, \quad \xi_2 = \begin{bmatrix} 0 & \xi \\ -\xi^{-1} & t_2 \end{bmatrix}.$$

The matrix $\omega_0 \in M_2(\mathbb{C})$ defined by

$$\omega_0 = \begin{bmatrix} t_3 & ((t_{13} - t_1 t_3)\xi + t_{23})\xi / (\xi^2 - 1) \\ ((t_{13} - t_1 t_3) + t_{23}\xi) / (\xi^2 - 1) & 0 \end{bmatrix}$$

satisfies (5.15). Any other $\omega \in \mathcal{W}$ must satisfy

$$\begin{aligned} \mathrm{tr}(\omega - \omega_0) &= 0, \\ \mathrm{tr}(\xi_2(\omega - \omega_0)) &= 0, \\ \mathrm{tr}(\xi_1(\omega - \omega_0)) &= 0. \end{aligned} \quad (5.19)$$

Lemma 5.1.4. *Any solution $\omega - \omega_0$ of (5.19) is a multiple of*

$$\mathrm{Lie}(\xi_1, \xi_2) = \xi_1 \xi_2 - \xi_2 \xi_1 = \begin{bmatrix} \xi^{-1} - \xi & -t_2 + t_1 \xi \\ -t_2 + \xi^{-1} t_1 & \xi - \xi^{-1} \end{bmatrix}.$$

Proof. The first equation in (5.19) asserts that $\omega - \omega_0$ lies in the subspace $\mathfrak{sl}(2)$, upon which the trace form is nondegenerate. The second and third equations assert that $\omega - \omega_0$ is orthogonal to ξ_1 and ξ_2 . By (5.16), ξ_1, ξ_2 and Id are linearly independent in $M_2(\mathbb{C})$, so the solutions of (5.19) form a one-dimensional linear subspace. The Lie product

$$\mathrm{Lie}(\xi_1, \xi_2) = \xi_1 \xi_2 - \xi_2 \xi_1$$

is nonzero and lies in $\mathfrak{sl}(2)$. Furthermore, for $i = 1, 2$,

$$\mathrm{tr}(\xi_i \xi_1 \xi_2) = \mathrm{tr}(\xi_i \xi_2 \xi_1)$$

implies that $\mathrm{Lie}(\xi_1, \xi_2)$ is orthogonal to ξ_1 and ξ_2 . The lemma follows. \square

Parametrize \mathcal{W} explicitly as $\omega = \omega_0 + s \mathrm{Lie}(\xi_1, \xi_2)$. By (2.15),

$$\det(\mathrm{Lie}(\xi_1, \xi_2)) = 4 - (t_1^2 + t_2^2 + t_{12}^2 - t_1 t_2 t_{12}) = 2 - \kappa(t_1, t_2, t_{12}) \neq 0.$$

By (5.16), the polynomial

$$\mathcal{W} \xrightarrow{\det} \mathbb{C}$$

is nonconstant, and hence onto. Taking

$$\omega_1 \in (\det|_{\mathcal{W}}^{-1})(1),$$

the proof of Proposition 5.1.1 is complete assuming (5.16).

The case when $4 - t_1^2 - t_2^2 - t_{12}^2 + t_1 t_2 t_{12} = 0$ remains. Then

$$t_i = a_i + (a_i)^{-1}$$

for $i = 1, 2$, for some $a_1, a_2 \in \mathbb{C}^*$. Then either

$$t_{12} = a_1 a_2 + (a_1 a_2)^{-1} \tag{5.20}$$

or

$$t_{12} = a_1 (a_2)^{-1} + (a_1)^{-1} a_2. \tag{5.21}$$

In the first case (5.20), set

$$\begin{aligned} \xi_1 &:= \begin{bmatrix} a_1 & t_{13} - a_1 t_3 \\ 0 & (a_1)^{-1} \end{bmatrix}, \\ \xi_2 &:= \begin{bmatrix} a_2 & t_{23} - a_2 t_3 \\ 0 & (a_2)^{-1} \end{bmatrix}, \\ \xi_3 &:= \begin{bmatrix} t_3 & -1 \\ 1 & 0 \end{bmatrix}, \end{aligned}$$

and in the second case (5.21), set

$$\begin{aligned} \xi_1 &:= \begin{bmatrix} (a_1)^{-1} & t_{13} - (a_1)^{-1} t_3 \\ 0 & a_1 \end{bmatrix}, \\ \xi_2 &:= \begin{bmatrix} a_2 & t_{23} - a_2 t_3 \\ 0 & (a_2)^{-1} \end{bmatrix}, \\ \xi_3 &:= \begin{bmatrix} t_3 & -1 \\ 1 & 0 \end{bmatrix} \end{aligned}$$

obtaining $(\xi_1, \xi_2, \xi_3) \in \mathrm{SL}(2, \mathbb{C})^3$ explicitly solving (5.13) and (5.14). The proof of Proposition 5.1.1 is complete.

5.2 The four-holed sphere

Let $\Sigma \approx \Sigma_{0,4}$ be the four-holed sphere, with boundary components A, B, C, D subject to the relation

$$ABCD = \mathrm{Id}.$$

The fundamental group is freely generated by

$$A = X_1, \quad B = X_2, \quad C = X_3$$

which represent three of the boundary components. The fourth boundary component is represented by an element

$$D := (X_1 X_2 X_3)^{-1},$$

satisfying the relation

$$ABCD = \text{Id}.$$

The resulting redundant presentation of the free group is

$$\pi = \langle A, B, C, D \mid ABCD = \text{Id} \rangle.$$

The elements

$$X := X_1 X_2, \quad Y := X_2 X_3, \quad Z := X_1 X_3$$

correspond to simple loops on Σ separating Σ into two three-holed spheres. (Compare Figure 8.) The (even more redundant) presentation

$$\pi = \langle A, B, C, D, X, Y, Z \mid ABCD = \text{Id}, X = AB, Y = BC, Z = CA \rangle$$

gives regular functions

$$a = t_1,$$

$$b = t_2,$$

$$c = t_3,$$

$$x = t_{12},$$

$$y = t_{23},$$

$$z = t_{13},$$

$$d = t_{123}$$

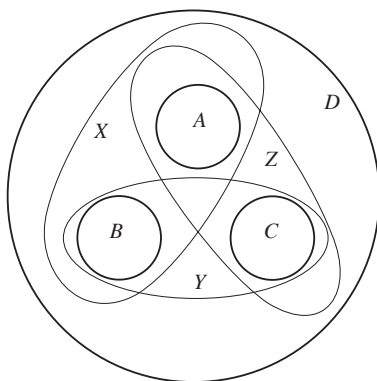


Figure 8. Seven simple curves on $\Sigma_{0,4}$.

generating the character ring. Using (5.1) to eliminate t_{132} as in (5.3), the product relation (5.2) implies

$$\begin{aligned} x^2 + y^2 + z^2 + xyz &= (ab + cd)x \\ &\quad + (ad + bc)y \\ &\quad + (ac + bd)z \\ &\quad + (4 - a^2 - b^2 - c^2 - d^2 - abcd). \end{aligned} \tag{5.22}$$

This leads to a presentation of the character ring as a quotient of the polynomial ring $\mathbb{C}[a, b, c, d, x, y, z]$ by the principal ideal (Φ) generated by

$$\begin{aligned} \Phi(a, b, c, d; x, y, z) &= x^2 + y^2 + z^2 + xyz \\ &\quad - (ab + cd)x - (ad + bc)y - (ac + bd)z \\ &\quad + a^2 + b^2 + c^2 + d^2 + abcd - 4. \end{aligned} \tag{5.23}$$

Thus the $\mathrm{SL}(2, \mathbb{C})$ -character variety is a quartic hypersurface in \mathbb{C}^7 , and for fixed boundary traces $(a, b, c, d) \in \mathbb{C}^4$, the relative $\mathrm{SL}(2, \mathbb{C})$ -character variety is the cubic surface in \mathbb{C}^3 defined by (5.22), as was known to Fricke and Vogt. (Compare Benedetto–Goldman [2], Goldman [29], Goldman–Neumann [34], Cantat–Loray [11] and Cantat [10], Iwasaki [41].)

The Fricke space of $\Sigma_{0,4}$. We identify the Fricke space $\mathfrak{F}(\Sigma_{0,4})$ in terms of trace coordinates.

Theorem 5.2.1. *The Fricke space of a four-holed sphere with boundary traces $a, b, c, d > 2$ is defined by the following conditions in $\mathbb{R}^4 \times \mathbb{R}^3$ for $(a, b, c, d) \in \mathbb{R}^4$ and $(x, y, z) \in \mathbb{R}^3$:*

$$\begin{aligned} a, b, c, d &\geq 2 \quad \text{and} \quad x < -2, \\ F^-, F^+ &> 0, \\ F^- F^+ &= \frac{(x^2 + a^2 + b^2 - abx - 4)(x^2 + c^2 + d^2 - cdx - 4)}{x^2 - 4}, \end{aligned}$$

where

$$\begin{aligned} F^- &= \sqrt{2-x} \left(y - z - \frac{(a-b)(d-c)}{2-x} \right) - \sqrt{-2-x} \left(y + z - \frac{(a+b)(d+c)}{-2-x} \right), \\ F^+ &= \sqrt{2-x} \left(y - z - \frac{(a-b)(d-c)}{2-x} \right) + \sqrt{-2-x} \left(y + z - \frac{(a+b)(d+c)}{-2-x} \right). \end{aligned}$$

Proof. For a given hyperbolic structure on Σ , the holonomy generators $\rho(A), \rho(B), \rho(C), \rho(D) \in \mathrm{PSL}(2, \mathbb{R})$ are hyperbolic or parabolic. Choose lifts

$$\tilde{\rho}(A), \tilde{\rho}(B), \tilde{\rho}(C), \tilde{\rho}(D) \in \mathrm{SL}(2, \mathbb{R})$$

which have positive trace. Since ρ is a representation $\pi_1(\Sigma) \rightarrow \mathrm{PSL}(2, \mathbb{R})$,

$$\tilde{\rho}(A)\tilde{\rho}(B)\tilde{\rho}(C)\tilde{\rho}(D) = \pm \mathrm{Id}.$$

We claim that $\tilde{\rho}(A)\tilde{\rho}(B)\tilde{\rho}(C)\tilde{\rho}(D) = \mathrm{Id}$. Since

$$\begin{aligned} \mathrm{tr}(\tilde{\rho}(A)) &\geq 2, \\ \mathrm{tr}(\tilde{\rho}(B)) &\geq 2, \\ \mathrm{tr}(\tilde{\rho}(C)) &\geq 2, \\ \mathrm{tr}(\tilde{\rho}(D)) &\geq 2, \end{aligned}$$

each of $\tilde{\rho}(A)$, $\tilde{\rho}(B)$, $\tilde{\rho}(C)$ and $\tilde{\rho}(D)$ lies in a unique one-parameter subgroup of $\mathrm{SL}(2, \mathbb{R})$. The corresponding embeddings define trivializations of the corresponding flat $\mathrm{PSL}(2, \mathbb{R})$ -bundle over each component of $\partial\Sigma$ as in Goldman [23], [27]. Namely, since each component $\partial_i(\Sigma)$ is a closed 1-manifold, lifting a homeomorphism $\mathbb{R}/\mathbb{Z} \rightarrow \partial_i(\Sigma)$ to

$$\mathbb{R} \longrightarrow \widetilde{\partial_i(\Sigma)}$$

the flat bundle over $\partial_i(\Sigma)$ with holonomy γ_i lifts to the quotient of the trivial principal $\mathrm{PSL}(2, \mathbb{R})$ -bundle $\mathbb{R} \times \mathrm{PSL}(2, \mathbb{R})$ by the \mathbb{Z} -action generated by

$$(t, g) \longmapsto (t + 1, \gamma_i g).$$

The corresponding trivialization is covered by the \mathbb{Z} -equivariant isomorphism

$$(t, g) \longmapsto (t, \exp(-t \log(\gamma_i)g)),$$

where

$$\{\exp(t \log(\gamma_i))\}_{t \in \mathbb{R}}$$

is the unique one-parameter subgroup of $\mathrm{PSL}(2, \mathbb{R})$ containing γ_i as above.

Since $\chi(\Sigma) = -2$, the Euler class of the representation ρ equals -2 and is even. The obstruction to lifting a representation to the double covering space

$$\mathrm{SL}(2, \mathbb{R}) \longrightarrow \mathrm{PSL}(2, \mathbb{R})$$

is the second Stiefel–Whitney class, which is the reduction of the Euler class modulo 2. Therefore $\tilde{\rho}$ defines a representation and $\tilde{\rho}(A)\tilde{\rho}(B)\tilde{\rho}(C)\tilde{\rho}(D) = \mathrm{Id}$ as claimed.

Furthermore, if ρ is a Fuchsian representation, then X is represented by a unique closed geodesic on Σ and

$$\rho(X) = \rho(A)\rho(B)$$

is hyperbolic. The relative Euler classes of the restriction of ρ to the subsurfaces complementary to X sum to ± 2 . Since they are constrained to equal $-1, 0, +1$, they both must be equal to $+1$ or both equal to -1 . (Compare [23], [27].) It follows that the trace $x = \mathrm{tr}(X) < -2$.

We study (5.22) using the following identity:

$$\begin{aligned}
 & 4(4 - x^2)\{x^2 + y^2 + z^2 + xyz - ((ab + cd)x + (ad + bc)y + (ac + bd)z) \\
 & \quad + (a^2 + b^2 + c^2 + d^2 + abcd - 4)\} \\
 & = (2 + x)\{(y - z)(2 - x) + (a - b)(c - d)\}^2 \\
 & \quad + (2 - x)\{(y + z)(2 + x) - (a + b)(c + d)\}^2 - 4\kappa_{a,b}(x)\kappa_{c,d}(x),
 \end{aligned} \tag{5.24}$$

where

$$\kappa_{p,q}(x) := x^2 + p^2 + q^2 - pqx - 4. \tag{5.25}$$

This function equals $\kappa(p, q, x) - 2$, where κ is the commutator trace function defined in (2.9).

When $x \neq \pm 2$, rewrite (5.22) using (5.24) as follows:

$$\begin{aligned}
 & \frac{2 + x}{4} \left((y + z) - \frac{(a + b)(d + c)}{2 + x} \right)^2 + \frac{2 - x}{4} \left((y - z) - \frac{(a - b)(d - c)}{2 - x} \right)^2 \\
 & = \frac{\kappa_{a,b}(x)\kappa_{c,d}(x)}{4 - x^2}.
 \end{aligned} \tag{5.26}$$

(Compare (3-3) of Benedetto–Goldman [2].)

We fix $a, b, c, d \geq 2$. As x varies, (5.26) defines a family of conics parametrized by x . For $x < -2$, this conic is a hyperbola, denoted $H_{a,b,c,d;x}$. The solutions of (5.26) for $a, b, c, d \geq 2$ and $x < -2$ fall into two connected components corresponding to the two components of the hyperbolas.

We explicitly describe these components. First observe that if $a, b \geq 2$ and $x < -2$, then

$$\kappa_{a,b}(x) > 16, \quad \kappa_{c,d}(x) > 16, \quad 4 - x^2 < 0,$$

so the left-hand side (5.26) is negative. For notational simplicity denote its opposite by $k = \kappa_{a,b,c,d;x}$:

$$k = \kappa_{a,b,c,d;x} := \frac{\kappa_{a,b}(x)\kappa_{c,d}(x)}{x^2 - 4} > 0.$$

Rewrite (5.26) as

$$\begin{aligned}
 & \frac{2 - x}{4} \left((y - z) - \frac{(a - b)(d - c)}{2 - x} \right)^2 \\
 & \quad - \frac{-2 - x}{4} \left((y + z) - \frac{(a + b)(d + c)}{2 + x} \right)^2 = k.
 \end{aligned}$$

Factoring the left-hand side of this equation, rewrite (5.26) as:

$$F^+(y, z) F^-(y, z) = k, \tag{5.27}$$

where the functions $F^\pm(y, z)$ are defined as

$$F^-(y, z) = F_{a,b,c,d;x}^-(y, z) := \sqrt{2-x} \left(y - z - \frac{(a-b)(d-c)}{2-x} \right) - \sqrt{-2-x} \left(y + z - \frac{(a+b)(d+c)}{-2-x} \right), \quad (5.28)$$

$$F^+(y, z) = F_{a,b,c,d;x}^+(y, z) := \sqrt{2-x} \left(y - z - \frac{(a-b)(d-c)}{2-x} \right) + \sqrt{-2-x} \left(y + z - \frac{(a+b)(d+c)}{-2-x} \right). \quad (5.29)$$

For fixed $a, b, c, d \geq 2$ and $x < -2$, the functions $F^\pm(y, z)$ are affine functions of y, z .

We identify each of the two components of the hyperbola $H_{a,b,c,d;x}$ defined by (5.27). One component, denoted $H_{a,b,c,d;x}^+$, is the intersection of $H_{a,b,c,d;x}$ with the open half plane

$$F_{a,b,c,d;x}^-(y, z) > 0.$$

Equivalently, $H_{a,b,c,d;x}^+$ is the intersection of $H_{a,b,c,d;x}$ with the open half-plane

$$F_{a,b,c,d;x}^+(y, z) > 0.$$

Similarly the other component $H_{a,b,c,d;x}^-$ is the intersection of $H_{a,b,c,d;x}$ with the open half-plane

$$F_{a,b,c,d;x}^-(y, z) < 0,$$

or, equivalently,

$$F_{a,b,c,d;x}^+(y, z) < 0.$$

The union of these hyperbola components correspond to values of the relative Euler class (compare [27]) as follows. Either

$$H^+ := \bigcup_{a,b,c,d \geq 2, x < -2} H_{a,b,c,d;x}^+$$

or

$$H^- := \bigcup_{a,b,c,d \geq 2, x < -2} H_{a,b,c,d;x}^-$$

corresponds to characters of representations with relative Euler class 0. The other component corresponds to representations with relative Euler class ± 1 . There is no way to distinguish between relative Euler class $+1$ and -1 since the characters are equivalence classes under the group $\mathrm{PGL}(2, \mathbb{R})$, which does not preserve orientation. To determine which one is which, it suffices to check one single example and use continuity of the integer-valued relative Euler class.

Here is an example whose relative Euler class is zero. Choose

$$\rho(A) = \rho(D)^{-1}, \quad \rho(B) = \rho(C)^{-1}$$

so that the relation

$$\rho(A)\rho(B)\rho(C)\rho(D) = \text{Id}$$

is trivially satisfied. Clearly such a representation depends only on the pair $\rho(A), \rho(B)$ which is arbitrary. Furthermore we restrict the boundary traces to satisfy

$$a = \text{tr}(\rho(A)) \geq 2, \quad b = \text{tr}(\rho(B)) \geq 2.$$

This space is connected and contains the character of the trivial representation, whose relative Euler class is zero. Now consider the specific example

$$\rho(A) := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \rho(B) := \begin{bmatrix} 1 & 0 \\ x-2 & 1 \end{bmatrix},$$

where $x < -2$ is arbitrary. Then $a = b = c = d = 2$ and

$$y = 2, \quad z = 4 - x.$$

In particular $y - z < 0$ and $y + z > 0$. and therefore (5.28) implies

$$F_{a,b,c,d;x}^{\pm} < 0$$

proving that this representation has a character in H^- , proving Theorem 5.2.1. \square

5.3 The two-holed torus

The two-holed torus admits a redundant geometric presentation corresponding to the ribbon graph depicted in Figure 9:

$$\langle A, B, U, X, Y \mid A = UXY, B = UYX \rangle.$$

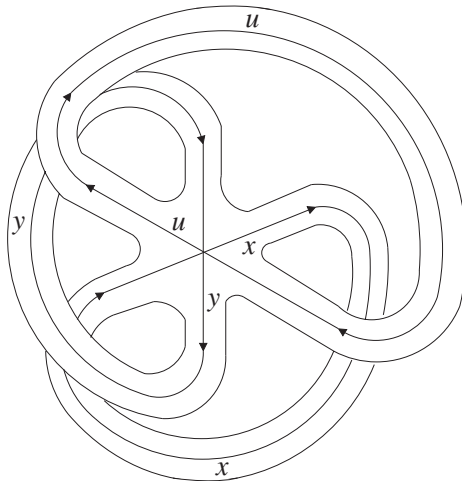


Figure 9. A ribbon graph representing a two-holed torus.

The 1-handles correspond to free generators U, X, Y and the boundary components correspond to the triple products

$$\begin{aligned} A &= UXY, \\ B &= UYX. \end{aligned}$$

Since the curves corresponding to X, Y, U in Figure 9 intersect transversely at the basepoint, the double products

$$\begin{aligned} V &= UX, \\ W &= UY, \\ Z &= XY \end{aligned}$$

are represented by simple loops as well. The Sum Relation (5.1) and the Product Relation (5.2) imply that the relative character variety for $\Sigma_{1,2}$ is defined by

$$\begin{aligned} a + b &= yv + xw + zu - uxy \\ ab &= x^2 + y^2 + u^2 + v^2 + w^2 + z^2 - xyz - yuw - uxv + vwz - 4. \end{aligned} \quad (5.30)$$

Button [9] gives defining inequalities for the Fricke space, where the boundary components are mapped to parabolics, as follows. First consider the \mathbb{R} -locus of the character variety, defined as the set of all $(a, b, u, x, y, z, v, w) \in \mathbb{R}^8$ satisfying (5.30) above, and

$$a = b = 2.$$

The regular neighborhood of the union of the loops corresponding to a pair of X, Y, U is an embedded one-holed torus. For example corresponding to the pair X, Y is a one-holed torus whose boundary corresponds to the commutator $[X, Y]$, and cuts Σ into the one-holed torus and a three-holed sphere. This commutator has trace $\kappa(x, y, z)$. Similarly the pair Y, U determines a separating curve whose corresponding trace function is $\kappa(y, u, w)$ and the pair U, X determines a separating curve whose corresponding trace function is $\kappa(u, x, v)$. The preceding discussions of the Fricke spaces of the one-holed torus and the three-holed sphere imply

$$\kappa(x, y, z) < -2, \quad \kappa(y, u, w) < -2, \quad \kappa(u, x, v) < -2.$$

Button [9] shows that these necessary conditions are sufficient, thus obtaining an explicit description of the Fricke–Teichmüller space of $\Sigma_{1,2}$ in terms of traces. (The reader should draw these curves on the ribbon graph depicted in Figure 9.)

5.4 Orientable double covering spaces

Let Σ be a nonorientable surface of $\chi(\Sigma) = -1$ and $\hat{\Sigma} \xrightarrow{\hat{\Pi}} \Sigma$ be its orientable covering space. There are two cases:

- $\Sigma \approx C_{0,2}$ and $\hat{\Sigma} \approx \Sigma_{0,4}$;

- $\Sigma \approx C_{1,1}$ and $\hat{\Sigma} \approx \Sigma_{1,2}$.

Then $\pi_1(\Sigma) \cong \mathbb{F}_2$ and $\pi_1(\hat{\Sigma}) \cong \mathbb{F}_3$. Denote a set of free generators of $\pi_1(\Sigma)$ by X_1, X_2 which correspond to orientation-reversing loops on Σ . The image of

$$\pi_1(\hat{\Sigma}) \xrightarrow{\hat{\Pi}_*} \pi_1(\Sigma) = \langle X_1, X_2 \rangle$$

equals the kernel of the homomorphism

$$\begin{aligned} \pi_1(\Sigma) &\longrightarrow \{\pm 1\}, \\ X_1 &\longmapsto -1, \\ X_2 &\longmapsto -1, \end{aligned}$$

which is freely generated by, for example,

$$\begin{aligned} Y_1 &= X_1^2, \\ Y_2 &= X_1^{-1} X_2^{-1}, \\ Y_3 &= X_2^2. \end{aligned}$$

The nontrivial deck transformation of $\hat{\Sigma} \xrightarrow{\hat{\Pi}} \Sigma$ is induced by the restriction of the inner automorphism $\text{Inn}(X_1)$ to $(\hat{\Pi})_*(\pi_1(\Sigma))$:

$$\begin{aligned} Y_1 &\longmapsto Y_1, \\ Y_2 &\longmapsto Y_3^{-1} Y_2^{-1} Y_1^{-1}, \\ Y_3 &\longmapsto Y_1 Y_2 Y_3 Y_2^{-1} Y_1^{-1}. \end{aligned}$$

The character ring of $\pi_1(\Sigma) \cong \mathbb{F}_2$ is the polynomial ring $\mathbb{C}[x_1, x_2, x_{12}]$. The character ring of $\pi_1(\hat{\Sigma}) \cong \mathbb{F}_3$ is the quotient

$$\mathbb{C}[y_1, y_2, y_3, y_{123}, y_{12}, y_{23}, y_{13}]/(\mathcal{J}),$$

where (\mathcal{J}) is the principal ideal generated by

$$\begin{aligned} &\Phi(y_1, y_2, y_3, y_{123}, y_{12}, y_{23}, y_{13}) \\ &= y_{12}^2 + y_{13}^2 + y_{23}^2 + y_{12}y_{13}y_{23} \\ &\quad - (y_1y_2 + y_3y_{123})y_{12} - (y_1y_{123} + y_2y_3)y_{23} \\ &\quad - (y_1y_3 + y_2y_{123})y_{13} + y_1^2 + y_2^2 + y_3^2 + y_{123}^2 + y_1y_2y_3y_{123} - 4 \end{aligned} \tag{5.31}$$

with Φ the polynomial (5.23). The automorphism

$$\text{Inn}(X_1)|_{(\hat{\Pi})_*(\pi_1(\Sigma))}$$

corresponding to the deck transformation induces the involution of character rings:

$$\begin{aligned} \mathfrak{R}_3 &\longleftrightarrow \mathfrak{R}_3, \\ y_1 &\longleftrightarrow y_1, \\ y_2 &\longleftrightarrow y_{123}, \\ y_3 &\longleftrightarrow y_3, \\ y_{12} &\longleftrightarrow y_{23}, \\ y_{13} &\longleftrightarrow y_1 y_3 - y_{13} - y_{12} y_{23} + y_{123} y_2, \\ y_{23} &\longleftrightarrow y_{12}, \\ y_{123} &\longleftrightarrow y_2. \end{aligned}$$

The covering space $\widehat{\Pi}$ induces the embedding of character rings:

$$\begin{aligned} \mathfrak{R}_3 &\hookrightarrow \mathfrak{R}_2, \\ y_1 &\longmapsto x_1^2 - 2, \\ y_2 &\longmapsto x_{12}, \\ y_3 &\longmapsto x_2^2 - 2, \\ y_{12} &\longmapsto x_1 x_2 - x_{12}, \\ y_{13} &\longmapsto x_1 x_2 x_{12} - x_1^2 - x_2^2 + 2, \\ y_{23} &\longmapsto x_1 x_2 - x_{12}, \\ y_{123} &\longmapsto x_{12}. \end{aligned}$$

The two topological types for Σ differ by their choice of *peripheral structure*:

- $\Sigma \approx C_{0,2}$ has two boundary components corresponding to

$$\delta_1 := Y_2 = X_1^{-1} X_2^{-1}, \quad \delta_2 = Y_1 Y_2 = X_1 X_2^{-1}.$$

- $\Sigma \approx C_{1,1}$ has one boundary component corresponding to

$$\delta := Y_1 Y_3 = X_1^2 X_2^2.$$

5.5 The two-holed cross-surface

The fundamental group of the two-holed cross-surface $C_{0,2}$ is free of rank two, with presentation

$$\pi_1(C_{0,2}) := \langle U, V, W, W' \mid W = UV, W' = V^{-1}U \rangle \cong \mathbb{F}_2.$$

The free generators U, V correspond to orientation-reversing simple curves on $C_{0,2}$ and W, W' correspond to the components of $\widehat{\partial C_{0,2}}$.

The orientable double covering-space $\widehat{C_{0,2}} \rightarrow C_{0,2}$ is connected, has four boundary components (since $C_{0,2}$ has two boundary components, each of which is orientable)

and has Euler characteristic $-2 = 2\chi(C_{0,2})$. Therefore $\widehat{C}_{0,2} \approx \Sigma_{0,4}$, the four-holed sphere, and has presentation

$$\pi = \langle A, B, C, D \mid ABCD = \text{Id} \rangle.$$

The corresponding monomorphism of fundamental groups is

$$\begin{aligned} \pi_1(\Sigma_{0,4}) &\hookrightarrow \pi_1(C_{0,2}), \\ A &\mapsto W = UV, \\ B &\mapsto W' = V^{-1}U, \\ C &\mapsto \text{Inn}(U^{-1})(W')^{-1} = U^{-2}VU, \\ D &\mapsto \text{Inn}(U^{-1})(W)^{-1} = U^{-1}V^{-1}. \end{aligned}$$

The character ring of $\pi_1(C_{0,2})$ is the polynomial ring $\mathfrak{R}_2 \cong \mathbb{C}[u, v, w]$ and the character ring of $\pi_1(\Sigma_{0,4})$ is the quotient of $\mathbb{C}[a, b, c, d, x, y, z]$ by the relation defined by (5.22). The induced homomorphism of character rings is

$$\begin{aligned} \mathfrak{R}_3 &\longrightarrow \mathfrak{R}_2 = \mathbb{C}[u, v, w], \\ a &\mapsto w, \\ b &\mapsto uv - w = w', \\ c &\mapsto uv - w = w', \\ d &\mapsto w, \\ x &\mapsto u^2 - 2, \\ y &\mapsto u^2 + v^2 + w^2 - uvw - 2, \\ z &\mapsto v^2 - u^2(u^2 + v^2 + w^2 - uvw - 2) - 2, \end{aligned}$$

evidently satisfying the defining equation (5.22) for the character variety of $\Sigma_{0,4}$.

5.6 The one-holed Klein bottle

The fundamental group of the one-holed Klein bottle $C_{1,1}$ is free of rank two, with presentation

$$\pi_1(C_{1,1}) := \langle P, Q, R, D \mid PQR = P^2Q^2D = \text{Id} \rangle \cong \mathbb{F}_2.$$

The free generators P, Q correspond to orientation-reversing simple curves on $C_{1,1}$ and D corresponds to $\partial C_{1,1}$.

The orientable double covering-space $\widehat{C}_{1,1} \rightarrow C_{1,1}$ is connected, has two boundary components (since $\partial C_{1,1}$ is connected and orientable) and has Euler characteristic $-2 = 2\chi(C_{1,1})$. Therefore $\widehat{C}_{0,2} \approx \Sigma_{1,2}$, the two-holed torus.

This covering-space $\Sigma_{1,2} \rightarrow C_{1,1}$ induces the monomorphism

$$\begin{aligned} \pi_1(\Sigma_{1,2}) &\hookrightarrow \pi_1(C_{1,1}), \\ U &\mapsto PQ, \\ X &\mapsto QP^{-1}, \\ Y &\mapsto P^2, \\ A &\mapsto PQ^2P \sim P^2Q^2, \\ B &\mapsto PQP^2QP^{-1} \sim P^2Q^2, \end{aligned}$$

where $\pi_1(\Sigma_{1,2})$ is presented as

$$\langle A, B, U, X, Y \mid A = UXY, B = UYX \rangle.$$

The character ring of $\pi_1(C_{1,1})$ is the polynomial ring $\mathfrak{R}_2 \cong \mathbb{C}[p, q, r]$ and the character ring of $\pi_1(\Sigma_{1,2})$ is the quotient of $\mathbb{C}[a, b, x, y, z, u, v, w]$ by the relations defined by (5.30). The covering space $\Sigma_{1,2} \rightarrow C_{1,1}$ induces the homomorphism of character rings:

$$\begin{aligned} \mathfrak{R}_3 &\longrightarrow \mathfrak{R}_2 = \mathbb{C}[p, q, r], \\ u &\mapsto r, \\ x &\mapsto pq - r, \\ y &\mapsto p^2 - 2, \\ v &\mapsto q^2 - 2, \\ w &\mapsto r, \\ z &\mapsto p(pr - q) - r, \\ a &\mapsto 2 - p^2 - q^2 + pr, \\ b &\mapsto 2 - p^2 - q^2 + pr, \end{aligned}$$

which evidently satisfies the relations of (5.30).

We briefly give a geometric description of the deck transformation of the double covering of the $\mathrm{SL}(2, \mathbb{C})$ -character variety V_3 of the rank three free group \mathbb{F}_3 .

Consider the elliptic involution ι of the torus $\Sigma_{1,0}$. Writing $\Sigma_{1,0}$ as the quotient $\mathbb{R}^2/\mathbb{Z}^2$, this involution is induced by the map

$$\begin{aligned} \mathbb{R}^2 &\longrightarrow \mathbb{R}^2, \\ u &\mapsto -u. \end{aligned}$$

This involution has four fixed points, and its quotient orbifold is S^2 with four branch points of order two. Choose a small disc $D \subset \Sigma_{1,0}$ such that D and its image $\iota(D)$ are disjoint. Then ι induces an involution on the complement

$$\Sigma_{1,0} \setminus (D \cup \iota(D)) \approx \Sigma_{1,2}.$$

This involution of the two-holed torus $\Sigma_{1,2}$ induces the involution of $\pi_1(\Sigma_{1,2}) = \langle U, X, Y \rangle$:

$$\begin{aligned} U &\longmapsto U^{-1}, \\ X &\longmapsto X^{-1}, \\ Y &\longmapsto Y^{-1}. \end{aligned}$$

The quotient orbifold is a disc with four branch points of order two. The corresponding involution of character varieties is the branched double covering (5.4) of the character variety V_3 over \mathbb{C}^6 described in Proposition 5.1.1.

5.7 Free groups of rank ≥ 3

The basic trace identity (2.6), the Sum Relation (5.1) and the Product Relation (5.2) imply that the trace polynomial f_w of any word $w(X_1, X_2, \dots, X_n)$ can be written in terms of trace polynomials of words

$$X_{i_1} X_{i_2} \dots X_{i_1} X_{i_r},$$

where $1 \leq i_1 < i_2 < \dots < i_r \leq n$. The following identity, which may be found in Vogt [76], implies that it suffices to choose $r \leq 3$:

$$\begin{aligned} 2t_{1234} &= t_1 t_2 t_3 t_4 + t_1 t_2 t_3 + t_2 t_3 t_4 + t_3 t_4 t_1 + t_4 t_1 t_2 \\ &\quad + t_{12} t_3 t_4 + t_{41} t_2 t_3 - t_{13} t_2 t_4 - t_1 t_2 t_3 t_4 - t_{12} t_3 t_4 - t_4 t_1 t_2 t_3 - t_{41} t_2 t_3. \end{aligned}$$

The $\mathrm{SL}(2, \mathbb{C})$ -character variety of a rank n free group has dimension $3n - 3$, as it corresponds to the quotient of the $3n$ -dimensional complex manifold $\mathrm{SL}(2, \mathbb{C})^n$ by the generically free action of the 3-dimensional group $\mathrm{PGL}(2, \mathbb{C})$. Thus the transcendence degree of the field of fractions of the *character ring* equals $3n - 3$. In contrast, the above discussion implies that this ring has

$$n + \binom{n}{2} + \binom{n}{3} = \frac{n(5 + n^2)}{6}$$

generators, considerably larger than the dimension $3n - 3$ of the character variety.

References

- [1] H. Bass, Letter to W. Goldman, April 1983. [626](#)
- [2] R. Benedetto and W. Goldman, The Topology of the Relative Character Varieties of a Quadruply-Punctured Sphere. *Experiment. Math.* 8 (1) (1999), 85–104. [670](#), [672](#)
- [3] L. Bers and F. Gardiner, Fricke spaces. *Adv. Math.* 62 (1986), 249–284. [612](#)
- [4] B. Bowditch, Markoff triples and quasi-Fuchsian groups. *Proc. London Math. Soc.* 77 (1998), 607–736.

- [5] S. Bratholdt and D. Cooper, On the topology of the character variety of a free group. *Rend. Istit. Mat. Univ. Trieste* 32 (suppl. 1) (2001), 45–53.
- [6] K. S. Brown, *Cohomology of groups*. Grad. Texts in Math. 87, Springer-Verlag, New York 1982. [629](#)
- [7] G. Brumfiel and H. Hilden, *SL(2)-representations of finitely presented groups*. Contemp. Math. 187, Amer. Math. Soc., Providence, RI, 1995.
- [8] P. Buser, *Geometry and Spectra of Compact Riemann Surfaces*. Progr. Math. 106, Birkhäuser, Boston 1992.
- [9] J. O. Button, Matrix representations and the Teichmüller space of the twice punctured torus. *Conform. Geom. Dyn.* 4 (2000), 97–107 (electronic). [675](#)
- [10] S. Cantat, Bers and Hénon, Painlevé and Schrödinger. Preprint, arXiv:0711.1727. [670](#)
- [11] S. Cantat and F. Loray, Holomorphic dynamics, Painlevé VI and character varieties. Preprint, arXiv:0711.1579. [670](#)
- [12] M. Culler, Lifting representations to covering groups. *Adv. Math.* 59 (1) (1986), 64–70. [630](#)
- [13] M. Culler and P. Shalen, Varieties of group representations and splittings of 3-manifolds. *Ann. of Math.* 117 (1983), 109–146. [627](#)
- [14] G. de Rham, Sur les polygones générateurs des groupes fuchsien. *Enseignement Math.* 17 (1971), 49–61.
- [15] C. Ehresmann, Sur les espaces localement homogènes. *Enseignement Math.* 35 (1936), 317–333. [612](#)
- [16] W. Fenchel, *Elementary Geometry in Hyperbolic Space*. De Gruyter Stud. Math. 11, Walter de Gruyter, Berlin 1989. [624](#), [635](#)
- [17] W. Fenchel and J. Nielsen, *Discontinuous groups of isometries in the hyperbolic plane*. De Gruyter Stud. Math. 29, Walter de Gruyter, Berlin 2003. [657](#)
- [18] F. Luo, Grothendieck’s reconstruction principle and 2-dimensional topology and geometry. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 733–765. [615](#)
- [19] C. Florentino, Invariants of 2×2 matrices, irreducible $SL(2, \mathbb{C})$ -characters and the Magnus trace map. *Geom. Dedicata* 121 (2006), 167–186. [665](#)
- [20] R. Fricke, Über die Theorie der automorphen Modulgruppen. *Nachr. Akad. Wiss. Göttingen* (1896), 91–101. [612](#), [618](#)
- [21] R. Fricke and F. Klein, *Vorlesungen über die Theorie der automorphen Funktionen*. Vol. I, Vol. II, Teubner, Leipzig 1897, 1912. [612](#), [618](#)
- [22] J. Gilman, Two-Generator Discrete Subgroups of $PSL(2, \mathbb{R})$. *Mem. Amer. Math. Soc.* 117 (561) (1995).
- [23] W. Goldman, Discontinuous groups and the Euler class. Doctoral dissertation, University of California, Berkeley 1980. [671](#)
- [24] W. Goldman, The symplectic nature of fundamental groups of surfaces. *Adv. Math.* 54 (1984), 200–225.
- [25] W. Goldman, Representations of fundamental groups of surfaces. In *Geometry and Topology* (College Park, Md., 1983/84), J. Alexander and J. Harer, eds., Lecture Notes in Math.

- 1167, Springer-Verlag, Berlin 1985, 95–117.
- [26] W. Goldman, Invariant functions on Lie groups and Hamiltonian flows invariant functions of surface group representations. *Invent. Math.* 85 (1986), 1–40.
- [27] W. Goldman, Topological components of spaces of representations. *Invent. Math.* 93 (3), (1988), 557–607. [630](#), [639](#), [642](#), [671](#), [673](#)
- [28] W. Goldman, Geometric structures and varieties of representations. In *The Geometry of Group Representations*, Contemp. Math. 74, Amer. Math. Soc., Providence, RI, 1988, 169–198. [651](#)
- [29] W. Goldman, Ergodic theory on moduli spaces. *Ann. of Math.* (2) 146 (1997), 1–33. [670](#)
- [30] W. Goldman, Action of the modular group on real $SL(2)$ -characters of a one-holed torus. *Geom. Topol.* 7 (2003), 443–486. [615](#), [651](#), [656](#)
- [31] W. Goldman, An exposition of results of Fricke. Preprint, math/0402103v2. [615](#)
- [32] W. Goldman, Projective Geometry on Manifolds. In preparation. [651](#)
- [33] W. Goldman, G. McShane, G. Stantchev, and S. Tan. In preparation. [660](#), [661](#)
- [34] W. Goldman and W. Neumann, Homological action of the modular group on some cubic moduli spaces. *Math. Res. Lett.* 12 (4) (2005), 575–591. [670](#)
- [35] W. Goldman, and G. Stantchev, Dynamics of the Automorphism Group of the $GL(2, \mathbb{R})$ -Characters of a Once-punctured Torus. Preprint, math.DG/0309072. [661](#)
- [36] F. Goñzález-Acuna and J. M. Montesinos-Amilibia, On the character variety of group representations in $SL(2, C)$ and $PSL(2, C)$. *Math. Z.* 214 (4) (1993), 627–652.
- [37] L. Greenberg, Homomorphisms of triangle groups into $PSL(2, C)$. In *Riemann surfaces and related topics* (State Univ. New York, Stony Brook, N.Y., 1978), Ann. Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 167–181.
- [38] W. J. Harvey, Spaces of Discrete Groups. In *Discrete Groups and Automorphic Functions*, Academic Press, London 1977, 295–347.
- [39] H. Helling, Diskrete Untergruppen vol $SL(2, \mathbb{R})$. *Invent. Math.* 17 (1972), 217–229.
- [40] R. Horowitz, Characters of free groups represented in the two-dimensional linear group. *Comm. Pure Appl. Math.* 25 (1972), 635–649.
- [41] K. Iwasaki, An area-preserving action of the modular group on cubic surfaces and the Painlevé VI equation. *Commun. Math. Phys.* 242 (1–2) (2003), 185–219. [670](#)
- [42] K. Iwasaki, and T. Uehara, An ergodic study of Painlevé VI. *Math. Ann.* 338 (2) (2007), 295–345.
- [43] T. Jørgensen, Compact 3-manifolds of constant negative curvature fibering over the circle. *Ann. of Math.* (2) 106 (1977), 61–72. [624](#)
- [44] L. Keen, Intrinsic Moduli on Riemann Surfaces. *Ann. of Math.* 84 (3) (1966), 404–420.
- [45] L. Keen, On Fricke Moduli. In *Advances in the Theory of Riemann Surfaces* (Proc. Conf., Stony Brook, N.Y., 1969), Ann. of Math. Stud. 66, Princeton University Press, Princeton, N. J., 1971, 205–224; A correction to: “On Fricke moduli”. *Proc. Amer. Math. Soc.* 40 (1973), 60–62.
- [46] L. Keen, A Rough Fundamental Domain for Teichmüller Spaces. *Bull. Amer. Math. Soc.* 83 (6) (1977), 1199–1226.

- [47] A. W. Knapp, Doubly generated Fuchsian groups. *Michigan Math. J.* 15 (1969), 289–304.
- [48] I. Kra, On lifting Kleinian groups to $SL(2, \mathbb{C})$. In *Differential geometry and complex analysis*, Springer-Verlag, Berlin 1985, 181–193. [630](#)
- [49] Lang, S., *Algebra*. Addison-Wesley, 1971. [626](#)
- [50] S. Lawton, Generators, relations and symmetries in pairs of 3×3 unimodular matrices. *J. Algebra* 313 (2) (2007), 782–801.
- [51] A. Lubotzky and A. Magid, *Varieties of representations of finitely generated groups*. Mem. Amer. Math. Soc. 336 (58) (1985).
- [52] F. Luo, Characters of $\mathfrak{sl}(2)$ -representations of groups. *J. Differential Geom.* 53 (3) (1999), 575–626.
- [53] C. Maclachlan and A. W. Reid, *The Arithmetic of Hyperbolic 3-Manifolds*. Grad. Texts in Math. 219, Springer-Verlag, New York 2003.
- [54] W. Magnus, Rings of Fricke characters and automorphism groups of free groups. *Math. Z.* 170 (1980), 91–103.
- [55] A. Marden, *Outer Circles*. Cambridge University Press, Cambridge 2007. [638](#)
- [56] B. Maskit, Parameters for Fuchsian groups. I: Signature (0, 4). In *Holomorphic functions and moduli* (D. Drasin et al., eds.), Vol. II, Math. Sci. Res. Inst. Publ. 11, Springer-Verlag, New York 1988, 251–265.
- [57] B. Maskit, Parameters for Fuchsian groups. I: Signature (1, 1). *Ann. Acad. Sic. Fenn. Ser. A I Math.* 14 (1989), 265–275.
- [58] B. Maskit, On Poincaré’s theorem for fundamental polygons. *Adv. Math.* 7 (1971), 219–230.
- [59] J. W. Milnor, On the existence of a connection with curvature zero. *Comment. Math. Helv.* 32 (1958), 215–223. [630](#)
- [60] J. W. Milnor, and J. Stasheff, *Characteristic classes*. Ann. Math. Studies 76, Princeton University Press, Princeton, N.J., 1974.
- [61] J. W. Morgan and P. B. Shalen, Valuations, trees, and degenerations of hyperbolic structures I. *Ann. of Math.* (2) 120 (3) (1984), 401–476. [643](#)
- [62] P. Newstead, *Introduction to moduli problems and orbit spaces*. Tata Institute of Fundamental Research, Bombay 1978.
- [63] S. P. K. Ng and S. P. Tan, The complement of the Bowditch space in the $SL(2, \mathbb{C})$ -character variety. *Osaka J. Math.* 44 (2007), 247–254.
- [64] P. Norbury, Lengths of geodesics on non-orientable hyperbolic surfaces. *Geom. Dedicata* 134 (2008), 153–176.
- [65] S. Patterson, On the cohomology of Fuchsian groups. *Glasgow Math. J.* 16 (2) (1975), 123–140. [630](#)
- [66] H. Petersson, Zur analytischen Theorie der Grenzkreisgruppen. *Math. Ann.* 115 (1) (1938), 518–572. [630](#)
- [67] C. Procesi, The invariant theory of $n \times n$ matrices. *Adv. Math.* 19 (1976), 306–381. [612](#)
- [68] J. Przytycki and A. Sikora, On skein algebras and $SL(2, \mathbb{C})$ -character varieties. *Topology* 39 (1) (2000), 115–148.

- [69] J. J. Rotman, *Advanced Modern Algebra*. Prentice Hall, Upper Saddle River, N.J., 2002. [626](#)
- [70] K. Saito, Algebraic representations of Teichmüller space. *Kodai Math. J.* 17 (1994), 609–626.
- [71] C. Series, The geometry of Markoff numbers. *Math. Intelligencer* 7 (1985), 20–29.
- [72] G. Stantchev, Action of the modular group on $GL(2, \mathbb{R})$ -characters on a once-punctured torus. Doctoral dissertation, University of Maryland 2002. [660](#), [661](#)
- [73] N. Steenrod, *The topology of fibre bundles*. Princeton University Press, Princeton, N. J., 1951.
- [74] S. P. Tan, Y. L. Wang and Y. Zhang, The $SL(2, \mathbb{C})$ -character variety of a one holed torus. *Electron. Res. Announc. Amer. Math. Soc.* 11 (2005), 103–111.
- [75] W. Thurston, *Three-dimensional geometry and topology.*, Vol. 1, Silvio Levy, ed., Princeton Math. Ser. 35, Princeton University Press, Princeton, N.J., 1997. [651](#)
- [76] H. Vogt, Sur les invariants fondamentaux des équations différentielles linéaires du second ordre. *Ann. Sci. École Norm. Sup.* (3) VI (Suppl. 3-72) (1889). [612](#), [615](#), [618](#), [680](#)
- [77] Z. X. Wen, Relations polynomiales entre les traces de produits de matrices. *C. R. Acad. Sci. Paris* 314 (1994), 99–104.
- [78] P. Will, Groupes libres, groupes triangulaires et tore épointé. Thèse, Université Paris VI.
- [79] P. Will, Cross-ratios and 2-generator subgroups of $PU(2, 1)$. *Canad. J. Math.*, to appear.
- [80] S. Wolpert, The Fenchel-Nielsen deformation. *Ann. of Math.* 115 (1982), 501–528. [657](#), [659](#)
- [81] S. Wolpert, On the symplectic geometry of deformations of a hyperbolic surface. *Ann. of Math.* (2) 117 (2) (1983), 207–234. [657](#)
- [82] S. Wolpert, On the Weil-Petersson geometry of the moduli space of curves. *Amer. J. Math.* 107 (4) (1985), 969–997. [657](#)

Chapter 16

Spin networks and $SL(2, \mathbb{C})$ -character varieties

Sean Lawton and Elisha Peterson

Contents

1	Introduction	685
2	Preliminaries	689
	2.1 Algebraic structure of the character variety $\mathcal{R}/SL(2, \mathbb{C})$	689
	2.2 A partial history of $SL(2, \mathbb{C})$ invariant theory	690
	2.3 Representation theory of $SL(2, \mathbb{C})$	691
3	The spin network calculus	692
	3.1 Spin networks and representation theory	693
	3.2 Basic diagram manipulations	696
	3.3 Symmetrizers and irreducible representations	698
	3.4 Symmetrizers and trivalent spin networks	701
	3.5 Trivalent diagram manipulations	703
4	Decomposition of $\mathbb{C}[G]$	707
	4.1 Central functions	707
	4.2 Proof of $\mathbb{C}[G]$ decomposition theorem	709
	4.3 Ring structure of $\mathbb{C}[G]^G$	713
5	Structure of $\mathbb{C}[G \times G]^G$	715
	5.1 Symmetry of central functions	717
	5.2 A recurrence relation for central functions	719
	5.3 Graded structure of the central function basis	723
	5.4 Multiplication of central functions	724
	5.5 Applications	726
	References	728

1 Introduction

The purpose of this chapter is to demonstrate the utility of a graphical calculus in the algebraic study of $SL(2, \mathbb{C})$ -representations of the fundamental group of an oriented surface of Euler characteristic -1 .

Let F_2 be a rank 2 free group, the fundamental group of both the three-holed sphere and the one-holed torus. The set $\mathcal{R} = \text{Hom}(F_2, SL(2, \mathbb{C}))$ of representations

inherits the structure of an algebraic set from $\mathrm{SL}(2, \mathbb{C})$. The subset of representations that are *completely reducible*, denoted by \mathcal{R}^{ss} , have closed orbits under conjugation. Consequently, the orbit space $\mathcal{R}^{ss}/\mathrm{SL}(2, \mathbb{C}) = \mathcal{R}/\mathrm{SL}(2, \mathbb{C})$ is an algebraic set referred to as the *character variety*. The character variety encodes Teichmüller space and other moduli of geometric structures [17].

Graphs known as *spin networks* permit a concise description of a natural additive basis for the coordinate ring of the character variety

$$\mathbb{C}[\mathcal{R}/\mathrm{SL}(2, \mathbb{C})] = \mathbb{C}[\mathcal{R}]^{\mathrm{SL}(2, \mathbb{C})}.$$

We will refer to the basis elements as *central functions*. The central functions are indexed by Clebsch–Gordan injections

$$V_c \hookrightarrow V_a \otimes V_b,$$

where $V_c = \mathrm{Sym}^c(\mathbb{C}^2)$ denotes an irreducible representation of $\mathrm{SL}(2, \mathbb{C})$. Our main results use the spin network calculus to describe a strong symmetry within the central function basis, a graphical means of computing the product of two central functions, and an algorithm for computing central functions. This provides a concrete description of the regular functions on the $\mathrm{SL}(2, \mathbb{C})$ -character variety of F_2 and a new proof of a classical result of Fricke, Klein, and Vogt.

We are motivated by a greater understanding of the invariant ring, and the subsequent knowledge of various geometric objects of interest encoded within the character variety. Consequently, the main results in this chapter concern the structure of the central function basis. The results and methods of this chapter may also provide new insight into gauge theoretic questions. However, we are most interested in a methodology and point of view that allows for generalizations to other Lie groups and other surface groups.

History of central functions and spin networks

The first reference to the central function basis in the literature appears in [2], where Baez used spin networks to describe a basis of quantum mechanical “state vectors.” He considered the basis abstractly, showing that the space of square integrable functions on a related space of connections modulo gauge transformations is spanned by a set of labelled graphs. He also demonstrated that the basis is orthonormal with respect to the L^2 inner product. His basis, when restricted to $\mathrm{SU}(2)$, is precisely the one under consideration here.

More recently, Florentino, Mourão, and Nunes use a like basis to produce distributions related to geometric quantization of moduli spaces of flat connections on a surface [13]. Adam Sikora has also used spin networks to study the character variety for $\mathrm{SL}(n, \mathbb{C})$, although without using the central function basis [30]. The construction of arbitrary rank $\mathrm{SL}(2, \mathbb{C})$ central functions is described in [25], while much of the diagrammatic theory required for the $\mathrm{SL}(n, \mathbb{C})$ case is covered in [8], [9], [25], [30].

The history of the diagrammatic calculus in this chapter is hard to trace, due to the historical difficulty in publishing papers making extensive use of figures. While it is likely that many works on diagrammatic notation have been lost over the years, the specific notation used in this chapter is due to Roger Penrose. In a 1981 letter to Predrag Cvitanović, a physicist who also used diagrams extensively, Penrose recalls developing the notation in the early 1950s while “trying to cope with Hodge’s lectures on differential geometry” [24].

Diagrammatic notations have also played an important role in modern physics. Feynman diagrams are probably the most famous example, but spin networks have also been used for many years, as a graphical description of quantum angular momentum [23]. The use of diagrams in physics is probably best summarized in [31]. Cvitanović also has a thorough description of such notations, which he calls *birdtracks* in [8], [9]. In his work, birdtracks play a starring role in a new classification of semi-simple Lie algebras. Using primitive invariants, which have unique diagrammatic depictions, the exceptional Lie algebras arise in a single series in a construction that he calls the “Magic Triangle.”

This chapter is organized as follows.

Section 2 gives some basic definitions and results from invariant theory, as well as a short history of $SL(2, \mathbb{C})$ invariant theory. It also covers necessary material from representation theory.

In Section 3, we introduce spin networks, which are special types of graphs that may be identified with functions between tensor powers of \mathbb{C}^2 . We give a full treatment of the *spin network calculus*, a powerful means for working with regular functions on $\mathcal{R}/SL(2, \mathbb{C})$.

Section 4 begins by constructing an additive basis for $\mathbb{C}[\mathcal{R}/SL(2, \mathbb{C})]$. This basis, denoted by $\{\chi^{a,b,c}\}$, is indexed by triples of nonnegative integers (a, b, c) satisfying the *admissibility condition*:

$$\frac{1}{2}(-a + b + c), \frac{1}{2}(a - b + c), \frac{1}{2}(a + b - c) \in \mathbb{N}.$$

The functions $\chi^{a,b,c} \in \mathbb{C}[\mathcal{R}/SL(2, \mathbb{C})]$ are central in

$$\text{End}(V_c) \hookrightarrow \text{End}(V_a) \otimes \text{End}(V_b),$$

and are referred to as *central functions*. The construction of the central function basis uses the decomposition

$$\mathbb{C}[SL(2, \mathbb{C})] \cong \sum_{n \geq 0} V_n^* \otimes V_n.$$

We include a constructive proof of this decomposition, since it is hard to find in the literature. The section concludes by examining the $SL(2, \mathbb{C})$ -central functions of a rank one free group.

Section 5 contains the main results of this chapter, which concern the case of a rank two free group. In this case, central functions may be written as polynomials in

three trace variables, a consequence of a theorem due to Fricke, Klein, and Vogt [14], [32]. The results we prove are summarized below.

- Theorem 5.2 describes a symmetry property of the central function basis: permuting the indices of a central function is equivalent to permuting the variables of its polynomial representation.
- Corollary 5.7 states that, with an appropriate definition of rank, any central function may be written in terms of at most four central functions of lower rank:

$$\begin{aligned} \chi^{a,b,c} = & x \cdot \chi^{a-1,b,c-1} - \frac{(a+b-c)^2}{4a(a-1)} \chi^{a-2,b,c} - \frac{(-a+b+c)^2}{4c(c-1)} \chi^{a,b,c-2} \\ & - \frac{(a+b+c)^2(a-b+c-2)^2}{16a(a-1)c(c-1)} \chi^{a-2,b,c-2}. \end{aligned}$$

Together with Theorem 5.2, this result gives an algorithm for computing central functions explicitly.

- Proposition 5.8 states that central functions are monic, and gives the leading term of the central function $\chi^{a,b,c}$.
- Proposition 5.9 describes a $\mathbb{Z}_2 \times \mathbb{Z}_2$ grading on the central function basis.
- Theorem 5.11 gives the coefficients in the expression of the product of two central functions as a sum of central functions, and therefore a precise description of the ring structure of $\mathbb{C}[\mathcal{R}]^{\text{SL}(2,\mathbb{C})}$ in terms of central functions.

Finally, as another consequence of the recurrence relation and Theorem 5.2, we provide a new constructive proof of the following classical theorem [14], [32]:

Theorem 5.12 (Fricke–Klein–Vogt Theorem). *Let $G = \text{SL}(2, \mathbb{C})$ act on $G \times G$ by simultaneous conjugation. Then*

$$\mathbb{C}[G \times G]^G \cong \mathbb{C}[t_x, t_y, t_z],$$

the complex polynomial ring in three indeterminates. In particular, every regular function $f : \text{SL}(2, \mathbb{C}) \times \text{SL}(2, \mathbb{C}) \rightarrow \mathbb{C}$ satisfying

$$f(\mathbf{x}_1, \mathbf{x}_2) = f(g\mathbf{x}_1g^{-1}, g\mathbf{x}_2g^{-1}) \quad \text{for all } g \in \text{SL}(2, \mathbb{C}),$$

can be written uniquely as a polynomial in the three trace variables $x = \text{tr}(\mathbf{x}_1)$, $y = \text{tr}(\mathbf{x}_2)$, and $z = \text{tr}(\mathbf{x}_1\mathbf{x}_2^{-1})$.

Acknowledgements. We would like to thank Bill Goldman for introducing this problem to us and for many helpful suggestions, including generously sharing his *Mathematica* notebooks with us. His correspondence with Nicolai Reshetikhin and Charles Frohman provided the foundation for the application of spin networks to this problem. Reshetikhin sketched proofs of both Theorems 5.2 and 5.11. This work has benefited from helpful conversations with Ben Howard, Tom Haines, and regular participation in the University of Maryland’s Research Interaction Teams. The first author has received research support from John Millson, Richard Schwartz, and the University of

Maryland's VIGRE grant. The second author has been supported by an NSF Graduate Fellowship. All diagrams in this chapter were generated by a suite of TikZ/PGF commands written by the second author.

We benefitted greatly from comments by Adam Sikora, Carlos Florentino, Athanase Papadopoulos, and the referee on early drafts of this chapter. Carlos Florentino provided many helpful corrections, and pointed out the significance of the symmetry in Theorem 5.2. He also suggested the correspondence given in Proposition 5.8. The referee gave valuable feedback regarding the organization and exposition.

2 Preliminaries

2.1 Algebraic structure of the character variety $\mathcal{R} // SL(2, \mathbb{C})$

The group $G = SL(2, \mathbb{C})$ has the structure of an irreducible algebraic set, since it is the zero set of the irreducible polynomial $\det(\mathbf{x}) - 1$. Since the product of two varieties is again a variety, the *representation variety* $\mathcal{R} = \text{Hom}(F_2, G) \cong G \times G$ of a rank 2 free group F_2 is an irreducible algebraic set as well. The *coordinate ring* of \mathcal{R} is

$$\mathbb{C}[\mathcal{R}] = \frac{\mathbb{C}[x_{ij}^k : 1 \leq i, j, k \leq 2]}{(\det(\mathbf{x}_1) - 1, \det(\mathbf{x}_2) - 1)}.$$

Stated otherwise, it is the free commutative polynomial ring in 8 indeterminates over \mathbb{C} subject to the ideal generated by the two polynomials $\det(\mathbf{x}_k) - 1$, where $\mathbf{x}_k = (x_{ij}^k)$ are called *generic matrices*.

There is an action of G on \mathcal{R} by simultaneous conjugation. Given $(\mathbf{x}_1, \mathbf{x}_2) \in G \times G$, then

$$g \cdot (\mathbf{x}_1, \mathbf{x}_2) = (g\mathbf{x}_1g^{-1}, g\mathbf{x}_2g^{-1}).$$

This is a *polynomial action*, since $\mathcal{R} \times G \rightarrow \mathcal{R}$ is a regular mapping.

Definition 2.1. The *ring of invariants* $\mathbb{C}[\mathcal{R}]^G$ consists of elements of the coordinate ring $\mathbb{C}[\mathcal{R}]$ which are invariant under the action of simultaneous conjugation:

$$\mathbb{C}[\mathcal{R}]^G = \{f \in \mathbb{C}[\mathcal{R}] : g \cdot f = f\}.$$

Recall that an algebraic group is *linearly reductive* if its finite dimensional rational representations are decomposable as direct sums of irreducible representations. Since $G = SL(2, \mathbb{C})$ is linearly reductive, the ring of invariants $\mathbb{C}[\mathcal{R}]^G = \{f \in \mathbb{C}[\mathcal{R}] : g \cdot f = f\}$ is finitely generated [10]. This implies that the space of maximal ideals of $\mathbb{C}[\mathcal{R}]^G$ is also an irreducible algebraic set, permitting the following definition:

Definition 2.2. The *G-character variety* of F_2 is the space of maximal ideals

$$\mathfrak{X} = \text{Spec}_{\max}(\mathbb{C}[\mathcal{R}]^G) = \mathcal{R} // G.$$

The character variety \mathfrak{X} is identified with conjugacy classes of *completely reducible* representations in \mathcal{R} [1], [27]. Procesi [26] has shown that $\mathbb{C}[\mathcal{R}]^G$ is generated by traces of products of matrix variables of word length less than or equal to three [26]. Hence $\mathbb{C}[\mathfrak{X}]$ is generated, although not minimally, by

$$\{\mathrm{tr}(\mathbf{x}_1), \mathrm{tr}(\mathbf{x}_2), \mathrm{tr}(\mathbf{x}_1\mathbf{x}_2), \mathrm{tr}(\mathbf{x}_1\mathbf{x}_2^2), \mathrm{tr}(\mathbf{x}_2\mathbf{x}_1^2)\}.$$

2.2 A partial history of $\mathrm{SL}(2, \mathbb{C})$ invariant theory

The following paragraphs provide a partial account of the long history of $\mathrm{SL}(2, \mathbb{C})$ invariant theory. Two pioneering papers on the subject were authored by Vogt in 1889 [32], and by Fricke and Klein in 1896 [14]. Both investigated the invariants of pairs of unimodular 2×2 matrices with respect to simultaneous conjugation. They showed this ring of invariants to be the free commutative polynomial ring in three indeterminates, given by the trace of each generic matrix and the trace of their product. This chapter concludes with a reproof of this classical result using the spin network calculus.

In 1972, Horowitz investigated the algebraic structure of this ring, saying that Fricke’s approach was principally analytic, and partially incomplete [20]. In 1980, Magnus made clear the priority of Vogt’s approach [32] and worked out the defining polynomial relations for an arbitrary number of matrices under simultaneous conjugation [22]. In 1983, Culler and Shalen defined the character variety and showed that it is in fact an algebraic set [6]; the set is the image under a “trace” map. González-Acuña and Montesinos-Amilibia showed in 1993 that the relations of Magnus in fact determine the algebraic set that Culler and Shalen had defined [19]. In 2001, Sikora, using results of Procesi [26], showed that the character variety of $\mathrm{SL}(n, \mathbb{C})$ can be realized as spaces of graphs subject to topologically motivated relations [30]. These graphs correspond to the spin networks discussed in this chapter when $n = 2$.

Closely related is the ring of invariants of *arbitrary* generic 2×2 matrices under simultaneous conjugation. The works of Procesi (1976) and Razmyslov (1974) generalized the work above to the case of $n \times n$ matrices [26], [28], and showed that the invariant ring is generated by traces of words in generic matrices. Methods from geometric invariant theory (see Dolgachev [10]) show that the character variety is the variety whose coordinate ring is the ring of invariants. Restricting to unimodular matrices gives like results for the unimodular ring of invariants. From this point of view, the character variety begins as an algebraic set and so is obviously closed. However, the defining relations and minimal generators are not at all obvious.

A central question in invariant theory is a description of the generators and relations of an invariant ring. Indeed, a theorem that characterizes the generators of an invariant ring is called a *first fundamental theorem*, and a theorem giving the relations is called a *second fundamental theorem*. In [26], [28] both Procesi and Razmyslov gave the two fundamental theorems, although they offered only sufficient generators and an implicit description of the relations.

It is much more difficult to determine *minimal* generators and *explicit* relations. In this more general context, which bears strongly on the unimodular case, minimal generators and defining relations for the invariants of an arbitrary number of generic 2×2 matrices were found only recently by Drensky in 2003 [11].

2.3 Representation theory of $\mathrm{SL}(2, \mathbb{C})$

The coordinate ring $\mathbb{C}[G]$ decomposes into a direct sum of tensor products of the finite-dimensional irreducible representations of G . We will use this decomposition, given explicitly by Theorem 4.1, to understand the coordinate ring of the character variety \mathcal{X} . To this end, we review the representation theory of G (see [3], [10], [15]).

The symmetric powers of the standard representation of G are all irreducible representations and moreover they comprise a complete list. Let $V_0 = \mathbb{C} = V_0^*$ be the trivial representation of G . Denote the standard basis for \mathbb{C}^2 by $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and the dual basis by $e_1^* = e_1^T$ and $e_2^* = e_2^T$. Then the standard representation and its dual are

$$V = V_1 = \mathbb{C}e_1 \oplus \mathbb{C}e_2 \quad \text{and} \quad V^* = V_1^* = \mathbb{C}e_1^* \oplus \mathbb{C}e_2^*,$$

respectively. Denote the symmetric powers of these representations by

$$V_n = \mathrm{Sym}^n(V) \quad \text{and} \quad V_n^* = \mathrm{Sym}^n(V^*).$$

Since V_n admits an invariant non-degenerate bilinear form, $V_n \cong (V_n)^*$.

Moreover, V_n^* is naturally isomorphic to $(V_n)^*$, so elements in V_n pair with elements in V_n^* . Denote the projection of $v_1 \otimes v_2 \otimes \cdots \otimes v_n \in V^{\otimes n}$ to V_n by $v_1 \circ v_2 \circ \cdots \circ v_n$. There exist bases for V_n and V_n^* , given by the elements

$$\mathfrak{n}_{n-k} = e_1^{n-k} e_2^k = \underbrace{e_1 \circ e_1 \circ \cdots \circ e_1}_{n-k} \circ \underbrace{e_2 \circ e_2 \circ \cdots \circ e_2}_k$$

and

$$\mathfrak{n}_{n-k}^* = (e_1^*)^{n-k} (e_2^*)^k = \underbrace{e_1^* \circ e_1^* \circ \cdots \circ e_1^*}_{n-k} \circ \underbrace{e_2^* \circ e_2^* \circ \cdots \circ e_2^*}_k,$$

respectively, where $0 \leq k \leq n$. In these terms, this pairing is given by

$$\mathfrak{n}_{n-k}^*(v_1 \circ v_2 \circ \cdots \circ v_n) = \frac{1}{n!} \sum_{\sigma \in \Sigma_n} (\mathfrak{n}_{n-k})^*(v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)}),$$

where Σ_n is the symmetric group on n elements. In particular,

$$\mathfrak{n}_{n-k}^*(\mathfrak{n}_{n-l}) = \frac{(n-k)!k!}{n!} \delta_{kl} = \delta_{kl} / \binom{n}{k}.$$

Let $g = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \in G$. Then the G -action on V_n is given by

$$\begin{aligned} g \cdot n_{n-k} &= (g_{11}e_1 + g_{21}e_2)^{n-k} (g_{12}e_1 + g_{22}e_2)^k \\ &= \sum_{\substack{0 \leq j \leq n-k \\ 0 \leq i \leq k}} \binom{n-k}{j} \binom{k}{i} (g_{11}^{n-k-j} g_{12}^{k-i} g_{21}^j g_{22}^i) n_{n-(i+j)}. \end{aligned}$$

For the dual, G acts on V_n^* in the usual way:

$$(g \cdot n_{n-k}^*)(v) = n_{n-k}^*(g^{-1} \cdot v) \quad \text{for } v \in V_n.$$

The tensor product $V_a \otimes V_b$, where $a, b \in \mathbb{N}$, is also a representation of G and decomposes into irreducible representations as follows:

Proposition 2.3 (Clebsch–Gordan formula).

$$V_a \otimes V_b \cong \bigoplus_{j=0}^{\min(a,b)} V_{a+b-2j}.$$

Finally, we give several versions of Schur’s Lemma, which will be used frequently.

Proposition 2.4 (Schur’s Lemma). *Let G be a group, V and W representations of G , and $f \in \text{Hom}_G(V, W)$ a non-zero G -equivariant linear mapping from V to W .*

- (1) *If V is irreducible, then f is injective.*
- (2) *If W is irreducible, then f is surjective.*
- (3) *If $V = W$ is irreducible, then f is a homothety.*
- (4) *Suppose V, W are irreducible:*

$$\text{if } V \cong W, \text{ then } \dim_{\mathbb{C}} \text{Hom}_G(V, W) = 1;$$

$$\text{if } V \not\cong W, \text{ then } \dim_{\mathbb{C}} \text{Hom}_G(V, W) = 0.$$

See [3] or [7] for proof of Propositions 2.3 and 2.4.

3 The spin network calculus

This section provides a self-contained introduction to spin networks and the spin network calculus. Our treatment employs a nonstandard definition of spin networks which is more natural when working with traces. This definition leads to different versions of the usual spin network relations in the literature [5], [8], [9], [21], [23], [31].

3.1 Spin networks and representation theory

At its heart, a spin network is a graph that is identified with a specific function between tensor powers of $V = \mathbb{C}^2$, the standard $SL(2, \mathbb{C})$ representation.

In order for this function to be well-defined, the edges incident to each vertex of the spin network must have a cyclic ordering. This ordering is often called a *ciliation*, since it is represented on paper by a small mark drawn between two of the edges. The edges adjacent to a ciliated vertex are ordered by proceeding in a clockwise fashion from this mark. For example, in the degree 2 case, there are two possible ciliations:

$$\begin{array}{c} \uparrow^1 \\ | \\ \downarrow^2 \end{array} \text{ and } \begin{array}{c} \uparrow^2 \\ | \\ \downarrow^1 \end{array}.$$

Definition 3.1. A *spin network* S is a graph with vertex set $S_i \sqcup S_o \sqcup S_v$ consisting of degree 1 ‘inputs’ S_i , degree 1 ‘outputs’ S_o and degree 2 ‘ciliated vertices’ S_v . If there are $k_i = |S_i|$ inputs and $k_o = |S_o|$ outputs, then S is identified with a function $f_S: V^{\otimes k_i} \rightarrow V^{\otimes k_o}$. If the spin network is *closed*, meaning $k_i = 0 = k_o$, it is identified with a complex scalar $f_S \in \mathbb{C}$.

Spin networks are drawn in *general position* inside an oriented rectangle with inputs at the bottom and outputs at the top. This convention allows us to equate the composition of functions $f_{S'} \circ f_S$ with the concatenation of diagrams $S' \circ S$ formed by placing S' on top of S .

For example, the following spin network has two ciliated vertices and represents a function from $V^{\otimes 5} \rightarrow V^{\otimes 3}$:

$$\begin{array}{c} \text{3 outputs} \\ \hline \text{Diagram of a spin network with 5 inputs and 3 outputs} \\ \hline \text{5 inputs} \end{array} = \left(\text{Cap} \right) \circ \left(\text{Cup} \right) \circ \left(\text{Crossing} \right).$$

Note that the marks on the local extrema do not indicate vertices of the graph, but are indicators of how to decompose the graph.

Since spin networks are just graphs with ciliations, it does not matter how the graph is represented inside the square. Strands may be moved about freely and ciliations may “slide” along the strands. As long as the endpoints remain fixed, the underlying spin network does not change.

Let $v, w \in V$ and let $\{e_1, e_2\}$ be the standard basis for \mathbb{C}^2 . The function f_S of a spin network S is computed by decomposing S into four *spin network component maps*:

- the *identity* $\left| \right|: V \rightarrow V, v \mapsto v$;
- the *cap* $\cap: V \otimes V \rightarrow \mathbb{C}, v \otimes w \mapsto v^T w$ (inner product);
- the *cup* $\cup: \mathbb{C} \rightarrow V \otimes V, 1 \mapsto e_1 \otimes e_1 + e_2 \otimes e_2$;

- the cap vertex $\frown: V \otimes V \rightarrow \mathbb{C}, v \otimes w \mapsto \det[v \ w]$.

For example, since \frown and \smile are the same ciliated graph,

$$\frown(v \otimes w) = \smile(v \otimes w) = \frown \circ \times(v \otimes w) = \frown(w \otimes v) = \det[w \ v].$$

The definition given here differs from the literature [5], [21], [23]. In particular, we omit the $i = \sqrt{-1}$ factor in the definition of \frown to gain an advantage in trace calculations. Also, the maps \frown and \smile are included in order to simplify the proof that f_S is well-defined.

Theorem 3.2. *The spin network function f_S is well-defined.*

Proof. We need to show that every decomposition of S into the component maps gives the same function.

If S has n ciliated vertices, then any decomposition of S into component maps has n occurrences of \frown . The remainder of the diagram consists of loops or arcs without any vertices. Two corresponding arcs in different decompositions will differ only by the insertion or deletion of a number of ‘kinks’ of the form $\smile \downarrow$. Finally, since

$$\smile \downarrow(v) = \frown \mid \circ \mid \smile(v) = \mid(v)$$

for all $v \in V$, these kinks do not change the resulting function. For alternate proofs, see [5], [21]. □

This theorem allows us to freely interpret a spin network S as a function. The computation of f_S will be easier once the functions for a few simple spin networks are known.

Proposition 3.3. *As spin network functions,*

- (1) the swap $\times: V \otimes V \rightarrow V \otimes V$ takes $v \otimes w \mapsto w \otimes v$;
- (2) the vertex on a straight line $\vdash: V \rightarrow V$ takes $v \mapsto \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} v$;
- (3) the vertex on a cup $\smile: \mathbb{C} \rightarrow V \otimes V$ takes $1 \mapsto e_1 \otimes e_2 - e_2 \otimes e_1$;
- (4) with opposite ciliations, $\frown = -\smile$, $\vdash = -\dashv$, and $\smile = -\smile$.

Proof. First (1) is the statement that crossings change only the order of the outputs. Statement (2) follows from, for $v = \begin{bmatrix} v^1 \\ v^2 \end{bmatrix}$:

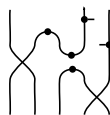
$$\begin{aligned} \vdash(v) &= \left(\frown \mid \right) \circ \left(\mid \smile \right) (v) = \left(\frown \mid \right) (v \otimes e_1 \otimes e_1 + v \otimes e_2 \otimes e_2) \\ &= \det[v \ e_1]e_1 + \det[v \ e_2]e_2 = -v^2 e_1 + v^1 e_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} v. \end{aligned}$$

Statement (3) is computed similarly, using the decomposition

$$\cup = \left(\left| \begin{array}{c} \curvearrowright \\ \cup \end{array} \right| \right) \circ (\cup \cup).$$

Finally, (4) follows from the observation $\curvearrowright = \bigcirc = -\curvearrowleft$, which has already been demonstrated. □

Given these facts, the function of the earlier example can be computed. The reader may check that the function given by



takes $e_1 \otimes e_2 \otimes e_2 \otimes e_1 \otimes e_2$ to $-e_2 \otimes e_2 \otimes e_2$.

The maps \curvearrowright and \cup are unnecessary for trace computations, and so we make the following assumption:

Convention 3.4. For the remainder of this chapter, the set of ciliated vertices will coincide exactly with the set of local extrema. The ciliations are usually omitted, with the understanding that

$$\cup = \cup: 1 \mapsto e_1 \otimes e_2 - e_2 \otimes e_1$$

and

$$\curvearrowright = \curvearrowright: v \otimes w \mapsto \det[v \ w].$$

Under this assumption, each straightened kink $\curvearrowright \leftrightarrow \left| \begin{array}{c} \curvearrowright \\ \cup \end{array} \right|$ introduces a sign, and more generally

$$\left(\begin{array}{c} \curvearrowright \\ \cup \end{array} \right)^n = (-1)^n \left(\begin{array}{c} \curvearrowright \\ \cup \end{array} \right)^n.$$

Thus, any diagram manipulation in which kinks are straightened must be done carefully.

Spin networks exhibit considerable symmetry, which can be exploited for calculations. For example:

Proposition 3.5. Let \mathbb{S} be a spin network with function $f_{\mathbb{S}}: V^{\otimes k_i} \rightarrow V^{\otimes k_o}$. Denote its images under reflection through vertical and horizontal lines by $\overleftarrow{\mathbb{S}}$ and \mathbb{S}^\updownarrow , respectively. Then

$$f_{\overleftarrow{\mathbb{S}}} = (-1)^{|\mathbb{S}_v|} \overleftarrow{f_{\mathbb{S}}}: V^{\otimes k_i} \rightarrow V^{\otimes k_o},$$

where $|\mathbb{S}_v|$ is the number of local extrema in the diagram and \overleftarrow{f} indicates that the ordering of inputs and outputs is reversed. Also, $f_{\mathbb{S}\dagger} = (f_{\mathbb{S}})^*$ where

$$(f_{\mathbb{S}})^*(v_1 \otimes \cdots \otimes v_{k_i}) = \sum_{e_b \in \mathcal{B}(V^{\otimes k_i})} (f_{\mathbb{S}}(e_b) \cdot (v_1 \otimes \cdots \otimes v_{k_o})) e_b,$$

where \cdot indicates the dot product with respect to the standard basis for $V^{\otimes k_o}$ and $\mathcal{B}(V^{\otimes k_i})$ is the basis for $V^{\otimes k_i}$. That is, $(f_{\mathbb{S}})^*$ and $f_{\mathbb{S}}$ are dual with respect to the standard inner product on V .

Proof. The first statement is an extension of the fact that reflecting \curvearrowright through a vertical line gives $\curvearrowleft = -\curvearrowright$.

For the second statement, consider $\mathbb{S} = \cup$. If $v_i = \begin{bmatrix} v_i^1 \\ v_i^2 \end{bmatrix}$, then

$$\begin{aligned} (f_{\mathbb{S}})^*(v_1 \otimes v_2) &= \cup(1) \cdot (v_1 \otimes v_2) = (e_1 \otimes e_2 - e_2 \otimes e_1) \cdot (v_1 \otimes v_2) \\ &= v_1^1 v_2^2 - v_1^2 v_2^1 = \det[v_1 \ v_2] = \curvearrowright(v_1 \otimes v_2). \end{aligned}$$

This computation, together with the corresponding one for $\mathbb{S} = \curvearrowright$, are sufficient to prove the second claim (see [25] for details). □

The next theorem, which follows from Proposition 3.5, describes how to apply these symmetries to relations among spin networks:

Theorem 3.6 (Spin network reflection theorem). *A relation*

$$\sum_m \alpha_m \mathbb{S}^m = 0$$

among some collection of spin networks $\{\mathbb{S}^m\}$ is equivalent to the same relation for the vertically reflected spin networks $\{\mathbb{S}^{\dagger m}\}$ and (up to sign) for the horizontally reflected spin networks $\{\overleftarrow{\mathbb{S}}^m\}$, that is

$$\sum_m \alpha_m \mathbb{S}^{\dagger m} = 0 \quad \text{and} \quad \sum_m \alpha_m (-1)^{|\mathbb{S}_v^m|} \overleftarrow{\mathbb{S}}^m = 0.$$

3.2 Basic diagram manipulations

In this section, we describe the *spin network calculus*, which governs diagram manipulations.

Proposition 3.7. *Any spin network can be expressed as a sum of diagrams with no crossings or loops. In particular,*

$$\curvearrowright = \left| \left| - \cup \right. \right.; \quad \bigcirc \mathbb{S} = \text{tr}(I)\mathbb{S} = 2\mathbb{S}. \tag{3.1}$$

The proof is given in [25]. The first of these relations is called the *fundamental binor identity*, and represents a fundamental type of structure in mathematics; it is the core concept in defining both the *Kauffman bracket skein module* in knot theory [4] and the *Poisson bracket* on the set of loops on a surface, which Goldman describes in [16]. It can also be identified with the *characteristic equation* for 2×2 matrices [25], [30].

Since 2×2 matrices act on V , the definition of spin networks may be extended to allow matrices to act on diagrams: \circlearrowleft is the action $v \mapsto \mathbf{x} \cdot v$. The corresponding action on the tensor product $V^{\otimes n}$ is represented by

$$\circlearrowleft \cdot \circlearrowleft \cdots \circlearrowleft (v_1 \otimes \cdots \otimes v_n) = \mathbf{x} v_1 \otimes \cdots \otimes \mathbf{x} v_n.$$

The matrices $\mathbf{x} \in SL(2, \mathbb{C})$ of interest in this chapter satisfy the following special property:

Proposition 3.8. *The spin network component maps $\left| \right.$, $\cup = \cup$, and $\cap = \cap$, and therefore all spin networks, are equivariant under the natural action of $SL(2, \mathbb{C})$ on V described above.*

Proof. The case for the identity $\left| \right.$ is clear, while

$$\begin{aligned} \circlearrowleft \circlearrowleft (v \otimes w) &= \det[\mathbf{x} v \ \mathbf{x} w] = \det(\mathbf{x} \cdot [v \ w]) \\ &= \det(\mathbf{x}) \cdot \det[v \ w] = 1 \cdot \det[v \ w] = \cap (v \otimes w) \end{aligned}$$

shows that $\cap \circ \mathbf{x} = \cap = \mathbf{x} \circ \cap$.

The proof for \cup follows by reflecting this relation. □

This means that matrices in such a diagram can “slide across” a vertex (local extremum) by simply inverting the matrix, so that

$$\text{if } \circlearrowleft = \mathbf{x}^{-1} \in SL(2, \mathbb{C}), \quad \text{then } \circlearrowleft \cup = \cup \circlearrowleft.$$

For a general matrix $\mathbf{x} \in M_{2 \times 2}$, the determinant is introduced in such relations since

$$\circlearrowleft \circlearrowleft = \det \left(\circlearrowleft \right) \cup. \text{ If } \mathbf{x} \text{ is invertible, this implies}$$

$$\circlearrowleft \cup = \det \left(\circlearrowleft \right) \circlearrowleft.$$

A *closed* spin network with one or more matrices is called a *trace diagram*, and may be identified with a map $G \times \cdots \times G \rightarrow \mathbb{C}$. One of the primary motivations for this chapter is the study of invariance properties of such maps. The simplest cases are given by:

Proposition 3.9. For $x \in M_{2 \times 2}$ and $\mathbb{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,

$$\bigcirc = 2 = \text{tr}(\mathbb{I}); \quad \textcircled{x} = \text{tr}(x); \quad \textcircled{x} \textcircled{x} = \det(x) \cdot \text{tr}(\mathbb{I}). \quad (3.2)$$

3.3 Symmetrizers and irreducible representations

Another important $\text{SL}(2, \mathbb{C})$ -equivariant map is the symmetrizer.

Definition 3.10. The *symmetrizer* $\begin{array}{c} ||\dots \\ |n| \\ ||\dots \end{array}$: $V^{\otimes n} \rightarrow V^{\otimes n}$ is the map

$$v_1 \otimes v_2 \otimes \dots \otimes v_n \mapsto \frac{1}{n!} \sum_{\sigma \in \Sigma_n} v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \dots \otimes v_{\sigma(n)}, \quad (3.3)$$

where $v_i \in V$ and Σ_n is the group of permutations on n elements.

For example,

$$\begin{array}{c} | \\ | \\ \blacksquare \\ | \\ | \end{array} = \frac{1}{2} \left(\begin{array}{c} | \\ | \\ | \\ | \end{array} + \begin{array}{c} \diagup \\ | \\ \diagdown \end{array} \right) = \begin{array}{c} | \\ | \\ | \\ | \end{array} - \frac{1}{2} \left(\begin{array}{c} \cup \\ \cup \end{array} \right);$$

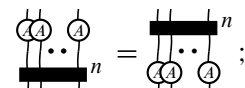
$$\begin{array}{c} | \\ | \\ | \\ \blacksquare \\ | \\ | \end{array} = \frac{1}{6} \left(\begin{array}{c} | \\ | \\ | \\ | \\ | \end{array} + \begin{array}{c} \diagup \\ | \\ \diagdown \end{array} + \begin{array}{c} \diagdown \\ | \\ \diagup \end{array} + \begin{array}{c} \diagup \\ \diagdown \end{array} + \begin{array}{c} \diagdown \\ \diagup \end{array} + \begin{array}{c} \cup \\ \cup \end{array} \right) \\ = \begin{array}{c} | \\ | \\ | \\ | \\ | \end{array} - \frac{2}{3} \left(\begin{array}{c} \cup \\ \cup \end{array} + \begin{array}{c} \cup \\ \cup \end{array} \right) - \frac{1}{3} \left(\begin{array}{c} \diagup \\ \diagdown \end{array} + \begin{array}{c} \diagdown \\ \diagup \end{array} \right)$$

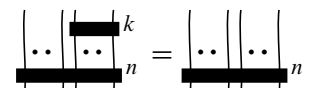
Note that the crossings are removed by applying the fundamental binor identity.

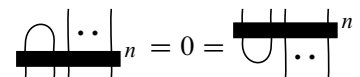
The defining equation (3.3) of $\begin{array}{c} ||\dots \\ |n| \\ ||\dots \end{array}$ should look familiar: its image is a subspace of $V^{\otimes n}$ isomorphic to the n th symmetric power $\text{Sym}^n V$, and thus it can be thought of as either the projection $\pi : V^{\otimes n} \rightarrow \text{Sym}^n V$ or as the inclusion $i : \text{Sym}^n V \rightarrow V^{\otimes n}$ (see [15], page 473).

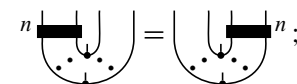
What does this mean for us? If a diagram from $V^{\otimes k_i}$ to $V^{\otimes k_o}$ has symmetrizers at its top and bottom, it can be thought of as a map between V_{k_i} and V_{k_o} . We freely interpret such spin networks as maps between tensor powers of these irreducible $\text{SL}(2, \mathbb{C})$ -representations.

Proposition 3.11 (Basic symmetrizer properties).

Invariance:  (3.4)

stacking relation:  (3.5)

capping/cupping:  (3.6)



symmetrizer sliding:  (3.7)

Proof. The first relation (3.4) is evident if one expands the symmetrizer in terms of permutations, since permutations are $SL(2, \mathbb{C})$ -equivariant.

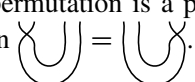
The *stacking relation* is the statement that symmetrizing the last k elements of a symmetric tensor has no effect, since they are already symmetric.


For the *capping and cupping relations*, notice that

$$\cap \circ \begin{array}{|c|} \hline 2 \\ \hline \end{array} (v \otimes w) = \cap (\frac{1}{2}(v \otimes w + w \otimes v)) = \frac{1}{2}(\det[v \ w] + \det[w \ v]) = 0.$$

This implies the general case because, by the stacking relation, one may insert  between \cap and . The other case is similar.

There are a number of ways to demonstrate (3.7). It follows by reflection (Proposition 3.5) or as a special case of $SL(2, \mathbb{C})$ -equivariance, since $\downarrow = \uparrow = \begin{array}{|c|} \hline \oplus \\ \hline \end{array}$ for

$\begin{array}{|c|} \hline \oplus \\ \hline \end{array} = g = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in SL(2, \mathbb{C})$. More directly, expand the symmetrizer into a sum of permutations. Since each permutation is a product of transpositions, then (3.7) follows from the simple relation . See [25] for more details. \square

We now move on to some more involved relations among symmetrizers. Although it is easy to write down an arbitrary  in terms of permutations, it is usually rather difficult to write it down in terms of diagrams without crossings (the Temperley-Lieb algebra). The next two propositions describe how to do exactly this. As such, they are a fundamental step in the proof of Theorem 5.6, which permits a fast computation of rank two central functions.

Proposition 3.12. *The symmetrizer $\begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}$ satisfies:*

$$\begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^n = \begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} - \binom{n-1}{n} \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} - \binom{n-2}{n} \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} + \dots \\ - \binom{n-i}{n} \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \cup \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} - \dots - \binom{1}{n} \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \cup \\ \cup \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1}. \tag{3.8}$$

Proof. If Σ_n is the group of permutations on the set $N_n = \{1, 2, \dots, n\}$, then

$$|\Sigma_n| = |N_n| |\Sigma_{n-1}|.$$

Interpret $|\Sigma_n|$ as the number of ways to arrange n people in a line. To do this, one may first select someone to be at the front of the line ($|N_n|$ choices), and then rearrange the remaining $n - 1$ people ($|\Sigma_{n-1}|$ choices).

In diagram form, the selection of someone to head the line corresponds to one of the diagrams

$$\begin{array}{c} | \\ | \\ | \end{array} \bullet \cdot |, \begin{array}{c} | \\ | \\ | \end{array} \cup \bullet \cdot |, \begin{array}{c} | \\ | \\ | \end{array} \cup \cup \bullet \cdot |, \dots, \begin{array}{c} | \\ | \\ | \end{array} \cup \cup \cup \bullet \cdot |, \dots, \begin{array}{c} | \\ | \\ | \end{array} \cup \cup \cup \cup \bullet \cdot |.$$

The arrangement of the remaining people corresponds to $\begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1}$. Thus, the diagrammatic form of the above interpretation is:

$$\begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^n = \frac{1}{n} \begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} \circ \left(\begin{array}{c} | \\ | \\ | \end{array} \bullet \cdot | + \begin{array}{c} | \\ | \\ | \end{array} \cup \bullet \cdot | + \begin{array}{c} | \\ | \\ | \end{array} \cup \cup \bullet \cdot | + \dots + \begin{array}{c} | \\ | \\ | \end{array} \cup \cup \cup \bullet \cdot | + \dots + \begin{array}{c} | \\ | \\ | \end{array} \cup \cup \cup \cup \bullet \cdot | \right).$$

Now, use the binor identity to remove crossings. Most of the resulting terms disappear, since any term whose cups are not in the ‘first position’ on top will vanish due to the *capping relation*. In particular:


$$\begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} \circ \begin{array}{c} \cup \\ \cup \\ \cup \end{array} \bullet \cdot | = \begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} - \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} - \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} - \dots - \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \cup \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1},$$

where i is the number of ‘kinks’ $\begin{array}{c} \cup \\ \cup \end{array}$ in $\begin{array}{c} \cup \\ \cup \\ \cup \end{array} \bullet \cdot |$ or 1 plus the number of kinks in $\begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \cup \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1}$. Finally, group the number of terms on the righthand side with the same number of kinks together: there will be $n - i - 1$ terms with i kinks. \square

Proposition 3.13. *$\begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}$ also satisfies the recurrence relations:*

$$\begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^n = \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} - \binom{n-i}{n} \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1}; \tag{3.9}$$

$$\begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^n = \begin{array}{c} \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1} - \binom{n-1}{n} \begin{array}{c} \bullet\bullet\bullet \\ | \\ \cup \\ \bullet\bullet\bullet \\ | \\ \bullet\bullet\bullet \end{array}^{n-1}. \tag{3.10}$$

Proof. Compose Relation (3.8) with $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^i \otimes \begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-i}$. This has no effect on the left-hand side, by the *stacking relation*. On the righthand side, all but one of the terms with a cap on the bottom vanish, due to the *capping relation*, since they will cap off either the $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^i$ or the $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-i}$. The one term which remains ‘caps between’ these two symmetrizers. The coefficient is $-\left(\frac{n-i}{n}\right)$ since in recurrence (3.8), i is equal to one more than the number of kinks  in $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-1}$.

Relation (3.10) is a special case of (3.9) for $i = 1$. □

The next relations follow directly from these recurrences:

Proposition 3.14 (Looping relations).

$$\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^n = \left(\frac{n+1}{n}\right) \begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-1}. \tag{3.11}$$

When k strands of $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^n$ are closed off:

$$k \left\{ \begin{array}{|c} \cdots \\ \hline \cdots \end{array}^n \right\} = \left(\frac{n+1}{n-k+1}\right) \begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-k}. \tag{3.12}$$

$$\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^n = n + 1. \tag{3.13}$$

Proof. Close off the left strand in (3.10) above. Then, $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^n$, $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-1}$, and $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-1}$ become $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^n$, $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-1}$ and $2 \begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-1}$ and $\begin{array}{|c} \cdots \\ \hline \cdots \end{array}^{n-1}$, respectively. Now collect terms to get (3.11), and proceed to (3.12) by applying the first relation k times. Finally, (3.13) is a special case of (3.12) with $k = n$. □

3.4 Symmetrizers and trivalent spin networks

Recall the Clebsch–Gordan decomposition (Proposition 2.3):

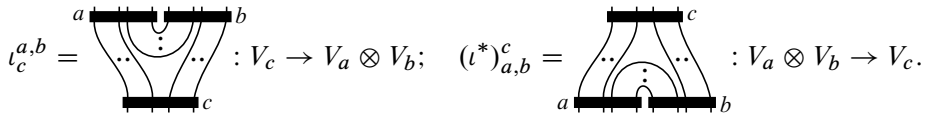
$$V_a \otimes V_b \cong \bigoplus_{c \in [a,b]} V_c, \quad [a, b] = \{a + b, a + b - 2, \dots, |a - b|\}.$$

The requirement $c \in [a, b]$ is equivalent to the following symmetric condition:

Definition 3.15. A triple (a, b, c) of nonnegative integers is *admissible*, and we write $c \in [a, b]$, if

$$\frac{1}{2}(-a + b + c), \quad \frac{1}{2}(a - b + c), \quad \frac{1}{2}(a + b - c) \in \mathbb{N}. \tag{3.14}$$

Two maps arise from the Clebsch–Gordan decomposition: an injection $\iota_c^{a,b} : V_c \rightarrow V_a \otimes V_b$ and a projection $(\iota^*)_{a,b}^c : V_a \otimes V_b \rightarrow V_c$. Both have simple diagrammatic depictions [5]:



The admissibility condition (3.14) is the requirement that there is a nonnegative number of strands connecting each pair of symmetrizers. These “strand numbers” appear frequently in diagram manipulations, and will be referenced by the Greek letters α, β, γ .

Convention 3.16. Given an admissible triple (a, b, c) , denote by α, β , and γ the total number of strands connecting V_b to V_c , V_a to V_c , and V_a to V_b , respectively. Also, denote by δ the total number of strands in the diagram. Then:

$$\alpha = \frac{1}{2}(-a + b + c), \quad \beta = \frac{1}{2}(a - b + c), \quad \gamma = \frac{1}{2}(a + b - c); \quad \delta = \frac{1}{2}(a + b + c).$$

Note that (a, b, c) is admissible if and only if $\alpha, \beta, \gamma \in \mathbb{N}$.

Convention 3.17. Because the maps $\iota_c^{a,b}$ and $(\iota^*)_{a,b}^c$ will be so important for the remainder of this chapter, we introduce a notation which simplifies their depiction. Let n lines with a symmetrizer be represented by one **thick** line labelled n , so that

$$\left. \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right\} \equiv \begin{array}{c} \vdots \\ \text{---} \\ \vdots \end{array} \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array}$$

Definition 3.18. A *trivalent spin network* S is a graph drawn on the plane with vertices of degree ≤ 3 and edges labelled by positive integers such that:

- 2-vertices are ciliated and coincide with local extrema;
- 3-vertices are drawn ‘up’ Υ or ‘down’ Λ ;
- any two edges meeting at a 2-vertex have the same label;
- the three labels adjacent to any vertex form an admissible triple.

If there are m input edges with labels l_i for $i = 1, \dots, m$ and n output edges with labels l'_i for $i = 1, \dots, n$, the network is identified with a map between tensor products of irreducible $SL(2, \mathbb{C})$ representations,

$$f_S : V_{l_1} \otimes \dots \otimes V_{l_m} \rightarrow V_{l'_1} \otimes \dots \otimes V_{l'_n}.$$

This map is computed by identifying S with a regular spin network using the following identifications:

$$\begin{aligned}
 \int^n &\equiv \left| \begin{array}{c} \vdots \\ \hline \vdots \end{array} \right|^n, & \int^n &\equiv \left| \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right|^n, & \cup^n &= \cup^n \equiv \left(\begin{array}{c} \cup \\ \vdots \\ \cup \end{array} \right)^n, \\
 \begin{array}{c} a \\ \diagdown \\ \diagup \\ c \end{array} b &\equiv \begin{array}{c} a \quad b \\ \hline \vdots \\ \hline c \end{array}, & \begin{array}{c} c \\ \diagdown \\ \diagup \\ a \end{array} b &\equiv \begin{array}{c} c \\ \hline \vdots \\ \hline a \quad b \end{array}.
 \end{aligned}$$

Note that ciliations are normally chosen to be on the local extrema, and degree-3 vertices, when expanded, also have a number of ciliated vertices. The need to keep track of these ciliations makes diagram manipulation a more delicate operation.

3.5 Trivalent diagram manipulations

This section describes in detail the relations which may be used to manipulate trivalent spin networks. For the remainder of this chapter, we assume that all sets of labels incident to a common vertex in a diagram are admissible. Moreover, whenever we sum over a label in a diagram, the sum is taken over all possible values of that label which make the requisite triples in the diagram admissible.

Any closed trivalent spin network may be interpreted as a constant. The simplest such diagrams are given by

Proposition 3.19. Let $\Theta(a, b, c) = \begin{array}{c} a \\ \diagdown \\ \diagup \\ b \end{array}^c$ and $\Delta(c) = \bigcirc^c$. Then $\Theta(a, b, c)$ is symmetric in $\{a, b, c\}$ and explicitly (recall the $\alpha, \beta, \gamma, \delta$ given in Convention 3.16):

$$\Delta(c) = c + 1 = \dim(V_c); \tag{3.15}$$

$$\Theta(a, b, c) = \frac{\left(\frac{-a+b+c}{2}\right)! \left(\frac{a-b+c}{2}\right)! \left(\frac{a+b-c}{2}\right)! \left(\frac{a+b+c+2}{2}\right)!}{a!b!c!} = \frac{\alpha!\beta!\gamma!(\delta+1)!}{a!b!c!}; \tag{3.16}$$

$$\Theta(1, a, a + 1) = \Delta(a + 1) = a + 2. \tag{3.17}$$

Proof. The first equation (3.15) is a consequence of the *looping relation* (3.11). That $\Theta(1, a, a + 1) = \Delta(a + 1)$ is a consequence of the *stacking relation*, and demonstrates (3.17). We refer the reader to [5] for the $\Theta(a, b, c)$ formula. \square

Ratios of Δ and Θ show up in the next two propositions, which tell us how to “pop bubbles” and how to “fuse together” two thick edges. The first demonstrates the usefulness of Schur’s Lemma (Proposition 2.4) in diagrammatic techniques.

Proposition 3.20 (Bubble identity). $\begin{array}{c} c \\ \diagdown \\ \diagup \\ d \end{array}^b = \left(\frac{\Theta(a,b,c)}{\Delta(c)} \int^c \right) \delta_{cd}$, where δ_{cd} is the Kronecker delta.

Proof. Schur’s Lemma requires $\begin{matrix} a \\ \diagup \\ d \\ \diagdown \\ b \end{matrix}^c = C \int^c \delta_{cd}$ for some constant C , since $\begin{matrix} a \\ \diagup \\ d \\ \diagdown \\ b \end{matrix}^c$ is a map between irreducible representations. This equation remains true if we “close off” the diagrams, giving:

$$\begin{matrix} a \\ \diagup \\ \circlearrowleft \\ \diagdown \\ b \end{matrix}^c = C \bigcirc^c \implies C = \frac{\Theta(a, b, c)}{\Delta(c)}. \quad \square$$

Proposition 3.21 (Fusion identities).

$$\begin{aligned} \begin{matrix} a \\ \diagup \\ \diagdown \\ b \end{matrix}^b &= \sum_{c \in [a, b]} \left(\frac{\Delta(c)}{\Theta(a, b, c)} \right) \begin{matrix} a \\ \diagup \\ c \\ \diagdown \\ b \end{matrix} \\ \begin{matrix} a \\ \diagdown \\ \diagup \\ b \end{matrix}^b &= \sum_{c \in [a, b]} (-1)^{\frac{1}{2}(a-b+c)} \left(\frac{\Delta(c)}{\Theta(a, b, c)} \right) \begin{matrix} a \\ \diagdown \\ c \\ \diagup \\ b \end{matrix}. \end{aligned}$$

Proof. Maps of the form $\begin{matrix} a \\ \diagup \\ c \\ \diagdown \\ b \end{matrix}$ for $c \in [a, b]$ form a basis for the space of $SL(2, \mathbb{C})$ -equivariant maps $V_a \otimes V_b \rightarrow V_a \otimes V_b$ [5]. Thus, we may express the first diagram as a linear combination:

$$\begin{matrix} a \\ \diagup \\ \diagdown \\ b \end{matrix}^b = \sum_{c \in [a, b]} C(c) \begin{matrix} a \\ \diagup \\ c \\ \diagdown \\ b \end{matrix}.$$

For a fixed $d \in [a, b]$, the constant $C(d)$ is computed by composing this expression with $\begin{matrix} a \\ \diagdown \\ d \\ \diagup \\ b \end{matrix}$:

$$\begin{aligned} \begin{matrix} a \\ \diagdown \\ d \\ \diagup \\ b \end{matrix} \circ \begin{matrix} a \\ \diagup \\ \diagdown \\ b \end{matrix}^b &= \sum_{c \in [a, b]} C(c) \begin{matrix} a \\ \diagdown \\ d \\ \diagup \\ b \end{matrix} \circ \begin{matrix} a \\ \diagup \\ c \\ \diagdown \\ b \end{matrix} \\ &= \sum_{c \in [a, b]} C(c) \left(\frac{\Theta(a, b, c)}{\Delta(c)} \right) \begin{matrix} a \\ \diagdown \\ d \\ \diagup \\ b \end{matrix} \circ \int^c \delta_{cd} \\ &= C(d) \left(\frac{\Theta(a, b, d)}{\Delta(d)} \right) \begin{matrix} a \\ \diagdown \\ d \\ \diagup \\ b \end{matrix} \implies C(d) = \frac{\Delta(d)}{\Theta(a, b, d)}. \end{aligned}$$

The second equation follows from the first and from Proposition 3.22 below:

$$\begin{aligned}
 \text{X}_{a,b}^a &= (-1)^b \mathcal{N}_{a,b}^a = \sum_{c \in [a,b]} (-1)^b \left(\frac{\Delta(c)}{\Theta(a,b,c)} \right) \text{Diagram}_1 \\
 &= \sum_{c \in [a,b]} (-1)^{\frac{1}{2}(a-b-c)} \left(\frac{\Delta(c)}{\Theta(a,b,c)} \right) \text{Diagram}_2 \\
 &= \sum_{c \in [a,b]} (-1)^{\frac{1}{2}(a-b+c)} \left(\frac{\Delta(c)}{\Theta(a,b,c)} \right) \text{Diagram}_3. \quad \square
 \end{aligned}$$

The identity $\text{Diagram}_1 = \text{Diagram}_2$ gives rise to the following compendium of sign changes through diagram manipulations:

Proposition 3.22.

$$\text{Diagram}_4^n = (-1)^n \text{Diagram}_5^n; \tag{3.18}$$

$$\text{Diagram}_6^c = (-1)^{\frac{1}{2}(a+b-c)} \text{Diagram}_7^c; \tag{3.19}$$

$$\text{Diagram}_8^c = (-1)^{\frac{1}{2}(-a+b+c)} \text{Diagram}_9^c; \tag{3.20}$$

$$\text{Diagram}_{10}^{a,b,c,d,e} = (-1)^{\frac{1}{2}(a+b+c+d-2e)} \text{Diagram}_{11}^{a,b,c,d,e}; \tag{3.21}$$

$$(-1)^{\frac{1}{2}(a+c)} \text{Diagram}_{12}^{a,b,c,d,e} = (-1)^{\frac{1}{2}(b+d)} \text{Diagram}_{13}^{a,b,c,d,e}; \tag{3.22}$$

$$\text{Diagram}_{14}^{a,b,c,d,e} = (-1)^{b+d-e} \text{Diagram}_{15}^{a,b,c,d,e}. \tag{3.23}$$

Proof. First, (3.18) is just a restatement of $\text{Diagram}_{16}^n = (-1)^n \text{Diagram}_{17}^n$, and (3.19) follows directly from the Proposition 3.5, since Diagram_7^c contains $\gamma = \frac{1}{2}(a + b - c)$ local

extrema and $\text{Diagram}_{18}^c = \text{Diagram}_{19}^c$.

For (3.20), notice that in the simplest case

$$\begin{array}{c} \text{---} \\ \diagup \quad \diagdown \\ \text{---} \end{array} = - \begin{array}{c} \text{---} \\ \diagdown \quad \diagup \\ \text{---} \end{array},$$

the negative sign comes from the strand on top of the diagram. Similarly, the general case for transforming $\begin{array}{c} c \\ \diagup \quad \diagdown \\ a \quad b \end{array}$ into $\begin{array}{c} c \\ \diagdown \quad \diagup \\ a \quad b \end{array}$ has a sign for each strand between b and c , giving $(-1)^\alpha = (-1)^{\frac{1}{2}(-a+b+c)}$. This identity is used twice to give (3.21).

Finally, (3.22) follows from:

$$\begin{array}{c} a \\ \diagup \quad \diagdown \\ d \quad c \end{array} \begin{array}{c} e \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagdown \quad \diagup \\ \quad \quad \end{array} = (-1)^e \begin{array}{c} a \\ \diagup \quad \diagdown \\ d \quad c \end{array} \begin{array}{c} e \\ \diagdown \quad \diagup \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagup \quad \diagdown \\ \quad \quad \end{array} = (-1)^{e+\frac{1}{2}(d+e-a+b+e-c)} \begin{array}{c} a \\ \diagup \quad \diagdown \\ d \quad c \end{array} \begin{array}{c} e \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagdown \quad \diagup \\ \quad \quad \end{array},$$

and (3.23) is given by combining (3.21) and (3.22). □

The above relations permit the definition of a “ $\frac{\pi}{4}$ -reflection” on certain types of diagrams, which will be important later:


Proposition 3.23. *If a relation consists entirely of terms of the form $\begin{array}{c} a \\ \diagup \quad \diagdown \\ d \quad c \end{array} \begin{array}{c} e \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagdown \quad \diagup \\ \quad \quad \end{array}$ and $\begin{array}{c} a \\ \diagdown \quad \diagup \\ d \quad c \end{array} \begin{array}{c} f \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagup \quad \diagdown \\ \quad \quad \end{array}$, then one may “reflect about the line through a and c ” in the following sense:*

$$\sum_e \alpha_e \begin{array}{c} a \\ \diagup \quad \diagdown \\ d \quad c \end{array} \begin{array}{c} e \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagdown \quad \diagup \\ \quad \quad \end{array} = \sum_f \beta_f \begin{array}{c} a \\ \diagdown \quad \diagup \\ d \quad c \end{array} \begin{array}{c} f \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \iff \sum_e \alpha_e \begin{array}{c} a \\ \diagdown \quad \diagup \\ b \quad c \end{array} \begin{array}{c} e \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} d \\ \diagdown \quad \diagup \\ \quad \quad \end{array} = \sum_f \beta_f \begin{array}{c} a \\ \diagup \quad \diagdown \\ b \quad c \end{array} \begin{array}{c} f \\ \diagdown \quad \diagup \\ \quad \quad \end{array} \begin{array}{c} d \\ \diagup \quad \diagdown \\ \quad \quad \end{array}.$$


Proof. By horizontally reflecting the first relation, using Theorem 3.6,

$$\begin{aligned} \sum_e \alpha_e \begin{array}{c} a \\ \diagup \quad \diagdown \\ d \quad c \end{array} \begin{array}{c} e \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagdown \quad \diagup \\ \quad \quad \end{array} &= \sum_f \beta_f \begin{array}{c} a \\ \diagdown \quad \diagup \\ d \quad c \end{array} \begin{array}{c} f \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} b \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \\ \iff \sum_e \alpha_e (-1)^{\frac{1}{2}(a+b+c+d-2e)} \begin{array}{c} b \\ \diagup \quad \diagdown \\ c \quad d \end{array} \begin{array}{c} e \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} a \\ \diagdown \quad \diagup \\ \quad \quad \end{array} &= \sum_f \beta_f (-1)^{\frac{1}{2}(a+b+c+d-2f)} \begin{array}{c} b \\ \diagup \quad \diagdown \\ c \quad d \end{array} \begin{array}{c} f \\ \diagdown \quad \diagup \\ \quad \quad \end{array} \begin{array}{c} a \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \\ \iff \sum_e \alpha_e \begin{array}{c} b \\ \diagup \quad \diagdown \\ c \quad d \end{array} \begin{array}{c} e \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} a \\ \diagdown \quad \diagup \\ \quad \quad \end{array} &= \sum_f \beta_f \begin{array}{c} b \\ \diagdown \quad \diagup \\ c \quad d \end{array} \begin{array}{c} f \\ \diagup \quad \diagdown \\ \quad \quad \end{array} \begin{array}{c} a \\ \diagup \quad \diagdown \\ \quad \quad \end{array}, \end{aligned}$$

where the signs cancel due to the admissibility conditions.

Now, add strands to both sides, so that the right side  becomes

$$\text{Diagram} = (-1)^{b+d-f} \text{Diagram}.$$

Likewise, on the left side,  becomes $(-1)^{b+d-e} \text{Diagram}$. Once again, admissibility implies that e and f must have the same parity, so these signs cancel. \square

Two alternate versions of this proposition follow (see [25]).

Corollary 3.24.

$$\sum_e \alpha_e \text{Diagram} = \sum_f \beta_f \text{Diagram} \iff \sum_e \alpha_e \text{Diagram} = \sum_f \beta_f \text{Diagram}$$

$$\sum_e \alpha_e \text{Diagram} = \sum_f \beta_f \text{Diagram}$$

$$\iff \sum_e \alpha_e (-1)^{\frac{1}{2}(e-b)} \text{Diagram} = \sum_f \beta_f (-1)^{\frac{1}{2}(d-f)} \text{Diagram}.$$

4 Decomposition of $\mathbb{C}[G]$

The following theorem is a consequence of the “unitary trick”[10], the Peter–Weyl Theorem, and the fact that the set of matrix coefficients of G is exactly its coordinate ring [7]. We offer a self-contained constructive proof in Section 4.2, since it gives an explicit correspondence between regular functions and spin networks.

Theorem 4.1. *There is a G -module isomorphism*

$$\mathbb{C}[G] \cong \sum_{n \geq 0} V_n^* \otimes V_n.$$

4.1 Central functions

Theorem 4.1 allows $\mathbb{C}[G \times G]^G$ to be described in terms of an additive basis of class functions that have an elegant realization as spin networks. Indeed, together with the

Clebsch–Gordan decomposition, it implies

$$\begin{aligned}
\mathbb{C}[G \times G] &\cong \mathbb{C}[G] \otimes \mathbb{C}[G] \\
&\cong \left(\sum_{a \geq 0} V_a^* \otimes V_a \right) \otimes \left(\sum_{b \geq 0} V_b^* \otimes V_b \right) \\
&\cong \sum_{a \geq 0} \sum_{b \geq 0} V_a^* \otimes V_a \otimes V_b^* \otimes V_b \\
&\cong \sum_{0 \leq a, b < \infty} (V_a^* \otimes V_b^*) \otimes (V_a \otimes V_b) \\
&\cong \sum_{0 \leq a, b < \infty} \left(\sum_{i=0}^{\min(a,b)} V_{a+b-2i}^* \right) \otimes \left(\sum_{j=0}^{\min(a,b)} V_{a+b-2j} \right) \\
&\cong \sum_{\substack{0 \leq a, b < \infty \\ 0 \leq i, j \leq \min(a,b)}} V_{a+b-2i}^* \otimes V_{a+b-2j}.
\end{aligned}$$

Since the above maps are G -equivariant,

$$\mathbb{C}[G \times G]^G \cong \sum_{\substack{0 \leq a, b < \infty \\ 0 \leq i, j \leq \min(a,b)}} (V_{a+b-2i}^* \otimes V_{a+b-2j})^G. \quad (4.1)$$

By Schur’s Lemma (Proposition 2.4),

$$\dim_{\mathbb{C}} (V_{a+b-2i}^* \otimes V_{a+b-2j})^G = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

so

$$\mathbb{C}[G \times G]^G \cong \sum_{\substack{0 \leq a, b < \infty \\ 0 \leq j \leq \min(a,b)}} \text{End}(V_{a+b-2j})^G.$$

Definition 4.2. Given the above isomorphism, for each $c \in [a, b]$ (see Definition 3.15), there exists a class function $\chi^{a,b,c} \in \mathbb{C}[G \times G]^G$ which corresponds to a generating homothety (unique up to scalar) in $\text{End}(V_c)^G$. We refer to the functions $\chi^{a,b,c}$ as *central functions*.

Denote by $\mathbb{C} \chi^{a,b,c} \subset \mathbb{C}[G \times G]^G$ the linear span over \mathbb{C} of $\chi^{a,b,c}$. Then (4.1) may be rewritten as

$$\mathbb{C}[G \times G]^G \cong \sum_{\substack{0 \leq a, b < \infty \\ c \in [a,b]}} \mathbb{C} \chi^{a,b,c}.$$

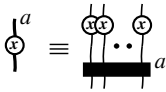
Thus, the central functions $\chi^{a,b,c}$ form an additive basis for the ring of regular functions on $\mathcal{X} = \text{Spec}_{\max}(\mathbb{C}[\mathcal{R}]^G) = \mathcal{R} // G$. In Section 5, we describe the multiplicative structure of $\mathbb{C}[G \times G]^G$ in terms of this basis.

The central functions may be described using the Clebsch–Gordan injection $\iota_c^{a,b} : V_c \hookrightarrow V_a \otimes V_b$:

$$\chi^{a,b,c}(\mathbf{x}_1, \mathbf{x}_2) = \text{tr}\left(\iota(\mathbf{c}_i^*)((\mathbf{x}_1, \mathbf{x}_2) \cdot \iota(\mathbf{c}_j))\right)_{ij},$$

where $\{\mathbf{c}_j\}$ is a basis for V_c . We will omit indices on ι when they are clear from context.

The functions $\chi^{a,b,c}$ take a natural diagrammatic form. If the matrix \mathbf{x} is represented diagrammatically by $\bigoplus : V \rightarrow V$, then its action on V_a can be represented by



A closed spin network with r different matrices is an invariant regular function $G^{\times r} \rightarrow \mathbb{C}$. In particular, since Υ and Λ are the Clebsch–Gordan injection and projection, respectively,

$$\chi^{a,b,c}(\mathbf{x}_1, \mathbf{x}_2) = \left(\begin{array}{c} a \quad b \\ \Upsilon(\mathbf{x}_1, \mathbf{x}_2) \\ c \end{array} \right) = a \left(\begin{array}{c} \mathbf{x}_1 \quad \mathbf{x}_2 \\ \Lambda \\ c \end{array} \right) b.$$

As a special case, setting $\mathbf{x}_1 = \mathbf{x}_2 = \mathbb{I}$, where \mathbb{I} is the identity matrix in G , gives $\chi^{a,b,c}(\mathbb{I}, \mathbb{I}) = \Theta(a, b, c)$.

4.2 Proof of $\mathbb{C}[G]$ decomposition theorem

Define

$$\Upsilon : \sum_{n \geq 0} V_n^* \otimes V_n \longrightarrow \mathbb{C}[G]$$

by linear extension of the mapping

$$V_{n-k}^* \otimes V_{n-l} \mapsto V_{n-k}^*(\mathbf{x} \cdot V_{n-l}),$$

where $\mathbf{x} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$ is a matrix variable.

Proposition 4.3. Υ is a well-defined G -equivariant morphism.

Proof. The image of Υ consists of regular functions since

$$\begin{aligned} V_{n-k}^*(\mathbf{x} \cdot V_{n-l}) &= V_{n-k}^*((x_{11}e_1 + x_{21}e_2)^{n-l}(x_{12}e_1 + x_{22}e_2)^l) \\ &= \sum_{\substack{i+j=k \\ 0 \leq i \leq n-l \\ 0 \leq j \leq l}} \binom{n}{k}^{-1} \binom{n-l}{i} \binom{l}{j} x_{11}^{n-l-i} x_{12}^{l-j} x_{21}^i x_{22}^j. \end{aligned}$$

Equivariance is verified by the calculation:

$$\begin{aligned} \Upsilon(g \cdot (n_{n-k}^* \otimes n_{n-l})) &= \Upsilon((g \cdot n_{n-k}^*) \otimes (g \cdot n_{n-l})) \\ &= (g \cdot n_{n-k}^*)(x \cdot (g \cdot n_{n-l})) = n_{n-k}^*((g^{-1}xg) \cdot n_{n-l}) \\ &= g \cdot n_{n-k}^*(x \cdot n_{n-l}) = g \cdot \Upsilon(n_{n-k}^* \otimes n_{n-l}). \quad \square \end{aligned}$$

There is a right action of G on $\mathbb{C}[G]$ given by $f \cdot g(x) = f(xg)$. Denote by $\mathbb{C}[G]_{\text{right}}$ the ring $\mathbb{C}[G]$ with this right action, to distinguish it from the conjugation action already imposed on $\mathbb{C}[G]$. Additionally, G acts on the left of $\text{Hom}_G(V_n, \mathbb{C}[G]_{\text{right}})$ by

$$(g \cdot \gamma)(v)(x) = \gamma_v(g^{-1}x),$$

where $\gamma_v = \gamma(v)$. This action is well-defined since

$$(g \cdot \gamma)(g' \cdot v)(x) = \gamma_{g' \cdot v}(g^{-1}x) = \gamma_v(g^{-1}xg') = ((g \cdot \gamma)(v)) \cdot g'(x).$$

The next two lemmas, whose proofs are deferred, define two additional maps which will be used to prove the theorem.

Lemma 4.4. *The map*

$$\Phi : \sum_{n \geq 0} \text{Hom}_G(V_n, \mathbb{C}[G]_{\text{right}}) \otimes V_n \longrightarrow \mathbb{C}[G]$$

defined by linearly extending the mappings $\gamma \otimes v \mapsto \gamma(v)$ is an isomorphism of G -modules.

Lemma 4.5. *Define the map $\Psi_n : V_n^* \rightarrow \text{Hom}_G(V_n, \mathbb{C}[G]_{\text{right}})$ by $w^* \mapsto \mathbf{F}_{w^*}$, where $\mathbf{F}_{w^*}(v)(x) = w^*(x \cdot v)$. Then the map*

$$\Psi : \sum_{n \geq 0} V_n^* \otimes V_n \longrightarrow \sum_{n \geq 0} \text{Hom}_G(V_n, \mathbb{C}[G]_{\text{right}}) \otimes V_n$$

given by $\Psi = \sum (\Psi_n \otimes \text{id})$ is an isomorphism of G -modules.

Assuming the above lemmas, Theorem 4.1 is equivalent to showing that the following diagram commutes:

$$\begin{array}{ccc} \sum_{n \geq 0} V_n^* \otimes V_n & \xrightarrow{\quad \Upsilon \quad} & \mathbb{C}[G] \\ & \searrow \Psi & \nearrow \Phi \\ & \sum_{n \geq 0} \text{Hom}_G(V_n, \mathbb{C}[G]_{\text{right}}) \otimes V_n & \end{array}$$

The proof of commutativity follows:

$$\Phi \circ \Psi(w^* \otimes v) = \Phi(\mathbf{F}_{w^*} \otimes v) = \mathbf{F}_{w^*}(v) = w^*(x \cdot v) = \Upsilon(w^* \otimes v). \quad \square$$

It remains to establish Lemmas 4.4 and 4.5. The proof of Lemma 4.4 requires some preliminary technical results.

Lemma 4.6. *Every regular function is contained in a finite-dimensional sub-representation of $\mathbb{C}[G]$.*

Proof of Lemma 4.6. The following $G \times G$ -action encompasses both the right and diagonal G -actions defined above. Let

$$\alpha : G \times G \times G \longrightarrow G$$

be defined by $(g_1, g_2, \mathbf{x}) \mapsto g_1 \mathbf{x} g_2^{-1}$, and further let

$$\alpha^* : \mathbb{C}[G] \longrightarrow \mathbb{C}[G \times G \times G] \cong \mathbb{C}[G]^{\otimes 3} \quad (4.2)$$

be defined by $f \mapsto f \circ \alpha$, the *pull-back* of regular functions on G to regular functions on $G \times G \times G$. For $f \in \mathbb{C}[G]$, (4.2) implies that there exist $n_f \in \mathbb{N}$ and regular functions f_i, f'_i, f''_i for $1 \leq i \leq n_f$ such that

$$\alpha^*(f) = \sum_{i=1}^{n_f} f_i \otimes f'_i \otimes f''_i.$$

Therefore

$$\alpha^*(f)(g_1^{-1}, g_2^{-1}, \mathbf{x}) = \sum_{i=1}^{n_f} f_i(g_1^{-1}) f'_i(g_2^{-1}) f''_i(\mathbf{x}).$$

On the other hand,

$$\alpha^*(f)(g_1^{-1}, g_2^{-1}, \mathbf{x}) = f(\alpha(g_1^{-1}, g_2^{-1}, \mathbf{x})) = f(g_1^{-1} \mathbf{x} g_2) = ((g_1, g_2) \cdot f)(\mathbf{x}),$$

which implies

$$(g_1, g_2) \cdot f = \sum_{i=1}^{n_f} f_i(g_1^{-1}) f'_i(g_2^{-1}) f''_i. \quad (4.3)$$

Let $(G \times G)f = \{(g_1, g_2) \cdot f : f \in G\}$ be the $G \times G$ -orbit of f , and let W_f be the linear subspace spanned over \mathbb{C} by $(G \times G)f$ in $\mathbb{C}[G]$. By (4.3), $\{f''_i\}$ is a spanning set for W_f , and so W_f is finite-dimensional. Clearly W_f is $G \times G$ -invariant, and so invariant with respect to the diagonal and right G -actions. Thus, it is a finite-dimensional sub-representation containing f . \square

Lemma 4.7. *$\mathbb{C}[G]$ is completely $G \times G$ -reducible.*

Proof of Lemma 4.7. Let \mathcal{J} be the set of direct sums of irreducible finite-dimensional sub-representations of $\mathbb{C}[G]$. \mathcal{J} is partially ordered by set inclusion and is nonempty. Thus, by Zorn's lemma there exists a maximal element $M \in \mathcal{J}$. If $M \neq \mathbb{C}[G]$, then consider any $f \notin M$. By Lemma 4.6, there exists a finite-dimensional sub-representation W_f that contains f . Let $K = \mathrm{SU}(2)$ be the maximal compact sub-

group of G . Restrict the action of $G \times G$ to $K \times K$ to find an invariant orthogonal complement to W_f in $M \cup W_f$. Denote this complement by M^\perp . Then $M^\perp \oplus W_f \in \mathcal{I}$, since $K \times K$ representations extend to $G \times G$ representations. Hence M is not maximal, which is a contradiction. Therefore $\mathbb{C}[G]$ is completely reducible with respect to the $G \times G$ -action, and so

$$\mathbb{C}[G] \cong \sum_{j \geq 0} c_j V_j,$$

where c_j is the (possibly infinite) multiplicity of V_j in $\mathbb{C}[G]$. This decomposition also holds for $\mathbb{C}[G]$ with both the right and diagonal actions since they are restrictions of the same $G \times G$ -action. \square

Proof of Lemma 4.4. By Lemma 4.7,

$$\Phi: \sum_{n \geq 0} (\text{Hom}_G(V_n, \mathbb{C}[G]_{\text{right}}) \otimes V_n) \longrightarrow \mathbb{C}[G]$$

is an isomorphism if and only if

$$\sum_{n \geq 0} \left(\sum_{j \geq 0} \text{Hom}_G(V_n, c_j V_j) \otimes V_n \right) \longrightarrow \sum_{j \geq 0} c_j V_j$$

is an isomorphism. By Schur's Lemma, this reduces to

$$\sum_{n \geq 0} (c_n \mathbb{C} \otimes V_n) \cong \sum_{n \geq 0} (\text{Hom}_G(V_n, c_n V_n) \otimes V_n) \longrightarrow \sum_{n \geq 0} c_n V_n.$$

However, this is the map sending $\sum \lambda \otimes v \mapsto \sum \lambda v$ for $\lambda \in \mathbb{C}$ and $v \in V_n$, which is canonically an isomorphism. \square

The final task is to show that Ψ is an isomorphism:

Proof of Lemma 4.5. Recall that

$$\Psi_n : V_n^* \longrightarrow \text{Hom}_G(V_n, \mathbb{C}[G]_{\text{right}})$$

was defined by $w^* \mapsto \mathbf{F}_{w^*}$, where $\mathbf{F}_{w^*}(v)(\mathbf{x}) = w^*(\mathbf{x} \cdot v)$. Ψ_n is well-defined since $\mathbf{F}_{w^*}(g \cdot v)(\mathbf{x}) = w^*(\mathbf{x} \cdot (g \cdot v)) = w^*((\mathbf{x}g) \cdot v) = \mathbf{F}_{w^*}(v)(\mathbf{x}g) = (\mathbf{F}_{w^*}(v)) \cdot g(\mathbf{x})$, and is G -equivariant because

$$\begin{aligned} \Psi_n(g \cdot w^*)(v)(\mathbf{x}) &= \mathbf{F}_{g \cdot w^*}(v)(\mathbf{x}) = (g \cdot w^*)(\mathbf{x} \cdot v) = w^*((g^{-1}\mathbf{x}) \cdot v) \\ &= \mathbf{F}_{w^*}(v)(g^{-1}\mathbf{x}) = (g \cdot \mathbf{F}_{w^*})(v)(\mathbf{x}) = g \cdot \Psi_n(w^*)(v)(\mathbf{x}). \end{aligned}$$

Since V_n^* is irreducible, Schur's Lemma implies Ψ_n is injective. We now show surjectivity. Consider $\gamma \in \text{Hom}_G(V_n, \mathbb{C}[G]_{\text{right}})$. For $\mathbb{I} \in G$, $\gamma(v)(\mathbb{I})$ is a linear functional on V_n . Hence there exists $w^* \in V_n^*$ such that $w^*(v) = \gamma(v)(\mathbb{I})$ for all $v \in V_n$. The following computation establishes that $\Psi_n(w^*) = \gamma$:

$$\mathbf{F}_{w^*}(v)(\mathbf{x}) = w^*(\mathbf{x} \cdot v) = \gamma(\mathbf{x} \cdot v)(\mathbb{I}) = (\gamma(v)) \cdot \mathbf{x}(\mathbb{I}) = \gamma(v)(\mathbb{I}\mathbf{x}) = \gamma(v)(\mathbf{x}).$$

Therefore Ψ_n is an isomorphism and so is $\Psi = \sum (\Psi_n \otimes \mathrm{id})$:

$$\sum_{n \geq 0} V_n^* \otimes V_n \cong \sum_{n \geq 0} (\mathrm{Hom}_G(V_n, \mathbb{C}[G]_{\mathrm{right}}) \otimes V_n). \quad \square$$

4.3 Ring structure of $\mathbb{C}[G]^G$

We have established

$$\mathbb{C}[G] \cong \sum_{n \geq 0} V_n^* \otimes V_n.$$

By Schur's Lemma and the fact that $V_n^* \otimes V_n \cong \mathrm{End}(V_n)$,

$$\mathbb{C}[G]^G \cong \sum_{n \geq 0} (V_n^* \otimes V_n)^G \cong \sum_{n \geq 0} \mathbb{C} \chi^n,$$

where $\chi^n \in \mathrm{End}(V_n)^G$ is a multiple of the identity.

The isomorphism $\mathrm{End}(V_n) \rightarrow V_n^* \otimes V_n$ is given by

$$n_{n-l}(n_{n-k})^T \mapsto \binom{n}{k} n_{n-k}^* \otimes n_{n-l}.$$

Therefore, the central function χ^n corresponds to an invariant function in $\mathbb{C}[G]^G$ by

$$\chi^n = \sum_{i=0}^n n_i (n_i)^T \mapsto \sum_{i=0}^n \binom{n}{i} n_i^* \otimes n_i \xrightarrow{\Upsilon} \sum_{i=0}^n \binom{n}{i} n_i^*(\mathbf{x} \cdot n_i).$$

We will freely identify χ^n with its image in $\mathbb{C}[G]^G$.

For example, the trivial representation V_0 gives $\chi^0 = 1$. The standard representation V_1 has diagonal matrix coefficients x_{11} and x_{22} , hence

$$\chi^1 = x_{11} + x_{22} = \mathrm{tr}(\mathbf{x}).$$

The remaining functions may be computed directly, or by using the following product formula:

Theorem 4.8 (Product formula).

$$\chi^a \chi^b = \sum_{c \in [a, b]} \chi^c \quad (4.4)$$

Proof. From the Clebsch–Gordan decomposition,

$$(V_a \otimes V_b)^* \otimes (V_a \otimes V_b) \cong \sum_{c, d \in [a, b]} V_c^* \otimes V_d.$$

Hence

$$\mathrm{End}(V_a \otimes V_b)^G \cong \sum_{c \in [a, b]} \mathrm{End}(V_c)^G$$

and the characters satisfy

$$\chi^a \chi^b = \chi_{(v_a \otimes v_b)} = \chi_{\oplus_c v_c} = \sum_{c \in [a,b]} \chi^c.$$

There is an alternate diagrammatic proof of this statement, which uses the fusion and bubble identities in Propositions 3.20 and 3.21. If the matrix x is represented by \textcircled{x} , then

$$\begin{aligned} \chi^a \chi^b &= \textcircled{\textcircled{x} b}^a = \sum_{c \in [a,b]} \left(\frac{\Delta(c)}{\Theta(a, b, c)} \right) \textcircled{\textcircled{x} b}^c \\ &= \sum_{c \in [a,b]} \left(\frac{\Delta(c)}{\Theta(a, b, c)} \right) \textcircled{\textcircled{x} b}^c = \sum_{c \in [a,b]} \left(\frac{\Delta(c)}{\Theta(a, b, c)} \right) \textcircled{x a}^c \\ &= \sum_{c \in [a,b]} \left(\frac{\Delta(c)\Theta(a, b, c)}{\Theta(a, b, c)\Delta(c)} \right) \textcircled{x}^c = \sum_{c \in [a,b]} \textcircled{x}^c = \sum_{c \in [a,b]} \chi^c. \quad \square \end{aligned}$$

The product formula (4.4) and the initial calculations of χ^0 and χ^1 may be used to show:

Theorem 4.9. $\mathbb{C}[G]^G \cong \mathbb{C}[t]$.

Proof. Consider the ring homomorphism $\Phi: \mathbb{C}[t] \rightarrow \mathbb{C}[G]^G$ defined by $f \mapsto f \circ \text{tr}$. Suppose $f(\text{tr}(g)) = 0$ for all $g \in G$. If $f \neq 0$, then since f has a finite number of zeros, $\text{tr}(g)$ must have a finite number of values. However,

$$\begin{bmatrix} t & 1 \\ -1 & 0 \end{bmatrix} \in G$$

for all values of t . Hence, $f = 0$ and Φ is injective. It remains to establish surjectivity. We have already shown $t \mapsto \chi^1$ and $1 \mapsto \chi^0$. Suppose $a \geq 2$ and χ^b is in the image of Φ for all $b < a$. Equation (4.4) implies $\chi^1 \chi^{a-1} = \chi^a + \chi^{a-2}$. Thus, by induction,

$$t \Phi^{-1}(\chi^{a-1}) - \Phi^{-1}(\chi^{a-2}) \mapsto \chi^a. \quad \square$$

The following closed formula for χ^n is given in [25]:

$$\chi^n(t) = \sum_{r=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^r \binom{n-r}{r} t^{n-2r}.$$

The characters χ^n may also be expressed as functions of eigenvalues, since χ^n is determined by its values on normal forms

$$\begin{bmatrix} \lambda & * \\ 0 & \lambda^{-1} \end{bmatrix} \in G.$$

Explicitly, $\begin{bmatrix} \lambda & * \\ 0 & \lambda^{-1} \end{bmatrix}$ acts on V_n by the matrix

$$\begin{bmatrix} \lambda^n & * & * & \cdots & * \\ 0 & \lambda^{n-2} & * & \cdots & * \\ \vdots & 0 & \ddots & * & * \\ 0 & \vdots & 0 & \lambda^{2-n} & * \\ 0 & 0 & \cdots & 0 & \lambda^{-n} \end{bmatrix}.$$

Hence,

$$\chi^n = \lambda^n + \lambda^{n-2} + \cdots + \lambda^{2-n} + \lambda^{-n} = \frac{\lambda^{n+1} - \lambda^{-n-1}}{\lambda - \lambda^{-1}} = [n+1]_\lambda,$$

where $[n+1]_\lambda$ is the quantized integer for $q = \lambda$.

5 Structure of $\mathbb{C}[G \times G]^G$

Recall the decomposition

$$\mathbb{C}[G \times G]^G \cong \sum_{\substack{a, b \in \mathbb{N} \\ c \in [a, b]}} \mathbb{C} \chi^{a, b, c},$$

where $\chi^{a, b, c}$ corresponds by Υ to the image of

$$\sum_{k=0}^c \mathbf{c}_k (\mathbf{c}_k)^T \mapsto \sum_{k=0}^c \binom{c}{k} \mathbf{c}_k^* \otimes \mathbf{c}_k$$

under the injection $V_c^* \otimes V_c \hookrightarrow V_a^* \otimes V_b^* \otimes V_a \otimes V_b$. This inclusion is determined by the Clebsch–Gordan injection $\iota : V_c \hookrightarrow V_a \otimes V_b$. Hence, an explicit formula for ι provides a means to compute $\chi^{a, b, c}$ directly. We freely use $\chi^{a, b, c}$ to denote its image in $\mathbb{C}[G \times G]^G$.

A few simple examples will motivate the construction of ι . For $k = 1, 2$, let $\mathbf{x}_k = [x_{ij}^k]$ be 2×2 matrix variables, and let

$$\begin{aligned} x &= \text{tr}(\mathbf{x}_1) = x_{11}^1 + x_{22}^1, \\ y &= \text{tr}(\mathbf{x}_2) = x_{11}^2 + x_{22}^2, \\ z &= \text{tr}(\mathbf{x}_1 \mathbf{x}_2^{-1}) = (x_{11}^1 x_{22}^2 + x_{22}^1 x_{11}^2) - (x_{12}^1 x_{21}^2 + x_{21}^1 x_{12}^2). \end{aligned}$$

Recall that the map $\cup : V_0 \hookrightarrow V_1 \otimes V_1$ given by

$$c_0 \mapsto a_1 \otimes b_0 - a_0 \otimes b_1$$

is invariant, using the notation defined in section 2.3. More generally, the injection $V_0 \hookrightarrow V_a \otimes V_a$ is given by

$$\cup^a : c_0 \mapsto \sum_{m=0}^a (-1)^m \binom{a}{m} a_{a-m} \otimes b_m. \quad (5.1)$$

Hence, $\chi^{0,0,0} = 1$ and $\chi^{1,1,0}$ may be computed by:

$$\begin{aligned} \chi^{1,1,0} &\mapsto c_0^* \otimes c_0 \\ &\mapsto (a_1^* \otimes b_0^* - a_0^* \otimes b_1^*) \otimes (a_1 \otimes b_0 - a_0 \otimes b_1) \\ &\mapsto (a_1^* \otimes a_1) \otimes (b_0^* \otimes b_0) - (a_0^* \otimes a_1) \otimes (b_1^* \otimes b_0) \\ &\quad - (a_1^* \otimes a_0) \otimes (b_0^* \otimes b_1) + (a_0^* \otimes a_0) \otimes (b_1^* \otimes b_1) \\ &\mapsto x_{11}^1 \otimes x_{22}^2 - x_{12}^1 \otimes x_{21}^2 - x_{21}^1 \otimes x_{12}^2 + x_{22}^1 \otimes x_{11}^2 \\ &\mapsto (x_{11}^1 x_{22}^2 + x_{22}^1 x_{11}^2) - (x_{12}^1 x_{21}^2 + x_{21}^1 x_{12}^2) = z. \end{aligned}$$

The representation V_c may be identified with a subset of $V^{\otimes c}$ via the equivariant maps

$$\begin{array}{ccc} & \text{Sym} & \\ & \curvearrowright & \\ V_c & & V^{\otimes c} \\ & \curvearrowleft & \\ & \text{Proj} & \end{array}$$

where $\text{Proj} \circ \text{Sym} = \text{id}$. Thus, when $c = a + b$, ι is given by the commutative diagram

$$\begin{array}{ccc} V^{\otimes c} & \xlongequal{\quad} & V^{\otimes a} \otimes V^{\otimes b} \\ \text{Sym} \uparrow & \curvearrowright & \downarrow \text{Proj} \otimes \text{Proj} \\ V_c & \xrightarrow{\quad \iota \quad} & V_a \otimes V_b. \end{array}$$

In particular,

$$\binom{c}{k} c_k \mapsto \sum_{\substack{0 \leq i \leq a \\ 0 \leq j \leq b \\ i+j=k}} \binom{a}{i} a_i \otimes \binom{b}{j} b_j. \quad (5.2)$$

For example, consider $\chi^{1,0,1}$. In this case, $c_0 \mapsto a_0 \otimes b_0$ and $c_1 \mapsto a_1 \otimes b_0$. Hence,

$$\begin{aligned} \chi^{1,0,1} &\mapsto c_0^* \otimes c_0 + c_1^* \otimes c_1 \mapsto (a_0^* \otimes a_0) \otimes (b_0^* \otimes b_0) + (a_1^* \otimes a_1) \otimes (b_0^* \otimes b_0) \\ &\mapsto x_{11}^1 \otimes 1 + x_{22}^1 \otimes 1 \mapsto x_{11}^1 + x_{22}^1 = x. \end{aligned}$$

A similar computation shows that $\chi^{0,1,1} \mapsto y$.

The general form of ι is determined by combining (5.1) and (5.2) in the following diagram:

$$\begin{array}{ccc}
 V_c & \xrightarrow{\iota} & V_\beta \otimes V_\alpha \\
 \downarrow \iota & \curvearrowright & \downarrow \text{id} \otimes \bigcup^\gamma \otimes \text{id} \\
 V_a \otimes V_b & \longleftarrow & V_\beta \otimes V_\gamma \otimes V_\gamma \otimes V_\alpha.
 \end{array}$$

It follows that the mapping $\iota: V_c \rightarrow V_a \otimes V_b$ is explicitly given by

$$\begin{aligned}
 \binom{c}{k} c_k &\mapsto \sum_{\substack{0 \leq i \leq \beta \\ 0 \leq j \leq \alpha \\ 0 \leq m \leq \gamma \\ i+j=k}} \binom{\beta}{i} a_i \otimes [(-1)^m \binom{\gamma}{m} a_{\gamma-m} \otimes b_m] \otimes \binom{\alpha}{j} b_j \\
 &\mapsto \sum_{\substack{0 \leq i \leq \beta \\ 0 \leq j \leq \alpha \\ 0 \leq m \leq \gamma \\ i+j=k}} (-1)^m \binom{\beta}{i} \binom{\alpha}{j} \binom{\gamma}{m} a_{i+\gamma-m} \otimes b_{j+m}.
 \end{aligned}$$

5.1 Symmetry of central functions

Our first theorem regarding central functions is a symmetry property that is essentially trivial in diagram form, despite being highly nontrivial algebraically. A portion of the Fricke–Klein–Vogt Theorem (5.12) is required to state the theorem. We begin with a diagrammatic proof of this classical result, in which the binor identity plays the role of the characteristic equation in the classical proof.

Lemma 5.1. *Each central function $\chi^{a,b,c}$ is associated to a unique polynomial $p_{a,b,c}$, denoted for all pairs $(x_1, x_2) \in G \times G$ by*

$$\chi^{a,b,c}(x_1, x_2) = p_{a,b,c}(\text{tr}(x_2), \text{tr}(x_1), \text{tr}(x_1 x_2^{-1})).$$

Proof. Expanding the symmetrizers in $\chi^{a,b,c}$ gives a collection of circles with matrix elements, each of which correspond to a product of traces of words in x_1 and x_2 , so it suffices to show that every loop can be reduced to a collection of loops containing one of x_1, x_2 , or $x_1 x_2^{-1}$.

This reduction depends entirely on the binor identity (3.1), which when composed

with $x_1 \otimes x_2 = \begin{array}{c} | \\ \bigcirc \\ | \end{array} \begin{array}{c} | \\ \bigcirc \\ | \end{array}$ gives:

$$\begin{array}{c} | \\ \bigcirc \\ | \end{array} \begin{array}{c} | \\ \bigcirc \\ | \end{array} = \begin{array}{c} | \\ \bigcirc \\ | \end{array} \begin{array}{c} | \\ \bigcirc \\ | \end{array} - \begin{array}{c} | \\ \bigcirc \\ | \end{array} \begin{array}{c} | \\ \bigcirc \\ | \end{array}. \tag{5.3}$$

Denote x_1^{-1} by \bar{x}_1 . Two special cases of (5.3) follow:

The first relation allows us to assume no loop has both x_1 and x_1^{-1} , while the second allows us to assume no loop has more than one of any matrix. The remaining cases are the traces $\text{tr}(x_1)$, $\text{tr}(x_2)$, $\text{tr}(x_1x_2)$, and $\text{tr}(x_1x_2^{-1})$. Finally, closing off (5.3) permits the reduction of $\text{tr}(x_1x_2)$:

$$\text{tr}(x_1x_2) = \text{tr}(x_1)\text{tr}(x_2) - \text{tr}(x_1x_2^{-1}). \quad \square$$

We can now prove the symmetry result. In the statement and proof below, $\sigma(\diamond_1, \diamond_2, \diamond_3)$ denotes the ordered triple $(\diamond_{\sigma(1)}, \diamond_{\sigma(2)}, \diamond_{\sigma(3)})$ obtained by applying a given permutation $\sigma \in \Sigma_3$ to the triple $(\diamond_1, \diamond_2, \diamond_3)$. This result was first outlined in [29].

Theorem 5.2 (Symmetry of central functions). *The family of polynomials*

$$\chi^{a,b,c}(x_1, x_2) = p_{a,b,c}(\text{tr}(x_2), \text{tr}(x_1), \text{tr}(x_1x_2^{-1}))$$

possesses the following symmetry:

$$p_{\sigma(a,b,c)}(y, x, z) = p_{a,b,c}(\sigma^{-1}(y, x, z)).$$

Proof. Define the following function $G \times G \times G \rightarrow \mathbb{C}$:

where the symmetrizer on the right is assumed to ‘wrap around’ to the one on the left (imagine this diagram being drawn on a cylinder). By construction this function is symmetric, in the sense that:

$$\chi_{\sigma(\alpha,\beta,\gamma)}\left(\sigma\left(\begin{array}{c} \circ \\ | \\ \circ \end{array} (x_1), \begin{array}{c} \circ \\ | \\ \circ \end{array} (x_2), \begin{array}{c} \circ \\ | \\ \circ \end{array} (x_3)\right)\right) = \chi_{\alpha,\beta,\gamma}\left(\begin{array}{c} \circ \\ | \\ \circ \end{array} (x_1), \begin{array}{c} \circ \\ | \\ \circ \end{array} (x_2), \begin{array}{c} \circ \\ | \\ \circ \end{array} (x_3)\right).$$

A central function $\chi^{a,b,c}(x_1, x_2)$ may be drawn as:

with the symmetrizers in the last two diagrams assumed to wrap around as before. Thus, $p_{a,b,c}(y, x, z) = \chi_{\alpha,\beta,\gamma}(\mathbf{x}_2, \mathbf{x}_1^{-1}, \mathbf{x}_1 \mathbf{x}_2^{-1})$ and so:

$$\begin{aligned} p_{\sigma(a,b,c)}(y, x, z) &= \chi_{\sigma(\alpha,\beta,\gamma)}(\mathbf{x}_2, \mathbf{x}_1^{-1}, \mathbf{x}_1 \mathbf{x}_2^{-1}) \\ &= \chi_{\alpha,\beta,\gamma}(\sigma^{-1}(\mathbf{x}_2, \mathbf{x}_1^{-1}, \mathbf{x}_1 \mathbf{x}_2^{-1})) \\ &= p_{a,b,c}(\sigma^{-1}(y, x, z)). \end{aligned} \quad \square$$

Table 1 contains six central functions illustrating this symmetry.

Table 1. Rank two central function symmetry.

$\chi^{1,2,3} = xy^2 - \frac{2}{3}(yz + x)$	$\chi^{3,2,1} = xz^2 - \frac{2}{3}(yz + x)$
$\chi^{2,3,1} = yz^2 - \frac{2}{3}(xz + y)$	$\chi^{1,3,2} = y^2z - \frac{2}{3}(xy + z)$
$\chi^{3,1,2} = x^2z - \frac{2}{3}(xy + z)$	$\chi^{2,1,3} = x^2y - \frac{2}{3}(xz + y)$

5.2 A recurrence relation for central functions

Define the *degree* of a central function to be:

$$\delta = \deg(\chi^{a,b,c}) = \frac{1}{2}(a + b + c).$$

We will obtain a recurrence relation for an arbitrary central function $\chi^{a,b,c}$ by manipulating diagrams to express the product

$$\text{tr}(\mathbf{x}_1) \cdot \chi^{a,b,c}(\mathbf{x}_1, \mathbf{x}_2)$$

as a sum of central functions. This formula can be rearranged to write $\chi^{a,b,c}$ as a linear combination of central functions *with lower degree*. There are three main ingredients to the diagram manipulations: the *bubble identity* and the *fusion identity* from Section 3.5, and two *recoupling formulae* which we prove in the following lemma.

Lemma 5.3. For $i = \frac{1}{2}(a + 1 - b + c)$ and appropriate triples admissible,

$$\begin{array}{c} 1 \\ \diagdown \\ c \end{array} \begin{array}{c} c-1 \\ \diagup \\ a \end{array} \begin{array}{c} \diagdown \\ b \end{array} = \begin{array}{c} 1 \\ \diagdown \\ c \end{array} \begin{array}{c} a \\ \diagup \\ a+1 \end{array} \begin{array}{c} \diagdown \\ b \end{array} + \left(\frac{a+b-c+1}{2(a+1)} \right) \begin{array}{c} 1 \\ \diagdown \\ c \end{array} \begin{array}{c} a \\ \diagup \\ a-1 \end{array} \begin{array}{c} \diagdown \\ b \end{array}; \quad (5.4)$$

$$\begin{array}{c} 1 \\ \diagdown \\ c \end{array} \begin{array}{c} c+1 \\ \diagup \\ a \end{array} \begin{array}{c} \diagdown \\ b \end{array} = \left(\frac{-a+b+c+1}{2(c+1)} \right) \begin{array}{c} 1 \\ \diagdown \\ c \end{array} \begin{array}{c} a \\ \diagup \\ a+1 \end{array} \begin{array}{c} \diagdown \\ b \end{array} + \left(\frac{(a+b+c+3)(a-b+c+1)}{4(a+1)(c+1)} \right) \begin{array}{c} 1 \\ \diagdown \\ c \end{array} \begin{array}{c} a \\ \diagup \\ a-1 \end{array} \begin{array}{c} \diagdown \\ b \end{array}. \quad (5.5)$$

Proof. Note that i is just the number of strands connecting $\begin{array}{|c|} \hline \cdots \\ \hline \end{array}^{a+1}$ to $\begin{array}{|c|} \hline \cdots \\ \hline \end{array}^c$ in $\begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a+1 \end{array} = \begin{array}{c} \diagup \\ \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a+1 \end{array}$. For (5.4), use $n = a + 1$ and i in recurrence relation (3.9) to get:

$$\begin{array}{|c|} \hline \cdots \\ \hline \end{array}^{a+1} = \begin{array}{|c|} \hline \cdots \\ \hline \end{array}^{a+1-i} \begin{array}{|c|} \hline \cdots \\ \hline \end{array}^{i-1} - \left(\frac{a+1-i}{a+1} \right) \begin{array}{|c|} \hline \cdots \\ \hline \end{array}^{a+1-i} \begin{array}{|c|} \hline \cdots \\ \hline \end{array}^i$$

Compose this equation with $\begin{array}{c} i \quad a+1-i \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \end{array}$ to get, via the *stacking relation* (3.5):

$$\begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a+1 \end{array} = \begin{array}{c} \diagup \\ \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a+1 \end{array} = \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ c-1 \end{array} - \left(\frac{a+1-i}{a+1} \right) \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a-1 \end{array},$$

which is the desired result.

To prove (5.5), notice that if we switch a and c in the previous relation, and apply a $\frac{\pi}{4}$ -reflection to the relation about the $1 \leftrightarrow b$ axis as in Proposition 3.23, then i is unchanged and the equation becomes:

$$\begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c+1 \quad b \\ \diagdown \quad \diagup \\ a \end{array} = \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a-1 \end{array} - \left(\frac{c+1-i}{c+1} \right) \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ c-1 \end{array}.$$

Rearrange this equation, and use (5.4) in its exact form to get:

$$\begin{aligned} \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c+1 \quad b \\ \diagdown \quad \diagup \\ a \end{array} &= \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a-1 \end{array} - \left(\frac{c+1-i}{c+1} \right) \left(\begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a+1 \end{array} + \left(\frac{a+1-i}{a+1} \right) \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a-1 \end{array} \right) \\ &= - \left(\frac{c+1-i}{c+1} \right) \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a+1 \end{array} + \left(1 - \frac{(a+1-i)(c+1-i)}{(a+1)(c+1)} \right) \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a-1 \end{array} \\ &= - \left(\frac{-a+b+c+1}{2(c+1)} \right) \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a+1 \end{array} + \left(\frac{(a+b+c+3)(a-b+c+1)}{4(a+1)(c+1)} \right) \begin{array}{c} 1 \\ \diagup \quad \diagdown \\ c \quad b \\ \diagdown \quad \diagup \\ a-1 \end{array}. \end{aligned}$$

To show the last computation, note that $a+1-i = \frac{1}{2}(a+b-c+1)$ and $c+1-i = \frac{1}{2}(-a+b+c+1)$, so the numerator of the last term is:

$$\begin{aligned} &4((a+1)(c+1) - (a+1-i)(c+1-i)) \\ &= 4(a+1)(c+1) - ((b+1) + (c-a))((b+1) - (c-a)) \\ &= 4(a+1)(c+1) - (b+1)^2 + (a-c)^2 \\ &= ((a+1) - (c+1))^2 + 4(a+1)(c+1) - (b+1)^2 \end{aligned}$$

$$\begin{aligned} &= ((a + 1) + (c + 1))^2 - (b + 1)^2 \\ &= (a + 1 + c + 1 + b + 1)(a + 1 + c + 1 - b - 1) \\ &= (a + b + c + 3)(a - b + c + 1). \end{aligned} \quad \square$$

The coefficients we have computed are examples of $6j$ -symbols, most easily defined to be the coefficients $\left[\begin{smallmatrix} a & b & f \\ d & c & e \end{smallmatrix} \right]'$ in the following *change of basis* equation:

$$\begin{array}{c} a & & b \\ & \diagdown & / \\ & e & \\ & / & \diagdown \\ d & & c \end{array} = \sum_{f \in [a,b] \cap [c,d]} \left[\begin{smallmatrix} a & b & f \\ d & c & e \end{smallmatrix} \right]' \cdot \begin{array}{c} a & & b \\ & / & \diagdown \\ & f & \\ & \diagdown & / \\ d & & c \end{array}.$$

We use a prime because we will need an alternate version later:

Definition 5.4. The $6j$ -symbols $\left[\begin{smallmatrix} a & b & f \\ d & c & e \end{smallmatrix} \right]$ are the coefficients given by

$$\begin{array}{c} a & & b & c \\ & / & \diagdown & \\ & e & & \\ & \diagdown & / & \\ d & & & \end{array} = \sum_{f \in [a,b] \cap [c,d]} \left[\begin{smallmatrix} a & b & f \\ d & c & e \end{smallmatrix} \right] \cdot \begin{array}{c} a & & b & c \\ & / & \diagdown & \\ & f & & \\ & \diagdown & / & \\ d & & & \end{array}.$$

Both versions given here differ from those in the literature [5], [21]. It is not hard to show, using Corollary 3.24, that

$$\left[\begin{smallmatrix} a & b & f \\ d & c & e \end{smallmatrix} \right]' = (-1)^{\frac{1}{2}(b+d-e-f)} \left[\begin{smallmatrix} a & b & f \\ d & c & e \end{smallmatrix} \right].$$

Thus, as a corollary to the above lemma we have the following $6j$ -symbols, given by replacing c with $c + 1$ or $c - 1$, which will be used to prove the next theorem:

Corollary 5.5.

$$\begin{aligned} \left[\begin{smallmatrix} 1 & a & a+1 \\ c+1 & b & c \end{smallmatrix} \right] &= 1; & \left[\begin{smallmatrix} 1 & a & a-1 \\ c+1 & b & c \end{smallmatrix} \right] &= -\frac{(a+b-c)}{2(a+1)}; \\ \left[\begin{smallmatrix} 1 & a & a+1 \\ c-1 & b & c \end{smallmatrix} \right] &= +\frac{(-a+b+c)}{2c}; & \left[\begin{smallmatrix} 1 & a & a-1 \\ c-1 & b & c \end{smallmatrix} \right] &= \frac{(a+b+c+2)(a-b+c)}{4(a+1)c}. \end{aligned}$$

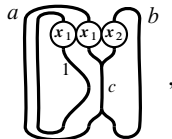
We can now prove the “multiplication by x ” formula.

Theorem 5.6.

$$\begin{aligned} x \cdot \chi^{a,b,c} &= \chi^{a+1,b,c+1} + \frac{(a+b-c)^2}{4a(a+1)} \chi^{a-1,b,c+1} + \frac{(-a+b+c)^2}{4c(c+1)} \chi^{a+1,b,c-1} \\ &\quad + \frac{(a+b+c+2)^2(a-b+c)^2}{16a(a+1)c(c+1)} \chi^{a-1,b,c-1}. \end{aligned} \tag{5.6}$$

This equation still holds for $a = 0$ or $c = 0$, provided we exclude the terms with a or c in the denominator.

Proof. Diagrammatically, $x \cdot \chi^{a,b,c}(x, y, z)$ is represented by



since $x = \text{tr}(x_1) = \textcircled{x_1}$ and multiplication is automatic on disjoint diagrams. Now manipulate the diagram to obtain a sum over χ 's with the following three steps.

First, apply the fusion identity to connect the lone $\textcircled{x_1}$ strand to the $\chi^{a,b,c}$:

$$\begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} c \text{---} \\ b \end{array} = \begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} c \text{---} \\ b \\ c+1 \end{array} + \frac{c}{c+1} \begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} c \text{---} \\ b \\ c-1 \end{array}, \quad (5.7)$$


where the coefficients are evaluated from

$$\frac{\Delta(c \pm 1)}{\Theta(1, c, c \pm 1)} = \frac{c \pm 1 + 1}{c + \frac{3}{2} \pm \frac{1}{2}}.$$

Second, use the $6j$ -symbols computed in Corollary 5.5 above to move the a strand from one side of the diagram to the other:

$$\begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} c \text{---} \\ b \\ c+1 \end{array} = \begin{array}{c} a+1 \\ \textcircled{x_1} \textcircled{x_2} \\ \text{---} \epsilon+1 \text{---} \\ b \end{array} + \frac{(a+b-c)^2}{4(a+1)^2} \begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} c \text{---} \\ b \\ a-1 \end{array} \quad (5.8)$$

$$\begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} c \text{---} \\ b \\ c-1 \end{array} = \frac{(-a+b+c)^2}{4c^2} \begin{array}{c} a+1 \\ \textcircled{x_1} \textcircled{x_2} \\ \text{---} \epsilon-1 \text{---} \\ b \end{array} + \frac{(a+b+c+2)^2(a-b+c)^2}{16(a+1)^2c^2} \begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} c \text{---} \\ b \\ a-1 \end{array}. \quad (5.9)$$

In each case, we are recoupling twice: once for the top piece  and once for the corresponding bottom piece. In doing this, we would actually get four terms, but since the $a \pm 1$ labels *must be the same* on both the top and the bottom (a consequence of Schur's Lemma or the bubble identity), two of the terms vanish.

In the final step, use the bubble identity to collapse the final pieces:

$$\begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} a+1 \text{---} \\ \text{---} a+1 \text{---} \\ b \\ c \pm 1 \end{array} = \left(\frac{\Theta(1, a, a+1)}{\Delta(a+1)} \right) \begin{array}{c} a+1 \\ \textcircled{x_1} \textcircled{x_2} \\ \text{---} \epsilon \pm 1 \text{---} \\ b \end{array} = \chi^{a+1, b, c \pm 1};$$

$$\begin{array}{c} a \\ \textcircled{x_1} \textcircled{x_1} \textcircled{x_2} \\ \text{---} 1 \text{---} \\ \text{---} a-1 \text{---} \\ \text{---} a-1 \text{---} \\ b \\ c \pm 1 \end{array} = \left(\frac{\Theta(1, a, a-1)}{\Delta(a-1)} \right) \begin{array}{c} a-1 \\ \textcircled{x_1} \textcircled{x_2} \\ \text{---} \epsilon \pm 1 \text{---} \\ b \end{array} = \left(\frac{a+1}{a} \right) \chi^{a-1, b, c \pm 1}.$$

At this point, obtaining (5.6) is simply a matter of multiplying the coefficients obtained in the previous formulae.

Now consider the special cases. For $a = 0$, since $b = c$ and consequently $\frac{c}{c+1} = \frac{(-a+b+c)^2}{4c(c+1)}$, the desired formula is exactly (5.7). Similarly, for $c = 0$, the desired formula is (5.8). \square

We find it interesting that, for all our discussion of signs introduced by non-topological invariance, all signs introduced are eventually squared and thus do not show up in this result.

We can rearrange the terms in (5.6) and re-index to get:

Corollary 5.7 (Central function recurrence). *Provided $a > 1$ and $c > 1$, we can write*

$$\begin{aligned} \chi^{a,b,c} = x \cdot \chi^{a-1,b,c-1} - \frac{(a+b-c)^2}{4a(a-1)} \chi^{a-2,b,c} - \frac{(-a+b+c)^2}{4c(c-1)} \chi^{a,b,c-2} \\ - \frac{(a+b+c)^2(a-b+c-2)^2}{16a(a-1)c(c-1)} \chi^{a-2,b,c-2}. \end{aligned}$$

The relation still holds for $a = 1$ or $c = 1$, provided we exclude the terms with $a - 1$ or $c - 1$ in the denominator.

The condition $a > 1, c > 1$ arises because decrementing a and c in (5.6) means $(a - 1, b, c - 1)$ must now be admissible. Also, note that formulae for multiplication by y and z may be obtained by applying the symmetry relation of Theorem 5.2. This fact is indispensable in our proof of Theorem 5.12.

5.3 Graded structure of the central function basis

The majority of the content in this section was suggested to us by Carlos Florentino [12] after he read an early draft of this chapter.

Recall the α, β, γ notation used earlier, and the notation

$$\chi_{\alpha,\beta,\gamma}(\mathbf{x}_2, \mathbf{x}_1^{-1}, \mathbf{x}_1 \mathbf{x}_2^{-1}) = \chi^{a,b,c}(\mathbf{x}_1, \mathbf{x}_2)$$

introduced in the proof of Theorem 5.2. The recurrence in Corollary 5.7 may be rewritten as

$$\begin{aligned} \chi_{\alpha,\beta,\gamma} = \chi_{0,1,0} \chi_{\alpha,\beta-1,\gamma} - \frac{\gamma^2}{a(a-1)} \chi_{\alpha+1,\beta-1,\gamma-1} - \frac{\alpha^2}{c(c-1)} \chi_{\alpha-1,\beta-1,\gamma+1} \\ - \frac{\delta^2(\beta-2)^2}{a(a-1)c(c-1)} \chi_{\alpha,\beta-2,\gamma}. \end{aligned}$$

The interchangeability of (a, α) and (c, γ) is guaranteed by the symmetry theorem.

Proposition 5.8. *The polynomial $\chi^{a,b,c} = \chi_{\alpha,\beta,\gamma}$ is monic, with highest degree monomial $x^\beta y^\alpha z^\gamma$.*

Proof. Induct on the degree $\delta = \alpha + \beta + \gamma$ of central functions. The statement is clearly true for the base cases, since $\chi_{0,0,0} = 1, \chi_{0,1,0} = x, \chi_{1,0,0} = y$, and

$\chi_{0,0,1} = z$. The recurrence relation implies that the highest order term of $\chi_{\alpha,\beta,\gamma}$ is x times the highest order term of $\chi_{\alpha,\beta-1,\gamma}$, hence $x(x^{\beta-1}y^\alpha z^\gamma) = x^\beta y^\alpha z^\gamma$. This fact, together with the appropriate symmetric facts for y and z , completes the induction. \square

The basis also preserves a certain grading on $\mathbb{C}[x, y, z]$. To define this grading, partition the standard basis $\mathcal{B} = \{x^a y^b z^c\}$ of this space as follows. Let $\text{gr}: \mathcal{B} \rightarrow \mathbb{Z}_2 \times \mathbb{Z}_2$ be defined by:

$$\text{gr}(x^a y^b z^c) = (a + c, b + c) \pmod 2.$$

If \mathcal{B} is considered as a semigroup under multiplication, then gr is a homomorphism since

$$\begin{aligned} \text{gr}(x^a y^b z^c) + \text{gr}(x^{a'} y^{b'} z^{c'}) &= (a + c, b + c) + (a' + c', b' + c') \pmod 2 \\ &= (a + a' + c + c', b + b' + c + c') \pmod 2 \\ &= \text{gr}(x^{a+a'} y^{b+b'} z^{c+c'}) \pmod 2. \end{aligned}$$

Therefore, gr defines a grading on this basis.

Proposition 5.9. *The basis $\{\chi^{a,b,c}\}$ respects the $\mathbb{Z}_2 \times \mathbb{Z}_2$ -grading on $\mathbb{C}[x, y, z]$ defined by gr , in the sense that*

$$\chi^{a,b,c} \in \mathbb{C}(\text{gr}^{-1}(a, b) \pmod 2).$$

Proof. This is another proof by induction on the degree δ . Clearly, $\chi^{0,0,0} = 1 \in \text{gr}^{-1}(0, 0)$, and likewise $\chi^{1,0,1} = x \in \text{gr}^{-1}(1, 0)$, $\chi^{0,1,1} = y \in \text{gr}^{-1}(0, 1)$, and $\chi^{1,1,0} = z \in \text{gr}^{-1}(1, 1)$. In the induction step, note that

$$(a, b) = (1, 0) + (a - 1, b) = (a - 2, b) \pmod 2,$$

so all terms on the righthand side of the recurrence relation in Corollary 5.7 have the same grading. Thus $\chi^{a,b,c} \in \text{gr}^{-1}(a, b)$. \square

5.4 Multiplication of central functions

It is not difficult to write down the formula for the product of two central functions, although the formula is by no means simple. The proof that follows was motivated by [29]. We begin with a lemma which encapsulates the most tedious diagram manipulations:

Lemma 5.10.

$$\begin{array}{c} a' \quad b \\ \diagdown \quad \diagup \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ c \quad c' \\ \diagdown \quad \diagup \\ a' \quad b \end{array} = \sum_{i,j,k,l,m} C_{j_1 k_1 l_1 . j_2 k_2 l_2 . m}^{abc, a' b' c'} \begin{array}{c} a' \quad b \\ \diagdown \quad \diagup \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ k_1 \quad l_1 \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ k_2 \quad l_2 \\ \quad \quad \quad \\ \diagdown \quad \diagup \\ a' \quad b \end{array},$$

where the coefficients are given by the formula

$$C_{j_1 k_1 l_1, j_2 k_2 l_2, m}^{abc a' b' c'} = \frac{\Theta(c, c', m)}{\Delta(m)} \prod_{i=1,2} \frac{\Delta(j_i)}{\Theta(a', b, j_i)} \cdot \begin{bmatrix} a & a' & k_i \\ c & j_i & b \end{bmatrix} \begin{bmatrix} b' & b & l_i \\ c' & j_i & a' \end{bmatrix} \begin{bmatrix} k_i & l_i & m \\ c & c' & j_i \end{bmatrix},$$

and the following 15 triples are assumed to be admissible:

$$(a, a', k_i), (b, b', l_i), (c, c', m), (a', b, j_i), (c, j_i, k_i), (c', j_i, l_i), (b, j_i, l_i), (k_i, l_i, m).$$

Proof. We will just demonstrate the diagram manipulation for the top half of the diagram, which by symmetry must be the same as for the bottom half. Combining these two manipulations and applying a bubble identity will give the desired result. We will save enumeration of admissible triples until after the manipulation, but keep a close eye on signs in the meantime.

$$\begin{aligned} & \begin{array}{c} a \quad b \\ \diagdown \quad / \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ c \quad c' \end{array} = \sum_j (-1)^{\frac{1}{2}(a'-b+j)} \frac{\Delta(j)}{\Theta(a', b, j)} \begin{array}{c} a \quad a' \quad b \\ \diagdown \quad / \quad \diagup \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ c \quad c' \end{array} \\ & = \sum_{j,k} (-1)^{\frac{1}{2}(a'-b+j)+j} \frac{\Delta(j)}{\Theta(a', b, j)} \begin{bmatrix} a & a' & k \\ c & j & b \end{bmatrix} \begin{array}{c} a \quad a' \quad b \\ \diagdown \quad / \quad \diagup \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ c \quad c' \end{array} \\ & = \sum_{j,k,l} (-1)^{\frac{1}{2}(a'-b-j)} \frac{\Delta(j)}{\Theta(a', b, j)} \begin{bmatrix} a & a' & k \\ c & j & b \end{bmatrix} \begin{bmatrix} b' & b & l \\ c' & j & a' \end{bmatrix} \begin{array}{c} a \quad a' \quad b \\ \diagdown \quad / \quad \diagup \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ c \quad c' \end{array} \\ & = \sum_{j,k,l} (-1)^{\frac{1}{2}(a'-b-j)+\frac{1}{2}(j+l-c')} \frac{\Delta(j)}{\Theta(a', b, j)} \begin{bmatrix} a & a' & k \\ c & j & b \end{bmatrix} \begin{bmatrix} b' & b & l \\ c' & j & a' \end{bmatrix} \begin{array}{c} a \quad a' \quad b \\ \diagdown \quad / \quad \diagup \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ c \quad c' \end{array} \\ & = \sum_{j,k,l,m} (-1)^{\frac{1}{2}(a'-b+c'-j-m)+l} \frac{\Delta(j)}{\Theta(a', b, j)} \begin{bmatrix} a & a' & k \\ c & j & b \end{bmatrix} \begin{bmatrix} b' & b & l \\ c' & j & a' \end{bmatrix} \begin{bmatrix} k & l & m \\ c & c' & j \end{bmatrix} \begin{array}{c} a \quad a' \quad b \\ \diagdown \quad / \quad \diagup \\ \quad \quad \quad \\ \diagup \quad \diagdown \\ c \quad c' \end{array} \end{aligned}$$

The (-1) terms all cancel in the end, a consequence of the fact that the following triples must be admissible:

$$(a, a', k), (b, b', l), (c, c', m), (a', b, j), (c, j, k), (c', j, l), (b, j, l), (k, l, m).$$

One computes the 13-parameter coefficients $C_{j_1 k_1 l_1, j_2 k_2 l_2, m}^{abc, a' b' c'}$ above by reflecting this result vertically, taking two sets of indices for the variables j, k, l, m on the two halves, and noting that the resulting bubble in the middle collapses with a factor of $\frac{\Theta(c, c', m)}{\Delta(m)}$ for $m = m_1 = m_2$. \square

With that out of the way, we can describe the central function multiplication table explicitly. Note the symmetry with respect to k, l, m , which is guaranteed by Theorem 5.2.

Theorem 5.11 (Multiplication of central functions). *The product of two central functions $\chi^{a,b,c}$ and $\chi^{a',b',c'}$ is given by*

$$\chi^{a,b,c} \chi^{a',b',c'} = \sum_{j_1, j_2, k, l, m} C_{j_1 k l m} C_{j_2 k l m} \frac{\Theta(a, a', k) \Theta(b, b', l) \Theta(c, c', m)}{\Delta(k) \Delta(l) \Delta(m)} \chi^{k, l, m},$$

where the sum is taken over admissible triples

$$(a, a', k), (b, b', l), (c, c', m), (a', b, j_i), (c, j_i, k), (c', j_i, l), (b, j_i, l), (k, l, m)$$

and the coefficients are given by

$$C_{j_i k l m} = \frac{\Delta(j_i)}{\Theta(a', b, j_i)} \begin{bmatrix} a & a' & k \\ c & j_i & b \end{bmatrix} \begin{bmatrix} b' & b & l \\ c' & j_i & a' \end{bmatrix} \begin{bmatrix} k & l & m \\ c & c' & j_i \end{bmatrix}.$$

Proof. By the previous lemma and the *bubble identity*, we have:

$$\begin{aligned} & \text{Diagram 1} = \sum_{j_1, k_1, l_1, j_2, k_2, l_2, m} C_{j_1 k_1 l_1, j_2 k_2 l_2, m}^{abc, a' b' c'} \text{Diagram 2} \\ & = \sum_{j_1, j_2, k, l, m} C_{j_1 k l m}^{abc, a' b' c'} \left(\frac{\Theta(a, a', k) \Theta(b, b', l)}{\Delta(k) \Delta(l)} \right) \text{Diagram 3} \\ & = \sum_{i, j, k, l} C_{j_1 k l m} C_{j_2 k l m} \frac{\Theta(a, a', k) \Theta(b, b', l) \Theta(c, c', m)}{\Delta(k) \Delta(l) \Delta(m)} \text{Diagram 4}. \end{aligned}$$

□

5.5 Applications

Spin networks offer a novel approach to a classical theorem of Fricke, Klein, and Vogt [14], [32]. We give here a new *constructive* proof which depends on the symmetry, recurrence, and multiplication formulae for central functions.

Theorem 5.12 (Fricke–Klein–Vogt Theorem). *Let $G = \text{SL}(2, \mathbb{C})$ act on $G \times G$ by simultaneous conjugation. Then*

$$\mathbb{C}[G \times G]^G \cong \mathbb{C}[t_x, t_y, t_z],$$

the complex polynomial ring in three indeterminates. In particular, every regular function $f : \mathrm{SL}(2, \mathbb{C}) \times \mathrm{SL}(2, \mathbb{C}) \rightarrow \mathbb{C}$ satisfying

$$f(\mathbf{x}_1, \mathbf{x}_2) = f(g\mathbf{x}_1g^{-1}, g\mathbf{x}_2g^{-1}) \quad \text{for all } g \in \mathrm{SL}(2, \mathbb{C}),$$

can be written uniquely as a polynomial in the three trace variables $x = \mathrm{tr}(\mathbf{x}_1)$, $y = \mathrm{tr}(\mathbf{x}_2)$, and $z = \mathrm{tr}(\mathbf{x}_1\mathbf{x}_2^{-1})$.

Proof. Define the ring homomorphism

$$\Gamma : \mathbb{C}[t_x, t_y, t_z] \rightarrow \mathbb{C}[G \times G]^G$$

by $f(t_x, t_y, t_z) \mapsto f(\mathrm{tr}(\mathbf{x}_1), \mathrm{tr}(\mathbf{x}_2), \mathrm{tr}(\mathbf{x}_1\mathbf{x}_2^{-1}))$.

We first show that Γ is injective. Suppose $f(\mathrm{tr}(\mathbf{x}_1), \mathrm{tr}(\mathbf{x}_2), \mathrm{tr}(\mathbf{x}_1\mathbf{x}_2^{-1})) = 0$ for all pairs $(\mathbf{x}_1, \mathbf{x}_2) \in G \times G$. Let $(\tau_x, \tau_y, \tau_z) \in \mathbb{C}^3$, $\epsilon_x = \begin{bmatrix} \tau_x & 1 \\ -1 & 0 \end{bmatrix}$, and $\eta_{y,z} = \begin{bmatrix} \tau_y & \frac{1}{\zeta} \\ -\zeta & 0 \end{bmatrix}$, where $\zeta + \zeta^{-1} = \tau_z$. Then

$$(\tau_x, \tau_y, \tau_z) = (\mathrm{tr}(\epsilon_x), \mathrm{tr}(\eta_{y,z}), \mathrm{tr}(\epsilon_x\eta_{y,z}^{-1})).$$

Hence $f = 0$ on \mathbb{C}^3 , $\mathrm{Ker}(\Gamma) = \{0\}$, and Γ is injective. This is the ‘‘Fricke slice’’ given by Goldman in [18].

It remains to show that Γ is surjective. Theorem 4.1 implies that the central functions form a basis for $\mathbb{C}[G \times G]^G$. Since $t_x \mapsto x$, $t_y \mapsto y$, and $t_z \mapsto z$, it suffices to show that every $\chi^{a,b,c}$ may be written as a polynomial in x , y , and z . This was already done via Lemma 5.1, but we provide here a *constructive* proof.

Proceed by induction on the degree $\delta = \frac{1}{2}(a + b + c)$ of a central function $\chi^{a,b,c}$. For the base cases $\delta = 0, 1$ recall our earlier computations demonstrating

$$\chi^{0,0,0} = 1, \quad \chi^{1,0,1} = x, \quad \chi^{0,1,1} = y, \quad \chi^{1,1,0} = z.$$

For $\delta > 0$, we may inductively assume that all central functions with degree less than δ are in $\mathbb{C}[x, y, z]$. The admissibility conditions imply that at least two out of the triple (a, b, c) are positive. Without loss of generality, using Theorem 5.2, we may assume that a and c are positive. In this case, the recurrence given by Corollary 5.7,

$$\begin{aligned} \chi^{a,b,c} &= x \cdot \chi^{a-1,b,c-1} - \frac{(a+b-c)^2}{4a(a-1)} \chi^{a-2,b,c} \\ &\quad - \frac{(-a+b+c)^2}{4c(c-1)} \chi^{a,b,c-2} - \frac{(a+b+c)^2(a-b+c-2)^2}{16a(a-1)c(c-1)} \chi^{a-2,b,c-2}, \end{aligned}$$

allows us to write $\chi^{a,b,c}$ in terms of central functions of lower degree, which by induction must be in $\mathbb{C}[x, y, z]$. Thus, $\chi^{a,b,c} \in \mathbb{C}[x, y, z]$, and we have established surjectivity. \square

The recurrence relations provide an algorithm for writing any $\chi^{a,b,c}$ as a polynomial in $\{x, y, z\}$. Conversely, in [25] the following formula is established, which may

be used to express any polynomial in $\mathbb{C}[x, y, z]$ in terms of central functions:

$$x^a y^b z^c = \sum_{\substack{r,s,t=0 \\ k,l,m}}^{\lfloor \frac{a}{2} \rfloor, \lfloor \frac{b}{2} \rfloor, \lfloor \frac{c}{2} \rfloor} \binom{a}{r} \binom{a}{r-1} \binom{b}{s} \binom{b}{s-1} \binom{c}{t} \binom{c}{t-1} \cdot \left(\frac{\Delta(l)\Delta(m)\Theta(a-2r,c-2t,k)}{\Delta(k)\Theta(a-2r,b-2s,m)\Theta(b-2s,c-2t,l)} \right) \begin{bmatrix} a-2r & c-2t & k \\ m & l & b-2s \end{bmatrix}^2 \chi^{k,l,m}.$$

Table 2 lists several central functions that were computed with Mathematica using Corollary 5.7. Only one function per triple of indices is listed; the others follow directly from Theorem 5.2.

Table 2. $SL(2, \mathbb{C})$ -central functions.

δ	$\chi^{a,b,c}$	$\chi_{\alpha,\beta,\gamma}$	$\mathfrak{p}_{a,b,c}(y, x, z)$
0	$\chi^{0,0,0}$	$\chi_{0,0,0}$	1
1	$\chi^{1,0,1}$	$\chi_{0,1,0}$	x
2	$\chi^{2,0,2}$	$\chi_{0,2,0}$	$x^2 - 1$
	$\chi^{1,1,2}$	$\chi_{1,1,0}$	$xy - \frac{1}{2}z$
3	$\chi^{3,0,3}$	$\chi_{0,3,0}$	$x^3 - 2x$
	$\chi^{2,1,3}$	$\chi_{1,2,0}$	$x^2y - \frac{2}{3}(xz + y)$
	$\chi^{2,2,2}$	$\chi_{1,1,1}$	$xyz - \frac{1}{2}(x^2 + y^2 + z^2) + 1$
4	$\chi^{4,0,4}$	$\chi_{0,4,0}$	$x^4 - 3x^2 + 1$
	$\chi^{3,1,4}$	$\chi_{1,3,0}$	$x^3y - \frac{3}{4}x^2z - \frac{1}{2}(3xy - z)$
	$\chi^{2,2,4}$	$\chi_{2,2,0}$	$x^2y^2 - xyz + \frac{1}{6}z^2 - \frac{1}{2}(x^2 + y^2) + \frac{1}{3}$
	$\chi^{3,2,3}$	$\chi_{1,2,1}$	$x^2yz - \frac{2}{3}(xz^2 + xy^2) - \frac{1}{2}x^3 - \frac{1}{9}(2yz - 13x)$

References

- [1] M. Artin, On Azumaya algebras and finite dimensional representations of rings. *J. Algebra* 11 (1969), 532–536.
- [2] J. Baez, Spin networks in gauge theory. *Adv. Math.* 117 (1996), 253–272.
- [3] T. Bröcker and T. tom Dieck, *Representations of Compact Lie Groups*. Grad. Texts in Math. 98, Springer-Verlag, New York 1985.

- [4] D. Bullock, C. Frohman, and J. Kania-Bartoszyńska, Understanding the Kauffman Bracket Skein Module. *J. Knot Theory Ramifications* 8 (1999), 265–277.
- [5] J. Carter, D. Flath, and M. Saito, *The Classical and Quantum $6j$ -Symbols*. Math. Notes 43, Princeton University Press, Princeton, NJ, 1995.
- [6] M. Culler and P. Shalen, Varieties of group representations and splittings of 3-manifolds. *Ann. of Math. (2)* 117 (1) (1983), 109–146.
- [7] R. Carter, G. Segal, and I. MacDonald, *Lectures on Lie Groups and Lie Algebras*. Lond. Math. Soc. Student Texts 32, Cambridge University Press, Cambridge 1995.
- [8] P. Cvitanović, *Group Theory: Birdtracks, Lie's, and Exceptional Groups*. Princeton University Press, Princeton, NJ, 2008; available online at <http://birdtracks.eu/>.
- [9] P. Cvitanović, H. Elvang, and A. D. Kennedy, Diagrammatic Young Projection Operators for $U(n)$. *J. Math. Phys.* 46 (4) (2005), 043501.
- [10] I. Dolgachev, *Lectures on Invariant Theory*. Lond. Math. Soc. Lecture Notes Series 296, Cambridge University Press, Cambridge 2003.
- [11] V. Drensky, Defining relations for the algebra of invariants of 2×2 matrices. *Algebr. Represent. Theory* 6 (2) (2003), 193–214.
- [12] C. Florentino, personal correspondence.
- [13] C. Florentino, J. Mourão, and J. P. Nunes, Theta functions, geometric quantization and unitary Schottky bundles. In *The Geometry of Riemann Surfaces and Abelian Varieties*, Contemp. Math. 397, Amer. Math. Soc., Providence, RI, 2006, 55–72.
- [14] R. Fricke and F. Klein, Über die theorie der automorphen modulgruppen. *Kgl. Ges. d. W. Nachrichten Math-Phys. Klasse* (1896), 91–93.
- [15] W. Fulton, and J. Harris, *Representation Theory: A First Course*. Grad. Texts in Math. 129, Springer-Verlag, New York 1991.
- [16] W. Goldman, Invariant functions on Lie groups and Hamiltonian flows of surface group representations. *Invent. Math.* 85 (1986), 263–302.
- [17] W. Goldman, *Introduction to character varieties*. Unpublished notes, 2003.
- [18] W. Goldman, Trace coordinates on Fricke spaces of some simple hyperbolic surfaces. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 611–684.
- [19] F. González-Acuña and J. Montesinos-Amilibia, On the character variety of group representations in $SL(2, \mathbb{C})$ and $PSL(2, \mathbb{C})$. *Math. Z.* 214 (1993), 627–652.
- [20] R. Horowitz, Characters of free groups represented in the two-dimensional special linear group. *Comm. Pure Appl. Math.* 25 (1972), 635–649.
- [21] L. Kauffman, *Knots and Physics*. Ser. Knots Everything 1, World Scientific, River Edge, NJ, 1991.
- [22] W. Magnus, Rings of Fricke characters and automorphism groups of free groups. *Math. Z.* 170 (1) (1980), 91–103.
- [23] R. Penrose, Applications of negative dimensional tensors. In *Combinatorial Mathematics and its Applications*, Academic Press, London 1971, 221–244.
- [24] R. Penrose, letter to P. Cvitanović, 1981.

- [25] E. Peterson, Trace Diagrams, Representation Theory, and Low-Dimensional Topology. PhD Thesis, University of Maryland, 2006.
- [26] C. Procesi, Invariant theory of $n \times n$ matrices. *Adv. in Math.* 19 (1976), 306–381.
- [27] C. Procesi, Finite dimensional representations of algebras. *Israel J. Math.* 19 (1974), 169–182.
- [28] Y. Razmyslov, Trace identities of full matrix algebras over a field of characteristic zero. *Izv. Akad. Nauk SSSR Ser. Mat.* 38 (1974), 723–756 (in Russian).
- [29] N. Reshetikhin, personal correspondence.
- [30] A. Sikora, $SL(n)$ -character varieties as spaces of graphs. *Trans. Amer. Math. Soc.* 353 (2001), 2773–2804.
- [31] G. E. Stedman, *Diagram Techniques in Group Theory*. Cambridge University Press, Cambridge 1990.
- [32] M. Vogt, Sur les invariants fondamentaux des equations differentielles linéaires du second ordre. *Ann. Sci. École Norm. Sup.* (3) 6 (1889), 281–282.

Part D

The Grothendieck–Teichmüller theory

Chapter 17

**Grothendieck's reconstruction principle and
2-dimensional topology and geometry**

Feng Luo

Contents

1	Introduction	734
2	A simple example of convex polygons	737
	2.1 Hierarchies of polygons and surfaces	738
	2.2 Reconstruction principle for polygons	739
	2.3 A solution to Problem 2	739
	2.4 Observable invariants	740
	2.5 Spherical and hyperbolic polygons	741
3	Teichmüller space, the mapping class group, and the space of measured laminations	742
	3.1 The example of the torus	743
	3.2 Thurston's compactification	744
4	Restriction maps and the reconstruction theorems	745
	4.1 Restriction maps	745
	4.2 Reconstruction theorems	745
	4.3 Thurston's embedding	746
	4.4 Level-0 surfaces	746
	4.5 Level-1 subsurfaces	747
	4.6 Constraints on level-1 surfaces	750
	4.7 Relationship between the 1-holed torus and the 4-holed sphere	750
	4.8 Grothendieck's view	750
5	The space of simple loops on surfaces and the modular structure	751
	5.1 The works of Dehn and Lickorish	751
	5.2 A modular structure	752
	5.3 Resolution of intersection	752
	5.4 A lemma of Lickorish	753
	5.5 Level-2 surfaces	753
	5.6 A multiplicative structure on the space of curve systems	754
6	Reduction to level-2 surfaces	755
	6.1 The restriction map and essential subsurfaces	755
	6.2 A reduction lemma	755
	6.3 Relationship to Grothendieck's idea	756

7	$SL(2, \mathbb{C})$ representation variety of surface groups	756
7.1	Reducible representations	757
7.2	The role of level-1 surfaces	757
7.3	A question on high dimension	758
	Appendix A. The statement of the reconstruction theorems for level-1 surfaces .	758
	Appendix B. The modular structure on the space of 3-holed sphere decompositions	760
	Appendix C. A simple proof that the multiplication produces a curve system . .	761
	Appendix D. A proof of the reduction lemma in §6.2	761
	References	763

1 Introduction

The goal of this chapter is an attempt to relate some ideas of Grothendieck in his *Esquisse d'un programme* [10] and some of the recent results on 2-dimensional topology and geometry. Especially, we shall discuss Teichmüller theory, the mapping class groups, the $SL(2, \mathbb{C})$ representation variety of surface groups, and Thurston's theory of measured laminations.

A prominent idea in surface theory is that to study a surface, one should consider all subsurfaces inside it. Indeed, there is a *hierarchy* of compact oriented surfaces of negative Euler number under (essential) inclusion. Each surface in the hierarchy is indexed by its *level* which is the number of disjoint simple loops needed to decompose it into 3-holed spheres (i.e., the complex dimension of the Teichmüller space of complete hyperbolic metrics of finite area). The first three levels in the hierarchy are listed as follows. The level-0 surface is the 3-holed sphere, the level-1 surfaces are the 1-holed torus and the 4-holed sphere, and the level-2 surfaces are the 2-holed torus and the 5-holed sphere.

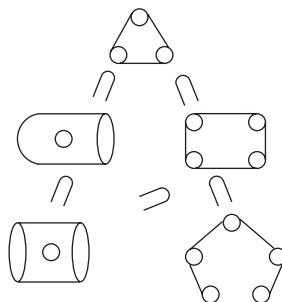


Figure 1. The first three levels of the hierarchy of surfaces.

One of the key ideas in [10] which we would like to discuss at length in this chapter is Grothendieck's reconstruction principle for the "Teichmüller tower". We quote the relevant paragraph in [10] below. On page 11, lines 2–6, Grothendieck wrote (with

English translation by L. Schneps) “The a priori interest of a complete knowledge of the first two levels of the tower is to be found in the principle that *the entire tower can be reconstructed from these two first levels*, in the sense that via the fundamental operation of ‘gluing’, level-1 gives the complete system of generators, and level-2 a complete system of relations.” One may interpret this principle broadly as follows. To study a structure (for instance, hyperbolic structure, complex projective structure, measured lamination, or linear representation of the surface group) and its moduli space on a surface, one should consider the restrictions of the structure to the level-1 subsurfaces and reconstruct the structure from its restrictions. The level-2 surfaces should serve as “relators” in the reconstruction process. For example, one may ask if the reconstruction principle holds for the characters of representations of the surface groups into the general linear group $GL(n, \mathbb{C})$. Namely, suppose f is a complex valued function defined on the fundamental group of the surface so that the restriction of f to the fundamental group of each essential level-1 subsurface is a $GL(n, \mathbb{C})$ -character. Is f the character of some $GL(n, \mathbb{C})$ representation of the surface group? In [27], we show that the answer is affirmative for $SL(2, \mathbb{C})$ representations of surface groups. It is interesting to note that this principle of reconstruction was taken as one of the basic axioms by physicists in conformal field theory ([30]).

The main theorems in [25], [26] state that the Teichmüller space and Thurston’s measured lamination space for surfaces obey the reconstruction principle. Also using the work of Gervais [7], we see that the mapping class group of a surface fits the principle as well [24]. These will be the main topics of this chapter. We shall also discuss some open questions arising from reading [10].

We remark that as far as we know, there is no precise definition of the Teichmüller tower in [10]. See also the books [22], [37]. What follows is my interpretation of Grothendieck’s reconstruction principle and there should be other ways of interpreting it (for instance in algebraic geometry).

One way to illustrate Grothendieck’s reconstruction principle is to consider convex planar n -sided polygons. According to the principle, to construct a convex n -sided ($n \geq 5$) polygon, one should consider all convex quadrilaterals inside the polygon (each vertex of the quadrilateral is a vertex of the polygon). The convex polygon is a union of these quadrilaterals by gluing along their overlaps. Now these quadrilaterals overlap in two different ways. An *essential overlap* of two quadrilaterals contains an edge or diagonal. Otherwise, they overlap inessentially (see Figure 2). The re-

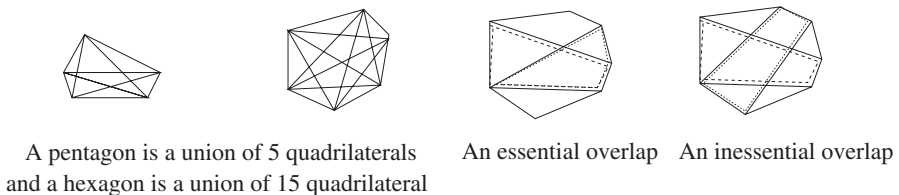


Figure 2

construction principle states that it suffices to glue quadrilaterals along the essential overlaps. The gluing along the inessential overlaps is a *consequence* of the gluing along essential overlaps (see §2 for more details). As a consequence, to study the geometry of the moduli space of convex polygons, it suffices to understand that of quadrilaterals.

The situation for surfaces is analogous to that of polygons where the 3-holed sphere corresponds to a triangle and the 1-holed torus and the 4-holed sphere correspond to quadrilaterals. (One should think of the polygons forming a hierarchy under inclusion. And the level of a polygon is the number of disjoint diagonals needed to decompose a polygon into triangles.) Thus according to the reconstruction principle, to construct a hyperbolic metric on a surface of negative Euler number, one should consider all (isotopy classes of) subsurfaces which are homeomorphic to the 1-holed torus or the 4-holed sphere. These subsurfaces overlap in two different patterns: an overlap is *essential* if there is a homotopically non-trivial loop in the overlap, otherwise it is *inessential* (see Figure 3). The reconstruction principle says that we can glue along essential overlaps to recover the original hyperbolic structure. To be more precise, assign to each level-1 subsurface a hyperbolic structure so that when two level-1 subsurfaces overlap essentially, they overlap geometrically (i.e., the geodesic lengths of all overlapping simple loops are the same in both level-1 surfaces). Then the reconstruction principle states that there exists a hyperbolic metric on the surface whose restrictions to level-1 subsurfaces are (isotopic to) the assigned hyperbolic structures. This is the main result established in [25].

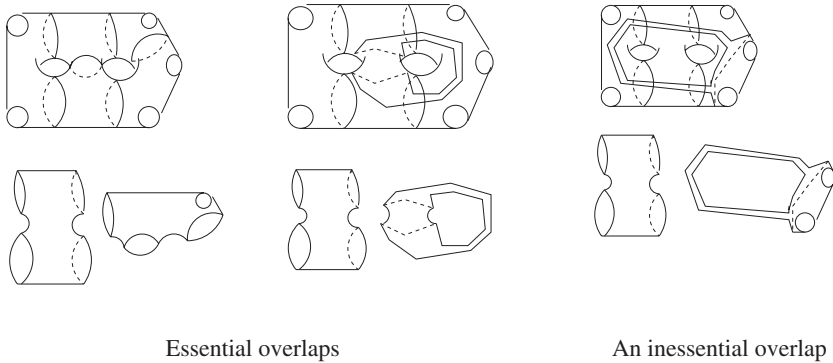


Figure 3

The organization of this chapter is as follows. In §2, we study the moduli space of convex polygons in details and use it to illustrate Grothendieck's principle. We also discuss ideal triangulations of surfaces. In §3, we recall basic facts about Teichmüller spaces, the mapping class groups, and Thurston's projective measured lamination spaces. It is well known that these three themes represent the geometric, algebraic, and topological aspects of surface theory. In the case of the torus, these three themes correspond to the upper-half space \mathbb{H} , the group $SL(2, \mathbb{Z})$, and the projective line

$\mathbb{R}P^1 = \mathbb{R} \cup \{\infty\}$ which appears as the boundary of \mathbb{H} . The natural action of the mapping class group corresponds to the actions of $SL(2, \mathbb{Z})$ on \mathbb{H} and on $\mathbb{R} \cup \{\infty\}$ by Möbius transformations. We shall also recall related topics for level-1 surfaces. In §4, we state the reconstruction theorems for Teichmüller spaces, measured lamination spaces and mapping class groups. In §5, we discuss the key ingredient in the proofs of the reconstruction theorem, namely simple loops on surfaces. We also recall the notion of $SL(2, \mathbb{Z})$ modular structure on a set. The role of modular structures, equivalently, of $(\mathbb{Q}P^1, SL(2, \mathbb{Z}))$ structures, is prominent in the reconstruction program as predicted by Grothendieck (see page 248–249 in [10] or §4.8). Topologists have known the role of modular configuration for simple loops on level-1 surfaces since the fundamental work of Max Dehn [4] in 1938. Dehn actually used such a structure to give an elegant derivation of the mapping class group of the 4-holed sphere (see §5.1). The special feature of a modular configuration is the huge symmetry built in the configuration. This is, in our view, one reason why the set of homotopy classes of simple loops on the surface is more useful than the fundamental group in establishing the reconstruction principle for many structures (see §4.7, §7.1 and Figure 9). In §6 we give a fairly general reason which indicates the special role played by level-2 surfaces in the reconstruction principle. In the last section, we discuss the characters of $SL(2, \mathbb{C})$ representations.

Acknowledgement. I would like to thank X. S. Lin for inviting me to write the paper for publication in 1998 and Athanase Papadopoulos for inviting me to contribute it as a chapter to the Handbook. Discussions with F. Bonahon, L. Keen, X.-S. Lin and C. Series have been very helpful for me in developing ideas in the chapter. I thank the referee, P. Landweber and A. Papadopoulos for careful reading of the manuscript and for suggestions on improving the exposition of the chapter. This work was partially supported by the NSF. The chapter is an updated version of an article that originally appeared in the *Communications in Contemporary Mathematics*, Vol. 1, No. 2, (1999), 125–153, and is printed with permission from the World Scientific Publishing Company.

2 A simple example of convex polygons

We shall illustrate the reconstruction principle and its applications by considering the configuration space of convex n -sided polygons. Let us begin with the following problem.

Problem 1. Describe the space $\mathcal{T}(n)$ of all convex n -sided polygons up to isometries. Here polygons have marked vertices and isometries preserve markings.

To be more precise, let us distinguish the topological (or combinatorial) and geometric aspects of the problem. By an n -sided polygon we mean a topological disk with n marked points (the vertices) on its boundary. A *convex structure* on a polygon is a

metric on the polygon which is isometric to a convex n -sided planar polygon so that the marked points correspond to the vertices. An *edge* in a convex polygon is a line segment joining two vertices, and a *diagonal* is an edge which does not join adjacent vertices. Given an n -sided polygon P , the space of all convex structures on P modulo isometries preserving the vertices is denoted by $\mathcal{T}(P)$ which is essentially $\mathcal{T}(n)$.

Having introduced these notations, we may rephrase the problem as follows.

Problem 2. Assign to each edge in a convex n -sided polygon a positive number. When does the assignment correspond to the edge lengths of an n -sided convex polygon?

The solution for triangles $n = 3$ is well known. The assignment must satisfy the triangular inequalities, namely, that the sum of two is larger than the third. For general n , the assignment must satisfy triangular inequalities over three edges forming a triangle and equations over six edges forming a quadrilateral. Grothendieck's reconstruction principle asserts that these are the set of all constraints, i.e., the quadrilaterals (= level-1 polygons) are the "generators" in building convex polygons.

2.1 Hierarchies of polygons and surfaces

It is instructive to compare the hierarchies of polygons and surfaces. One first observes that the isometry class of a convex polygon is determined by the lengths of all edges. The corresponding fact in hyperbolic geometry is a result of Fricke–Klein [5], that the isometry class of a hyperbolic metric on a surface is determined by the lengths of simple geodesic loops. The solution for $\mathcal{T}(3)$ is given by $\mathcal{T}(3) = \{(a_1, a_2, a_3) \in \mathbb{R}^3 \mid a_i + a_j > a_k\}$ reflecting the fact that a triangle is determined up to isometry by its three edge lengths subject to the triangular inequalities. The corresponding fact in hyperbolic geometry is the well-known theorem of Fricke–Klein [5], that a hyperbolic metric on a 3-holed sphere is determined up to isometry by the three lengths of the boundary geodesics, and these lengths subject no constraints. For $n \geq 4$, an old way of solving the problem for $\mathcal{T}(n)$ is to triangulate the n -sided polygon by $(n - 3)$ edges, i.e., one uses the triangle as the basic building block. This corresponds to the Fenchel–Nielsen decomposition of surfaces into 3-holed spheres (using the 3-holed sphere as the basic building block). In this way, one parametrizes the convex polygon by the lengths of the edges in the triangulation. These lengths have to satisfy complicated inequalities due to the convexity. Unlike the Fenchel–Nielsen coordinates for Teichmüller space which can be used to express the Weil–Petersson symplectic form by Wolpert's formula [43], the length coordinates for convex polygons seem to be less useful in extracting geometric information about $\mathcal{T}(n)$ except that they can be used to show that $\mathcal{T}(n)$ is a real analytic manifold diffeomorphic to \mathbb{R}^{2n-3} .

2.2 Reconstruction principle for polygons

Now Grothendieck's reconstruction principle asserts that quadrilaterals are the basic building blocks. Thus given an n -sided polygon P , one considers all quadrilaterals whose edges are edges in P . If the polygon has a convex structure, each quadrilateral in P becomes a convex quadrilateral. These convex quadrilaterals satisfy the obvious consistency condition:

- (*) If two quadrilaterals overlap essentially (i.e., there is an edge in the overlap), then the convex quadrilaterals overlap geometrically, i.e., the corresponding lengths of edges in both convex quadrilaterals are the same.

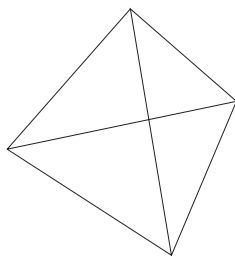
It turns out that the condition (*) is also sufficient to recover the convex polygon for obvious reason.

Reconstruction principle for polygons. To construct a convex n -sided polygon with $n \geq 5$, it suffices to assign to each quadrilateral in the polygon a convex structure so that the assignment satisfies the consistency condition (*).

2.3 A solution to Problem 2

In terms of the reconstruction principle, the solution to Problem 2 is simply that the assignment must be realized by a convex quadrilateral for each choice of six edges forming a quadrilateral.

This principle essentially reduces the study of $\mathcal{T}(n)$ to that of $\mathcal{T}(4)$. To understand $\mathcal{T}(4)$, one uses the lengths of the six edges of a quadrilateral. First of all, the lengths satisfy the triangular inequalities, that is, that the sum of two lengths is larger than the third one over each of the four triangles in the quadrilateral. By a simple calculation, one shows that these six lengths satisfy the following constraint:



$$(1) \quad \sum (x^2y + xy^2) + \sum xyz = \sum xyz$$

$x \quad \diagdown \quad y$

$x \quad \diagup \quad z$
 y

$x \quad \diagdown \quad z$
 y

Figure 4

where the sums are over all specified subgraphs (of the complete graph on 4 vertices) whose edges are labelled by the squares of their lengths. The convexity condition is equivalent to the following *Largest Root Condition*. Fix five lengths and think of (1) as a quadratic equation in the square of the length of the remaining diagonal. It has two real roots and the convexity condition says that the largest root is the square of the length. As a consequence of (1), the Largest Root Condition, the triangular inequalities, and the reconstruction principle, one obtains a complete solution to Problem 2.

2.4 Observable invariants

Given a convex polygon P , the observable invariants of P seem to be the area of the polygon, the lengths of edges and the angles of intersections of edges. These define the “observable” area, length and angle functions on the configuration space $\mathcal{T}(P)$. To be more precise, fixing an isotopy class of an edge e in P (resp. a pair of isotopy classes of intersecting edges), one defines a length function (resp. angle function) from $\mathcal{T}(P)$ to \mathbb{R} by sending a convex structure to the length (resp. angle) of e in the convex structure. These naturally defined functions seem to play an important role in the geometry of the configuration space $\mathcal{T}(P)$. And indeed they do. Here is one way to see it using Thurston’s invariant of oriented triangles.

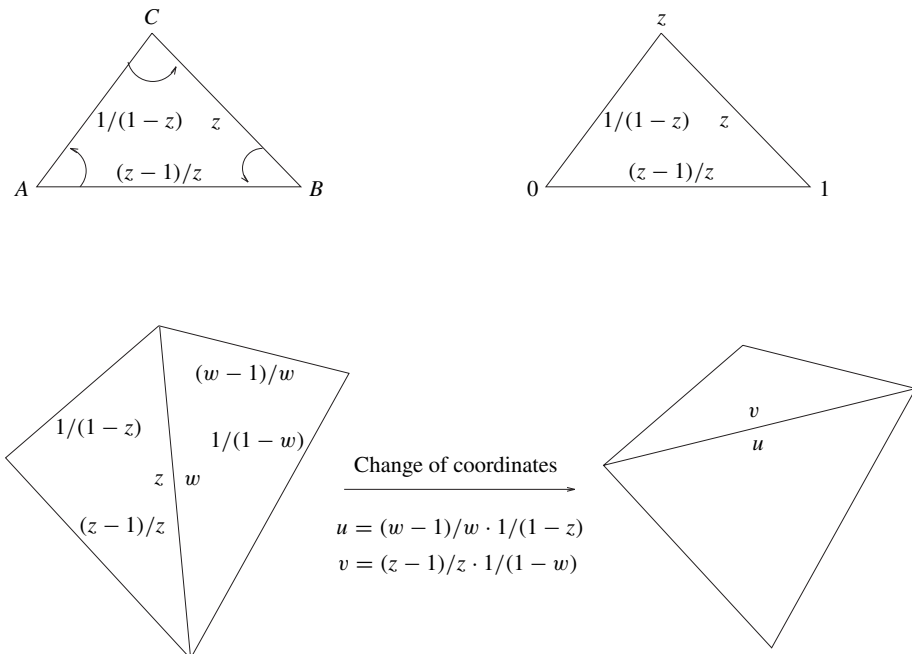


Figure 5. Right-hand orientation in the plane.

Suppose $\triangle ABC$ is an oriented triangle in the plane so that the cyclic order (A, B, C) is the orientation. Then the *Thurston invariant* z_A of the triangle at edge BC is defined to be the complex number $\frac{C-A}{B-A}$. The edge invariants $z_B = \frac{A-B}{C-B}$ and $z_C = \frac{B-C}{A-C}$ for AC and AB are given by $z_B = 1/(1 - z_A)$ and $z_C = 1/(1 - z_B) = (z_A - 1)/z_A$ respectively. In particular $z_A z_B z_C = -1$. Evidently if two oriented triangles differ by a similarity transformation ($f(w) = aw + b$, $a, b \in \mathbb{C}$, $a \neq 0$), then their Thurston invariants are the same. For an oriented convex quadrilateral with a marked diagonal, one defines the Thurston invariant to be the pair $(z, w) \in \mathbb{C} \times \mathbb{C}$ where each coordinate is the invariant of a triangle at the marked diagonal. For instance, in terms of Thurston's invariants, parallelograms are exactly those convex quadrilaterals with Thurston invariant (z, z) . A simple calculation shows that if the marked diagonal is changed, then the invariant becomes (u, v) where $u = \frac{w-1}{w(1-z)}$ and $v = \frac{z-1}{(1-w)z}$ as shown in Figure 5.

Furthermore, the convexity condition is equivalent to either all z, w, u, v are in \mathbb{H} or all z, w, u, v are in the lower-half plane $\{t \in \mathbb{C} \mid \bar{t} \in \mathbb{H}\}$. As a consequence of the transformation formulas, one sees that the space of similarity classes of convex quadrilaterals has a natural complex structure. Combining these with the reconstruction principle, one obtains the fact that the projectivized space $\mathcal{T}(n)/\mathbb{R}_+$ of similarity classes of convex polygons has a natural complex structure so that angle functions and the logarithm of the ratio of the length functions are pluriharmonic. Furthermore, the space $\mathcal{T}(n)/\mathbb{R}_+$ can be explicitly described. This result itself is not surprising since another way of parametrizing $\mathcal{T}(n)/\mathbb{R}_+$ is by taking vertices as coordinates. But the fact that the complex structure is built on that of $\mathcal{T}(4)/\mathbb{R}_+$ seems to be interesting. Evidently these "observable" length and angle functions also exist on Teichmüller spaces. It is natural to ask if these functions are somehow related to the complex structure of Teichmüller space.

2.5 Spherical and hyperbolic polygons

The reconstruction principle also holds for hyperbolic or spherical convex polygons. Thus the same picture holds in these cases as well. The most interesting case seems to be one of the ideal polygons in hyperbolic plane where one assigns to each oriented ideal quadrilateral with a marked diagonal the Bonahon–Thurston shearing coordinate ([2], [39] and [40]). Recall that the shearing coordinate is defined as follows. The *mid-point* of an edge in an ideal triangle is the point of tangency between the inscribed circle and the edge. Given an oriented ideal quadrilateral with a marked diagonal, the Bonahon–Thurston coordinate for the marked quadrilateral is the exponential of the signed hyperbolic distance from the left mid-point to the right mid-point of the diagonal. Note that the coordinate is independent of the choice of the orientation on the diagonal. If one changes the diagonal, the coordinate changes to its inverse. The change of coordinate formula for other four edges is given in Figure 6.

Note that the transformation formulas are real algebraic. Since each non-closed surface has an ideal triangulation, this gives an easy way to parametrize the Teichmüller

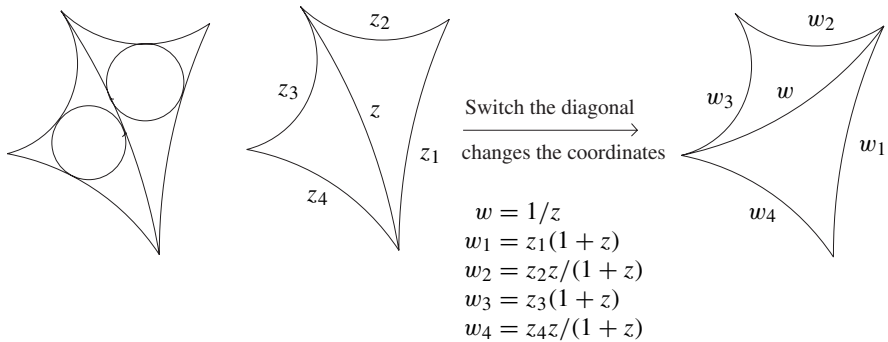


Figure 6. Right-hand orientation in the plane.

space of the surface using Bonahon–Thurston coordinates. As a consequence, one proves easily that Teichmüller space is a real analytic manifold diffeomorphic to a Euclidean space. (This seems to be one of the quickest ways of showing that the Teichmüller space of a non-closed surface is contractible. The other proof using the Fenchel–Nielsen coordinate seems to be always running into the technical difficulties of showing that the Fenchel–Nielsen coordinates associated to different 3-holed sphere decompositions differ by a diffeomorphism.) These shearing coordinates are closely related to Penner’s coordinates for decorated Teichmüller spaces (see [34]). See also [31] for related material on measured laminations.

3 Teichmüller space, the mapping class group, and the space of measured laminations

Given a compact orientable surface Σ with or without boundary, there are three themes naturally associated to the surface. Namely, the Teichmüller space $\mathcal{T}(\Sigma)$, the mapping class group $\Gamma(\Sigma)$, and the space $\mathcal{S}(\Sigma)$ of isotopy classes of unoriented simple loops not homotopic to a point (or its completion, Thurston’s space of measured laminations). These three themes represent the geometric, algebraic, and topological aspects of surface theory. Recall that the mapping class group $\Gamma(\Sigma) = \text{Homeo}^+(\Sigma, \partial\Sigma)/\text{Iso}$ is the group of orientation preserving self-homeomorphisms modulo isotopies so that the boundary of the surface is fixed pointwise by the homeomorphisms and the isotopies. For a surface of negative Euler number, the Teichmüller space $\mathcal{T}(\Sigma)$ is the space of all hyperbolic metrics with geodesic boundary on the surface modulo isometries isotopic to the identity. These three themes interact with each other in the sense that the mapping class group acts naturally on both $\mathcal{T}(\Sigma)$ and $\mathcal{S}(\Sigma)$ by pull back, and the space $\mathcal{S}(\Sigma)$ appears in Thurston’s compactification of the Teichmüller space $\mathcal{T}(\Sigma)$.

3.1 The example of the torus

One may illustrate these three themes and their interaction by the classical example of the oriented torus $\Sigma_{1,0}$ where the Teichmüller space $\mathcal{T}(\Sigma_{1,0})$ is defined to be the space of all flat metrics modulo similarity maps isotopic to the identity. In his doctoral thesis in 1913, J. Nielsen proved that two homologous homeomorphisms (resp. simple loops) of the torus are isotopic. Thus the space of simple loops $\mathcal{S}(\Sigma_{1,0})$ can be identified naturally with the set of primitive elements in $H_1(\Sigma_{1,0}, \mathbb{Z})$ modulo ± 1 , and the mapping class group $\Gamma(\Sigma_{1,0})$ is naturally isomorphic to the automorphism group $\text{Aut}^+(H_1(\Sigma_{1,0}, \mathbb{Z}))$. Define a *marking* on $\Sigma_{1,0}$ to be a pair of oriented simple loops (a, b) intersecting transversely at one point. Fix a marking (a, b) on $\Sigma_{1,0}$. Their homology classes $[a], [b]$ form a basis for the first homology group $H_1(\Sigma_{1,0}, \mathbb{Z})$. In terms of the basis, one can identify $\mathcal{S}(\Sigma_{1,0})$ with $\mathbb{Q}P^1 = \mathbb{Q} \cup \{\infty\}$ by sending each primitive class $\pm(p[a] + q[b])$ to its “slope” p/q . One can also identify the automorphism group $\text{Aut}^+(H_1(\Sigma_{1,0}, \mathbb{Z}))$ with $\text{SL}(2, \mathbb{Z})$. The natural action of the mapping class group $\Gamma(\Sigma_{1,0})$ on $\mathcal{S}(\Sigma_{1,0})$ becomes the standard action of $\text{SL}(2, \mathbb{Z})$ on the rationals by fractional linear transformations. The marking (a, b) can also be used to parametrize the Teichmüller space $\mathcal{T}(\Sigma_{1,0})$ as follows. Fix a flat metric d on $\Sigma_{1,0}$. We isotope a and b into two d -geodesics \hat{a} and \hat{b} which intersect at one point p . Let θ be the angle measured from \hat{a} to \hat{b} at p in the orientation of the surface and let l_a and l_b be the lengths of the geodesics \hat{a} and \hat{b} . Assign to the flat metric d the complex number $z_d = \frac{l_a e^{i\theta}}{l_b}$ in the upper-half plane \mathbb{H} . Evidently the invariant z_d depends only on the similarity class of the flat metric d . Thus one obtains a maps π_m from the Teichmüller space $\mathcal{T}(\Sigma_{1,0})$ to \mathbb{H} . This map is a bijection since the inverse can be constructed by sending $z \in \mathbb{H}$ to the torus $\mathbb{C}/(\mathbb{Z} + z\mathbb{Z})$ with marking corresponding to 1 and z . Note that the invariant z_d is independent of the orientations on a and b . Furthermore, the pair (z_d, \bar{z}_d) is the Thurston invariant of the parallelogram obtained by cutting the flat torus $(\Sigma_{1,0}, d)$ open along the geodesics \hat{a}, \hat{b} . Now if we are given a different marking $m' = (a', b')$, there is an $\text{SL}(2, \mathbb{Z})$ matrix A which sends $[a]$ to $[a']$ and $[b]$ to $\pm[b']$. A simple calculation shows that two invariants π_m and $\pi_{m'}$ are related by A acting as a fractional linear transformation on \mathbb{H} . Thus the Teichmüller space $\mathcal{T}(\Sigma_{1,0})$ can be naturally identified with \mathbb{H} so that the action of the mapping class group becomes the standard action of $\text{SL}(2, \mathbb{Z})$ on \mathbb{H} by fractional linear transformations. In short, the three themes $\mathcal{T}(\Sigma_{1,0})$, $\Gamma(\Sigma_{1,0})$, and $\mathcal{S}(\Sigma_{1,0})$ for the torus are exactly $(\mathbb{H}, \text{SL}(2, \mathbb{Z}), \mathbb{Q}P^1)$. It is interesting to note that the complex structure on the Teichmüller space makes both the angle function and the logarithm of ratio of length functions pluriharmonic. Indeed, by fixing a marking $m = (a, b)$ on the torus $\Sigma_{1,0}$, one obtains a fundamental domain map $f_m: \mathcal{T}(\Sigma_{1,0}) \rightarrow \mathcal{T}(4)/\mathbb{R}_+$ by sending the similarity class $[d]$ to the parallelogram based on \hat{a} and \hat{b} which forms a fundamental domain for the flat metric. The complex structure on $\mathcal{T}(\Sigma_{1,0})$ makes the map f_m holomorphic, i.e., holomorphic motions in the Teichmüller space correspond to the homomorphic motions of the fundamental domains. The same phenomenon does not seem to hold for the complex structure on the Teichmüller space of a closed

surface of higher genus with hyperbolic metrics ([19], [44]). (A natural choice of fundamental domains for closed surfaces of higher genus seems to be those associated to chains in the surface. See Maskit [29] for more information.) However, there is evidence indicating that the complex structure of Teichmüller space is more closely related to the singular flat metrics on the surface. To be more precise, given a closed Riemann surface Σ of genus g and a fixed point on the surface, there exists a unique singular flat metric of area 1 in the conformal class of Σ so that its singularity is at the fixed point having cone angle $2\pi(2g - 1)$.

3.2 Thurston's compactification

There is a natural compactification of the upper-half plane \mathbb{H} by the extended real line $\mathbb{R}P^1 = \mathbb{R} \cup \{\infty\}$ where the action of $SL(2, \mathbb{Z})$ extends continuously. This produces a compactification of the Teichmüller space of the torus. Thurston's deep work on surface theory shows that the same compactification also exists for all surfaces. We shall discuss briefly Thurston's work in this section. See [6] and [41] for more details. To begin with, a proper 1-dimensional submanifold s in a compact surface Σ is called a *curve system* if no component of s is homotopic into $\partial\Sigma$ relative to $\partial\Sigma$. The set of isotopy classes of all curve systems on Σ is denoted by $CS(\Sigma)$ and was introduced by Dehn who called it *the Arithmetic field* of the surface. In the case of a torus, the set $CS(\Sigma)$ is naturally identified with the set of all non-zero lattice points in $H_1(\Sigma_{1,0}, \mathbb{Z})$ modulo ± 1 . There exists a quadratic pairing on $CS(\Sigma)$ given by the geometric intersection number $I(\alpha, \beta) = \min\{|a \cap b| \mid a \in \alpha, b \in \beta\}$. For the torus, the pairing is $I((p, q), (p', q')) = |pq' - p'q|$ which is the absolute value of the canonical symplectic form on \mathbb{Z}^2 . This pairing satisfies the homogeneity and non-degenerate property in the sense that $I(k_1\alpha_1, k_2\alpha_2) = k_1k_2I(\alpha_1, \alpha_2)$ ($k_i \in \mathbb{Z}_+$ and $k_i\alpha_i$ means k_i copies of the curve system α_i), and for each α there exists β so that $I(\alpha, \beta) \neq 0$. Thurston's space of measured laminations $ML(\Sigma)$ is the completion of $CS(\Sigma)$ with respect to the pairing I . In linear algebra, given a non-degenerate quadratic form ω on a lattice L of rank r , one can form a completion of (L, ω) by canonically embedding L into R^r so that the form w extends continuously on R^r . If the form is definite, the simplest way to construct the completion is by formally extending ω to $\mathbb{Q}L$ and taking the metric completion of $\mathbb{Q}L$. If the form ω is not definite, one may embed L into the infinite dimensional space \mathbb{R}^L (with the product topology) by sending $x \in L$ to the linear function $\pi(x) = \omega(\cdot, x)$. The canonical completion is given by taking the closure of the set $\mathbb{Q}\pi(L)$. Since the form ω is non-degenerate, the Riesz representation theorem says that the closure is isomorphic to a vector space \mathbb{R}^r and the form ω extends continuously to the closure. Thurston's completion of $(CS(\Sigma), I)$ is an analogous construction. The space $CS(\Sigma)$ is embedded into $\mathbb{R}^{S(\Sigma)}$ by sending α to the *intersection function* $Th(\alpha) = I(\cdot, \alpha)$ and the closure of $\mathbb{Q}_+Th(CS(\Sigma))$ is defined to be the completion, the space of measured laminations $ML(\Sigma)$. Thurston proved a remarkable theorem that the space $ML(\Sigma)$ is homeomorphic to a Euclidean space and the quadratic pairing extends continuously to

$\text{ML}(\Sigma)$ ([3], [6], [35], [39]). Since Thurston's completion is canonically constructed, the mapping class group $\Gamma(\Sigma)$ acts continuously on $\text{ML}(\Sigma)$. In the case of the torus, $\text{ML}(\Sigma)$ is canonically $H_1(\Sigma_{1,0}, \mathbb{R})/\pm 1$ and the action of the mapping class group $\text{SL}(2, \mathbb{Z})$ is the standard action. The projectivized space $\text{PML} = (\text{ML}(\Sigma) - 0)/\mathbb{R}_{>0}$ is Thurston's compactification of Teichmüller space.

4 Restriction maps and the reconstruction theorems

4.1 Restriction maps

A compact subsurface Σ' in Σ is called *essential* if no component of $\partial\Sigma'$ is null homotopic in Σ . If Σ' is essential with negative Euler number, there exists a natural *restriction map* from the Teichmüller space $\mathcal{T}(\Sigma)$ to $\mathcal{T}(\Sigma')$ (resp. from $\text{ML}(\Sigma)$ to $\text{ML}(\Sigma')$). The restriction map is defined as follows. Given a hyperbolic metric d on Σ , we isotope the open surface $\text{int}(\Sigma')$ (the interior of Σ') to an open subsurface Σ'' so that Σ'' is bounded by disjoint simple geodesics. The metric completion of $(\Sigma'', d|_{\Sigma''})$ is a hyperbolic metric d' on Σ' with geodesic boundary. The restriction map sends $[d]$ to $[d']$. The restriction map for measured laminations is defined similarly (see [26]). The key step is to define the restriction map from the space of curve systems $\text{CS}(\Sigma)$ to $\text{CS}(\Sigma')$. Given $\alpha \in \text{CS}(\Sigma)$, choose a representative $a \in \alpha$ so that the number of components of $a \cap \Sigma'$ is minimal. Then the restriction map sends $[a]$ to $[a|_{\Sigma'}]$. The restriction maps are natural in the sense that if we are given two essential subsurfaces $\Sigma_1 \subset \Sigma_2 \subset \Sigma$, then the composition of restrictions is the restriction.

To state the reconstruction theorems, we say that two essential subsurfaces Σ_1 and Σ_2 *overlap essentially* if there is a non-trivial simple loop which is isotopic into both Σ_1 and Σ_2 . If furthermore both Σ_1 and Σ_2 are level-1 subsurfaces, then their possible intersection surfaces are either essential annuli, or an essential 3-holed sphere or they are isotopic.

4.2 Reconstruction theorems

With this preparation, we can state the main theorems in [25], [26], and [24] as follows. These can be considered as establishing Grothendieck's reconstruction principle for Teichmüller spaces, measured lamination spaces and mapping class groups.

Theorem 4.1 (Reconstruction of Teichmüller spaces and measured lamination spaces). *Each hyperbolic metric (resp. measured lamination) on a surface of level at least 2 is constructed uniquely up to isotopy by assigning a hyperbolic metric (resp. measured lamination) to each essential level-1 subsurface so that when two level-1 subsurfaces overlap essentially, the restrictions of the metrics (resp. measured laminations) to their intersection are isotopic. Furthermore, the restriction of the hyperbolic metric (resp.*

measured lamination) on the surface to each level-1 essential subsurface is isotopic to the assigned one.

For the mapping class group, it was a theorem of Dehn and Lickorish that the mapping class group $\Gamma(\Sigma)$ is finitely generated by Dehn twists along simple loops. The main result in [24], based on the work of [7], states that

Theorem 4.2 (Reconstruction of the mapping class group). *Each orientation preserving self-homeomorphism of a surface which fixes the boundary pointwise is isotopic to a composition of finitely many Dehn twists so that the composition is unique modulo cancellation laws supported in subsurfaces of level 1.*

Earlier work on the subject was done by Hatcher and Thurston [16] who showed among other things that the subsurfaces can be taken to be genus 2 with 3 holes and Gervais [7] who proved that subsurface can be taken to be genus 1 with 2 holes. Theorem 4.2 is just a simplification of the work of Gervais.

4.3 Thurston's embedding

Before discussing the related theorems for surfaces of levels 0 or 1, let us recall Thurston's embeddings of Teichmüller space $\mathcal{T}(\Sigma)$ and the measured lamination space $\text{ML}(\Sigma)$. Given an isotopy class $[d] \in \mathcal{T}(\Sigma)$, the *geodesic length function* $l_d: \mathcal{S}(\Sigma) \rightarrow \mathbb{R}_{\geq 0}$ sends an isotopy class of simple loop to the length of its geodesic representative. For a measured lamination $m \in \text{ML}(\Sigma)$, the *geometric intersection number function*, or simply *intersection function* $I_m: \mathcal{S}(\Sigma) \rightarrow \mathbb{R}_{\geq 0}$ is given by $I_m(\alpha) = I(\alpha, m)$. Thurston's embedding $\text{Th}: \mathcal{T}(\Sigma) \rightarrow \mathbb{R}_{\geq 0}^{\mathcal{S}(\Sigma)}$ (resp. $\text{Th}: \text{ML}(\Sigma) \rightarrow \mathbb{R}_{\geq 0}^{\mathcal{S}(\Sigma)}$) sends the isotopy class of a metric to its geodesic length function, i.e., $\text{Th}([d]) = l_d$ (resp. sends a measured lamination to its intersection function). The fact that the map Th is injective for Teichmüller space was a result of Fricke and Klein. The works of Okumura [32], [33] and Schmutz [36] determine the smallest finite set $F \subset \mathcal{S}(\Sigma)$ so that the restriction $l_d|_F$ determines the metric d . See also Hamenstädt [11] who simplified Schmutz's proof. A result of Thurston shows that $9g - 9$ simple loops suffice to determine the intersection function for closed surfaces of genus g ([39], [6]). But the number $9g - 9$ is not the smallest. Hamenstädt in [11] showed that $6g - 5$ simple loops suffice. It is unknown whether $6g - 6$ suffice.

4.4 Level-0 surfaces

For the level-0 surface, i.e., the 3-holed sphere, the space of simple loops $\mathcal{S}(\Sigma_{0,3})$ consists of isotopy classes of the three boundary components. The Teichmüller space $\mathcal{T}(\Sigma_{0,3})$, the measured lamination space $\text{ML}(\Sigma_{0,3})$ and the mapping class group $\Gamma(\Sigma_{0,3})$ can be described as follows. By a theorem of Fricke and Klein mentioned

before, each isotopy class of a hyperbolic metric on $\Sigma_{0,3}$ is determined by the lengths of three boundary components and these lengths are subject to no constraints, i.e., $\text{Th}(\mathcal{T}(\Sigma_{0,3})) = \mathbb{R}_{>0}^{S(\Sigma_{0,3})}$. For an exposition of the work of Fricke and Klein, we refer the reader to Chapter 15 of this Handbook written by W. Goldman, [9]. The work of Thurston shows that each measured lamination is determined by its intersection number with three boundary components and these three numbers subject to no constraints, i.e., $\text{Th}(\text{ML}(\Sigma_{0,3})) = \mathbb{R}_{>0}^{S(\Sigma_{0,3})}$. Dehn proved in 1938 that $\Gamma(\Sigma_{0,3})$ is isomorphic to the free abelian group on three generators which are the Dehn twists on three boundary components.

With these results on the 3-holed sphere, one can restate Theorem 4.1 in an equivalent form as follows. Given a surface Σ of level at least 2, a real valued function on the set of simple loops $S(\Sigma)$ is a geodesic length function (resp. an intersection function) if and only if for each essential surface Σ' of level 1, the restriction of the function to $S(\Sigma')$ is.

4.5 Level-1 subsurfaces

For level-1 surfaces Σ , i.e., the 1-holed torus $\Sigma_{1,1}$ and 4-holed sphere $\Sigma_{0,4}$, the set of simple loops $S(\Sigma)$, the Teichmüller space $\mathcal{T}(\Sigma)$ and the mapping class group $\Gamma(\Sigma)$ are essentially the same as those of the torus. To be more precise, let us consider the subset $S'(\Sigma) \subset S(\Sigma)$ of isotopy classes of simple loops which are not homotopic into the boundary $\partial\Sigma$, i.e., $S'(\Sigma) = S(\Sigma) \cap \text{CS}(\Sigma)$. There exists a natural bijection i_* between $S'(\Sigma_{1,1})$ and $S(\Sigma_{1,0})$ induced by the inclusion map from $\Sigma_{1,1}$ to $\Sigma_{1,0}$. This isomorphism preserves the intersection pairing. For the 4-holed sphere $\Sigma_{0,4}$, there exists a natural isomorphism P^* between $S'(\Sigma_{0,4})$ and $S'(\Sigma_{1,1})$ which satisfies $I(\alpha, \beta) = 2I(P^*(\alpha), P^*(\beta))$. It is defined as follows. Let τ be a hyperelliptic involution on the 1-holed torus $\Sigma_{1,1}$ and let $P: \Sigma_{1,1} \rightarrow \Sigma_{1,1}/\tau$ be the quotient map where $\Sigma_{1,1}/\tau$ is the disc with three cone points of order two (an orbifold). It is well known that the hyperelliptic involution τ preserves the isotopy class of each simple loop and τ commutes with each homeomorphism up to isotopy. Let the 4-holed sphere $\Sigma_{0,4}$ be the subsurface of $\Sigma_{1,1}/\tau$ with three small disc neighborhoods of the cone points removed. Then the isomorphism P^* from $S'(\Sigma_{0,4})$ to $S'(\Sigma_{1,1})$ sends the isotopy class $[a]$ to $[b]$ where b is a component of $P^{-1}(a)$. To summarize, for a level-1 surface Σ , there exists a bijection π from $S'(\Sigma)$ to $\mathbb{Q}P^1$ so that $\pi(\alpha) = p/q$ and $\pi(\beta) = p'/q'$ satisfy $pq' - p'q = \pm 1$ if and only if $I(\alpha, \beta) = 1$ for $\Sigma_{1,1}$ and 2 for $\Sigma_{0,4}$. Draw a hyperbolic geodesic in the upper-half plane ending at p/q and p'/q' when $pq' - p'q = \pm 1$. One obtains the so called "modular configuration" (see Figure 7). Call three elements in $S'(\Sigma)$ forming an triangle if they correspond to the vertices of an ideal triangle in the modular configuration and call four elements in $S'(\Sigma)$ forming a quadrilateral if they correspond to the vertices of an ideal quadrilateral. The modular structure on the space of simple loops $S'(\Sigma)$ for level-1 surfaces was known

to Fricke and Klein ([19]) and to Dehn ([4]) who used the rational numbers to code the set $S'(\Sigma)$. See also [16], [38], [41] and others.

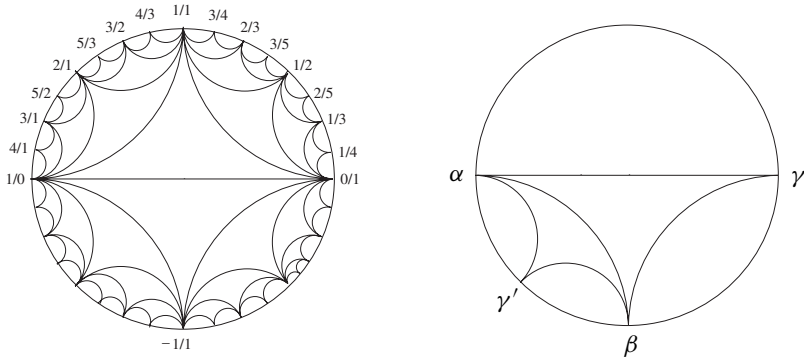


Figure 7

A special feature of the modular configuration is the huge symmetry built in the configuration. This is, in our view, one reason why the set of homotopy class of simple loops on the surface is more useful than the fundamental group in establishing the reconstruction principle for many structures. Suppose we take four vertices $(\alpha, \beta, \gamma; \gamma')$ forming a quadrilateral in $S'(\Sigma)$ so that both (α, β, γ) and (α, β, γ') are triangles. Then there is an orientation reversing involution of $\mathbb{Q}P^1$ leaving the quadrilateral invariant and interchanging γ and γ' . This involution is realized by an orientation reversing involution of the surface Σ which is the reflection of Figure 8 (where $\gamma = \alpha\beta$) about the yz -plane. On the other hand, given any triangle (α, β, γ)

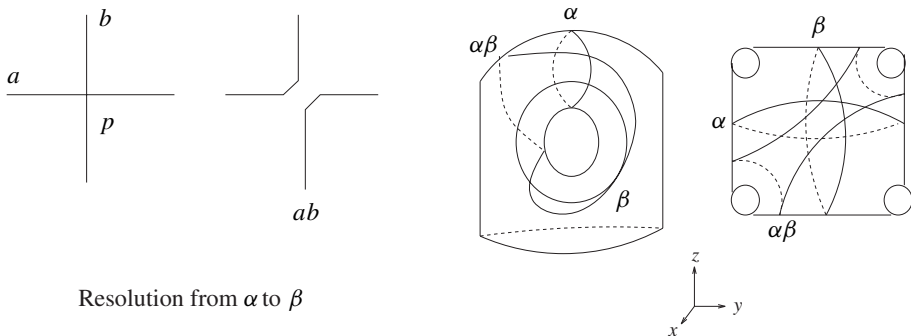


Figure 8. Right-hand orientation on the front faces

in the modular configuration $S'(\Sigma)$, there is a \mathbb{Z}_3 action on $\mathbb{Q}P^1$ which permutes the three vertices. Thus there is a \mathbb{Z}_3 action on the surface Σ permuting the isotopy classes. This symmetry is illustrated in Figure 9 below where the 1-holed torus is the Seifert

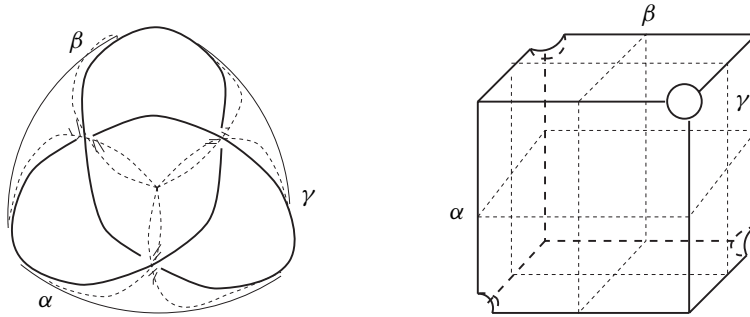


Figure 9. The three-fold symmetry in the modular configuration viewed in three-space.

surface of the trefoil knot and the 4-holed sphere is the truncated boundary surface of a cube. The symmetry involved in the 4-holed sphere is huge which is difficult to visualize in Figure 8. Indeed, as one can see from Figure 9, any permutation of the four boundary components is realized by a homeomorphism preserving the set $\{\alpha, \beta, \gamma\}$, i.e., the permutation group on four letters acts on $\Sigma_{0,4}$ preserving the set $\{\alpha, \beta, \gamma\}$.

As an application of this 24-fold symmetry, we consider the trace relations for $SL(2, \mathbb{C})$ matrices. In this case, the analogous question to Problem 2 for triangles and quadrilaterals is the following. Given three complex numbers a, b, c , do there exist two matrices $A, B \in SL(2, \mathbb{C})$ so that $\text{tr}(A) = a$, $\text{tr}(B) = b$ and $\text{tr}(AB) = c$? It is well known that the answer is positive. Next, in analogy to six edge lengths of a quadrilateral, given seven complex numbers $a_1, a_2, a_3, a_{12}, a_{23}, a_{31}$ and a_{123} , under what condition do there exist three $SL(2, \mathbb{C})$ matrices A_1, A_2 , and A_3 so that $\text{tr}(A_i) = a_i$, $\text{tr}(A_i A_j) = a_{ij}$ and $\text{tr}(A_1 A_2 A_3) = a_{123}$? A solution by Fricke–Klein [5] and Vogt [42] was the following. These three matrices exist if and only if $a_{123}^2 - a_{123}(a_1 a_{23} + a_2 a_{31} + a_3 a_{12} - a_1 a_2 a_3) + a_1^2 + a_2^2 + a_3^2 + a_{12}^2 + a_{23}^2 + a_{31}^2 + a_{12} a_{23} a_{31} - a_1 a_2 a_{12} - a_2 a_3 a_{23} - a_3 a_1 a_{31} - 4 = 0$. This equation, as it stands, is quite complicated. One can easily notice the 3-fold symmetry of the equation under cyclic permutation of $\{a_1, a_2, a_3\}$. In fact, there exists a 24-fold symmetry in the equation. Namely, the equation is invariant under any permutation of $\{a_1, a_2, a_3, a_{123}\}$. This can be seen using the modular configuration. Indeed, if we choose the generators of the fundamental group of the 4-holed sphere carefully (see for instance, Figure 5 in [25]), then the four boundary components are represented by $x_1, x_2, x_3, x_1 x_2 x_3$ and three simple loops forming a triangle in the modular configuration by $x_1 x_2, x_2 x_3$ and $x_3 x_1$. The first equation in Theorem A.4 (b) in Appendix A is a rewriting of the above polynomial equation in terms of the modular configuration.

4.6 Constraints on level-1 surfaces

Teichmüller space, measured lamination space and the mapping class groups for level-1 surfaces can be explicitly constructed from the modular configuration on the set of simple loops. To be more precise, the geodesic length functions, the geometric intersection number functions and the Dehn twists satisfy universal relations at the vertices of triangles and quadrilaterals in $\mathbb{Q}P^1$. Furthermore, these universal relations form a complete set of relations. See Appendix A for the list of universal relations.

The relations for the Dehn twists were found by Dehn in [4]. D. Johnson [18] independently rediscovered the lantern relation (relation (IV) in Theorem A.3 in Appendix A) in 1979. Dehn also proved that these relations are complete for the mapping class group of level-1 surfaces. The relations for the geodesic length function were essentially discovered by Fricke and Klein [5] and Vogt [42] (though they were not stated in terms of modular relations). These are derived from the trace identities for $SL(2, \mathbb{C})$ matrices. That the set of all relations is complete was proved by Keen [20] for the 1-holed torus and was proved in [25] for 4-holed spheres using Maskit combination theorem. The relations for measured laminations are just degenerations of the relations for hyperbolic metrics and they are shown to be complete in [26].

4.7 Relationship between the 1-holed torus and the 4-holed sphere

The relationship between Teichmüller spaces and the mapping class groups among level-1 surfaces becomes clearer if one considers the Teichmüller spaces $\mathcal{T}_{1,1}$ and $\mathcal{T}_{0,4}$ of complete hyperbolic metrics with cups ends (on the open surface), and the reduced mapping class group $\Gamma^*(\Sigma)$ which is the quotient of the mapping class group by the subgroup generated by Dehn twists on boundary components. The key fact is that the hyperelliptic involution τ on $\Sigma_{1,1}$ induces the identity map on both the Teichmüller space $\mathcal{T}_{1,1}$ and $\mathcal{S}(\Sigma_{1,1})$ and is in the center of the mapping class group. Indeed, one has a natural biholomorphism between $\mathcal{T}_{1,1}$ and $\mathcal{T}_{0,4}$ induced by the pull back map $P: \Sigma_{1,1} \rightarrow \Sigma_{1,1}/\tau$. A natural isomorphism from $\Gamma^*(\Sigma_{1,1})$ to $\Gamma(\Sigma_{1,0}) = SL(2, \mathbb{Z})$ is induced by inclusion of $\Sigma_{1,1}$ to $\Sigma_{1,0}$. Since the hyperelliptic involution τ commutes with each homeomorphism, there is a monomorphism from the reduced mapping class group $\Gamma^*(\Sigma_{0,4})$ to $\Gamma^*(\Sigma_{1,1})/\langle\tau\rangle = PSL(2, \mathbb{Z})$ whose image is the principal congruence subgroup of order 2.

4.8 Grothendieck's view

It is instructive to read [10] on related topics. We cite the paragraph on page 248ff. in [10] (with English translation by L. Schneps). “There is a striking analogy, and I am certain it is not merely formal, between this principle and the analogous principle of Demazure for the structure of reductive algebraic groups, if we replace the term ‘level’ or ‘modular dimension’ with ‘semi-simple rank of the reductive group’. The

link becomes even more striking if we recall that the mapping class group $\Gamma_{1,1}^*$ is no other than $SL(2, \mathbb{Z})$, i.e., the group of integral points of the simple group scheme of 'absolute' rank 1 $SL(2)_{\mathbb{Z}}$. Thus, the fundamental building block for the Teichmüller tower is essentially the same as for the tower of reductive groups of all ranks – a group of which, moreover, we may say that it is doubtless in all the essential disciplines of mathematics.”

5 The space of simple loops on surfaces and the modular structure

Unlike subsurfaces in a surface, simple loops on surfaces have been the focus of more attention for a long time. Indeed, most of the surface problems can be reduced to ones concerning simple loops and the proofs of Theorems 4.1 and 4.2 are no exception. The topological investigation of the set $\mathcal{S}(\Sigma)$ of isotopy classes of essential simple loops began in Dehn's work [4] on the mapping class groups. As an example of use of simple loops to solve surface problems, let us recall the elegant proof of Dehn that the (reduced) mapping class group $\Gamma^*(\Sigma_{0,4})$ is the free group on two generators generated by Dehn twists on two simple loops intersecting at two points. Dehn first observed that the set $\mathcal{S}'(\Sigma_{0,4})$ of essential simple loops not homotopic into the boundary forms the modular configuration $\mathbb{Q}P^1$ and the mapping class group $\Gamma^*(\Sigma_{0,4})$ acts on the modular configuration faithfully preserving both the modular relation and the orientation. Thus $\Gamma^*(\Sigma_{0,4})$ is a subgroup of the modular group $PSL(2, \mathbb{Z})$. Since each boundary component of $\Sigma_{0,4}$ is fixed by the mapping class group elements, $\Gamma^*(\Sigma_{0,4})$ is actually in the principal congruence subgroup of level 2 generated by the two matrices $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$. But these two matrices correspond to two Dehn twists mentioned above.

To go from simple loops to subsurfaces, one takes the regular neighborhood of a union of simple loops. In this way, it can be shown for instance that given any two level-1 essential subsurfaces A, B , there is a sequence of level-1 essential subsurfaces starting from A and ending at B so that any two adjacent level-1 subsurfaces overlap in an essential level-0 subsurface.

5.1 The works of Dehn and Lickorish

The works of Dehn [4] and Lickorish [21] already suggested strongly that level-1 subsurfaces are fundamental in simplifying the intersections of two simple loops. Indeed, Lemma 2 in [21] states that if two simple loops a, b satisfy either $|a \cap b| \geq 3$ or $|a \cap b| = 2$ with non-zero algebraic intersection number, then there is a Dehn twist which sends b to a new loop having fewer intersection points with a . Thus the only situations which cannot be simplified are: 1) a, b are disjoint, 2) a intersects b at one point and 3) a intersects b at two points of different intersection signs. In these cases, the lowest-level connected subsurface which contains both a and b is either a level-0

or a level-1 subsurface. Furthermore, the pair of curves a, b satisfying condition 2) or 3) corresponds to the basic relation in the modular configuration.

5.2 A modular structure

We mentioned in several places the notion of modular structure. Here is a formal definition after Thurston's geometric structures on manifolds.

Definition 5.1. (a) A $(\mathbb{Q}P^1, \text{PSL}(2, \mathbb{Z}))$ modular structure on a set X is a maximal collection of charts $\{(U_i, \phi_i) \mid \phi_i : U_i \rightarrow \mathbb{Q}P^1 \text{ is injective}\}$ so that the following hold.

(1) $X = \bigcup_i U_i$.

(2) The transition function $\phi_i \phi_j^{-1}$ is the restriction of an element in $\text{PSL}(2, \mathbb{Z})$.

(b) A modular structure on a set X is called *compact* if the group of bijections of X which preserve the modular structure acts on X with finite orbits.

It seems that compactness is essential for developing a useful “function theory” on a set with a modular structure. All interesting examples that we encounter have compact modular structures.

For an oriented surface Σ of level at least 1, the set $S'(\Sigma)$ of isotopy classes of non-boundary parallel essential simple loops on Σ has a natural compact modular structure invariant under the action of the mapping class group. A special collection of charts for the modular structure is given by $(S'(\Sigma'), \phi_{\Sigma'})$ where Σ' is an essential level-1 subsurface and $\phi_{\Sigma'} : S'(\Sigma') \rightarrow S'(\Sigma_{1,1}) = \mathbb{Q}P^1$ is a bijection induced by either an orientation preserving homeomorphism or by an orientation preserving quotient map (see §4.6). To see that Condition (2) in the definition holds, one simply notes that if two essential level-1 subsurfaces intersect at two non-homotopic simple loops, then they are isotopic. To see the compactness, we note that the mapping class group acts on $S'(\Sigma)$ preserving the modular structure and the action of the mapping class group has finite orbits. Thus we can talk about triangles and quadrilaterals in $S'(\Sigma)$. Furthermore, since the set of rational numbers $\mathbb{Q}P^1$ has a natural orientation invariant under $\text{PSL}(2, \mathbb{Z})$, we can talk about oriented triangles in $S'(\Sigma)$.

Another example of a compact modular structure is the set of isotopy classes of 3-holed sphere decompositions of a surface. This set is related to the Heegaard splittings of 3-manifolds and Weil–Petersson geometry. See Appendix B for more detail.

5.3 Resolution of intersection

One way to see the modular structure on the space of simple loops $S(\Sigma)$ is to use the notion of resolution of intersection points. Recall that two rational numbers p/q and p'/q' are modular related if $pq' - p'q = \pm 1$ and are denoted by $p/q \perp p'/q'$. Two isotopy classes α and β of curves on surfaces corresponding to a modular related pair

are denoted by $\alpha \perp \beta$ or $\alpha \perp_0 \beta$. Here $\alpha \perp \beta$ means that their intersection number $I(\alpha, \beta) = 1$ and $\alpha \perp_0 \beta$ means that $I(\alpha, \beta) = 2$ so that their algebraic intersection number is zero. To find out the vertices of ideal triangles based on $\{\alpha, \beta\}$, we use the resolutions of intersections. Recall that surfaces are oriented and simple loops are not oriented. If a, b are two arcs intersecting at one point transversely, then the *resolution of $a \cup b$ at the intersection point from a to b* is defined as follows. Fix any orientation on a and use the orientation on the surface to determine an orientation on b . Then resolve the intersection according to the orientations (see Figure 8). The resolution is independent of the orientation chosen on a . If $\alpha \perp \beta$ or $\alpha \perp_0 \beta$, take $a \in \alpha$ and $b \in \beta$ so that $|a \cap b| = I(\alpha, \beta)$. Then the curve obtained by resolving all intersection points in $a \cap b$ from a to b is again an essential non-boundary parallel simple loop. We denote its isotopy class by $\alpha\beta$. One sees easily that positively oriented triangles and quadrilaterals in the modular structure on $S'(\Sigma)$ are exactly $(\alpha, \beta, \alpha\beta)$ and $(\alpha, \beta, \alpha\beta, \beta\alpha)$. If $\alpha \perp \beta$ or $\alpha \perp_0 \beta$, we use $\partial(\alpha \cup \beta)$ to denote the isotopy class of the boundary of a regular neighborhood of $a \cup b$. In terms of these notations, all universal relations for the geodesic length functions, the intersection functions and the Dehn twists are expressed in terms of $\alpha, \beta, \alpha\beta, \beta\alpha$ and the components of $\partial(\alpha \cup \beta)$. For instance, the relations for the Dehn twists are: 1) if $\alpha \perp \beta$, then $D_\alpha D_\beta = D_\beta D_{\alpha\beta}$ and $(D_\alpha D_\beta D_{\alpha\beta})^4 = D_{\partial(\alpha \cup \beta)}$, and 2) if $\alpha \perp_0 \beta$, then $D_\alpha D_\beta D_{\alpha\beta} = D_{\partial(\alpha \cup \beta)}$ (the lantern relation). Since the modular relation $(\mathbb{Q}P^1, \perp)$ has a \mathbb{Z}_3 -symmetry leaving an ideal triangle invariant, we obtain $\alpha(\beta\alpha) = (\alpha\beta)\alpha = \beta$.

5.4 A lemma of Lickorish

One of the most useful properties of the modular structure on $S'(\Sigma)$ is the following lemma (Lemma 7 in [25]) which generalizes Lickorish's Lemma 2 in [21]. It states that given two intersecting elements $\alpha, \beta \in S'(\Sigma)$ which are not related by the modular relation \perp or \perp_0 , then we can write $\beta = \gamma_1 \gamma_2$ with $\gamma_1 \perp \gamma_2$ or $\gamma_1 \perp_0 \gamma_2$ so that (1) $I(\alpha, \gamma_i) < I(\alpha, \beta)$ and $I(\alpha, \gamma_2 \gamma_1) < I(\alpha, \beta)$ for $i = 1, 2$ and (2) if $\gamma_1 \perp_0 \gamma_2$, then for each component δ of $\partial(\gamma_1 \cup \gamma_2)$ we have $I(\alpha, \delta) < I(\alpha, \beta)$. As an easy consequence, one shows that the reconstruction principle for Teichmüller spaces follows from Theorem 4.1 for level-2 surfaces. As another consequence, one shows that the space $\mathcal{S}(\Sigma)$ is finitely generated in the following strong sense: There is a finite subset X_0 in $\mathcal{S}(\Sigma)$ so that $\mathcal{S}(\Sigma) = \bigcup_{n=0}^{\infty} X_n$ where $X_{i+1} = X_i \cup \{\alpha \mid \alpha = \gamma_1 \gamma_2 \text{ where either (1) } \gamma_1 \perp \gamma_2, \text{ and } \gamma_1, \gamma_2, \gamma_2 \gamma_1 \text{ are in } X_i \text{ or (2) } \gamma_1 \perp_0 \gamma_2 \text{ and } \gamma_1, \gamma_2, \gamma_2 \gamma_1 \text{ and each component of } \partial(\gamma_1 \cup \gamma_2) \text{ are in } X_i\}$.

5.5 Level-2 surfaces

The proof of the reconstruction theorem for level-2 surfaces has always been one of the key steps in establishing the reconstruction principle. In dealing with simple loops on level-2 surfaces, the following collection of five curves (the pentagon relation)

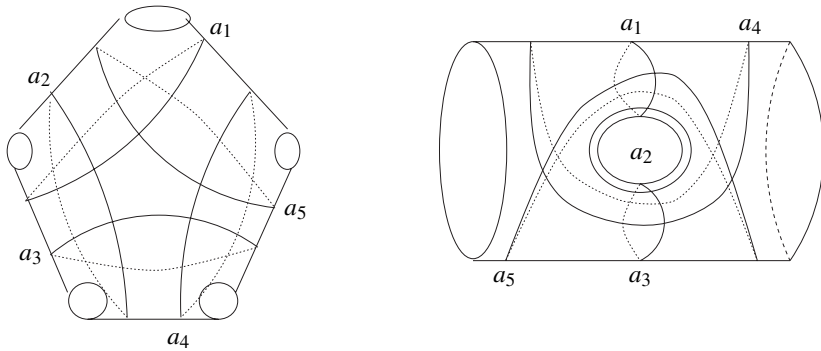


Figure 10. The pentagon relations.

$\{\alpha_1, \dots, \alpha_5 \text{ in } S'(\Sigma) \mid I(\alpha_i, \alpha_j) = 0 \text{ for indices } |i - j| \not\equiv 1 \pmod{5}\}$ (see Figure 10) appears constantly and plays an important role. For the 5-holed sphere, one has the following relation $\alpha_i \alpha_{i+1} \alpha_{i+2} = \alpha_{i+3} \alpha_{i+4}$ and for the 2-holed torus, we have $\alpha_1 \alpha_2 \alpha_3 \alpha_4 = \alpha_3 \alpha_2 \alpha_1$ (see [26]). These five curves for the 5-holed sphere were first observed by Dehn in [4] who showed that the Dehn twists on them generate the reduced mapping class group of the 5-holed sphere. Furthermore, these five curves are rigid in the sense that any other collection of five curves with the same disjointness property is the image of $\{\alpha_i\}$ under a homeomorphism ([28]).

5.6 A multiplicative structure on the space of curve systems

We finish this section with an application of the notion of intersection resolving to a multiplicative structure on the space of curve systems $\text{CS}(\Sigma)$. Given two curve systems a, b on an oriented surface with $|a \cap b| = I([a], [b])$, the multiplication ab is defined to be the 1-dimensional submanifold obtained by resolving all intersection points in $a \cap b$ from a to b . It can be shown that ab is again a curve system (see Appendix C for a simple proof when a, b contain no arcs). This induces a multiplicative structure on $\text{CS}(\Sigma)$ by defining $\alpha\beta = [ab]$ where $a \in \alpha, b \in \beta$ and $|a \cap b| = I(\alpha, \beta)$. For instance the Dehn twist on a simple loop α applied to β is given by $D_\alpha(\beta) = \alpha^k \beta$ where $k = I(\alpha, \beta)$. This multiplication is natural with respect to the action of the mapping class group and is highly non-commutative. Indeed, if α contains no arc component, then $\alpha\beta = \beta\alpha$ implies $I(\alpha, \beta) = 0$. As a consequence of this, one obtains a new proof of a result of Ivanov [17] that Dehn twists on two intersecting isotopy classes of simple loops can never commute up to isotopy. The most interesting property of the multiplication seems to be the ‘‘cancellation law’’ saying that if each component of α is not an arc and intersects β , then $\alpha(\beta\alpha) = (\alpha\beta)\alpha = \beta$. This is a generalization of the \mathbb{Z}_3 -symmetry in the modular configuration. As an application of the cancellation law, let us prove a weak form of a result of Thurston [41] that if α and β are two surface filling simple loops (i.e., $I(\alpha, \gamma) + I(\beta, \gamma) > 0$ for all $\gamma \in$

$S'(\Sigma)$), then the self-homeomorphism $f = D_\alpha^{-1}D_\beta$ does not leave any curve system invariant up to isotopy (in fact Thurston proved that f is pseudo-anosov). Indeed, if f leaves an element $\gamma \in \text{CS}(\Sigma)$ invariant, then $D_\alpha(\gamma) = D_\beta(\gamma)$, i.e., $\alpha^k\gamma = \beta^l\gamma$. Now multiply the equation by γ from the left and use the cancellation law. One obtains a contradiction to the surface filling property.

6 Reduction to level-2 surfaces

The goal of this section is to establish a fairly general criterion to reducing problems concerning all surfaces to that of level-2 surfaces.

6.1 The restriction map and essential subsurfaces

We shall begin by some abstract definitions. Let Y be a set. Given a subset X of the set $Y^{S(\Sigma)}$ of all maps from $S(\Sigma)$ to Y , we say that the subset X has *property RP* if for any decomposition of the surface Σ into a union of two essential subsurfaces A_1 and A_2 which overlap in a level-0 essential subsurface, then the restriction map from $Y^{S(\Sigma)}$ to $Y^{S(A_1) \cup S(A_2)}$ is injective on X . In other words, if f, g are two elements in X such that $f|_{S(A_1) \cup S(A_2)} = g|_{S(A_1) \cup S(A_2)}$, then $f = g$. For simplicity, we call the function $f|_{S(A_i)}$ the *restriction* of f to A_i . We say that a subset $X \subset Y^{S(\Sigma)}$ with property RP is *complete* if for any two elements f_1 and f_2 in the restrictions of X to A_1 and A_2 so that their restrictions to the overlap $A_1 \cap A_2$ are the same, then there exists an element $f \in X$ whose restriction to A_i is f_i for $i = 1, 2$. For instance, the set of all geodesic length functions and the set of all intersection functions have complete property RP. This is equivalent to the following gluing lemma for hyperbolic metrics and measured laminations. Namely, suppose the surface Σ is a union of two essential subsurfaces A_1 and A_2 which overlap in an essential level-0 surface. If we are given two hyperbolic metrics d_i on A_i whose restrictions to the overlap of A_1 with A_2 are isotopic, then there is a hyperbolic metric unique up to isotopy on the surface Σ whose restriction to A_i is isotopic to d_i . The same gluing lemma holds for measured laminations. Note that $\text{SL}(2, \mathbb{C})$ characters do not have property RP due to the existence of reducible representations. The mapping class group $\Gamma(\Sigma)$ considered as a subset of $S(\Sigma)^{S(\Sigma)}$ does not have property RP either. But if one modifies the definition of $S(\Sigma)$ by taking the isotopy classes of all *oriented* simple loops, then the mapping class group has complete property RP.

6.2 A reduction lemma

The main reduction lemma says the following. If X is a subset of $Y^{S(\Sigma)}$ such that for each level-2 essential subsurface Σ' the restriction of X to $Y^{S(\Sigma')}$ has property RP, then X has property RP. See Appendix D for a proof of this reduction lemma. As a consequence of this reduction lemma, we have the following fact. Suppose Σ is a

surface of level at least three and X and X' are two subsets of $Y^{S(\Sigma)}$ so that for each level-2 essential subsurface Σ' the restrictions of X and X' to $Y^{S(\Sigma')}$ are the same. If furthermore $X \subset X'$ and X has complete property RP, then $X = X'$. To see this, we use induction on the levels of subsurfaces. First of all, by the reduction lemma, both X and X' have property RP. Now to show $X' \subset X$, take an element $x' \in X'$ and decompose Σ into a union of two essential surfaces A_1 and A_2 of smaller levels so that they overlap in a level-0 surface. By the induction hypothesis, we find two elements x_1 and x_2 which are in the restrictions of X to A_1 and A_2 so that x_i is the restriction of x' to A_i . But the restrictions of x_i to the overlap are the same, namely, it is the restriction of x' to the overlap. Thus by the completeness, there is an element x in X whose restrictions to A_i is x_i . Thus $x = x'$ by property RP. This shows $X' \subset X$.

By taking X to be the set of all geodesic length functions and X' to be the set of all real valued functions on $S(\Sigma)$ which satisfy the universal relations in Theorem A.1 in Appendix A, we see that the reconstruction principle for all Teichmüller spaces follows from that for level-2 surfaces.

Also, the reduction lemma shows that the problem on the automorphisms of the curve complex of a surface is essentially a problem on level-2 surfaces (see [28]).

6.3 Relationship to Grothendieck's idea

The above gives some hints on the special role played by level-2 surfaces. It also supports Grothendieck's principle that in the reconstruction process "relations are supported in level-2 surfaces".

7 $SL(2, \mathbb{C})$ representation variety of surface groups

An $SL(2, \mathbb{C})$ representation of a group is a homomorphism of the group into $SL(2, \mathbb{C})$. The character of the representation sends each group element to the trace of the representation matrix. If the group is the fundamental group of a surface, by using a result of Fricke and Klein [5] and Vogt [42], one shows that the character function is determined by its restriction to the set $S(\Sigma)$ of homotopy classes of simple loops. The main result in [27] shows that the character function on $S(\Sigma)$ satisfies the reconstruction principle, i.e., except for finitely many (at least 2^{n-1}) exceptional functions defined on $S(\Sigma_{0,n})$ for $n \geq 5$, a function on $S(\Sigma)$ is an $SL(2, \mathbb{C})$ character if and only if for each essential level-1 subsurface Σ' in Σ the restriction of the function to $S(\Sigma')$ is an $SL(2, \mathbb{C})$ character. An exceptional function $f : S(\Sigma_{0,n}) \rightarrow \mathbb{C}$ satisfies the following:

- (1) $f(S(\Sigma_{0,n})) = \{2, -2\}$,
- (2) for each level-1 subsurface, the restriction of f to the subsurface is a character,
- (3) there exists a level-2 subsurface Σ' so that $f|_{S(\Sigma')}$ is exceptional.

All exceptional functions are constructed from the basic one defined on $S(\Sigma_{0,5})$ which sends b_i to 2 and all others to -2 . There are no representations whose characters are these exceptional functions.

Given a surface Σ of level 1, $SL(2, \mathbb{C})$ characters on $S(\Sigma)$ are characterized by the trace identities on the vertices of triangles and quadrilaterals in the modular relation (see Appendix A for the exact statement). Thus the space of all $SL(2, \mathbb{C})$ characters of a surface group can be explicitly described. The reconstruction theorem also holds for any $SL(2, K)$ characters where the field K is quadratically complete (i.e., each quadratic equation with coefficients in K has roots in K).

7.1 Reducible representations

The main difficulty in establishing the reconstruction principle for $SL(2, \mathbb{C})$ characters is due to the existence of reducible representations. Recall that an $SL(2, \mathbb{C})$ representation is *reducible* if it leaves a 1-dimensional linear subspace in \mathbb{C}^2 invariant. Unlike the discrete faithful subgroups in $SL(2, \mathbb{R})$ which occur in Teichmüller theory, there are many irreducible representations of a surface group so that its restriction to a subgroup coming from an essential subsurface of negative Euler number is reducible. Now the reduction lemma (see §6.2) is valid only for representations so that their restrictions to the fundamental group of the intersection surface are irreducible. Thus one should choose the decomposition of a surface Σ as a union of two subsurfaces Σ_1 and Σ_2 carefully. It turns out that the following which plays a key role in choosing the decomposition of a surface is true. Namely, a representation of a surface group into $SL(2, \mathbb{C})$ is irreducible if and only if its restriction to the subgroup of an Euler number-1 essential subsurface is irreducible [27].

As a consequence, we obtain the following result concerning $SL(2, K)$ characters on any group. Suppose K is a field so that each quadratic equation with coefficients in K has roots in K . Given a group G , we are interested in finding all $SL(2, K)$ characters on G . In his work on $SL(2, \mathbb{R})$ characters, Helling [15] introduced the notion of *trace function*. Recall that a K valued function f on G is a trace function if any two elements x, y in G , $f(xy) + f(xy^{-1}) = f(x)f(y)$ and $f(\text{id}) = 2$. Evidently all $SL(2, K)$ characters on G are trace functions due to the trace identity $\text{tr}(AB) + \text{tr}(AB^{-1}) = \text{tr}(A)\text{tr}(B)$. One consequence of the characterization theorem is that each trace function is also a character.

7.2 The role of level-1 surfaces

The role of level-1 surfaces among all surfaces is similar to the role of 2-generator groups among all groups. For instance, by Jorgensen's inequality, a non-elementary subgroup in $SL(2, \mathbb{C})$ is discrete if and only if each of its 2-generator subgroups is discrete. The reconstruction theorem for Teichmüller space says that a faithful representation of a surface group into $SL(2, \mathbb{R})$ is discrete if and only if its restriction to each subgroup of its level-1 subsurface is discrete and uniformizing a surface of the

same type. It is natural to ask if a similar description exists for discrete close surface subgroup of $SL(2, \mathbb{C})$.

7.3 A question on high dimension

It is interesting to ask if the reconstruction principle holds for representations of surfaces groups into the general linear group $GL(n, \mathbb{C})$. To be more precise, suppose f is a complex valued function defined on the conjugacy classes of the fundamental group of the surface so that the restriction of f to the conjugacy classes of the fundamental group of each essential level-1 subsurface is a $GL(n, \mathbb{C})$ -character. Is f the character of some $GL(n, \mathbb{C})$ representation of the surface group?

Appendix A. The statement of the reconstruction theorems for level-1 surfaces

Given a hyperbolic metric d on a surface, the *trace* of the metric d is the function $2 \cosh(l_d/2)$ where l_d is the geodesic length function associated to d .

Theorem A 1. (a) For the surface $\Sigma_{1,1}$ with $b = \partial \Sigma_{1,1}$, a function $t : S(\Sigma_{1,1}) \rightarrow \mathbb{R}_{\geq 2}$ is a trace function of a hyperbolic metric if and only if the following hold:

$$\prod_{i=1}^3 t(\alpha_i) = \sum_{i=1}^3 t^2(\alpha_i) + t(b) - 2$$

and

$$t(\alpha_3)t(\alpha'_3) = \sum_{i=1}^2 t^2(\alpha_i) + t(b) - 2$$

where $(\alpha_1, \alpha_2, \alpha_3)$ and $(\alpha_1, \alpha_2, \alpha'_3)$ are distinct ideal triangles in $S'(\Sigma_{1,1})$.

(b) For the surface $\Sigma_{0,4}$ with $\partial \Sigma_{0,4} = \bigcup_{i=1}^4 b_i$, a function $t : S(\Sigma_{0,4}) \rightarrow \mathbb{R}_{\geq 2}$ is a trace function of a hyperbolic metric if and only if for each ideal triangle $(\alpha_1, \alpha_2, \alpha_3)$ so that (α_i, b_j, b_k) bounds a $\Sigma_{0,3}$ in $\Sigma_{0,4}$ the following hold:

$$\prod_{i=1}^3 t(\alpha_i) = \sum_{i=1}^3 t^2(\alpha_i) + \sum_{j=1}^4 t^2(b_j) + \prod_{j=1}^4 t(b_j) + \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^4 t(\alpha_i)t(b_j)t(b_k) - 4$$

and

$$t(\alpha_3)t(\alpha'_3) = \sum_{i=1}^2 t^2(\alpha_i) + \sum_{j=1}^4 t^2(b_j) + \prod_{j=1}^4 t(b_j) + \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^4 t(\alpha_i)t(b_j)t(b_k) - 4$$

where $(\alpha_1, \alpha_2, \alpha'_3)$ and $(\alpha_1, \alpha_2, \alpha_3)$ are two distinct ideal triangles in $S'(\Sigma_{0,4})$.

Part (a) of Theorem A.1 was a result of Fricke–Klein [5] and Keen [20], and part (b) was proved in [25]. See also [9].

Theorem A 2. (a) For the surface $\Sigma_{1,1}$, a function $f: \mathcal{S}(\Sigma_{1,1}) \rightarrow \mathbb{R}_{\geq 0}$ is an intersection function if and only if the following hold:

$$f(\alpha_1) + f(\alpha_2) + f(\alpha_3) = \max_{i=1,2,3} (2f(\alpha_i), f([\partial\Sigma_{1,1}]))$$

where $(\alpha_1, \alpha_2, \alpha_3)$ is an ideal triangle, and

$$f(\alpha_3) + f(\alpha'_3) = \max(2f(\alpha_1), 2f(\alpha_2), f([\partial\Sigma_{1,1}]))$$

where $(\alpha_1, \alpha_2, \alpha_3)$ and $(\alpha_1, \alpha_2, \alpha'_3)$ are two distinct ideal triangles.

(b) For the surface $\Sigma_{0,4}$ with $\partial\Sigma_{0,4} = b_1 \cup b_2 \cup b_3 \cup b_4$, a function $f: \mathcal{S}(\Sigma_{0,4}) \rightarrow \mathbb{R}_{\geq 0}$ is an intersection function if and only if for each ideal triangle $(\alpha_i, \alpha_2, \alpha_3)$ so that (α_i, b_s, b_r) bounds a $\Sigma_{0,3}$ in $\Sigma_{0,4}$ the following hold:

$$\sum_{i=1}^3 f(\alpha_i) = \max_{\substack{1 \leq i \leq 3 \\ 1 \leq s \leq 4}} (2f(\alpha_i), 2f(b_s), \sum_{s=1}^4 f(b_s), f(\alpha_i) + f(b_s) + f(b_r)),$$

$$f(\alpha_3) + f(\alpha'_3) = \max_{\substack{1 \leq i \leq 2 \\ 1 \leq s \leq 4}} (2f(\alpha_i), 2f(b_s), \sum_{s=1}^4 f(b_s), f(\alpha_i) + f(b_s) + f(b_r))$$

where $(\alpha_1, \alpha_2, \alpha_3)$ and $(\alpha_1, \alpha_2, \alpha'_3)$ are two distinct ideal triangles.

Theorem A.2 was proved in [26].

Below is the statement of the presentation of the mapping class group for all surfaces of negative Euler number (see [27], [7]).

Theorem A 3. For a compact oriented surface Σ of negative Euler number, the mapping class group $\Gamma(\Sigma)$ has the following presentation.

Generators: $\{D_\alpha \mid \alpha \in \mathcal{S}(\Sigma)\}$.

Relations:

- (I) $D_\alpha D_\beta = D_\beta D_\alpha$ if $\alpha \cap \beta = \emptyset$,
- (II) $D_{\alpha\beta} = D_\alpha D_\beta D_\alpha^{-1}$ if $\alpha \perp \beta$,
- (III) $(D_\alpha D_\beta D_{\alpha\beta})^4 = D_{\partial(\alpha \cup \beta)}$ if $\alpha \perp \beta$,
- (IV) $D_\alpha D_\beta D_{\alpha\beta} = D_{\partial(\alpha, \beta)}$ if $\alpha \perp_0 \beta$.

The characterization of the $SL(2, \mathbb{C})$ characters for surface group representations is given by the following. The theorem is proved by Fricke and Klein [5] and Vogt [42], thought stated in different terminologies. See for instance [8] and [27].

Theorem A 4. (a) For the surface $\Sigma_{1,1}$ with $b = \partial\Sigma_{1,1}$, a function $t : \mathcal{S}(\Sigma_{1,1}) \rightarrow \mathbb{C}$ is an $\mathrm{SL}(2, \mathbb{C})$ trace function if and only if the following hold:

$$\prod_{i=1}^3 t(\alpha_i) = \sum_{i=1}^3 t^2(\alpha_i) - t(b) - 2$$

and

$$t(\alpha_3) + t(\alpha'_3) = t(\alpha_1)t(\alpha_2)$$

where $(\alpha_1, \alpha_2, \alpha_3)$ and $(\alpha_1, \alpha_2, \alpha'_3)$ are distinct ideal triangles in $\mathcal{S}'(\Sigma_{1,1})$.

(b) For the surface $\Sigma_{0,4}$ with $\partial\Sigma_{0,4} = \bigcup_{i=1}^4 b_i$, a function $t : \mathcal{S}(\Sigma_{0,4}) \rightarrow \mathbb{C}$ is an $\mathrm{SL}(2, \mathbb{C})$ trace function if and only if for each ideal triangle $(\alpha_i, \alpha_2, \alpha_3)$ so that (α_i, b_j, b_k) bounds a $\Sigma_{0,3}$ in $\Sigma_{0,4}$ the following hold:

$$\prod_{i=1}^3 t(\alpha_i) = -\sum_{i=1}^3 t^2(\alpha_i) - \sum_{j=1}^4 t^2(b_j) - \prod_{j=1}^4 t(b_j) + \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^4 t(\alpha_i)t(b_j)t(b_k) + 4$$

and

$$t(\alpha_3) + t(\alpha'_3) = -t(\alpha_1)t(\alpha_2) + \frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^4 t(\alpha_i)t(b_j)t(b_k)$$

where $(\alpha_1, \alpha_2, \alpha'_3)$ and $(\alpha_1, \alpha_2, \alpha_3)$ are two distinct ideal triangles in $\mathcal{S}'(\Sigma_{0,4})$.

Appendix B. The modular structure on the space of 3-holed sphere decompositions

The other natural example of compact modular structure is the set $\mathrm{HD}(\Sigma)$ of all isotopy classes of 3-holed sphere decompositions of a surface Σ . The charts are constructed as follows. Suppose (a_1, \dots, a_k) is an element in $\mathrm{HD}(\Sigma)$. Take an essential subsurface Σ' of level 1 so that all but one, say a_i , of the coordinates are disjoint from Σ' . Now the chart associated to Σ' is the set of elements $\{(a_1, \dots, a_{i-1}, b_i, a_{i+1}, \dots, a_k) \in \mathrm{HD}(\Sigma) | b_i \in \mathcal{S}'(\Sigma')\}$ with chart map sending the element to the slope of b_i . Again if two charts overlap in two elements, they coincide. A result of Hatcher–Thurston [16] says that given any two elements in $\mathrm{HD}(\Sigma)$ there is a sequence of charts whose union contains these two elements so that any two adjacent charts overlap in at least one element. On the other hand, each element in $\mathrm{HD}(\Sigma)$ determines a handlebody structure on the surface Σ obtained by attaching 2-cells to the components of the 3-holed sphere decomposition and then 3-cells. Evidently if two elements in $\mathrm{HD}(\Sigma)$ lie in a chart associated to a 4-holed sphere, then they determine the same handlebody structure. The main result in [28] shows that the converse is also true. Namely, if two elements in $\mathrm{HD}(\Sigma)$ determine the same handlebody structure, then there is a sequence

of charts associated to 4-holed spheres whose union contains these two elements so that any two adjacent charts overlap in at least one element.

Appendix C. A simple proof that multiplication produces a curve system

For simplicity, let us assume that the surface is closed (see [26] for general cases). We will give a simple proof of the following fact used in §5. Namely, if a and b are two curve systems that intersect minimally in their isotopy classes, then the 1-dimensional submanifold ab obtained by resolving all intersection points in $a \cup b$ from a to b is again a curve system.

Suppose otherwise that the 1-submanifold ab contains a null homotopic component c . By the Jordan curve theorem, we see that c is the boundary of an embedded disk D in the surface. Replacing c by components of ab in the interior of D if necessary, we may assume that c is the "inner-most" component, i.e., there are no other components of ab in the interior of D . By the definition of resolution, the disk D is obtained as follows. There exists a collection of components A_1, \dots, A_k of $\Sigma - (a \cup b)$ so that, after we resolve all intersection points of a and b , these components A_1, \dots, A_k are joint at some of their vertices to form D . Each component A_i is an open disc since c is the inner-most. The boundary of A_i consists of arcs in a and b , and the corners (or vertices) of A_i correspond to the intersection points of a and b . Let us call each A_i a polygon bounded by sides in a and b alternatively. Since a intersects b minimally within their isotopy classes, each A_i has at least four sides. Now by the definition of the resolution, the disc D is obtained by resolving corners of A_i 's from a to b . Considering the resolutions at the vertices along the boundary of A_i , one sees that corners open and close alternatively in a cyclic order on the boundary. (See Figure 11.) Form a graph in D by assigning a vertex to each A_i , so that this vertex lies in A_i and joining an edge between two vertices if their corresponding polygons A_i and A_j have the same vertex which is opened by the resolution. Then, on one hand, the graph is a tree since it is homotopic to the disk D . On the other hand, each vertex of the graph has valence at least two since the valence of a vertex is half of the number of sides of the corresponding polygon A_i (by the alternating property). This contradicts the fact that a tree must have a vertex of valence one.

Appendix D. A proof of the reduction lemma in §6.2

We shall prove the following reduction lemma stated in §6.2. Suppose Σ is a surface of level at least 3. If X is a subset of $Y^{S(\Sigma)}$ so that for each level-2 essential subsurface Σ' the restriction of X to $Y^{S(\Sigma')}$ has property RP, then X has property RP.

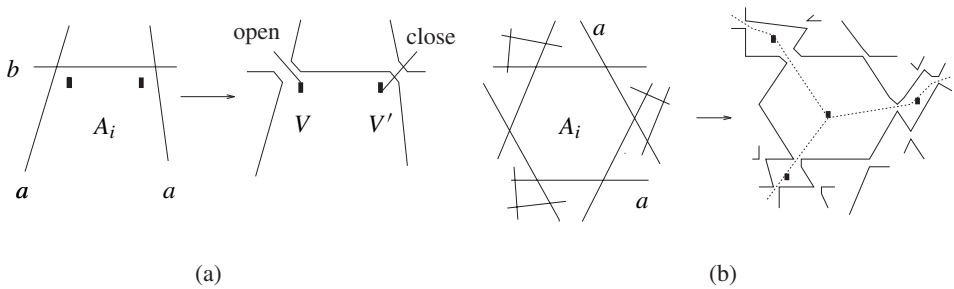


Figure 11. The pentagon relations.

To begin the proof, suppose the surface Σ is decomposed into a union of two essential subsurfaces A_1 and A_2 overlapping in an essential level-0 surface and we are given two elements f, g in X whose restrictions to A_i are the same. The goal is to show that $f = g$. To this end, let us construct a 3-holed sphere decomposition $\{a_1, \dots, a_k\}$ of the surface Σ so that each 3-holed sphere in the decomposition is either in A_1 or in A_2 and $A_1 \cap A_2$ is bounded by a_i 's. Thus if s is an element in $\mathcal{S}(\Sigma)$ which intersects only one element of $\{a_1, \dots, a_k\}$, then s is in $\mathcal{S}(A_i)$ for $i = 1$ or 2 . In particular $f(s) = g(s)$. Now suppose we make an elementary move on $\{a_1, \dots, a_k\}$ to produce a new 3-holed sphere decomposition $\{b_1, \dots, b_k\}$ where all but one of b_i are a_i and the exceptional component, say b_j is modular related to a_j (i.e., $b_j \perp a_j$ or $b_j \perp_0 a_j$) (these moves were introduced in the appendix of [16]). We claim that if $f(s) = g(s)$ for all elements s which intersect at most one of $\{a_1, \dots, a_k\}$ and if $\{b_1, \dots, b_k\}$ is obtained from $\{a_1, \dots, a_k\}$ by an elementary move, then $f(s) = g(s)$ for all elements s which intersect at most one of $\{b_1, \dots, b_k\}$. Indeed, by the property RP for level-2 surfaces, we see that $f(s) = g(s)$ for all elements s inside any level-2 subsurface which is bounded by elements in $\{a_1, \dots, a_k\}$. Now if s is an isotopy class which intersects at most one element in $\{b_1, \dots, b_k\}$, then s intersects at most two elements in $\{a_1, \dots, a_k\}$. Thus the isotopy class s is in a level-2 subsurface which is bounded by elements in $\{a_1, \dots, a_k\}$. Thus $f(s) = g(s)$. Now by the result in [16] that any two 3-holed sphere decompositions of the surface are related by a finite sequence of elementary moves, it follows that $f(s) = g(s)$ for any element s in $\mathcal{S}(\Sigma)$.

In view of the importance of the 3-holed sphere decompositions, it is tempting to make a 2-dimensional cell-complex Z based on 3-holed sphere decompositions of the surfaces as follows. The vertices of Z are the isotopy classes of 3-holed sphere decompositions of the surface and the edges are those pair of vertices related by an elementary move. Now attach a 2-cell to each 5-gon associated to each pentagon relation (see Figure 8), a 2-cell to each 4-gon associated to four elementary moves which are supported in two disjoint level-1 surfaces, and a 2-cell to each 3-gon associated to three elementary moves supported in a level-1 surface. This cell-complex was implicitly introduced in the appendix of [16]. The simple connectivity of this cell-complex

seems to be asserted in [16]. Hatcher defined, later on, a related complex, called the *pants decomposition complex*, and he showed that it is simply connected (see [13] and [14]). See also [12] for related topics.

References

- [1] J. S. Birman, *Braids, links, and mapping class groups*. Ann. of Math. Stud. 82, Princeton University Press, Princeton, NJ, 1975.
- [2] F. Bonahon, Shearing hyperbolic surfaces, bending pleated surfaces and Thurston's symplectic form. *Ann. Fac. Sci. Toulouse Math.* (6) 5 (2) (1996), 233–297. [741](#)
- [3] F. Bonahon, Bouts des variétés hyperboliques de dimension 3. *Ann. of Math.* (2) 124 (1) (1986), 71–158. [745](#)
- [4] M. Dehn, *Papers on group theory and topology*. J. Stillwell, ed., Springer-Verlag, New York 1987. [737](#), [748](#), [750](#), [751](#), [754](#)
- [5] R. Fricke, R. Klein, *Vorlesungen über die Theorie der automorphen Functionen*. Vol. 2, Teubner, Leipzig 1897. [738](#), [749](#), [750](#), [756](#), [759](#)
- [6] A. Fathi, F. Laudenbach and V. Poénaru (eds.), *Travaux de Thurston sur les surfaces*. *Astérisque* 66–67 (1979). [744](#), [745](#), [746](#)
- [7] S. Gervais, Presentation and central extensions of mapping class groups. *Trans. Amer. Math. Soc.* 348 (8) (1996), 3097–3132. [735](#), [746](#), [759](#)
- [8] W. Goldman, Topological components of spaces of representations. *Invent. Math.* 93 (3) (1988), 557–607. [759](#)
- [9] W. Goldman, Trace coordinates on Fricke spaces of some simple hyperbolic surfaces. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 611–684. [747](#), [759](#)
- [10] A. Grothendieck, Esquisse d'un programme. In *Geometric Galois Actions*, 1. Around Grothendieck's "Esquisse d'un programme" (P. Lochak, L. Schneps, eds.), London Math. Soc. Lecture Note Ser. 242, Cambridge University Press, Cambridge 1997, 5–48. [734](#), [735](#), [737](#), [750](#)
- [11] U. Hamenstädt, Parametrizations of Teichmüller space and its Thurston boundary. In *Geometric analysis and nonlinear partial differential equations* (S. Hildebrandt and H. Karcher, eds.), Springer-Verlag, Berlin 2003, 81–88. [746](#)
- [12] J. Harer, Stability of the homology of the mapping class groups of orientable surfaces. *Ann. of Math.* (2) 121 (2) (1985), 215–249. [763](#)
- [13] A. Hatcher, Pants decomposition of surfaces. Preprint, 1999. [763](#)
- [14] A. Hatcher, P. Lochak, L. Schneps, On the Teichmüller tower of mapping class groups. *J. Reine Angew. Math.* 521 (2000), 1–24. [763](#)
- [15] H. Helling, Diskrete Untergruppen von $SL_2(R)$. *Invent. Math.* 17 (1972), 217–229. [757](#)
- [16] A. Hatcher, W. Thurston, A presentation for the mapping class group of a closed orientable surface. *Topology* 19 (1980), 221–237. [746](#), [748](#), [760](#), [762](#), [763](#)

- [17] N. Ivanov, Automorphisms of Teichmüller modular groups. In *Topology and geometry* (O. Ya. Viro, ed.), Lecture Notes in Math. 1346, Springer-Verlag, Berlin 1988, 190–270. [754](#)
- [18] D. Johnson, Homeomorphisms of a surface which act trivially on homology. *Proc. Amer. Math. Soc.* 75 (1) (1979), 119–125. [750](#)
- [19] L. Keen, private communication. [744](#), [748](#)
- [20] L. Keen, Intrinsic moduli on Riemann surfaces. *Ann. of Math.* 84 (1966), 405–420. [750](#), [759](#)
- [21] R. A. Lickorish, A representation of oriented combinatorial 3-manifolds. *Ann. of Math.* 72 (1962), 531–540. [751](#), [753](#)
- [22] P. Lochak, L. Schneps, eds., *Geometric Galois Actions*. 1. Around Grothendieck’s “Esquisse d’un programme”, London Math. Soc. Lecture Note Ser. 242, Cambridge University Press, Cambridge 1997. [735](#)
- [23] F. Luo, On Heegaard diagrams. *Math. Res. Lett.* 4 (1997), 365–373.
- [24] F. Luo, A presentation of the mapping class group. *Math. Res. Lett.* 4 (1997), 735–739. [735](#), [745](#), [746](#)
- [25] F. Luo, Geodesic length functions and Teichmüller spaces. *J. Differential Geom.* 48 (1998), 275–317. [735](#), [736](#), [745](#), [749](#), [750](#), [753](#), [759](#)
- [26] F. Luo, Simple loops on surfaces and their intersection numbers, *Math. Res. Lett.* 5 (1998), 47–56. [735](#), [745](#), [750](#), [754](#), [759](#), [761](#)
- [27] F. Luo, Characters of $SL(2)$ representations of groups. *J. Differential Geom.* 53 (3) (1999), 575–626. [735](#), [756](#), [757](#), [759](#)
- [28] F. Luo, Automorphisms of the complex of curves. *Topology* 39 (2) (2000), 283–298. [754](#), [756](#), [760](#)
- [29] B. Maskit, A picture of moduli space. *Invent. Math.* 126 (2) (1996), 341–390. [744](#)
- [30] C. Moore, N. Seiberg, Polynomial equations for rational conformal field theories. *Phys. Lett. B.* 212 (1988), 451–460. [735](#)
- [31] L. Mosher, Tiling the projective foliation space of a punctured surface. *Trans. Amer. Math. Soc.* 306 (1988), 1–70. [742](#)
- [32] Y. Okumura, On the global real analytic coordinates for Teichmüller spaces. *J. Math. Soc. Jap.* 42 (1990), 91–101. [746](#)
- [33] Y. Okumura, Global real analytic length parameters for Teichmüller spaces. *Hiroshima Math. J.* 26 (1) (1996), 165–179 [746](#)
- [34] R. Penner, The decorated Teichmüller space of punctured surfaces. *Comm. Math. Phys.* 113 (2) (1987), 299–339. [742](#)
- [35] M. Rees, An alternative approach to the ergodic theory of measured foliations on surfaces. *Ergodic Theory Dynam. Systems* 1 (4) (1981), 461–488. [745](#)
- [36] P. Schmutz, Die Parametrisierung des Teichmüllerraumes durch geodätische Längenfunktionen. *Comment. Math. Helv.* 68 (1993), 278–288. [746](#)
- [37] L. Schneps, ed., *The Grothendieck Theory of Dessins d’Enfants*. London Math. Soc. Lecture Note Ser. 200, Cambridge University Press, Cambridge 1994. [735](#)

- [38] C. Series, The modular surface and continued fractions. *J. London Math. Soc.* 31 (2) (1985), 69–80. [748](#)
- [39] W. Thurston, *The Geometry and Topology of Three-Manifolds*. Princeton University Lecture Notes, 1979. [741](#), [745](#), [746](#)
- [40] W. Thurston, Minimal stretch maps between hyperbolic surfaces. Preprint, 1986. [741](#)
- [41] W. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc.* 19 (2) (1988), 417–438. [744](#), [748](#), [754](#)
- [42] H. Vogt, Sur les invariants fondamentaux des équations différentielles linéaires du second ordre. *Ann. Sci. École Norm. Sup.* (3) 6 Suppl. (1889), 3–72. [749](#), [750](#), [756](#), [759](#)
- [43] S. Wolpert, Fenchel-Nielsen deformation. *Ann. of Math.* (2) 115 (3) (1982), 501–528. [738](#)
- [44] S. Wolpert, private communication. [744](#)
- [45] S. A. Wolpert, The Weil–Petersson metric geometry. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 47–64.

Chapter 18

Dessins d'enfants and origami curves

Frank Herrlich and Gabriela Schmithüsen

Contents

1	Introduction	767
2	From Riemann surfaces to algebraic curves	769
3	Dessins d'enfants	771
4	The Galois action on dessins d'enfants	778
	4.1 The action on dessins	778
	4.2 Fields of definition and moduli fields	780
	4.3 Faithfulness	781
	4.4 Galois invariants	782
	4.5 The action on \hat{F}_2	783
	4.6 The action on the algebraic fundamental group	785
	4.7 The Grothendieck–Teichmüller group	787
5	Origamis	788
	5.1 Introduction to origamis	788
	5.2 Teichmüller curves	790
6	Galois action on origamis	792
7	A dessin d'enfants on the origami curve	793
8	Dessins d'enfants related to boundary points of origami curves	796
	8.1 Cusps of origami curves	796
	8.2 The dessin d'enfants associated to a boundary point	798
	8.3 Examples	800
	8.3.1 The origami $L_{2,2}$	800
	8.3.2 General L -shaped origamis	802
	8.3.3 The quaternion origami	804
	8.3.4 The characteristic origami of order 108	805
	References	807

1 Introduction

In this chapter, we give an introduction to the theory of dessins d'enfants. They provide a charming concrete access to a special topic of arithmetic geometry: Curves defined over number fields can be described by such simple combinatorial objects as graphs embedded into topological surfaces. Dessins d'enfants are in some sense an answer of Grothendieck to the beautiful Theorem of Belyi, which characterises curves

defined over number fields by the existence of certain coverings of the projective line. Grothendieck was fascinated by the fact that such a covering is completely determined by the preimage of the real interval $[0, 1]$ and called this a dessin d'enfants. As one consequence that especially attracted people one has an action of the absolute Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the set of dessins which is faithful. Therefore in principle all the information on $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is hidden in some mysterious way in these combinatorial objects. The study of dessins d'enfants leads to the Grothendieck–Teichmüller group in which $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ injects. It is still an open question whether these two groups are equal or not.

In the next three sections we introduce dessins d'enfants and the Galois action on them. We begin in Section 2 with a review of the correspondence between closed Riemann surfaces and regular complex projective curves. Since the link between these two fields is an essential tool which is used throughout the whole chapter, we provide a sketch of the proof. In Section 3 we give characterizations of dessins d'enfants in terms of Belyi pairs, graphs embedded into surfaces, ribbon graphs, monodromy homomorphisms and subgroups of the free group on two generators and explain how to get from one of these descriptions to the other. Section 4 is devoted to the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on dessins d'enfants. We review some of the known results on faithfulness and Galois invariants and explain how it gives rise to an action on the algebraic fundamental group \hat{F}_2 of the three-punctured sphere. The explicit description of how $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the topological generators leads us to the definition of the Grothendieck–Teichmüller group. We finish the section by indicating how $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ embeds into this group.

In the second part of the chapter we turn to connections between origamis and dessins d'enfants. Similar to the latter, origamis are given by combinatorial data and define arithmetic objects, more precisely curves in moduli space which are defined over $\overline{\mathbb{Q}}$. Following the same approach as for dessins, one can study the action of the absolute Galois group on them. Besides these analogies, origamis and dessins are linked by several explicit constructions.

Section 5 gives an introduction to origamis and explains how they define curves in the moduli space M_g of smooth algebraic curves of genus g . We call them origami curves; they are in fact special examples of Teichmüller curves. In Section 6 we describe the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on them and state some known results. The last two sections present two explicit constructions of dessins d'enfants associated to a given origami. Section 7 interprets the origami curve itself as a dessin. In Section 8 we associate a dessin to every cusp of an origami curve. We illustrate these constructions by several nice examples.

The subject of dessins d'enfants has been treated from different points of view in several survey articles, as e.g. [34], [40] and [19, Chapter 2] to mention only a few. A collection of articles on dessins d'enfants including many explicit examples is contained in [33]. More on origamis can be found e.g. in [20] and [31] and the references therein. Almost all results in this chapter were known previously, with the exception of the examples in the last sections.

Acknowledgments. We would like to thank Pierre Lochak, who has initially drawn our attention to the beautiful topic of origamis and their relations to dessins d'enfants. Furthermore, we thank André Kappes and Florian Nisbach for proofreading. The second author is indebted to the LANDESSTIFTUNG Baden-Württemberg for facilitating the analysis entailed in this chapter.

2 From Riemann surfaces to algebraic curves

One fascinating aspect of the theory of dessins d'enfants is that it touches two different fields of mathematics, namely algebraic geometry and complex geometry. The bridge between these two fields is built on the following observation which was already understood by Riemann himself: The classes of closed Riemann surfaces and of regular complex projective curves can be considered to be the same. More precisely, we have an equivalence between the following three categories (see e.g. [28, Theorem 7.2]) and [12, Corollary 6.12]):

- closed Riemann surfaces with non-constant holomorphic maps;
- function fields over \mathbb{C} of transcendence degree 1 with \mathbb{C} -algebra homomorphisms;
- regular complex projective curves with dominant algebraic morphisms.

Recall that a *function field* over a field k is a finitely generated extension field of k . We give here only a brief outline of the above equivalences and refer for further readings to literature in complex geometry (e.g. [7, §16], [24, IV, 1]) and algebraic geometry (e.g. [12]).

In a first step we describe how to get from the category of closed Riemann surfaces to the category of function fields over \mathbb{C} of transcendence degree 1. Let X be a Riemann surface and $\mathbb{C}(X)$ the field of meromorphic functions from X to \mathbb{C} . Then $\mathbb{C}(X)$ is a function field: The fact that $\mathbb{C}(X)$ has transcendence degree 1 essentially follows from the Riemann–Roch theorem. Recall that the theorem determines for a divisor D on X the dimension of the complex vector space $L(D) = H^0(X, \mathcal{O}_D)$ of meromorphic functions f satisfying $\text{div}(f) \geq -D$. It states in particular that if the divisor D is effective, i.e. $D = \sum_i a_i P_i$ with $a_i \geq 0$, then $\dim(L(D)) \leq 1 + \deg D$.

Suppose now that the degree of $\mathbb{C}(X)$ were greater or equal to 2. Then there would exist two algebraically independent meromorphic functions f and g . Let P_1, \dots, P_k be the poles of f and Q_1, \dots, Q_m be the poles of g , with degrees a_1, \dots, a_k and b_1, \dots, b_m respectively. One picks the divisor $D = \sum_i a_i P_i + \sum_j b_j Q_j$. By the definition of D we have, for $i + j \leq n$, $f^i g^j \in L(nD)$. Since f and g are algebraically independent, we have that all the $f^i g^j$ are linearly independent. Therefore $\dim(L(nD)) \geq (n^2 + 3n + 2)/2$. On the other hand, one obtains from the Riemann–Roch theorem that $\dim(L(nD)) \leq 1 + \deg(nD) = 1 + n \deg(D)$. These two inequalities give a contradiction for n large enough. Hence the transcendence degree of $\mathbb{C}(X)$ is ≤ 1 . Equality follows from the fact that each compact Riemann surface

admits a non-constant meromorphic function. The Riemann–Roch theorem precisely guarantees the existence of meromorphic functions. E.g. if we fix a divisor of degree greater or equal to $g + 1$, then $\dim(L(D)) \geq 2$, therefore we have a non-constant meromorphic function in $L(D)$. Altogether we have seen that each closed Riemann surface defines a function field of transcendence degree 1. Furthermore a non-constant holomorphic function defines a morphism of \mathbb{C} -algebras between the function fields by pulling back the rational functions. We have thus constructed a contravariant functor from the category of closed Riemann surfaces to the category of function fields of transcendence degree 1.

The equivalence between function fields and regular projective complex curves is described e.g. in [12, Chapter I]. In fact the statement holds a bit more generally. One may replace the field \mathbb{C} by any algebraically closed field k . Similarly as before one obtains a function field $k(C)$ of degree 1 starting from an algebraic curve C over k . In this case $k(C)$ is the field of all rational functions from C to $\mathbb{P}^1(k)$, the projective line over k . Two algebraic varieties are birationally equivalent if and only if they have the same function field [12, I Corollary 4.5] and nonsingular curves are birationally equivalent if and only if they are isomorphic [12, Proposition 6.8]. Hence it remains to show that for each function field K of degree 1 over k one can construct a projective regular curve C whose function field is K . This construction is described in [12, I, §6]. It is based on the following observation: Each point p on an algebraic curve C defines a discrete valuation ring whose quotient field is $K = k(C)$, namely the local ring \mathcal{O}_p of germs of regular functions on C near p . The main idea is to identify the points of the curve with the valuation rings which they induce, in order to reconstruct the projective curve C from its function field $k(C)$.

Hence, given a function field K , we take the set C_K of discrete valuation rings of K . We want to think of its elements R as points of the algebraic curve that we are going to construct. First, C_K becomes a topological space by taking the finite sets and the whole space to be the closed sets. Next, we can cover C_K by affine regular curves as follows. Suppose that R is a point of C_K , i.e. R is a discrete valuation ring in K . Hence K is the quotient field of R and R is a local ring of dimension 1. Let m_R be its maximal ideal. We want to define an affine curve Y together with an embedding of Y into C_K , such that the image contains the point R . We pick an arbitrary $y \in R \setminus k$ and define B to be the integral closure of $k[y]$ in K . It follows from commutative algebra that B is contained in R , it is a Dedekind domain and a finitely generated k -algebra. Thus B is in particular the affine coordinate ring of an affine regular curve Y . Finally, we want to construct an injective continuous map from Y to C_K . Recall from algebraic geometry that the points in Y correspond to the maximal ideals of B . Let Q be in Y and n_Q the corresponding maximal ideal of B . Then B_{n_Q} is a local ring in K , and indeed a discrete valuation ring. Hence we may map $Q \in Y$ to $B_{n_Q} \in C_K$. This gives a continuous map from Y to C_K . Let m_Q be the unique maximal ideal in B_{n_Q} . Then $n_Q = m_Q \cap B$. Hence the map is injective. Furthermore R is in the image, since $m_R \cap B$ is a maximal ideal in B .

One then shows that C_K with this structure is a regular projective curve.

In order to close the circle between the three categories, one constructs for each regular complex projective curve C a closed Riemann surface X with the same function field: Suppose that C lies in $\mathbb{P}^n(\mathbb{C})$. $\mathbb{P}^n(\mathbb{C})$ becomes a complex manifold using the natural cover by $\mathbb{A}^n(\mathbb{C})$'s for charts to \mathbb{C}^n . (Be aware that $\mathbb{A}^n(\mathbb{C})$ and \mathbb{C}^n do not have the same topology. Therefore one speaks of the *Zariski topology* and the *complex topology* of $\mathbb{P}^n(\mathbb{C})$ and later also of the curve C). C is the zero set of finitely many homogeneous polynomials f_1, \dots, f_m . Since C is regular, the Jacobian matrices $(\frac{\partial f_i}{\partial x_j}(p))_{i,j}$ are invertible for all points p on C . The implicit functions theorem, together with the fact that the complex dimension of C is 1, provides us locally with a function from \mathbb{C} to \mathbb{C} , which is invertible. Its inverse map is a chart for C . C becomes a closed Riemann surface X with these chart maps. Finally, one shows that the function fields are the same by checking that rational functions on C become meromorphic functions on the Riemann surface X and vice versa.

We will use this equivalence between the category of closed Riemann surfaces and the category of regular complex projective curves throughout the whole chapter. Observe in particular that the Riemann sphere corresponds to the projective line $\mathbb{P}^1(\mathbb{C})$ under this identification.

3 Dessins d'enfants

In this section we give a brief introduction to dessins d'enfants. They are a nice way to describe coverings $\beta: X \rightarrow \mathbb{P}^1(\mathbb{C})$ from a closed Riemann surface X to the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ which are ramified at most over the three points $0, 1$ and ∞ . Such coverings are called *Belyi morphisms*. One reason why they are particularly interesting is the famous Theorem of Belyi. This theorem establishes a connection between complex Riemann surfaces X , which allow a Belyi morphism, and projective algebraic curves C which are defined over the algebraic closure $\overline{\mathbb{Q}}$ of \mathbb{Q} . As described in Section 2 we identify the closed Riemann surface X with the corresponding projective regular curve C defined over \mathbb{C} . C is defined over $\overline{\mathbb{Q}}$ if it can be described as the zero set of polynomials whose coefficients lie in $\overline{\mathbb{Q}}$. Observe that in this case the curve C actually is defined over a number field, since a curve can be defined by finitely many polynomials and therefore there exists a finite field extension of \mathbb{Q} which contains all coefficients. Therefore Belyi morphisms provide a tool for studying complex curves over number fields.

Theorem 3.1 (Theorem of Belyi, [3]). *Let X be a regular complex projective curve. Then X is defined over $\overline{\mathbb{Q}}$ if and only if there exists a finite morphism $\beta: X \rightarrow \mathbb{P}^1(\mathbb{C})$ from X to the projective line $\mathbb{P}^1(\mathbb{C})$ which is ramified at most over $0, 1$ and ∞ .*

It follows from the proof of the theorem that if the condition of the theorem holds, we can choose the morphism β such that it is defined over $\overline{\mathbb{Q}}$. Therefore in the

following, if we call β a *Belyi morphism*, we will always assume that it is defined over $\overline{\mathbb{Q}}$.

The surprising part of Belyi's result was the only-if direction. Belyi gave an elementary but tricky algorithm for how to calculate the morphism β . For the if part of the proof, Belyi referred to a very general result of A. Weil. Later on, more direct proofs were given by B. Köck in [18] in the language of algebraic geometry and by J. Wolfart in [39] using uniformisation theory. We shall sketch the main idea of the proof in Section 4.2.

The theorem makes it particularly desirable to describe Belyi morphisms β as simply as possible. Fortunately, this can be done using "objects so simple that a child learns them while playing" (Grothendieck in [9]). In the following we present several methods on how to describe β and give an idea of the proofs of why they are all equivalent and how one can retrieve β from them. A nice and broad overview on these ideas as well as more references can be found e.g. in [33].

Let (X, β) be a *Belyi pair*, i.e. a closed Riemann surface X together with a Belyi morphism $\beta: X \rightarrow \mathbb{P}^1(\mathbb{C})$. We say that two Belyi pairs (X_1, β_1) and (X_2, β_2) are *equivalent* if there exists an isomorphism $f: X_1 \rightarrow X_2$ such that $\beta_2 \circ f = \beta_1$. We consider Belyi pairs up to this equivalence relation.

Proposition 3.2. *A Belyi pair (X, β) is up to equivalence uniquely determined by:*

- *a dessin d'enfants (defined below) up to equivalence;*
- *a bipartite connected ribbon graph up to equivalence;*
- *a monodromy map $\alpha: F_2 \rightarrow S_d$ for some $d \geq 1$, i.e. a transitive action of F_2 on $\{1, \dots, d\}$, up to conjugation in S_d ;*
- *a finite index subgroup of F_2 up to conjugation.*

Here F_2 denotes the free group on two generators and S_d the symmetric group on d objects. The first part of Proposition 3.2, namely the equivalence between Belyi pairs and dessins, is often called the *Grothendieck correspondence*. In the following we will sketch the proof of the proposition by explaining how to pass from one description to the next.

From a Belyi pair to a dessin. One starts from the observation that a Belyi pair (X, β) naturally defines a bipartite graph G on the surface X : Let I be the closed segment on the real line \mathbb{R} between 0 and 1. Then its preimage $\beta^{-1}(I)$ is a graph on X . Its vertices are the preimages of the two points 0 and 1. It carries a natural bipartite structure: we may colour all preimages of 0 with one colour (e.g. black) and all preimages of 1 with another colour (e.g. white). It is a striking fact which we will see in the rest of this section that the graph embedded into the topological surface already carries enough information. It uniquely determines the Belyi pair (X, β) up to equivalence and thus in particular the complex structure on X .

Furthermore, one observes that $X - G$ decomposes into components each of them containing precisely one preimage of ∞ . The holomorphic map β restricted to one of the components is therefore ramified at most in one point and hence at this point is locally of the form $z \mapsto z^n$. Its image is an open cell. Therefore the component itself is an open cell and thus holomorphically equivalent to the open unit disk. Altogether, the graph G decomposes the surface into open cells containing precisely one preimage of ∞ .

Example 3.3. In Figure 1 we show the dessin on the elliptic curve

$$C : y^2 = x(x - 1)(x - \lambda_0)$$

with $\lambda_0 = 1/2 + (\sqrt{3}/2)i$. The curve C has an automorphism of order 3. The Belyi morphism β is the quotient map with respect to this automorphism.

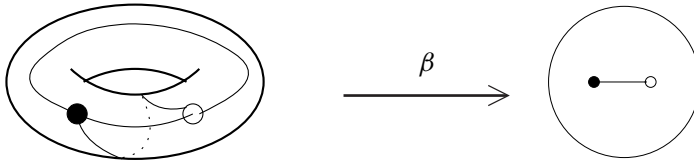


Figure 1. A Belyi morphism and its dessin d'enfants.

Definition 3.4. A *dessin d'enfants* is a bipartite connected graph G which is embedded into an orientable closed topological surface X , such that it fills the surface, i.e. $X \setminus G$ is a union of open cells. Two dessins d'enfants (X_1, G_1) and (X_2, G_2) are called *equivalent* if there exists a homeomorphism $f : X_1 \rightarrow X_2$ such that $f(G_1) = G_2$.

Dessins and bipartite ribbon graphs. Ribbon graphs are a handy way to describe dessins. Let D be a dessin, i.e. $D = (G, i)$, where G is a connected graph and $i : G \hookrightarrow X$ is a continuous embedding of G into a closed topological surface X . We start from the observation that the abstract graph G does not uniquely determine the dessin. One can e.g. embed the same graph into surfaces of different genera, see Example 3.6. How much information do we have to add to the graph in order to nail down the dessin? It turns out that it suffices to assign to each vertex a cyclic permutation of the edges which are adjacent to the vertex. To simplify notations, we divide each edge into two half edges and number them with $1, \dots, 2d$, where d is the number of edges of the graph. For each vertex v of G we take a chart (U, φ) of a small neighbourhood U of v in X to the plane \mathbb{R}^2 such that the image of $G \cap U$ is a star with the vertex $\varphi(v)$ as centre. Imagine we circle anticlockwise around the vertex in $\varphi(U)$. Let π_v be the cyclic permutation which denotes the order in which we meet the images of the half edges adjacent to v . Hence π_v is in the symmetric group S_{2d} . For the dessin in Figure 1 we obtain e.g. the cyclic permutations $(1\ 3\ 5)$ and $(2\ 4\ 6)$, if we label the half edges as in Figure 2.

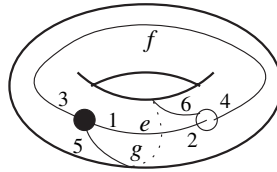


Figure 2. Labelling the half edges of the dessin in Figure 1.

Definition 3.5. A ribbon graph (G, O) – often also called *fat graph* – is a connected graph G together with a ribbon structure $O = \{\pi_v \mid v \text{ a vertex of } G\}$, which assigns to each vertex v of G a cyclic permutation π_v of the half edges adjacent to v . Two ribbon graphs (G_1, O_1) and (G_2, O_2) are called *equivalent* if there exists an isomorphism $h : G_1 \rightarrow G_2$ of graphs such that the pull back of O_2 is equal to O_1 .

Let π be the product of all the π_v 's and τ the transposition which maps each half edge to the other half edge that belongs to the same edge. Then the tuple (π, τ) determines the ribbon graph.

Recall that by the definition of dessins the graph G fills the surface X , i.e. $X \setminus G$ consists of disjoint open cells C_1, \dots, C_s . Observe that we obtain the edges of the cycle bounding a cell C clockwise successively by taking the edges on which the half edges $e, \tau\pi(e), (\tau\pi)^2(e), \dots$ lie. Here e is the half edge at the beginning of an edge in the cycle, where the cycle carries the natural anti-clockwise orientation. For example, for the dessin in Figure 2 we obtain one cell which is bounded by the cycle $(e f g e f g)$.

One gets the dessin back from the ribbon graph doing the reverse procedure: Each cycle $(e, \tau\pi(e), \dots, (\tau\pi)^k(e))$ defines a cycle in the graph which is the union of the corresponding edges. One glues a cell to each such cycle. Then each edge is on the boundary of precisely two cells (which may coincide) and one obtains a closed surface X in which G is embedded.

Example 3.6. In Figure 3 we show two ribbon graphs (G_1, O_1) and (G_2, O_2) .

Observe that the two ribbon graphs have the same underlying graph, but the ribbon structures are different and they define different surfaces. The second one is the ribbon graph from Figure 2.

For both ribbon graphs we have $\tau = (1\ 2)(3\ 4)(5\ 6)$. Hence for the first graph we obtain $\tau\pi = (1\ 6)(2\ 3)(4\ 5)$ and for the second graph we have $\tau\pi = (5\ 4\ 1\ 6\ 3\ 2)$. Thus, in the first case we obtain three cells: The first one is bounded by e_3 and e_1 , the second one is bounded by e_3 and e_2 and the third one is bounded by e_2 and e_1 . Gluing the disks along their edges gives a genus 0 surface. This can be checked with a short Euler characteristic calculation. As we already saw above, we obtain in the second case one cell bounded clockwise by the edges e_1, e_2, e_3, e_1, e_2 and e_3 , and a surface of genus 1.

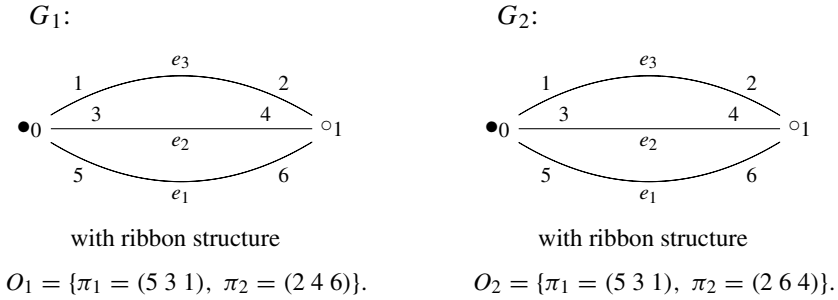


Figure 3. Two ribbon graphs.

Hence we may equivalently talk about dessins or about bipartite ribbon graphs. One can check that the respective equivalence relations match each other.

Remark 3.7. The constructions above define a bijection between the set of equivalence classes of dessins and the set of equivalence classes of bipartite connected ribbon graphs. Furthermore we described a natural way to assign to each Belyi pair an equivalence class of dessins.

In order to see how we can retrieve the Belyi pair from a given dessin or a given ribbon graph, it is convenient to introduce monodromy maps.

Monodromy maps and subgroups of F_2 . Recall that for an unramified degree d covering $p: X^* \rightarrow Y^*$ of surfaces we obtain the *monodromy map* $\alpha: \pi_1(Y^*) \rightarrow S_d$ to the symmetric group S_d on d letters as follows: Fix a point $y \in Y^*$. Call its d preimages x_1, \dots, x_d . For $[c] \in \pi_1(Y^*, y)$ map $i \in \{1, \dots, d\}$ to j , if x_j is the end point of the lift of c to X , which starts in x_i . The resulting map α is independent of the chosen point y and of the choice of the labeling of its preimage up to composition with a conjugation in S_d .

Let us now consider the natural embedding $\pi_1(X^*) \hookrightarrow \pi_1(Y^*)$ induced by p and let U be its image. U depends on the chosen base points of the fundamental groups only up to conjugation. Hence we may assume that the base point of $\pi_1(X^*)$ is the preimage of the base point of $\pi_1(Y^*)$ labeled by 1. Then the image of $\pi_1(X^*) \hookrightarrow \pi_1(Y^*) \xrightarrow{\alpha} S_d$ is the stabilizer $\text{Stab}_{S_d}(1)$ of 1 in S_d and U is its full preimage in $\pi_1(Y^*)$. Hence one obtains U directly from α , namely $U = \alpha^{-1}(\text{Stab}_{S_d}(1))$. Conversely given U one obtains α as follows: $\pi_1(Y^*)$ acts on the d cosets Ug_i of U in $\pi_1(Y^*)$ by multiplication from the right; α is the induced action from the right on the indices.

Starting now from a Belyi pair (X, β) , we obtain an unramified cover by removing the three ramification points $0, 1$ and ∞ from $\mathbb{P}^1(\mathbb{C})$ and all their preimages from X . We denote the resulting punctured surfaces by \mathbb{P} and X^* , respectively. We fix an isomorphism between $\pi_1(\mathbb{P})$ and F_2 , the free group in two generators. Then $p: X^* \rightarrow$

$\tilde{\mathbb{P}}$ is an unramified covering and defines a monodromy map from $F_2 \cong \pi_1(\tilde{\mathbb{P}})$ to S_d (where d is the degree of p) or equivalently a finite index subgroup U of $F_2 \cong \pi_1(\tilde{\mathbb{P}})$.

Finally we describe how to retrieve the Belyi pair from the subgroup U . The main ingredient that we use is the universal covering theorem. Let us choose a universal covering $u: \tilde{H} \rightarrow \tilde{\mathbb{P}}$. By the theorem we may identify $F_2 \cong \pi_1(\tilde{\mathbb{P}})$ with the group of deck transformations $\text{Deck}(\tilde{H}/\tilde{\mathbb{P}})$. By the same theorem each finite index subgroup of $\pi_1(\tilde{\mathbb{P}})$ defines an unramified covering β from some surface X^* to $\tilde{\mathbb{P}}$ such that it induces an embedding $\text{Deck}(\tilde{H}/X^*) \hookrightarrow \text{Deck}(\tilde{H}/\tilde{\mathbb{P}})$ whose image is the subgroup U . There is a unique complex structure on X^* which makes β holomorphic, namely the lift of the complex structure on $\tilde{\mathbb{P}}$ via β . It follows from the classical theory of Riemann surfaces that there is a unique closed Riemann surface X which is the closure of X^* . It is obtained by filling in one point for each puncture. Furthermore β can be extended in a unique way to $\beta: X \rightarrow \mathbb{P}^1(\mathbb{C})$.

One can check that all this is independent of the choices that we did in between up to the equivalence relations, that the equivalence relations fit together and that the constructions are inverse to each other.

Remark 3.8. The above constructions define bijections between the set of equivalence classes of Belyi pairs, the set of conjugacy classes of group homomorphisms $F_2 \rightarrow S_d$ which are transitive actions, and the set of conjugacy classes of finite index subgroups of F_2 .

As a last step, we have to show how we can relate dessins and ribbon graphs to monodromy maps or finite index subgroups of F_2 .

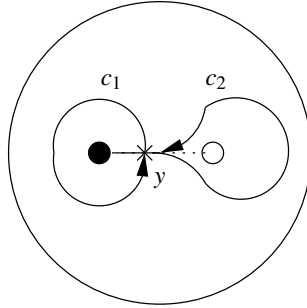
From a dessin D to a Belyi pair (X, β) . Let D be a dessin and $(G, O = \{\pi_1, \dots, \pi_s\})$ the corresponding ribbon graph from Remark 3.7. How can we retrieve the monodromy of β from these data? Recall that G is bipartite and the vertices are coloured: the preimages of 0 are black and those of 1 are white. We may also colour the half edges used in the construction of Remark 3.7 with the colour of the vertex which lies on them. Observe that π acts on the set E_{black} of black half edges and the set E_{white} of white half edges separately. Thus we can decompose $\pi = \pi_{\text{black}} \circ \pi_{\text{white}}$ with $\pi_{\text{black}} \in \text{Perm}(E_{\text{black}})$ and $\pi_{\text{white}} \in \text{Perm}(E_{\text{white}})$.

Let us now choose a base point $y \in \tilde{\mathbb{P}}$ on the segment between 0 and 1 close to 0. Hence all its preimages x_i lie on black half edges. Furthermore we pick two curves c_1 and c_2 as generators of $\pi_1(\tilde{\mathbb{P}}) \cong F_2$, where c_1 is a simple closed circle around 0 and c_2 is a simple closed circle around 1; both starting in y and both anti-clockwise. (See Figure 4.) By the definition of π (see Remark 3.7), the monodromy $\alpha(c_1)$ is the permutation π_{black} and the monodromy $\alpha(c_2)$ is the permutation $\tau \pi_{\text{white}} \tau$. Here we identify the point x_i with the black half edge on which it lies.

Hence, we may assign to a dessin the monodromy map

$$F_2 \rightarrow S_d, \quad x \mapsto \pi_{\text{black}}, \quad y \mapsto \tau \pi_{\text{white}} \tau.$$

Again one can check that this construction is inverse to the construction given in Remark 3.7 and the equivalence relations fit together.

Figure 4. Generators of the fundamental group of \mathbb{P}^2 .

Furthermore it follows from the above construction that given the monodromy map $\alpha: F_2 \rightarrow S_d$ of a Belyi pair (X, β) , one obtains the corresponding bipartite ribbon graph (G, O) directly as follows: Label the black half edges with $1, \dots, d$ and the white half edges with $d+1, \dots, 2d$. Then (G, O) is described by the two permutations

$$\tau: i \mapsto d+i \quad \text{and} \quad \pi = \pi_{\text{black}} \circ \pi_{\text{white}}$$

with

$$\pi_{\text{black}}: i \mapsto \alpha(x)(i) \quad \text{and} \quad \pi_{\text{white}}: d+i \mapsto d+\alpha(y)(i).$$

Remark 3.9. The above construction defines a bijection between the set of equivalence classes of dessins and the set of equivalence classes of Belyi pairs. This map is the inverse map to the one described before Remark 3.7.

With Remark 3.9 we have finished the outline of the proof of Proposition 3.2.

It follows in particular that we can describe a Belyi pair (X, β) or equivalently the corresponding dessin D by a pair of permutations (σ_1, σ_2) , namely $\sigma_1 = \alpha(c_1)$ and $\sigma_2 = \alpha(c_2)$, where $\alpha: F_2 \rightarrow S_d$ is the monodromy map. We will say the dessin *has monodromy* (σ_1, σ_2) . This description is unique up to simultaneous conjugation with an element in S_d . Furthermore the group generated by σ_1 and σ_2 acts transitively on $\{1, \dots, d\}$ and each pair of permutations with this property defines a Belyi pair.

The genus of a dessin. Suppose that a dessin (X, β) of degree d has monodromy (σ_1, σ_2) . The dessin naturally defines a two-dimensional cell complex. By the construction in Remark 3.9 we have:

- The black vertices are in one-to-one correspondence with the cycles in σ_1 . Denote their number by s_1 .
- The white vertices are in one-to-one correspondence with the cycles in σ_2 . Denote their number by s_2 .
- The faces of the complex are in one-to-one correspondence with the cycles in $\sigma_1 \circ \sigma_2$. Denote their number by f .

Hence, we can calculate the genus as follows:

$$g = \frac{2 - \chi}{2} \quad \text{with } \chi = s_1 + s_2 - d + f.$$

Definition 3.10. We call g as above the *genus of the dessin* D .

4 The Galois action on dessins d'enfants

One of the original motivations to study dessins d'enfants was the hope to get new insights into the structure of the “absolute” Galois group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ of the algebraic closure $\overline{\mathbb{Q}}$ of the rational number field \mathbb{Q} . This hope came from the fact that, as a consequence of the Grothendieck correspondence between dessins d'enfants and Belyi pairs explained in the previous section, $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the set of dessins d'enfants. We shall see that this action is faithful, so in principle, all information about $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is somehow contained in the dessins d'enfants. Unfortunately, except for very special cases, it is so far not known how to describe the action of a Galois automorphism on a dessin in terms of the combinatorial data that determine the dessin. Nevertheless this approach led to many beautiful results concerning e. g. the faithfulness of the action on special classes of dessins d'enfants. Perhaps the most conceptual outcome of the investigation of the Galois action on dessins is the embedding of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ into the Grothendieck–Teichmüller group \widehat{GT} .

4.1 The action on dessins

In this section we explain the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on dessins d'enfants by saying how it acts on Belyi pairs.

By the theorem of Belyi, every Riemann surface X that admits a Belyi morphism β is defined over a number field and thus in particular over $\overline{\mathbb{Q}}$. As explained in the first paragraph of Section 3, this means that, as an algebraic curve, X can be described as the zero set of polynomials with coefficients in $\overline{\mathbb{Q}}$. The fancier language of modern algebraic geometry expresses this property by saying that X admits a morphism of finite type $\varphi: X \rightarrow \text{Spec}(\overline{\mathbb{Q}})$ to the one point scheme $\text{Spec}(\overline{\mathbb{Q}})$. Such a φ is called a *structure morphism* of X .

Every Galois automorphism $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ induces an automorphism σ^* of $\text{Spec}(\overline{\mathbb{Q}})$. Composing it with the structure morphism φ gives a new structure morphism ${}^\sigma\varphi := (\sigma^{-1})^* \circ \varphi: X \rightarrow \text{Spec}(\overline{\mathbb{Q}})$. We call ${}^\sigma X$ the scheme X endowed with the structure morphism ${}^\sigma\varphi$. In more elementary language, ${}^\sigma X$ is obtained from X by applying σ to the polynomials defining X . In general, X and ${}^\sigma X$ are not isomorphic as $\overline{\mathbb{Q}}$ -schemes or as Riemann surfaces, i.e. there is in general no isomorphism making the following diagram commutative:

$$\begin{array}{ccc}
 X & \overset{?}{\dashrightarrow} & \sigma X \\
 \searrow \varphi & & \swarrow \sigma\varphi \\
 & \text{Spec}(\overline{\mathbb{Q}}) &
 \end{array}$$

Example 4.1. Let E be an elliptic curve over \mathbb{C} , in other words a Riemann surface of genus 1. E can be embedded into the projective plane as the zero set of a Weierstrass equation $y^2 = x^3 + ax + b$ (or rather its homogenisation). It is defined over a number field if and only if $a, b \in \overline{\mathbb{Q}}$. A Belyi map for E is obtained e. g. by applying Belyi's algorithm to the four critical values of the projection $\beta_0: E \rightarrow \mathbb{P}^1, (x, y) \mapsto x$. The elliptic curve ${}^\sigma E$ is the zero set of $y^2 = x^3 + \sigma(a)x + \sigma(b)$. It is well known that Weierstrass equations define isomorphic Riemann surfaces if and only if their j -invariants agree. Thus ${}^\sigma E$ is isomorphic to E if and only if $j(E) = j(a, b) = \frac{a^3}{4a^3 + 27b^2}$ is fixed by σ .

To describe the Belyi map $\sigma\beta: \sigma X \rightarrow \mathbb{P}^1$ that gives the image of the Belyi pair (X, β) under σ , we first look at the characterization of X as the zero set of polynomials f_1, \dots, f_k in variables x_1, \dots, x_n : then β is, at least locally, also given as a polynomial in x_1, \dots, x_n with coefficients in $\overline{\mathbb{Q}}$, and $\sigma\beta$ is obtained by applying σ to the coefficients of this polynomial.

The description of $\sigma\beta$ in terms of schemes is as follows: Let $\pi: \mathbb{P}^1 \rightarrow \text{Spec}(\overline{\mathbb{Q}})$ denote the (fixed) structure morphism of the projective line \mathbb{P}^1 ; π is related to the structure morphism φ of X by the equation

$$\varphi = \pi \circ \beta.$$

Since \mathbb{P}^1 clearly is defined over \mathbb{Q} , for every $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ the induced automorphism σ^* of $\text{Spec}(\overline{\mathbb{Q}})$ lifts to an automorphism ρ_σ of \mathbb{P}^1 . Then we have

$$\sigma\beta = \rho_{\sigma^{-1}} \circ \beta.$$

This is summarized in the following commutative diagram:

$$\begin{array}{ccccc}
 X & & & & \\
 \searrow \varphi & & \searrow \sigma\varphi & & \\
 & \text{Spec}(\overline{\mathbb{Q}}) & \xrightarrow{(\sigma^{-1})^*} & \text{Spec}(\overline{\mathbb{Q}}) & \\
 \downarrow \beta & \nearrow \pi & & \nearrow \pi & \\
 \mathbb{P}^1 & \xrightarrow{\rho_{\sigma^{-1}}} & \mathbb{P}^1 & & \\
 \nearrow \sigma\beta & & & &
 \end{array}$$

4.2 Fields of definition and moduli fields

Before studying properties of the Galois action on dessins d'enfants, we shortly digress for the following question: Given a Riemann surface X , what is the smallest field over which X can be defined?

In general we say that a variety (or scheme) X/K over a field K can be defined over a subfield $k \subset K$ if there is a scheme X_0/k over k such that X is obtained from X_0 by extension of scalars: $X = X_0 \times_k K$. In this case, we call k a *field of definition* for X .

For example, a Riemann surface can always be defined over a field K which is finitely generated over \mathbb{Q} . Namely, considered as an algebraic curve, X is the zero set of finitely many polynomials, and we may take K to be the extension field of \mathbb{Q} which is generated by the finitely many coefficients of these polynomials.

It is not true in general that there is a unique smallest subfield of K over which a given variety X/K can be defined. Therefore we cannot speak of “the field of definition” of X . But there is another subfield of K associated with X , called the *moduli field*, which is uniquely determined by X and turns out to be closely related to fields of definition:

Definition 4.2. Let $\text{Aut}(\mathbb{C})$ be the group of all field automorphisms of \mathbb{C} . For a Riemann surface X denote by $U(X)$ the subgroup of all $\sigma \in \text{Aut}(\mathbb{C})$ for which ${}^\sigma X$ is isomorphic to X . The fixed field $M(X) \subset \mathbb{C}$ of $U(X)$ is called the *moduli field* of X .

There are two rather straightforward observations about moduli fields:

Remark 4.3. Let X be a Riemann surface of genus g .

a) If $k \subset \mathbb{C}$ is a field of definition for X , then $M(X) \subseteq k$.

b) Assume that X can be defined over $\overline{\mathbb{Q}}$ and let $[X]$ be the corresponding point in the moduli space $M_{g,\overline{\mathbb{Q}}}$ of regular projective curves defined over $\overline{\mathbb{Q}}$ (considered as a variety over $\overline{\mathbb{Q}}$). Recall that $M_{g,\overline{\mathbb{Q}}}$ is obtained from a variety $M_{g,\mathbb{Q}}$ which is defined over \mathbb{Q} by extension of scalars. Then the orbit of $[X]$ under the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $M_{g,\overline{\mathbb{Q}}}$ gives a closed point $[X]_{\mathbb{Q}}$ in the variety $M_{g,\mathbb{Q}}$ whose residue field is isomorphic to $M(X)$.

Proof. a) If X is defined over k and if $\sigma \in \text{Aut}(\mathbb{C})$ fixes k , then id_X is an isomorphism between ${}^\sigma X$ and X . Thus $\{\sigma \in \text{Aut}(\mathbb{C}) \mid \sigma|_k = \text{id}_k\} \subseteq U(X)$, hence $M(X) \subseteq k$.

b) (Sketch) Let $V \subset M_{g,\mathbb{Q}}$ be an affine neighbourhood of $[X]_{\mathbb{Q}}$ and let A be its affine coordinate ring. Then $[X]_{\mathbb{Q}}$ corresponds to a maximal ideal m in A , and $k = A/m$ is its residue field. In $A \otimes k$, m decomposes into maximal ideals m_1, \dots, m_d which are in bijection with the points in the Galois orbit of $[X]$. Thus the fixed field of the stabilizer of, say, m_1 in $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is $A \otimes k/m_1 = k$. \square

The relation between the field of moduli and fields of definition of a Riemann surface is much closer than indicated in part a) of the remark:

Proposition 4.4. Any Riemann surface can be defined over a finite extension of its moduli field.

This result is proved in [39]. For a proof in the language of algebraic geometry that holds for curves over any field, see [10]. Further results on moduli fields, in particular on the moduli field of a Belyi pair, can be found in [18]. There it is shown, among other nice properties, that for “most” curves, the moduli field is also a field of definition. The precise statement is that $X/\text{Aut}(X)$ can be defined over $M(X)$ for any curve X of genus $g \geq 2$. This implies in particular that X can be defined over $M(X)$ if X admits no nontrivial automorphism. In this case, which holds for a generically chosen Riemann surface of genus ≥ 3 , the moduli field is the unique smallest field of definition.

Proposition 4.4 plays a key role in the proof of the “if”-direction of Belyi’s theorem. As explained in Section 3 one has to show that a Riemann surface can be defined over $\overline{\mathbb{Q}}$ if it admits a finite covering $\beta: X \rightarrow \mathbb{P}^1(\mathbb{C})$ which is ramified at most over $0, 1$ and ∞ . Observe that, up to isomorphism, there are only finitely many coverings $Y \rightarrow \mathbb{P}^1(\mathbb{C})$ from some Riemann surface Y of a fixed degree that are unramified outside $0, 1, \infty$ (see [18, Proposition 3.1] for an elementary proof of this fact). It follows that the moduli field of β and hence in particular that of X is a finite extension of \mathbb{Q} . From Proposition 4.4 we then conclude that X can be defined over a number field.

4.3 Faithfulness

We have established an action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on Belyi pairs by defining $\sigma \cdot (X, \beta)$ to be the Belyi pair $(\sigma X, \sigma \beta)$ for $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Example 4.1 shows that this action is faithful, since for every Galois automorphism $\sigma \neq \text{id}$ we can find $a, b \in \overline{\mathbb{Q}}$ such that $\sigma(j(a, b)) \neq j(a, b)$ and thus σE is not isomorphic to E , where E is the elliptic curve with Weierstrass equation $y^2 = x^3 + ax + b$. Translating the Galois action to dessins d’enfants via the Grothendieck correspondence we deduce:

Proposition 4.5. *The action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on dessins d’enfants is faithful.*

Several nice examples for this Galois action on dessins are worked out in the manuscript [40] by J. Wolfart; he attributes the following one to F. Berg: Let $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ be an element that maps the primitive 20th root of unity $\zeta = e^{\pi i/10}$ to ζ^3 . Then σ maps the left hand dessin in Figure 5 to the right hand one:

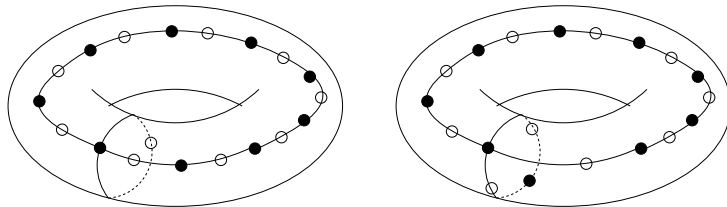


Figure 5. Two Galois equivalent dessins which are not isomorphic.

The dessin on the left lies on the elliptic curve $y^2 = (x + 1)(x - 1)(x - \cos \frac{\pi}{10})$, whereas the right hand dessin lies on $y^2 = (x + 1)(x - 1)(x - \cos \frac{3\pi}{10})$. The Belyi map is in both cases the composition of the projection $\beta_0(x, y) = x$ with the square $T_5^2(z)$ of the fifth Chebyshev polynomial.

In the proof of Proposition 4.5 we have shown more precisely that the action is faithful on dessins of genus 1. The same faithfulness result holds for the Galois action on dessins of any fixed genus $g \geq 1$. This can be seen for example using hyperelliptic curves: for mutually distinct numbers a_1, \dots, a_{2g} in $\mathbb{P}^1(\mathbb{C})$, the (affine) equation $y^2 = (x - a_1) \dots (x - a_{2g})$ defines a nonsingular curve X of genus g . The automorphism $(x, y) \mapsto (x, -y)$ is called the *hyperelliptic involution* on X ; the quotient map is the projection $(x, y) \mapsto x$. It is a covering $X \rightarrow \mathbb{P}^1$ of degree 2, ramified exactly over a_1, \dots, a_{2g} . Two hyperelliptic curves with equations $y^2 = (x - a_1) \dots (x - a_{2g})$ and $y^2 = (x - a'_1) \dots (x - a'_{2g})$ are isomorphic if and only if there is a Möbius transformation that maps the set $\{a_1, \dots, a_{2g}\}$ to the set $\{a'_1, \dots, a'_{2g}\}$.

A hyperelliptic curve is defined over $\overline{\mathbb{Q}}$ if all the a_i are algebraic numbers. In this case, for $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, the curve ${}^\sigma X$ is given by the equation $y^2 = (x - \sigma(a_1)) \dots (x - \sigma(a_{2g}))$. It is then easy, if $\sigma \neq \text{id}$, to choose a_1, \dots, a_{2g} in such a way that there is no Möbius transformation that maps the a_i to the $\sigma(a_j)$. An explicit way to find suitable a_i 's is explained in [1].

With a bit more work, it is also possible to show that the Galois action on genus 0 dessins is faithful. Since all Riemann surfaces of genus zero are isomorphic to the projective line, it is not possible to find, as in the case of higher genus, a Riemann surface X such that ${}^\sigma X \not\cong X$. Rather one has to provide, for a given $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, $\sigma \neq \text{id}$, a rational function $\beta(z)$ such that ${}^\sigma \beta$ is not equivalent to β , i.e. not of the form $\beta \circ \rho$ for some Möbius transformation ρ . L. Schneps [34] showed that one can always find a suitable polynomial. The dessin d'enfants obtained from a polynomial is a planar graph whose complement in the plane is connected, hence the dessin is a tree. Schneps' result thus is

Proposition 4.6. *The Galois action on trees is faithful.*

Using a similar argument as for the hyperelliptic curves, F. Armknecht [1] gave an alternative proof of this result. L. Zapponi [42] improved the result to trees of diameter at most 4.

4.4 Galois invariants

To understand the Galois action on dessins d'enfants one can look for *Galois invariants*, i.e. properties of a dessin that remain unchanged under all Galois automorphisms. The idea, or rather the dream, is to find a complete list of invariants; then two dessins d'enfants would be Galois conjugate if and only if they agreed on all the data from the list. Unfortunately such a list is not known up to now.

But several Galois invariants are known and can at least help distinguishing different orbits. The most fundamental invariants are derived from the correspondence of dessins d'enfants with Belyi pairs: If (X, β) is a Belyi pair and $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, there is a bijection between the ramification points of β on X and the ramification points of $\sigma\beta$ on σX ; moreover this bijection preserves the ramification orders. It therefore follows from the Riemann–Hurwitz formula that X and σX have the same genus. Translating these remarks to the corresponding dessin D and observing that the ramification points of β over $0, 1$ and ∞ correspond to the black vertices, the white vertices and the cells of D , respectively, we obtain:

Proposition 4.7. *The genus and the valency lists of a dessin d'enfants are Galois invariants.*

Recall that the genus of a dessin d'enfants $D = (G, i)$ is the genus of the surface onto which the dessin is drawn, see Definition 3.10. D has 3 valency lists: one for the black vertices, one for the white vertices, and one for the cells. These lists contain an entry for each vertex (resp. cell), and the entry is the valency of this vertex (resp. cell).

A famous example that these invariants do not suffice to separate Galois orbits is “Leila’s flower”, see [34], [41]. A few more subtle Galois invariants are known: the automorphism group of D , properties of the action of $\text{Aut}(D)$ on vertices or edges (like “regularity”); Zapponi [41] introduced the *spin structure* of a dessin and showed that it is a Galois invariant and in particular that it separates the two non-equivalent versions of Leila’s flower.

4.5 The action on \hat{F}_2

Recall that \mathbb{P} is the projective line $\mathbb{P}^1(\mathbb{C})$ with the three points $0, 1$ and ∞ removed. We saw in Proposition 3.2 that dessins d'enfants correspond bijectively to finite unramified coverings of \mathbb{P} and thus to (conjugacy classes of) finite index subgroups of $F_2 = \pi_1(\mathbb{P})$. In this section we explain how the Galois action on dessins induces an action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on \hat{F}_2 , the profinite completion of F_2 , and thus an embedding of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ into $\text{Aut}(\hat{F}_2)$.

We restrict our attention to dessins for which the associated covering of \mathbb{P} is Galois. The corresponding subgroup of F_2 is then normal, and we have no ambiguity “up to conjugation”. Moreover the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on finite index subgroups of F_2 can also be interpreted as an action on the set of finite quotient groups F_2/N , where N runs through the normal subgroups of F_2 . These finite quotient groups form a projective system of finite groups, with projections $F_2/N' \rightarrow F_2/N$ coming from inclusions $N' \subset N$. The inverse limit of this projective system is \hat{F}_2 , the profinite completion of F_2 .

The action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on \hat{F}_2 can be described quite explicitly. We sketch the approach by Y. Ihara, P. Lochak and M. Emsalem, see [17] and [6]; the details are worked out in e.g. [29]. Let x and y be generators of $F_2 = \pi_1(\mathbb{P})$ that correspond to

loops around 0 and 1, respectively. Their residue classes in the finite quotients F_2/N of F_2 define elements $(x \bmod N)_N$ and $(y \bmod N)_N$ of \hat{F}_2 , that we still denote by x and y . They are called *topological generators* since the subgroup they generate is dense in the profinite (or Krull) topology of \hat{F}_2 . Note that the group theoretical and the topological data are related as follows: if N is a finite index normal subgroup of F_2 which corresponds to the normal covering $p: Y \rightarrow \mathbb{P}^1$, then the order of $x \bmod N$ in F_2/N is the ramification index of p above 0, i.e. the l.c.m. of the ramification indices of the points in the fibre $p^{-1}(0)$; we denote this number by $e(N)$. With this notation at hand we can state the announced result:

Proposition 4.8. *For $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and x and y the topological generators of \hat{F}_2 described above we have*

$$\sigma \cdot x = x^{\chi(\sigma)} \quad \text{and} \quad \sigma \cdot y = f_\sigma^{-1} y^{\chi(\sigma)} f_\sigma.$$

The element $f_\sigma \in \hat{F}_2$ in the second formula will be explained at the end of this subsection; $x^{\chi(\sigma)}$ is the element of \hat{F}_2 defined by $x_N^{\chi(\sigma)} = (x \bmod N)^{\chi_{e(N)}(\sigma)}$, N running through the finite index normal subgroups of F_2 , where for a positive integer e , $\chi_e: \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow (\mathbb{Z}/e\mathbb{Z})^\times$ is the *cyclotomic character*, i.e. $\chi_e(\sigma) = n$ if $\sigma(\zeta_e) = \zeta_e^n$ for a primitive e -th root of unity ζ_e . Note that $\chi(\sigma) = (\chi_{e(N)}(\sigma))_N$ can be considered as an element of $\hat{\mathbb{Z}}^\times$.

The starting point for the proof of Proposition 4.8 is the equivalence of the following categories:

- finite normal coverings of $\mathbb{P}^1(\mathbb{C})$ unramified outside 0, 1 and ∞ ;
- finite normal coverings of $\mathbb{P}^1(\overline{\mathbb{Q}})$ unramified outside 0, 1 and ∞ ;
- finite normal holomorphic unramified coverings of \mathbb{P}^1 ;
- finite Galois extensions of $\overline{\mathbb{Q}}(T)$ unramified outside $T, T - 1$ and $\frac{1}{T}$.

The first equivalence is a consequence of Belyi’s theorem, the others are standard results on Riemann surfaces and algebraic curves (cf. Section 2 and the paragraph before Remark 3.8).

A crucial technical tool in the proof is the notion of a *tangential base point* of a Riemann surface X . It consists of a point together with a direction in this point. For the fundamental group with respect to a tangential base point, only closed paths are considered that begin and end in the prescribed direction. For example, we denote by $\vec{01}$ the tangential base point of $\mathbb{P}^1(\mathbb{C})$ which is located at 0 and whose direction is the positive real axis. We take the element $x \in \pi_1(\mathbb{P}^1, \vec{01})$ to be a small loop around 0, that begins and ends in 0 in the direction towards 1. (See Figure 6.)

Another important tool is the field $\mathcal{P}_{\vec{u}}$ of convergent Puiseux series in a tangential base point \vec{u} . For $\vec{u} = \vec{01}$, these are series of the form $\sum_{n=k}^\infty a_n T^{\frac{n}{e}}$ for some integer k , some positive integer e and complex coefficients a_n , such that the series converges in some punctured neighbourhood of 0. These Puiseux series then define meromorphic functions in a neighbourhood of 0 that is slit along the real line from 0 to 1.

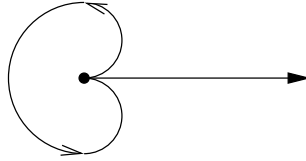


Figure 6. The generator x in the tangential base point $\vec{u} = \vec{0}1$.

Given a covering $p: X \rightarrow \mathbb{P}^1$ that is possibly ramified over 0 , the function field $\mathbb{C}(X)$ of X can be embedded into $\mathcal{P}_{\vec{u}}$ as follows: fix a point $v \in X$ above 0 and choose a local coordinate z in v such that p is given by $z \mapsto z^e$ in a neighbourhood of v . For a meromorphic function $f \in \mathbb{C}(X)$, let $f(z) = \sum a_n z^n$ be the Laurent expansion in v , and take $\sum a_n T^{\frac{n}{e}}$ to be its image in $\mathcal{P}_{\vec{u}}$, where e is the ramification index of p in v . If ζ is an e -th root of unity, we get another embedding by sending $\sum a_n z^n$ to $\sum a_n \zeta^n T^{\frac{n}{e}}$. These embeddings of $\mathbb{C}(X)$ into $\mathcal{P}_{\vec{u}}$ correspond bijectively to the tangential base points in v that are mapped to \vec{u} by z .

Lifting the (small!) loop x via p to X with starting point v we again get a closed path, but it may end in v in a direction different from the starting one. In this way we get an action of x on the tangential base points over \vec{u} and hence on the embeddings of $\mathbb{C}(X)$ into $\mathcal{P}_{\vec{u}}$.

Now let $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. To describe $\sigma \cdot x$, we have to specify, for each Belyi pair (X, β) , the embedding of $\mathbb{C}(X)$ into $\mathcal{P}_{\vec{u}}$ induced by $\sigma \cdot x$. By the above equivalences of categories, it suffices to take the function field $\overline{\mathbb{Q}}(X)$ and Puiseux series with coefficients in $\overline{\mathbb{Q}}$. Then σ acts on the coefficients of the Puiseux series, and an embedding that maps $f \in \overline{\mathbb{Q}}(X)$ to the series $\sum a_n T^{\frac{n}{e}}$ is transformed by $\sigma \cdot x$ into the embedding

$$f \mapsto \sum a_n T^{\frac{n}{e}} \mapsto \sum \sigma^{-1}(a_n) T^{\frac{n}{e}} \mapsto \sum \sigma^{-1}(a_n) \zeta^n T^{\frac{n}{e}} \mapsto \sum a_n \sigma(\zeta)^n T^{\frac{n}{e}}$$

where ζ is the root of unity corresponding to x . Since $\sigma(\zeta) = \zeta^{\chi_e(\sigma)}$, this shows the first formula of Proposition 4.8.

The second formula is proved similarly using a small loop y around 1 . The difference is that here we need the path t from 0 to 1 along the real line to make y into a closed path around \vec{u} . But t can also be interpreted as acting on embeddings of $\overline{\mathbb{Q}}(X)$ into the field of Puiseux series. Working with fundamental groupoids instead of the fundamental group, we can calculate $\sigma \cdot t$ in a similar way as $\sigma \cdot x$. The element f_σ in the formula then turns out to be $t^{-1} \sigma \cdot t$.

4.6 The action on the algebraic fundamental group

At first glance the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on \hat{F}_2 described in the previous section might look very special. But in fact it is an explicit example of the very general and conceptual construction of Galois actions on algebraic fundamental groups. We shall briefly explain this relation in this section.

The *algebraic fundamental group* $\pi_1^{\text{alg}}(X)$ of a scheme X is defined as the projective limit of the Galois (or deck transformation) groups of the finite normal étale coverings of X . In general, a morphism of schemes is called *étale* if it is “smooth of relative dimension 0”. If X is an algebraic curve over the complex numbers, this property is equivalent to the usual notion of an unramified covering. So in this case the projective system defining $\pi_1^{\text{alg}}(X)$ is the system of the finite quotient groups of the topological fundamental group $\pi_1(X)$. It follows that the algebraic fundamental group of a Riemann surface is the profinite completion of its topological fundamental group, cf. [27, p. 164].

In the proof of Proposition 4.8 we used the equivalence of four categories, namely the normal coverings of \mathbb{P}^1 as a variety over \mathbb{C} resp. $\overline{\mathbb{Q}}$ that are unramified outside 0, 1 and ∞ , the unramified coverings of \mathbb{P}^1 , and the suitably ramified Galois extensions of the function field $\overline{\mathbb{Q}}(T)$ of $\mathbb{P}^1_{\overline{\mathbb{Q}}}$. In all four categories, to every object there is associated a finite group (the Galois group of the covering resp. the field extension). The morphisms in the respective category make these groups into a projective system. The inverse limits of these systems are resp. the algebraic fundamental groups of $\mathbb{P}^1_{\mathbb{C}}$ and $\mathbb{P}^1_{\overline{\mathbb{Q}}}$, the profinite completion \hat{F}_2 of the topological fundamental group $\pi_1(\mathbb{P}^1) = F_2$, and the Galois group of $\Omega/\overline{\mathbb{Q}}(T)$, where Ω is the maximal Galois field extension of $\overline{\mathbb{Q}}(T)$ which is unramified outside $T, T - 1$ and $\frac{1}{T}$. As a corollary to Proposition 4.8 we thus obtain:

Remark 4.9. We have the following chain of group isomorphisms:

$$\pi_1^{\text{alg}}(\mathbb{P}^1_{\mathbb{C}}) \cong \pi_1^{\text{alg}}(\mathbb{P}^1_{\overline{\mathbb{Q}}}) \cong \hat{F}_2 \cong \text{Gal}(\Omega/\overline{\mathbb{Q}}(T)).$$

From the chain of Galois extensions $\mathbb{Q}(T) \subset \overline{\mathbb{Q}}(T) \subset \Omega$ we obtain the exact sequence

$$1 \rightarrow \text{Gal}(\Omega/\overline{\mathbb{Q}}(T)) \rightarrow \text{Gal}(\Omega/\mathbb{Q}(T)) \rightarrow \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow 1$$

of Galois groups (since $\text{Gal}(\overline{\mathbb{Q}}(T)/\mathbb{Q}(T)) \cong \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$). Using the isomorphisms of Remark 4.9, we obtain the following special case of Grothendieck’s exact sequence of algebraic fundamental groups, cf. [27, Theorem 8.1.1]:

$$1 \rightarrow \pi_1^{\text{alg}}(\mathbb{P}^1_{\overline{\mathbb{Q}}}) \rightarrow \pi_1^{\text{alg}}(\mathbb{P}^1_{\mathbb{Q}}) \rightarrow \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow 1.$$

This exact sequence provides us a priori with an outer action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $\pi_1^{\text{alg}}(\mathbb{P}^1_{\overline{\mathbb{Q}}})$, i.e. a group homomorphism from $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ to the outer automorphism group $\text{Out}(\hat{F}_2) = \text{Aut}(\hat{F}_2)/\text{Inn}(\hat{F}_2)$ of $\hat{F}_2 \cong \pi_1^{\text{alg}}(\mathbb{P}^1_{\overline{\mathbb{Q}}})$. The additional information that we obtain from the explicit results in Section 4.5 is that the sequence splits, and that the outer action thus is in fact a true action. In other words, the construction in Section 4.5 corresponds to a particular splitting homomorphism $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \pi_1^{\text{alg}}(\mathbb{P}^1_{\overline{\mathbb{Q}}})$.

4.7 The Grothendieck–Teichmüller group

In the last two sections we established a group homomorphism $\tau: \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow \text{Aut}(\hat{F}_2)$ coming from the Galois action on dessins. It follows from Proposition 4.5 that τ is injective. In this way, $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is embedded into a group whose definition does not refer to field extensions or number theory; but since $\text{Aut}(\hat{F}_2)$ is a very large group that is not well understood, there is not much hope that this embedding alone can shed new light on the structure of the group $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

In his paper [4], V. Drinfel'd defined a much smaller subgroup of $\text{Aut}(\hat{F}_2)$, which still contains the image of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ under τ . He called this group the *Grothendieck–Teichmüller group* and denoted it by \widehat{GT} . It is still an open question whether $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is equal to \widehat{GT} . In this section we present the definition of \widehat{GT} and indicate how $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is embedded into \widehat{GT} .

We saw in Proposition 4.8 that for $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, the automorphism $\tau(\sigma) \in \text{Aut}(\hat{F}_2)$ is completely determined by the “exponent” $\lambda_\sigma = \chi(\sigma) \in \hat{\mathbb{Z}}^\times$ and the “conjugator” $f_\sigma \in \hat{F}_2$. The explicit knowledge of f_σ makes it possible to show that it acts trivially on abelian extensions of $\mathbb{Q}(T)$ and therefore that f_σ is contained in the (closure of the) commutator subgroup \hat{F}'_2 of \hat{F}_2 , see [17, Proposition 1.5] or [29, Section 4.4]. The composition of automorphisms implies that pairs in $\hat{\mathbb{Z}}^\times \times \hat{F}'_2$ that come from Galois automorphisms, are multiplied according to the rule

$$(\lambda, f) \cdot (\mu, g) = (\lambda\mu, fF_{\lambda,f}(g)), \tag{4.1}$$

where $F_{\lambda,f}$ is the endomorphism of \hat{F}_2 which is induced by $x \mapsto x^\lambda$ and $y \mapsto f^{-1}y^\lambda f$. Motivated by his investigations of braided categories Drinfel'd found some natural conditions to impose on such pairs (λ, f) :

Definition 4.10. a) Let \widehat{GT}_0 be the set of pairs $(\lambda, f) \in \hat{\mathbb{Z}}^\times \times \hat{F}'_2$ that satisfy

$$\begin{aligned} \text{(I)} \quad & \theta(f) f = 1, \\ \text{(II)} \quad & \omega^2(fx^m) \omega(fx^m) fx^m = 1, \end{aligned}$$

where $m = \frac{1}{2}(\lambda - 1)$ and θ resp. ω are the automorphisms of \hat{F}_2 defined by $\theta(x) = y$, $\theta(y) = x$ resp. $\omega(x) = y$, $\omega(y) = (xy)^{-1}$.

b) Let \widehat{GT}_0 be the group of elements in \widehat{GT}_0 that are invertible for the composition law (4.1).

c) The *Grothendieck–Teichmüller group* \widehat{GT} is the subgroup of \widehat{GT}_0 of elements that satisfy the further relation

$$\text{(III)} \quad \rho^4(\tilde{f}) \rho^3(\tilde{f}) \rho^2(\tilde{f}) \rho(\tilde{f}) \tilde{f} = 1$$

which takes place in the profinite completion \hat{K}_5 of the pure braid group K_5 on five strands. This group is generated by elements $x_{i,i+1}$ for $i \in \mathbb{Z}/5\mathbb{Z}$, and ρ is the automorphism that maps $x_{i,i+1}$ to $x_{i+3,i+4}$; finally $\tilde{f} = f(x_{1,2}, x_{2,3})$.

It is not obvious from the definition that \widehat{GT}_0 is a group, more precisely the group of all elements in \widehat{GT}_0 that induce automorphisms on \widehat{F}_2 . The proofs of these facts can be found in [21] and [35]; a careful proof with all details is contained in [8].

The relation between the Galois group and the Grothendieck–Teichmüller group is stated in

Theorem 4.11. *Via the homomorphism τ , $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ becomes a subgroup of \widehat{GT} .*

That the pairs $(\lambda_\sigma, f_\sigma)$ coming from Galois automorphisms satisfy the first two relations (I) and (II) can be shown using the explicit computations of the action on \widehat{F}_2 , see e.g. [6] or [29, Section 5.1] for a detailed version. The proof of the third relation is a bit more complicated; we refer to [17].

5 Origamis

5.1 Introduction to origamis

In the world of dessins d’enfants we study finite unramified holomorphic coverings $\beta: X^* \rightarrow \overline{\mathbb{P}}$ between Riemann surfaces. We have seen in Section 3 that such a covering is up to equivalence completely determined by the covering $R^* \rightarrow \overline{S}$ of the underlying topological surfaces. It is very tempting to generalize this and look at general finite unramified coverings between *punctured closed surfaces*, i.e. closed surfaces with finitely many points removed. It turns out that choosing the once-punctured torus $E^* = E \setminus \infty$ (∞ some point on the torus E) as base surface instead of \overline{S} is in some sense the next “simplest” case. Following the spirit and the denominations of [20], we call a covering $p: R \rightarrow E$ ramified at most over the point ∞ an *origami*. Note that this defines the unramified covering $R^* \rightarrow E^*$, where $R^* = R \setminus p^{-1}(\infty)$, and conversely each finite unramified cover of E^* is obtained in this way. Similarly as for Belyi pairs we call two origamis $O_1 = (p_1: R_1 \rightarrow E)$ and $O_2 = (p_2: R_2 \rightarrow E)$ *equivalent*, if there exists some homeomorphism $f: R_1 \rightarrow R_2$ such that $p_2 \circ f = p_1$.

The first observation is that the different combinatorial descriptions of a topological covering $R^* \rightarrow \overline{S}$ explained in Section 3 smoothly generalize to arbitrary unramified coverings of punctured closed surfaces. In the case of origamis we obtain the equivalent descriptions stated in Proposition 5.1. The generalization of a dessin d’enfants can be done as follows: In the case of the three-punctured sphere \overline{S} , we used that we obtain a cell if we remove the interval $I = [0, 1]$ from \overline{S} . For the once-punctured torus E^* we remove two simple closed curves a and b starting in the puncture as shown in Figure 7.

The cell that we obtain in this way is bounded in E by four edges labeled with a and b . We identify it with a quadrilateral. Similarly as described in Section 3, we have for an origami $p: R^* \rightarrow E^*$ that $R^* \setminus (p^{-1}(a) \cup p^{-1}(b))$ decomposes into a finite union of quadrilaterals. Unlike the case of dessins, the map p restricted to $R^* \setminus (p^{-1}(a) \cup p^{-1}(b))$ is unramified and the number of quadrilaterals is the degree d of p . We retrieve the surface R^* by gluing the quadrilaterals. Hereby only edges

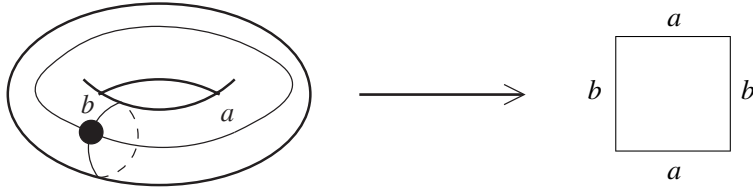


Figure 7. Removing two simple closed curves from the torus E gives a cell.

labeled with the same letter a or b may be glued. Furthermore we have to respect orientations. Altogether this leads to the following “origami-rules”: Glue finitely many copies of the Euclidean unit square such that

- each left edge is glued to a unique right edge and vice versa;
- each upper edge is glued to a unique lower one and vice versa;
- we obtain a connected surface R .

R has a natural covering map $p: R \rightarrow E$ by mapping each square to one square which forms the torus E . The map p is unramified except possibly above ∞ , which is the one point on E that results from the vertices of the square. Thus $p: R \rightarrow E$ is an origami. Note that for the moment we are only interested in the topological covering p and it would not be necessary to take Euclidean unit squares which endows R in addition with a metric.

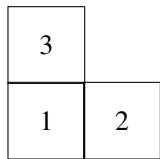
It is remarkable that by some fancy humour of nature the fundamental groups of \mathbb{P}^1 and E^* are both the same abstract group, the free group F_2 in two generators.

Proposition 5.1. *An origami $p: R^* \rightarrow E^*$ is up to equivalence uniquely determined by the following:*

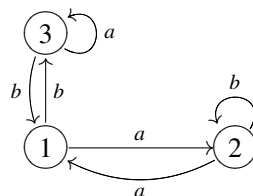
- A surface obtained from gluing Euclidean unit squares according to the “origami rules” (see above).
- A finite oriented graph whose edges are labeled with a and b such that each vertex has precisely two incoming edges and two outgoing edges one of which is labeled with a and one with b , respectively.
- A monodromy map $\alpha: F_2 \rightarrow S_d$ for some $d \geq 1$ up to conjugation in S_d . We denote $\sigma_a = \alpha(a)$ and $\sigma_b = \alpha(b)$. Similarly as in Section 3 we require a monodromy map by definition to induce a transitive action on $\{1, \dots, d\}$.
- A finite index subgroup U of F_2 up to conjugation in F_2 .

The equivalences stated in Proposition 5.1 are carried out in detail e.g. in [32, Section 1]. Thus we restrict here to giving the different descriptions for an example.

Example 5.2. In the following we describe the origami, commonly known as $L_{2,2}$, in the different ways assembled in Proposition 5.1.



Gluing squares according to the origami-rules: Opposite edges are glued.



The finite graph which describes the origami.

The monodromy map is the map $\alpha : F_2 \rightarrow S_3$ which is given by $a \mapsto \sigma_a = (1\ 2)$ and $b \mapsto \sigma_b = (1\ 3)$ and a corresponding subgroup of F_2 is $U = \langle a^2, b^2, aba^{-1}, bab^{-1} \rangle$.

A short Euler characteristic calculation shows that for this example the surface R has genus 2. The covering map $p : R \rightarrow E$ has degree 3 and the puncture ∞ has one preimage on R .

5.2 Teichmüller curves

So far, we have only considered coverings between topological surfaces. A crucial point of the theory of dessins d'enfants is that the three-punctured sphere \mathbb{S} has a unique complex structure as a Riemann surface. Therefore choosing a finite unramified covering $\beta : R^* \rightarrow \mathbb{S}$ defines a closed Riemann surface of genus $g = \text{genus}(R)$: Take the unique complex structure on the sphere and lift it via p to R^* . For the so obtained Riemann surface X^* there is a unique closed Riemann surface X into which we can embed X^* holomorphically. Hence, β defines the point $[X]$ in M_g , respectively $[X^*]$ in $M_{g,n}$, where M_g is the moduli space of regular complex curves of genus g , $M_{g,n}$ is the moduli space of regular complex curves with n marked points and n is the number of points in $X \setminus X^*$. Recall from algebraic geometry that M_g and $M_{g,n}$ are themselves complex varieties. In fact they are obtained by base change from schemes defined over \mathbb{Z} . By Belyi's Theorem the image points $[X] \in M_g$, respectively $[X^*] \in M_{g,n}$ are points defined over $\overline{\mathbb{Q}}$.

How can we generalize this construction for origamis? Since we have a one-dimensional family of complex structures on the torus E , an origami $O = (p : S \rightarrow E)$ will define a collection of Riemann surfaces depending on one complex parameter. More generally, an unramified cover $p : R_1^* \rightarrow R_2^*$ between punctured closed surfaces naturally defines the holomorphic and isometric embedding

$$\iota_p : \mathcal{T}(R_2^*) \hookrightarrow \mathcal{T}(R_1^*), \quad [\mu] \mapsto [p^* \mu], \tag{5.1}$$

from the Teichmüller space $\mathcal{T}(R_2^*)$ to the Teichmüller space $\mathcal{T}(R_1^*)$, which maps a complex structure μ on R_2^* to the complex structure $p^* \mu$ on R_1^* obtained as pull back via p . We now project the image $B := \iota_p(\mathcal{T}(R_2^*))$ to $M_{g,n}$ and further to M_g . How do the images in the moduli spaces look like? Can we describe their geometry based on the combinatorial data of the map p with which we started?

In the following we restrict to the case of origamis. Thus we obtain an embedding $\iota_p: \mathbb{H} \cong \mathcal{T}_{1,1} \hookrightarrow \mathcal{T}_{g,n}$ which is holomorphic and isometric. Such a map is called a *Teichmüller embedding* and its image in Teichmüller space is called a *Teichmüller disk*. Teichmüller disks arise in general from the following construction, which is described in detail and with further hints to literature e.g. in [15]: Let X be a compact Riemann surface together with a flat structure ν on it; i.e. we have an atlas on $X \setminus \{P_1, \dots, P_n\}$ for finitely many points P_i such that all transition maps are locally of the form $z \mapsto \pm z + c$ with some constant c . Suppose furthermore that the P_i 's are cone singularities of ν . Then each matrix $A \in \mathrm{SL}_2(\mathbb{R})$ induces a new flat structure ν_A by composing each chart with the affine map $z \mapsto A \cdot z$. This defines a map

$$\iota_\nu: \mathbb{H} \cong \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R}) \rightarrow \mathcal{T}_g, \quad [A] \mapsto [\nu_A] \quad (5.2)$$

which is in fact a holomorphic and isometric embedding, i.e. it is a Teichmüller embedding. It is a nice feature that for an origami $O = (p: S \rightarrow E)$ the surface S comes with a flat structure: One identifies E with $\mathbb{C}/(\mathbb{Z} \oplus \mathbb{Z}i)$. This quotient carries a natural flat structure induced by the Euclidean structure on \mathbb{C} . It is actually a translation structure, i.e. the transition maps are of the form $z \mapsto z + c$. Note that in the description of origamis with the “origami-rules” we obtain the translation structure for free, if we glue the edges of the unit squares via translations. The translation surfaces arising in this way are often called *square tiled surfaces*. It is not hard to see that for an origami O the induced maps ι_ν defined in (5.2) and $\mathrm{proj}_{g,n} \circ \iota_p$ (with ι_p from (5.1)) from \mathbb{H} to \mathcal{T}_g are equal (see e.g. [31, p. 11]); here $\mathrm{proj}_{g,n}: \mathcal{T}_{g,n} \rightarrow \mathcal{T}_g$ is the natural projection obtained by forgetting the marked points. In the following we will therefore denote the map $\mathrm{proj}_{g,n} \circ \iota_p = \iota_\nu$ just by ι_O .

The study of Teichmüller disks has lead to vivid research activities connecting different mathematical fields such as dynamical systems, algebraic geometry, complex analysis and geometric group theory. Many different authors have contributed to this field in the last years with a multitude of interesting results (see e.g. [13] in Volume I of this Handbook, [26] in this volume, or [15] for comments on literature). Important impacts to this topic were already given in [36]. An important tool for the study of Teichmüller disks is the *Veech group*, which was introduced in [37]. For a translation surface (X, ν) one takes the *affine group* $\mathrm{Aff}(X, \nu)$ of diffeomorphisms which are locally affine. The Veech group $\Gamma(X, \nu)$ is its image in $\mathrm{SL}_2(\mathbb{R})$ under the derivative map D , which maps each affine diffeomorphism to its linear part. The article [26] in this volume gives a more detailed introduction to Veech groups, an overview on recent results and hints to more literature. In Theorem 5.3 we list the properties of Veech groups that we will use. It is a collection of results contributed by different authors, which we have learned mainly from [37], [5] and [23]. Section 2.4 in [15] contains a quite detailed summary of them and further references. An important ingredient is the fact that if we have a translation structure and pull it back by an affine diffeomorphism f , it is changed by composing each chart with the affine map $z \mapsto Az$, where A is the inverse of the derivative of f . Therefore the elements in the mapping class group which come from affine diffeomorphisms stabilize the image Δ of the Teichmüller

embedding ι_ν . One shows that in fact, they form the full stabilizer of Δ . Furthermore the group $\text{Trans}(X, \mu) = \{f \in \text{Aff}(X, \mu) \mid D(f) = \text{identity matrix}\}$ acts trivially on Δ and $\Gamma(X, \nu) \cong \text{Aff}(X, \nu)/\text{Trans}(X, \nu)$.

Theorem 5.3. *Let X be a compact Riemann surface and ν a translation structure on X with finitely many cone singularities. Let $\iota = \iota_\nu: \mathbb{H} \hookrightarrow \mathcal{T}_g$ be the corresponding Teichmüller embedding, Δ its image in \mathcal{T}_g , $p_g: \mathcal{T}_g \rightarrow M_g$ the natural projection and Γ_g the mapping class group for genus g . Then we have:*

- $\text{Stab}_{\Gamma_g}(\Delta) \cong \text{Aff}(X, \nu)$.
- $p_g|_\Delta$ factors through the quotient map $q: \Delta \rightarrow \Delta/\Gamma(X, \nu)$, i.e. we obtain a map $n: \Delta/\Gamma(X, \nu) \rightarrow M_g$ with $p_g|_\Delta = n \circ q$.
- The image of Δ in M_g is an algebraic curve C if and only if the Veech group $\Gamma(X, \nu)$ is a lattice in $\text{SL}_2(\mathbb{R})$. If this is the case, C is called a Teichmüller curve, and n is a birational map. Therefore it is the normalization of C . C is birationally equivalent to a mirror image of $\mathbb{H}/\Gamma(X, \nu)$.

In the following we will only consider Teichmüller embeddings coming from origamis. In this case, the Veech group is commensurable to $\text{SL}_2(\mathbb{Z})$ and thus a lattice in $\text{SL}_2(\mathbb{R})$. It turns out to be useful given an origami $O = (p: X \rightarrow E)$, to consider only affine diffeomorphisms which preserve $p^{-1}(\infty)$. The image of this group is in fact a subgroup of $\text{SL}_2(\mathbb{Z})$. Following the notations in [31], we denote it by $\Gamma(O)$ and call it the Veech group of the origami O . If we replace \mathcal{T}_g by $\mathcal{T}_{g,n}$ and M_g by $M_{g,n}$ in Theorem 5.3, then $\Gamma(O)$ becomes the effective stabilizing group of $\Delta \subseteq \mathcal{T}_{g,n}$. [30] describes an algorithm which computes $\Gamma(O)$.

Coming back to the question asked at the beginning of this section, we state that in the case of an origami the image of the map ι_p in M_g is an algebraic curve which comes from a Teichmüller disk. In the following sections we study these curves, which we call *origami curves*. More precisely we point out some explicit relations between them and dessins d'enfants.

6 Galois action on origamis

In [20], Lochak suggested to study the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on origamis in some sense as a generalization of the action on dessins d'enfants following the spirit of Grothendieck's *Esquisse d'un programme*. Recall from Section 4 that for each $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and each projective curve X defined over $\overline{\mathbb{Q}}$, we obtain a projective curve ${}^\sigma X$. This actually defines an action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $M_{g, \overline{\mathbb{Q}}}$, the moduli space of regular projective curves which are defined over $\overline{\mathbb{Q}}$.

In the following we want to make the definition of an action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on origamis more precise: Let $O = (p: R \rightarrow E)$ be an origami with $\text{genus}(R) = g$. Recall that O defines a whole family of coverings $p_A: X_A \rightarrow E_A$ ($A \in \text{SL}_2(\mathbb{R})$)

between Riemann surfaces. It follows from Theorem 5.3 that two coverings p_A and $p_{A'}$ are equivalent, if and only if A and A' are mapped to the same point on $\tilde{C}(O) = \mathbb{H}/\Gamma(O)$, where $\Gamma(O)$ is the Veech group; furthermore X_A and $X_{A'}$ are isomorphic, if and only if the two matrices are mapped to the same point on the possibly singular curve $C(O)$. In particular we may parametrize the family of coverings by the elements t of $\tilde{C}(O)$ and denote them as $p_t: X_t \rightarrow E_t$. In the following we will restrict to those t for which $p_t: X_t \rightarrow E_t$ is defined over $\overline{\mathbb{Q}}$. We denote this subset of $\tilde{C}(O)$ by $\tilde{C}_{\overline{\mathbb{Q}}}(O)$ and similarly we write $C_{\overline{\mathbb{Q}}}(O)$.

Let us now pick some $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. One immediately has two ideas on how σ could act on origami curves; both lead at first glance to a problem:

- $C = C_{\overline{\mathbb{Q}}}(O)$ is mapped to its image ${}^\sigma C_{\overline{\mathbb{Q}}} = \{\sigma(X_t) \mid t \in \tilde{C}_{\overline{\mathbb{Q}}}(O)\}$. Is the image again an origami curve? Or more precisely is there some origami ${}^\sigma O$ such that ${}^\sigma C_{\overline{\mathbb{Q}}} = C_{\overline{\mathbb{Q}}}({}^\sigma O)$?
- For $p_t: X_t \rightarrow E_t$ (defined over $\overline{\mathbb{Q}}$) define ${}^\sigma p_t$ similarly as ${}^\sigma \beta$ in Section 4. Each ${}^\sigma p_t$ defines an origami. Do they all lead to the same origami curve?

In [25, Proposition 3.2] Möller showed that the two approaches lead to the same unique origami curve ${}^\sigma C$. We denote the corresponding origami by ${}^\sigma O$, i.e. ${}^\sigma C = C({}^\sigma O)$.

The basic ingredient of the proof in [25] is to consider the Hurwitz space of all coverings with the same ramification behaviour as p for a given origami $O = (p: X \rightarrow E)$. By a result of Wewers in [38], one obtains a smooth stack over \mathbb{Q} . The covering p lies in a connected component of it, whose image in moduli space is the origami curve $C(O)$. Möller deduces from this that $C(O)$ is defined over a number field and that one has the natural action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ described above.

The Galois action on origamis is faithful in the following sense: For each σ in $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ there exists an origami O such that $C(O) \neq {}^\sigma C(O)$. This is shown in [25, Theorem 5.4]. The proof uses the faithfulness of the action of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on dessins of genus 0 (see Proposition 4.6). Starting with a Belyi morphism $\beta: \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$ with ${}^\sigma \beta \not\cong \beta$, one takes the fibre product of β with the degree 2 morphism $E \rightarrow \mathbb{P}^1(\mathbb{C})$, where E is an elliptic curve which is defined over the fixed field of σ in $\overline{\mathbb{Q}}$. Precomposing the obtained morphism with the normalization and postcomposing with multiplication by 2 on E , gives an origami as desired.

This is a nice example for some interplay going on between origamis and dessins in the way it was proposed in [20]. In the next two sections we describe two further ways, how origamis and dessins can be related.

7 A dessin d'enfants on the origami curve

Let $O = (p: R \rightarrow E)$ be an origami and let $\Gamma(O)$ be its Veech group. In this section we consider the corresponding Teichmüller curve in the moduli space $M_{g,n}$ of n -punctured curves and denote it by $C(O)$. As always, g is the genus of R and n is

the number of preimages of the ramification point $\infty \in E$. Let $\tilde{C}(O)$ be the quotient $\mathbb{H}/\Gamma(O)$. Recall from Theorem 5.3 that $\tilde{C}(O)$ is the normalization of $C(O)$.

The quotient $\tilde{C}(O)$ naturally defines a dessin d’enfants, as it was pointed out in [20, Proof of Proposition 3.2]: $\Gamma(O)$ is a finite index subgroup of $SL_2(\mathbb{Z})$. Thus we obtain a finite covering $q: \mathbb{H}/\Gamma(O) \rightarrow \mathbb{H}/SL_2(\mathbb{Z}) \cong \mathbb{A}^1(\mathbb{C})$. We may fill in cusps and extend q to a finite covering $q: X \rightarrow \mathbb{P}^1(\mathbb{C})$ of closed Riemann surfaces. This covering has ramification at most above three points of $\mathbb{P}^1(\mathbb{C})$: the two ramification points of the map $\mathbb{H} \rightarrow \mathbb{H}/SL_2(\mathbb{Z})$ and the cusp $\infty = \mathbb{P}^1(\mathbb{C}) \setminus \mathbb{A}^1(\mathbb{C})$. Hence q is a Belyi morphism. Applying once more the Theorem of Belyi, one obtains for free that the complex curve $\tilde{C}(O)$ is defined over $\overline{\mathbb{Q}}$.

The dessin corresponding to q is obtained quite explicitly from this description, as we explain in the following. Recall that $SL_2(\mathbb{Z})$ is generated by the two matrices

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We take our favourite fundamental domain for $SL_2(\mathbb{Z})$, namely the ideal triangle with vertices $P = \zeta_3 = e^{\frac{2\pi i}{3}}$, $Q = \zeta_3 + 1$ and the cusp $R = \infty$, see Figure 8. Recall that P

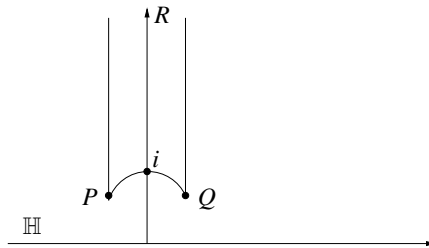


Figure 8. Fundamental domain of $SL_2(\mathbb{Z})$.

is a fixed point of the matrix $S \circ T$, which is of order 3 in $PSL_2(\mathbb{Z})$. Furthermore, i is a fixed point of the order 2 matrix S and thus a further hidden vertex of the fundamental domain. Finally, the transformation T maps the edge PR to QR and S maps the edge Pi to Qi . We obtain $\mathbb{P}^1(\mathbb{C})$ by “gluing” PR to QR and Pi to Qi and filling in the cusp at ∞ .

In order to make the dessin explicit, we identify the image of P on $\mathbb{P}^1(\mathbb{C})$ with 0, the image of i with 1 and the image of the cusp with ∞ . The geodesic segment PQ is then mapped to our interval I ; its preimage $q^{-1}(I)$ on X is the dessin.

The algorithm in [30] gives the Veech group Γ by a system G of generators and a system C of coset representatives. C is in fact a Schreier-transversal with respect to the generators S and T of $SL_2(\mathbb{Z})$, i.e. each element in C is given as a word in S and T such that each prefix of it is also in C . Therefore C defines a connected fundamental domain F of Γ which is the union of translates of the triangle PQR ; for each coset we obtain one translate. The identification of the boundary edges of F are given by the generators in G . Thus the fundamental domain F is naturally tessellated by triangles,

which indicate the Belyi morphism. The dessin is the union of all translates of the edge PQ .

In the following we describe the dessin for an example. We take the origami D drawn in Figure 9, which is studied in [32]. The Veech group $\Gamma = \Gamma(D)$ and the

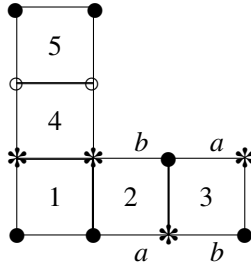


Figure 9. The origami D . Edges with the same label and unlabeled edges that are opposite are glued.

fundamental domain of Γ are given in Section 3 of [32]. The index of Γ in $SL_2(\mathbb{Z})$ is 24 and the quotient \mathbb{H}/Γ is a surface of genus 0 with six cusps. Figure 10 shows a fundamental domain of $\Gamma(D)$.

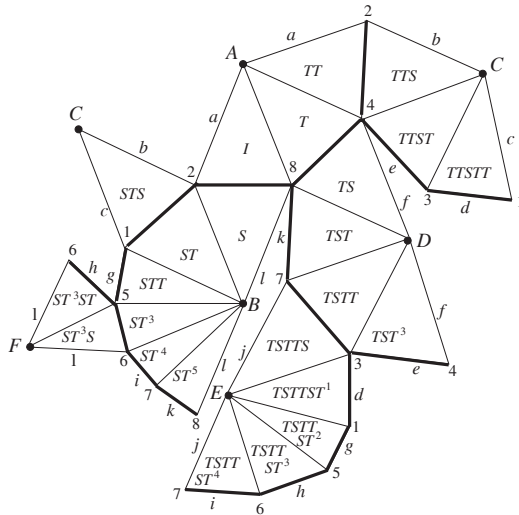


Figure 10. The fundamental domain for the Veech group $\Gamma(D)$.

We use a schematic diagram: Each triangle represents a translate of the triangle PQR . The vertices labeled with A, \dots, F are the cusps. The thickened edges form the dessin. The planar graph is redrawn in Figure 11. This picture matches its embedding into $\mathbb{P}^1(\mathbb{C})$.

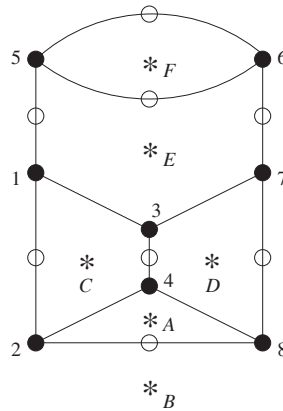


Figure 11. The dessin on the origami curve $\tilde{C}(D)$.

8 Dessins d'enfants related to boundary points of origami curves

Let $O = (p: X \rightarrow E)$ be an origami of genus $g \geq 2$ and $C(O)$ the corresponding origami curve in the moduli space M_g . Recall that the algebraic variety M_g can be compactified by a projective variety \overline{M}_g , the *Deligne–Mumford compactification*, which classifies stable Riemann surfaces, i.e. surfaces with “nodes” (see below for a precise definition). The closure $\overline{C(O)}$ of $C(O)$ in \overline{M}_g is a projective curve; its boundary $\partial C(O) = \overline{C(O)} - C(O)$ consists of finitely many points, called the *cusps* of the origami curve.

In this section we shall associate in a natural way dessins d'enfants to the cusps of origami curves.

8.1 Cusps of origami curves

There is a general procedure to determine the cusps of algebraic curves in moduli space, called *stable reduction*. We first recall the notion of a stable Riemann surface:

Definition 8.1. A one-dimensional connected compact complex space X is called *stable Riemann surface* if

- (i) every point of X is either smooth or has a neighbourhood which is analytically isomorphic to $\{(z, w) \in \mathbb{C}^2 \mid z \cdot w = 0\}$ (such a point is called a *node*), and
- (ii) every irreducible component of X that is isomorphic to $\mathbb{P}^1(\mathbb{C})$ intersects the other components in at least three points.

Now let C_0 be an algebraic curve in M_g and $x \in \partial C$ a cusp of C . We may assume that C_0 is smooth (by removing the finitely many singular points of C_0) and that also $C = C_0 \cup \{x\}$ is smooth (by passing to the normalization). Next we assume that we have a family $\pi_0: \mathcal{C}_0 \rightarrow C_0$ of smooth curves over C_0 , i.e. a proper flat morphism π_0 such that the fibre $X_c = \pi_0^{-1}(c)$ over a point $c \in C_0$ is isomorphic to the compact Riemann surface which is represented by c (for this we may have to pass to a finite covering of C_0). The stable reduction theorem (see [11, Proposition 3.47]) states that, after passing to another finite covering C' of C (which can be taken totally ramified over x), the family $\mathcal{C}_0 \times_{C_0} C'_0$ extends to a family $\pi: \mathcal{C} \rightarrow C'$ of stable Riemann surfaces, and that the stable Riemann surface $X_\infty = \pi^{-1}(x)$, that occurs as fibre over the cusp x , is independent of the choice of C' .

Although the proof of the stable reduction theorem is constructive, this construction usually becomes quite involved: First examples are discussed in [11, Section 3C]; a particularly nice example for the cusp of an origami curve is worked out in [2].

If the algebraic curve C_0 in M_g is a Teichmüller curve, there is a much more direct way to find the stable Riemann surface associated to a cusp, avoiding the stable reduction theorem. This construction is based on the description of Jenkins–Strebel rays in [22] and worked out in detail in [15, Section 4.1]. The basic observation is that for every cusp x of a Teichmüller curve C there is a direction on the flat surface X defining C in which X is decomposed into finitely many cylinders; this direction is associated to a Jenkins–Strebel differential on X . The stable Riemann surface corresponding to the cusp is now obtained by contracting the core curves of these cylinders. See [15, Section 4.2] for a proof of this result.

In the special case of a Teichmüller curve coming from an origami, the construction is particularly nice: Let $O = (p: X \rightarrow E)$ be an origami as above. The squares define a translation structure on X and divide it into horizontal cylinders, which we denote by C_1, \dots, C_n . The core lines c_1, \dots, c_n of these cylinders are the connected components of the inverse image $p^{-1}(a)$ of the horizontal closed path a on the torus E . Contracting each of the closed paths c_i to a point x_i turns X into a surface X'_∞ which is smooth outside x_1, \dots, x_n . It is described in [15, Section 4.1] how to put, in a natural way, a complex structure on X'_∞ . Then X'_∞ satisfies the above Definition 8.1, except perhaps (ii). If an irreducible component of X'_∞ violates (ii), we can contract this component to a single point and obtain a complex space which still satisfies (i). After finitely many such contractions we obtain a stable Riemann surface X_∞ . This process of contracting certain components is called “stabilizing”. For simplicity we used here the horizontal cylinders. But the construction is the same for any direction in which there is a decomposition into cylinders.

If we apply this construction to the torus E itself, we obtain a surface E_∞ which has a single node and whose geometric genus is zero. This surface is known as *Newton’s node* and can algebraically be described as the singular plane projective curve with affine equation $y^2 = x^3 - x^2$.

Note that in the above construction, the covering p naturally extends to a covering $p_\infty: X_\infty \rightarrow E_\infty$, which is ramified at most over the critical point ∞ of p (or, to be

precise, the point on E_∞ that corresponds to ∞ on E), and over the node. This is illustrated in the following picture for the origami W from [16]:

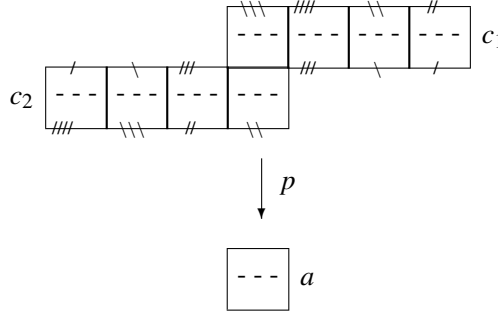


Figure 12. The origami covering for the cusp of W .

8.2 The dessin d'enfants associated to a boundary point

The construction in 8.1 leads in a natural way to a dessin d'enfants, as was observed in [20, Section 3.1], where it is attributed to L. Zapponi. Let, as before, $O = (p: X \rightarrow E)$ be an origami of genus $g \geq 2$ and $C(O)$ the corresponding origami curve in M_g . Furthermore let $x \in \partial C(O)$ be a cusp and X_∞ the stable Riemann surface that is represented by x . Denote by X_1, \dots, X_n the irreducible components of X_∞ and by $p_\infty: X_\infty \rightarrow E_\infty$ the covering discussed at the end of the previous section. For each $i = 1, \dots, n$, the restriction of p_∞ to X_i gives a finite covering $p_i: X_i \rightarrow E_\infty$. For the degrees d_i of p_i we have the obvious relation

$$\sum_{i=1}^n d_i = d = \deg(p).$$

Now let C_i be the normalization of X_i ($i = 1, \dots, n$). Then p_i induces a covering $f_i: C_i \rightarrow \mathbb{P}^1(\mathbb{C})$ (which is the normalization of E_∞).

Proposition 8.2. *For every boundary point x of the origami curve $C(O)$ and each irreducible component X_i of the stable Riemann surface X_∞ , the covering $f_i: C_i \rightarrow \mathbb{P}^1(\mathbb{C})$ is a Belyi morphism.*

Proof. We already noticed in Section 8.1 that the covering $p_\infty: X_\infty \rightarrow E_\infty$ is ramified at most over the critical point ∞ of p and over the node. The normalization map $\mathbb{P}^1(\mathbb{C}) \rightarrow E_\infty$ maps two different points to the node, so each f_i can be ramified over these two points, and otherwise only over the inverse image of ∞ . □

In Section 3 we explained that a dessin d'enfants is completely determined by the monodromy map of the corresponding Belyi map β , i.e. two permutations σ_0 and σ_1 in S_d , where d is the degree of β .

Similarly, an origami $O = (p: X \rightarrow E)$ is also determined by two permutations σ_a and σ_b , see Proposition 5.1. Recall that they describe the gluing of the squares in horizontal resp. vertical direction. Thus a horizontal cylinder consists of the squares in a cycle of σ_a , and a vertical cylinder consists of the squares in a cycle of σ_b .

There is a nice relation between the permutations σ_a and σ_b of the origami O and the permutations σ_0 and σ_1 of the dessin d'enfants associated to the boundary point x on $\partial C(O)$ which is obtained by contracting the centre lines of the horizontal cylinders. It was first made explicit (but not published) by Martin Möller as follows:

Proposition 8.3. *Let $O = (p: X \rightarrow E)$ be an origami of degree d , and let σ_a, σ_b be the corresponding permutations in S_d . Then the dessin d'enfants associated to the horizontal boundary point on $C(O)$ is defined by*

$$\sigma_0 = \sigma_a, \quad \sigma_1 = \sigma_b \sigma_a \sigma_b^{-1}.$$

In this proposition, we use a slightly more general definition of a dessin d'enfant. We allow the surface to be not connected. The dessin is then by definition the union of the dessins on the irreducible components described above in Proposition 8.2.

Proof. Recall the construction of the covering $p_\infty: X_\infty \rightarrow E_\infty$ and the Belyi map $f_\infty: \bigcup_{i=1}^n X_i \rightarrow \mathbb{P}^1(\mathbb{C})$: E_∞ is obtained from the torus E by contracting the horizontal path a to a single point, the node of E_∞ . Let U be a neighbourhood of the node, analytically isomorphic to $\{(z, w) \in \mathbb{C}^2 \mid |z| \leq 1, |w| \leq 1, z \cdot w = 0\}$. U is the union of two closed unit disks U_0, U_1 which are glued together at their origins. In the normalization $\mathbb{P}^1(\mathbb{C})$ of E_∞ , the node has two preimages, and the preimage of U is the disjoint union of the two disks U_0 and U_1 . The loops l_0 and l_1 can be taken as simple loops in U_0 resp. U_1 around the origin. On E_∞ , l_0 and l_1 are the images of parallels a_0 and a_1 of a , one above a , the other below:

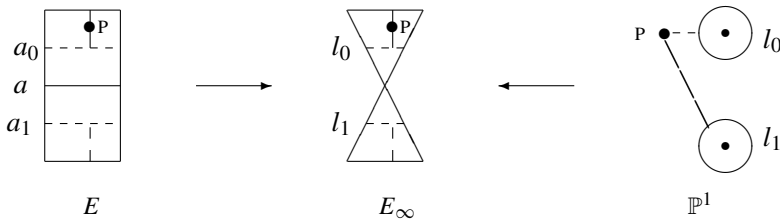


Figure 13. The loops on E, E_∞ and \mathbb{P}^1 .

Since all our loops have to be considered as elements of the respective fundamental groups, we have to choose base points in $\mathbb{P}^1(\mathbb{C}), E_\infty$, and E . Since l_0 and l_1 may not pass through the origin (resp. the node), a_0 and a_1 may not intersect a . Therefore, if we choose the base point P as in the figure, a_0 is homotopic to a , but a_1 is homotopic to bab^{-1} .

Finally we have to lift a_0 and a_1 to X_∞ resp. $\bigcup_{i=1}^n X_i$ and write down the order in which we traverse the squares if we follow the irreducible components of these lifts. Thereby clearly the lift of a_0 induces σ_a , whereas the lift of a_1 induces $\sigma_b \sigma_a \sigma_b^{-1}$. \square

The Belyi map $f_\infty: \bigcup_{i=1}^n X_i \rightarrow \mathbb{P}^1(\mathbb{C})$ can also be described directly in a very explicit way: In the above proof, $E_\infty - \{\text{node}\}$ is obtained by gluing U_0 and U_1 along their boundaries (with opposite orientation). We may assume that the distinguished point ∞ , over which the origami map p is ramified, lies on this boundary, and that, for the given Euclidean structure, the boundary has length 1. In this way we have described an isomorphism between $E_\infty - \{\text{node}\}$ and $\ddot{\mathbb{P}}$.

Now let C_1, \dots, C_n be the horizontal cylinders of the origami surface X . Contracting the centre line c_i of C_i to a point turns $C_i - c_i$ into the union of two punctured disks $U_{0,i}$ and $U_{1,i}$. If C_i consists of d_i squares, the boundary of $U_{0,i}$ and $U_{1,i}$ has length d_i , and is subdivided by the squares into d_i segments of length 1. The Belyi map f_∞ is obtained by mapping each $U_{0,i}$ to U_0 and each $U_{1,i}$ to U_1 in such a way that the lengths are preserved. Thus in standard coordinates, the restriction of f_∞ to $U_{0,i}$ is $z \mapsto z^{d_i}$.

8.3 Examples

8.3.1 The origami $L_{2,2}$. The smallest origami with a surface X of genus > 1 (actually 2) is the one called $L_{2,2}$ in Example 5.2; it is also the smallest one in the family $L_{n,m}$ of L -shaped origamis defined in Section 8.3.2 below. The origami map $p: L_{2,2} \rightarrow E$ is of degree 3 and totally ramified over the point $\infty \in E$ (the vertex of the square). As explained in the previous section, the same holds for the covering $p_\infty: X_\infty \rightarrow E_\infty$ of the degenerate surfaces corresponding to the boundary points in the horizontal direction. As $X = L_{2,2}$ has 2 cylinders in the horizontal direction, X_∞ has 2 singular points which both are mapped by p_∞ to the node of E_∞ . X_∞ is irreducible, and its geometric genus is 0. Thus the normalization of X_∞ is $\mathbb{P}^1(\mathbb{C})$, and the induced map $f_\infty: \mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$ is of degree 3. The two points of the normalization of E_∞ that lie over the node (and which we normalized to be 0 and 1), both have two preimages under f_∞ , one ramified, the other not. Thus we obtain the following dessin for the Belyi map f_∞ :



Figure 14. The dessin for a cusp of $L_{2,2}$.

Since f_∞ is totally ramified over ∞ , we can take it to be a polynomial. If we further normalize it so that 0 is a ramification point, we find that f_∞ is of the form

$$f_\infty(x) = x^2(x - a)$$

for some $a \in \mathbb{C}$. The derivative of f_∞ is

$$f'_\infty(x) = 3x^2 - 2ax = x(3x - 2a),$$

thus the other ramification point of f_∞ in \mathbb{C} is $\frac{2a}{3}$. Since the corresponding critical value is 1, we must have

$$1 = f_\infty\left(\frac{2a}{3}\right) = \frac{4a^2}{9}\left(-\frac{1}{3}a\right) = \left(\frac{a}{3}\right)^3 \cdot (-4)$$

or

$$a = 3 \cdot \sqrt[3]{-\frac{1}{4}}.$$

All three choices of the third root lead to the same dessin, as can be seen from the following observation: The polynomial $f_a(x) = f_\infty(x) = x^2(x - a)$ has its zeroes at 0 and a , and takes the value 1 at $\frac{2}{3}a$ and $-\frac{a}{3}$, as can easily be checked. The cross ratio of these four points is -8 , hence rational. This means that for all possible choices of a , the Belyi map f_a is equivalent to $f_a \circ \sigma_a$, where the Möbius transformation σ_a is determined by

$$\sigma_a(0) = 0, \quad \sigma_a(1) = -\frac{1}{3}a, \quad \sigma_a(\infty) = a,$$

and consequently $\sigma_a(-8) = \frac{2}{3}a$. An easy calculation shows

$$\sigma_a(x) = \frac{ax}{x - 4} \quad \text{and} \quad f_a \circ \sigma_a(x) = -27 \frac{x^2}{(x - 4)^3}.$$

Note that $f_a \circ \sigma_a$ has a triple pole (at 4), a double zero at 0 (and another zero at ∞), and it takes the value 1 with multiplicity 2 at -8 (and a third time at 1).

It was shown in [31] that the origami curve $C(L_{2,2})$ has only one further cusp besides the one just discussed. It corresponds to cylinders in the “diagonal” direction $(1, 1)$. In fact, there is only one cylinder in this direction (of length 3), and by taking this direction to be horizontal, the origami looks like

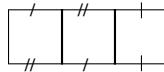


Figure 15. Another view on $L_{2,2}$.

where as usual edges with the same marking are glued.

The corresponding singular surface X_∞ has one irreducible component with one singular point. Its normalization is an elliptic curve E_0 which admits an automorphism of order 3 (induced by the cyclic permutation of the three “upper” and the three “lower” triangles of X_∞). This property uniquely determines E_0 : It is the elliptic curve with Weierstrass equation $y^2 = x^3 - 1$ and j -invariant 0.

The corresponding dessin d'enfants is

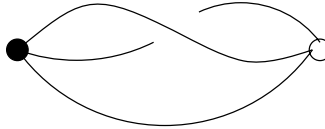


Figure 16. The dessin of a boundary point of $C(L_{2,2})$.

It is the same ribbon graph as G_2 in Example 3.6. The Belyi map f for this dessin is, up to normalization, the quotient map for the automorphism of order 3. If E_0 is given in Weierstrass form as above, this automorphism is the map $(x, y) \mapsto (\zeta_3 x, y)$, where $\zeta_3 = -\frac{1}{2} + \frac{i}{2}\sqrt{3}$ is a primitive third root of unity. Such a quotient map is $(x, y) \mapsto y$. It is easily seen to be totally ramified over $i, -i$ and ∞ . To make the critical values $0, 1$ and ∞ , we have to compose with the linear map $z \mapsto \frac{i}{2}(z - i)$. This shows that our Belyi map is

$$f(x, y) = \frac{i}{2}(y - i).$$

8.3.2 General L -shaped origamis. Denote by $L_{n,m}$ the L -shaped origami with n squares in the horizontal and m squares in the vertical direction:

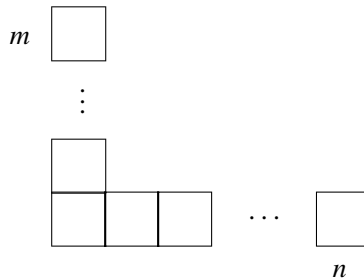


Figure 17. The origami $L_{n,m}$; opposite edges are glued.

These origamis have been studied from several points of view by Hubert and Lelièvre, Schmithüsen, and others. The genus of $L_{n,m}$ is 2, independent of n and m . The index of the Veech group gets larger if n and m increase, and the genus of $C(L_{n,m})$ can be arbitrarily large. Also the number of cusps of $C(L_{n,m})$ grows with n and m .

In this section we only discuss the cusp of $L_{n,m}$ which is obtained by contracting the core lines of the horizontal cylinders. The resulting singular surface X_∞ has $m - 1$ irreducible components: there is one component that contains the cylinder of length n and also the upper half of the top square. All other components consist of the upper half of one square, together with the lower half of the next square. Each such

component is a projective line that intersects two of the other components. Moreover such a component contains a vertex, i.e. a point which is mapped to ∞ by f_∞ (but not ramified). It follows that the Belyi map f_i corresponding to such a component is the identity map $\mathbb{P}^1(\mathbb{C}) \rightarrow \mathbb{P}^1(\mathbb{C})$.

Thus the only interesting irreducible component of X_∞ is the one that contains the “long” horizontal cylinder. For simplicity we only discuss the case where there are no components of the other type, i.e. $m = 2$. In this case we have, as for $L_{2,2}$, exactly 2 singular points on X_∞ . They are both mapped to the node of E_∞ by p_∞ , one unramified, the other one with ramification order n (note that the degree of p_∞ is $n + 1$, the number of squares of $L_{n,2}$). As in the previous subsection, this picture is preserved if we pass to the normalization. Thus $f_\infty^{-1}(0)$ and $f_\infty^{-1}(1)$ both consist of 2 points, one unramified, the other one ramified of order n , and the dessin looks as follows:

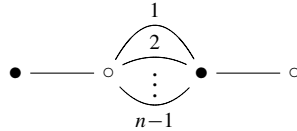


Figure 18. The dessin at the cusp of $L_{n,2}$.

The $n - 1$ cells of the dessin correspond to the fact that $L_{n,2}$ has $n - 1$ different vertices: one of order 3 and $n - 2$ of order 1. Therefore the two vertices of order one of the dessin lie in the same cell.

Note that there is only one dessin of genus 0 with these properties, namely two vertices of order n and two vertices of order one, which are in the same cell. Hence our dessin is completely determined by its Galois invariants. This implies in particular that the moduli field of the dessin is \mathbb{Q} .

It is also possible to determine explicitly the associated Belyi map: To simplify the calculation, we first exhibit a rational function with a zero and a pole of order n , and in addition a simple zero and a simple pole; later we shall change the roles of 1 and ∞ to get the proper Belyi map. So we begin with a rational function of the type

$$f_0(x) = x^n \cdot \frac{x - 1}{x - c}.$$

The condition that f_0 has a further ramification point of order 3 implies that f'_0 has a double zero somewhere. A straightforward calculation shows that this happens if and only if the parameter c has the value

$$c = c_n = \left(\frac{n - 1}{n + 1}\right)^2.$$

The corresponding ramification point is

$$v_n = \sqrt{c_n} = \frac{n - 1}{n + 1}.$$

Since we want the critical value in this point to be 1, we have to replace f_0 by

$$f_1(x) = b_n^{-1} \cdot f_0(x) \quad \text{with } b_n = -\left(\frac{n-1}{n+1}\right)^{n-1} = -c_n^{\frac{n-1}{2}}.$$

Now we interchange 1 and ∞ (keeping 0 fixed); to give the final function a nicer form, we bring the zeroes to 0 and ∞ , and the places where the value 1 is taken to 1 and a fourth point which is determined by the cross ratio of the zeroes and poles of f_1 , i.e. 0, 1, ∞ and c_n ; it turns out to be $d_n = 1 - \frac{1}{c_n}$. Altogether we replace f_1 by

$$f_n = \beta \circ f_1 \circ \sigma, \quad \text{where } \beta(x) = \frac{x}{x-1} \text{ and } \sigma(x) = \frac{x}{x-d_n}.$$

The final result is

$$f_n(x) = \frac{\gamma_n x^n}{\gamma_n x^n - (x-d_n)^n (x-1)} \quad \text{with } \gamma_n = \left(\frac{n+1}{n-1}\right)^{n+1}.$$

By construction, f_n has a triple pole; it turns out to be $p_n = \frac{2n}{n-1}$. Putting in the values of the constants we find e. g.

$$f_2(x) = \frac{-27x^2}{(x-4)^3} \quad \text{and} \quad f_3(x) = \frac{-16x^3}{(x-3)^3(x+1)}.$$

8.3.3 The quaternion origami. Let W be the quaternion origami which was illustrated at the end of Section 8.1 and studied in detail in [16]. It has genus 3, and the origami map $p: W \rightarrow E$ is a normal covering of degree 8 with Galois group Q_8 , the classical quaternion group. Its Veech group is $SL_2(\mathbb{Z})$, which implies that the origami curve $C(W)$ in \bar{M}_3 has only one cusp. As indicated in Figure 12, this cusp corresponds to a stable curve W_∞ with two irreducible components, both nonsingular of genus 1; the components intersect transversely in two points. Both components of W_∞ admit an automorphism of order 4 and are therefore isomorphic to the elliptic curve E_{-1} with Weierstrass equation $y^2 = x^3 - x$. The normalization of W_∞ then consists of two copies of E_{-1} . On each of them, p induces a Belyi map $f: E_{-1} \rightarrow \mathbb{P}^1(\mathbb{C})$ of degree 4, which is totally ramified over the 2 points that map to the node of E_∞ (these are the points of intersection with the other component). Over ∞ we have two points on E_{-1} , both ramified of order 2.

Thus the corresponding dessin d'enfants is the one shown in Figure 19. The Belyi map in this case is a quotient map for the automorphism c of order 4, which acts by $(x, y) \mapsto (-x, iy)$. Such a quotient map is $(x, y) \mapsto x^2$; it is ramified at the four 2-torsion points of E_{-1} : two of them are the fixed points of c , the other two are exchanged by c . The critical values are 0, 1 and ∞ , but not in the right order: To have the values 0 and 1 at the fixed points of c we have to change the roles of 1 and ∞ in $\mathbb{P}^1(\mathbb{C})$, and then obtain the Belyi map $f: E_{-1} \rightarrow \mathbb{P}^1(\mathbb{C})$ as

$$f(x, y) = \frac{x^2}{x^2 - 1}$$

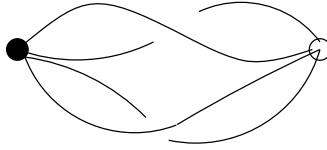


Figure 19. The dessin of the boundary point of $C(W)$.

or, in homogeneous coordinates,

$$f(x : y : z) = (x^2 : x^2 - z^2) = (y^2 + xz : y^2).$$

8.3.4 The characteristic origami of order 108. Our last example in this section is the origami B with 108 squares which corresponds to a normal origami covering $p: B \rightarrow E$ with Galois group

$$G = \{(\sigma_1, \sigma_2, \sigma_3) \in S_3 \times S_3 \times S_3 \mid \prod_{i=1}^3 \text{sign}(\sigma_i) = 1\}.$$

As for W in the previous section, the Veech group of B is $SL_2(\mathbb{Z})$. It was the first normal origami of genus > 1 that was discovered to have the full group $SL_2(\mathbb{Z})$ as Veech group. It is studied in detail in [2] and also (more shortly) in [14].

The genus of B is 37; the horizontal cylinders all have length 6. Contracting their core lines gives a stable curve B_∞ with 6 irreducible components, each nonsingular of genus 4. Each of the irreducible components intersects three others in two points each. The intersection graph of B_∞ is shown in Figure 20.

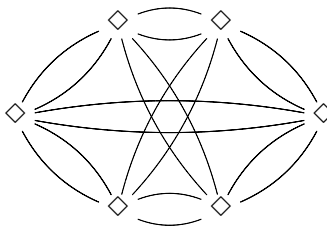


Figure 20. The intersection graph of the 108 origami.

Since the group G acts transitively on the irreducible components of B_∞ , they are all isomorphic. Let us denote by C one of them. The stabilizer of C in G is a subgroup H of order 18. The quotient map $f: C \rightarrow C/H = \mathbb{P}^1(\mathbb{C})$ is the Belyi map corresponding to this (unique) cusp of the origami curve $C(B)$.

The ramification of f over ∞ comes from the fixed points of the elements of H . There are two different subgroups of order 3 that have 3 fixed points each, and no other fixed points. The other ramification points lie over the two points in $\mathbb{P}^1(\mathbb{C})$, that

are mapped to the node of E_∞ . Hence they are the 6 points where C meets other components, and each of them has ramification order 6. These considerations show by the way that the genus of C is in fact 4, since by Riemann–Hurwitz we have

$$2g - 2 = 18 \cdot (-2) + 6 \cdot (6 - 1) + 6 \cdot (3 - 1) = -36 + 42 = 6.$$

On the original origami B , the component C corresponds to 36 half squares. The 18 upper halves among them are the lower halves of three horizontal cylinders, and in the same way, the 18 lower halves contributing to C are the upper halves of three other cylinders. The core lines of these six cylinders give the six ramification points of f that lie over 0 and 1. The precise picture looks as follows:

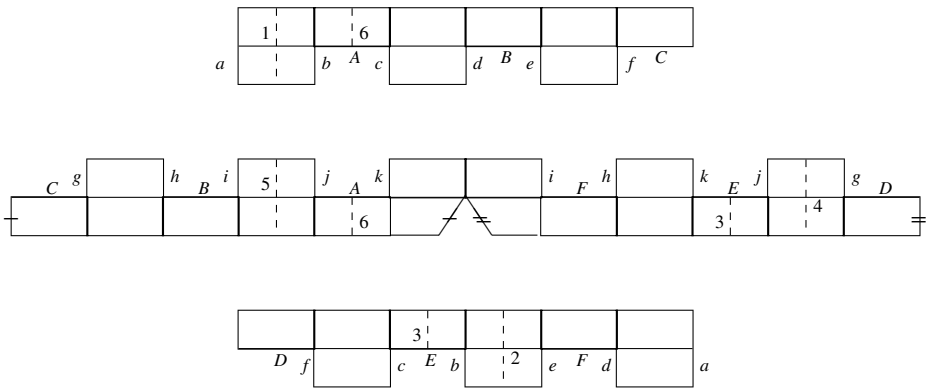


Figure 21. The 36 half squares of which a component of the curve B_∞ is composed. Vertical gluings are indicated by capital letters, horizontal gluings by small letters. The dashed lines are, in the order 1, . . . , 6, the boundary of one of the six cells of the dessin.

In each row of the figure, the upper horizontal edges give one vertex of the dessin (corresponding to a point lying over 0). The lower edges of the second row give two vertices over 1, and the third vertex comes from the six lower edges in the first and the last row.

The 18 edges of the dessin are vertical centre lines of the squares; some of them are shown in the figure. Each of the three “upper” vertices is connected to two of the “lower” vertices by three edges each, and not connected to other vertices. The order in which the edges leave the vertices is determined by the horizontal gluing of the squares.

One way of describing the resulting dessin d’enfants is to consider its cells and their gluing. Since $f^{-1}(\infty)$ consists of 6 points of ramification order 6, our dessin has 6 cells, and each of them is a hexagon. In the origami, these hexagons are found as follows: begin with an arbitrary edge (i.e. a vertical centre line of a square); at its end point, go one square to the right and continue with the edge that starts at its centre.

Go on like this until you reach the first edge again. The figure shows one example for this. Note that the 6 vertices of this hexagon are all different. By symmetry this holds for all 6 hexagons. The way how these hexagons have to be glued can be read off from the origami. Thus finally we find the following dessin, in which, as in the pictures of origamis, edges with the same label have to be glued:

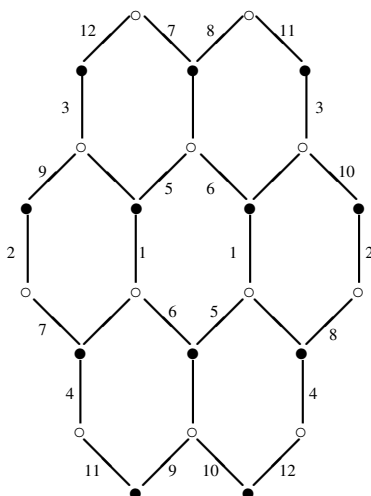


Figure 22. The dessin d'enfants to the cusp of the 108 origami. The surface consists of the six outer hexagons, with edges glued as indicated by the labels.

References

- [1] F. Armknecht, Über die Aktion der absoluten Galoisgruppe auf hyperelliptischen Kinderzeichnungen. Diplomarbeit, Universität Karlsruhe, 2001. [782](#)
- [2] O. Bauer, Stabile Reduktion und Origamis. Diplomarbeit, Universität Karlsruhe, 2005. [797](#), [805](#)
- [3] G. V. Belyi, On Galois extensions of a maximal cyclotomic field. *Izv. Akad. Nauk SSSR Ser. Mat.* 43 (1979), 267–276. [771](#)
- [4] V. Drinfel'd, On quasitriangular quasi-Hopf algebras and a group closely connected to $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. *Leningrad Math. J.* 2 (1991), 829–860. [787](#)
- [5] C. J. Earle and F. P. Gardiner, Teichmüller disks and Veech's \mathcal{F} -structures. *Contemp. Math.* 201 (1997), 165–189. [791](#)
- [6] M. Emsalem and P. Lochak, The action of the absolute Galois group on the moduli space of spheres with four marked points. In *The Grothendieck Theory of Dessins d'Enfants* (L. Schneps, ed.), London Math. Soc. Lecture Note Ser. 200, Cambridge University Press, Cambridge 1994, 307–321. [783](#), [788](#)

- [7] O. Forster, *Lectures on Riemann Surfaces*. Grad. Texts in Math. 81, Springer-Verlag, New York 1981. [769](#)
- [8] S. Gemmrich, On the Grothendieck-Teichmüller group and the Teichmüller tower of mapping class groups. Diplomarbeit, Universität Karlsruhe, 2004. [788](#)
- [9] A. Grothendieck, Esquisse d'un programme. In *Geometric Galois Actions*, 1. Around Grothendieck's "Esquisse d'un programme" (P. Lochak, L. Schneps, eds.), London Math. Soc. Lecture Note Ser. 242, Cambridge University Press, Cambridge 1997, 5–48. [772](#)
- [10] H. Hammer and F. Herrlich, A remark on the moduli field of a curve, *Arch. Math.* 81 (2001), 5–10. [781](#)
- [11] J. Harris and I. Morrison, *Moduli of Curves*. Grad. Texts in Math. 187, Springer-Verlag, New York 1998. [797](#)
- [12] R. Hartshorne, *Algebraic Geometry*. Grad. Texts in Math. 52, Springer-Verlag, New York 1977. [769](#), [770](#)
- [13] W. Harvey, Teichmüller spaces, triangle groups and Grothendieck dessins. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 249–292. [791](#)
- [14] F. Herrlich, Teichmüller curves defined by characteristic origamis. *Contemp. Math.* 397 (2006), 133–144. [805](#)
- [15] F. Herrlich and G. Schmithüsen, On the boundary of Teichmüller disks in Teichmüller and in Schottky space. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 293–349. [791](#), [797](#)
- [16] F. Herrlich and G. Schmithüsen, An extraordinary origami curve. *Math. Nachrichten* 281 (2) (2008), 219–237. [798](#), [804](#)
- [17] Y. Ihara, On the embedding of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ into \widehat{GT} . In *The Grothendieck Theory of Dessins d'Enfants* (L. Schneps, ed.), London Math. Soc. Lecture Note Ser. 200, Cambridge University Press, Cambridge 1994, 289–306. [783](#), [787](#), [788](#)
- [18] B. Köck, Belyi's theorem revisited. *Beitr. Algebra Geom.* 45 (1) (2004), 253–265. [772](#), [781](#)
- [19] S. K. Lando and A. K. Zvonkin, Graphs on surfaces and their applications. In *Low-Dimensional Topology*, II, Encyclopaedia Math. Sci. 141, Springer-Verlag Berlin 2004. [768](#)
- [20] P. Lochak, On arithmetic curves in the moduli spaces of curves. *J. Inst. Math. Jussieu* 4 (3) (2005), 443–508. [768](#), [788](#), [792](#), [793](#), [794](#), [798](#)
- [21] P. Lochak and L. Schneps, The Grothendieck-Teichmüller group and automorphisms of braid groups. In *The Grothendieck Theory of Dessins d'Enfants* (L. Schneps, ed.), London Math. Soc. Lecture Note Ser. 200, Cambridge University Press, Cambridge 1994, 323–358. [788](#)
- [22] H. Masur, On a class of geodesics in Teichmüller space. *Ann. of Math.* 102 (1975), 205–221. [797](#)
- [23] C. T. McMullen, Billiards and Teichmüller curves on Hilbert modular surfaces. *J. Am. Math. Soc.* 16 (4) (2003) 857–885. [791](#)
- [24] R. Miranda, *Algebraic Curves and Riemann Surfaces*. Grad. Stud. Math. 5, Amer. Math. Soc., Providence, RI, 1995. [769](#)

- [25] M. Möller, Teichmüller curves, Galois actions and \widehat{GT} -relations. *Math. Nachr.* 278 (9) (2005), 1061–1077. [793](#)
- [26] M. Möller, Affine groups of flat surfaces. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 369–387. [791](#)
- [27] J. P. Murre, *Lectures on an Introduction to Grothendieck's Theory of the Fundamental Group*. Tata Institute, Bombay 1967. [786](#)
- [28] E. Reyssat, *Quelques Aspects des Surfaces de Riemann*. Progr. Math. 77, Birkhäuser, Boston 1989. [769](#)
- [29] G. Schmithüsen, Die Aktion von $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ auf der algebraischen Fundamentalgruppe von \mathbb{P}^1 ohne drei Punkte. Diplomarbeit, Universität Karlsruhe 2001. [783](#), [787](#), [788](#)
- [30] G. Schmithüsen, An algorithm for finding the Veech group of an origami. *Experiment. Math.* 13 (4) (2004), 459–472. [792](#), [794](#)
- [31] G. Schmithüsen, Veech Groups of Origamis. PhD thesis, Universität Karlsruhe 2005. [768](#), [791](#), [792](#), [801](#)
- [32] G. Schmithüsen, Origamis with non congruence Veech groups. In *Proceedings of 34th Symposium on Transformation Groups* (Yokohama 2006), Wing Co., Wakayama 2007, 31–55. [789](#), [795](#)
- [33] L. Schneps, ed., *The Grothendieck Theory of Dessins d'Enfants*. London Math. Soc. Lecture Note Ser. 200, Cambridge University Press, Cambridge 1994. [768](#), [772](#)
- [34] L. Schneps, Dessins d'enfants on the Riemann sphere. In *The Grothendieck Theory of Dessins d'Enfants*. London Math. Soc. Lecture Note Ser. 200, Cambridge University Press, Cambridge 1994, 47–77. [768](#), [782](#), [783](#)
- [35] L. Schneps, The Grothendieck-Teichmüller group \widehat{GT} : a survey. In *Geometric Galois Actions, 1. Around Grothendieck's "Esquisse d'un programme"* (P. Lochak, L. Schneps, eds.), London Math. Soc. Lecture Note Ser. 242, Cambridge University Press, Cambridge 1997, 183–203. [788](#)
- [36] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)* 19 (2) (1988), 417–431. [791](#)
- [37] W. A. Veech, Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. Math.* 97 (3) (1989), 553–583. [791](#)
- [38] S. Wewers, Constructing Hurwitz spaces. PhD thesis, Universität Essen 1998. [793](#)
- [39] J. Wolfart, The 'obvious' part of Belyi's theorem and Riemann surfaces with many automorphisms. In *Geometric Galois Actions, 1. Around Grothendieck's "Esquisse d'un programme"* (P. Lochak, L. Schneps, eds.), London Math. Soc. Lecture Note Ser. 242, Cambridge University Press, Cambridge 1997, 97–112. [772](#), [781](#)
- [40] J. Wolfart, Kinderzeichnungen und Uniformisierung. Unpublished manuscript, 2001. [768](#), [781](#)
- [41] L. Zapponi, Fleurs, arbres et cellules: un invariant galoisien pour une famille d'arbres. *Compositio Math.* 122 (2000), 113–133. [783](#)
- [42] L. Zapponi, Galois action on diameter four trees. Unpublished manuscript, 2000. [782](#)

The Teichmüller theory of the solenoid

Dragomir Šarić

Contents

1	Introduction	811
2	The compact solenoid	817
	2.1 The profinite completion	819
	2.2 The G -tagged compact solenoid	819
3	Complex structures and hyperbolic metrics on the compact solenoid	820
4	The G -tagged non-compact solenoid	821
5	The Teichmüller space of the compact solenoid	823
	5.1 The universal coverings of complex compact solenoids	824
	5.2 Beltrami coefficients and holomorphic quadratic differentials on the compact solenoid	825
	5.3 The restriction map π_l	826
	5.4 The complex Banach manifold structure on $\mathcal{T}(\mathcal{S}_G)$	828
6	The Reich–Strebel inequality	828
7	The Teichmüller-type extremal maps	833
8	The modular group of the compact solenoid	838
9	The Teichmüller space of the non-compact solenoid	840
10	The decorated Teichmüller space of the non-compact solenoid	842
11	A presentation for the modular group of the non-compact solenoid	846
12	Elements of $\text{Mod}(\mathcal{S}_{\text{nc}})$ with small non-zero dilatations	853
13	Some open problems	854
	References	855

1 Introduction

The compact solenoid \mathcal{S} (also called the *universal hyperbolic solenoid*) was introduced by Sullivan [44] as a universal object in the category of all pointed, unbranched, finite-sheeted coverings of a (base) closed surface of genus at least two (\mathcal{S} can be thought of as a “universal closed surface”). The compact solenoid \mathcal{S} is independent (as a topological space) of the choice of the base surface in the definition (as long as the genus is at least two).

More explicitly, the compact solenoid \mathcal{S} is the inverse limit of the system of all pointed, unbranched, finite-sheeted coverings of a closed surface of genus at least two. Alternatively, one can consider a tower of coverings in place of all coverings. Then the inverse limits of any two infinite towers of pointed, unbranched, finite-sheeted coverings are homeomorphic as long as the intersection of all fundamental groups in each tower (when considered as subgroups of the fundamental group of the base surface) is trivial. A particularly interesting tower is obtained by defining the n -th covering in the tower to have fundamental group equal to the intersection of all subgroups of index at most n .

Another description of the compact solenoid \mathcal{S} is that it is a principal fiber bundle over a closed surface of genus at least two. The fibers are homeomorphic to a Cantor set with a topological group structure such that the base surface fundamental group is realized as a dense subgroup of the fiber group. If the base surface is given a fixed hyperbolic metric then the compact solenoid is explicitly realized as follows (see Section 2.2 or [38]). Let G be a co-compact subgroup of the Möbius group acting on the unit disk \mathbb{D} which uniformizes the base surface and let $\omega \subset \mathbb{D}$ be a fundamental polygon for G . Then there exists a Cantor set \hat{G} with the structure of a topological group and an injective homomorphism $G \hookrightarrow \hat{G}$ whose image is dense in \hat{G} (the group \hat{G} is defined in Section 2 and later in the Introduction). The compact solenoid \mathcal{S} is the quotient of $\omega \times \mathbb{D}$ by the action of finitely many elements of G which pairwise identify the sides of ω . The action of these elements on $\omega \times \mathbb{D}$ is given by the side pairing Möbius action on the ω -factor and by the right multiplication in the group \hat{G} using the identification of G with its image in \hat{G} (on the \hat{G} -factor). Thus a neighborhood of a point in the compact solenoid \mathcal{S} is given by the product of the open fundamental polygon $\hat{\omega}$ and a Cantor set \hat{G} . The boundary sides of ω are identified with the corresponding boundary sides of ω but on different “levels”, i.e., the second factors (in \hat{G}) are different. The path component of a single $\omega \times \{t\}$, for a fixed $t \in \hat{G}$, in the quotient approaches any point of \mathcal{S} arbitrary close (because the image of G in \hat{G} is dense).

The compact solenoid \mathcal{S} is locally homeomorphic to a 2-disk times a Cantor set; each leaf (i.e., a path component) of \mathcal{S} is dense in \mathcal{S} and it is homeomorphic to the unit disk; a distinguished leaf of \mathcal{S} is called the *baseleaf*. Moreover, \mathcal{S} has a unique transverse measure, i.e., a holonomy invariant measure on each transverse set, which is induced by the Haar measure on the fiber group. The holonomy map is given by the action of the base surface group on the fiber group via its natural identification as a subgroup of the fiber.

We give some motivation for the study of the compact solenoid \mathcal{S} . The Ehrenpreis conjecture [13] states that for any two closed non-conformal Riemann surfaces of the same genus greater than 1 and for any $\varepsilon > 0$ there exist two finite-sheeted, unbranched, conformal covers that are $(1 + \varepsilon)$ -quasiconformal. Since the universal cover of both surfaces is the unit disk, this question can be interpreted as to whether finite covers approximate the universal cover. Instead of considering two Riemann surfaces at the same time and finding their appropriate finite covers, it is (at least) conceptually more

appropriate to have all Riemann surfaces in a single space. The space is the union of properly normalized embeddings in the universal Teichmüller space $\mathcal{T}(\mathbb{D})$ of the Teichmüller spaces of all closed Riemann surfaces covering the base surface. The group of all isomorphisms between finite index subgroups of the fundamental group (called the *commensurator* of the surface group) acts naturally on the above union and the Ehrenpreis conjecture is equivalent to the statement that the action has dense orbits in the union [32]. It is natural to take the closure of the union in the universal Teichmüller space to obtain a Banach manifold and the action of the commensurator extends naturally to the closure [32].

Sullivan noticed the connection with the compact solenoid \mathcal{S} : instead of considering Riemann surfaces of different genera as points in a single space $\mathcal{T}(\mathbb{D})$ as well as their limit points in $\mathcal{T}(\mathbb{D})$, it is natural to form a single topological space (the compact solenoid \mathcal{S}) using all finite coverings of a base surface and to express Riemann surfaces of different genera as well as their limit points in $\mathcal{T}(\mathbb{D})$ as different complex structures on the compact solenoid \mathcal{S} . Then the Ehrenpreis conjecture is equivalent to the statement that the action of the commensurator group $\text{Comm}(\pi_1(S))$ of the base surface group $\pi_1(S)$ on the Teichmüller space $\mathcal{T}(\mathcal{S})$ of the compact solenoid \mathcal{S} has dense orbits [5], [32].

C. Odden [38] showed that the modular group $\text{Mod}(\mathcal{S})$ of the compact solenoid \mathcal{S} is isomorphic to the commensurator group $\text{Comm}(\pi_1(\mathcal{S}))$ of the fundamental group $\pi_1(S)$ of the base surface S . This is in an analogy with the classical statement that the group of outer isomorphisms of the closed surface group is the mapping class group of the surface ([8], [33], [2]). Thus, the compact solenoid \mathcal{S} is a natural space for which the commensurator group is its modular group. From the group theoretic point of view, the modular group $\text{Mod}(\mathcal{S}) \equiv \text{Comm}(\pi_1(S))$ describes the “hidden symmetries” of the surface group [26].

We are also interested in studying complex structures on the compact solenoid \mathcal{S} from the viewpoint of the complex analytic theory of Teichmüller spaces. The Teichmüller space $\mathcal{T}(\mathcal{S})$ is a first example of a Teichmüller space which is an infinite-dimensional but separable complex Banach manifold. Recall that Teichmüller spaces of Riemann surfaces are either finite dimensional complex manifolds or infinite-dimensional non-separable Banach manifolds. It appears that the complex analytic and the metric structure of $\mathcal{T}(\mathcal{S})$ is quite different from the Teichmüller spaces of geometrically finite as well as geometrically infinite Riemann surfaces. Inverse limit spaces commonly appear in dynamics ([44], [45], [24], [31]) and the compact solenoid is a first non-trivial example of an inverse limit with interesting Teichmüller space.

The non-compact solenoid \mathcal{S}_{nc} (also called the *punctured solenoid*) is the inverse limit of the system of all pointed, unbranched, finite-sheeted coverings of a base punctured surface with negative Euler characteristic [37]. The covering surfaces are punctured with the covering maps sending punctures to punctures. If we fill in the punctures, the covering maps become finitely branched at the punctures. Therefore, the branching in the covering tower is restricted by allowing it only over the punctures of the base surface (unlike for towers of rational maps where branching appears to be

“wild” [24]). The inverse limit \mathcal{S}_{nc} is a non-compact space because we do not include the backward orbits of punctures in the space, each leaf is homeomorphic to the unit disk \mathbb{D} and the ends of each leaf are universal covers of neighborhoods of punctures on surfaces, i.e., the ends are horoballs with the induced non-standard topology from \mathcal{S}_{nc} . The analog of the Ehrenpreis conjecture for punctured surfaces asks whether every two finite Riemann surfaces have finite covers which are $(1 + \varepsilon)$ -quasiconformal. This is equivalent to the statement that the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ has dense orbits in the Teichmüller space $\mathcal{T}(\mathcal{S}_{\text{nc}})$ of the non-compact solenoid \mathcal{S}_{nc} . (The Ehrenpreis conjecture has recently been proved for the punctured surfaces [21] and the normalized Weil–Petersson metric, which is an important result. However, the conjecture is still open for the Teichmüller metric.) In analogy to the compact case, the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ is isomorphic to a subgroup of the commensurator of the base punctured surface group which preserves the peripheral elements [37]. The existence of ends of leaves allows for a combinatorial decomposition of the (decorated) Teichmüller space of \mathcal{S}_{nc} (see [37]) which gives a better understanding of the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ of the non-compact solenoid than of the modular group of the compact solenoid \mathcal{S} .

In this chapter, we survey results on the Teichmüller space $\mathcal{T}(\mathcal{S})$ of the compact solenoid \mathcal{S} regarding its metric structure with respect to the Teichmüller metric and its complex structure. We also survey results on the modular group $\text{Mod}(\mathcal{S})$ of the compact solenoid \mathcal{S} and the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ of the non-compact solenoid \mathcal{S}_{nc} . We give more details below.

In Section 2 we give different equivalent definitions of the compact solenoid \mathcal{S} . In addition to defining \mathcal{S} as an inverse limit space, we define it as a principal fiber bundle space as follows. For a fixed Fuchsian group G uniformizing a closed Riemann surface of genus at least two, we define a profinite group completion \hat{G} of G with respect to the profinite metric. The profinite metric on G is defined by [38]

$$d_{\text{pf}}(A, B) = e^{-\frac{1}{n}},$$

where AB^{-1} is an element of all subgroups of G of index at most n , and there exists a subgroup of G of index $n + 1$ which does not contain AB^{-1} . Then the G -tagged compact solenoid \mathcal{S}_G is the quotient of $\mathbb{D} \times \hat{G}$ by an action of G , where G acts by Möbius maps on the unit disk component \mathbb{D} and shifts the levels by acting on \hat{G} by right translations on the group. The G -tagged solenoid \mathcal{S}_G is homeomorphic to the compact solenoid \mathcal{S} . The natural map from $\mathbb{D} \times \hat{G}$ to the Riemann surface \mathbb{D}/G obtained by “forgetting” the second coordinate and by mapping the first coordinate to its orbit under G projects to a map from the quotient $(\mathbb{D} \times \hat{G})/G = \mathcal{S}_G$ onto \mathbb{D}/G ; the fibers of the map are homeomorphic to \hat{G} . Thus \mathcal{S}_G is a \hat{G} -fiber bundle over \mathbb{D}/G .

A complex structure on the compact solenoid \mathcal{S} is by definition an atlas whose transition maps when restricted to local leaves are holomorphic and are continuous for the transverse variations of local leaves [44]. Candel [7] proved a uniformization theorem for laminations which, in particular, implies that each transversely continuous conformal structure on \mathcal{S} has a unique transversely continuous leafwise hyperbolic metric representative. The compact solenoid \mathcal{S} is a fiber bundle over a closed surface

such that the restriction to any leaf of the fiber map is the universal covering of the base surface. Thus, any complex structure on the base surface lifts to a complex structure on the compact solenoid \mathcal{S} . Each lifted complex structure has a sub-atlas whose transition maps are constant in the (Cantor set) transverse directions. Nag and Sullivan [32] showed that every such complex structure is obtained by forming a G -tagged solenoid \mathcal{S}_G , for some uniformizing Fuchsian group G of a closed Riemann surface (see Section 3 for more details). The G -tagged punctured solenoid is formed similarly by using the punctured Riemann surface uniformizing Fuchsian group G (see Section 4).

The Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the compact solenoid \mathcal{S}_G consists of all marked complex solenoids $f: \mathcal{S}_G \rightarrow \mathcal{X}$ up to post-composition by conformal maps and up to homotopy, where G is fixed and f is a differentiable, quasiconformal map (see Definitions 5.1 and 5.3). (Equivalently, the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ is a quotient of the space of smooth Beltrami coefficients on \mathcal{S}_G continuous in the transverse directions.) The Teichmüller distance of $[f] \in \mathcal{T}(\mathcal{S}_G)$ to the basepoint $[\text{id}] \in \mathcal{T}(\mathcal{S}_G)$ is the infimum of the logarithm of the quasiconformal constants of maps in the homotopy class $[f]$ of the map f . The Teichmüller metric is not degenerate, namely $\mathcal{T}(\mathcal{S}_G)$ is a Hausdorff space (see [44]; see Section 5 for an alternative proof).

The restrictions of the pull-backs of complex structures on the marked solenoids $f: \mathcal{S}_G \rightarrow \mathcal{X}$ to the baseleaf l of \mathcal{S}_G defines a map $\pi_l: \mathcal{T}(\mathcal{S}_G) \rightarrow \mathcal{T}(\mathbb{D})$. This map is a homeomorphism onto its image (a proof is sketched in [44]; see Section 5.2 for an alternative proof). In fact, a consequence of Theorem 7.1 and McMullen's solution [30] to Kra's theta conjecture is that π_l is a bi-Lipschitz map onto its image with constant $1/3$.

The study of the Teichmüller metric on $\mathcal{T}(\mathcal{S}_G)$ starts with the Reich–Strebel inequality (see [42] and Section 6) which estimates the (complex) distortion (i.e., the Beltrami coefficient) of a quasiconformal self-map of the compact solenoid \mathcal{S}_G which is homotopic to the identity in terms of the leafwise Euclidean structures given by the restrictions of holomorphic quadratic differentials on the leaves of \mathcal{S}_G . The Reich–Strebel inequality is a non-trivial generalization of Grötzsch's length-area method for determining extremal maps between rectangles. In this chapter we give a different proof of the Reich–Strebel inequality from the proof in [42] (see Theorem 6.1 and its proof).

The consequences of the Reich–Strebel inequality give a better understanding of the Teichmüller metric on $\mathcal{T}(\mathcal{S}_G)$. In Section 6, we summarize consequences related to the infinitesimal structure of $\mathcal{T}(\mathcal{S}_G)$ from [42]. In particular, a Beltrami differential is tangent to a trivial path of Beltrami coefficients (i.e., it represents a trivial infinitesimal deformation) if and only if it is zero when paired with all holomorphic quadratic differentials on \mathcal{S}_G (see Theorem 6.2). In Section 7, we analyze extremal maps in a given homotopy (Teichmüller) class. A consequence of the Reich–Strebel inequality is that Teichmüller-type maps (i.e., vertical stretch maps in the natural parameter of a holomorphic quadratic differential on \mathcal{S}_G) are extremal in their corresponding homotopy (Teichmüller) classes (see Theorem 7.1). Moreover, the natural inclusion

map from the Teichmüller space of a closed surface into $\mathcal{T}(\mathcal{S}_G)$ obtained by lifting a complex structure on the surface to \mathcal{S}_G is an isometry (see Corollary 7.3).

We also give an account of the question of the existence of Teichmüller-type extremal maps in a given Teichmüller class considered in a joint work of the author with A. Epstein and V. Markovic [14]. The results on extremal maps for Riemann surfaces fall into two cases; either every point in Teichmüller space has a Teichmüller-type representative for closed and finite punctured surfaces – Teichmüller’s theorem, or an open, dense subset of Teichmüller space has Teichmüller-type representatives for geometrically infinite surfaces [22]. Therefore, in both cases, a large subset of the Teichmüller space has Teichmüller-type representatives. For the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the compact solenoid \mathcal{S}_G the situation is quite different. In fact, a generic point in $\mathcal{T}(\mathcal{S}_G)$ does not have Teichmüller-type representatives, i.e., only a set of the first kind in $\mathcal{T}(\mathcal{S}_G)$ in the sense of Baire has Teichmüller-type representatives (see [14], or Theorem 7.4 together with a brief account of the proof.) We also give a necessary condition for a point in $\mathcal{T}(\mathcal{S}_G)$ to have a Teichmüller-type representative (see [14], or Corollary 7.5).

In Section 8, we survey basic results on the modular group (see [38], [28]). The modular group $\text{Mod}(\mathcal{S}_G)$ is isomorphic to the commensurator group of the base surface group (see [38] or Theorem 8.3). In a joint work with V. Markovic, we established that there exist orbits of $\text{Mod}(\mathcal{S}_G)$ in $\mathcal{T}(\mathcal{S}_G)$ with accumulation points [28]; and that finite subgroups of $\text{Mod}(\mathcal{S}_G)$ are cyclic and mapping class like (i.e., they are lifts of self-maps of closed surfaces) [28].

In Section 9, we give a quasiconformal definition of the Teichmüller space $\mathcal{T}(\mathcal{S}_{\text{nc}})$ of the non-compact solenoid \mathcal{S}_{nc} and an equivalent representation-theoretic definition from our joint work with R. Penner (see [37]). In Section 10, we define the decorated Teichmüller space $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ of the non-compact solenoid \mathcal{S}_{nc} and give its parametrization in terms of lambda lengths (see our work with R. Penner [37] or Theorem 10.3). We also describe a convex hull construction for decorations on the punctured solenoid and show that a dense, open subset of $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ is combinatorially interesting (see [37] or Theorem 10.6 for the punctured solenoid; for punctured surfaces see [15], [35]; see [34] for the universal Teichmüller space; see [20] for a related construction for punctured surfaces).

In Section 11, we give a generating set for $\text{Mod}(\mathcal{S}_{\text{nc}})$ in terms of Whitehead homeomorphisms and $\text{PSL}_2(\mathbb{Z})$ (see our work with R. Penner [37] or Theorems 11.3, 11.4 and 11.5). Moreover, we define a natural triangulation 2-complex, show that it is connected and simply connected, and show that the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ acts cellularly on it (see our joint work with S. Bonnot and R. Penner [6] or Theorem 11.6). Using the triangulation complex, we give a presentation for $\text{Mod}(\mathcal{S}_{\text{nc}})$ (see [6] or Theorem 11.7).

Acknowledgement. Most of the recent results in this chapter were obtained in various collaborations of the author with S. Bonnot, A. Epstein, V. Markovic and R. Penner. I am indebted to them for the insights and efforts in our joint explorations. I am grateful to F. Bonahon, F. Gardiner, W. Goldman, L. Keen, M. Lyubich, J. Millson, J. Milnor,

D. Sullivan and S. Wolpert for the various discussion we had regarding the surveyed material. This chapter was written while I was visiting the Department of Mathematics of the University of Maryland. I am grateful to the department for its warm hospitality during my visit.

2 The compact solenoid

In this section we give two equivalent definitions of the compact solenoid \mathcal{S} which is usually called the universal hyperbolic solenoid [44], [32], [38].

Let (S_0, x_0) be a fixed closed surface of genus at least two with basepoint x_0 . Consider all finite-degree, unbranched, pointed covers $\pi_i: (S_i, x_i) \rightarrow (S_0, x_0)$ up to isomorphisms of covers. The family of such covers has a natural partial ordering “ \leq ” defined by

$$(\pi_i, S_i, x_i) \leq (\pi_j, S_j, x_j)$$

if there exists a pointed, unbranched, finite-degree cover $\pi_{i,j}: (S_j, x_j) \rightarrow (S_i, x_i)$ such that

$$\pi_j = \pi_i \circ \pi_{i,j}.$$

Given two arbitrary covers $\pi_i: (S_i, x_i) \rightarrow (S_0, x_0)$ and $\pi_j: (S_j, x_j) \rightarrow (S_0, x_0)$ from the above family, there exists a third cover $\pi_k: (S_k, x_k) \rightarrow (S_0, x_0)$ such that $(\pi_i, S_i, x_i), (\pi_j, S_j, x_j) \leq (\pi_k, S_k, x_k)$. Namely, the family of all covers $\pi_i: (S_i, x_i) \rightarrow (S_0, x_0)$ is inverse directed; thus the inverse limit of the family is well defined. Sullivan [44] introduced the *compact solenoid* \mathcal{S} by

$$\mathcal{S} = \varprojlim (S_i, x_i).$$

By definition, $\mathcal{S} \subset \prod_{i \in I} S_i$, where I is the index set of coverings, consists of all $y = (y_i)_{i \in I} \in \prod_{i \in I} S_i$ such that whenever $(S_i, x_i) \leq (S_j, x_j)$ then $\pi_{i,j}(y_j) = y_i$. The product space $\prod_{i \in I} S_i$ is compact in the Tychonov topology because each S_i is compact. The subset \mathcal{S} is closed in $\prod_{i \in I} S_i$ and therefore it is also a compact space.

The compact solenoid \mathcal{S} is universal in the sense that it does not depend on the base surface S_0 . Namely, if we take the inverse limit of all finite-degree unbranched covers of another closed surface S'_0 of genus at least two then it is homeomorphic to \mathcal{S} . (This follows from the fact that the inverse limit of any given cofinal subsystem of covers is homeomorphic to the inverse limit of the original system of covers. Recall that a subsystem of covers is *cofinal* if any surface in the original system is covered by a surface of the subsystem.) To show that \mathcal{S} is independent of the base surface, it is enough to note that any two such inverse systems of covers have homeomorphic cofinal subsystems because there exist two surfaces in these two systems that are homeomorphic.

The universal property of the compact solenoid enables us to consider a tower of covers of a closed surface of genus at least two instead of all finite covers (as long as the

intersection of all fundamental groups in the tower when identified via pointed covers with subgroups of the base surface is the trivial group; this is required for a tower to be cofinal by the residual finiteness of the base surface group). For example, we can consider the system of covers given by the tower of covers whose n -th level surface has fundamental group equal to the intersection of all index at most n subgroups of the base surface (see [38]). The choice of the subgroup at the n -th level uniquely (up to isomorphism) determines the pointed cover of the base surface. For each $m > n$ the group at level m is a subgroup of the group at level n . Thus the system of covers is a tower and its inverse limit is homeomorphic to \mathcal{S} .

For convenience, we work with the above tower of covers from now on. Thus we can replace the index set I for the covers by the natural numbers \mathbb{N} , where $\pi_j: S_j \rightarrow S_0$ factors through a cover $\pi_i: S_i \rightarrow S_0$ if and only if $j > i$. Then a point y in \mathcal{S} is given by a backward sequence $y = (y_0, y_1, y_2, \dots)$ with respect to the tower of covers, namely $y_i \in S_i$ and $\pi_{i,i+1}(y_{i+1}) = y_i$ for $i \in \mathbb{N} \cup \{0\}$. A neighborhood of a point in \mathcal{S} is homeomorphic to a (2-disk) \times (Cantor set). To see this, note that by the definition of the Tychonov topology a neighborhood of a point $y = (y_0, y_1, y_2, \dots)$ in $\prod_{i \in \mathbb{N}} S_i$ is the set $V(y)$ consisting of all $z = (z_0, z_1, z_2, \dots)$ such that each z_i is in a small ball $U_i(y_i)$ with center $y_i \in S_i$ for all $i < i_0$, with $i_0 \in \mathbb{N}$ fixed, and the rest of the coordinates of z are arbitrary. If $y \in \mathcal{S}$ then a neighborhood $V(y) \subset \mathcal{S}$ is given by successively taking a single lift $U_i(y_i) \subset S_i$ of a ball $U_0(y_0) \subset S_0$ for all $i < i_0$, where $i_0 \in \mathbb{N}$ is fixed, and the rest of the coordinates of the points in $V(y)$ belong to all lifts $\pi_i^{-1}(U_0)$ such that π_{i,i_0} maps them into U_{i_0} , for $i \geq i_0$. The lifts to the tower $\{S_i\}_{i \in \mathbb{N}}$ of a ball in S_0 are enumerated by the locally finite tree of all possibilities of lifts from S_i to S_{i+1} for $i \in \mathbb{N}$. The local structure of \mathcal{S} is given by taking a 2-disk for each infinite path (without backtracking) in the tree with the induced product topology, where the 2-disk has the standard topology, and points in two 2-disks for two different infinite paths are close if they are close as points in the 2-disk and if the infinite paths follow the same finite paths for a long time. The set of all infinite paths without backtracking is a Cantor set and we have completely described the local structure of \mathcal{S} .

A path component of \mathcal{S} is called a *leaf*. A *local leaf* is a path component in any local chart of the above form (2-disk) \times (Cantor set), namely a local leaf is a 2-disk. Therefore, a (global) leaf of the compact solenoid \mathcal{S} is a surface. There is a natural projection $\Pi_i: \mathcal{S} \rightarrow S_i$, for $i \in \mathbb{N} \cup \{0\}$, to any surface in the tower of covers given by

$$\Pi_i(y_0, y_1, y_2, \dots) = y_i,$$

for $y = (y_0, y_1, y_2, \dots) \in \mathcal{S}$. Since the intermediate covers $\pi_{i,i+1}$ in the tower are unbranched, the restriction of the projection Π_i to each leaf is an unbranched covering. We claim that each leaf is simply connected. If a leaf of \mathcal{S} is not simply connected, then a closed curve which is not homotopic to a point maps under each Π_i to a curve on S_i which is not homotopic to a point. However, the covers in the tower are chosen so that each closed homotopically non-trivial curve on any surface cannot be lifted to a closed curve in a high enough cover. (If one considers all finite covers in the definition

of \mathcal{S} , this follows because the fundamental group of S_0 is residually finite.) Thus each leaf is a simply connected unbranched cover of closed surfaces. Namely, each leaf of \mathcal{S} is homeomorphic to the unit disk and the restriction of the natural projection to each leaf is the universal covering map.

2.1 The profinite completion

Denote by G the fundamental group of S_0 . We define the profinite metric on G as follows. Let G_n be the intersection of all subgroups of G of index at most n . There are only finitely many such subgroups and their intersection G_n is also of finite index. (It is possible that $G_n = G_{n+1}$ for some n and we ignore the repeating groups.) From now on, $\{G_n\}_{n \in \mathbb{N}}$ is a sequence of decreasing (as sets) subgroups of G of finite index. Each G_n is a characteristic subgroup of G and in particular a normal subgroup. Since G is residually finite, it follows that $\bigcap_{n \in \mathbb{N}} G_n = \{\text{id}\}$. We define the *profinite distance* of $A, B \in G$ by

$$d_{\text{pf}}(A, B) = e^{-\frac{1}{n}},$$

where $AB^{-1} \in G_n \setminus G_{n+1}$. In particular, an element of G is close to the identity in the profinite metric d_{pf} if it belongs to G_n for n large.

We denote by \hat{G} the metric completion of G in the profinite metric d_{pf} (see [38]). Each point of \hat{G} is an equivalence class of Cauchy sequences in (G, d_{pf}) . The multiplication of two sequences is given by multiplying corresponding elements and the product of two Cauchy sequences is Cauchy. The operation of multiplying equivalence classes of Cauchy sequences is well defined and \hat{G} is a group with respect to multiplication. The group \hat{G} is homeomorphic to the Cantor set and there is a natural injective homomorphism of G into \hat{G} obtained by mapping $A \in G$ into the equivalence class of the constant sequence (A, A, A, \dots) . The image of G is dense in \hat{G} .

Since \hat{G} is a compact topological group, there exists a unique left and right translation measure m on \hat{G} such that $m(\hat{G}) = 1$. The measure m is called Haar measure and it is a positive Radon measure.

2.2 The G -tagged compact solenoid

At this point we fix a Fuchsian group G such that the Riemann surface \mathbb{D}/G has genus at least two, where \mathbb{D} is the unit disk. We describe the compact solenoid \mathcal{S} using the profinite group \hat{G} . Consider the product $\mathbb{D} \times \hat{G}$. The action of $A \in G$ on $\mathbb{D} \times \hat{G}$ is defined by

$$A(z, t) = (Az, tA^{-1}),$$

where $(z, t) \in \mathbb{D} \times \hat{G}$ and A acts by hyperbolic isometries on the disk component and by right multiplication by A^{-1} on the group \hat{G} component. By the universality of the compact solenoid \mathcal{S} , the quotient $(\mathbb{D} \times \hat{G})/G$ is homeomorphic to the compact

solenoid \mathfrak{S} (see [38]). The natural projection $\Pi: (\mathbb{D} \times \hat{G})/G \rightarrow \mathbb{D}/G$ is given by forgetting the second coordinate. Thus the fiber over a point in \mathbb{D}/G is homeomorphic to \hat{G} . The orbit in $\mathbb{D} \times \hat{G}$ under G of a single disk $\mathbb{D} \times \{t\}$ is a leaf of the solenoid. We define the orbit of $\mathbb{D} \times \{\text{id}\}$ to be the baseleaf and the orbit of $(0, \text{id})$ to be the basepoint. After fixing the baseleaf and the basepoint, each fiber has a unique identification with \hat{G} and the projection $\Pi: (\mathbb{D} \times \hat{G})/G \rightarrow \mathbb{D}/G$ is a \hat{G} -bundle.

We define the G -tagged compact solenoid \mathfrak{S}_G by

$$\mathfrak{S}_G = (\mathbb{D} \times \hat{G})/G.$$

Let $\omega \subset \mathbb{D}$ be a fundamental polygon for the action of G on \mathbb{D} . Then $\omega \times \hat{G}$ is a fundamental set for the action of G on $\mathbb{D} \times \hat{G}$. The action of G identifies a boundary side of $\omega \times \{t\}$ with a boundary side of $\omega \times \{tA^{-1}\}$, where $A \in G$ identifies the boundary side of ω onto another boundary side of ω . The group G is countable while \hat{G} is an uncountable set. Since G glues together the ω -pieces to make a single leaf, we conclude that $\mathfrak{S}_G \approx \mathfrak{S}$ has uncountably many leaves.

The holonomy of the leaves of the G -tagged compact solenoid \mathfrak{S}_G is given by the right translation of the group G in the group \hat{G} . Since m is a translation invariant measure on the group \hat{G} , we conclude that m induces a holonomy invariant transverse measure on the compact solenoid \mathfrak{S}_G .

3 Complex structures and hyperbolic metrics on the compact solenoid

A local chart of the compact solenoid \mathfrak{S} is homeomorphic to a (2-disk) \times (Cantor set). A transition function between two local charts is a homeomorphism from an open subset of a (2-disk) \times (Cantor set) onto another such set. In particular, the restriction of the transition map to each 2-disk is a homeomorphism and the family of homeomorphisms varies continuously in the Cantor set direction for the C^0 -topology on continuous maps.

A *complex structure* on the compact solenoid \mathfrak{S} is a choice of charts such that transition maps are holomorphic when restricted to each local leaf and vary continuously in the Cantor set direction for the C^0 -topology. Since maps are holomorphic, the continuous variation in the C^0 -topology implies continuous variation in the C^∞ -topology.

A *hyperbolic metric* on the compact solenoid \mathfrak{S} is an assignment of a metric of curvature -1 to each local leaf such that it varies continuously in the Cantor set direction. Namely, there is a choice of an atlas whose transition functions are leafwise isometries and vary continuously in the Cantor set direction, and the metric in the charts has curvature -1 on each local leaf.

It follows from the work of Candel [7] that any conformal structure on the compact solenoid \mathfrak{S} contains a unique hyperbolic metric. Any complex structure on the compact solenoid \mathfrak{S} corresponds to a conformal structure and any conformal structure gives a

unique complex structure by the continuous dependence on the parameters of the solution of the Beltrami equation (see Ahlfors–Bers [1]).

The above construction of the G -tagged compact solenoid provides an example of a complex structure on \mathcal{S} as well as a hyperbolic metric (by simply inducing the complex structure and the hyperbolic metric on the leaves of \mathcal{S} from the unit disk \mathbb{D}). The local charts of \mathcal{S} are chosen to be of the form $D \times \hat{G}$, where $D \subset \mathbb{D}$ is a small hyperbolic disk such that no two points of $D \times \hat{G}$ are in the same orbit of G . The complex structure on the unit disk \mathbb{D} gives complex charts for \mathcal{S} such that the transition maps between any two charts $D \times \hat{G}$ and $D_1 \times \hat{G}$ are constant in \hat{G} (namely, they are given by Möbius maps $A \in G$) and therefore continuous. The hyperbolic metric on \mathbb{D} gives a hyperbolic metric on \mathcal{S} which is also constant in the \hat{G} direction.

Complex structures on \mathcal{S} whose transition maps are locally constant in the Cantor set direction are called *transversely locally constant* (TLC) complex structures. (It is enough to find a subfamily of charts which cover \mathcal{S} for which transition maps are constant in the Cantor set direction.) Similarly, a hyperbolic metric on \mathcal{S} is TLC if there exists a cover of \mathcal{S} by charts in which the hyperbolic metric is locally constant in the Cantor set direction. It is a fact that any TLC complex structure (hyperbolic metric) is obtained by taking a G -tagged solenoid, where G is a Fuchsian group uniformization of a closed surface of a (possibly large) genus greater than one. This follows by the compactness of \mathcal{S} and the fact that each transverse direction corresponds to the profinite completion of a finite index subgroup of the fundamental group of a genus two surface (see also [32]). Therefore, the set of all TLC complex structures on \mathcal{S} is given by lifting complex structures on Riemann surfaces. Sullivan [44] showed that any complex structure on \mathcal{S} can be approximated by TLC complex structure in the C^0 -topology, which is equivalent to the C^∞ -topology.

4 The G -tagged non-compact solenoid

We introduce the non-compact solenoid \mathcal{S}_{nc} (see [37]). Since we require that the topological ends of leaves are well-behaved, our construction immediately assigns a hyperbolic metric on \mathcal{S}_{nc} . It will follow that \mathcal{S}_{nc} has finite area in an appropriate sense.

Let $G < \text{PSL}_2(\mathbb{Z})$ be such that \mathbb{D}/G is the modular, once punctured torus. Denote by \hat{G} the profinite completion of G . The action of G on $\mathbb{D} \times \hat{G}$ is given by $A(z, t) = (Az, tA^{-1})$ for $z \in \mathbb{D}$, $t \in \hat{G}$ and $A \in G$. We define the *non-compact solenoid* \mathcal{S}_{nc} by

$$\mathcal{S}_{\text{nc}} = (\mathbb{D} \times \hat{G})/G.$$

A leaf of \mathcal{S}_{nc} is the orbit under G of a single disk $\mathbb{D} \times \{t\}$. Let ω be a fundamental polygon for the action of G on \mathbb{D} such that the boundary edges are infinite geodesics which project to the geodesics on the torus \mathbb{D}/G connecting the puncture to itself. Then $\omega \times \hat{G}$ is a fundamental set for the action of G on $\mathbb{D} \times \hat{G}$. The identifications

by G on $\omega \times \hat{G}$ are identifying only the boundary edges in pairs on different levels according to the G -action on \hat{G} .

Any compact subset of \mathcal{S}_{nc} is a subset of a compact set of the form $((\mathbb{D}_r \cap \omega) \times \hat{G})/G$, where \mathbb{D}_r , $0 < r < 1$, is the Euclidean disk of radius r with center 0 and $\mathbb{D}_r \cap \omega$ is a compact subset of ω . The complement of $((\mathbb{D}_r \cap \omega) \times \hat{G})/G$ when restricted to a leaf $(\mathbb{D} \times \{t\})/G \equiv \mathbb{D}$ as is given by the G -orbit of a single horoball ζ in \mathbb{D} centered at the fixed point of a parabolic element of G . In the topology induced from \mathcal{S} on the set $G\{\zeta\}$ each horoball accumulates onto itself since it is preserved by the action of an infinite cyclic group generated by the parabolic element of G with the fixed point at the center of the horoball. A fundamental set η in the horoballs for the action of the cyclic group is the intersection of the horoball and the region between a geodesic with endpoint at the center of the horoball and its image under the generating parabolic map. Then the corresponding points in η and $C^n(\eta)$ are close for $n \in \mathbb{Z}$ with $|n|$ large, where $C \in G$ is the generating parabolic element with fixed point at the center of the horoball. Moreover, the corresponding points in the horoballs ζ and $A(\zeta)$ are close provided that $A \in G_n$ for n large. Therefore, each leaf of \mathcal{S}_{nc} has countably many topological ends corresponding to the fixed points of the parabolics in G . The non-compact solenoid \mathcal{S}_{nc} has only one end given by the equivalence class of the set $(G(\zeta) \times \hat{G})/G$.

Given a local chart of the form (2-disk) \times (Cantor set), there is a transversal identification of local charts. We consider only local charts of the form $D \times \hat{G} \subset \mathbb{D} \times \hat{G}$ for D a hyperbolic disk sufficiently small such that the projection map from $D \times \hat{G}$ to \mathcal{S}_{nc} is a homeomorphism. The transverse identification of local leaves is an isometry because the hyperbolic metric is constant in \hat{G} and it extends to an isometric identification of global leaves. This identification is specified by fixing two local leaves of two global leaves. In the above identification, the ends of leaves correspond to each other. The ends are called “*punctures*” by abuse of notation. We say that two punctures of \mathcal{S}_{nc} are close if they correspond to each other under an identification of the leaves on which they reside, where the identification is specified by two local leaves which are close in a given chart.

The above construction gives a hyperbolic metric on the leaves of \mathcal{S}_{nc} which is transversely locally constant. We will consider an arbitrary non-compact solenoid \mathcal{X} with a hyperbolic metric on leaves which varies continuously in the transverse direction together with a marking map $f: \mathcal{S}_{\text{nc}} \rightarrow \mathcal{X}$. The marking f is a homeomorphism which is quasiconformal and differentiable on leaves, varies continuously in the transverse direction for the C^1 -topology on differentiable maps and for the quasiconformal topology when global leaves are identified using local charts as above. In particular, the supremum of quasiconformal constants over the leaves is bounded. The end of \mathcal{S}_{nc} is homeomorphically mapped onto the end of \mathcal{X} . Moreover, the intersection of a leaf of \mathcal{S}_{nc} with the end is quasi-isometrically mapped onto the corresponding leaf of \mathcal{X} . Therefore, our notion of ends being close on the TLC non-compact solenoid \mathcal{S}_{nc} is transferable to an arbitrary non-compact marked solenoid $f: \mathcal{S}_{\text{nc}} \rightarrow \mathcal{X}$.

5 The Teichmüller space of the compact solenoid

We define the Teichmüller space $\mathcal{T}(\mathcal{S})$ of the compact solenoid \mathcal{S} . Let G be a fixed Fuchsian group such that \mathbb{D}/G is a closed Riemann surface of genus at least two. Let \mathcal{S}_G be the G -tagged compact solenoid with the induced complex structure from \mathbb{D}/G . The complex structure on the solenoid \mathcal{S}_G is a TLC complex structure.

Definition 5.1. A homeomorphism $f: \mathcal{S} \rightarrow \mathcal{X}$ of a complex compact solenoid \mathcal{S} onto a complex compact solenoid \mathcal{X} is said to be *quasiconformal* if it is differentiable and quasiconformal on each leaf and if it varies continuously in the transverse direction in the C^1 -topology on the C^1 -maps.

By the above definition, the composition $g \circ f$ of two quasiconformal maps $f: \mathcal{S} \rightarrow \mathcal{X}$ and $g: \mathcal{X} \rightarrow \mathcal{Y}$ is quasiconformal.

Remark 5.2. Since \mathcal{S} is compact, it follows that the continuity in the C^1 -topology for the variations on the local leaves implies the continuity in the quasiconformal topology on the global leaves. It is necessary to require smoothness of quasiconformal maps in order to preserve quasiconformality under the composition. One is tempted to require that Beltrami coefficients of leafwise quasiconformal maps vary continuously in the transverse direction in the essential supremum norm. However, the chain rule for Beltrami coefficients shows that the composition of two such maps does not satisfy the same continuity property unless the quasiconformal maps have additional C^1 smoothness and continuity in the transverse direction in the C^1 -topology.

Definition 5.3. The *Teichmüller space* $\mathcal{T}(\mathcal{S}_G)$ of the compact G -tagged solenoid \mathcal{S}_G consists of all quasiconformal maps $f: \mathcal{S}_G \rightarrow \mathcal{X}$ up to an equivalence. Two quasiconformal maps $f, g: \mathcal{S}_G \rightarrow \mathcal{X}, \mathcal{Y}$ are *Teichmüller equivalent* if there exists a conformal map $c: \mathcal{X} \rightarrow \mathcal{Y}$ such that $g^{-1} \circ c \circ f: \mathcal{S}_G \rightarrow \mathcal{S}_G$ is homotopic to the identity. Denote by $[f] \in \mathcal{T}(\mathcal{S}_G)$ the Teichmüller class of the quasiconformal map $f: \mathcal{S}_G \rightarrow \mathcal{X}$, i.e., all quasiconformal maps homotopic to f up to post-composition by conformal maps.

Since the transverse set T for \mathcal{S} is totally disconnected, any homotopy does not mix the leaves. Any two homotopic quasiconformal maps of a complex compact solenoid are isotopic through uniformly bounded quasiconformal maps [28, Theorem 3.1].

Definition 5.4. The *Teichmüller distance* d_T on $\mathcal{T}(\mathcal{S}_G)$ is given by

$$d_T([f], [g]) = \inf_{f_1 \in [f], g_1 \in [g]} 1/2 \log K(f_1 \circ g_1^{-1}),$$

where $K(f)$ is the supremum of the quasiconformal constants of the restrictions of $f: \mathcal{S} \rightarrow \mathcal{X}$ to the leaves of \mathcal{S} . Since f is transversely continuous in the C^1 -topology and since each leaf is dense in \mathcal{S} , we conclude that $K(f)$ is equal to the quasiconformal

constant on each leaf of \mathcal{S} . In particular, the restriction of f to each leaf has the same quasiconformal constant.

Sullivan [44] showed that the Teichmüller (pseudo-)metric is a genuine metric, i.e., that it is not degenerate. We give an alternative proof in this section.

5.1 The universal coverings of complex compact solenoids

Recall that the G -tagged complex solenoid \mathcal{S}_G is given by the quotient of $\mathbb{D} \times \hat{G}$ under the action of G . The complex structure and the hyperbolic metric on \mathcal{S}_G are inherited from \mathbb{D} and they are transversely locally constant. The natural quotient map

$$\pi : \mathbb{D} \times \hat{G} \rightarrow (\mathbb{D} \times \hat{G})/G \equiv \mathcal{S}_G$$

is a local homeomorphism which is leafwise conformal and which varies continuously in \hat{G} for the C^0 -topology on continuous maps (which is equivalent to the C^∞ -topology on conformal maps). The space $\mathbb{D} \times \hat{G}$ is globally much simpler (a product) than \mathcal{S}_G . Thus we consider $\mathbb{D} \times \hat{G}$ as a complex “universal covering” of a TLC solenoid \mathcal{S}_G and we consider G as the covering group with its action on $\mathbb{D} \times \hat{G}$.

Let $f : \mathcal{S}_G \rightarrow \mathcal{X}$ be a quasiconformal map, where \mathcal{X} is a hyperbolic compact solenoid not necessarily TLC. We form a complex universal covering for \mathcal{X} using the marking map f . We recall (see [42]) that there exists a chart $(U \times T, \psi)$ of \mathcal{X} , where U a disk with center 0, such that $\psi \circ f(\{0\} \times \hat{G}) = \{0\} \times T$. Then f induces a homeomorphism of \hat{G} and T . Consider a family of maps $\pi_t^{\mathcal{X}} : \mathbb{D} \rightarrow \mathcal{X}$, for $t \in T$, such that π_t is an isometry onto a leaf of \mathcal{X} (with its hyperbolic metric), $\psi \circ \pi_t^{\mathcal{X}}(0) = 0$ and $(\psi \circ \pi_t^{\mathcal{X}})'(0) > 0$. Then the maps $\pi_t^{\mathcal{X}}$ fit together to a single map

$$\pi^{\mathcal{X}} : \mathbb{D} \times T \rightarrow \mathcal{X},$$

defined by $\pi^{\mathcal{X}}(\cdot, t) := \pi_t^{\mathcal{X}}(\cdot)$. The map $\pi^{\mathcal{X}}$ is a local homeomorphism and a leafwise isometry. We consider $\mathbb{D} \times T$ as a hyperbolic (or a complex) “universal covering” with $\pi^{\mathcal{X}}$ as a cover map [42].

There is a well-defined lift

$$\tilde{f} : \mathbb{D} \times \hat{G} \rightarrow \mathbb{D} \times T$$

of the map $f : \mathcal{S}_G \rightarrow \mathcal{X}$ given by the formula $\tilde{f}(z, t) := (\pi_t^{\mathcal{X}})^{-1} \circ f \circ \pi(z, t)$ (see [42]).

The group G acts on $\mathbb{D} \times \hat{G}$ as a covering group of \mathcal{S}_G and we use \tilde{f} to introduce a conformal covering group for \mathcal{X} . Since

$$\pi^{\mathcal{X}} \circ \tilde{f} = f \circ \pi,$$

it follows that

$$\pi^{\mathcal{X}} \circ \tilde{f} \circ A = \pi^{\mathcal{X}} \circ \tilde{f}$$

for all $A \in G$. Let $(z, t) \in \mathbb{D} \times \hat{G}$, $\tilde{f}(z, t) = (w_1, t_1) \in \mathbb{D} \times T$ and $(\tilde{f} \circ A)(z, t) = (w_2, t_2) \in \mathbb{D} \times T$. By the above,

$$\pi^{\mathcal{X}}(w_2, t_2) = \pi^{\mathcal{X}}(w_1, t_1).$$

Consequently,

$$(\pi^{\mathcal{X}})^{-1}(\pi^{\mathcal{X}}(w_1, t_1))$$

contains (w_2, t_2) . Note that $(\pi_{t_2}^{\mathcal{X}})^{-1} \circ \pi_{t_1}^{\mathcal{X}}$ is an isometry of $\mathbb{D} \times \{t_1\}$ onto $\mathbb{D} \times \{t_2\}$ and $((\pi_{t_2}^{\mathcal{X}})^{-1} \circ \pi_{t_1}^{\mathcal{X}})(w_1) = w_2$. We induce an action of $A \in G$ on T by its natural action (by right multiplication) on \hat{G} via the identification $\psi \circ f \circ \pi : \hat{G} \equiv T$. We introduce a covering map $A_{\mathcal{X}}$ on the universal covering $\mathbb{D} \times T$ of \mathcal{X} corresponding to A by

$$A_{\mathcal{X}}(z, t) = ((\pi_{tA^{-1}}^{\mathcal{X}})^{-1} \circ \pi_t^{\mathcal{X}}(z), tA^{-1}),$$

where $t, tA^{-1} \in T \equiv \hat{G}$. The covering map $A_{\mathcal{X}}$ is an isometry on each leaf. Moreover, $A_{\mathcal{X}}$ is transversely continuous and $\tilde{f} \circ A = A_{\mathcal{X}} \circ \tilde{f}$ from the definition (see [42]). Then we define $G_{\mathcal{X}} := \tilde{f}G\tilde{f}^{-1}$ to be the covering group of \mathcal{X} , namely $(\mathbb{D} \times T)/G_{\mathcal{X}}$ is conformally equivalent to \mathcal{X} .

5.2 Beltrami coefficients and holomorphic quadratic differentials on the compact solenoid

Given a quasiconformal map $f : \mathcal{S}_G \rightarrow \mathcal{X}$, there is a corresponding leafwise smooth (i.e., C^1) Beltrami coefficient $\mu = \frac{\partial \tilde{f}}{\partial \bar{f}}$ which is continuous for the transverse variations in the local charts for the C^1 -topology on C^1 -maps. The lift \tilde{f} has Beltrami coefficient $\tilde{\mu}$ (which is the lift of μ) and it satisfies

$$\tilde{\mu}(z, t) = \tilde{\mu}(Az, tA^{-1}) \frac{\overline{A'(z)}}{A'(z)}, \tag{5.1}$$

for $A \in G$.

More generally, if $g : \mathcal{X} \rightarrow \mathcal{Y}$ is a quasiconformal map of complex compact solenoids then there exists a lift $\tilde{g} : \mathbb{D} \times T_1 \rightarrow \mathbb{D} \times T_2$ to their universal covers. The Beltrami coefficient ν of g lifts to the Beltrami coefficient $\tilde{\nu}$ on $\mathbb{D} \times T_1$ such that

$$\tilde{\nu}(z, t) = \tilde{\nu}(A_{\mathcal{X}}(z, t)) \frac{\overline{A'_{\mathcal{X}}(z, t)}}{A'_{\mathcal{X}}(z, t)},$$

for $A_{\mathcal{X}} \in G_{\mathcal{X}}$, where $A_{\mathcal{X}}(z, t) = (A_t^{\mathcal{X}}(z), tA^{-1})$ and $A'_{\mathcal{X}}(z, t)$ is the leafwise derivative. Note that $A'_{\mathcal{X}}(z, t)$ depends on t .

By the compactness of \mathcal{S}_G , the continuity in the local charts for the transverse variations of a Beltrami coefficient μ on \mathcal{S}_G implies that

$$\|\tilde{\mu}(\cdot, t) - \tilde{\mu}(\cdot, t_1)\|_{\infty} \rightarrow 0 \tag{5.2}$$

as $t \rightarrow t_1$, for all $t_1 \in \hat{G}$. In the opposite direction, a Beltrami coefficient $\tilde{\mu}$ on $\mathbb{D} \times \hat{G}$ which is leafwise C^1 , which is continuous in the C^1 -topology for the variations in \hat{G} on the compact subsets of $\mathbb{D} \times \hat{G}$ and which satisfies (5.2) is the lift of the Beltrami coefficient of a quasiconformal map $f: \mathcal{S}_G \rightarrow \mathcal{X}$, where \mathcal{X} is determined by $\tilde{\mu}$.

Let $\tilde{f}^{\tilde{\mu}}$ denote the leafwise solution to the Beltrami equation with the coefficient $\tilde{\mu}$ on $\mathbb{D} \times \hat{G}$ normalized such that 1, i and -1 are fixed on each leaf. Then $\tilde{f}^{\tilde{\mu}}$ conjugates the action of G on $\mathbb{D} \times \hat{G}$ to the action of $G^{\tilde{\mu}} = \tilde{f}^{\tilde{\mu}} \circ G \circ (\tilde{f}^{\tilde{\mu}})^{-1}$ on $\mathbb{D} \times \hat{G}$. Let $\mathcal{X}^{\tilde{\mu}} = (\mathbb{D} \times \hat{G})/G^{\tilde{\mu}}$ be the induced complex solenoid. Then $\tilde{f}^{\tilde{\mu}}$ projects to a quasiconformal map $f^{\tilde{\mu}}: \mathcal{S}^G \rightarrow \mathcal{X}^{\tilde{\mu}}$.

Definition 5.5. A *transversely locally constant* (TLC) Beltrami coefficient on a TLC compact solenoid \mathcal{S} is a leafwise Beltrami coefficient which is constant in the transverse direction in some atlas of local charts.

Definition 5.6. A *holomorphic quadratic differential* φ on a complex compact solenoid \mathcal{X} is a leafwise holomorphic quadratic differential which varies continuously in the local chart in the transverse direction in the C^0 -topology. Equivalently, a leafwise holomorphic function $\tilde{\varphi}$ on the universal cover $\mathbb{D} \times T$ of \mathcal{X} is a lift of a holomorphic quadratic differential if

$$\tilde{\varphi}(z, t) = \tilde{\varphi}(A_{\mathcal{X}}(z, t))A'_{\mathcal{X}}(z, t)^2 \tag{5.3}$$

for $A_{\mathcal{X}} \in G_{\mathcal{X}}$ and if

$$\|\tilde{\varphi}(\cdot, t) - \tilde{\varphi}(\cdot, t_1)\|_{\text{Bers}} \rightarrow 0 \tag{5.4}$$

as $t \rightarrow t_1$, where $\|f\|_{\text{Bers}} := \sup_{z \in \mathbb{D}} |\rho^{-2}(z) f(z)|$ with ρ the Poincaré density on \mathbb{D} (see [42]).

Definition 5.7. A *transversely locally constant* (TLC) holomorphic quadratic differential on a complex compact solenoid \mathcal{S} is a leafwise holomorphic quadratic differential which is constant in the transverse direction in some atlas of local charts.

Using the above notion of Beltrami coefficients on the universal cover of \mathcal{S}_G we give an equivalent definition of the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$.

Definition 5.8. The *Teichmüller space* $\mathcal{T}(\mathcal{S}_G)$ of the compact G -tagged solenoid \mathcal{S}_G consists of all smooth Beltrami coefficients $\tilde{\mu}$ on $\mathbb{D} \times \hat{G}$ which vary continuously in the C^1 -topology on compact subsets of $\mathbb{D} \times \hat{G}$ and which satisfy (5.1) and (5.2) modulo an equivalence relation. Two Beltrami coefficients $\tilde{\mu}$ and $\tilde{\nu}$ are (Teichmüller) equivalent if there exists a conformal map $c: \mathcal{X}^{\tilde{\mu}} \rightarrow \mathcal{X}^{\tilde{\nu}}$ such that $(f^{\tilde{\nu}})^{-1} \circ c \circ f^{\tilde{\mu}}: \mathcal{S}_G \rightarrow \mathcal{S}_G$ is isotopic to the identity map.

5.3 The restriction map π_l

We recall the definition of the restriction map $\pi_l: \mathcal{T}(\mathcal{S}_G) \rightarrow \mathcal{T}(\mathbb{D})$ from [44]. Given a quasiconformal map $f: \mathcal{S}_G \rightarrow \mathcal{X}$, the restriction to the baseleaf $f|_l: l \rightarrow f(l)$

maps l to the leaf $f(l) \subset \mathcal{X}$. We fix a conformal identification $l \equiv \mathbb{D}$ and take an arbitrary conformal identification $f(l) \equiv \mathbb{D}$. Then $f|_l: \mathbb{D} \rightarrow \mathbb{D}$ is well defined up to post-composition with a conformal map of \mathbb{D} (because of the choice $f(l) \equiv \mathbb{D}$). This gives a well-defined element of the universal Teichmüller space $\mathcal{T}(\mathbb{D})$. Sullivan [44] showed that π_l is injective. We give a different proof.

Theorem 5.9. *The map $\pi_l: \mathcal{T}(\mathcal{S}_G) \rightarrow \mathcal{T}(\mathbb{D})$ is injective.*

Proof. It is enough to show that if $\pi_l([f])$ is trivial in $\mathcal{T}(\mathbb{D})$ then $[f] \in \mathcal{T}(\mathcal{S}_G)$ is trivial. Let $\tilde{f}: \mathbb{D} \times \hat{G} \rightarrow \mathbb{D} \times T$ be a lift of $f: \mathcal{S}_G \rightarrow \mathcal{X}$ to the universal coverings. Denote by $G_{\mathcal{X}}$ the covering group of \mathcal{X} . Our assumption implies that $\tilde{f}|_{S^1 \times \{\text{id}\}}$ is a Möbius map. By the invariance of \tilde{f} we conclude that $\tilde{f}|_{S^1 \times \{A\}} = (A_{\mathcal{X}})^{-1} \circ \tilde{f}|_{S^1 \times \{\text{id}\}} \circ A$ is also a Möbius map, for each $A \in G$. Since G is dense in \hat{G} , we conclude that \tilde{f} is a Möbius map on the boundary $S^1 \times \{t\}$, $t \in \hat{G}$, of each leaf.

Thus, when restricted to a leaf, \tilde{f} is homotopic to a Möbius map (where different leaves can give different Möbius maps). We need to show that there is a homotopy F_t , $0 \leq t \leq 1$, of \tilde{f} to Möbius maps on leaves such that $F_1 = \tilde{f}$, $F_t|_{S^1 \times \hat{G}} = \tilde{f}|_{S^1 \times \hat{G}}$ for each t , and the Beltrami coefficients $\tilde{\mu}_t$ of F_t satisfy (5.1) and (5.2), for each t .

Let $\tilde{\mu}$ be the Beltrami coefficient of \tilde{f} . Then we consider a path of Beltrami coefficients $t \mapsto \tilde{\nu}_t = t\tilde{\mu}$, for $0 \leq t \leq 1$, which satisfy (5.1) and (5.2). Then $\tilde{\nu}_t$ converges to the trivial (i.e., zero) Beltrami coefficient as $t \rightarrow 0$ and the path of properly normalized solutions $t \mapsto \tilde{f}^{\tilde{\nu}_t}$ give a homotopy from $\tilde{f}^{\tilde{\nu}_1} = \tilde{f}$ to the Möbius maps. However, it is not guaranteed that $\tilde{f}^{\tilde{\nu}_t}$ extends to the Möbius maps (determined by \tilde{f}) on the boundaries $S^1 \times \hat{G}$ for $0 < t < 1$. Let h_t be the boundary map for $\tilde{f} \circ (\tilde{f}^{\tilde{\mu}_t})^{-1}$. Let g_t be the leafwise barycentric extensions of h_t , for $0 \leq t \leq 1$. Then $g_1 = g_0 = \text{id}$ because $h_1 = h_0 = \text{id}$ on the boundary (by the properties of the barycentric extension [9]). The Beltrami coefficients of $g_t \circ \tilde{f}^{\tilde{\nu}_t}$ satisfy (5.1) and (5.2) for each t (again by the properties of the barycentric extension [9]) and the path $t \mapsto g_t \circ \tilde{f}^{\tilde{\nu}_t}$ gives a homotopy from $\tilde{f} = g_1 \circ \tilde{f}^{\tilde{\nu}_1}$ to the leafwise Möbius maps. (The idea of using barycentric extensions to find homotopies first appears in [12] for plane domains, and it is utilized in [28] to show that homotopic maps of the compact solenoid are isotopic as well.) \square

Sullivan [44] showed that the Teichmüller metric on $\mathcal{T}(\mathcal{S}_G)$ is a genuine metric. We use the above theorem to give an alternative argument.

Theorem 5.10. *The Teichmüller metric on the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the compact solenoid is a genuine metric, i.e., $\mathcal{T}(\mathcal{S}_G)$ is a Hausdorff space for the Teichmüller metric.*

Proof. Note that $\pi_l: \mathcal{T}(\mathcal{S}_G) \rightarrow \mathcal{T}(\mathbb{D})$ is a contracting map with respect to the Teichmüller metrics on $\mathcal{T}(\mathcal{S}_G)$ and $\mathcal{T}(\mathbb{D})$. Since the Teichmüller metric on $\mathcal{T}(\mathbb{D})$ is a genuine metric the theorem follows. \square

5.4 The complex Banach manifold structure on $\mathcal{T}(\mathcal{S}_G)$

The Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ embeds as an open subset in a complex Banach vector space as follows. Denote by $\tilde{f}: \mathbb{D} \times \hat{G} \rightarrow \mathbb{D} \times T$ the lift to the universal covering of a quasiconformal map $f: \mathcal{S}_G \rightarrow \mathcal{X}$. The Bers embedding for the universal Teichmüller space assigns to each $\tilde{f}|_{\mathbb{D} \times \{t\}}$ a holomorphic quadratic differential $\tilde{\varphi}|_{\mathbb{D} \times \{t\}}$. The holomorphic quadratic differential $\tilde{\varphi}$ satisfies (5.4) because of the continuous dependence on the parameters of the solutions to the Beltrami equation [1] and it satisfies (5.3) because the Beltrami coefficient of \tilde{f} satisfies (5.1).

We denote by $B(\mathcal{S}_G)$ the space of all holomorphic quadratic differentials on \mathcal{S}_G which vary continuously in the transverse direction in the local charts for the C^0 -topology. Note that $B(\mathcal{S}_G)$ is conformally isometric to the space of all leafwise holomorphic functions on $\tilde{\varphi}: \mathbb{D} \times \hat{G} \rightarrow \mathbb{C}$ that are uniformly leafwise Bers bounded, i.e., $\sup_{t \in \hat{G}} \|\tilde{\varphi}|_{\mathbb{D} \times \{t\}}\|_{\text{Bers}} < \infty$, and that vary continuously in the transverse direction for Bers norm, i.e., $\|\tilde{\varphi}|_{\mathbb{D} \times \{t\}} - \tilde{\varphi}|_{\mathbb{D} \times \{t_1\}}\|_{\text{Bers}} \rightarrow 0$ as $t \rightarrow t_1$, for each $t_1 \in \hat{G}$, and that are invariant under the action of G , i.e., they satisfy (5.3) (see [42]).

Therefore, we obtained a map $\Phi: \mathcal{T}(\mathcal{S}_G) \rightarrow B(\mathcal{S}_G)$ which is injective because the Bers map for the universal Teichmüller space is injective and the restriction map π_l is injective. Moreover, Φ is a homeomorphism onto an open subset of $B(\mathcal{S}_G)$ (see [42] for details). Note that $\Phi: \mathcal{T}(\mathcal{S}_G) \rightarrow B(\mathcal{S}_G)$ is the quotient of the holomorphic map $\tilde{\Phi}: U_s^\infty(\mathcal{S}_G) \rightarrow B(\mathcal{S}_G)$, where $U_s^\infty(\mathcal{S}_G)$ is the unit ball in the space $L_s^\infty(\mathcal{S}_G)$ of all leafwise smooth, transversely continuous Beltrami differentials with the essential supremum norm and where $\tilde{\Phi}$ is obtained by taking the leafwise Bers embedding construction as above. Thus we define $\Phi: \mathcal{T}(\mathcal{S}_G) \rightarrow B(\mathcal{S}_G)$ to be a complex global chart for $\mathcal{T}(\mathcal{S}_G)$. For details see Sullivan [44].

Since $\mathcal{T}(\mathcal{S}_G)$ has a complex structure, there is a well-defined Kobayashi pseudometric on $\mathcal{T}(\mathcal{S}_G)$. The *Kobayashi pseudometric* is the largest metric on $\mathcal{T}(\mathcal{S}_G)$ which makes all holomorphic maps from the unit disk with the Poincaré metric into $\mathcal{T}(\mathcal{S}_G)$ weakly contracting. It is a well-known fact that the Kobayashi pseudometric coincides with the Teichmüller metric for the Teichmüller spaces of Riemann surfaces (see [41], [17]). We showed that the same is true for $\mathcal{T}(\mathcal{S}_G)$ [42].

Theorem 5.11. *On the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the universal hyperbolic solenoid \mathcal{S}_G , the Kobayashi pseudometric equals the Teichmüller metric. In particular, the Kobayashi pseudometric is a metric.*

6 The Reich–Strebel inequality

The study of the Teichmüller metric on Teichmüller spaces of Riemann surfaces depends on the Reich–Strebel inequality which is a (highly non-trivial) generalization of the length-area method for finding extremal maps between quadrilaterals. We give a proper generalization of the Reich–Strebel inequality for the marked compact solenoid

\mathcal{X} from [42]. If φ is a transversely continuous holomorphic quadratic differential then $|\varphi|$ is a leafwise area form on \mathcal{X} . The product $|\varphi|dm$ is a measure on \mathcal{X} . Recall that m is the Haar measure on the profinite completion group \hat{G} of G and that the transverse sets in the local charts for \mathcal{X} are identified with \hat{G} .

Definition 6.1. Let φ be a holomorphic quadratic differential on a complex compact solenoid \mathcal{X} . Then

$$\|\varphi\|_{L^1(\mathcal{X})} := \iint_{\mathcal{X}} |\varphi| dm.$$

Definition 6.2. The space of all holomorphic quadratic differentials on a complex compact solenoid \mathcal{X} is called $A(\mathcal{X})$.

The proof of the Reich–Strebel inequality for the closed solenoid used a careful approximation argument (of holomorphic quadratic differentials and complex solenoids by TLC holomorphic quadratic differentials on TLC complex solenoids) in [42] and we give a different proof below utilizing an idea of Gardiner [17, Section 2] for the proof in the closed surface case.

Definition 6.3. A Beltrami coefficient μ on a complex solenoid \mathcal{X} is called *Teichmüller trivial* if it is equivalent to the trivial coefficient 0, i.e., the solution of the Beltrami equation is homotopic to a conformal map.

Theorem 6.4 (Reich–Strebel inequality). *Let φ be a holomorphic quadratic differential on the solenoid \mathcal{X} and let μ be a Teichmüller trivial Beltrami coefficient. Then*

$$\|\varphi\|_{L^1(\mathcal{X})} \leq \int_{\mathcal{X}} \frac{|1 + \mu \frac{\varphi}{|\varphi|}|^2}{1 - |\mu|^2} |\varphi| dm. \tag{6.1}$$

Proof. Let $\varphi \in A(\mathcal{X})$ and let $f : \mathcal{X} \rightarrow \mathcal{X}$ be the quasiconformal map whose Beltrami coefficient is μ . Then f is homotopic to the identity on \mathcal{X} and its restriction to each leaf is homotopic to the identity. Since φ is a holomorphic function on each leaf, the set of zeroes of φ on each leaf is at most countable and they accumulate at the boundary of the leaf. Thus, the set of critical vertical (as well as horizontal) trajectories is countable on each leaf and does not influence the integration of $|\varphi|$ on compact subsets of a leaf.

For a given arc $\beta \subset \mathcal{X}$, we denote by $h_\varphi(\beta)$ the height of β , namely the length in the metric $|\operatorname{Im}(\sqrt{\varphi(z, \bar{t}})dz)|$ given in the local chart. We claim that there exists $M > 0$ such that for any compact segment β on a non-critical vertical trajectory we have

$$h_\varphi(\beta) \leq h_\varphi(f(\beta)) + M. \tag{6.2}$$

Let $t \mapsto f_t$ be a homotopy from $f_0 = \operatorname{id}$ to $f_1 = f$. Recall that if γ is a path in \mathcal{X} connecting the endpoints of β then $h_\varphi(\beta) \leq h_\varphi(\gamma)$ (see, for example, [43] or [17, Lemma 2, page 41]).

Let p be the initial point and let q be the terminal point of β . We define a path γ connecting the endpoints of β by taking $\gamma_0 : t \mapsto f_t(p)$ followed by $f(\beta)$ followed

by $\gamma_1 : t \mapsto f_{1-t}(q)$. Then

$$h_\varphi(\beta) \leq h_\varphi(\gamma_0) + h_\varphi(f(\beta)) + h_\varphi(\gamma_1).$$

We define the displacement function $d : \mathcal{X} \rightarrow \mathbb{R}$ for the map f . Since f is homotopic to the identity and since the transverse set is totally disconnected, it follows that f fixes each leaf. Then $d(p)$, for $p \in \mathcal{X}$, is defined by taking the leafwise distance in the metric $\sqrt{|\varphi|}|dz|$ from p to $f(p)$. The displacement function d is continuous because φ varies continuously for the transverse variations in charts and f is continuous as well for the transverse variations. Since \mathcal{S} is a compact space, there exists a maximum M_1 for the displacement function d . Then, from the above inequality, we obtain the desired inequality (6.2) by using the above triangle inequality for heights, by observing that the height h_φ of a curve is shorter than the distance in the above metric $\sqrt{|\varphi|}|dz|$ and by taking $M = 2M_1$.

We claim that each ray of a non-critical vertical trajectory of φ is of infinite length. To see this, assume that a ray r of a non-critical vertical trajectory is of finite length in the $\sqrt{|\varphi|}|dz|$ metric, namely $h_\varphi(r) = h < \infty$. Then let $0 < u_n < h$ be an increasing sequence of parameters for r with $u_n \rightarrow h$ such that $r(u_n)$ converges to a point $q \in \mathcal{X}$ (there is a convergent sequence by the compactness of \mathcal{X}). Then either q belongs to the same leaf as r or to a different leaf. We consider both cases below.

If q belongs to the same leaf as r then a standard argument shows that q must be a zero of φ [43], [17]. This implies that r is critical which is a contradiction.

If q belongs to another leaf, then q must be a zero of φ as well. Otherwise, there would exist a neighborhood of q in \mathcal{X} in which φ does not have any zeroes. This neighborhood contains the product of a Cantor set and a disk with fixed radius in the metric $\sqrt{|\varphi|}|dz|$. But r has to enter this neighborhood intersecting the disks of half the radius infinitely many times. This implies that r has an infinite length which is a contradiction.

Therefore q is a zero of φ . Then there exists a neighborhood of q in \mathcal{X} consisting of disks with small fixed radius around a transverse neighborhood of q such that all zeroes of φ in this neighborhood are in the disks of $1/3$ the radius. Note that the vertical ray r has to enter infinitely many times in the smaller disks and exit the larger disk. In particular, r crosses infinitely many times the annulus whose outer boundary is the boundary of the larger disk and whose inner boundary is the boundary of the smaller disk. The holomorphic quadratic differential has no zeros in the annulus. It follows that the length of r is infinite, which is again a contradiction. Thus r has infinite length.

At this point we modify the standard arguments in [17] to the compact solenoid \mathcal{X} . On the set of points $p \in \mathcal{X}$ which do not lie on the critical vertical trajectories of φ , we define the function

$$g(p) = h_\varphi(f(\beta_p))$$

where β_p is a compact vertical segment with center p and length b . By a change of variable, we obtain

$$g(p) = \int_{\beta_p} |\operatorname{Im}(\sqrt{\psi} dz)|,$$

where $\psi = (\varphi \circ f) f_z^2 (1 - \frac{\mu\varphi}{|\varphi|})^2$ is a quadratic differential on \mathcal{X} . At this point, we write all the integration in terms of the natural parameter $\zeta = \xi + i\eta$ for φ . Then

$$\int_{\mathcal{X}} g(p) d\xi d\eta dm = b \int_{\mathcal{X}} |\operatorname{Im}\sqrt{\psi}| d\xi d\eta dm$$

by Fubini's theorem. By (6.2), we obtain $b - M \leq \int_{\beta_p} |\operatorname{Im}(\sqrt{\psi} d\zeta)|$ for $p \in \mathcal{X}$. By integrating both sides of the above inequality over \mathcal{X} with respect to the measure $d\xi d\eta dm$, we obtain

$$\frac{b - M}{b} \int_{\mathcal{X}} d\xi d\eta dm \leq \int_{\mathcal{X}} |\operatorname{Im}\sqrt{\psi(\zeta)}| d\xi d\eta dm.$$

By letting $b \rightarrow \infty$ and inserting $|\sqrt{\varphi(\zeta)}| = 1$ under the integral on the right, we obtain

$$\int_{\mathcal{X}} |\varphi| dm \leq \int_{\mathcal{X}} |\sqrt{\varphi}\sqrt{\psi}| dm,$$

and after substituting the expression for ψ and using Cauchy–Schwarz's inequality, we obtain the desired inequality called the *Reich–Strebel inequality*. \square

We consider equivalence classes of Beltrami coefficients on the compact solenoid \mathcal{S}_G as elements of the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$. If $f: \mathcal{S}_G \rightarrow \mathcal{X}$ is a marked solenoid and μ is a Beltrami coefficient on \mathcal{X} , then there is a marked solenoid $f^\mu \circ f: \mathcal{S}_G \rightarrow \mathcal{X}^\mu$ such that the Beltrami coefficient of $f^\mu: \mathcal{X} \rightarrow \mathcal{X}^\mu$ is μ . Then the class of the Beltrami coefficient of $f^\mu \circ f$ determines a point in $\mathcal{T}(\mathcal{S}_G)$. In this sense, we consider the class of a Beltrami coefficient on a marked solenoid \mathcal{X} as an element of $\mathcal{T}(\mathcal{S}_G)$.

A derivative of a path of Beltrami coefficients on a marked compact solenoid \mathcal{X} is called a *Beltrami differential* (when the derivative exists) and it is considered as a representative of a tangent vector to $\mathcal{T}(\mathcal{S}_G)$ at the marked point \mathcal{X} . A Beltrami differential has finite essential supremum norm while a Beltrami coefficient has essential supremum norm less than 1.

One important question is when do two Beltrami differentials on \mathcal{X} represent the same tangent vector. The Reich–Strebel inequality gives the answer (see [42]) similar to the Riemann surface case. We say that a Beltrami coefficient μ on a complex solenoid X is *infinitesimally trivial* if

$$\int_X \mu\varphi dm = 0$$

for each holomorphic quadratic differential φ on \mathcal{X} .

Theorem 6.5. *A smooth Beltrami differential v on a complex solenoid \mathcal{X} is infinitesimally trivial if and only if there exists a holomorphic curve μ_s of Teichmüller trivial smooth Beltrami coefficients on \mathcal{X} such that $\mu_s = sv + O(s^2)$ in the essential supremum norm on \mathcal{X} .*

Denote by $L_s^\infty(\mathcal{X})$ the space of all smooth Beltrami differentials on \mathcal{X} that vary continuously in the transverse direction for the C^1 -topology. Denote by $N(\mathcal{X})$ the space of infinitesimally trivial smooth Beltrami differentials. The above theorem identifies the space of tangent vectors at $[f : \mathcal{G}_G \rightarrow \mathcal{X}] \in \mathcal{T}(\mathcal{G}_G)$ with $L_s^\infty(\mathcal{X})/N(\mathcal{X})$. Since $L_s^\infty(\mathcal{X})$ and $N(\mathcal{X})$ are not complete, it is not obvious that the tangent space is a complete vector space.

Recall that $A(\mathcal{X})$ is the space of all (transversely continuous) holomorphic quadratic differentials on \mathcal{X} . We introduce a surjective continuous linear map $P : L_s^\infty(\mathcal{X}) \rightarrow A(\mathcal{X})$, where $A(\mathcal{X})$ is equipped with the Bers norm. Note that $L_s^\infty(\mathcal{X})$ is identified with the space of all essentially bounded leafwise smooth function $\tilde{\mu}$ on the universal cover $\mathbb{D} \times T$ of \mathcal{X} that are continuous for the transverse variations in the C^1 -topology and for the essential supremum norm, i.e.,

$$\|\tilde{\mu}(z, t) - \tilde{\mu}(z, t_1)\|_\infty \rightarrow 0$$

as $t \rightarrow t_1$ for all $t_1 \in T$, and that satisfy

$$\tilde{\mu}(z, t) = \tilde{\mu}(A_{\mathcal{X}}(z, t)) \frac{\overline{A'_{\mathcal{X}}(z, t)}}{A'_{\mathcal{X}}(z, t)}$$

for all $A_{\mathcal{X}} \in G_{\mathcal{X}}$.

Then $P : L_s^\infty(\mathcal{X}) \rightarrow A(\mathcal{X})$ is defined by taking leafwise Bers' reproducing formula and noting that the invariance of $\tilde{\mu}$ with respect to $G_{\mathcal{X}}$ gives the invariance of the leafwise holomorphic functions $P(\tilde{\mu})$ with respect to $G_{\mathcal{X}}$. The transverse continuity of $P(\tilde{\mu})$ follows by the continuity of the Bers' reproducing formula.

We showed in [42] that P induces a linear isomorphism \bar{P} from the tangent space at the point $[f : \mathcal{G}_G \rightarrow \mathcal{X}] \in \mathcal{T}(\mathcal{G}_G)$ onto $A(\mathcal{X})$.

Corollary 6.6. *The map $P : L_s^\infty(\mathcal{X}) \rightarrow A(\mathcal{X})$ induces a continuous linear isomorphism from the normed space $L_s^\infty(\mathcal{X})/N(\mathcal{X})$ onto the Banach space $A(\mathcal{X})$ equipped with the Bers norm. Consequently, the tangent space $L_s^\infty(\mathcal{X})/N(\mathcal{X})$ at any point $[f : \mathcal{G}_G \rightarrow \mathcal{X}] \in \mathcal{T}(\mathcal{G}_G)$ is a complex Banach space.*

Thus the tangent space to $\mathcal{T}(\mathcal{G}_G)$ has a nice interpretation in terms of the harmonic Beltrami differentials as in the case of Teichmüller spaces of Riemann surfaces. We considered in [42] to which extent the duality between the integrable holomorphic quadratic differentials and tangent vectors carries from Teichmüller spaces of Riemann surfaces to $\mathcal{T}(\mathcal{G}_G)$. It is worth noting that $A(\mathcal{X})$ is a complete space in the Bers norm and it is not complete in the L^1 -norm. This accounts for the difference from the Riemann surface case.

Theorem 6.7. *The dual $A^*(\mathcal{X})$ for L^1 -norm on $A(\mathcal{X})$ is strictly larger than the tangent space at $[f : \mathcal{S}_G \rightarrow \mathcal{X}] \in \mathcal{T}(\mathcal{S}_G)$.*

Denote by $A^1(\mathcal{X})$ the space of integrable, a.e. leafwise holomorphic quadratic differentials on \mathcal{X} . Then $A(\mathcal{X}) \subsetneq A^1(\mathcal{X})$ and we showed in [42] the density statement in the L^1 -norm.

Theorem 6.8. *The closure of $A(\mathcal{X})$ for the L^1 -norm is equal to $A^1(\mathcal{X})$.*

7 The Teichmüller-type extremal maps

The Teichmüller distance between a point $[f : \mathcal{S}_G \rightarrow \mathcal{X}] \in \mathcal{T}(\mathcal{S}_G)$ and the basepoint $[\text{id}] \in \mathcal{T}(\mathcal{S}_G)$ is the infimum of the logarithms of the quasiconformal constants of all maps homotopic to f . A map $f_1 \in [f]$ is called *extremal* if it has the least quasiconformal constant in the homotopy class $[f]$. If $f_1 \in [f]$ is extremal then

$$d_T([f], [\text{id}]) = 1/2 \log K(f_1).$$

A Beltrami coefficient μ on \mathcal{S}_G is called *extremal* if its corresponding quasiconformal map is extremal.

Given a holomorphic quadratic differential φ on \mathcal{X} , the Beltrami coefficient $k \frac{|\varphi|}{\varphi}$ is called a *Teichmüller-type* Beltrami coefficient. The corresponding quasiconformal map $f^{k \frac{|\varphi|}{\varphi}}$ is called a *Teichmüller-type* map; the quasiconformal constant of $f^{k \frac{|\varphi|}{\varphi}}$ is $K = \frac{1+k}{1-k}$; in the natural parameter $\zeta = \sqrt{\varphi}$, $f^{k \frac{|\varphi|}{\varphi}}$ is given by stretching the horizontal direction by a factor \sqrt{K} and by shrinking the vertical direction by a factor $1/\sqrt{K}$.

An important consequence of the Reich–Strebel inequality is that the Teichmüller-type Beltrami coefficients are extremal in their classes (see [42]). In fact, a path of Teichmüller-type Beltrami coefficients gives a geodesic in $\mathcal{T}(\mathcal{S}_G)$.

Theorem 7.1. *Let $f : \mathcal{S}_G \rightarrow \mathcal{X}$ be a quasiconformal map and let $\varphi \neq 0$ be a holomorphic quadratic differential on \mathcal{X} . Then the path $t \mapsto t \frac{|\varphi|}{\varphi}$, $-1 < t < 1$, of Teichmüller type Beltrami coefficients on \mathcal{X} gives a geodesic (in the Teichmüller metric) through the point $[f] \in \mathcal{T}(\mathcal{S}_G)$. In addition, any two points on this geodesic have no other geodesics connecting them.*

Remark 7.2. Note that $\varphi \in A(\mathcal{X})$ can have zeros on \mathcal{X} which makes Teichmüller-type Beltrami coefficient discontinuous at these points. Strictly speaking a Teichmüller-type Beltrami coefficient does not belong to a Teichmüller class of smooth Beltrami coefficients on \mathcal{X} . However, this is a technical difficulty which was addressed in [42]. In fact, any zero of φ on a leaf of \mathcal{X} has a neighborhood in \mathcal{X} such that each local leaf has at least one zero. It can happen that a multiple zero of φ on one leaf is a limit of several simple zeros of φ on nearby leaves. The idea is to replace the Teichmüller-type

Beltrami coefficient in such small neighborhoods of zeros of φ by a smooth Beltrami coefficient such that the new global Beltrami coefficient on \mathcal{X} is smooth. This can be done in such a way that the restriction to each leaf of the original Teichmüller-type Beltrami coefficient and the new smooth Beltrami coefficient represent the same point in the universal Teichmüller space $\mathcal{T}(\mathbb{D})$ and that the sequence of new smooth Beltrami coefficients (obtained by shrinking the neighborhoods of zeros of φ to a zero area set) converges to the Teichmüller-type Beltrami coefficient uniformly on the compact subsets of the complement of the set of zeros of each leaf. Moreover, the essential supremum norm of the approximating sequence approaches the norm of the Teichmüller-type Beltrami coefficient (see [42, Proposition 5.1]). Thus, Teichmüller-type Beltrami coefficients are “well” approximated by smooth Beltrami coefficients and we can consider them as elements of Teichmüller classes as well.

Remark 7.3. We noted that the union of the lifts of the Teichmüller spaces of all finite unbranched coverings of the base surface to the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the compact solenoid \mathcal{S}_G is dense in $\mathcal{T}(\mathcal{S}_G)$. Moreover, if the covering surface S_1 is covered by another covering surface S_2 then $\mathcal{T}(S_1)$ embeds by isometry into $\mathcal{T}(S_2)$ (a consequence of the Teichmüller’s theorem for surfaces). One can consider a metric on $\mathcal{T}(\mathcal{S}_G)$ to be the “limit” metric of the Teichmüller metrics on the union of the Teichmüller spaces of finite coverings. The above theorem says that the Teichmüller metric on $\mathcal{T}(\mathcal{S}_G)$ (induced by taking the quasiconformal constants of the quasiconformal maps between the compact solenoids) agrees with the “limit” metric (because they agree on a dense subset). In particular, the extremal quasiconformal map between two TLC complex solenoids is given by the lift of the extremal maps between the surfaces (note that the Teichmüller class contains quasiconformal maps which are not lifts of maps between surfaces).

We note that Definition 5.8 is equivalent to the following definition of $\mathcal{T}(\mathcal{S}_G)$ because each leaf is dense in \mathcal{S}_G . Let G be a Fuchsian group such that \mathbb{D}/G is a closed surface and let G_n be the intersection of all subgroups of G of index at most n . Then G_n is a finite index characteristic subgroup of G .

Definition 7.4. The Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the compact solenoid \mathcal{S}_G is the space of all smooth Beltrami coefficients μ on the unit disk \mathbb{D} which are “almost invariant” under G , i.e., which satisfy

$$\sup_{A \in G_n} \|\mu - A^*(\mu)\|_\infty \rightarrow 0$$

as $n \rightarrow \infty$, up to the Teichmüller equivalence in the universal Teichmüller space $\mathcal{T}(\mathbb{D})$.

Remark 7.5. The proof of Theorem 7.1 uses the Reich–Strebel inequality in an essential way. It is important that we have a transverse measure m on \mathcal{S}_G in order to be able to integrate leafwise holomorphic quadratic differentials on \mathcal{S}_G . If we use

Definition 7.4 for $\mathcal{T}(\mathcal{S}_G)$, then the Teichmüller metric is defined in terms of the quasiconformal constants of the quasiconformal maps of the unit disk \mathbb{D} . If we consider a holomorphic quadratic differential φ on \mathbb{D} such that the Teichmüller type Beltrami coefficient $k \frac{|\varphi|}{\varphi}$ is almost invariant, then it seems difficult to directly show that it is extremal among all equivalent almost invariant Beltrami coefficients. Thus, even though Definition 7.4 is simpler than Definition 5.8, it seems beneficial to work with the later definition when studying extremal maps.

Any TLC Beltrami coefficient $\tilde{\mu}$ on \mathcal{S}_G is a lift of a Beltrami coefficient μ on a closed Riemann surface S_n in the tower of Riemann surfaces defining a TLC complex structure of \mathcal{S} which is possibly different from the fixed TLC structure obtained from \mathbb{D}/G . By Teichmüller's theorem for closed surfaces, there exists $0 < k < 1$ and $\varphi \in A(S_n)$ such that $k \frac{|\varphi|}{\varphi} \in [\mu]$. Then φ lifts to a TLC holomorphic quadratic differential $\tilde{\varphi} \in A(\mathcal{S}_G)$ and $k \frac{|\tilde{\varphi}|}{\tilde{\varphi}} \in [\tilde{\mu}]$. By the above theorem, we get immediately that $d_T([\tilde{\mu}], [0]) = d_T([\mu], [0])$. In other words [42],

Corollary 7.6. *Let S be a closed Riemann surface such that the TLC complex structure on the compact solenoid \mathcal{S}_G can be obtained by lifting the complex structure of S . Then the natural inclusion map*

$$i: \mathcal{T}(S) \rightarrow \mathcal{T}(\mathcal{S}_G)$$

obtained by mapping Beltrami coefficients on S to their lifts on \mathcal{S}_G is an isometry for the Teichmüller metrics.

The Teichmüller space $\mathcal{T}(S)$ of a closed surface S is a finite-dimensional complex manifold. Any two points $[f: S \rightarrow S_1]$ and $[g: S \rightarrow S_2]$ in $\mathcal{T}(S)$ are connected by a unique Teichmüller-type geodesic path $t \mapsto [t \frac{|\varphi|}{\varphi}]$, $0 \leq t \leq k$, for $\varphi \in A(S_1)$ and some $0 < k < 1$.

On the other hand, the Teichmüller space $\mathcal{T}(\mathbb{D})$ of the unit disk \mathbb{D} is an infinite-dimensional non-separable complex Banach manifold. There are points in $\mathcal{T}(\mathbb{D})$ which are not connected by a Teichmüller-type geodesic path. However, Lacic [22] observed that an open, dense subset of $\mathcal{T}(\mathbb{D})$ is connected by a Teichmüller-type geodesic to the basepoint $[0] \in \mathcal{T}(\mathbb{D})$.

The Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the universal hyperbolic solenoid \mathcal{S}_G is also an infinite-dimensional complex Banach manifold, but it is separable. This is the first example of a separable Teichmüller space which is the “smallest” possible infinite-dimensional space. Moreover, even though each leaf is non-compact, the solenoid \mathcal{S}_G is a compact space. In addition, the union of lifts of Teichmüller spaces of all closed surfaces of genus at least two is dense in $\mathcal{T}(\mathcal{S}_G)$ (see Nag–Sullivan [32] or [42]) and we showed in the above corollary that each such point is connected to the basepoint by a Teichmüller-type geodesic path. Based on the above remarks, one would hope that each point in $\mathcal{T}(\mathcal{S}_G)$ is connected by a Teichmüller-type geodesic to the basepoint. If not, at least one would expect this to be true for a large subset of $\mathcal{T}(\mathcal{S}_G)$. However, the situation for $\mathcal{T}(\mathcal{S}_G)$ is unexpectedly different (see [14]).

Theorem 7.7. *The set of points in the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the compact solenoid \mathcal{S}_G which do not have a Teichmüller-type extremal representative is generic in $\mathcal{T}(\mathcal{S}_G)$. That is, the set of points that do have a Teichmüller-type representative is of the first kind in the sense of Baire with respect to the Teichmüller metric.*

Proof. For the benefit of the reader, we give a short description of the ideas involved in the proof. The key idea is to exploit the difference between the L^1 -norm and the Bers norm on the space of transversely continuous holomorphic quadratic differentials $A(\mathcal{S}_G)$ on \mathcal{S}_G . In particular, $A(\mathcal{S}_G)$ is complete for the Bers norm and incomplete for the L^1 -norm.

We sketch the proof that there exist points in $\mathcal{T}(\mathcal{S}_G)$ which do not have a Teichmüller-type Beltrami coefficient representatives. The proof that they are generic is just an easy modification.

Assume on the contrary that all points in $\mathcal{T}(\mathcal{S}_G)$ have Teichmüller-type representatives. Let $A_1 = \{\varphi \in A(\mathcal{S}_G); \|\varphi\|_{L^1} = 1\}$ and let $A_1(N) = \{\varphi \in A_1; \|\varphi\|_{\text{Bers}} \leq N\}$, where $N \in \mathbb{N}$. Then $A_1 = \bigcup_{N=1}^{\infty} A_1(\mathcal{X})(N)$. We define a map

$$\pi : \mathcal{T}(\mathcal{S}_G) \rightarrow A_1 \cup \{0\}, \quad \pi([0]) = 0 \text{ and } \pi([\mu]) = \varphi \text{ if } [\mu] \neq [0],$$

where $k \frac{|\varphi|}{\varphi} \in [\mu]$ and φ is normalized such that $\|\varphi\|_{L^1(\mathcal{X})} = 1$. Then

$$\mathcal{T}(\mathcal{S}_G) = \bigcup_{N=1}^{\infty} \pi^{-1}(A_1(N)) \cup \{[0]\}.$$

We recall (see [14, Proposition 4.2]) that each $\pi^{-1}(A_1(N)) \cup [0]$ is closed in $\mathcal{T}(\mathcal{S}_G)$ under our assumption above. To see this, note that if $[k_n \frac{|\varphi_n|}{\varphi_n}] \rightarrow [k \frac{|\varphi|}{\varphi}]$ then $k_n \rightarrow k$ and $\int_{\mathcal{S}_G} \frac{|\varphi_n|}{\varphi_n} \varphi \, dm \rightarrow 1$ as $n \rightarrow \infty$, by the Reich–Strebel inequality. We assume that $[k_n \frac{|\varphi_n|}{\varphi_n}] \in \pi^{-1}(A_1(N))$. Then $\int_{\mathcal{S}_G} \frac{|\varphi_n|}{\varphi_n} \varphi \, dm \rightarrow 1$ implies that $\varphi \in A_1(N)$, i.e., $[k \frac{|\varphi|}{\varphi}] \in \pi^{-1}(A_1(N))$.

This implies that at least one $\pi^{-1}(A_1(N))$ is of the second kind in the sense of Baire and hence it has an interior. We obtain a contradiction by showing that each $\pi^{-1}(A_1(N))$ is nowhere dense, hence is of the first kind in the sense of Baire.

The rest of the proof depends on a geometric construction. Assume that $\pi^{-1}(A_1(N))$ has an interior. Let $[\mu]$ be a TLC point in the interior, which is equivalent to $k \frac{|\tilde{\varphi}|}{\tilde{\varphi}}$, where $\tilde{\varphi}$ is a lift of a holomorphic quadratic differential φ on a closed Riemann surface S . Denote by S_b a surface obtained by cutting S along a non-separating simple closed geodesic b . We consider a \mathbb{Z}_n -cover S_n of S obtained by cyclically gluing n copies of the surface S_b . Let $0 < r < 1$ and denote by $S_{n,r}$ the $[rn]/n$ portion of S_n which is made out of $[rn]$ neighboring copies of S_b , where $[rn]$ is the greatest integer which is at most rn . The boundary of $S_{n,r}$ consists of two curves which are copies of b . Let φ_n be a quadratic differential on S_n obtained by lifting φ on the $S_{n,r}$ part and defining it to be zero on the $S_n \setminus S_{n,r}$ part. Let $\tilde{\varphi}_n$ be the lifted quadratic differential to \mathcal{S}_G . Note that φ_n and $\tilde{\varphi}_n$ are piecewise holomorphic. It turns out that $\tilde{\varphi}_n$ can be approximated

by holomorphic quadratic differential $\tilde{\psi}_n$ on \mathcal{S}_G in the L^1 -norm such that $\tilde{\psi}_n$ is a lift of a holomorphic quadratic differential ψ_n on S_n (see [14, Lemma 4.3]).

Let $\tilde{S}_{n,r}$ denote the pre-image of $S_{n,r}$ in the solenoid \mathcal{S}_G . Then $\alpha(\tilde{S}_{n,r}) = [nr]/n$, where α is the product measure of the leafwise hyperbolic area measure and the transverse Haar measure m multiplied by an appropriate constant such that $\alpha(\mathcal{S}_G) = 1$. We keep the notation $[\mu]$ for the fixed TLC point in the interior of $\pi^{-1}(A_1(N))$. Let v_n on \mathcal{S}_G be defined by $v_n = (1+r)k\frac{|\tilde{\varphi}|}{\tilde{\varphi}}$ on $\tilde{S}_{n,r}$ and $v_n = k\frac{|\tilde{\varphi}|}{\tilde{\varphi}}$ on $\mathcal{S}_G - \tilde{S}_{n,r}$. By the Reich–Strebel inequality, the Beltrami coefficient v_n when considered as a functional on $A(\mathcal{S}_G)$ is close to achieving its norm on a holomorphic quadratic differential $\tilde{\psi}_n^*$ which is “similar” to $\tilde{\psi}_n$ in the L^1 sense. More precisely, the integral of $|\tilde{\psi}_n^*|$ when coupled with the Haar measure over $\mathcal{S}_G - \tilde{S}_{n,r}$ is converging to zero as $n \rightarrow \infty$.

If r is small enough then $[v_n] \in \pi^{-1}(A_1(N))$. This implies that $\tilde{\psi}_n^*$ is in $A_1(N)$. This is a contradiction with

$$1 = \int_{\mathcal{S}_G} |\tilde{\psi}_n^*| dm \leq \|\tilde{\psi}_n^*\|_{\text{Bers}} \alpha(\tilde{S}_{n,r}) + \int_{\mathcal{S}_G - \tilde{S}_{n,r}} |\tilde{\psi}_n^*| dm$$

because the right side can be made arbitrary small for n large and r small enough. Therefore, our starting assumption that all points have Teichmüller-type extremal representatives is not correct.

To show the stronger statement that the set of points which have Teichmüller-type Beltrami coefficient representatives is of the first kind, it is enough to assume that it is of the second kind and use this set instead of the whole $\mathcal{T}(\mathcal{S}_G)$ in the above argument. \square

We recall that each point $[\mu] \in \mathcal{T}(\mathcal{S}_G)$ is approximated by a sequence $[\mu_n] \in \mathcal{T}(\mathcal{S}_G)$ of TLC points. Each μ_n is Teichmüller equivalent to a unique Teichmüller-type Beltrami coefficient $k_n \frac{|\varphi_n|}{\varphi_n}$, where φ_n is a TLC holomorphic quadratic differential on \mathcal{S}_G . We say that $[\mu]$ is *well-approximated* by the TLC sequence $[\mu_n]$ if

$$\sum_{n=1}^{\infty} \|k_n \varphi_n - k_{n+1} \varphi_{n+1}\|_{\text{Bers}} < \infty.$$

Theorem 7.8. *If a non locally transversely constant point in $\mathcal{T}(\mathcal{S}_G)$ is well-approximated by transversely locally constant points then it contains a Teichmüller-type extremal Beltrami coefficient representative.*

Remark 7.9. We note that the above two theorems have counterparts in the infinitesimal setting. Namely, a generic vector in the tangent space at the basepoint $[0] \in \mathcal{T}(\mathcal{S}_G)$ does not achieve its norm on $A(\mathcal{S}_G)$ (when considered as a linear functional on $A(\mathcal{S}_G)$), namely it cannot be represented by a Teichmüller-type Beltrami coefficient $k\frac{|\varphi|}{\varphi}$, for $k > 0$ and $\varphi \in A(\mathcal{S}_G)$ (see [14, Theorem 3]). A well-approximated non TLC vector in the tangent space at the basepoint $[0] \in \mathcal{T}(\mathcal{S}_G)$ does achieve its norm on $A(\mathcal{S}_G)$ (see [14, Theorem 2’]).

The set of real numbers which are not well-approximated by rational numbers is of full Lebesgue measure on the real line. From Theorem 7.7 and Theorem 7.8, we immediately obtain a similar statement for well-approximation with TLC points in $\mathcal{T}(\mathcal{S}_G)$ (see [14]).

Corollary 7.10. *The set of points in $\mathcal{T}(\mathcal{S}_G)$ which are not well-approximated by transversely locally constant marked complex structures is generic in $\mathcal{T}(\mathcal{S}_G)$.*

8 The modular group of the compact solenoid

The following definition was given by C. Odden [38]:

Definition 8.1. The *modular group* $\text{Mod}(\mathcal{S}_G)$ consists of all quasiconformal self-maps of \mathcal{S}_G which preserve the baseleaf up to isotopy.

The modular group $\text{Mod}(\mathcal{S}_G)$ acts on the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ by

$$[f : \mathcal{S}_G \rightarrow \mathcal{X}] \mapsto [f \circ g^{-1} : \mathcal{S}_G \rightarrow \mathcal{X}],$$

where $[g : \mathcal{S}_G \rightarrow \mathcal{S}_G] \in \text{Mod}(\mathcal{S}_G)$ and $[f] \in \mathcal{T}(\mathcal{S}_G)$.

Definition 8.2 (see [5], [38]). A *partial automorphism* of the fundamental group $G = \pi_1(S_0)$ is an isomorphism between two finite index subgroups of G . Two partial automorphisms $\psi_1 : K_1 \rightarrow H_1$ and $\psi_2 : K_2 \rightarrow H_2$ are said to be *equivalent* if they agree on the intersection of their domains. The *virtual automorphism group* $\text{Vaut}(G)$ of the surface group G is by definition the group of equivalence classes of partial automorphisms. Note that the virtual automorphism group is also called the (abstract) *commensurator group* $\text{Comm}(G)$ of the surface group G and we use this notation in the rest of the chapter.

In [5], a natural group in which each element is given by two non-isomorphic pointed covers of the same degree of the base surface (S_0, x_0) is shown to act on the union of Teichmüller spaces of all closed surfaces of genus at least two, namely the subset of $\mathcal{T}(\mathcal{S}_G)$ consisting of all TLC points. The above group is naturally isomorphic to the commensurator group $\text{Comm}(G)$ of the surface group $G = \pi_1(S_0)$. The action is isometric for the Teichmüller distance on the union of Teichmüller spaces of all closed surfaces of genus at least two and it extends by continuity to the action on the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$. One should note that our definition of the Teichmüller metric on $\mathcal{T}(\mathcal{S}_G)$ does not guarantee that the above union embeds isometrically in $\mathcal{T}(\mathcal{S}_G)$; this is a consequence of the Reich–Strebel theorem for \mathcal{S}_G (see Corollary 7.6). However, we do not need to use Corollary 7.6 to show that a continuous extension is possible; it is enough to note that the Teichmüller metric on the above union is bi-Lipschitz (with constant 1/3) to the Teichmüller metric on $\mathcal{T}(\mathcal{S}_G)$ (which is a

consequence of the standard result comparing the Teichmüller metric on Teichmüller space of a Riemann surface with the restriction of Teichmüller metric of the universal Teichmüller space $\mathcal{T}(\mathbb{D})$ to its embedding into $\mathcal{T}(\mathbb{D})$ due to McMullen [30], [18]).

The following theorem (see [38]) gives a natural interpretation of the commensurator group $\text{Comm}(G)$ in terms of the solenoid.

Theorem 8.3. *Let \mathcal{S}_G be the G -tagged compact solenoid. Fix an identification of the baseleaf of \mathcal{S}_G with \mathbb{D} . Then the modular group $\text{Mod}(\mathcal{S}_G)$ is isomorphic to the commensurator group $\text{Comm}(G)$ of the base surface group G . The isomorphism is given by the restriction of $\text{Mod}(\mathcal{S}_G)$ to the baseleaf.*

The group of baseleaf preserving conformal maps of \mathcal{S}_G (which is a subgroup of the modular group $\text{Mod}(\mathcal{S}_G)$) is identified with the commensurator group $\text{Comm}_{\text{PSL}_2(\mathbb{R})}(G)$ of G in $\text{PSL}_2(\mathbb{R})$ [38], where $\text{Comm}_{\text{PSL}_2(\mathbb{R})}(G)$ consists of all $M \in \text{PSL}_2(\mathbb{R})$ for which there exist two finite index subgroups K and H of G such that $MKM^{-1} = H$. There are two cases, either G is an arithmetic group in which case $\text{Comm}_{\text{PSL}_2(\mathbb{R})}(G)$ is dense in $\text{PSL}_2(\mathbb{R})$ or G is not arithmetic in which case $\text{Comm}_{\text{PSL}_2(\mathbb{R})}(G)$ is a finite extension of G . In both cases, the group of conformal maps of the G -tagged solenoid $(\mathbb{D} \times \hat{G})/G$ is infinite (because it contains G in both cases), unlike for Riemann surfaces where it is finite. (Note that G acts non-trivially on $\mathcal{T}(\mathcal{S}_G)$ even though it acts trivially on $\mathcal{T}(\mathbb{D}/G)$.) Biswas and Nag [4] showed that the action of $\text{Comm}_{\text{PSL}_2(\mathbb{R})}(G)$ on the G -tagged solenoid is ergodic (with respect to the product of the hyperbolic area measure on leaves and the transverse measure) if and only if G is arithmetic. For any Fuchsian uniformizing group G of a closed Riemann surface, a G -tagged solenoid represents the lift of the complex structure on \mathbb{D}/G to \mathcal{S}_G . Thus, the isotropy group (in $\text{Mod}(\mathcal{S}_G)$) of a marked TLC point in $\mathcal{T}(\mathcal{S}_G)$ is always infinite. We showed [28] that the isotropy group of any non-TLC point in $\mathcal{T}(\mathcal{S}_G)$ is infinite as well. The basic idea was to show that the right action of the conformal covering group $G_{\mathcal{X}}$ for a non-TLC solenoid \mathcal{X} commutes with the left action of $G_{\mathcal{X}}$.

If a sequence of homeomorphisms of a closed surface converges uniformly on compact subsets to the identity, then the elements of its tail are isotopic to the identity. In [28] we showed a corresponding statement for the solenoid \mathcal{S}_G .

Theorem 8.4. *Let \mathcal{S}_G be a TLC complex solenoid and let $f_n: \mathcal{S}_G \rightarrow \mathcal{S}_G$ be a sequence of baseleaf preserving quasiconformal self maps of \mathcal{S}_G that uniformly converges to the identity map. Then there exists n_0 such that f_n is homotopic to a baseleaf preserving conformal self map $c_n: \mathcal{S}_G \rightarrow \mathcal{S}_G$, for all $n > n_0$.*

A classical result on closed surfaces states that any two homeomorphisms which are homotopic are isotopic. Moreover, any two quasiconformal maps of two Riemann surfaces (possibly geometrically infinite) which are homotopic through bounded homotopy are isotopic through bounded quasiconformal isotopy, namely the quasiconformal constants of maps in the isotopy are uniformly bounded. We showed [28] similar result for the solenoid.

Theorem 8.5. *Let $f: X \rightarrow Y$ and $g: X \rightarrow Y$ be two homotopic quasiconformal maps of complex solenoids X and Y . Then f and g are isotopic by a uniformly quasiconformal isotopy.*

We also considered the orbits of $\text{Mod}(\mathcal{S}_G)$ in $\mathcal{T}(\mathcal{S}_G)$. It is an observation of Sullivan that the Ehrenpreis conjecture is equivalent to the statement that orbits of $\text{Mod}(\mathcal{S}_G)$ are dense. In a joint work with Markovic, we showed the weaker statement that orbits have accumulation points [28].

Theorem 8.6. *There exists a dense subset of $\mathcal{T}(\mathcal{S}_G)$ such that the orbit of the modular group $\text{Mod}(\mathcal{S}_G)$ of any point in this subset has accumulation points in $\mathcal{T}(\mathcal{S}_G)$. This subset contains only non-TLC points.*

An element h of $\text{Mod}(\mathcal{S}_G)$ is called *mapping class like* if h conjugates a finite index subgroup K of the base surface group G onto itself, i.e., $hKh^{-1} = K$. C. Odden [38] showed that if a power h^n , $n \neq 0$, is mapping class like then h is mapping class like.

The Nielsen realization problem states that any finite subgroup of the modular group of a closed surface is realized as a conformal group of a homeomorphic Riemann surface. We showed in [28] a version of the Nielsen realization problem for the solenoid \mathcal{S}_G .

Theorem 8.7. *Any finite subgroup of $\text{Mod}(\mathcal{S}_G)$ is cyclic and mapping class like. Consequently, elements of $\text{Mod}(\mathcal{S}_G)$ which are not mapping class like are of infinite order.*

9 The Teichmüller space of the non-compact solenoid

Let $G < \text{PSL}_2(\mathbb{Z})$ be such that \mathbb{D}/G is the once punctured modular torus (which we shall henceforth call the “modular torus”).

Definition 9.1. The G -tagged non-compact solenoid \mathcal{S}_{nc} is the quotient of $\mathbb{D} \times \hat{G}$ by the action of G , where $A(z, t) := (Az, tA^{-1})$ for $(z, t) \in \mathbb{D} \times \hat{G}$ and $A \in G$. The base leaf of \mathcal{S}_{nc} is $(\mathbb{D} \times \{\text{id}\})/G$.

Definition 9.2. An arbitrary non-compact marked complex solenoid is a complex solenoid \mathcal{X} together with a differentiable, quasiconformal map $f: \mathcal{S}_{\text{nc}} \rightarrow \mathcal{X}$ which is continuous in the transverse direction in the local charts for the C^1 -topology, and whose leafwise Beltrami coefficients are continuous in the transverse direction for the essential supremum norm when nearby leaves are identified using the canonical identifications coming from the G -tagged TLC complex structure of \mathcal{S}_{nc} .

The requirement that Beltrami coefficients are close on the whole leaves as opposed to being close in local charts is necessary because \mathcal{S}_{nc} is non-compact. For marked

compact solenoids we obtain the same property from the continuity in local charts because of the compactness. In a joint work with R. Penner we introduced in [37] the Teichmüller space $\mathcal{T}(\mathcal{S}_{nc})$ of the non-compact solenoid \mathcal{S}_{nc} as follows.

Definition 9.3. The Teichmüller space $\mathcal{T}(\mathcal{S}_{nc})$ of the non-compact solenoid \mathcal{S}_{nc} is the space of all differentiable, quasiconformal maps $f: \mathcal{S}_{nc} \rightarrow \mathcal{X}$ from the G -tagged solenoid to an arbitrary non-compact complex solenoid \mathcal{X} up to conformal maps of the range and up to homotopy, where f is required to be continuous in the transverse direction in the local charts in the C^1 -topology and the leafwise Beltrami coefficients of f are required to vary continuously on the global leaves in the essential supremum norm when leaves are canonically identified using the G -tagged complex structure of \mathcal{S}_{nc} .

The definition of $\mathcal{T}(\mathcal{S}_{nc})$ is justified by the following density theorem analogous to the finite surface case (see [37]).

Theorem 9.4. *The union of the lifts of the Teichmüller spaces of all finite punctured hyperbolic surfaces covering the modular torus is dense in the Teichmüller space $\mathcal{T}(\mathcal{S}_{nc})$ of the non-compact solenoid \mathcal{S}_{nc} .*

In [37] we introduced a representation definition of the Teichmüller space $\mathcal{T}(\mathcal{S}_{nc})$ as follows. Consider the space $\text{Hom}(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))$ of all functions $\rho: G \times \hat{G} \rightarrow \text{PSL}_2(\mathbb{R})$ satisfying the following three properties:

Property 1. ρ is continuous.

Property 2. (G -equivariance). For each $\gamma_1, \gamma_2 \in G$ and $t \in \hat{G}$, we have

$$\rho(\gamma_1 \circ \gamma_2, t) = \rho(\gamma_1, t\gamma_2^{-1}) \circ \rho(\gamma_2, t).$$

Property 3. For every $t \in \hat{G}$, there is a quasiconformal mapping $\phi_t: \mathbb{D} \rightarrow \mathbb{D}$ depending continuously on $t \in \hat{G}$ so that for every $\gamma \in G$, the following diagram commutes, where $\rho(\gamma, t) \circ \phi_t(z) = \phi_{t\gamma^{-1}} \circ \gamma(z)$:

$$\begin{array}{ccc}
 \mathbb{D} \times \hat{G} & \xrightarrow{(z,t) \mapsto (\gamma z, t\gamma^{-1})} & \mathbb{D} \times \hat{G} \\
 \downarrow \phi_t \times \text{id} & & \downarrow \phi_{t\gamma^{-1}} \times \text{id} \\
 \mathbb{D} \times \hat{G} & \xrightarrow{(\phi_t(z), t) \mapsto (\rho(\gamma, t) \circ \phi_t(z) = \phi_{t\gamma^{-1}} \circ \gamma(z), t\gamma^{-1})} & \mathbb{D} \times \hat{G}.
 \end{array}$$

Since G is discrete, ρ is continuous if and only if it is continuous in its second variable. Therefore, it is enough to require continuity in the second variable in Property 1. Property 2 is a kind of homomorphism property of ρ mixing the leaves; notice

in particular that taking $\gamma_2 = \text{id}$ gives $\rho(\text{id}, t) = \text{id}$ for all $t \in \hat{G}$. Property 3 mandates that for each $t \in \hat{G}$, ϕ_t conjugates the standard action of $\gamma \in G$ on $\mathbb{D} \times \hat{G}$ at the top of the diagram to the action

$$\gamma_\rho: (z, t) \mapsto (\rho(\gamma, t)z, t\gamma^{-1})$$

at the bottom, and we let $G_\rho = \{\gamma_\rho: \gamma \in G\} \approx G$. Notice that the action of G_ρ on $\mathbb{D} \times \hat{G}$ extends continuously to an action on $(\mathbb{D} \cup S^1) \times \hat{G}$. We finally define the solenoid (with marked hyperbolic structure)

$$\mathfrak{S}_\rho = (\mathbb{D} \times_\rho \hat{G}) = (\mathbb{D} \times \hat{G})/G_\rho.$$

Define the group $\text{Cont}(\hat{G}, \text{PSL}_2(\mathbb{R}))$ to be the collection of all continuous maps $\alpha: \hat{G} \rightarrow \text{PSL}_2(\mathbb{R})$, where the product of two $\alpha, \beta \in \text{Cont}(\hat{G}, \text{PSL}_2(\mathbb{R}))$ is taken pointwise $(\alpha\beta)(t) = \alpha(t) \circ \beta(t)$ in $\text{PSL}_2(\mathbb{R})$. An element $\alpha \in \text{Cont}(\hat{G}, \text{PSL}_2(\mathbb{R}))$ acts on $\rho \in \text{Hom}(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))$ according to

$$(\alpha\rho)(\gamma, t) = \alpha(t\gamma^{-1}) \circ \rho(\gamma, t) \circ \alpha^{-1}(t).$$

We introduced the topology on $\text{Hom}(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))$ as follows. Consider the natural metric d on $\text{PSL}_2(\mathbb{R})$ induced by identifying it with the unit tangent bundle of the unit disk \mathbb{D} . Let $\rho_1, \rho_2 \in \text{Hom}(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))$ and let $\gamma_1, \dots, \gamma_j \in G$ be a generating set of G . The distance between ρ_1 and ρ_2 is given by

$$\max_{1 \leq i \leq j, t \in \hat{G}} d(\rho_1(\gamma_i, t), \rho_2(\gamma_i, t)). \quad (9.1)$$

This metric is not canonical, but any such two metrics induce the same topology. Note that $\text{Hom}'(G \times \hat{G}, \text{PSL}_2(\mathbb{R})) := \text{Hom}(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))/\text{Cont}(\hat{G}, \text{PSL}_2(\mathbb{R}))$ is equipped with the quotient topology of the above topology on $\text{Hom}(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))$. We showed that $\text{Hom}'(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))$ is naturally homeomorphic to $\mathcal{T}(\mathfrak{S}_{\text{nc}})$ ([37]).

Theorem 9.5. *There is a natural homeomorphism of the Teichmüller space $\mathcal{T}(\mathfrak{S}_{\text{nc}})$ of the solenoid \mathfrak{S}_{nc} with*

$$\text{Hom}'(G \times \hat{G}, \text{PSL}_2(\mathbb{R})),$$

given by assigning to each $\rho \in \text{Hom}'(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))$ the corresponding marked hyperbolic solenoid \mathfrak{S}_ρ .

10 The decorated Teichmüller space of the non-compact solenoid

We introduced in [37] the decorated Teichmüller space $\tilde{\mathcal{T}}(\mathfrak{S}_{\text{nc}})$ of the punctured solenoid \mathfrak{S}_{nc} . Points in $\tilde{\mathcal{T}}(\mathfrak{S}_{\text{nc}})$ are decorations of (homotopy classes of) marked hyperbolic structures up to isometries. It is convenient to use the presentation definition of the Teichmüller space $\mathcal{T}(\mathfrak{S}_{\text{nc}})$ for assigning decorations to hyperbolic metrics.

Recall that a puncture on \mathcal{S}_{nc} is an end of a single leaf of \mathcal{S}_{nc} . Since \mathcal{S}_{nc} is a G -tagged solenoid with $G < \mathrm{PSL}_2(\mathbb{Z})$ the punctured torus group, an end has an explicit description in the universal cover $\mathbb{D} \times \hat{G}$. Denote by $\bar{\mathbb{Q}} \subset S^1$ the set of fixed points of the parabolic elements of G . Then the set of lifts to $\mathbb{D} \times \hat{G}$ of ends of \mathcal{S}_{nc} is identified with $\bar{\mathbb{Q}} \times \hat{G}$.

Given a quasiconformal map $f: \mathcal{S}_{nc} \rightarrow \mathcal{X}$, the images of the ends in \mathcal{S}_{nc} are the ends of \mathcal{X} . A decoration on \mathcal{X} is easiest to understand in terms of a presentation description $\rho \in \mathrm{Hom}(G \times \hat{G}, \mathrm{PSL}_2(\mathbb{R}))$. Let \mathcal{S}_ρ be a hyperbolic solenoid obtained from the representation ρ with corresponding quasiconformal map $\phi: \mathcal{S}_{nc} \rightarrow \mathcal{S}_\rho$.

In [37] we described the punctures of \mathcal{S}_ρ using the representation ρ . The quasiconformal map $\phi: \mathbb{D} \times \hat{G} \rightarrow \mathbb{D} \times \hat{G}$ extends continuously to a leafwise quasi-symmetric map $\phi: S^1 \times \hat{G} \rightarrow S^1 \times \hat{G}$. Recall that $\bar{\mathbb{Q}} \subset S^1$ parametrizes the endpoints of the standard triangulation of \mathbb{D} invariant under $\mathrm{PSL}_2(\mathbb{Z})$. We say that a point $(p, t) \in S^1 \times \hat{G}$ is a ρ -puncture if $\phi^{-1}(p, t) \in \bar{\mathbb{Q}}$, and a puncture of \mathcal{S}_ρ itself is the G_ρ -orbit of a ρ -puncture. A ρ -horocycle at a ρ -puncture (p, t) is the horocycle in $\mathbb{D} \times \{t\}$ centered at (p, t) and a horocycle on \mathcal{S}_ρ is the G_ρ -orbit of a ρ -horocycle.

We introduce an identification of horocycles with points in the light cone in Minkowski three space. Recall that Minkowski three space is \mathbb{R}^3 with the indefinite pairing $\langle \cdot, \cdot \rangle$ whose quadratic form is $x^2 + y^2 - z^2$ for $(x, y, z) \in \mathbb{R}^3$. The upper sheet of the hyperboloid $\mathbb{H} := \{w = (x, y, z); \langle w, w \rangle = -1, z > 0\}$ is a model for the hyperbolic plane and rays in the positive light cone $L^+ := \{u = (x, y, z) : \langle u, u \rangle = 0, z > 0\}$ are identified with boundary points to the hyperbolic plane. The hyperbolic distance between $w_1, w_2 \in \mathbb{H}$ is equal to $\cosh \langle w_1, w_2 \rangle$. The set of horocycles in \mathbb{H} is identified with points of the positive light cone L^+ by the duality $w \mapsto \{u \in \mathbb{H}; \langle w, u \rangle = -1\}$ (see [35]). A topology on the set of horocycles is induced by the correspondence with L^+ with its natural topology as a subset of \mathbb{R}^3 .

Definition 10.1. A decoration on \mathcal{S}_ρ , or a *decorated hyperbolic structure* on \mathcal{S}_ρ , is a function $\tilde{\rho}: G \times \hat{G} \times \bar{\mathbb{Q}} \rightarrow \mathrm{PSL}_2(\mathbb{R}) \times L^+$, where

$$\tilde{\rho}(\gamma, t, q) = \rho(\gamma, t) \times h(t, q)$$

with $\rho(\gamma, t) \in \mathrm{Hom}(G \times \hat{G}, \mathrm{PSL}_2(\mathbb{R}))$, which satisfies the following conditions:

Property 4. For each $t \in \hat{G}$, the image $h(t, \bar{\mathbb{Q}}) \subseteq L^+$ is discrete and the center of the horocycle $h(t, q)$ is $\phi_t(q)$, for all $(t, q) \in \hat{G} \times \bar{\mathbb{Q}}$ (using here the identification of L^+ with the space of horocycles).

Property 5. For each $q \in \bar{\mathbb{Q}}$, the restriction $h(\cdot, q): \hat{G} \rightarrow L^+$ is a continuous function from \hat{G} to L^+ .

Property 6. $h(t, q)$ is ρ invariant in the sense that

$$\rho(\gamma, t)(h(t, q)) = h(t\gamma^{-1}, \rho(\gamma, t)q).$$

We introduced in [37] the decorated Teichmüller space $\tilde{\mathcal{T}}(\mathcal{S}_{nc})$ as follows. Let $\mathrm{Hom}(G \times \hat{G} \times \bar{\mathbb{Q}}, \mathrm{PSL}_2(\mathbb{R}) \times L^+)$ denote the space of all decorated hyperbolic

structures satisfying the properties above. We define a topology on $\text{Hom}(G \times \hat{G} \times \bar{\mathbb{Q}}, \text{PSL}_2(\mathbb{R}) \times L^+)$. A neighborhood of $\tilde{\rho}(\gamma, t, q) = \rho(\gamma, t) \times h(t, q)$ consists of all $\tilde{\rho}_1(\gamma, t, q) = \rho_1(\gamma, t) \times h_1(t, q)$ such that ρ_1 belongs to a chosen neighborhood of ρ in $\text{Hom}(G \times \hat{G}, \text{PSL}_2(\mathbb{R}))$, and the maps $h_1(\cdot, q): \hat{G} \rightarrow L^+$ and $h(\cdot, q): \hat{G} \rightarrow L^+$ are close in the supremum norm, for each $q \in \bar{\mathbb{Q}}$. The above condition and the invariance Property 6 implies that the set $h_1(t, \bar{\mathbb{Q}})$ is close to the set $h(t, \bar{\mathbb{Q}})$ in the Hausdorff metric on the closed subsets of \mathbb{R}^3 , for each $t \in \hat{G}$.

Definition 10.2. The decorated Teichmüller space $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ is the quotient

$$\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}}) := \text{Hom}(G \times \hat{G} \times \bar{\mathbb{Q}}, \text{PSL}_2(\mathbb{R}) \times L^+) / \text{Cont}(\hat{G}, \text{PSL}_2(\mathbb{R})),$$

where $\alpha: \hat{G} \rightarrow \text{PSL}_2(\mathbb{R})$ acts on $\tilde{\rho}$ by

$$(\alpha\tilde{\rho})(\gamma, t, q) = (\alpha(t\gamma^{-1}) \circ \rho(\gamma, t) \circ \alpha^{-1}(t)) \times (\alpha(t)h(t, q)).$$

It is immediate that the forgetful map $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}}) \rightarrow \mathcal{T}(\mathcal{S}_{\text{nc}})$ is a continuous surjection (see [37, Proposition 5.2]).

Given two horocycles in the hyperbolic plane, consider a geodesic connecting their centers. The horocycles intersect the geodesic at two points and the *lambda length* of the pair is defined as $\sqrt{2 \exp \delta}$, where δ is the signed length of the arc of the geodesic between the two points (see [35], [36]). The sign of δ is positive if the arc is outside the horoballs and it is negative if the arc is inside the horoballs. If $u, v \in L^+$ represent the horocycles then the lambda length is given by $\lambda(u, v) = \sqrt{-\langle u, v \rangle}$.

Let τ_* be the Farey tessellation of the unit disk (see, for example, [34], [37], [36]). Then the vertices of τ_* are at $\bar{\mathbb{Q}}$ and $\tau_* \times \hat{G}$ is a tessellation of the universal cover $\mathbb{D} \times \hat{G}$ of \mathcal{S}_{nc} . Given a decoration $\tilde{\rho} = (\rho, h) \in \tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$, we consider the image tessellation $\phi(\tau_* \times \hat{G})$ of the universal cover $\mathbb{D} \times \hat{G}$ of \mathcal{S}_ρ (where ϕ is the union of quasiconformal maps from Property 3). Then there is an assignment of lambda length $\lambda(e, t)$ to each edge (e, t) , $e \in \tau_*$ and $t \in \hat{G}$, in the tessellation $\tau_* \times \hat{G}$ by

$$\lambda(e, t) = \lambda(h(p, t), h(q, t)),$$

where p, q are the endpoints of e . Thus we obtain a lambda length map $\lambda: \tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}}) \rightarrow (\mathbb{R}_{>0}^{\tau_*})^{\hat{G}}$, where $(\mathbb{R}_{>0}^{\tau_*})^{\hat{G}}$ are maps from \hat{G} into the function space $\mathbb{R}_{>0}^{\tau_*}$ (see [37]). We consider the supremum norm over edges in τ_* on the function space $\mathbb{R}_{>0}^{\tau_*}$. Let $\text{Cont}(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$ be the space of continuous functions in the compact-open topology. In other words,

$$f \in \text{Cont}(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$$

if

$$\sup_{e \in \tau_*} |f(t)(e) - f(t_1)(e)| \rightarrow 0$$

as $t \rightarrow t_1$, for all $t_1 \in \hat{G}$. Moreover, we define $\text{Cont}^G(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$ to be the set of G -invariant functions f in $\text{Cont}(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$, i.e., $f \in \text{Cont}^G(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$ if $f \in \text{Cont}(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$

and

$$f(tA^{-1})(A(e)) = f(t)(e),$$

for $A \in G$ and $t \in \hat{G}$. In [37] we obtain

Theorem 10.3. *The assignment of lambda lengths*

$$\lambda : \tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}}) \rightarrow \text{Cont}^G(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$$

is a surjective homeomorphism. Namely, $\text{Cont}^G(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$ parametrizes the decorated Teichmüller space $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$.

A direct corollary to the above theorem is (see [37])

Corollary 10.4. *The union of the lifts of the decorated Teichmüller spaces of all finite punctured surfaces covering the modular torus is dense in the decorated Teichmüller space $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ of the non-compact solenoid \mathcal{S}_{nc} .*

We consider the convex hull construction introduced in [15] and further utilized in [35] for punctured surfaces and in [34] for the universal Teichmüller space. The construction in [35] gives a decomposition of the decorated Teichmüller space of a punctured surface similar to [20]. Our approach is based on the universal Teichmüller space construction from [34] where the construction of [20] does not work.

A lambda length function $f \in \mathbb{R}_{>0}^{\tau_*}$ is said to be *pinched* if there exists $M > 1$ such that

$$1/M \leq f(e) \leq M$$

for all $e \in \tau_*$ (see [34]). Let $h \in (L^+)^{\bar{\mathbb{Q}}}$ and assume that the corresponding lambda length function $\lambda : e \mapsto \lambda(h(e))$, $e \in \tau_*$, is pinched. Consider the image $h(\bar{\mathbb{Q}}) \subset L^+$ of h and let $\mathcal{C}(h(\bar{\mathbb{Q}}))$ denote its convex hull as a subset of \mathbb{R}^3 . Then the results from [34] give that $h(\bar{\mathbb{Q}})$ is a discrete and radially dense subset of L^+ . Moreover, $h : \bar{\mathbb{Q}} \rightarrow L^+$ projects to a map $\bar{h} : \bar{\mathbb{Q}} \rightarrow S^1$ which extends to a quasisymmetric homeomorphism of S^1 . In addition, the set of faces of the boundary $\partial \mathcal{C}(h(\bar{\mathbb{Q}}))$ of the convex hull $\mathcal{C}(h(\bar{\mathbb{Q}}))$ consists of Euclidean polygons which meet along their boundary edges, the set of faces is locally finite and boundary edges of faces of $\partial \mathcal{C}(h(\bar{\mathbb{Q}}))$ project to a locally finite geodesic lamination on the hyperbolic plane \mathbb{H} whose geodesics have endpoints in $\bar{\mathbb{Q}}$ (see [34] for more details and proofs).

A decoration $\tilde{\rho} \in \tilde{\mathcal{S}}(\mathcal{S}_{\text{nc}})$ of the non-compact solenoid \mathcal{S}_{nc} gives a lambda length function $\lambda(\tilde{\rho}) \in \text{Cont}^G(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$. Namely, we obtain a Cantor set of lambda lengths $\lambda(\tilde{\rho})(t) : \tau_* \rightarrow \mathbb{R}_{>0}$, for $t \in \hat{G}$, and note that the lambda lengths are pinched uniformly in $t \in \hat{G}$ by the compactness of \hat{G} and the transverse continuity of $\tilde{\rho}$ (see [34, Lemma 6.1]). The above convex hull construction applied to each leaf $\mathbb{D} \times \{t\}$ of the universal cover $\mathbb{D} \times \hat{G}$ gives a Cantor set of convex hulls which in turn produce a Cantor set of geodesic laminations on $\mathbb{D} \times \hat{G}$ which are invariant under the action of G .

Denote by $\tau_{\tilde{\rho}}$ such obtained leafwise geodesic lamination on $\mathbb{D} \times \hat{G}$. The endpoints of geodesics in $\tau_{\tilde{\rho}}$ lie in $\overline{\mathbb{Q}} \times \hat{G}$ and we call such a geodesic lamination a *tesselation* if all complementary regions are ideal triangles. In general, the complementary regions of $\tau_{\tilde{\rho}}$ on leaves can be arbitrary ideal hyperbolic polygons. A tesselation τ of $\mathbb{D} \times \hat{G}$ such that the restriction to each leaf $\tau(t) \subset \mathbb{D}$, $t \in \hat{G}$, is invariant under some finite index subgroup K of G is called a *TLC tesselation*. Equivalently, τ is a TLC tesselation of $\mathbb{D} \times \hat{G}$ if it is a lift of a tesselation on a Riemann surface \mathbb{D}/K , for some finite index subgroup $K < G$.

Definition 10.5. Let τ be a leafwise geodesic lamination on $\mathbb{D} \times \hat{G}$. Denote by $C(\tau)$ the set of all decorations for which the convex hull construction produces λ , i.e.,

$$C(\tau) := \{\tilde{\rho} \in \tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}}); \partial\mathcal{C}(\lambda(\tilde{\rho})) = \tau\}.$$

We showed in [37] that generically in $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ convex hull constructions yield TLC tesselations. In more details,

Theorem 10.6. *The subset $C(\tau)$ of $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ is open for each TLC tesselation τ , and $\bigcup_{\tau} C(\tau)$ is a dense open subset of $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$, where the union is over all TLC tesselations τ .*

11 A presentation for the modular group of the non-compact solenoid

We define the *modular group* $\text{Mod}(\mathcal{S}_{\text{nc}})$ of the non-compact solenoid \mathcal{S}_{nc} to consist of (analogously to the compact solenoid) all quasiconformal differentiable baseleaf preserving self-maps of \mathcal{S}_{nc} up to isotopy (see [37]). We showed in [37] an appropriate version of the characterization of $\text{Mod}(\mathcal{S}_{\text{nc}})$ similar to the compact solenoid (see [38]). As in Section 9, let $G < \text{PSL}_2(\mathbb{Z})$ be the once punctured modular torus group. Recall that the commensurator of G consists of all isomorphisms of finite-index subgroups of G modulo a natural equivalence (see Section 8, Definition 8.2).

Theorem 11.1. *The restriction to the baseleaf of $\text{Mod}(\mathcal{S}_{\text{nc}})$ gives an isomorphism of $\text{Mod}(\mathcal{S}_{\text{nc}})$ with the subgroup of the commensurator group of G consisting of elements which map parabolic elements of G onto parabolic elements of G .*

The action of the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ on the decorated Teichmüller space $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ preserves the decomposition into sets $C(\tau)$, for τ a leafwise geodesic lamination on the solenoid, or equivalently a G -invariant geodesic lamination on the universal cover. It is convenient to consider TLC tesselations only. Then, as a consequence of the above theorem, the modular group preserves the subspace of TLC tesselations. We showed in [37] that an analogue of the Ehrenpreis conjecture in the decorated Teichmüller space $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ is not correct.

Theorem 11.2. *The quotient $\bigcup_{\tau} C(\tau)/\text{Mod}(\mathcal{S}_{\text{nc}})$ is Hausdorff, where the union is over all TLC tessellations τ . Moreover, no orbit under $\text{Mod}(\mathcal{S}_{\text{nc}})$ of a point in $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$ is dense in $\tilde{\mathcal{T}}(\mathcal{S}_{\text{nc}})$.*

From now on, we restrict the action of $\text{Mod}(\mathcal{S}_{\text{nc}})$ to the baseleaf. Then $\text{Mod}(\mathcal{S}_{\text{nc}})$ preserves the space of all TLC tessellations on \mathbb{D} , i.e., it preserves the space of lifts of all ideal hyperbolic triangulations of all Riemann surfaces \mathbb{D}/K , where $K < G$ is of finite index. The Farey tessellation τ_* on \mathbb{D} is a TLC tessellation which will be considered as a basepoint in our considerations.

We showed a transitivity statement for the family of TLC tessellations, or equivalently for the family $\{C(\tau)\}_{\tau}$, where τ belongs to all TLC tessellations (see [37]).

Theorem 11.3. *$\text{Mod}(\mathcal{S}_{\text{nc}})$ acts transitively on $\{C(\tau) : \tau \text{ is TLC}\}$.*

Proof. We give a brief description of the proof. It is enough to show that any TLC tessellation of the unit disk \mathbb{D} is mapped onto any other TLC tessellation of \mathbb{D} by a homeomorphism of S^1 which conjugates one finite index subgroup of G onto another finite index subgroup of G . Such a homeomorphism of S^1 induces an element of the commensurator group of G which preserves parabolics, and conversely any element of the commensurator group of G which preserves parabolics is induced by a homeomorphism of S^1 . Recall that a TLC tessellation of \mathbb{D} is a lift of an ideal triangulation of \mathbb{D}/K , where $K < G$ is of finite index. In particular, the set of ideal vertices of the lifted TLC tessellation of \mathbb{D} is $\bar{\mathbb{Q}}$ and the tessellation is K -invariant.

Moreover, it is enough to show that the Farey tessellation τ_* can be mapped by a homeomorphism inducing a parabolics-preserving element of the commensurator group of G onto any other TLC tessellation of \mathbb{D} . Let τ be an arbitrary TLC tessellation of \mathbb{D} which is invariant under a finite index subgroup K of G . We define a *characteristic map* for τ by giving an identification of the edges of τ_* and τ as follows (see [34]). We choose the edge e_0 of τ_* which joins -1 and 1 and orient it from -1 to 1 . Such a distinguished oriented edge is called a DOE. We choose an arbitrary edge e of τ and give it an arbitrary orientation; e is a DOE of τ . The characteristic map is built by induction. We first identify DOEs e_0 and e with orientations. The construction of the map proceeds by identifying complementary ideal triangles of τ_* and τ according to their relative positions with respect to DOEs e_0 and e ; in fact, the identifications of the triangles uniquely determine an identification of the edges of τ_* and τ . The DOEs $e_0 \in \tau_*$ and $e \in \tau$ separate \mathbb{D} into left and right half-disks according to their orientations. We identify the immediate left triangle Δ_0 of τ_* with respect to e_0 to the immediate left triangle Δ of τ with respect to e . This forces the identification of boundary edges of Δ_0 and Δ such that the edges with endpoints at the initial points of DOEs get identified and the edges at terminal points of DOEs get identified. To proceed with the construction of the map, we give orientations to both edges of both triangles Δ_0 and Δ such that the triangles are on the right of the edges. Then we continue the identifications of the triangles on the immediate left of the two edges in τ_* with the triangles on the immediate left of the two edges in τ as above. This process continues

indefinitely on the left side of DOEs and we do similar identifications on the right side of DOEs. It is not hard to see that the characteristic map between the edges of τ_* and τ extends to an order-preserving map h from the set of ideal boundary points $\mathbb{Q} \subset S^1$ onto itself. Then the characteristic map h extends to a homeomorphism of S^1 because the ideal points of the tessellations are dense in S^1 and the map is order-preserving (see [37]).

We show that the characteristic map $h: S^1 \rightarrow S^1$ conjugates a finite index subgroup H of G onto K (see [37]). Let ω be an ideal fundamental polygon for K whose boundary edges are in τ . Then $h^{-1}(\omega)$ is an ideal polygon with boundary edges in τ_* . The boundary sides of ω are identified in pairs by elements of K and we consider the corresponding boundary sides pairs in $h^{-1}(\omega)$. Since $\mathrm{PSL}_2(\mathbb{Z})$ acts freely and transitively on the oriented edges of τ_* , there exist unique maps in $\mathrm{PSL}_2(\mathbb{Z})$ which identify corresponding boundary side pairs of $h^{-1}(\omega)$ with the correct orientation such that the quotient is homeomorphic to \mathbb{D}/K . Let H be the subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ generated by these elements. Then h conjugates H onto K (see [37]). \square

We consider the isotropy group in $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ of a single TLC tessellation of \mathbb{D} . A basic result states that any orientation preserving homeomorphism of S^1 which setwise fixes the Farey tessellation is necessarily an element of $\mathrm{PSL}_2(\mathbb{Z})$ (see [37, Lemma 7.3]). Then

Theorem 11.4. *The isotropy subgroup in $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ of τ , for τ a TLC tessellation, is quasiconformally conjugate to $\mathrm{PSL}_2(\mathbb{Z})$. The isotropy subgroup of τ_* is $\mathrm{PSL}_2(\mathbb{Z})$.*

Let τ be a TLC tessellation of \mathbb{D} which is invariant under $K < G$. Fix an edge e of τ . Then e is on the boundary of exactly two complementary ideal triangles of τ . The union of the two triangles is an ideal quadrilateral $P \subset \mathbb{D}$ one of whose diagonals is e . If no two edges in the set $K\{e\}$ are immediate neighbors, then the operation of changing diagonals $K\{e\}$ along the orbit $K\{P\}$ of quadrilaterals is well defined and produces a new TLC tessellation which is also invariant under K . Such an operation is called a *K-equivariant Whitehead move* (see [37]). This is a lift to the unit disk \mathbb{D} of a classical Whitehead move on surface \mathbb{D}/K considered by Hatcher and Thurston [19], Harer [20] and Penner [35]. Penner [34] also considered Whitehead moves on \mathbb{D} without the equivariance property.

The above transitivity result implies that any TLC tessellation of \mathbb{D} can be mapped by an element of $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ to its image under an equivariant Whitehead move. An element of $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ which achieves this is not unique; the ambiguity is up to precomposition by an element of $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ which fixes the initial tessellation. If we are given a DOE e_1 on the initial TLC tessellation τ , then a DOE e_2 on the image tessellation τ_1 under a K -equivariant move on $K\{e\}$ is determined by $e_2 := e_1$ if $e_1 \notin K\{e\}$, or otherwise $e_2 := f_1$, where f_1 is the other diagonal in the quadrilateral containing e_1 oriented such that (e_1, f_1) is a positive basis at their intersection point. In this case the element of $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$, called the *Whitehead homeomorphism*, is uniquely determined by mapping DOE onto DOE. Let h_τ and h_{τ_1} be the characteristic maps

for τ and τ_1 , namely $h_\tau(\tau_*) = \tau$, $h_\tau(e_0) = e_1$, $h_{\tau_1}(\tau_*) = \tau_1$ and $h_{\tau_1}(e_0) = e_2$, where $e_0 = (-1, 1)$ is DOE of τ_* . Then the above Whitehead homeomorphism is given by $h_{\tau_1} \circ h_\tau^{-1}$ (see [37]).

A basic fact due to Thurston, Hatcher, Harer and Penner is that any two ideal triangulations of a punctured surface are connected by a sequence of Whitehead moves. Therefore, if two TLC tessellations are invariant under K then they can be connected by a sequence of K -equivariant moves. If one TLC tessellation is invariant under K_1 and the other is invariant under K_2 then they can be connected by $(K_1 \cap K_2)$ -equivariant Whitehead moves (because they are both invariant under $K_1 \cap K_2$). This transitivity of all equivariant Whitehead moves on the set of TLC tessellations allows us to give generators of $\text{Mod}(\mathcal{S}_{\text{nc}})$ (see [37]). A composition of Whitehead homeomorphisms is called *geometric* if they are all K -equivariant, for a fixed subgroup K , and they are formed from a geometric sequence of Whitehead moves. We obtained

Theorem 11.5 ([37]). *Any element of the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ can be written as a composition $w \circ \gamma$, where $\gamma \in \text{PSL}_2(\mathbb{Z})$ and w is a geometric composition of K -equivariant Whitehead homeomorphisms for some fixed K .*

In joint work with S. Bonnot and R. Penner we give a presentation of the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ in [6]. We first define a 2-complex called the *triangulation complex* \mathcal{X} (see [6]). The vertices are all TLC tessellations of the unit disk \mathbb{D} . We already showed that $\text{Mod}(\mathcal{S}_{\text{nc}})$ preserves the set of vertices \mathcal{X}_0 and it acts transitively on them. The Farey tessellation τ_* is the basepoint of \mathcal{X} .

The set of edges \mathcal{X}_1 is first defined at the base point τ_* . A vertex $\tau \in \mathcal{X}_0$ is connected to the basepoint τ_* by an edge if τ is obtained from τ_* by a single K -equivariant Whitehead move, for some finite index subgroup K of G . An edge at an arbitrary $\tau \in \mathcal{X}_0$ is the image under $h_\tau \in \text{Mod}(\mathcal{S}_{\text{nc}})$ of an edge at the basepoint. Therefore, an edge connecting arbitrary $\tau, \tau_1 \in \mathcal{X}_0$ is obtained by a single “generalized” Whitehead move, namely the move is equivariant under a conjugate of K , $[G : K] < \infty$, by $h_\tau : S^1 \rightarrow S^1$ which induces an element of $\text{Comm}(G)$. The difference from a (regular) Whitehead move is that h_τ conjugates a proper subgroup K_1 of K onto another subgroup H_1 of G and the move is along the orbit of an edge for $h_\tau K h_\tau^{-1}$ which is not a subgroup of G . However, the generalized $h_\tau K h_\tau^{-1}$ -equivariant Whitehead move can be decomposed into finitely many H_1 -equivariant Whitehead moves. On the other hand, the image at the basepoint of an edge at an arbitrary point is necessarily obtained by a (regular) Whitehead move [6]. Thus we do not introduce new edges at the basepoint τ_* . The set of edges \mathcal{X}_1 is invariant under $\text{Mod}(\mathcal{S}_{\text{nc}})$ by definition.

The two cells \mathcal{X}_2 are introduced first at the basepoint τ_* . There are three kinds of two cells.

The *square two cells* are defined by adding a two cell to each cycle of four edges which are based at τ_* and have the following properties. The four edges are given by Whitehead moves equivariant with respect to the same finite index subgroup K of G . We assume that $e_1, e_2 \in \tau_*$ are two edges such that their corresponding orbits $K\{e_1\}$ and $K\{e_2\}$ have no pairs (whose one element is from $K\{e_1\}$ and the other

is from $K\{e_2\}$ of adjacent edges in τ_* . (Since we take a K -equivariant Whitehead move for e_i , we implicitly assume that $K\{e_i\} \subset \tau_*$ does not have adjacent edges in τ_* , for $i = 1, 2$.) Let f_1, f_2 be the other diagonals in the two quadrilaterals in $(\mathbb{D} - \tau_*) \cup \{e_1\}, (\mathbb{D} - \tau_*) \cup \{e_2\}$ containing e_1, e_2 . Then we form a TLC tessellation τ_1 by performing a K -equivariant Whitehead move on τ_* along e_1 ; we form a TLC tessellation τ_2 by performing a K -equivariant Whitehead move on τ_1 along e_2 ; we form a TLC tessellation τ_3 by performing a K -equivariant Whitehead move on τ_2 along f_1 ; and we return to τ_* by performing a K -equivariant Whitehead move on τ_3 along f_2 . The corresponding edges $E_1 = (\tau_*, \tau_1), E_2 = (\tau_1, \tau_2), E_3 = (\tau_2, \tau_3)$ and $E_4 = (\tau_3, \tau_*)$ make a closed path. We add a square two cell to \mathcal{X} whose boundary is the above closed edge path.

The *pentagon two cells* are defined by adding a two cell whose boundary is a closed edge path of length five as follows. Let K be a finite index subgroup of G and let e_1, e_2 be two adjacent edges in τ_* . Assume that $e_1 \notin K\{e_2\}$. Let P be the pentagon in $(\mathbb{D} - \tau_*) \cup \{e_1, e_2\}$; the orbit of pentagons $K\{P\}$ has pairwise mutually disjoint interiors with possible identifications of their boundaries. We define a closed edge path of length five by Whitehead moves: E_1 is given by a Whitehead move along $K\{e_1\}$ where $K\{e_1\} \mapsto K\{f_1\}$; E_2 is given by a Whitehead move along $K\{e_2\}$ where $K\{e_2\} \mapsto K\{f_2\}$; E_3 is given by a Whitehead move along $K\{f_1\}$ where $K\{f_1\} \mapsto K\{f_3\}$; E_4 is given by a Whitehead move along $K\{f_2\}$ where $K\{f_2\} \mapsto K\{e_1\}$; and E_5 is given by a Whitehead move along $K\{f_3\}$ where $K\{f_3\} \mapsto K\{e_2\}$ (this is the classical pentagon relation on a surface lifted to \mathbb{D} ; see, for example, [35], [34], [37]). We add a pentagon two cell whose boundary is such an edge path.

The *coset two cells* are defined by subdividing a single equivariant Whitehead move into several equivariant Whitehead moves as follows. Let K be a finite index subgroup of G and let K_1 be a finite index subgroup of K . Let $e \in \tau_*$ be such that no two edges in the orbit $K\{e\}$ are adjacent in τ_* . The *long edge* E is given by K -equivariant Whitehead move along $K\{e\}$. The *short edges* are given by K_1 -equivariant Whitehead moves as follows. Since $k := [K : K_1] < \infty$, there exists finitely many $e_1, e_2, \dots, e_k \in K\{e\}$ such that $e_i \notin K_1\{e_j\}$, for $i \neq j$, and $\bigcup_{i=1}^k K_1\{e_i\} = K\{e\}$. We define a sequence of short edges E_1, \dots, E_k by $E_i = (\tau_{i-1}, \tau_i)$, where τ_i , for $i = 1, 2, \dots, k$, is obtained from τ_{i-1} by performing a K_1 -equivariant Whitehead move on τ_{i-1} along $K_1\{e_i\}$ and $\tau_0 = \tau_*$. The edge path E_1, E_2, \dots, E_k starts at τ_* and ends at the endpoint of E . Thus E_1, \dots, E_k, E is a closed edge path based at τ_* and we add a coset two cell whose boundary is the given path (see [6]).

A general two cell in \mathcal{X} is the image under $\text{Mod}(\mathcal{G}_{\text{nc}})$ of a two cell based at τ_* . It turns out that an image of a square or a pentagon two cell based at τ_* under $\text{Mod}(\mathcal{G}_{\text{nc}})$ whose one vertex on its boundary is τ_* is of the same form as above. Namely, all the edges are Whitehead moves equivariant under a subgroup of $\text{PSL}_2(\mathbb{Z})$, while an image of a coset two cell under $\text{Mod}(\mathcal{G}_{\text{nc}})$ has generalized Whitehead moves as edges whenever the long edge does not limit at τ_* . It may happen that two short edges limit at τ_* . The modular group $\text{Mod}(\mathcal{G}_{\text{nc}})$ preserves the set of two cells \mathcal{X}_2 by its definition (see [6]).

We showed that the triangulation complex is simply connected [6].

Theorem 11.6. *The modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ acts cellularly on the triangulation complex \mathcal{X} . The triangulation complex \mathcal{X} is connected and simply connected.*

In [6] we give a presentation of the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ using its action on \mathcal{X} . We already showed that $\text{Mod}(\mathcal{S}_{\text{nc}})$ acts transitively on the vertices of \mathcal{X} and that the isotropy group of τ_* is $\text{PSL}_2(\mathbb{Z})$. Therefore, each orbit of an edge contains an edge with one endpoint at τ_* . To give a presentation, it is necessary to find the isotropy groups of edges. There are two types of edges with one endpoint in τ_* , the set \mathcal{E}^+ of edges which are not inverted by the action of $\text{Mod}(\mathcal{S}_{\text{nc}})$ and the set \mathcal{E}^- of edges which are inverted by the action of $\text{Mod}(\mathcal{S}_{\text{nc}})$. If $E \in \mathcal{E}^+$ is obtained by a K -equivariant Whitehead move then the isotropy group of E has to be contained in $\text{PSL}_2(\mathbb{Z})$ and it contains K . In fact, the isotropy group of E is a finite extension K' of K by elliptic elements in $\text{PSL}_2(\mathbb{Z})$ which preserve the other vertex τ of the edge $E = (\tau_*, \tau)$. Let $E_1 = (\tau_*, \tau_1) \in \mathcal{E}^-$ be an edge reversed by $\text{Mod}(\mathcal{S}_{\text{nc}})$, where τ_1 is obtained by a K_1 -equivariant Whitehead move and fixed by $K'_1 > K$. Then the isotropy group of E_1 is generated by $K'_1 < \text{PSL}_2(\mathbb{Z})$ which does not reverse the orientation of E_1 and by $k \in \text{Mod}(\mathcal{S}_{\text{nc}})$ which reverses the orientation, where k is mapping class like (i.e., k conjugates a finite index subgroup of G onto itself) and $k^2 \in K'_1 - K_1$ is elliptic fixing an edge in τ_* which implies $k^4 = \text{id}$ (see [6]).

We choose a single Whitehead move for each edge $E = (\tau_*, \tau)$ starting at τ_* by taking $e_0 = (-1, 1)$ to be a DOE of τ_* and defining a DOE of the other vertex τ as above. The set of these Whitehead moves together with $\text{PSL}_2(\mathbb{Z})$ generate $\text{Mod}(\mathcal{S}_{\text{nc}})$. We describe relations coming from two cells in \mathcal{X} for the chosen generating set.

Let Q be a square two cell based at τ_* with edges $E_i = (\tau_{i-1}, \tau_i)$, for $i = 1, \dots, 4$, with $\tau_0 = \tau_4 = \tau_*$ such that τ_i are K -equivariant. Let $h_1 \in \text{Mod}(\mathcal{S}_{\text{nc}})$ be the Whitehead homeomorphism corresponding to E_1 , h_2 the Whitehead homeomorphism corresponding to the edge $E'_2 = (\tau_*, h_1^{-1}(\tau_*))$, h_3 the Whitehead homeomorphism corresponding to the edge $E'_3 = (\tau_*, (h_1 \circ h_2)^{-1}(\tau_*))$, and h_4 the Whitehead homeomorphism corresponding to the edge $E'_4 = (\tau_*, (h_1 \circ h_2 \circ h_3)^{-1}(\tau_*))$. If $e_0 \notin K\{e_1, e_2\}$, where $K\{e_1\}$ and $K\{e_2\}$ are orbits which get changed in the definition of Q , then $h_1 \circ \dots \circ h_4(e_0) = e_0$ which implies that

$$h_1 \circ \dots \circ h_4 = \text{id}. \quad (11.1)$$

If $e_0 \in K\{e_1, e_2\}$ then $h_1 \circ \dots \circ h_4(e_0) = \bar{e}_0$, where \bar{e}_0 is the opposite orientation of e_0 , which implies that

$$h_1 \circ \dots \circ h_4 = s_{e_0}, \quad (11.2)$$

where $s_{e_0} \in \text{PSL}_2(\mathbb{Z})$ is an elliptic element reversing the orientation of e_0 .

Let P be a pentagon two cell based at τ_* with boundary edge path $E_i = (\tau_{i-1}, \tau_i)$, for $i = 1, \dots, 5$, where $\tau_0 = \tau_5 = \tau_*$ and τ_i are K -equivariant. Let $K\{e_1, e_2\}$ be the orbits which get changed to obtain P . Let h_i be the Whitehead move corresponding

to $(h_1 \circ \cdots \circ h_{i-1})^{-1}(E_i)$ as defined above. Then we have a pentagon relation

$$h_1 \circ \cdots \circ h_5 = \text{id}, \quad (11.3)$$

whenever $e_0 \notin K\{e_1, e_2\}$. We get

$$h_1 \circ \cdots \circ h_5 = \gamma_{e_0, \bar{e}_2}, \quad (11.4)$$

when $e_0 \in K\{e_1\}$, where $\gamma_{e_0, \bar{e}_2} \in \text{PSL}(\mathbb{Z})$ maps e_0 onto \bar{e}_2 and \bar{e}_2 has orientation opposite to e_2 . Finally, we get

$$h_1 \circ \cdots \circ h_5 = \gamma_{e_0, \bar{e}_1}, \quad (11.5)$$

when $e_0 \in K\{e_2\}$, where $\gamma_{e_0, \bar{e}_1} \in \text{PSL}(\mathbb{Z})$ maps e_0 onto \bar{e}_1 and \bar{e}_1 has orientation opposite to e_1 .

Let C be a coset two cell given by a long edge determined by a Whitehead move along $K\{e\}$ and by short edges with respect to $K_1 < K$, where $n = [K : K_1] < \infty$. We note that given K_1 , there are $n!$ paths of short edges connecting the two endpoints of the long edge. If $e_0 \notin K\{e\}$ then we obtain the coset relation

$$h \circ h_1 \circ \cdots \circ h_n = \text{id}, \quad (11.6)$$

where h is the Whitehead homeomorphism corresponding to the long edge and h_i is the Whitehead homeomorphism corresponding to the image $(h \circ h_1 \circ \cdots \circ h_{i-1})^{-1}(E_i)$ of the i -th short edge E_i . If $e_0 \in K\{e\}$ then we obtain coset relation

$$h \circ h_1 \circ \cdots \circ h_n = s_{e_0}, \quad (11.7)$$

where $s_{e_0} \in \text{PSL}_2(\mathbb{Z})$ reverses the orientation of e_0 .

We obtained [6] a presentation for $\text{Mod}(\mathcal{S}_{\text{nc}})$ as follows.

Theorem 11.7. *The modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ is generated by the isotropy subgroup $\text{PSL}_2(\mathbb{Z})$ of the basepoint $\tau_* \in \mathcal{X}$, the isotropy subgroups $\Gamma(E)$ for $E \in \mathcal{E}^\pm$, and by the Whitehead homeomorphism g_E for $E \in \mathcal{E}^+$ chosen as above. The following relations on these generators give a complete presentation of $\text{Mod}(\mathcal{S}_{\text{nc}})$:*

- a) *The inclusions of $\Gamma(E)$ into $\text{PSL}_2(\mathbb{Z})$, for $E \in \mathcal{E}^+$, given by $\Gamma(E) = K'$, where the terminal endpoint of E is invariant under the finite-index subgroup $K' < \text{PSL}_2(\mathbb{Z})$.*
- b) *The inclusions of $\Gamma^+(E)$ into $\text{PSL}_2(\mathbb{Z})$, for $E \in \mathcal{E}^-$, given by $\Gamma(E) = K'$, where the terminal endpoint of E is invariant under the finite-index subgroup $K' < \text{PSL}_2(\mathbb{Z})$.*
- c) *The relations introduced by the boundary edge-paths of two-cells in \mathcal{F} given by the equations (11.3), (11.4), (11.5), (11.1), (11.2), (11.6) and (11.7).*
- d) *The redundancy relations: for any two edges E and E' in \mathcal{E}^\pm and for any $\gamma \in \text{PSL}_2(\mathbb{Z})$ such that $\gamma(E) = E'$, we have the relation*

$$g_{E'} \circ \gamma' = \gamma \circ g_E,$$

where γ' is the unique element of $\mathrm{PSL}_2(\mathbb{Z})$ that satisfies $\gamma'(e_0) = e'_1$ with $e'_1 = g_{E'}^{-1}(\gamma(e_0))$.

Remark 11.8. The redundancy relations d) in the above theorem are introduced because we used more generators than necessary. We could have used Whitehead homeomorphisms of representatives of orbits of edges based at τ_* instead. Then we would not have to add relations d). However, it is not easy to give a proper enumeration of such orbits which would necessarily complicate the relations in c). Thus, for the sake of simplicity of relations, we used a larger set of generators in the above theorem.

We also showed [6] that $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ has no center.

Theorem 11.9. *The modular group $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ of the non-compact solenoid $\mathcal{S}_{\mathrm{nc}}$ has trivial center.*

12 Elements of $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ with small non-zero dilatations

In a recent joint work with V. Markovic [29], we showed the following

Theorem 12.1. *For every $\varepsilon > 0$ there exist two finite index subgroups of $\mathrm{PSL}_2(\mathbb{Z})$ which are conjugated by a $(1 + \varepsilon)$ -quasisymmetric homeomorphism of the unit circle and this conjugation homeomorphism is not conformal.*

To construct the above groups and the quasisymmetric map, we use the generators of the modular group $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ introduced in [37] (see also the proof of Theorem 11.3).

We obtained [29] the following corollary to the above theorem.

Corollary 12.2. *Let \mathbb{T}_0 denote the modular torus. Then for every $\varepsilon > 0$ there are finite-degree, regular coverings $\pi_1: M_1 \rightarrow \mathbb{T}_0$ and $\pi_2: M_2 \rightarrow \mathbb{T}_0$, and a $(1 + \varepsilon)$ -quasiconformal homeomorphism $F: M_1 \rightarrow M_2$ that is not homotopic to a conformal map.*

The following corollary is an interpretation of Theorem 12.1 in terms of the Teichmüller space $\mathcal{T}(\mathcal{S}_{\mathrm{nc}})$ of the non-compact solenoid $\mathcal{S}_{\mathrm{nc}}$. This is a significant progress in understanding the quotient $\mathcal{T}(\mathcal{S}_{\mathrm{nc}})/\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$ for the Teichmüller metric.

Corollary 12.3. *The closure in the Teichmüller metric of the orbit (under the modular group $\mathrm{Mod}(\mathcal{S}_{\mathrm{nc}})$) of the basepoint in $\mathcal{T}(\mathcal{S}_{\mathrm{nc}})$ is strictly larger than the orbit. Moreover, the closure of this orbit is a perfect set and thus uncountable.*

13 Some open problems

We discuss some open question concerning the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ and the modular group $\text{Mod}(\mathcal{S}_G)$ of the solenoid.

As we already mentioned, a conjecture by L. Ehrenpreis states that given any two closed Riemann surfaces of genus at least two and given any $\varepsilon > 0$ there exist unbranched, finite-sheeted, holomorphic covers of these surfaces that are $(1 + \varepsilon)$ -quasiconformal. D. Sullivan gave the following equivalent formulation in terms of the compact solenoid:

1. Is it true that the modular group $\text{Mod}(\mathcal{S}_G)$ has dense orbits (for the Teichmüller metric) in the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ of the compact solenoid \mathcal{S}_G ?

We also considered the Teichmüller space $\mathcal{T}(\mathcal{S}_{\text{nc}})$ and the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ of the noncompact solenoid \mathcal{S}_{nc} . Therefore we can ask the analogous question in this setup:

2. Is it true that the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ has dense orbits (for the Teichmüller metric) in the Teichmüller space $\mathcal{T}(\mathcal{S}_{\text{nc}})$ of the noncompact solenoid \mathcal{S}_{nc} ?

It is interesting to note that a positive answer to question 1 does not immediately give a positive answer to question 2. This is easiest to understand in terms of the original formulation by Ehrenpreis. To see this, assume for the moment that any two closed surfaces have unbranched, finite-sheeted, holomorphic covers which are quasiconformal with arbitrary small dilatation. When considering two punctured surfaces, one is tempted to fill in the punctures and find unbranched holomorphic covers of the compactified surfaces which are quasiconformal with small dilatation. However, the problem is that the quasiconformal map does not necessarily send the lifts of the punctures on one surface to the lifts of the punctures on the other surface.

We considered the Teichmüller metric on $\mathcal{T}(\mathcal{S}_G)$ and the existence of geodesics between points. If a map is of Teichmüller-type then we showed that it is extremal and that there is a unique geodesic between the point determined by the Teichmüller-type map and the base point of $\mathcal{T}(\mathcal{S}_G)$. Moreover, we showed that only a small subset of $\mathcal{T}(\mathcal{S}_G)$ has Teichmüller-type representatives. We ask

3. Is it true that any point in the Teichmüller space $\mathcal{T}(\mathcal{S}_G)$ has an (unique) extremal representative?

If the answer is positive, then any two points are connected by a (unique) geodesic. Even if the answer is negative, it is still possible to have geodesics connecting a point in $\mathcal{T}(\mathcal{S}_G)$ without an extremal representative to the base point.

4. Is it possible to connect any two points in $\mathcal{T}(\mathcal{S}_G)$ by a (unique) geodesic?

We also established a sufficient condition for a point in $\mathcal{T}(\mathcal{S}_G)$ to have a Teichmüller-type representative. The condition is given in terms of the approximating sequence of TLC structures. We ask for additional sufficient conditions.

5. Is there a sufficient condition for a point $[f] \in \mathcal{T}(\mathcal{S}_G)$ expressed only in terms of the geometry of the point $[f]$ to have a Teichmüller-type representative?

A classical statement about duality of the cotangent and tangent space for Teichmüller spaces of Riemann surfaces is false for $\mathcal{T}(\mathcal{S}_G)$. We therefore ask

6. Does the tangent space $L_s^\infty(\mathcal{S}_G)/N(\mathcal{S}_G)$ at the basepoint of $\mathcal{T}(\mathcal{S}_G)$ have a pre-dual?

It is a classical fact that any biholomorphic map of the Teichmüller space of a finite Riemann surface is given by the geometric action of an element of the extended mapping class group. This is recently proved for all infinite Riemann surfaces as well [27] (see also [41], [11], [10], [23] and Chapter 2 of this volume [16]). We ask analogous question of $\mathcal{T}(\mathcal{S}_G)$.

7. Does every biholomorphism (isometry) of $\mathcal{T}(\mathcal{S}_G)$ arise by the geometric action of the full mapping class group $\text{Mod}_{\text{full}}(\mathcal{S}_G)$ (i.e., homotopy classes of self maps of \mathcal{S}_G not necessarily fixing the baseleaf and allowing orientation reversing elements)?

We considered the modular group $\text{Mod}(\mathcal{S}_{\text{nc}})$ of the noncompact solenoid \mathcal{S}_{nc} and found an explicit set of generators and a presentation. We ask the analogous question for the compact solenoid.

8. Find an explicit set of generators of $\text{Mod}(\mathcal{S}_G)$ of the compact solenoid \mathcal{S}_G .

9. Find a presentation of $\text{Mod}(\mathcal{S}_G)$.

We expect that these modular groups are infinitely generated.

10. Show that $\text{Mod}(\mathcal{S}_G)$ and $\text{Mod}(\mathcal{S}_{\text{nc}})$ are infinitely generated.

Recall that an element of $\text{Mod}(\mathcal{S}_G)$ is called mapping class like if it is a lift of a self map of a closed surface. C. Odden [38] asked the following question:

11. Is it true that $\text{Mod}(\mathcal{S}_G)$ and $\text{Mod}(\mathcal{S}_{\text{nc}})$ are generated by mapping class like elements?

Study properties of $\text{Mod}(\mathcal{S}_G)$. In particular,

12. Is there a classification of the elements of $\text{Mod}(\mathcal{S}_G)$ according to their actions on \mathcal{S}_G similar to the Thurston's classification of the mapping class group elements of a closed surface?

References

- [1] L. V. Ahlfors and L. Bers, Riemann's mapping theorem for variable metrics. *Ann. of Math.* 72 (2) (1960), 385–404. [821](#), [828](#)
- [2] R. Baer, Isotopien auf Kurven von orientierbaren, geschlossenen Flächen. *J. Reine Angew. Math.* 159 (1928), 101–116. [813](#)

- [3] I. Biswas and S. Nag, Weil-Petersson geometry and determinant bundles on inductive limits of moduli spaces. In *Lipa's legacy* (New York, 1995), Contemp. Math. 211, Amer. Math. Soc., Providence, RI, 1997, 51–80.
- [4] I. Biswas and S. Nag, Limit constructions over Riemann surfaces and their parameter spaces, and the commensurability group actions. *Selecta Math.* (N.S.) 6 (2000), 185–224. [839](#)
- [5] I. Biswas, S. Nag and D. Sullivan, Determinant bundles, Quillen metrics and Mumford isomorphisms over the universal commensurability Teichmüller space. *Acta Math.* 176 (2) (1996), 145–169. [813](#), [838](#)
- [6] S. Bonnot, R. C. Penner and D. Šarić, A presentation for the baseleaf preserving mapping class group of the punctured solenoid. *Algebr. Geom. Topol.* 7 (2007), 1171–1199. [816](#), [849](#), [850](#), [851](#), [852](#), [853](#)
- [7] A. Candel, Uniformization of surface laminations. *Ann. Sci. École Norm. Sup.* (4) 26 (4) (1993), 489–516. [814](#), [820](#)
- [8] M. Dehn, *Papers on group theory and topology*. Springer-Verlag, New York 1987. [813](#)
- [9] A. Douady and C. J. Earle, Conformally natural extension of homeomorphisms of the circle. *Acta Math.* 157 (1–2) (1986), 23–48. [827](#)
- [10] C. Earle and F. Gardiner, Geometric isomorphisms between infinite-dimensional Teichmüller spaces. *Trans. Amer. Math. Soc.* 348 (3) (1996), 1163–1190. [855](#)
- [11] C. Earle and I. Kra, On holomorphic mappings between Teichmüller spaces. In *Contributions to analysis*, Academic Press, New York 1974, 107–124. [855](#)
- [12] C. Earle and C. McMullen, Quasiconformal isotopies. In *Holomorphic Functions and Moduli I* (Berkeley, CA, 1986), Math. Sci. Res. Inst. Publ. 10, Springer-Verlag, New York 1988, 143–154. [827](#)
- [13] L. Ehrenpreis, Cohomology with bounds. In *Symposia Mathematica IV* (INDAM, Rome, 1968/69), Academic Press, London 1970, 389–395. [812](#)
- [14] A. Epstein, V. Markovic and D. Šarić, Extremal maps for the universal hyperbolic solenoid. Preprint, arXiv:math.DS/0604411. [816](#), [835](#), [836](#), [837](#), [838](#)
- [15] D. B. A. Epstein and R. C. Penner, Euclidean decompositions of non-compact hyperbolic manifolds. *J. Differential Geom.* 27 (1988) 67–80. [816](#), [845](#)
- [16] A. Fletcher and V. Markovic, Infinite dimensional Teichmüller spaces. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2009, 65–91. [855](#)
- [17] F. Gardiner, *Teichmüller theory and quadratic differentials*. Pure and Appl. Math., John Wiley & Sons, Inc., New York 1987. [828](#), [829](#), [830](#)
- [18] F. Gardiner and N. Lakic, *Quasiconformal Teichmüller Theory*. Math. Surveys Monogr. 76, Amer. Math. Soc., Providence, RI, 2000. [839](#)
- [19] A. Hatcher and W. Thurston, A presentation for the mapping class group of a closed orientable surface. *Topology* 19 (1980), 221–237. [848](#)
- [20] J. Harer, The virtual cohomological dimension of the mapping class group of an orientable surface. *Invent. Math.* 84 (1986) 157–176. [816](#), [845](#), [848](#)
- [21] J. Kahn and V. Markovic, Random ideal triangulations and the Weil-Petersson distance between finite degree covers of punctured Riemann surfaces. Preprint 2008, arXiv:0806.2304v1. [814](#)

- [22] N. Lakic, Strebel points. *Contemp. Math.* 211 (1997), 417–431. [816](#), [835](#)
- [23] N. Lakic, An isometry theorem for quadratic differentials on Riemann surfaces of finite genus. *Trans. Amer. Math. Soc.* 349 (7) (1997), 2951–2967. [855](#)
- [24] M. Lyubich and Y. Minsky, Laminations in holomorphic dynamics. *J. Differential Geom.* 47 (1997), 17–94. [813](#), [814](#)
- [25] A. Marden and K. Strebel, The heights theorem for quadratic differentials on Riemann surfaces. *Acta Math.* 153 (3–4) (1984), 153–211.
- [26] D. Margalit and C. Leininger, Abstract commensurators of braid groups. *J. Algebra* 299 (2) (2006), 447–455. [813](#)
- [27] V. Markovic, Biholomorphic maps between Teichmüller spaces. *Duke Math. J.* 120 (2) (2003), 405–431. [855](#)
- [28] V. Markovic and D. Šarić, Teichmüller Mapping class group of the universal hyperbolic solenoid. *Trans. Amer. Math. Soc.* 358 (6) (2006), 2637–2650. [816](#), [823](#), [827](#), [839](#), [840](#)
- [29] V. Markovic and D. Šarić, The Teichmüller distance between finite index subgroups of $PSL_2(\mathbb{Z})$. *Geom. Dedicata* 136 (2008), 145–165. [853](#)
- [30] C. McMullen, Amenability, Poincaré series and quasiconformal maps. *Invent. Math.* 91 (1988) 259–272. [815](#), [839](#)
- [31] C. McMullen, Thermodynamics, dimension and the Weil-Petersson metric. *Invent. Math.* 173 (2) (2008), 365–425. [813](#)
- [32] S. Nag and D. Sullivan, Teichmüller theory and the universal period mappings via quantum calculus and the $H^{1/2}$ space of the circle. *Osaka J. Math.* 32 (1995), 1–34. [813](#), [815](#), [817](#), [821](#), [835](#)
- [33] J. Nielsen, Untersuchungen zur Theorie der geschlossenen zweiseitigen Flächen I. *Acta Math.* 50 (1927), 189–358. [813](#)
- [34] R. C. Penner, Universal constructions in Teichmüller theory. *Adv. Math.* 98 (1993), 143–215. [816](#), [844](#), [845](#), [847](#), [848](#), [850](#)
- [35] R. C. Penner, The decorated Teichmüller space of punctured surfaces. *Comm. Math. Phys.* 113 (1987), 299–339. [816](#), [843](#), [844](#), [845](#), [848](#), [850](#)
- [36] R. C. Penner, Surfaces, circles and solenoids. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 205–221. [844](#)
- [37] R. C. Penner and D. Šarić, Teichmüller theory of the punctured solenoid. *Geom. Dedicata* 132 (2008), 179–212. [813](#), [814](#), [816](#), [821](#), [841](#), [842](#), [843](#), [844](#), [845](#), [846](#), [847](#), [848](#), [849](#), [850](#), [853](#)
- [38] C. Odden, The baseleaf preserving mapping class group of the universal hyperbolic solenoid. *Trans. Amer. Math. Soc.* 357 (2005), 1829–1858. [812](#), [813](#), [814](#), [816](#), [817](#), [818](#), [819](#), [820](#), [838](#), [839](#), [840](#), [846](#), [855](#)
- [39] E. Reich and K. Strebel, Teichmüller mappings which keep the boundary point-wise fixed. In *1971 Advances in the theory of Riemann surfaces* (Stony Brook, N.Y., 1969), Ann. of Math. Stud. 66, Princeton University Press, Princeton, N.J., 365–367.
- [40] E. Reich and K. Strebel, Extremal quasiconformal mappings with given boundary values. *Contributions to analysis*, Academic Press, New York 1974, 375–391.

- [41] H. Royden, Automorphisms and isometries of Teichmüller space. In *1971 Advances in the theory of Riemann surfaces* (Stony Brook, N.Y., 1969), Ann. of Math. Stud. 66, Princeton University Press, Princeton, N.J., 369–383. [828](#), [855](#)
- [42] D. Šarić, On Quasiconformal deformations of the universal hyperbolic solenoid. *J. Anal. Math.* 105 (2008), 303–343. [815](#), [824](#), [825](#), [826](#), [828](#), [829](#), [831](#), [832](#), [833](#), [834](#), [835](#)
- [43] K. Strebel, *Quadratic differentials*. Ergeb. Math. Grenzgeb. (3) 5, Springer-Verlag, Berlin 1984. [829](#), [830](#)
- [44] D. Sullivan, Linking the universalities of Milnor-Thurston, Feigenbaum and Ahlfors-Bers. In *Topological methods in modern mathematics* (L. Goldberg and A. Phillips, eds.), Publish or Perish, Houston, TX, 1993, 543–563. [811](#), [813](#), [814](#), [815](#), [817](#), [821](#), [824](#), [826](#), [827](#), [828](#)
- [45] D. Sullivan, Bounds, quadratic differentials, and renormalization conjectures. In *American Mathematical Society centennial publications*, Vol. II, Amer. Math. Soc., Providence, RI, 1992. [813](#)

List of Contributors

Riccardo Benedetti, Dipartimento di Matematica, Largo B. Pontecorvo 5, 56127 Pisa, Italy; email: benedett@dm.unipi.it

Francesco Bonsante, Dipartimento di Matematica, Via Ferrata 1, 27100 Pavia, Italy; email: francesco.bonsante@unipv.it

David Dumas, Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7045, U.S.A.; email: ddumas@math.uic.edu

Alastair Fletcher, Department of Mathematics, University of Glasgow, G12 8QW, United Kingdom; email: a.fletcher@maths.gla.ac.uk

Koji Fujiwara, Graduate School of Information Science, Tohoku University, Aoba, Sendai 980-8579, Japan; email: fujiwara@math.is.tohoku.ac.jp

William M. Goldman, Department of Mathematics, University of Maryland, College Park, MD 20742-4015, U.S.A.; email: wmg@math.umd.edu

Frank Herrlich, Institut für Algebra und Geometrie, Universität Karlsruhe, 76128 Karlsruhe, Germany; email: herrlich@math.uni-karlsruhe.de

Yoichi Imayoshi, Department of Mathematics, Osaka City University, Osaka 558-8585, Japan; email: imayoshi@sci.osaka-cu.ac.jp

Nariya Kawazumi, Department of Mathematical Sciences, University of Tokyo, Tokyo 153-8914, Japan; email: kawazumi@ms.u-tokyo.ac.jp

Yoshikata Kida, Mathematical Institute, Tohoku University, Sendai 980-8758, Japan; email: kida@math.tohoku.ac.jp

Sadayoshi Kojima, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Ohokayama, Meguro, Tokyo 152-8552, Japan; email: sadayosi@is.titech.ac.jp

Mustafa Korkmaz, Department of Mathematics, Middle East Technical University, 06531 Ankara, Turkey; email: korkmaz@metu.edu.tr

Sean Lawton, Departamento de Matemática, Instituto Superior Técnico, Lisbon, Portugal, and Department of Mathematics, University of Maryland, College Park, MD 20742-4015, U.S.A.; email: slawton@math.ist.utl.pt

Feng Luo, Department of Mathematics, Rutgers University, Piscataway, NJ 08854, U.S.A.; email: fluo@math.rutgers.edu

Vladimir Markovic, Mathematics Department, Stony Brook University, Stony Brook, NY 11794-3651, U.S.A.; email: markovic@math.sunysb.edu

Martin Möller, Max-Planck-Institut für Mathematik, Vivatsgasse 7, 53111 Bonn, Germany; email: moeller@mpim-bonn.mpg.de

Gabriele Mondello, Imperial College of London, Department of Mathematics, Huxley Building, South Kensington Campus, London SW7 2AZ, United Kingdom; email: g.mondello@imperial.ac.uk

Athanase Papadopoulos, Institut de Recherche Mathématique Avancée, CNRS et Université de Strasbourg, 7 rue René Descartes, 67084 Strasbourg Cedex, France; email: papadopoulos@math.u-strasbg.fr

Luis Paris, Institut de Mathématiques de Bourgogne – UMR 5584 du CNRS, Université de Bourgogne, BP 47870, 21078 Dijon Cedex, France; email: lparis@u-bourgogne.fr

Elisha Peterson, Department of Mathematical Sciences, West Point, NY 10996, U.S.A.; email: elisha.peterson@usma.edu

Dragomir Šarić, Department of Mathematics, Queens College of the City University of New York, 65-30 Kissena Blvd., Flushing, NY 11367, U.S.A.; email: dragomir.saric@qc.cuny.edu

Gabriela Schmithüsen, Institut für Algebra und Geometrie, Universität Karlsruhe, 76128 Karlsruhe, Germany; email: schmithuesen@math.uni-karlsruhe.de

Robert Silhol, Département de Mathématique, Université Montpellier II, Place E. Bataillon, 34095 Montpellier Cedex 5, France; email: rs@math.univ-montp2.fr

András I. Stipsicz, Rényi Institute of Mathematics, Budapest, Hungary, and Department of Mathematics, Columbia University, New York, NY 10027, U.S.A.; email: stipsicz@math-inst.hu

Scott A. Wolpert, Department of Mathematics, University of Maryland, College Park, MD 20742-4015, U.S.A.; email: saw@math.umd.edu

Index

- abelian differential, 370
- absolute Galois group, 778
- absolutely continuous on lines, 66
- accepted state, 418
- accessory parameters, 136
- achiral Lefschetz fibration, 290
- action
 - affine, 332
 - amenable, 336
 - essentially free, 298, 313
 - non-singular, 301, 313
 - standard, 319
- admissible curve, 479
- admissible system of cut curves,
 - 102, 110
 - precise, 111
 - reduced, 102
- admissible triple, 701
- AdS bending cocycle, 591
- affine geometry
 - complex, 512
- affine group, 369, 791
- algebraic fundamental group, 786
- algorithm for $SL(2, \mathbb{C})$ -traces of
 - reduced words in \mathbb{F}_2 , 620
- allowable map, 117
- almost harmonic map, 496
- alphabet, 418
- amenable group, 332
- amenable groupoid, 335
- amphidrome, 111
- Anti de Sitter space, 542
- antipodal involution, 493
- approximately invariant mean, 339
- Arakelov–Green function, 233
- arc, 157
- arc complex, 157
- arc system
 - automorphism of, 160
 - spinal, 191
- arc/graph duality, 167, 170, 186
- Artin group, 403
 - irreducible, 404
- Artin group of spherical type, 404
- Artin monoid, 410
- Artin representation, 402
- Artin system, 403
- asymptotic region, 596
- atom, 413
- atomic monoid, 412
- augmented Teichmüller space, 50, 52
 - Fenchel–Nielsen coordinates, 50
- automatic group, 419
- automorphic form, 141
- automorphism group
 - very large, 137
- base locus, 273
- Beltrami coefficient
 - transversely locally constant, 826
- Beltrami differential, 67, 177
 - harmonic, 177
 - solenoid, 831
 - infinitesimally trivial, 831
- Beltrami differential equation, 67
- Belyi morphism, 771, 798
- Belyi pair, 772
- bending, 559
- bending cocycle, 482, 579
- bending deformation, 481
- bending line, 470, 472
- Bergman kernel, 84
- Bergman space, 73, 82
 - geometric isometry, 78
 - separating points, 89
- Bers embedding, 72, 489

- Bers fiber space, 465
- Bers norm, 71
- Bers region, 51, 53, 56, 57
- Bers space, 71
- Bers Teichmüller space, 99
- biautomatic group, 419
- binor identity, 696
- birdtracks, 687
- bitangentiable, 476, 501
- black hole, 598
- Bonahon–Thurston coordinates, 741
- bordification, 174
- Borel space, 308
 - standard, 298, 308
- boundary map, 173
- boundary point of a Teichmüller curve, 796
- boundary weight map, 161
- bounded cohomology, 261
 - of a group, 244
 - of a space, 263
- bounded cohomology class, 263
- bounded generation, 255
- braid, 392, 398
- braid diagram, 392
- braid group, 393, 399
 - surface, 399
- braid twist, 400
- branched projective structure, 479
- broken Lefschetz fibration, 292
- bubble identity, 703
- bumping, 487
- canonical reduction system (CRS), 346, 348
- canonical representation, 407
- canonical stratification, 470, 472, 499
- Cantor set, 812
- Carathéodory distance, 89, 100
- CAT(0) metric space, 52
- Cauchy development, 586
- Cauchy surface, 539
- causal structure, 539
- Cayley–Hamilton theorem, 619
- center of grafting, 489
- central function, 686, 708
 - degree, 719
 - diagrammatic, 709
 - grading, 723
 - product, 713, 714, 721, 726
 - rank one, 713
 - rank two, 715
 - recurrence, 723
 - symmetry, 718
 - table, 719, 728
- chamber
 - hyperplane arrangement, 425
- character ring, 612
- character variety, 477, 478, 686, 689
- character variety of \mathbb{F}_2 , geometric version, 630
- character variety of \mathbb{F}_3 is a double covering, 663
- character variety of \mathbb{F}_3 is a hypersurface, 663, 665
- characteristic equation, 697
- Chern form, 222
- chord diagram, 436
- ciliation, 693
- circle, 460, 511, 515
- circle configuration, 515
- circle packing, 516
- class
 - generalized Witten, 204
 - Miller, 178
 - Mumford–Morita, 179
 - Witten, 204
- class function, 243
- Clebsch–Gordan formula, 692, 701
- co-collapsing map, 472, 494, 496, 497
- cocycle
 - ME, 318
 - OE, 314
- coherent complement, 416
- collapsing map, 472, 494, 496, 497

- combinatorial
 - forgetful map, 210
 - gluing map, 209
 - subcomplex, 199
- combinatorial data, 274
- common orthogonal geodesic, 625
- commutator length, 243, 282
 - stable, 282
- commutator trace, 621
- compact solenoid, 811
- compactification
 - grafting, 494
 - of measured laminations, 491
 - of quadratic differentials, 492
 - of Teichmüller space, 491
 - Schwarzian, 497
- comparison map, 244, 263
- complement, 415
- completely reduced, 103
- completely reducible representation, 686
- complex affine geometry, 512
- complex conjugation, 511
- complex dilatation, 220
- complex earthquake, 481
- complex projective structure, 458
- complex projective geometry, 511
- complex projective line, 511
- complex projective structure, 566
- complex structure
 - transversely locally constant, 821
- configuration space, 395, 399
- conformal metric, 474
- conical singularity, 603
- conjugacy problem, 419
- conjugate, 314
- constant braid, 393
- constant sectional curvature, 539
- contact structure, 291
- continued fractional parameter, 521
- convex core, 488
- convex core map, 544
- convex hull, 469
- convex pleated plane, 469
- coordinate map, 500
- coordinate ring, 686, 689
 - decomposition, 707
- copy
 - of a path, 253
 - of a word, 248
- cosmological constant, 538
- cosmological time, 567
- counting function, 249
- Coxeter complex, 428
- Coxeter graph, 403
- Coxeter graph of spherical type, 404
- Coxeter graph without triangle, 430
- Coxeter group, 403
 - irreducible, 404
- Coxeter matrix, 403
- Coxeter system, 403
- Coxeter triangle group, 629
- cross ratio, 524
- cross ratio parameter, 525
- cross ratio parameter space, 526
- cross-surface, 617
- crown, 545
- curvature concentration, 519
- curve
 - hyperelliptic, 133
 - modular, 144
 - nodal, 173
 - stable, 173
- curve complex, 49, 53, 251, 304
- curve defined over a number field, 771
- curve system, 744
- cusp of an origami curve, 796
- cyclic group, 618
- cyclic ordering, 160
- cyclically reduced word, 245
- cyclotomic character, 784
- de Sitter space, 470, 541, 650
- decorated hyperbolic structure
 - solenoid, 843
- decorated Teichmüller space
 - solenoid, 844

- defect, 242
- deformation, 117
 - isomorphism, 117
- deformation space, 117
- degenerate standard spacetime, 587
- Dehn twist, 175, 176, 276, 434
- Deligne–Mumford
 - augmentation, 175
 - compactification, 173, 796
- density function, 474
- dessin d’enfants, 772, 773, 793, 796
 - genus of, 777
- developing map, 459, 460, 462–464, 466, 467, 469, 470, 474, 485–487, 513, 540
- development-holonomy pair, 459, 463, 464
- diffeomorphism
 - elliptic, 371
 - parabolic, 371
 - pseudo-Anosov, 371
 - reducible, 371
- dihedral representation, 636
- dilatation coefficient, 371
- direction, 371
- discrete holonomy, 487, 490
- discrete valuation ring, 770
- discriminant, 397, 412
- Donaldson theorem (Lefschetz pencil), 278
- dual tree, 473
- dualizing sheaf, 178
- Earle class, 229
- earthquake, 481, 559
- earthquake flow, 562
- Ehrenpreis conjecture, 812
- Ehresmann connection, 221
- Einstein equation, 538
- Einstein metric, 539
- element
 - pseudo-Anosov, 306
 - pure, 306
 - reducible, 306
- elementary
 - subgroup, 248
 - word-hyperbolic group, 247
- elementary contraction, 168, 169
- elementary representation, 478
- elliptic curve
 - Belyi map for, 779
- elliptic modular transformation, 102
- embedded projective structure, 469
- energy of a map, 494
- enhanced bundle, 554
- enhanced length spectrum, 550
- enhanced measure spectrum, 556
- enhanced Teichmüller space, 549
- equimeasurable, 79
- equivalence relation
 - discrete measured, 312
 - quotient, 313
 - recurrent, 315
- equivalent in $N(G)$, 100
- equivalent in $Q_{\text{norm}}(G)$, 99
- equivalent marked surfaces, 99
- equivariant harmonic map, 495, 496
- essential arc, 400
- essential arrangement, 425
- essential circle, 434
- euclidean geometry, 512
- event, 538
- exact sequence of algebraic fundamental groups, 786
- exotic quasi-Fuchsian projective structure, 486
- exotic Fuchsian projective structure, 483
- extended mapping class group, 252
- extension to Coxeter group,
 - geometric proof, 638
- Faber’s formula, 180
- facet
 - hyperplane arrangement, 424
 - pleated plane, 470, 482

- family
 - flat, 172
 - smooth, 171
 - stable, 175
 - universal, 178
- fat graph, 774
- fellow traveler property, 419
- Fenchel–Nielsen coordinates, 49, 176, 738
 - one-holed torus, 659
 - Weil–Petersson metric expansion, 57
- fiber, 273
 - irreducible, 274
 - reducible, 274
 - regular, 274
 - singular, 274
 - type of a reducible singular, 274
- fiber sum, 274
- field of definition, 780
- field of meromorphic functions, 769
- finite state automaton, 418
- first fundamental form, 540
- flat conformal structure, 472
- flat structure, 791
- flat surface, 369, 371
- foliation
 - horizontal, 492
 - measured, 305
 - minimal measured, 305
 - projective measured, 305
 - pseudo-Anosov, 306
 - vertical, 492
- foliation map, 493, 498
- forgetful map, 173, 461, 527
- forgetting projection, 550
- four-holed sphere
 - trace coordinates, 668
- free group on two generators, 772, 789
- free monoid, 407
- Fricke space, 647
- Fricke space of four-holed sphere, 670
- Fricke space of one-holed torus, 654
- Fricke space of three-holed sphere, 651
- Fricke–Klein–Vogt Theorem, 717, 726
- Fuchsian center, 489, 490
- Fuchsian projective structure, 483
- Fuchsian space, 478
- fully biautomatic group, 419
- function
 - genus, 164, 174
 - marking, 174
- function field, 769
- fundamental binor identity, 696
- fundamental domain, 298, 315, 338
- fundamental group of $\hat{\mathbb{P}}$, 775
- fusion identity, 704
- future, 538
- future boundary, 586
- (G, X) -structure, 460
- Galois action
 - faithfulness of, 781, 793
 - on origamis, 792
 - on trees, 782
 - on \hat{F}_2 , 783
 - on Belyi pairs, 778
- Galois group
 - absolute, 778
- Galois invariants, 782
- Garside element, 410, 413
- Garside group, 413
- Garside monoid, 413
- Gauss map, 576
- Gauss–Codazzi equation, 539
- generating set, 816
- generic matrices, 689
- genus of a dessin, 777
- genus of a Lefschetz fibration, 274
- geodesic lamination, 550
- geodesic language, 418
- geodesic-length function, 56
 - convexity, 54
 - Fenchel–Nielsen coordinates, 49
 - Weil–Petersson gradient pairings, 54
- geodesically automatic group, 419
- geometric braid, 392

- geometric intersection, 161
- geometric representation, 434
- globally hyperbolic spacetime, 539
- globally trivial, 97
- Gompf theorem (symplectic 4-manifold), 278
- grafting, 195, 467–470, 472–475, 477, 483, 494, 501, 559, 567
 - conformal, 467, 475, 477
 - projective, 468, 477, 483, 494, 501
- grafting compactification, 494
- graph
 - bicolored, 163
 - dual, 174
 - labelled, 174
 - reduced, 181
 - merging, 164
 - quotient, 163
 - reduced, 163
 - ribbon, 162, 646, 772, 774
 - dual, 166, 170
 - enriched, 164, 168
 - metrized, 167
 - nonsingular enriched, 165
 - topological realization, 166
 - sub-, 163
- gravity, 538
- Green operator, 226
- Grothendieck correspondence, 772
- Grothendieck–Teichmüller group, 787
- group
 - amenable, 332
 - Artin, 403
 - automatic, 419
 - biautomatic, 419
 - braid, 393
 - commensurator, 813, 838
 - Coxeter, 403
 - exact, 362
 - fully biautomatic, 419
 - geodesically biautomatic, 419
 - perfect, 244, 282
 - surface braid, 399
 - uniformly perfect, 244, 282
- group of fractions, 413
- groupoid, 309
 - amenable, 335
 - discrete measured, 311
 - of infinite type, 316
 - principal, 313
- growth rate, 284
- (G, X) -structure, 540
- Hain–Reed function, 233
- half-plane, 648
- Harer’s stability, 207
- harmonic map, 457, 475, 494–497, 500
- Hecke operators, 145
- hexagon orbifold, 631
- H -hull, 569
- hierarchy, 734
- Hodge $*$ -operator, 220
- Hodge bundle, 179, 224
- Hodge decomposition, 227
- holomorphic family of Riemann surfaces, 94, 97
 - globally trivial, 97
 - holomorphically equivalent, 97
 - isomorphic, 97
 - locally trivial, 97
 - trivial family, 97
- holomorphic quadratic differential, 513
- holomorphically equivalent families, 97
- holonomy, 540
- holonomy map, 478
- holonomy representation, 459, 460, 513
- holonomy transversality, 501
- homogeneous quasi-homomorphism, 242
- homology with local coefficients, 432
- homotopical monodromy, 104
- homotopical monodromy of a representation, 104
- homotopy long exact sequence, 397
- Hopf differential, 495–497, 499
- Hopfian group, 402

- horizon, 597
- horizontal cylinder, 797
- horizontal foliation, 492
- horizontal lamination, 492
- Hubbard–Masur theorem, 492
- Hurwitz
 - equivalence, 280
 - move, 279
- hyperbolic (δ -), 246
- hyperbolic geometry, 512
- hyperbolic isometry, 247, 650
- hyperbolic metric
 - transversely locally constant, 821
- hyperbolic modular transformation, 102
- hyperbolic three-space, 634
- hyperbolization map, 527
- hyperelliptic component, 371
- hyperelliptic curve, 133
- hyperelliptic mapping class group, 288
- hyperplane arrangement, 424

- ideal triangulation, 159, 548
- independent pseudo-Anosov
 - elements, 252
- initial singularity, 574
- initial state, 418
- injectivity of character map for \mathbb{F}_2 , 625
- intersection diagram, 436
- intersection graph, 805
- intersection number, 744
- invariant Borel map, 336
- invariant theory, 690
- inverse limit, 812
- involution, 632, 633
- irreducible Artin group, 404
- irreducible Coxeter group, 404
- irreducible fiber, 274
- irreducible group, 254
- irreducible modular transformation, 102
- irreducible representation, 626, 628
- islands of quasi-Fuchsian holonomy, 489
- isomorphic, 97
- isomorphism, 117
 - of projective structures, 458
- iterated transition function, 418
- Johnson homomorphism, 229, 230
 - extended, 230
- KAT solution, 520
- Kauffman Bracket Skein Module, 697
- Kleinian group, 522
- Kobayashi distance, 100
- Kobayashi metric, 75, 473
- Kodaira surface, 98
- Kodaira–Spencer map, 220
- $K(\pi, 1)$ space, 398

- lamination signature, 554
- language
 - regular, 418
- large group, 265
- lattice, 254, 297
- leaf, 812
- Lefschetz fibration, 273
 - achiral, 290
 - broken, 292
 - genus of a, 274
 - relatively minimal, 274
- Lefschetz pencil, 273, 278
- length
 - as intersection number, 161
 - extremal, 185
 - horocyclic, 191
 - hyperbolic, 175
 - in a ribbon graph, 168
- length/twist parameters, 546
- level, 734
- Lie product, 638
- Lie product, geometric interpretation, 637
- lifting representations to $SL(2, \mathbb{C})$, 630
- limit set, 596
- linear fractional transformation, 511
- linear twist, 115
- LKB representation, 430
- locally trivial, 97
- locally $CAT(0)$, 495

- locally convex pleated plane, 470
- locally convex pleated plane, 482
- locally convex pleated surface, 488
- Lorentzian grafting, 567

- Möbius transformation, 511
- Magnus expansion, 231
 - harmonic, 231
- map
 - extremal, 815
- mapping class
 - pure, 306
- mapping class group, 48, 51–53, 81, 157, 251, 275, 303, 398
 - extended, 303
 - hyperelliptic, 288
- marked length spectrum, 548
- marked measure spectrum, 552
- marked projectively equivalent, 512
- marked Riemann surface, 98
 - equivalent, 99
- marking, 512
- maximal globally hyperbolic spacetimes, 540
- maximal round disk, 470
- ME coupling, 298
- measurable Banach bundle, 334
- measure
 - invariant, 311
 - quasi-invariant, 311
- measure equivalence (ME), 298
- measure space
 - standard, 309
 - standard finite, 309
- measure space isomorphism, 309
- measured geodesic lamination, 551
- measured lamination, 468
 - horizontal, 492
 - projective, 491
 - vertical, 492
- metric, 168
- minimal volume, 262
- Minkowski space, 541

- modular curve, 144
- modular structure, 737, 752
 - compact, 752
- moduli field, 780
- moduli map, 104
- moduli space, 117, 172
- modulus, 185
 - reduced, 185
- momentum, 597
- monodromy, 275, 777, 798
 - factorization, 276
 - map, 772, 775, 789
 - representation, 275
 - transitive, 281
- monodromy representation, 435, 459
- monoid
 - atomic, 412
 - Garside, 413
- Mordell conjecture, 105
- Morita–Mumford class, 218
- morphism
 - of a nonsingular graph, 165
- multi black hole, 599
- multi white hole, 599
- multi-curve, 251, 253
- multiplication, 754
- Mumford conjecture, 207

- nearest-point projection, 470
- negative Dehn twist, 109
- negative root, 407
- negative type, 95, 111
- nerve, 515
- Newton class, 228
- node, 116
- non-amphidrome, 111
- non-compact solenoid, 813
- non-elementary representation, 478
- nonpositively curved (NPC), 474
- norm
 - of a quadratic differential, 466
 - on a monoid, 413
- normal form, 418

- normal form for characters of \mathbb{F}_2 , 644
 normal form in a Garside monoid, 417
 Novikov additivity, 283

 open book decomposition, 291
 orbit equivalence (OE), 265, 314
 Ore conditions, 414
 orientable double covering space, 675
 origami, 382, 788
 rules, 789
 Veech group of, 792
 origami curve, 792
 cusp of, 796
 orthogonality, 493
 osculation, 462, 463
 Osgood–Stowe Schwarzian derivative, 500

 pants decomposition, 176, 479, 488, 546, 657
 pants graph, 49, 53
 parabolic modular transformation, 102
 parabolic type, 110
 parameter
 length, 176
 twist, 176
 part, 102, 116
 particle, 602
 past, 538
 past boundary, 586
 past part of a AdS spacetime, 587
 period coordinates, 372
 period map, 225
 periodic, 110
 Perron–Vannier representation, 437
 Peter–Weyl Theorem, 707
 planar graph, 774
 plaque, 470, 481–483
 pleated plane
 convex, 469
 locally convex, 470, 482
 pleated surface
 locally convex, 488
 pleating map, 528

 pluricanonical map, 141
 Poincaré series, 143
 Poincaré theta series, 84
 Poisson bracket, 697
 positive root, 407
 precise admissible system of curves, 111
 presentation, 816
 product of subgroups, 258
 product relation, 663, 665
 profinite completion, 783
 projection, 191, 514
 projective deformation, 480
 projective geometry
 complex, 511
 projective line, 771
 complex, 511
 projective measured lamination, 491
 projective metric, 473
 projective Riemann surface, 512
 projective structure, 512
 branched, 479
 complex, 458
 embedded, 469
 exotic Fuchsian, 483
 exotic quasi-Fuchsian, 486
 Fuchsian, 483
 quasi-Fuchsian, 465
 relative, 465
 standard Fuchsian, 461
 standard quasi-Fuchsian, 486
 projectively immersed disk, 473
 proper simplex, 157
 pseudo-Anosov, 252
 pseudo-hyperbolic modular
 transformation, 102
 pseudo-periodic map, 94, 110
 amphidrome, 111
 negative type, 95
 non-amphidrome, 111
 of negative type, 111
 of parabolic type, 110
 periodic, 110

- reducible with all component maps being of finite order, 110
 - screw number, 111
 - standard form, 115
 - surface transformation of
 - algebraically finite type, 110
- Puiseux series, 784
- pure braid group, 393
- pure gravity, 539
- pure mapping class, 306
- quadratic differential
 - critical point of, 183
 - holomorphic, 513
 - integrable holomorphic, 177
 - Jenkins–Strebel, 185, 188
 - meromorphic, 182
 - regular point of, 183
 - transversely locally constant, 826
- quake cocycle, 561
- quakebend, 481
- quasi-Fuchsian component, 486
- quasi-Fuchsian section, 465
- quasi-Fuchsian space, 478
- quasi-geodesic, 247
- quasi-homomorphism, 242
- quasi-invariant measure, 311
- quasiconformal, 66
- quasimorphism, 242
- quaternion origami, 804
- \mathbb{R} -tree, 473
- rank three free group, 680
- rational
 - 4-manifold, 283
 - elliptic surface, 280
- Rauch’s variational formula, 225
- real character, 643
- realizing path, 249
- recoupling formula, 719
- reduced expression, 408
- reduced map, 102
- reduced norm, 186
- reduced word, 415
- reducible
 - element, 252
 - map, 102
 - subgroup, 253
- reducible fiber, 274
- reducible with all component maps being of finite order, 110
- reductive, 689
- regular domain, 574
- regular fiber, 274
- regular language, 418
- regular projective curve, 769
- Reich–Strebel inequality, 815, 829
- relative Euler class, 671
- relative projective structure, 465
- relative tangent bundle, 218
- relatively minimal Lefschetz fibration, 274
- relaxed signature, 556
- representation, 103, 756
 - reducible, 757
- representation variety, 689
- rescaling, 567
- residually finite, 257
- residually finite group, 402
- residue, 222
 - of a quadratic differential, 222
- resolution, 752
- restriction, 103, 311
- resultant, 397
- retraction on the singularity, 575
- reversible word, 415
- ribbon graph, 162, 646, 772, 774
- Riemann sphere, 511
- Riemann surface
 - finite analytic type, 73
 - infinite analytic type, 74
- Riemann surface with nodes, 116
 - allowable map of, 117
 - deformation of, 117
 - deformation space of, 117

- moduli space, 117
 - node of, 116
 - part of, 116
- Riemann–Roch theorem, 769
- right hexagon orbifold, 631
- Rigidity theorem, 104
- ring of invariants, 689
- root system, 407
- round disk, 470, 578
- ruled
 - 4-manifold, 283
 - surface, 280
- saddle connection, 371
- Salvetti complex, 425
- saturation, 311
- Schur’s Lemma, 692, 703, 708
- Schwarzian compactification, 497
- Schwarzian derivative, 69, 133, 461, 462, 500, 514
- Schwarzian equation, 134, 463
- Schwarzian of a projective structure, 464
- Schwarzian parameterization, 461
- screw number, 111
- section, 274, 514
- self-bumping, 487
- semi-stable reduction, 173
- separation, 257
- Shafarevich conjecture, 105
- shape operator, 540
- shear parameters, 548
- Siegel upper halfspace, 224
- σ -operator, 160
- simple element, 410, 417
- simple graph, 517
- simple root, 407
- simplicial part, 551
- simplicial volume, 262
- singly degenerate, 490
- singular fiber, 274
- singularity, 371
- $6j$ -symbols, 721
- sliding circuits, 420
- small type Coxeter graph, 430
- solenoid
 - compact, 811, 817
 - extremal map, 815
 - fiber bundle description, 812
 - G -tagged, 814
 - G -tagged compact, 820
 - G -tagged non-compact, 821
 - modular group, 813, 838
 - non-compact, 813, 840
- spacetime, 538
- special twist, 115
- spherical geometry, 511
- spin network, 693
 - calculus, 696
 - component maps, 693
 - equivariance, 697
 - looping relation, 701
 - reflection, 695, 696, 706, 707
 - trivalent, 702
- spine, 191, 194
- sporadic surface, 252
- square tiled surface, 382, 791
- stability of quasi-geodesics, 247
- stable commutator length, 243, 282
- stable reduction, 796
- stable Riemann surface, 796
- stable torsion length, 284
- standard Anti de Sitter spacetime, 586
- standard de Sitter spacetime, 584
- standard form, 115
- state, 418
- Stein structure, 291
- straightening, 492
- strata, 470, 472–475, 499
- stratification, 174
 - canonical, 470
- strongly uniformizable, 529
- structure morphism, 778
- subgroup
 - elementary, 248
 - irreducible amenable (IA), 307

- reducible, 253, 307
 - sufficiently large, 253, 307
- subgroupoid, 310
 - irreducible amenable (IA), 343
 - normal, 350
 - reducible, 343
- subsurface
 - complementary, 157
 - invisible, 158
 - visible, 158, 169, 194
- sum relation, 663, 664
- superrigidity, 254
- support plane, 470
- surface
 - decorated, 191
 - stable hyperbolic, 192
 - truncated, 191
- surface braid group, 399
- surface bundle, 289
- surface transformation of algebraically finite type, 110
- surjectivity of character map for \mathbb{F}_2 , 622
- SU(2)-representation, 643
- syllable length, 287
- Sylvester matrix, 396
- symmetric language, 418
- symmetrizer, 698
 - invariance, 699
 - recurrence, 700
 - stacking, 699
- symplectic basis, 226
- symplectic manifold, 278
- system of annuli, 185
- systole, 161, 168

- T-symmetry, 601
- tangent map, 476
- tangentiable, 476
- tangential base point, 784
- Taubes theorem (Lefschetz fibration), 284
- tautological
 - ring, 179
 - system, 179
 - section, 173
- Teichmüller curve, 790, 792
- Teichmüller disk, 90, 791
- Teichmüller distance, 100
 - solenoid, 823
- Teichmüller embedding, 791
- Teichmüller metric, 69, 76
 - definition, 48
 - geodesic-length comparison expressions, 57
- Teichmüller modular group, 100
 - of a finitely generated Fuchsian group, 100
 - of a Riemann surface, 100
- Teichmüller modular transformation, 100
 - elliptic, 102
 - homotopical monodromy, 104
 - hyperbolic, 102
 - of a Riemann surface, 100
 - parabolic, 102
 - pseudo-hyperbolic, 102
 - topological monodromy, 104
- Teichmüller space, 68, 99, 171, 193, 289, 513, 543, 813
 - automorphism group, 81
 - Bers Teichmüller space, 99
 - complex structure, 73
 - of a finitely generated Fuchsian group, 99
 - of a Riemann surface, 99
 - solenoid, 823
 - universal, 68
- Teichmüller space of the torus, 743
- Teichmüller-type map, 815
- Temperley-Lieb algebra, 699
- Theorem of Belyi, 771
- thick–thin decomposition, 48
- thimble, 275
- three-by-three real symmetric matrix, 644
- three-dimensional inner product space, 640

- three-dimensional orthogonal representation, 639
- Thurston boundary, 305
- Thurston compactification, 305, 491
- Thurston coordinates, 528
- Thurston invariant, 741
- Thurston map, 498
- Thurston metric, 473, 499
- tight geodesic, 260
- time orientation, 538
- Tits alternative, 248, 253
- topological generators, 784
- topological monodromy, 94, 104
- topological monodromy of a representation, 104
- topological realization
 - of a graph, 166
 - of an enriched ribbon graph, 169
- topology of surfaces, 617
- Torelli group, 229, 277
- torsion length, 284
 - stable, 284
- torus
 - once-punctured, 788
- trace, 617
- trace diagram, 697
- trace function, 757
- trace function of a metric, 758
- trajectory
 - closed, 183
 - critical, 183
 - horizontal, 183
 - periodic, 184
 - vertical, 183
- transition function, 418
- translation length, 247
- translation structure, 791
- transvection, 442, 443
- transverse measure, 812
- triangulation complex, 816
- trivalent spin network, 702
- trivial family, 97
- two-holed torus, 674
- type of a hyperbolic surface, 545
- type of a reducible singular fiber, 274
- uniformization, 175
- uniqueness property (language), 418
- unitary trick, 707
- universal Teichmüller space, 68
- valence
 - hyperbolic surface, 191
- valency list, 783
- valency of an oriented simple closed curve, 111
- vanishing cycle, 275
- Veech group, 369, 791
- vertex
 - invisible, 181
 - marked, 164
 - nodal, 164
 - special, 164
 - visible, 181
- vertical lamination, 492
- vertical foliation, 492
- very large automorphism group, 137
- Virasoro cocycle, 223
- virtual automorphism group, 838
- virtual duality group, 162
- visible equivalence, 181
- wall, 427
- weak orbit equivalence (WOE), 315
- weight, 160
 - projective, 160
- weighted part, 551
- Weil–Petersson
 - bivector field, 194
 - completion, 177
 - form, 178, 194
 - Hermitian pairing, 177
 - metric, 178, 194
- Weil–Petersson metric, 47, 222
 - CAT(0) geometry, 52
 - Alexandrov tangent cone, 59
 - Brock’s quasi isometric model, 53

- convexity, [54](#)
- current research themes, [47](#)
- definition, [48](#)
- dense geodesics, [56](#)
- distance to augmentation set, [56](#)
- geodesic-length comparison
 - expressions, [57](#)
- model metric, [58](#)
- Weil–Petersson pairing, [222](#)
- Whittaker’s conjecture, [137](#)
- Wick rotation, [566](#)
- word length, [407](#)
- word problem, [408](#)
- word-hyperbolic group, [246](#)
- WPD, weak proper discontinuity, [252](#)
- wrapping invariant, [485](#)