

IRMA Lectures in Mathematics and Theoretical Physics 11

Series Editor: Vladimir G. Turaev

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

IRMA Lectures in Mathematics and Theoretical Physics

Edited by Vladimir G. Turaev

This series is devoted to the publication of research monographs, lecture notes, and other materials 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.)

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

Handbook of Teichmüller Theory Volume I

Athanase Papadopoulos
Editor



European Mathematical Society

Editor:

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

2000 Mathematics Subject Classification: Primary 30-00, 32G15, 30F60; secondary 30C62, 57N16, 53A35, 20F65, 30F20, 30F25, 30F10, 30F30, 30F40, 30F45, 14H15, 20H10, 30F15, 53B35, 57M07, 57M60, 14H60, 14D20, 57M20, 20F38.

ISBN 978-3-03719-029-6

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.

© 2007 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 \TeX files: I. Zimmermann, Freiburg
Printed in Germany

9 8 7 6 5 4 3 2 1

Foreword

In a broad sense, the subject of Teichmüller theory is the study of moduli spaces for geometric structures on surfaces. This subject makes important connections between several areas in mathematics, including low-dimensional topology, hyperbolic geometry, dynamical systems theory, differential geometry, algebraic topology, representations of discrete groups in Lie groups, symplectic geometry, topological quantum field theory, string theory, and there are others.

The central object in this theory is the Teichmüller space of a surface. As is well-known, this space can be seen from different points of view. It is a space of equivalence classes of conformal structures on the surface, a space of equivalence classes of hyperbolic metrics on this surface, and a space of equivalence classes of representations of the fundamental group of the surface into a Lie group (primarily, $\mathrm{PSL}(2, \mathbb{R})$, but there are several others). These three points of view are equally important in the study of Teichmüller space and, in several situations, they cannot be clearly separated. Furthermore, Teichmüller space inherits from these different points of view various structures, including several interesting metrics (Teichmüller, Weil–Petersson, Thurston, Bergman, Carathéodory, Kähler–Einstein, McMullen, etc.), a natural complex structure, a symplectic structure, a real analytic structure, the structure of an algebraic set, cellular structures, various boundary structures, a natural discrete action by the mapping class group, a quantization theory of its Poisson and symplectic structures, a measure-preserving geodesic flow, a horocyclic flow, and the list of structures goes on and on. Teichmüller theory is growing at a fantastic rate, and the richness of this theory is to a large extent a consequence of the diversity and the richness of the structures that Teichmüller space itself carries.

The purpose of this Handbook is to give a comprehensive picture of the classical and of the recent developments in Teichmüller theory. The range of this picture will hopefully include all the aspects mentioned above although, because new ideas and new connections come out regularly in this theory, the picture will necessarily be incomplete. Nevertheless we hope that the Handbook will reflect the beauty, the liveliness and the richness of Teichmüller theory.

I have tried to divide this first volume of the Handbook into parts. There were several possibilities, none of which imposed itself as being more natural or more efficient than the others. I finally ended up with the following division into four parts.¹

- The metric and the analytic theory, 1
- The group theory, 1
- Surfaces with singularities and discrete Riemann surfaces
- The quantum theory, 1

¹The titles of some parts are followed by the numeral 1 because they will be continued in Volume II of the Handbook.

Volume II of the Handbook will contain more material on the analytic and the metric theory, on the group theory, on the quantum theory, as well as sections on cluster algebras, on representation theory and higher Teichmüller theory, on complex projective structures, on the Teichmüller geodesic flow and other dynamical aspects, and on the Grothendieck–Teichmüller theory.

I would like to take this opportunity to thank Vladimir Turaev for his encouragement in this project, Manfred Karbe and Irene Zimmermann for their invaluable collaboration in the final editing process and, of course, all the authors who contributed to the Handbook. It was a pleasure to work with all these people.

Strasbourg, April 2007

Athanase Papadopoulos

Contents

Foreword.....	v
Introduction to Teichmüller theory, old and new <i>by Athanase Papadopoulos</i>	1
Part A. The metric and the analytic theory, 1	
Chapter 1. Harmonic maps and Teichmüller theory <i>by Georgios D. Daskalopoulos and Richard A. Wentworth</i>	33
Chapter 2. On Teichmüller’s metric and Thurston’s asymmetric metric on Teichmüller space <i>by Athanase Papadopoulos and Guillaume Th��ret</i>	111
Chapter 3. Surfaces, circles, and solenoids <i>by Robert C. Penner</i>	205
Chapter 4. About the embedding of Teichmüller space in the space of geodesic H��lder distributions <i>by Jean-Pierre Otal</i>	223
Chapter 5. Teichmüller spaces, triangle groups and Grothendieck dessins <i>by William J. Harvey</i>	249
Chapter 6. On the boundary of Teichmüller disks in Teichmüller and in Schottky space <i>by Frank Herrlich and Gabriela Schmith��sen</i>	293
Part B. The group theory, 1	
Chapter 7. Introduction to mapping class groups of surfaces and related groups <i>by Shigeyuki Morita</i>	353
Chapter 8. Geometric survey of subgroups of mapping class groups <i>by Lee Mosher</i>	387

Chapter 9. Deformations of Kleinian groups <i>by Albert Marden</i>	411
Chapter 10. Geometry of the complex of curves and of Teichmüller space <i>by Ursula Hamenstädt</i>	447
Part C. Surfaces with singularities and discrete Riemann surfaces	
Chapter 11. Parameters for generalized Teichmüller spaces <i>by Charalampos Charitos and Ioannis Papadoperakis</i>	471
Chapter 12. On the moduli space of singular euclidean surfaces <i>by Marc Troyanov</i>	507
Chapter 13. Discrete Riemann surfaces <i>by Christian Mercat</i>	541
Part D. The quantum theory, 1	
Chapter 14. On quantizing Teichmüller and Thurston theories <i>by Leonid O. Chekhov and Robert C. Penner</i>	579
Chapter 15. Dual Teichmüller and lamination spaces <i>by Vladimir V. Fock and Alexander B. Goncharov</i>	647
Chapter 16. An analog of a modular functor from quantized Teichmüller theory <i>by Jörg Teschner</i>	685
Chapter 17. On quantum moduli space of flat $\mathrm{PSL}_2(\mathbb{R})$ -connections <i>by Rinat M. Kashaev</i>	761
List of Contributors	783
Index	785

Introduction to Teichmüller theory, old and new

Athanase Papadopoulos

Teichmüller theory is one of those few wonderful subjects which bring together, at an equally important level, fundamental ideas coming from different fields. Among the fields related to Teichmüller theory, one can surely mention complex analysis, hyperbolic geometry, the theory of discrete groups, algebraic geometry, low-dimensional topology, differential geometry, Lie group theory, symplectic geometry, dynamical systems, number theory, topological quantum field theory, string theory, and there are many others.

Let us start by recalling a few definitions.

Let $S_{g,p}$ be a connected orientable topological surface of genus $g \geq 0$ with $p \geq 0$ punctures. Any such surface admits a complex structure, that is, an atlas of charts with values in the complex plane \mathbb{C} and whose coordinate changes are holomorphic. In the classical theory, one considers complex structures $S_{g,p}$ for which each puncture of $S_{g,p}$ has a neighborhood which is holomorphically equivalent to a punctured disk in \mathbb{C} . To simplify the exposition, we shall suppose that the orientation induced on $S_{g,p}$ by the complex structure coincides with the orientation of this surface. Homeomorphisms of the surface act in a natural manner on atlases, and two complex structures on $S_{g,p}$ are said to be equivalent if there exists a homeomorphism of the surface which is homotopic to the identity and which sends one structure to the other. The surface $S_{g,p}$ admits infinitely many non-equivalent complex structures, except if this surface is a sphere with at most three punctures.

To say things precisely, we introduce some notation. Let $\mathcal{C}_{g,p}$ be the space of all complex structures on $S_{g,p}$ and let $\text{Diff}^+(S_{g,p})$ be the group of orientation-preserving diffeomorphisms of $S_{g,p}$. We consider the action of $\text{Diff}^+(S_{g,p})$ by pullback on $\mathcal{C}_{g,p}$. The quotient space $\mathcal{M}_{g,p} = \mathcal{C}_{g,p}/\text{Diff}^+(S_{g,p})$ is called Riemann's moduli space of deformations of complex structures on $S_{g,p}$. This space was considered by G. F. B. Riemann in his famous paper on Abelian functions, *Theorie der Abel'schen Functionen*, Crelle's Journal, Band 54 (1857), in which he studied moduli for algebraic curves. In that paper, Riemann stated, without giving a formal proof, that the space of deformations of equivalence classes of conformal structures on a closed orientable surface of genus $g \geq 2$ is of complex dimension $3g - 3$.

The Teichmüller space $\mathcal{T}_{g,p}$ of $S_{g,p}$ was introduced in the 1930s by Oswald Teichmüller. It is defined as the quotient of the space $\mathcal{C}_{g,p}$ of complex structures by the group $\text{Diff}_0^+(S_{g,p})$ of orientation-preserving diffeomorphisms of $S_{g,p}$ that are isotopic to the identity. The group $\text{Diff}_0^+(S_{g,p})$ is a normal subgroup of $\text{Diff}^+(S_{g,p})$, and the quotient group $\Gamma_{g,p} = \text{Diff}^+(S_{g,p})/\text{Diff}_0^+(S_{g,p})$ is called the mapping class group of $S_{g,p}$ (sometimes also called the modular group, or the Teichmüller modular group)

of $S_{g,p}$. The mapping class group $\Gamma_{g,p}$ acts on the Teichmüller space $\mathcal{T}_{g,p}$, and the quotient of $\mathcal{T}_{g,p}$ by this action is Riemann's moduli space $\mathcal{M}_{g,p}$.

During a remarkably brief period of time (1935–1941), Teichmüller wrote about thirty papers which laid the foundations of the theory which now bears his name. After Teichmüller's death in 1943 (at the age of 30), L. V. Ahlfors, L. Bers, H. E. Rauch and several of their students and collaborators started a project that provided a solid grounding for Teichmüller's ideas. The realization of this project took more than two decades, during which the whole complex-analytic theory of Teichmüller space was built. In the 1970s, W. P. Thurston opened a new and wide area of research by introducing beautiful techniques of hyperbolic geometry in the study of Teichmüller space and of its asymptotic geometry. Thurston's work highlighted this space as a central object in the field of low-dimensional topology. Thurston's ideas are still being developed and extended today by his students and followers. In the 1980s, there also evolved an essentially combinatorial treatment of Teichmüller and moduli spaces, involving techniques and ideas from high-energy physics, namely from string theory. The current research interests in Teichmüller theory from the point of view of mathematical physics include the quantization of Teichmüller space using the Weil–Petersson symplectic and Poisson geometry of this space, as well as gauge-theoretic extensions of these structures. The quantization theories are also developed with the purpose of finding new invariants of hyperbolic 3-manifolds.

My aim in the rest of this introduction is twofold: first, to give an exposition of the structure of Teichmüller theory that is studied in this volume of the Handbook, and second, to give an overview of the material contained in this volume. The goal is to give a flavour of the subject. Of course, the exposition will be very sketchy at some places.¹

1 An overview of the structure of Teichmüller space

Teichmüller space $\mathcal{T}_{g,p}$ has a natural topology which makes it homeomorphic to an open ball of dimension $6g - 6 + 2p$. It has a complex structure which is induced by an embedding of this space as a bounded pseudo-convex domain of holomorphy in a Banach space. It carries several interesting metrics, in addition to the usual metrics that are associated to a bounded domain of holomorphy viz. the Carathéodory, the Kobayashi and the Bergman metrics. It also has a symplectic and a Poisson structure, various boundary structures and several other structures which I now briefly describe.

¹In order not to overload this introduction, I decided not to give the complete references for the results that I cite. Most of the references are given in the chapters that follow this introduction and, furthermore, with the availability of electronic supports, it is nowadays easy to find the exact reference for a result, given its statement, its author's name and the year it was written.

1.1 Complex structure

Let $S_{g,p}$ be as before a surface of genus $g \geq 0$ with $p \geq 0$ punctures. If the Euler characteristic of $S_{g,p}$ is negative, then the Teichmüller space $\mathcal{T}_{g,p}$ of $S_{g,p}$ is a complex manifold of complex dimension $3g - 3 + p$. The bases of the complex analytic theory of Teichmüller space were developed by Ahlfors and Bers. There are several ways of defining the complex structure of Teichmüller space. All of them use deep results from analysis, and giving the details of this theory would be far beyond the scope of this introduction. We only mention that there is a famous embedding (called Bers' embedding) which realizes $\mathcal{T}_{g,p}$ as an open bounded pseudo-convex (but not strictly pseudo-convex) domain in \mathbb{C}^{3g-3+p} . This embedding depends on the choice of a basepoint in Teichmüller space. Classically, the holomorphic cotangent space at a point of Teichmüller space is described by the Kodaira–Spencer deformation theory, which parametrizes the deformations of a conformal structure in terms of the sheaf of holomorphic vector fields on the given Riemann surface. In Bers' embedding of $\mathcal{T}_{g,p}$, the holomorphic cotangent space at a point is identified with the vector space of holomorphic quadratic differentials with simple poles at the punctures on a Riemann surface representing the point. By the theory of quasiconformal mappings, any complex structure on a surface is specified by a Beltrami differential of norm less than one, and this leads to a description of the holomorphic tangent space at a point of Teichmüller space as a vector space of Beltrami differentials of norm less than one divided by a subspace of differentials which induce trivial deformations. We note that besides providing a description of the complex structure of Teichmüller space, these descriptions of the tangent and cotangent spaces are used to give an infinitesimal description of the Teichmüller metric which we recall below, and that as such, the tangent and cotangent spaces are Banach spaces and not Hilbert spaces, which reflects the fact that the Teichmüller metric is a Finsler metric and not a Riemannian metric. We also note that the descriptions of the tangent and cotangent spaces by holomorphic quadratic differentials and Beltrami differentials are key tools for the definition of Hermitian metrics on Teichmüller space, for instance the Weil–Peterson metric which we discuss below.

It follows from the definitions that the mapping class group $\Gamma_{g,p}$ preserves this complex structure, that is, it acts on Teichmüller space by biholomorphic maps. Conversely, by a result of H. L. Royden, any biholomorphic automorphism of Teichmüller space is induced by an element of the mapping class group. The quotient moduli space $\mathcal{M}_{g,p}$ is a complex orbifold of dimension $3g - 3 + p$.

1.2 Metric structures

Teichmüller space carries various metrics, each arising naturally from a particular viewpoint on the space. There are three metrics on Teichmüller space which will be considered in some detail in this volume: Teichmüller's metric, the Weil–Peterson metric and Thurston's asymmetric metric (which, because of its asymmetry, is not a

metric in the usual sense of the word). I will say a few words about each of these metrics.

Teichmüller's metric. This metric is obtained by first defining the distance between two conformal structures g and h on the surface $S_{g,p}$ to be $\frac{1}{2} \inf_f \log K(f)$, where the infimum is taken over all quasiconformal homeomorphisms $f: (S_{g,p}, g) \rightarrow (S_{g,p}, h)$ that are isotopic to the identity and where $K(f)$ is the quasiconformal dilatation of f . Teichmüller showed that the infimum is realized by a quasiconformal homeomorphism, and he gave a description of this homeomorphism in terms of a quadratic differential on the domain conformal surface $(S_{g,p}, g)$. This distance function on the set of conformal structures is invariant by the action of the group of diffeomorphisms isotopic to the identity on each factor, and it induces a distance function on Teichmüller space $\mathcal{T}_{g,p}$, which is Teichmüller's metric. Teichmüller's metric is a complete Finsler metric which is not Riemannian unless the surface is a torus, in which case Teichmüller space, equipped with Teichmüller's metric, is isometric to the 2-dimensional hyperbolic plane. Teichmüller's metric is geodesically convex, that is, any two points are joined by a unique geodesic segment. The metric is also uniquely geodesic, that is, the geodesic segment joining two arbitrary points is unique. Furthermore, any geodesic segment can be extended in a unique way (up to parametrization) to a geodesic line. For some time, it was believed that Teichmüller's metric was nonpositively curved in the sense of Busemann, that is, that the distance function between two parametrized geodesics is convex. Indeed, in 1959, S. Kravetz published a paper containing this statement as one of the main results. In 1971, M. Linch found a mistake in the arguments of Kravetz and in 1975, H. Masur showed that the statement is false, by providing examples of two geodesic rays whose associated distance function is bounded, and therefore is not convex. By a result of Masur and Wolf (1995), Teichmüller's metric is not Gromov hyperbolic. Despite these facts, much of the work that has been done on Teichmüller's metric is motivated by analogies with metrics on simply connected manifolds of negative curvature. We can mention in this connection that there is a Teichmüller geodesic flow which preserves a natural measure, which induces a quotient geodesic flow on moduli space, and that this quotient flow, like the geodesic flow of a finite volume hyperbolic manifold, is ergodic. (This result was obtained independently in 1982 by H. Masur and W. Veech.)

The mapping class group $\Gamma_{g,p}$ of $S_{g,p}$ acts on $\mathcal{T}_{g,p}$ by isometries of the Teichmüller metric and, conversely, by a result of Royden (1970), every isometry group of the Teichmüller metric is induced by an element of the mapping class group. At the same time, Royden showed that the Teichmüller metric can be recovered directly from the complex structure of Teichmüller space, by proving that this metric coincides with the Kobayashi metric of this space. (In fact, Royden proved the first result as a corollary of the second.)

The Weil–Petersson metric. This metric is a Riemannian metric which is also closely related to the complex structure of Teichmüller space. Its definition starts with an L^2 -norm on the space of quadratic differentials at each point X of Teichmüller space,

considered as the holomorphic cotangent space at that point. The definition of this norm makes use of the hyperbolic metric of a surface representing this point. More precisely, this norm is given by the formula $\|\phi\|^2 = \int_X \rho^{-2}(z)|\phi(z)|^2|dz|^2$, with $\rho(z)|dz|$ being the length element on the point X represented by a hyperbolic surface. The Weil–Petersson inner product on the tangent space is then defined by taking a dual of this L^2 inner product, the duality between the tangent and cotangent spaces being defined through a natural pairing between quadratic differentials and Beltrami differentials. Several chapters of this Handbook deal with this metric, and we shall list here some of its important geometric features. L. Ahlfors showed that the Weil–Petersson metric is Kähler. It has variable negative sectional curvature (a result proved independently by A. Tromba and S. Wolpert). It is geodesically convex but it is not complete (a result obtained by Masur and by Wolpert). Masur studied the completion of this metric, and he described the points on the frontier of this completion as Riemann surfaces with nodes². More precisely, Masur identified the Weil–Petersson completion with the augmented Teichmüller space, whose frontier is a stratified union of lower-dimensional spaces, each of which is the Teichmüller space of a nodal surface. He showed that in some precise sense the tangential component of the Weil–Petersson metric on Teichmüller space extends to the Weil–Petersson metric of these boundary Teichmüller spaces. The sectional curvature of the Weil–Petersson metric is unbounded (its supremum is zero and its infimum is $-\infty$), except in the cases where the complex dimension of Teichmüller space is 1. Ahlfors showed that the holomorphic sectional curvature of the Weil–Petersson metric is negative. Wolpert showed that the holomorphic sectional curvature and the Ricci curvature of this metric have negative upper bounds (proving a conjecture made by Royden), with these upper bounds expressible in terms of the topological type of the surface. The mapping class group acts by biholomorphic isometries on Teichmüller space equipped with the Weil–Petersson metric, and this metric descends to a metric on moduli space. Masur and Wolf proved (2002) that conversely, every Weil–Petersson isometry is induced by an element of the mapping class group, except for a few surfaces of low genus. The untreated cases were completely analyzed later on by J. Brock and D. Margalit (2004). J. Brock and B. Farb (2004) showed that the Weil–Petersson metric is not Gromov hyperbolic, except if the surface $S_{g,p}$ is a torus with at most two punctures or a sphere with at most five punctures. In all the other cases, the space has higher rank in the sense of Gromov. S. Yamada (2001) showed that the Weil–Petersson completion of the Weil–Petersson metric is a CAT(0)-space (that is, a non-positively curved space in the sense of Cartan–Alexandrov–Toponogov). Brock initiated a new point of view on the

²We note that surfaces with nodes appear naturally in complex analysis. A node is the simplest type of a singularity of a complex curve. More precisely, while a smooth point of a complex curve is a point which admits a neighborhood which is biholomorphically equivalent to an open disk in \mathbb{C} , a node is a point which admits a neighborhood which is biholomorphically equivalent to a space obtained by taking a disjoint union of two disks in \mathbb{C} and gluing them along a point. Riemann surfaces with nodes are complex curves whose points are either smooth points or nodes, and they appear naturally as degenerations of smooth complex curves (that is, of Riemann surfaces). Thus, it is not surprising that surfaces with nodes appear as boundary structures of spaces of Riemann spaces.

Weil–Petersson metric, by establishing a quasi-isometry between this metric and the metric on the vertices of the pants graph of the surface induced by the simplicial metric on the 1-skeleton. This work has connections with volumes of quasi-Fuchsian hyperbolic three-manifolds, and it also leads to beautiful results on the global behaviour of Weil–Petersson geodesics, quasi-geodesics of the Weil–Petersson metric being in correspondence with quasi-geodesics in the pants graph. There is a result which has a similar flavour, on the comparison between quasi-geodesics of the Teichmüller metric, and the quasi-geodesics in the complex of curves, which is described in Chapter 10 by Ursula Hamenstädt.

There have been other recent breakthroughs on the Weil–Petersson metric, by S. Wolpert, R. Wentworth, S. Yamada, M. Mirzakhani, B-Y. Chen, Z. Huang, joint work by S. Y. Cheng and S. T. Yau and by K. Liu, X. Sun and S. T. Yau, and by others.

We also mention that the Weil–Petersson metric can be recovered through Bonahon’s embedding of Teichmüller space into the space of geodesic currents.

Recent work on the Weil–Petersson metric will be reported on in Volume II of this Handbook.

Thurston’s asymmetric metric. This metric was introduced by Thurston in 1986, in his paper *Minimal Stretch maps between hyperbolic surfaces* (which is still in an unpublished form). Here, one considers Teichmüller space as the set of isotopy classes of complete finite-area hyperbolic structures on $S_{g,p}$. If g and h are such structures, one defines their “distance” to be the logarithm of the infimum of Lipschitz constants of all diffeomorphisms $\varphi: (S_{g,p}, g) \rightarrow (S_{g,p}, h)$ which are isotopic to the identity, the Lipschitz constant of φ being, by definition, $\sup \frac{d_h(\varphi(x), \varphi(y))}{d_g(x, y)}$, the supremum being taken over all distinct pairs of points x and y in $S_{g,p}$. This distance function on the set of hyperbolic structures is invariant by the action on each factor of the group $\text{Diff}_0^+(S_{g,p})$ of diffeomorphisms of $S_{g,p}$ that are isotopic to the identity, and it induces a function on the Teichmüller space $\mathcal{T}_{g,p}$ of $S_{g,p}$, which we call Thurston’s asymmetric metric. This is not a metric in the usual sense of the word since, as its name indicates, it does not satisfy the symmetry axiom. Despite this asymmetry (and in some cases because of this asymmetry), Thurston’s metric has beautiful geometric properties. Several of these properties are expounded in Chapter 2 of this volume.

Let us note here that the study of asymmetric metrics has a long history, and we can mention in this respect the name of Herbert Busemann, who did a systematic study of such metrics, starting in the 1940s. The methods of synthetic geometry in the sense of Busemann can be used to investigate Thurston’s asymmetric metric, for instance in the study of the behaviour of its geodesics, its isometries, its visual boundaries and so on. The techniques that are used by Thurston in the study of his asymmetric metric are those of basic hyperbolic geometry, and it is most satisfying to see that with these basic elementary techniques, a beautiful picture of a whole theory arises. There are several good questions which one can ask about this metric, and some of them are listed in Chapter 2.

The Carathéodory, the Kobayashi and the Bergman metrics. The Carathéodory and the Kobayashi “metrics” are semi-metrics³ that are classically defined on any complex space, and that are invariant under biholomorphic self-mappings of this space. In the case of Teichmüller space, these semi-metrics are genuine metrics. Let us first recall the definitions, since they are easy to state and are most appealing from the point of view of synthetic geometry.

Let X be a complex space and let D be the unit disk in \mathbb{C} equipped with its Poincaré metric. The Carathéodory semi-metric on X is defined by the formula $d_{\mathcal{C}}(x, y) = \sup_f d_D(f(x), f(y))$ for all x and y in X , where the supremum is taken over all holomorphic maps $f: X \rightarrow D$.

The Kobayashi semi-metric on X is defined by a dual construction. One first defines a map $d'_{\mathcal{K}}: X \times X \rightarrow \mathbb{R}$ by $d'_{\mathcal{K}}(x, y) = \inf_{f,a,b} d_D(a, b)$ for all x and y in X , the infimum being taken over all holomorphic maps $f: D \rightarrow X$ and over all a and b in D satisfying $f(a) = x$ and $f(b) = y$. The map $d'_{\mathcal{K}}$ does not necessarily satisfy the triangle inequality, and the Kobayashi semi-metric $d_{\mathcal{K}}$ is defined as the largest semi-metric on X satisfying $d_{\mathcal{K}} \leq d'_{\mathcal{K}}$.

We already mentioned that H. L. Royden proved in 1970 that the Kobayashi metric of Teichmüller space coincides with the Teichmüller metric. In 1974, C. J. Earle proved that the Carathéodory metric on Teichmüller space is complete, and he raised the question of whether this metric coincides with the Teichmüller metric. He also asked the weaker question of whether this metric on Teichmüller space is proper (that is, whether every closed bounded set is compact). Both questions were answered by S. L. Krushkal'. In 1976, Krushkal' proved that the Carathéodory metric on Teichmüller space is proper, and in 1981 he proved that the Carathéodory metric does not coincide with the Teichmüller metric, unless the complex dimension of this space is one.

We also note that the Carathéodory and the Kobayashi metrics on a complex space have infinitesimal descriptions, and that in general they are Finsler and not Riemannian metrics.

The Bergman metric is a Kähler semi-metric that is associated to any bounded domain of holomorphy. Its definition uses the Bergman Kernel. The Bergman metric of Teichmüller space (seen as a bounded domain through Bers' embedding) is a genuine metric.

By a result of K. T. Hahn (1976), the Carathéodory metric on any bounded domain in \mathbb{C}^N is bounded from above by the Bergman metric. Therefore, by using Earle's result on the completeness of the Carathéodory metric, the Bergman metric is complete. B.-Y. Chen showed (2004) that the Bergman metric of Teichmüller space is equivalent to the Teichmüller metric in the sense that there exists a positive constant C such that $\frac{1}{C}d_T \leq d_B \leq Cd_T$, where d_T and d_B denote respectively the Teichmüller and the Bergman metrics. This result was also obtained, at about the same time, by K. Liu, X. Sun and S.-T. Yau.

³Recall that a semi-metric $d: X \times X \rightarrow [0, \infty]$ satisfies the axioms of a metric except possibly for the separation axiom $d(x, y) = 0 \Rightarrow x = y$.

A survey of the Carathéodory, the Kobayashi and the Bergman metrics on Teichmüller space is contained in a chapter by A. Fletcher and V. Markovitch in Volume II of this Handbook.

Other metrics. There are many other interesting metrics on Teichmüller space. We mention a few of them, although they are not considered in this volume. S. Y. Cheng, and S. T. Yau showed that there exists a complete Kähler–Einstein metric of negative Ricci curvature on any bounded pseudo-convex domain in \mathbb{C}^N (1975). Using Bers’ embedding, this result shows that there is a unique complete Kähler–Einstein metric on Teichmüller space. This metric is invariant under the action of the mapping class group. It has constant negative scalar curvature, it descends to a complete Kähler–Einstein metric on moduli space and it has been thoroughly studied by S. Y. Cheng, N. Mok and S. T. Yau, and by algebraic geometers. In 2004, K. Liu, X. Sun and S.-T. Yau showed that this Kähler–Einstein metric is equivalent to the Teichmüller metric (a result which had been conjectured by Yau in 1986), which implies that the Bergman and the Kähler–Einstein metrics are equivalent.

We also mention a metric defined in the late 1990s by C. McMullen’s (a metric which is now called McMullen’s metric), which is a complete Kähler metric of bounded sectional curvature. It is obtained as a kind of interpolation between the Teichmüller and the Weil–Peterson metric. McMullen showed that this metric is equivalent to Teichmüller’s metric and that it is Kähler-hyperbolic in the sense of Gromov.

Other complete Kähler metrics on Teichmüller space have been considered quite recently by K. Liu, X. Sun and S.-T. Yau. These authors asked a number of interesting questions about these metrics: computation of the L^2 -cohomology groups, Ricci flow between these metrics, index theory, representation of the mapping class group in the middle-dimensional L^2 -cohomology, and so on.

Let us note that a recent paper by S.-K. Yeung (Quasi-isometry of metrics on Teichmüller space, 2005) reports on the state of the art concerning quasi-isometries and equivalences between several of the metrics on Teichmüller space that we mentioned above. The same paper adds new results on that subject.

1.3 Symplectic and Poisson structures

Teichmüller space has a natural symplectic structure ω which is induced by its Weil–Peterson Kähler metric. This structure is invariant by the action of the mapping class group and it descends to a symplectic structure on moduli space. The most geometrically appealing description of this structure is certainly the one given by Scott Wolpert in the early 1980s, in terms of the Fenchel–Nielsen coordinates. Specifically, if $\alpha_1, \dots, \alpha_n$ is a set of homotopy classes of simple closed curves defining a pants decomposition of the surface, and if τ_1, \dots, τ_n and l_1, \dots, l_n are respectively the twist and length parameters associated to these curves, then it is known that the parameters (τ_i, l_i) , $i = 1, \dots, n$ are global coordinates for Teichmüller space; they are called the Fenchel–Nielsen (twist-length) coordinates. Wolpert proved that the Weil–Peterson

symplectic form is given by $w = \sum_{i=1}^n d\tau_i \wedge dl_i$ (which implies in particular that this twist-length description of the symplectic structure is independent of the choice of the pants decomposition). We note that in the definition of the twist parameter, one should be careful in counting right twists positively, and in twisting at time t a total length of t .

Wolpert obtained several other beautiful formulae concerning this symplectic structure, and we state a few of them. Let $t(\alpha)$ denote the Fenchel–Nielsen vector field on Teichmüller space associated to a homotopy class of simple closed curves α on the surface and $l(\alpha)$ denote the hyperbolic length function on \mathcal{T} associated to α . Then, Wolpert proved the following:

– A “duality formula”: $\omega(t(\alpha), \cdot) = -dt(\alpha)$.

– A “cosine formula”: $w(t(\alpha), t(\beta)) = t(\alpha)l(\beta) = \sum_p \cos \theta_p$. Here, $t(\alpha)l(\beta)$ is the natural action of a vector field on a function, p varies over the set of intersection points of two fixed representatives of α and β which are in minimal position and θ_p is the angle of the intersection at the point p of these representatives, with appropriate orientations.

– A formula for the Poisson bracket. First, Wolpert defines a normalized vector field $T_\alpha = 4(\sinh l(\alpha)/2)t(\alpha)$. The T_α ’s form a Lie algebra over the integers, and the formula for the Lie bracket is $[T_\alpha, T_\beta] = \sum_p T_{(\alpha_p\beta)^+} - T_{(\alpha_p\beta)^-}$, where p varies as above over the set of intersection points of two fixed closed curves in minimal position representing α and β and where $(\alpha_p\beta)^+$ and $(\alpha_p\beta)^-$ are homotopy classes of closed curves obtained by modifying by local surgeries the union of the curves representing α and β at the intersection point p .

We finally note that there are nice formulae for the pull-back of the symplectic form ω on decorated Teichmüller space of a punctured surface, obtained by Penner, which also have a geometric flavour. As we shall see below, Penner’s formulae have been used in the quantization theory of Teichmüller spaces of punctured surfaces.

1.4 Boundary structures

Although the topology of Teichmüller space is very simple (the space is homeomorphic to a ball), the boundary structure of this space is very rich and highly nontrivial. In fact, there are several interesting boundaries. Every time Teichmüller space is embedded in some function space (for instance, by Bers’ embedding as a bounded domain in a Banach space, or by Thurston’s embedding in the space of geodesic length functions, or by Bonahon’s embedding in the space of geodesic currents, or by its embeddings in various spaces of representations), one can define a boundary structure for this space, by taking the closure of the image of the embedding. In all the cases mentioned, the boundary points have a geometric significance, as degenerate Riemann surfaces (for instance surfaces with nodes), or as projective classes of measured foliations, or as degenerate representations defined by group actions on \mathbb{R} -trees, and so on. In some cases, the closure of the image is compact, and the boundary structure defines a compactification of Teichmüller space. In some cases, the boundary is homeomorphic

to a sphere, and in other cases it is not. It can happen that the boundary structure of Teichmüller space defines a boundary structure to Riemann's moduli space and vice versa. We shall review here some of the boundary structures of Teichmüller space and moduli space.

Thurston's boundary. Thurston's boundary of Teichmüller space is induced from the embedding of this space in the space $P\mathbb{R}_+^{\mathcal{S}}$ of projective classes of nonnegative functions on the set \mathcal{S} of homotopy classes of essential simple closed curves on the surface, by means of the geodesic length functions. The boundary points are projective classes of compactly supported measured geodesic laminations, embedded in $P\mathbb{R}_+^{\mathcal{S}}$ by means of the geometric intersection functions. The mapping class group action on Teichmüller space extends continuously to the union of Teichmüller space with its Thurston boundary.

We note that Thurston's boundary can also be recovered using Bonahon's embedding of Teichmüller space into the space of geodesic currents.

Bers' boundary. Bers' boundary of Teichmüller space is induced from the Bers embedding of this space in the Banach space of quadratic differentials on a given Riemann surface. Bers showed that the closure of this embedding is compact, and this defines Bers' compactification of Teichmüller space. Bers' embedding depends upon the choice of a basepoint in Teichmüller space, but any two such embeddings are biholomorphically equivalent. However, Kerckhoff and Thurston showed that the boundary structure that one obtains in this way depends upon the choice of the basepoint. They proved that the mapping class group action on Teichmüller space does not extend continuously to an action on the union of this space with its Bers boundary. In fact, Kerckhoff and Thurston's work (1990) gives a new description of Bers' boundary in the setting of geometric convergence in the space of representations of the fundamental group of the surface in $\mathrm{PSL}(2, \mathbb{C})$. Their results on Bers' boundary are obtained after translating the questions into the context of quasi-Fuchsian groups. By Kerckhoff and Thurston's results, Bers' boundary and Thurston's boundary are distinct.

The visual boundaries. Different visual boundaries of Teichmüller space can be associated to the various metrics on this space. We recall that the visual boundary of a complete metric space (X, d) is the space of equivalence classes of geodesic rays, where two geodesic rays $\gamma_1: [0, \infty[\rightarrow X$ and $\gamma_2: [0, \infty[\rightarrow X$ are considered equivalent if the set $\{d(\gamma_1(t), \gamma_2(t)), t \in [0, \infty[\}$ is bounded. As a matter of fact, one usually considers the visual boundary at a given point x in X , by taking the set of geodesic rays emanating from x . There is a natural topology (the "cone topology") on the visual boundary at x . In the case where the space X is uniquely geodesic, there is a topology on the union of X with its visual boundary at x (which makes this visual "boundary" indeed a boundary). This topology is obtained by embedding X into the space of all (that is, finite or infinite) geodesics emanating from x , each point in X being identified with the geodesic that joins it to x . Of course, the first natural question

arising from such a definition is: To what extent does the visual boundary depend on the choice of x ?

The visual boundary of Teichmüller's space equipped with Teichmüller's metric was studied in the late 1970s by Kerckhoff, who showed that the action of the mapping class group on Teichmüller space does not extend continuously to the Teichmüller visual boundary.

The Weil–Petersson visual boundary has recently been studied by J. Brock, who worked on the analogies of this visual boundary with the boundary defined by taking the completion of the Weil–Petersson metric, and with the Bers boundary of Teichmüller space. Masur had already worked on the comparison between these last two boundaries in 1982. Brock's work starts with the known observation that some Weil–Petersson geodesic rays have finite length, a fact related to the non-completeness of the Weil–Petersson metric. In this work, the definition of the visual boundary is adapted to the setting of a non-complete space. The finite-length geodesics can be made to converge to surfaces with cusps, and this convergence is comparable to the convergence to surfaces with nodes that appears in the works of Bers and Abikoff. Brock also showed that the action of the mapping class group on Teichmüller space does not extend continuously to the Weil–Petersson visual boundary. Again, this is analogous to the result by Kerckhoff and Thurston that we mentioned above, showing that the action of the mapping class group does not extend continuously to Bers' boundary. Brock showed that despite these analogies, the Weil–Petersson boundary is distinct from Bers' boundary.

Finally, we note that the visual boundary of Thurston's asymmetric metric is another interesting object which so far has not been given the attention it deserves.

Other boundary structures. There are boundary structures for Riemann's moduli space, some of them induced from boundary structures for Teichmüller space, and some of them belonging only to the realm of moduli space. The compactification of moduli space by stable curves is a projective variety (studied as such by Wolpert and others). It is a quotient space of Teichmüller space equipped with its bordification by surfaces with nodes (sometimes called augmented Teichmüller space, with boundary called Abikoff's augmented boundary) by the action of the mapping class group. Note that this boundary structure of Teichmüller space is not a compactification of this space, but the quotient of this structure by the mapping class group is a compactification of moduli space.

Sometimes, the same boundary structure for Teichmüller space is studied from various points of view. When Riemann surfaces are viewed as algebraic surfaces, the boundary structure can be studied in terms of degeneration of algebraic surfaces, and so on. For instance, the augmented moduli space (and finite coverings of this space), whose boundary points are surfaces with nodes, has been studied by Deligne, Mumford, Knudsen and Mayer from the point of view of algebraic geometry, by Abikoff from the point of view of (geometric) complex analysis, and by Marden in relation with hyperbolic 3-manifold theory. Due to the fact that people working in

different domains use different languages, it is sometime difficult to see the relations between the various results.

Let us finally mention that there are boundary structures for Schottky space, for general representation spaces, and for several other spaces related to Teichmüller space.

1.5 The harmonic maps approach

The study of harmonic maps is a classical subject in analysis and geometry, and these maps have been used in the last three decades as an important tool in the study of Teichmüller space. The definition of a harmonic map is based on the notion of the energy of a map. The energy of a map between two (say compact smooth) manifolds is the integral over the domain of the square of the derivative of the map. This quantity is regarded as a measure of the average squared stretching of the map. A map between two Riemannian manifolds is harmonic if it is a critical point of the energy functional. The collection of maps between the two manifolds over which the critical point is searched for is usually taken to be the set of maps in a given homotopy class of maps. Harmonic maps are solutions of an Euler–Lagrange equation, defined by a second-order elliptic partial differential operator. In some sense, harmonic maps generalize closed geodesics (which are harmonic maps with the domain manifold being the circle). They also generalize totally geodesic maps and harmonic functions (which are harmonic maps whose range is Euclidean space). The study of Teichmüller space using harmonic maps started with the work of M. Gerstenhaber and H. E. Rauch in 1954. In particular, there are harmonic map approaches to the Teichmüller metric and to the Weil–Petersson metric. The list of people that have worked in this domain is very long; just to give a few names, we mention E. Reich, C. Earle, J. Eells, J. Jost, A. Tromba, M. Wolf, Y. Minsky, C. Mese, G. Daskalopoulos, R. Wentworth and L. Katzarkov.

The classes of harmonic maps that have been used in Teichmüller theory include harmonic maps between surfaces, harmonic maps from surfaces to \mathbb{R} -trees (as limits of harmonic maps between surfaces; these maps can be used to describe degenerations of elements of Teichmüller space), and, more recently, harmonic maps from surfaces to Teichmüller space itself (equipped with various metrics).

A basic feature of the study of harmonic maps between surfaces is that the energy of such a map only depends on the conformal class of the domain surface, and not on its metric. In the harmonic maps approach to Teichmüller theory, the domain surface is a Riemann surface and the target surface is generally equipped with a hyperbolic metric. Sometimes, the target surface is equipped with a singular flat metric. For instance, E. Kuwert (1996) proved that a Teichmüller map is the unique harmonic map in its isotopy class when the target surface is equipped with the singular flat metric defined by the terminal quadratic differential of the map.

Thus, one can distinguish the following three approaches to Teichmüller theory:

- the approach where the elements of Teichmüller space are homotopy classes of conformal structures on a surface, and the optimal maps between two such structures are those that have minimum quasiconformal constant;

– the approach where the elements of Teichmüller space are homotopy classes of hyperbolic structures, the optimal maps between two such structures being those that have minimum Lipschitz constant;

– the harmonic maps approach where the domain surface is equipped with a conformal structure and the target surface is equipped with a hyperbolic structure, and where the harmonic maps are the optimal maps in the sense that they minimize energy.

Of course, there are very good questions concerning relations between the three approaches. The harmonic maps approach has the advantage of using, at the same time, the conformal and the hyperbolic point of view on Teichmüller space, and making links between them.

M. Wolf and A. E. Fischer together with A. Tromba developed theories that describe Teichmüller space in terms of harmonic maps. The approach of Fischer and Tromba uses techniques of analysis, whereas Wolf's approach is more geometric. Wolf gave a harmonic map description of the Weil–Petersson metric. He also studied harmonic maps between a fixed Riemann surface and families of degenerating hyperbolic surfaces. His approach includes a harmonic map description of Thurston's boundary. This approach involves maps from surfaces to \mathbb{R} -trees, and it uses the work of Gromov and Schoen on harmonic maps from surfaces to singular spaces. We can also mention here the work of Y. Minsky, who studied the behaviour of families of harmonic maps between surfaces when the domain surface varies along a Teichmüller geodesic, and other related problems.

We finally mention joint work by Daskalopoulos, Katzarkov and Wentworth on harmonic maps from surfaces to Teichmüller space equipped with the Weil–Petersson metric and with McMullen's metric, and joint work by Daskalopoulos, Dostoglou and Wentworth in which they give a harmonic map interpretation of the Morgan–Shalen compactification of the character variety of representations of the fundamental group of a surface into $SL(2, \mathbb{C})$.

By looking at the work done by the various people working in the area, one gets the impression that all the important features of Teichmüller theory can be recovered using the theory of harmonic maps.

1.6 The group theory

There are several classes of groups that are intimately connected to Teichmüller theory. Of course, one thinks primarily of mapping class groups of surfaces and their subgroups, but one can consider more generally the mapping class groups of general C^∞ manifolds, and their subgroups. Another important class of groups related to Teichmüller theory is the class of Kleinian groups, with the related Fuchsian, quasi-Fuchsian, Schottky groups, and other groups defined by representations of the fundamental group of the surface. Of course, all these groups are also studied for their own sake, that is, independently of their relation to Teichmüller space.

Mapping class groups. The groups that play the most prominent role in Teichmüller theory are certainly the mapping class groups.⁴ Important special cases of mapping class groups include the mapping class groups of punctured disks, which can be identified with braid groups.

The study of the mapping class group has a long history that starts with that of low-dimensional topology. There is a large number of interesting open problems related to the algebraic structure of mapping class groups, to their action on various spaces, and to the structure of their subgroups. An impressive list of such problems appears in a book edited by B. Farb (American Mathematical Society, 2006). An important subgroup of the mapping class group is the Torelli subgroup, consisting of the mapping classes that induce the identity on the homology of the surface, but there are several others. In Chapter 8 of this volume, L. Mosher gives an overview of the geometry and dynamics of actions of several classes of subgroups of the mapping class group on Teichmüller space equipped with its Thurston boundary.

As already mentioned, the mapping class group of a surface is also the isometry group of the Teichmüller metric and of the Weil–Petersson metric on the Teichmüller space of that surface, and it is conceivable that similar results hold for other metrics on that space. The most enlightening study of the mapping class group is certainly the one that Thurston made through the analysis of the action of that group on the compactification of Teichmüller space by the space of projective classes of measured foliations on the surface. Thurston’s topological classification of the elements of the mapping class group, which is surveyed in Chapter 8 by Mosher, has counterparts in terms of the metric structures of Teichmüller space. In this connection, we recall that after Thurston’s work was completed, Bers worked out a similar classification of mapping classes which is based on the action of the mapping class group equipped with the Teichmüller metric. Bers obtained this classification by analyzing the minimal set and the displacement function associated to a mapping class acting by isometries. A similar metric classification was recently obtained by Daskalopoulos and Wentworth, this time with respect to the Weil–Petersson metric. This classification is surveyed in Chapter 1 of this volume.

There are several other interesting actions of the mapping class group that were studied in the last few decades. For instance, Hatcher and Thurston proved in 1980 that the mapping class group of a closed surface is finitely presented, by studying the action of that group on a two-dimensional simplicial complex whose vertices are cut systems on the surface, that is, systems of isotopy classes of disjoint closed curves whose complement is a sphere with holes. The edges of this complex correspond to certain moves between cut systems, called elementary moves, and the two-dimensional cells correspond to certain cycles of moves, which are of three types: triangles, rectangles and pentagons. E. Irmak and M. Korkmaz proved recently that the automorphism

⁴There are some variations in the terminology of mapping class groups. Usually, the term *extended mapping class group* of a surface designates the group of isotopy classes of diffeomorphisms of that surface. The mapping class group is then the group of isotopy classes of orientation-preserving diffeomorphisms of an oriented surface. There are also various subclasses, depending on whether the surface is closed or not, with or without boundary components, whether the group fixes a distinguished point or not, and so on.

group of the Hatcher–Thurston complex of a compact oriented surface of positive genus is isomorphic to the extended mapping class group of the surface modulo its center. Other actions of mapping class groups which have been thoroughly studied include the action on the complex of curves, a flag complex whose n -simplices are collections of $(n + 1)$ distinct isotopy classes of essential disjoint simple closed curves on the surface. This complex was introduced in 1979 by W. Harvey, with the idea of including it as a boundary structure of Teichmüller space. We also mention the actions of the mapping class group on the complex of pants, on the complex of nonseparating curves, on the complex of separating curves and the complex of domains, and there are several others. We already mentioned that Brock found a precise relation between the Weil–Petersson metric on Teichmüller space and the metric on the vertices of the complex of pants, induced by the simplicial metric on the 1-skeleton. The coarse geometries of these complexes equipped with their natural simplicial metric have also been studied for their own sake, that is, independently of their relation to the mapping class group. In Chapter 10 of this volume, Ursula Hamenstädt studies the action of the mapping class group on the complex of curves. She gives a new proof of a result of Masur and Minsky stating that this complex is hyperbolic in the sense of Gromov, and she discusses the relation between the geometry of this complex and the geometry of Teichmüller space.

Outer automorphism groups. An inner automorphism of a group G is an automorphism of the form $g \mapsto h^{-1}gh$, where h is a fixed element of G . The outer automorphism group of G is the quotient group of the automorphism group of G by the action of the inner automorphism group of G .

Automorphism and outer automorphism groups of free groups are basic objects of study in combinatorial group theory. They were already extensively investigated by J. Nielsen and W. Magnus, in the first quarter of the twentieth century. There is a close relation between mapping class groups of surfaces and outer automorphism groups. For instance, in the particular case of a closed surface, the extended mapping class group of the surface is isomorphic to the outer automorphism group of its fundamental group. In the case of a surface with one puncture, the fundamental group is a free group, and there is a natural map from the extended mapping class group of the surface to the outer automorphism group of its fundamental group. This map is explicated in Chapter 7, written by S. Morita. The study of the outer automorphism of a free group acquired a very geometric character after the introduction in the 1980s, by M. Culler and K. Vogtmann, of a space called *Outer space*, on which this group acts. There are parallels between the action of the outer automorphism of a free group on Outer space, and the action of the mapping class group on Teichmüller space.

Kleinian groups A Kleinian group G is a discrete group of orientation-preserving Möbius transformation of hyperbolic 3-space \mathbb{H}^3 .⁵ A Fuchsian group is a Kleinian

⁵The terminology here is not universally established. For instance, some authors ask that a Kleinian group acts properly discontinuously on some non-empty open subset on the Riemann sphere seen as the boundary at infinity of \mathbb{H}^3 , and some other authors require that the group is finitely generated.

group whose action on $S^2 = \partial\mathbb{H}^3$ preserves a round disk. A quasi-Fuchsian group is obtained from a Fuchsian group by conjugation of the action by a quasiconformal homeomorphism of \mathbb{H}^2 . The quotient space of the action of a Kleinian group on \mathbb{H}^3 is a hyperbolic 3-orbifold (and it is a manifold if Γ is torsion-free). Under certain conditions, the picture of the action can be made more complete: the Kleinian group on $\mathbb{H}^3 \cup S^2$ gives rise to a hyperbolic 3-manifold with nonempty boundary consisting of a finite number of surfaces of finite type carrying a canonical conformal structure, and the deformation theory of the hyperbolic 3-manifold can be studied through the deformation theory of its boundary surfaces. More precisely, given a Kleinian group G , it has an induced action on the boundary at infinity $\partial\mathbb{H}^3 = S^2$. This action decomposes S^2 into the union of two subsets, the ordinary set and the limit set. The *ordinary set* (or domain of discontinuity) $\Omega \subset S^2$ is the largest open subset of the sphere on which the group G acts properly discontinuously. The *limit set* Λ of the action is the set of accumulation points in S^2 of the orbit of any point in \mathbb{H}^3 . From the point of view of the theory of hyperbolic 3-manifolds with boundary, one considers the action of G on the union $\mathbb{H}^3 \cup \Omega$. Ahlfors' finiteness theorem ensures that under some mild hypotheses, the quotient of $\mathbb{H}^3 \cup \Omega$ by this action is a 3-manifold with boundary, with the boundary consisting of a finite union of connected Riemann surfaces of finite type. The 3-manifold is equipped with a quotient hyperbolic structure induced from that of \mathbb{H}^3 , and each boundary component is equipped with a complex structure (in fact, with a complex projective structure) induced from that of S^2 . There is an interplay between the deformation theory of the hyperbolic 3-manifold and the union of the Teichmüller spaces of the boundary components. The classical approach to this study consists in applying Teichmüller's quasiconformal deformation theory in succession to the boundary components. For instance, in the case where the Kleinian group is a quasi-Fuchsian group, a theorem of Bers says that the hyperbolic structure of the 3-manifold is completely determined by the conformal structure of the boundary, which as a matter of fact consists of two conformal structures on the same topological surface. Thus, quasi-Fuchsian space is parametrized by the product of two copies of Teichmüller space.

We also note that the relation of Teichmüller theory with Kleinian groups involves the deformation theory of the projective structure of the boundary, and that the general deformation theory of Kleinian groups can be studied from the point of view of representation theory.

1.7 Representation theory

By the uniformization theorem, every Riemann surface of negative Euler characteristic can be defined as a complex structure induced by some hyperbolic metric. In the case of a surface without boundary, the hyperbolic metric can be regarded, by lifting the hyperbolic structure to the universal cover, as a discrete group of orientation-preserving isometries of hyperbolic 2-space \mathbb{H}^2 , and such a group is well-defined up to conjugacy. From this observation, Teichmüller space can be studied as a set of conjugacy classes

of representations of the fundamental group of the surface into the group $\mathrm{PSL}(2, \mathbb{R})$ of orientation-preserving isometries of \mathbb{H}^2 .

More generally, one can study the moduli spaces of representations of the fundamental group of a surface S in a Lie group G , that is, the space of homomorphisms of $\pi_1(S)$ into G up to conjugation. The classical case is $G = \mathrm{PSL}(2, \mathbb{R})$, as described above. The Teichmüller space of a closed surface S of genus ≥ 2 is a connected component of this moduli space. This component consists of the conjugacy classes of discrete and faithful representations. W. Goldman showed that in this case, the number of connected components of the moduli space of representations is equal to $4g - 3$.

There are other important cases for the Lie group G . These include the case where $G = \mathrm{PSL}(2, \mathbb{C})$ (which is the setting of Kleinian representations, which includes several subsettings, like quasi-Fuchsian representations, Schottky representations, and so on). There is also the case where G is a split semisimple real group, whose study is the subject of “higher Teichmüller theory”. It is worth saying a few words about this theory, since it has recently attracted much interest.

It is generally considered that higher Teichmüller theory was initiated by the work of Nigel Hitchin, who proved in the early 1990s that the space of conjugacy classes of discrete faithful representations in the space of conjugacy classes of all representations of the fundamental group of a compact oriented surface of genus ≥ 2 in a semisimple split real Lie group (namely $\mathrm{SL}(n, \mathbb{R})$ or $\mathrm{Sp}(2n, \mathbb{R})$) is a contractible component, a result which is regarded as a generalization of the fact that Teichmüller space is contractible. As mentioned above, in the case $n = 2$, this component is the Teichmüller space of the surface. In the case $n = 3$, the corresponding component is, by a result of Goldman and Choi, the moduli space of convex real projective structures on the surface. There are several good questions in higher Teichmüller theory, some of them independent of classical Teichmüller theory; some concern the action of the mapping class group on components other than the one discovered by Hitchin. V. Fock and A. Gontcharov recently studied “positive” representations of the fundamental group of a closed surface into a split semisimple algebraic group G with trivial center. They proved that these representations are faithful and discrete and that the moduli space of such representations is an open cell in the space of all representations. For $G = \mathrm{SL}(2, \mathbb{R})$, this space is the classical Teichmüller space. Higher Teichmüller theory has also been developed by F. Labourie. M. Burger, A. Iozzi and A. Weinhard recently undertook a study of representations of the fundamental group of the surface into semisimple Lie groups of Hermitian type. Let us note that one can also study representations of fundamental groups of surfaces in compact Lie groups like $\mathrm{SU}(2)$; in this case, the space of representations is compact.

1.8 Dessins d’enfants

A natural question in the algebro-geometric point of view on Teichmüller space is to understand the Riemann surfaces that can be defined by polynomials with coefficients

in number fields. For instance, one would like to understand how the collection of such surfaces is embedded in Riemann's moduli space, or how to recognize the algebraic and, more specially, the arithmetic surfaces. There are also interesting questions about the automorphisms of such surfaces, e.g. what are the stabilizers of these surfaces in the mapping class group, and so on.

Grothendieck's introduction of his theory of dessins d'enfants was a major step in this direction. This theory makes connections between Teichmüller theory, complex algebraic geometry, arithmetic geometry and the study of a class of combinatorial objects on the surface.

We now briefly discuss this theory, and the related space of Belyi affine curves. This space is an analytic variety equipped with a natural action of the *absolute Galois group* (also called the *universal Galois group* $\text{Gal}(\overline{\mathbb{Q}})$) which, by definition, is the automorphism group of the field $\overline{\mathbb{Q}}$ of algebraic numbers. We note that the absolute Galois group $\text{Gal}(\overline{\mathbb{Q}})$ is considered as one of the most mysterious groups in mathematics, and for this reason, producing new actions of this group is always interesting.

Grothendieck's approach to Teichmüller theory is based on the idea that the combinatorial techniques of hyperbolic geometry (starting with surface decompositions into hyperbolic pairs of pants) have a parallel in algebraic geometry over $\overline{\mathbb{Q}}$. In this approach, dessins d'enfants play a prominent role. A *dessin d'enfant* ("child's drawing") is a connected graph embedded in a compact topological surface, whose complementary components are simply connected and whose edges are colored black and white such that no edge has its two vertices of the same color.

This theory is outlined in Grothendieck's *Esquisse d'un programme*, a manuscript which has been widely circulated in the form of mimeographed notes since it was written in 1984 and which finally appeared in print (together with an English translation) in 1997.

The correspondence between dessins d'enfants and algebraic curves over $\overline{\mathbb{Q}}$ is a consequence of a theorem of G. Belyi. This theorem states that any algebraic curve over $\overline{\mathbb{Q}}$ can be realized as a ramified finite covering of the Riemann sphere $\mathbb{P}^1(\mathbb{C})$, with ramification locus contained in the set $\{0, 1, \infty\}$. In this way, an algebraic curve over $\overline{\mathbb{Q}}$ defines a dessin d'enfant, obtained as the preimage of $[0, 1] \subset \mathbb{P}^1(\mathbb{C})$ by the covering map. The color of a vertex is specified according to whether the image of that vertex is 0 or 1. The relation with algebraic curves over \mathbb{C} stems then from the fact that a complex algebraic curve X can be represented by an algebraic curve over $\overline{\mathbb{Q}}$ if and only if there exists a non-constant holomorphic map $f: X \rightarrow \mathbb{P}^1(\mathbb{C})$ whose critical values are in $\overline{\mathbb{Q}}$. The surprising contribution of Belyi was to reduce the number of critical values to three. This result was first announced at the Helsinki ICM in 1978. It had been conjectured by Grothendieck, who later on declared that he was amazed by the simplicity of Belyi's proof of this result. The converse result, that is, the fact that to a dessin d'enfant one can associate a ramified covering over $\mathbb{P}^1(\mathbb{C})$ is due to Grothendieck himself (who proved it using Belyi's work).

Grothendieck's ideas on Teichmüller space have been essentially exploited by algebraic geometers, because the original language of Grothendieck is that of algebraic

geometry, although Grothendieck claimed on several occasions (in particular in his *Esquisse d'un programme*) that he was amazed by Thurston's approach to Teichmüller space, and that he would have liked to see his theory expressed in a similar geometric language.

Volume II of this Handbook contains a chapter on the theory of dessins d'enfants, written by F. Herrlich and G. Schmithüsen.

Links between Grothendieck's theory and the geometric Teichmüller theory will certainly become more apparent in the near future.

1.9 Physics and quantum theory

E. Verlinde and H. Verlinde conjectured in 1990 that the space of conformal blocks of Liouville conformal field theory (which is a version of a two-dimensional Einstein quantum gravity) can be obtained by the quantization of the Teichmüller space of the corresponding Riemann surface. This was motivated by the developments of the quantization of Chern–Simons theory. The quantization theory of Teichmüller space makes the Verlindes idea precise and it provides an explicit Hilbert space and observable algebra for 2-dimensional quantum gravity.

More explicitly, the quantization of Teichmüller space $\mathcal{T}_{g,p}$ produces a one-parameter family of noncommutative $*$ -algebras, with an action of the mapping class group of the surface by $*$ -algebra automorphisms. These algebras are parametrized by a quantization parameter \hbar , having the property that when $\hbar = 0$ the algebra coincides with the original commutative algebra of functions on $\mathcal{T}_{g,p}$ while the first derivative of the commutator with respect to \hbar at zero gives the Weil–Petersson Poisson structure on $\mathcal{T}_{g,p}$.

Quantization theories of Teichmüller space were developed independently during the 1990s in joint work by L. Chekhov and V. Fock, and by R. Kashaev. The two theories are essentially equivalent up to technical details.

The quantization theory developed by Chekhov and Fock considers the Teichmüller space of a surface with holes as a (degenerate) Poisson manifold and uses Thurston's shear coordinates on that surface, whereas the quantization theory developed by Kashaev considers the decorated Teichmüller space and uses Penner's λ -length coordinates.

The quantization procedures produce (for $\hbar \neq 0$) semi-simple infinite-dimensional $*$ -algebras, and unitary projective representations of a subgroup of the mapping class group in the $*$ -representation spaces of these algebras. Technically, this representation acts on a Hilbert space of functions on \mathbb{R}^n by certain compositions of Fourier transformations, multiplications by quadratic forms and multiplication by quantum dilogarithms.

Let us also mention that Kashaev recently developed a quantization theory of the moduli space of flat $\mathrm{PSL}(2, \mathbb{R})$ -connections on a punctured surface with parabolic holonomy around the punctures. This moduli space is closely related to the classical

Teichmüller space of the surface, and in fact, two of its connected components are isomorphic to Teichmüller space.

J. Teschner's work proves the Verlinde's conjecture relating the quantization theories developed by Chekhov–Fock and by Kashaev to the Liouville conformal field theory.

Furthermore, the quantization of Teichmüller space may lead to new invariants of hyperbolic three-manifolds. A famous example in this respect is the volume conjecture by R. Kashaev, H. Murakami and J. Murakami which expresses the volume of any hyperbolic knot as a certain limit of a specific value of the colored Jones invariant associated with the quantum group $SU(2)_q$, which in turn can be computed using Kashaev's dilogarithmic quantization theory.

Thus, the quantization theories establish new links between Teichmüller theory, algebra, algebraic geometry, representation theory and mathematical physics.

The present volume contains chapters written by the various people involved in the quantization theory, namely, Chekhov, Fock, Goncharov, Kashaev, Penner and Teschner.

We mention that S. Baseilhac and R. Benedetti worked out a finite-dimensional combinatorial version of the quantization of Teichmüller space, with the aim of constructing new invariants for 3-manifolds, and with a view towards relating these results to the volume conjecture (with the hope of proving this conjecture). Their techniques are partly inspired by those used by V. Turaev in his work on Chern–Simons theory. Volume II of this Handbook will contain a chapter written by Baseilhac and Benedetti on that subject.

Let us also note that F. Bonahon and X. Liu worked out a finite-dimensional version of the quantization theory of Chekhov and Fock.

2 An overview of the content of this volume

The overview of the various sections of this Handbook that I will give now is also intended to give an idea of the large variety of ideas that Teichmüller theory involves.

2.1 The metric and the analytic theory, 1

Chapter 1, written by Georgios Daskalopoulos and Richard Wentworth, gives an overview of several important aspects of Teichmüller theory from the point of view of the theory of harmonic maps. The subjects that are treated in this chapter include the geometry of the Weil–Petersson metric (geodesics, curvature, isometries, completion, length functionals, convexity and superrigidity) and the study of group actions on \mathbb{R} -trees. In particular, the authors describe a harmonic map construction of the Teichmüller map, a harmonic map interpretation of the theorem of Hubbard and Masur on the existence and uniqueness of a quadratic differential with a given

equivalence class of horizontal foliation, and a harmonic map interpretation of the Thurston–Morgan–Shalen compactification of Teichmüller space. The chapter also contains an introduction to several basic topics in Teichmüller theory, including the uniformization theorem, quasiconformal maps, Beltrami differentials, the solution to the Beltrami equation, several proofs (including one that is based on harmonic maps) of Teichmüller’s theorem stating that Teichmüller space is a cell, Nielsen realization, a Higgs bundle interpretation of Teichmüller space in relation to flat $SL(2, \mathbb{C})$ connections, and holomorphic convexity. The authors also discuss similarities and differences between the mapping class group and arithmetic lattices.

To make the exposition more accessible, the authors also give an introduction to the theory of harmonic maps, with an emphasis on harmonic maps between surfaces, between surfaces and trees and between surfaces and Teichmüller space. The exposition includes some classical results by Gerstenhaber and Rauch, the results of Eells and Sampson on the existence of harmonic maps in a given homotopy class of maps in the case where the range is compact and has nonpositive sectional curvature, Hartmann’s uniqueness result in the case where the curvature is negative, and an overview of the relatively recent theory of Gromov and Schoen on harmonic maps between singular spaces and of the recent work by C. Mese on the proof of a conjecture made in 1954 by Gerstenhaber and Rauch on realizing extremal energy maps by Teichmüller maps.

Chapter 2 by Guillaume Th  ret and myself has two parts. The first part concerns Teichm  ller’s metric on Teichm  ller space, and the second part concerns Thurston’s asymmetric metric on this space. Teichm  ller’s metric regards Teichm  ller space from the point of view of conformal geometry, whereas Thurston’s asymmetric metric is based on considerations on Teichm  ller space from the point of view of hyperbolic geometry. The two metrics are Finsler metrics, and they are studied in the same chapter with the aim of drawing parallels and differences between them. This chapter contains basic facts about the two metrics and about the asymptotic geometry they induce on Teichm  ller space. The main results that are presented concern the behaviour of geodesic rays for the two metrics, in particular the convergence or non-convergence of certain classes of geodesic rays to points in Thurston’s boundary. This includes results on the limiting behaviour of stretch lines, which are geodesics for Thurston’s asymmetric metric, and on the limiting behaviour of anti-stretch lines, which are stretch lines traversed in the negative direction, and which in general are not geodesic lines for Thurston’s asymmetric metric. The chapter also contains a review of a parametrization of Teichm  ller space by a space of measured geodesic laminations that was introduced by Thurston, and which Thurston calls “cataclysm coordinates”.

The results concerning Teichm  ller’s metric which are surveyed in Chapter 2 are classical, whereas some of the results concerning Thurston’s asymmetric metric are new. We also discuss some open problems related to Thurston’s asymmetric metric.

Chapter 3, written by Robert Penner, contains an exposition of a generalized notion of a decorated hyperbolic structure.

A *decoration* of a hyperbolic surface with cusps is the choice of a closed horocycle around each cusp. The space of homotopy classes of decorated hyperbolic structures is a fibre bundle over Teichmüller space, called decorated Teichmüller space. This space was introduced by R. Penner in 1987, and the idea turned out to be extremely fruitful. A particularly useful set of parameters that Penner defined are the λ -length coordinates of this space. These are defined as signed distances between distinguished horocycles, measured on the edges of an ideal triangulation of the surface. There is a very simple and useful description of the pull-back of the Weil–Petersson symplectic form on decorated Teichmüller space in terms of λ -length coordinates. These coordinates, together with their transformation laws (the so-called “Ptolemy equations”), were used successfully in the study of the combinatorial structure of Teichmüller space. For instance, Penner used them to construct a mapping class group invariant cell-decomposition of decorated Teichmüller space and to compute Weil–Petersson volumes of moduli spaces. We already mentioned that the decorated theory, together with the λ -length coordinates, have been used as essential tools in the quantization theory of Teichmüller space, in particular in the work of R. Kashaev.

In Chapter 3, λ -length coordinates are reviewed and are generalized in two directions, namely, in the setting of homeomorphisms of the circle, and in what the author calls the decorated Teichmüller theory of the punctured solenoid. Let us say a few words about these two theories.

Universal Teichmüller theory first appeared in the work of Ahlfors and Bers. This theory can be formulated in terms of quasisymmetric homeomorphisms of the unit circle. These mappings arise as homeomorphisms which are boundary values of quasiconformal maps of the hyperbolic disk. In the context of universal Teichmüller theory, quasisymmetric homeomorphisms of the unit circle are used to parametrize hyperbolic structures on the unit disk relative to the boundary, which in particular parametrize Teichmüller spaces of all finite type surfaces. λ -length coordinates are used to parametrize the space of cosets of the subgroup of Möbius transformations in the group of orientation-preserving homeomorphisms of the circle, which is identified with a suitable space of tessellations of the hyperbolic disk. Some elements of universal Teichmüller theory are surveyed in Chapter 3 of this volume.

The (hyperbolic) solenoid \mathcal{H} was introduced in the 1990s by D. Sullivan as an inverse limit of the space of all branched covering of a closed Riemann surface. More precisely, let (S, x) be a pointed surface of genus > 1 . If $\pi_i : (S_i, x_i) \rightarrow (S, x)$ and $\pi_j : (S_j, x_j) \rightarrow (S, x)$ are two pointed coverings of (S, x) , then we write $\pi_i \leq \pi_j$ if the covering π_j factors through the covering $\pi_{j,i} : (S_j, x_j) \rightarrow (S_i, x_i)$ with $\pi_j = \pi_i \circ \pi_{j,i}$. Equipped with the relation \leq , the set of all coverings of (S, x) is inverse directed, and the solenoid \mathcal{H} is the inverse limit of this directed set. This space \mathcal{H} is equipped with a topology, viz. the subspace topology induced from the product topology on the infinite product of all closed surfaces in the finite coverings of (S, x) . With this topology, the local structure of \mathcal{H} is that of a plane times a Cantor set. The solenoid \mathcal{H} can also be equipped with a complex structures, and in fact, like surfaces of negative Euler characteristic, it can be equipped with a family of inequivalent complex structures.

The space of complex structures on the solenoid was studied by Sullivan and others. There is a Teichmüller space $\mathcal{T}(\mathcal{H})$ of \mathcal{H} , which can be defined as a certain closure of the stack of the Teichmüller spaces of all closed surfaces of genus > 1 . This space admits a complete metric, in analogy to Teichmüller's metric. The mapping class group of the solenoid \mathcal{H} is the set of isotopy classes of quasiconformal maps $h: \mathcal{H} \rightarrow \mathcal{H}$, and there is a natural action of this group on $\mathcal{T}(\mathcal{H})$. Volume II of this Handbook contains a chapter on the solenoid, written by D. Šarić.

In Chapter 3, Penner discusses joint work with Šarić on the punctured solenoid. In analogy to Sullivan's definition, the punctured hyperbolic solenoid is defined in this setting as the inverse limit of all finite unbranched pointed covers, the branching being permitted only over the punctures. The Teichmüller space of the punctured solenoid is a separable Banach space admitting a complete Teichmüller metric. This is a universal object in the sense that one can canonically embed in this space every classical Teichmüller space over a hyperbolic punctured surface. There are again λ -length coordinates on an appropriate decorated Teichmüller space of the punctured solenoid, and the structure of decorated Teichmüller space for finite type surfaces survives *mutatis mutandis* for this punctured solenoid.

Chapter 4 by Jean-Pierre Otal contains an exposition of the quasiconformal deformation theory of Riemann surfaces, of Bers' construction of the complex structure of Teichmüller space and of the theory of geodesic currents and of Hölder distributions. The space of geodesic currents is a subspace of the space of Hölder distributions. Both spaces were introduced by F. Bonahon in the setting of Fuchsian groups of finite co-area. According to Bonahon's definition, a geodesic current is a positive locally finite measure on the space of geodesic lines in the universal covering of the surface, which is invariant by the action of the group of covering transformations on that space. We recall that if S^1 is the boundary at infinity of the unit disk, considered as the universal covering of the surface, then the space of geodesic lines in the universal covering of the surface is naturally identified with the set $(S^1 \times S^1 \setminus \Delta)/(\mathbb{Z}/2)$, where Δ is the diagonal set of $S^1 \times S^1$ and where $\mathbb{Z}/2$ acts by exchanging coordinates. The space of geodesic currents is a complete uniform space. The terminology of currents was used by Bonahon after Sullivan who introduced it in the study of dynamical systems, in analogy to de Rham's currents. Bonahon defined a topological embedding of Teichmüller space into the space of geodesic currents, which he called the Liouville map. This embedding gives a new and unifying point of view on Teichmüller space, from which one can recover, for instance, Thurston's compactification by adjoining to this space the space of asymptotic rays, which are identified with measured geodesic laminations on the surface. The space of Hölder distributions is the dual space of the space of Hölder continuous functions with compact support. The definition of the space of Hölder distributions was extended to the case of an arbitrary Fuchsian group by Šarić, who showed that the Liouville map, generalized to that setting, is also a topological embedding. As in Bonahon's setting, Šarić defined a Thurston-type boundary by adding to the image of the Liouville map the set of asymptotic rays, which can be identified with bounded measured laminations on the surface, but here, the closure of

the image is in general not compact. Chapter 4 contains an exposition of this work of Šarić. The last part of this chapter contains new results by Otal on the analyticity of the Liouville map.

In Chapter 5, William Harvey considers the complex theory of moduli space of complex algebraic curves, and its relation with algebraic geometry and arithmetic geometry. This study mainly involves the following two interrelated topics:

— A discussion of Grothendieck's theory of dessins d'enfants and of Belyi's complex affine algebraic curves, with the action of the absolute Galois group of algebraic numbers $\text{Gal}(\overline{\mathbb{Q}})$ on these objects. One of the advantages of Harvey's review of Grothendieck's theory is that it is done by means of the classical tools of the theory of Teichmüller space, that is, the techniques introduced by Ahlfors and Bers.

— A study of Teichmüller disks in Teichmüller space, of their stabilizer groups and of their images in moduli space. In particular, the author addresses the question of the existence and the description of Teichmüller disks with large stabilizer groups, and he presents a construction of such disks that arises from hyperbolic tessellations of surfaces, defined by Hecke triangle groups and their subgroups.

We recall that a Teichmüller disk is a holomorphic and isometric embedding of the unit disk in the complex plane equipped with its hyperbolic metric into Teichmüller space equipped with Teichmüller's metric. Teichmüller disks were already studied by Teichmüller himself, and their study was revived by works of Thurston, Veech, McMullen and others.

As a matter of fact, Harvey describes several relations between Grothendieck's theory and the study of Teichmüller disks. The first relation is through a classical construction, due to Thurston, of pseudo-Anosov maps as products of Dehn twists along simple closed curves whose union fills the surface. This construction provides both a dessin d'enfant and a Teichmüller disk. Harvey then describes a general procedure for constructing Teichmüller curves in moduli space and he gives a characterization of all the curves of a given genus that are definable over $\overline{\mathbb{Q}}$. He presents a theorem stating that for any point in moduli space that corresponds to a curve defined over a number field (that is, a finite extension of \mathbb{Q}), there exists a Teichmüller disk of an arithmetic nature passing through it.

Chapter 5 also contains background material on the Ahlfors–Bers deformation theory of complex structures on surfaces as solutions of Beltrami partial differential equations, and an overview of some elements of the theory of universal Teichmüller space. The author also mentions works by C. McMullen, by P. Hubert, S. Lelièvre and T. A. Schmidt on Veech disk, and by P. B. Cohen and J. Wolfart on deformations of algebraic curves.

In Chapter 6, Frank Herrlich and Gabriela Schmithüsen study, among others, the following three questions:

- When is the image of a Teichmüller disk in moduli space an algebraic curve?
- What is the limiting behaviour of such a disk at the boundary of Teichmüller space and in Schottky space and its compactification?

— What are the stabilizer groups of Teichmüller disks in the mapping class group, and what is their relation with discrete subgroups of $\mathrm{PSL}(2, \mathbb{R})$ such as the Veech groups?

The authors present three different approaches to Teichmüller disks. First, such a disk can be considered as a complex geodesic in Teichmüller space, and in particular as a collection of (real) geodesic rays. The second approach starts with a (singular) flat structure that is induced on a Riemann surface by a holomorphic quadratic differential; a Teichmüller disk is then obtained by an affine deformation of this structure. Finally, a Teichmüller disk can also be obtained by varying the Beltrami differential of the quasiconformal mapping corresponding to a point in Teichmüller space. The authors show that all three approaches lead to the same object.

If the image of a Teichmüller disk in moduli space is an algebraic curve, it is called a Teichmüller curve. Whether this happens or not can be seen from the Veech group, which is introduced in this chapter geometrically as a Fuchsian group and equivalently as a subgroup of the mapping class group.

Chapter 6 also contains a report on W. Abikoff's construction of the augmented Teichmüller space and on an alternative point of view on that space, due to V. Braungardt, using a general complex-analytic device. As a matter of fact, Braungardt's construction is done in the category of locally complex ringed spaces.

A central aspect of Chapter 6 is the investigation of limit points of Teichmüller curves in the Abikoff–Braungardt boundary. The main result is that these points can be described as endpoints of Strebel rays corresponding to parabolic elements in the Veech group. The techniques build on Masur's analysis of Strebel rays.

Finally, a partial compactification of Schottky space is introduced in two ways. The relation between Teichmüller space, Schottky space, moduli space, their boundaries and the groups acting on them are worked out. Using the boundary of Schottky space, properties of the image of a Teichmüller disk in Schottky space are obtained.

2.2 The group theory, 1

In Chapter 7, Shigeyuki Morita studies several classes of groups which either generalize mapping class groups of surfaces, or are subgroups of mapping class groups. The classes of groups that are involved in this study include the following:

— Diffeotopy groups (that is, groups of isotopy classes of diffeomorphisms) of arbitrary smooth manifolds. When the manifold is a surface, then the diffeotopy group is the extended mapping class of this surface (that is, the group of isotopy classes of diffeomorphisms which do not necessarily preserve an orientation).

— Groups of isotopy classes of diffeomorphisms of smooth manifolds that preserve a volume form, respectively a symplectic form.

— The Torelli group, that is the subgroup of the mapping class group consisting of isotopy classes which act trivially on homology. Morita discusses the action of the mapping class group on the homology of the surface, and he presents properties of the Torelli group, including the fundamental work of D. Johnson on that group.

— The outer automorphism group of a free group. The exposition starts with a description of work done by Jacob Nielsen on this outer automorphism group. The outer automorphism group of a free group acts on a space introduced in the 1980s by M. Culler and K. Vogtmann, called *Outer space*. This space, with respect to its action by the outer automorphism group, is considered to play a role analogous to that of Teichmüller space, with respect to its action by the mapping class group. Morita explains important parallels between the two theories.

In this chapter, Morita presents important aspects of the algebraic structure of all these groups, in particular on the interplay between their cohomology groups, and he discusses some related open problems.

Chapter 8 by Lee Mosher expounds on the dynamics of subgroups of the mapping class group from the point of view of their action on Teichmüller space compactified by its Thurston boundary. The author presents important dynamical, geometric and algebraic properties of this action, with the aim of highlighting analogies between these actions and the actions of discrete groups of isometries of hyperbolic 3-space \mathbb{H}^3 on the union of that space with its boundary at infinity. From the purely dynamical point of view, the analogy starts with the notions of limit set and of domain of discontinuity. From the geometric point of view, the analogy is based on the notions of convex-compact subgroups and of Schottky subgroups. Teichmüller disks intervene in this study as isometric images of hyperbolic planes in Teichmüller space, their stabilizers being analogous to Fuchsian subgroups which in the classical case are isometry groups of \mathbb{H}^3 that stabilize isometrically embedded 2-dimensional hyperbolic spaces. From the algebraic point of view, the author presents the Tits alternative for subgroups of the mapping class group (proved by J. McCarthy) and the Leininger–Ried combination theorem which provides a method for building closed surface subgroups of mapping class groups. This last theorem is reminiscent of Maskit’s combination theorem that is used in the construction of discrete subgroups of isometries of hyperbolic spaces by combining simplest subgroups. The chapter also contains a survey of Thurston’s classification theorem of mapping classes.

Chapter 9 by Albert Marden concerns the deformation theory of Kleinian groups, specifically the interplay between the deformation space of a hyperbolic 3-manifold and the Teichmüller space of its boundary surfaces. In his report, Marden recalls several deep results related to this theory, including Thurston’s hyperbolization theorem, Mostow rigidity theorem and a theorem of Sullivan on the relation between the conformal structure of a simply connected region in the Riemann sphere and the geometry of a certain surface embedded in \mathbb{H}^3 . Sullivan’s result says that under some mild conditions, the boundary of the convex hull in \mathbb{H}^3 of the complement of an open subset U of the Riemann sphere (seen as the boundary at infinity of \mathbb{H}^3) is quasiconformally equivalent to U , with a quasiconformality constant being a universal constant. Chapter 9 includes a report on Bers’ embedding of Teichmüller space into the Banach space of quadratic differentials, which in Kleinian groups theory leads to the simultaneous uniformization theorem. It also includes a report on the augmented

space of representation which, when quotiented by the mapping class group action, gives a compactification of moduli space. The discussion on the representation space includes a review of the various notions of convergence in the study of the representation variety of the 3-manifold group in $\mathrm{PSL}(2, \mathbb{C})$.

Chapter 9 also contains a review of earthquakes and of their generalization to complex earthquakes, and the recent activity on that subject that stems from Sullivan's theorem and that has been carried out by D. Epstein, A. Marden and V. Markovic.

Chapter 10, by Ursula Hamenstädt, concerns the geometry of the complex of curves of a surface and its relation to the geometry of Teichmüller space. The natural action on this complex by the mapping class group of the surface has been thoroughly investigated in the last two decades by several people. The bulk of Chapter 10 is about the coarse geometry of this complex. Hamenstädt gives a new proof of a theorem due to H. Masur and Y. Minsky stating that the complex of curves is Gromov hyperbolic, and she reports on a result due to E. Klarreich describing the Gromov boundary of this complex. She describes a map from Teichmüller space to the complex of curves which is coarsely equivariant with respect to the action of the mapping class group on both spaces, with the property that the image under this map of a Teichmüller geodesic is a quasi-geodesic in the complex of curves.

Chapter 10 also contains an exposition of some basic facts on surface topology, namely, a description of the Hausdorff topology of the space of (not necessarily measured) laminations and of train track coordinates for this space.

2.3 Surfaces with singularities and discrete Riemann surfaces

Singular metrics on surfaces of constant curvature with conical singularities have been studied by several authors, including A. D. Alexandrov, W. P. Thurston, W. A. Veech, I. Rivin and B. Bowditch. Such structures naturally appear in the study of cellular decompositions of Teichmüller space and of moduli space, but also in other geometric contexts such as the study of patterns of circles. The existence and uniqueness results of M. Troyanov and his work on the analytic theory of spaces of singular flat metrics are often quoted as basic references on that subject.

In Chapter 11 of this volume, Marc Troyanov describes classical and new results on the deformation theory of flat metrics with cone singularities on surfaces. The chapter starts with a general introduction to the deformation theory of geometric structures on compact manifolds with boundary, based on the notions of developing map and holonomy homomorphism. The chapter also contains an introduction to the representation theory of a finitely presented group into a real algebraic group. This theory is applied to the special case of flat metrics with cone singularities on surfaces, in which the representation group is the group $\mathrm{SE}(2)$ of rigid motions of the Euclidean plane. The developing map and holonomy homomorphism of such structures are used to define an explicit structure on Teichmüller space that makes it a real algebraic variety, and a corresponding orbifold structure on the quotient moduli space. In the

special case where the surface has genus zero, this structure has already been studied by Deligne and Mostow, using techniques of algebraic geometry, and by Thurston. The techniques used by Troyanov are close to those of Thurston.

In Chapter 12, Charalampos Charitos and Ioannis Papadoperakis consider generalizations of hyperbolic structures (that is, metrics of constant curvature -1) on surfaces. This chapter is divided into two distinct parts that concern respectively hyperbolic structures with cone singularities on surfaces, and hyperbolic structures on simplicial 2-complexes.

The hyperbolic structures with cone singularities that are considered here have their cone angles $\geq 2\pi$. With this condition, the surfaces satisfy the CAT(-1) condition (the so-called Cartan–Alexandrov–Toponogov negative curvature condition), and the techniques of the theory of CAT(-1)-spaces can therefore be used in this study. The authors work out in detail coordinates for the moduli space of these structures in the case where the surface is a pair of pants with a unique cone point. In this case, the parameter space is homeomorphic to \mathbb{R}^6 .

Let us note that a pair of pants equipped with a hyperbolic metric with a unique cone point is the simplest surface that one can study in this context. In some sense, this case is a building block for the general theory of hyperbolic structures with conical singular points and with cone angles $\geq 2\pi$. Indeed, it is conceivable that the space of hyperbolic metrics with cone singularities can be parametrized by decomposing the surface into pairs of pants, in analogy to the case of non-singular hyperbolic structures, except that in this singular setting one has to deal with several kinds of degeneration of pairs of pants.

The other class of singular hyperbolic structures studied in Chapter 12 is the class of complete length metrics on 2-dimensional simplicial complexes in which each 2-simplex with its vertices deleted is isometric to a hyperbolic ideal triangle. Such a metric space is called a 2-dimensional ideal simplicial complex, and it also satisfies the CAT(-1)-condition. The authors describe parameters for the moduli space of ideal simplicial complexes, using lengths of closed geodesics, in analogy to Thurston’s parametrization of the Teichmüller space of a surface by embedding it in the space $\mathbb{R}_+^{\mathcal{S}}$ of functions on the set \mathcal{S} of homotopy classes of essential simple closed curves on the surface.

Chapter 13, by Christian Mercat, is an introduction to the theory of discrete Riemann surfaces. Here, a discrete Riemann surface is a topological surface equipped with a cell-decomposition whose faces are quadrilaterals, with positive weights assigned to the diagonals in such a way that for each quadrilateral the product of the weights of the two diagonals is equal to one. A discrete holomorphic map on a discrete Riemann surface is a function on the set of vertices satisfying a discrete version of the Cauchy–Riemann equations. These definitions lead to the notions of differential forms, holomorphic forms, wedge products, Dirichlet energy, a Hodge-star operator, harmonicity and period matrices for discrete Riemann surfaces. Using these notions, there are discrete analogs of a number of classical theorems in Riemann surface

theory, e.g. Hodge decomposition theorems, existence theorems for holomorphic forms with prescribed holonomies, and several approximation theorems. The results that are presented in Chapter 13 include a convergence theorem for the period matrices as a smooth Riemann surface is approximated by continuous ones, and results connecting a discrete version of the exponential function to the Bäcklund (or Darboux) transform for discrete holomorphic maps. The author also mentions a relation between discrete analytic functions and circle patterns.

2.4 The quantum theory of Teichmüller space, 1

In Chapter 14, Leonid Chekhov and Robert Penner present details of the Chekhov–Fock version of the quantization theory of the Teichmüller space of a punctured surface. They furthermore present results towards the quantization of its Thurston boundary, whose elements are projective measured foliations of compact support, and they show that the required operatorial limits exist weakly in the special case of the once-punctured torus. They introduce two classes of quantum operators, defined on the set of homotopy classes of essential simple closed curves on the surface, namely, quantum versions of the geodesic length operators associated to hyperbolic structures and quantum versions of the intersection function operators associated to measured foliations of compact support. Relating the quantization of Teichmüller space to the quantization of Thurston’s boundary is realized by showing that if a sequence of hyperbolic structures (g_n) converges to a projective class of a measured foliation λ , then the sequence of quantum operators associated to (g_n) converges weakly to the quantum operator associated to λ . This chapter also contains some basic material on Teichmüller theory, including an introduction to measured foliations, to train tracks and to shear coordinates associated to ideal triangulations.

Chapter 15 by Vladimir Fock and Alexander Goncharov is a report on mostly introductory material related to the geometry of surfaces with boundary and with distinguished points on their boundaries (which the authors call *ciliated surfaces*), and of the corresponding spaces of measured laminations. The authors give detailed descriptions of coordinates for these spaces, with formulae for the Poisson and the symplectic structures on Teichmüller and lamination spaces, of pairing between lamination and Teichmüller spaces, and of the action of the mapping class group in these coordinates. Their exposition also contains an interpretation of lamination spaces as tropical limits of Teichmüller spaces.

Chapter 16 by Jörg Teschner contains a construction of a modular functor out of quantum Teichmüller space. A modular functor is a functor from the category of Riemann surfaces with isotopy classes of embeddings as morphisms, to the category of vector spaces with linear maps as morphisms. In this work, the vector spaces are infinite-dimensional. Modular functors are natural generalizations of representations of mapping class groups. In his work, Teschner extends the set of operators corresponding to morphisms of a surface into itself (and arising from the quantum

Teichmüller space) to the morphisms between different surfaces in a compatible way. This chapter also contains a review of the coordinates for Teichmüller space that were used by Penner, by Fock and by Kashaev, as well as a self-contained presentation of the quantization theory of Teichmüller space.

Chapter 17 by Rinat Kashaev is a review of his quantization theory of the moduli space of irreducible flat $\mathrm{PSL}(2, \mathbb{R})$ -connections on a punctured surface, or, equivalently, of the space of conjugacy classes of irreducible representations of the fundamental group of the surface in $\mathrm{PSL}(2, \mathbb{R})$ with parabolicity conditions at the punctures. Teichmüller space embeds as a connected component of this space. The moduli space of irreducible flat $\mathrm{PSL}(2, \mathbb{R})$ -connections is equipped with a symplectic form defined by Goldman, which restricts to the Weil–Petersson form on Teichmüller space. This quantization theory leads to an infinite dimensional projective unitary representation of the mapping class group.

Part A

The metric and the analytic theory, 1

Chapter 1

Harmonic maps and Teichmüller theory

Georgios D. Daskalopoulos and Richard A. Wentworth

Contents

1	Introduction	34
2	Teichmüller space and extremal maps	37
	2.1 The Teichmüller theorems	37
	2.1.1 Uniformization and the Fricke space	38
	2.1.2 Quasiconformal maps	39
	2.1.3 Quadratic differentials and Teichmüller maps	40
	2.1.4 The Teichmüller space	42
	2.1.5 Metric definition of Teichmüller space	44
	2.2 Harmonic maps	46
	2.2.1 Definitions	47
	2.2.2 Existence and uniqueness	49
	2.2.3 Two dimensional domains	51
	2.2.4 A second proof of Teichmüller’s theorem	53
	2.3 Singular space targets	54
	2.3.1 The Gerstenhaber–Rauch approach	54
	2.3.2 \mathbb{R} -trees	57
	2.3.3 Harmonic maps to NPC spaces	59
3	Harmonic maps and representations	63
	3.1 Equivariant harmonic maps	63
	3.1.1 Reductive representations	63
	3.1.2 Measured foliations and Hopf differentials	66
	3.2 Higgs bundles and character varieties	69
	3.2.1 Stability and the Hitchin–Simpson theorem	69
	3.2.2 Higgs bundle proof of Teichmüller’s theorem	72
	3.2.3 The Thurston–Morgan–Shalen compactification	74
4	Weil–Petersson geometry and mapping class groups	77
	4.1 Weil–Petersson geodesics and isometries	77
	4.1.1 The Weil–Petersson metric and its completion	78
	4.1.2 The mapping class group	81
	4.1.3 Classification of Weil–Petersson isometries	83
	4.2 Energy of harmonic maps	85
	4.2.1 Nielsen realization	85

4.2.2	Properness of the energy	86
4.2.3	Convexity of energy and length functionals	87
4.2.4	Further applications	88
5	Harmonic maps to Teichmüller space	90
5.1	Existence of equivariant harmonic maps	90
5.1.1	Maps to the completion	90
5.1.2	Surface domains	92
5.1.3	Holomorphic maps from Riemann surfaces	93
5.2	Superrigidity	95
5.2.1	The Ivanov–Farb–Kaimanovich–Masur theorem	95
5.2.2	Harmonic maps from singular domains	96
	References	99

1 Introduction

Teichmüller theory is rich in applications to topology and physics. By way of the mapping class group the subject is closely related to knot theory and three-manifolds. From the uniformization theorem, Teichmüller theory is part of the more general study of Kleinian groups and character varieties. Conformal field theory and quantum cohomology make use of the algebraic and geometric properties of the Riemann moduli space.

At the same time, analytic techniques have been important in Teichmüller theory almost from the very beginning of the subject. Extremal maps and special metrics give alternative perspectives to moduli problems and clarify our understanding of a wide range of results. In some cases they can be used to obtain new properties.

The goal of this exposition is to present some of the more recent activity using analysis, and in particular harmonic maps, in the context of Teichmüller theory, representations of surface groups, mapping class groups, and Weil–Petersson geometry. Topics have been selected in order to illustrate the theme that the analytic and topological points of view complement each other in a useful way. For example, we shall present four different proofs of the fact that Teichmüller space is a cell, and we shall discuss the recent completion of a harmonic maps approach to Teichmüller’s existence and uniqueness theorems on extremal quasiconformal maps. Instead of a systematic survey of the subject, we have chosen to present the ideas behind the results through examples and in a rather informal way. There are very few proofs, but hopefully the references given at the end will provide the interested reader sufficient recourse for more details.

We make no attempt to exhaust all aspects of this subject. In particular, no mention is made of the work on quasiconformal harmonic maps of the disk and Schoen’s conjecture (see [117], [118], [159], [181], [190]), or of the universal Teichmüller space in general. Other topics that have been covered in great detail in the literature

have also been omitted or only briefly touched upon. For example, there is little discussion of the complex analytic theory of Teichmüller space, the Bers embedding, Royden's theorem on automorphisms, etc. We refer instead to the other chapters in this Handbook, in particular to [62] by A. Fletcher and V. Markovic. For the same reason, our summary of Weil–Petersson geometry is rather brief, and instead we refer to Wolpert's recent survey [206].

Finally, while we have tried to give complete and accurate references to the results stated in this chapter, given the expanse of the subject there will inevitably be omissions. For these we offer our apologies in advance. Two useful surveys of earlier results on harmonic maps are [50] and [166]. Relatively recent general texts on Teichmüller theory are [1], [86], [147]. The point of view taken in Tromba's book [186] is especially relevant to the material presented here. For an interesting account of Teichmüller's life and work, see Abikoff [2].

Georgios D. Daskalopoulos was partially supported by NSF Grant DMS-0204191. Richard A. Wentworth was partially supported by NSF Grants DMS-0204496 and DMS-0505512.

Notation

For simplicity, we shall deal with connected compact oriented surfaces without boundary and of genus $p \geq 2$. The notation we shall use is the following: S will denote the underlying smooth surface, and j will denote a complex structure on S . Hence, a Riemann surface is a pair (S, j) . The hyperbolic metric on S will be denoted by σ . Since it is uniquely determined by and uniquely determines the complex structure, the notation (S, j) and (S, σ) will both be understood to represent a Riemann surface structure. When the complex structure is understood we shall use letters S and R alone to denote Riemann surfaces, and hopefully this will not cause confusion. The following are some of the commonly used symbols in this chapter:

- id = identity map;
- \mathbf{I} = identity endomorphism;
- $f \sim f'$ homotopic maps;
- $\deg(f)$ = the degree of a map between surfaces;
- $K(f)$ = the dilatation of a quasiconformal map (Section 2.1.2);
- $\Gamma = \pi_1(S)$, or $= \pi_1(M)$ for a manifold M ;
- \tilde{M} = the universal cover of M ;
- Ω^p = the space of smooth p -forms;
- Λ = a Fuchsian group (Section 2.1.1);
- D = the unit disk in \mathbb{C} ;
- \mathbb{H} = the upper half plane in \mathbb{C} ;

- \mathbb{H}^3 = hyperbolic 3-space $\simeq \mathrm{SL}(2, \mathbb{C}) / \mathrm{SU}(2)$;
- $i(a, b)$ = the geometric intersection number of simple closed curves a, b on S ;
- $\ell_c[\sigma]$ = the length of a simple closed curve on S with respect to the hyperbolic metric σ ;
- \mathcal{F} = a measured foliation on S (Section 2.3.2);
- $i([c], \mathcal{F})$ = the intersection number of an isotopy class of simple closed curves with a measured foliation \mathcal{F} (see Section 2.3.2);
- $i(\mathcal{F}_1, \mathcal{F}_2)$ = the intersection number of a pair of measured foliations (see Section 2.3.2);
- $T_{\mathcal{F}}$ = the \mathbb{R} -tree dual to a measured foliation \mathcal{F} (Section 2.3.2);
- $\mathrm{MF}(S)$ (resp. $\mathrm{PMF}(S)$) = the spaces of measured (resp. projective measured) foliations on S (Section 2.3.2);
- K_S = the canonical line bundle on a Riemann surface S ;
- χ_S = the Euler characteristic of S ;
- ∇ = the covariant derivative, or a connection on a vector bundle V ;
- d_{∇} = the de Rham operator, twisted by a connection ∇ ;
- ∇_H = the Chern connection on a holomorphic bundle with hermitan metric H (Section 3.2.1);
- F_{∇} = the curvature of a connection ∇ ;
- Δ = the Laplace–Beltrami operator;
- μ = a Beltrami differential (Section 2.1.2);
- $\|\mu\|_{\infty}$ = the L^{∞} norm of a Beltrami differential μ ;
- φ = a holomorphic quadratic differential (Section 2.1.3);
- $\|\varphi\|_1$ (resp. $\|\varphi\|_2$) = the L^1 (resp. L^2) norms of a quadratic differential φ (see eqs. (2.6) and (2.17));
- T_{φ} = the \mathbb{R} -tree dual to the horizontal foliation of φ (Section 2.3.2);
- $\mathrm{QD}(S)$ = the space of holomorphic quadratic differentials;
- $\mathfrak{F}(S)$ = the Fricke space (Section 2.1.1);
- $\chi(\Gamma, r)$ (resp. $\chi(\Gamma)$) = the $\mathrm{SL}(r, \mathbb{C})$ (resp. $\mathrm{SL}(2, \mathbb{C})$) character varieties of Γ (Section 3.2.1);
- $\mathcal{T}(S)$ = Teichmüller space (Section 2.1.4);
- $\overline{\mathcal{T}}(S)$ = the Weil–Petersson completion of $\mathcal{T}(S)$ (Section 4.1.1);
- $\mathrm{Diff}(S)$, $\mathrm{Diff}^+(S)$, $\mathrm{Diff}_0(S)$ = the diffeomorphisms, orientation preserving diffeomorphisms, and diffeomorphisms connected to the identity of a surface S ;
- $\mathrm{Mod}(S)$ = the mapping class group (Section 4.1.2);

- $\mathcal{M}(S)$ = the Riemann moduli space (Section 4.1.2);
- $\overline{\mathcal{M}}(S)$ = the Deligne–Mumford compactification of $\mathcal{M}(S)$ (Section 4.1.2);
- $d_{\mathcal{T}}$ = the Teichmüller metric on $\mathcal{T}(S)$ (see eq. (2.9));
- d_{wp} = the Weil–Petersson metric on $\mathcal{T}(S)$ (Section 4.1.1);
- $\text{Iso}(X)$ = the isometry group of a metric space (X, d) ;
- ∂X = the ideal boundary of an NPC space X (Section 3.1);
- \mathcal{L}_{ρ} = the translation length function of a representation (3.1);
- $\mathcal{L}_{wp}[\phi]$ = the Weil–Petersson translation length of $[\phi] \in \text{Mod}(S)$ (see eq. (4.4));
- H^1 (resp. H^1_{loc}) = the Sobolev space of square integrable (resp. locally square integrable) functions with square integrable (resp. locally square integrable) distributional derivatives;
- $e(f)$ = the energy density of a map f (see eq. (2.19));
- $\pi_{\alpha\beta}$ = the directional energy tensor (see eq. (2.48));
- $E(f)$ = the energy of a map f (see eq. (2.20));
- $\text{End}(V)$ (resp. $\text{End}_0(V)$) = the endomorphism (resp. traceless endomorphism) bundle of a complex vector bundle V (Section 3.2.1);
- $\text{ad}(V)$ (resp. $\text{ad}_0(V)$) = the skew-hermitian (resp. traceless skew-hermitian) endomorphism bundle of a hermitian vector bundle V (Section 3.2.1);
- Φ = a Higgs field (Section 3.2.1);
- $\mathfrak{M}(S, r)$ = the moduli space of polystable Higgs bundles of rank r on S (Section 3.2.1).

2 Teichmüller space and extremal maps

2.1 The Teichmüller theorems

In this section we give a summary of the basics of Teichmüller theory from the point of view of quasiconformal maps. In Section 2.1.1 we review the uniformization theorem and the Fricke space. In Section 2.1.2 we introduce quasiconformal maps, Beltrami differentials, and we state the basic existence theorem for solutions to the Beltrami equation. We also formulate the extremal problem. In Section 2.1.3 we review quadratic differentials, Teichmüller maps and Teichmüller’s existence and uniqueness theorems. In Section 2.1.4 we define the Teichmüller space based on a Riemann surface and discuss the first approach to Teichmüller’s theorem on the contractibility of Teichmüller space. The proof that we give here is based on the notion of extremal maps, i.e. quasiconformal maps that minimize dilatation in their homotopy class. The connection between extremal and harmonic maps will be explained in Section 2.3.1. Finally, in Section 2.1.5, we provide an alternative definition of Teichmüller space via hyperbolic metrics.

2.1.1 Uniformization and the Fricke space. The famous *uniformization theorem* of Poincaré, Klein, and Koebe states that every closed Riemann surface S of genus at least 2 is biholomorphic to a quotient \mathbb{H}/Λ , where \mathbb{H} denotes the upper half plane and Λ is a group of holomorphic automorphisms of \mathbb{H} acting freely and properly discontinuously. Such a group can be identified with a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$, i.e. a *Fuchsian group* (cf. [63], [86]).

On \mathbb{H} we have the *Poincaré metric*

$$ds_{\mathbb{H}}^2 = \frac{|dz|^2}{(\mathrm{Im} z)^2}.$$

Under the biholomorphism $\mathfrak{h}: \mathbb{H} \rightarrow D$ given by $\mathfrak{h}(z) = (z-i)/(z+i)$, $ds_{\mathbb{H}}^2 = \mathfrak{h}^* ds_D^2$, where

$$ds_D^2 = \frac{4|dz|^2}{(1-|z|^2)^2}.$$

By a straightforward calculation the curvature of the Poincaré metric is constant equal to -1 , and by Pick's theorem its isometry group is $\mathrm{PSL}(2, \mathbb{R})$ (cf. [86]). In particular, this metric descends to the quotient \mathbb{H}/Λ . Hence, every Riemann surface of genus ≥ 2 has a hyperbolic metric, and this metric is unique. On the other hand, *any* Riemannian metric induces a unique complex structure. This is a consequence of Gauss' theorem on the existence of *isothermal coordinates*: if (S, g) is an oriented surface with Riemannian metric g , then S admits a unique complex structure j such that in local complex coordinates $g = g(z)|dz|^2$, where $g(z)$ is a smooth, positive (local) function. Hence, specifying a complex structure on the topological surface S is equivalent to specifying a hyperbolic metric. We shall use Greek letters, e.g. $\sigma = \sigma(z)|dz|^2$, to distinguish the hyperbolic from arbitrary Riemannian metrics g .

Let $\mathfrak{F}(S)$ denote the Fricke space of conjugacy classes of discrete embeddings $\Gamma = \pi_1(S) \rightarrow \mathrm{PSL}(2, \mathbb{R})$. Then $\mathfrak{F}(S)$ inherits a topology as a *character variety* (cf. [33], [72], [71] and Section 3.2.1 below, and the chapter by W. M. Goldman in Volume II of this Handbook [72]). The idea is to choose a *marking* of the genus p surface S , namely, a presentation

$$\Gamma = \langle a_1, \dots, a_p, b_1, \dots, b_p : \prod_{i=1}^p [a_i, b_i] = \mathrm{id} \rangle.$$

where the a_i and b_i are represented by simple closed curves on S with *geometric intersection numbers* satisfying $i(a_i, b_j) = \delta_{ij}$, $i(a_i, a_j) = i(b_i, b_j) = 0$. A homomorphism $\rho: \Gamma \rightarrow \mathrm{PSL}(2, \mathbb{R})$ is determined by specifying $2p$ elements $A_i, B_i \in \mathrm{PSL}(2, \mathbb{R})$ satisfying the relation $\prod_{i=1}^p [A_i, B_i] = \mathbf{I}$. A naive dimension count (which can easily be made precise at irreducible representations ρ) suggests that the dimension of the space of such homomorphisms is $6p - 3$. Since $\mathrm{PSL}(2, \mathbb{R})$ acts by conjugation, producing a 3-dimensional orbit, we have $\dim \mathfrak{F}(S) = 6p - 6$. Indeed, since the Fricke space consists of *discrete* embeddings, a more precise analysis can be given which realizes $\mathfrak{F}(S)$ as a subset of \mathbb{R}^{6p-6} (cf. [1]).

Proposition 2.1. *The Fricke space $\mathfrak{F}(S)$ is embedded in \mathbb{R}^{6p-6} .*

It is this embedding (the details of which will not be important) that we shall use to define the topology on $\mathfrak{F}(S)$. We shall see below that $\mathfrak{F}(S)$ is homeomorphic to Teichmüller space (Theorem 2.9).

2.1.2 Quasiconformal maps. An orientation preserving homeomorphism f of a domain $\Omega \subset \mathbb{C}$ into \mathbb{C} is called K -*quasiconformal* (or K -*qc*) if

- (1) f is of Sobolev class H_{loc}^1 , i.e. the distributional derivatives $f_z, f_{\bar{z}}$ are locally square integrable on Ω ;
- (2) there exists a constant $0 \leq k < 1$ such that $|f_{\bar{z}}| \leq k|f_z|$, almost everywhere on Ω , where $K = (1+k)/(1-k)$.

The infimum of $K \geq 1$ such that f is K -qc is called the *dilatation* of f , and it is denoted by $K(f)$. Clearly, 1-qc is equivalent to conformal.

An orientation preserving homeomorphism $f: S \rightarrow R$ between two Riemann surfaces is called K -qc if its lift to the universal cover $\tilde{f}: \mathbb{H} \rightarrow \mathbb{H}$ is K -qc. We define the dilatation $K(f)$ of f to be $K(\tilde{f})$. Given such a map f , let

$$\mathcal{QC}[f] = \{f': S \rightarrow R : f' \text{ is a qc homeomorphism homotopic to } f\}. \quad (2.1)$$

The main extremal problem in Teichmüller theory is a generalization to closed surfaces of Grötzsche's problem for rectangles (see [1]): given a qc map $f: S \rightarrow R$, let

$$K^*[f] = \inf_{f' \in \mathcal{QC}[f]} K(f'). \quad (2.2)$$

Teichmüller's extremal problem. Is $K^*[f]$ realized as the dilatation of a qc map, and if so, what are the properties of the map?

A qc homeomorphism f such that $K(f) = K^*[f]$ is called an *extremal map*. The existence of extremal maps is a relatively easy consequence of compactness properties of quasiconformal maps. The emphasis of this problem is therefore on the uniqueness and characterization of extremal maps. We shall give Teichmüller's answer to this question in the next section.

Choose coordinates (U, z) on S and (V, w) on R and set $F = w \circ f \circ z^{-1}$. Define the *Beltrami coefficient* of f with respect to the choice of coordinates by

$$\mu_f = \mu_f(z) d\bar{z} \otimes (dz)^{-1} = F_{\bar{z}}/F_z d\bar{z} \otimes (dz)^{-1}.$$

By (2), $|\mu_f(z)| < 1$ almost everywhere. The above expression is independent of the choice of coordinates w and transforms tensorially with respect to coordinate changes in z . More precisely, μ_f may be regarded as an L^∞ -section of the bundle $\bar{K}_S \otimes K_S^{-1}$, where K_S is the canonical line bundle of S . Notice, however, that $|\mu_f(z)|$ is independent of a choice of conformal coordinates. Set $\|\mu\|_\infty$ to be the essential supremum of $|\mu_f|$ over S .

Let $\mathcal{B}(S)$ denote the Banach space of L^∞ -sections of $\bar{K}_S \otimes K_S^{-1}$ with the L^∞ -norm. Set

$$\mathcal{B}_1(S) = \{\mu \in \mathcal{B}(S) : \|\mu\|_\infty < 1\}.$$

For any qc map $f: S \rightarrow R$ we associate $\mu_f \in \mathcal{B}_1(S)$. If $S = \mathbb{H}/\Lambda$ a Beltrami differential on S can be identified with an L^∞ function $\tilde{\mu}_f$ on \mathbb{H} satisfying the equations of automorphy

$$\tilde{\mu}_f(\gamma z) \frac{\overline{\gamma'(z)}}{\gamma'(z)} = \tilde{\mu}_f(z), \quad z \in \mathbb{H}, \gamma \in \Lambda. \quad (2.3)$$

Furthermore, qc homeomorphisms \tilde{f} of \mathbb{H} whose Beltrami coefficients satisfy (2.3) give deformations of Fuchsian groups via

$$\Lambda \rightsquigarrow \Lambda_\mu : \gamma \in \Lambda \mapsto \tilde{f} \circ \gamma \circ \tilde{f}^{-1} \in \mathrm{PSL}(2, \mathbb{R}). \quad (2.4)$$

Specifying the Beltrami coefficient and solving for a qc map is called *Beltrami's equation*. The following is the fundamental existence theorem for solutions to Beltrami's equation. The seminal reference is Ahlfors [5]. See also [86, Chapter 4].

Theorem 2.2. *For any Beltrami differential $\mu \in \mathcal{B}_1(\mathbb{C})$ there exists a unique qc homeomorphism f^μ of \mathbb{H} , extending continuously to $\hat{\mathbb{H}} = \mathbb{H} \cup \{\infty\}$, whose Beltrami coefficient is $\mu_{f^\mu} = \mu$, and which fixes the points 0, 1, and ∞ . Furthermore, f^μ depends complex analytically on μ .*

Corollary 2.3. *For any Beltrami differential $\mu \in \mathcal{B}_1(S)$ there exists a unique qc homeomorphism $f^\mu: S \rightarrow R$, for some Riemann surface R . More precisely, if $S = \mathbb{H}/\Lambda$, then $R = \mathbb{H}/\tilde{f}^\mu \circ \Lambda \circ (\tilde{f}^\mu)^{-1}$, where \tilde{f}^μ is the solution in Theorem 2.2 for the pullback Beltrami differential. Furthermore, f^μ depends complex analytically on μ .*

Hence, Beltrami differentials can be used to parametrize the Fricke space $\mathfrak{F}(S)$. Of course, there is an infinite dimensional family of Beltrami differentials giving conjugate Fuchsian groups.

2.1.3 Quadratic differentials and Teichmüller maps. By a *holomorphic quadratic differential* on a Riemann surface S we mean a holomorphic section of the line bundle K_S^2 . Set $\mathrm{QD}(S) = H^0(S, K_S^2)$. By the Riemann–Roch theorem, $\mathrm{QD}(S)$ is a complex vector space of dimension $3p - 3$, where p is the genus of S . If $\varphi \in \mathrm{QD}(S)$, then in local conformal coordinates (centered at z_0 , say) $\varphi = \varphi(z)dz^2$, where $\varphi(z)$ is a local holomorphic function. By a coordinate change we can write $\varphi = z^k dz^2$, where $k = 0, 1, 2, \dots$ is the order of vanishing of φ at z_0 . The coordinate system

$$w(z) = \int_{z_0}^z \sqrt{\varphi} = \int_{z_0}^z \sqrt{\varphi(z)} dz = \frac{2}{k+2} z^{\frac{k+2}{2}}$$

will be called the φ -coordinates around z_0 (if m is odd, this is multi-valued). Writing $w = u + iv$, the foliations $v = \text{constant}$ and $u = \text{constant}$ are called the *horizontal* and *vertical foliations* of φ , respectively.

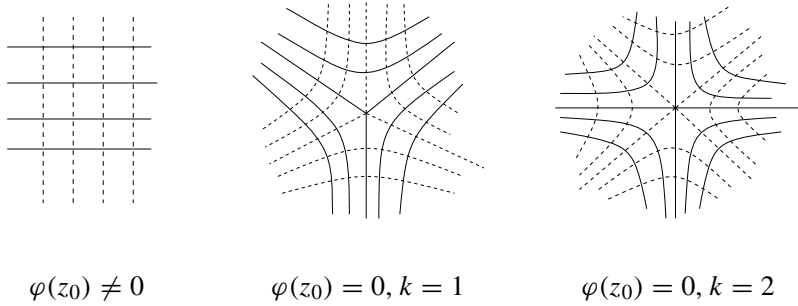


Figure 1

For more details, we refer to Strebel’s treatise on quadratic differentials [179].

A holomorphic quadratic differential on $S = \mathbb{H}/\Lambda$ is given by $\tilde{\varphi} = \tilde{\varphi}(z)dz^2$, where $\tilde{\varphi}$ is a holomorphic function on \mathbb{H} satisfying the equations of automorphy

$$\tilde{\varphi}(\gamma z)\gamma'(z)^2 = \tilde{\varphi}(z), \quad z \in \mathbb{H}, \gamma \in \Lambda. \tag{2.5}$$

Set

$$\text{QD}_1(S) = \{\varphi \in \text{QD}(S) : \|\varphi\|_1 < 1\},$$

where $\|\cdot\|_1$ denotes the L^1 -norm

$$\|\varphi\|_1 = \int_S |\varphi(z)| dx dy. \tag{2.6}$$

Given $\varphi \in \text{QD}_1(S) \setminus \{0\}$, we say that a qc homeomorphism $f : S \rightarrow R$ is a *Teichmüller map* for φ if the Beltrami coefficient of f satisfies

$$\mu_f = k \frac{\tilde{\varphi}}{|\varphi|}, \quad k = \|\varphi\|_1. \tag{2.7}$$

We are now in a position to give Teichmüller’s solution to the extremal problem stated in the previous section. First, a *Teichmüller map is uniquely extremal*.

Theorem 2.4 (Teichmüller’s Uniqueness Theorem). *Let $f : S \rightarrow R$ be a Teichmüller map. Then every $f' \in \mathcal{QC}[f]$ satisfies*

$$\|\mu_{f'}\|_\infty \geq \|\mu_f\|_\infty \quad (\text{equivalently, } K(f') \geq K(f)).$$

Moreover, the equality holds if and only if $f' = f$.

The second result asserts that *Teichmüller maps always exist*.

Theorem 2.5 (Teichmüller’s Existence Theorem). *In the homotopy class of every qc homeomorphism $f : S \rightarrow R$ there is either a conformal map or a Teichmüller map.*

Theorems 2.4 and 2.5 were first stated by Teichmüller (see [182]). His papers are “generally considered unreadable” (Abikoff, [1, p. 36]). Subsequent proofs were given in [3] and [13] (see also [66], [77]). Below we outline a proof of these two fundamental results based on harmonic maps to singular spaces (see Section 2.3.1).

Teichmüller maps are essentially affine with respect to a natural choice of coordinates (see [86]):

Theorem 2.6. *Fix $\varphi \in \text{QD}_1(S) \setminus \{0\}$, $k = \|\varphi\|_1 < 1$, and let $f: S \rightarrow R$ be a Teichmüller map for φ . Then there exists a unique holomorphic quadratic differential ψ on R satisfying the following conditions.*

- (1) *If z is a zero of φ , then $f(z)$ is a zero of ψ of the same order.*
- (2) *If z is not a zero of φ and ζ is a φ -coordinate about z , then there exists a ψ -coordinate w at $f(z)$ such that*

$$w \circ f = \frac{\zeta + k\bar{\zeta}}{1 - k}. \quad (2.8)$$

The quadratic differentials φ and ψ are called the *initial* and *terminal* differentials of the Teichmüller map f , respectively.

2.1.4 The Teichmüller space. We now come to the definition of Teichmüller space. Let S be a closed Riemann surface of genus $p \geq 2$. Consider triples (S, f, R) , where R is a Riemann surface and $f: S \rightarrow R$ is an orientation preserving diffeomorphism. Triples (S, f_1, R_1) and (S, f_2, R_2) are said to be equivalent if $f_2 \circ f_1^{-1}: R_1 \rightarrow R_2$ is homotopic to a biholomorphism. The set of all equivalence classes $[S, f, R]$ of triples (S, f, R) is denoted $\mathcal{T}(S)$ and is called the *Teichmüller space based on S* . The definition of $\mathcal{T}(S)$ turns out to be independent of the complex structure on S (see Theorem 2.7 below). Since any homeomorphism (in particular quasiconformal ones) is homotopic to a diffeomorphism, one obtains the same space if one considers pairs (S, f, R) where f is quasiconformal. This is a point of subtlety when dealing with Riemann surfaces with punctures.

Restricting as we are to the case of closed surfaces, Teichmüller space may be regarded as parametrizing complex structures up to biholomorphisms connected to the identity. Indeed, if $S_0 = (S, j_0)$ denotes the basepoint and (S, j) is another complex structure on the underlying surface S , then by choosing $f = \text{id}$ and $R = (S, j)$ there is an associated point $[j] = [S_0, \text{id}, R] \in \mathcal{T}(S)$. Two points $[S_0, \text{id}, R_1]$ and $[S_0, \text{id}, R_2]$ obtained in this way are equivalent if and only if (S, j_1) and (S, j_2) are biholomorphic via a map connected to the identity. Conversely, given any triple (S_0, f, R) , let j denote the pullback by f of the complex structure on R to the underlying surface S . Then by definition $f: (S, j) \rightarrow R$ is a biholomorphism; hence, $(S_0, \text{id}, (S, j))$ is equivalent to (S_0, f, R) . With this understood, we sometimes represent points in $\mathcal{T}(S)$ by equivalence classes $[j]$.

Given $[j_1], [j_2] \in \mathcal{T}(S)$, recall that $\mathcal{QC}[\text{id}]$ is the set of all qc homeomorphisms $(S, j_1) \rightarrow (S, j_2)$ homotopic to the identity. The *Teichmüller metric* $d_{\mathcal{T}}$ is defined by

$$d_{\mathcal{T}}([j_1], [j_2]) = \inf_{f \in \mathcal{QC}[\text{id}]} \log K(f). \quad (2.9)$$

For the next result we refer to [86, §5.1].

Theorem 2.7. *$\mathcal{T}(S)$ is a complete metric space with respect to the Teichmüller metric $d_{\mathcal{T}}$. Furthermore, given $[S, f, R] \in \mathcal{T}(S)$, the map $[f]_*: \mathcal{T}(S) \rightarrow \mathcal{T}(R)$ given by $[S, f', R'] \mapsto [R, f' \circ f^{-1}, R']$ is an isometry.*

Henceforth, the topology on $\mathcal{T}(S)$ is that given by the metric $d_{\mathcal{T}}$. Also, in light of the theorem we identify all Teichmüller spaces independent of the choice of base point. Now we are ready for Teichmüller's third result.

Theorem 2.8 (Teichmüller's theorem). *$\mathcal{T}(S)$ is homeomorphic to a cell of dimension $6p - 6$.*

By Corollary 2.3 on solutions to Beltrami's equation, Teichmüller maps with initial differential φ exist for any $\varphi \in \text{QD}_1(S)$. Hence, we may define a map

$$\tau: \text{QD}_1(S) \longrightarrow \mathcal{T}(S), \quad \tau(\varphi) = [S, f, R], \quad (2.10)$$

where f is a Teichmüller map for $\varphi \neq 0$, and $f = \text{id}$, $R = S$, for $\varphi = 0$. Theorem 2.8 follows from

Theorem 2.9. *The map τ in (2.10) is a homeomorphism. Moreover, $\mathcal{T}(S)$ is homeomorphic to $\mathfrak{F}(S)$.*

Proof. First, note that there is a natural bijection $F: \mathcal{T}(S) \rightarrow \mathfrak{F}(S)$ defined as follows: given $[S, f, R] \in \mathcal{T}(S)$, by the uniformization theorem applied to the Riemann surface R there is a discrete embedding $\rho_R: \pi_1(R) \rightarrow \text{PSL}(2, \mathbb{R})$, determined up to conjugation. Since the diffeomorphism f induces an isomorphism $f_*: \Gamma = \pi_1(S) \xrightarrow{\sim} \pi_1(R)$, we obtain a discrete embedding $\rho = \rho_R \circ f_*: \Gamma \rightarrow \text{PSL}(2, \mathbb{R})$. Notice that if $[S, f_1, R_1] = [S, f_2, R_2]$, then the corresponding homomorphisms are conjugate. Hence, there is a well-defined point $F[S, f, R] \in \mathfrak{F}(S)$. Conversely, given a discrete embedding $\rho: \Gamma \rightarrow \text{PSL}(2, \mathbb{R})$, consider the Riemann surface $R = \mathbb{H}/\rho(\Gamma)$. The Poincaré polygon theorem realizes the boundary of a fundamental domain for the action of $\rho(\Gamma)$ as the lift of simple closed curves α_i, β_i on R satisfying the relations $i(\alpha_i, \beta_j) = \delta_{ij}$, $i(\alpha_i, \alpha_j) = i(\beta_i, \beta_j) = 0$ (cf. [11]). The identification of a_i, b_i with α_i, β_i fixes a homotopy class of diffeomorphisms $f: S \rightarrow R$, and it is clear that $F[S, f, R] = [\rho]$. Hence, F is a bijection. Moreover, F is continuous by Corollary 2.3, since a qc map of small dilatation is close to the identity, hence the corresponding deformation of the Fuchsian groups is small. Consider the following

diagram

$$\begin{array}{ccc} \mathrm{QD}_1(S)_G & \xrightarrow{\tau} & \mathcal{T}(S) \\ & \searrow & \downarrow F \\ & & \mathfrak{F}(S) \end{array}$$

where $G = F \circ \tau$. By Theorems 2.4 and 2.5, τ is a bijection. It is also continuous. Indeed, $d_{\mathcal{T}}(\tau(0), \tau(\varphi)) = \log((1+k)/(1-k))$, where $k = \|\varphi\|_1$ (recall that $\tau(0) = S$), so τ is clearly continuous at the origin. Continuity at general points follows from the change of basepoints in Theorem 2.7. It follows that G is a continuous bijection. By the embedding $\mathfrak{F}(S) \hookrightarrow \mathbb{R}^{6p-6}$ (Proposition 2.1) and Invariance of Domain, G is a homeomorphism; hence, so are F and τ . \square

We have proven Teichmüller's Theorem via his existence and uniqueness results (Theorems 2.4 and 2.5). The proof uses the Fricke space $\mathfrak{F}(S)$ and the finite dimensionality of the space of holomorphic quadratic differentials. In Sections 2.2.4, 3.2.2, and 4.2.2 we shall give three alternative proofs of Theorem 2.8 using harmonic maps and the metric description of Teichmüller space.

2.1.5 Metric definition of Teichmüller space. Let S be an oriented surface of genus $p \geq 2$. Let $\mathrm{Met}_{\mathrm{hyp}}(S)$ be the space of metrics with constant curvature -1 . This has a smooth structure inherited as a smooth submanifold of the space $\mathrm{Met}(S)$ of all smooth metrics on S . As discussed in Section 2.1.1, a hyperbolic metric defines a complex structure on S via Gauss' theorem, and conversely, in every conformal class of metrics compatible with a given complex structure there is a unique hyperbolic metric. The group $\mathrm{Diff}_0(S)$ of diffeomorphisms isotopic to the identity acts on $\mathrm{Met}_{\mathrm{hyp}}(S)$ by pullback. Define

$$\mathcal{T}_{\mathrm{hyp}}(S) = \mathrm{Met}_{\mathrm{hyp}}(S)/\mathrm{Diff}_0(S), \quad (2.11)$$

with the quotient topology. By constructing a slice for the action of $\mathrm{Diff}_0(S)$ on $\mathrm{Met}_{\mathrm{hyp}}(S)$ it is not hard to prove (see [48], [60], [154], [186])

Proposition 2.10. $\mathcal{T}_{\mathrm{hyp}}(S)$ is a smooth manifold of dimension $6p - 6$.

To elaborate on this statement, we review the description of the tangent and cotangent spaces to $\mathcal{T}(S)$ (for this approach, cf. [41], [60]). Let ∇ denote the covariant derivative for a metric g on S . On the tangent space $T_g\mathrm{Met}(S)$ there is a natural L^2 -pairing:

$$\langle \delta g, \delta g' \rangle = \int_S (g^{\alpha\beta} g^{\mu\nu} \delta g_{\alpha\mu} \delta g'_{\beta\nu}) d \mathrm{vol}_{(S,g)} \quad (2.12)$$

where the metrics and variations are expressed with respect to local coordinates $\{x^\alpha\}$, $z = x^1 + ix^2 = x + iy$, and repeated indices are summed. For σ a hyperbolic metric, the condition that $\delta\sigma$ be tangent to $\mathrm{Met}_{\mathrm{hyp}}(S)$ is

$$0 = (-\Delta + 1) \mathrm{Tr}(\delta\sigma) + \nabla^\alpha \nabla^\beta (\delta\sigma_{\alpha\beta}), \quad (2.13)$$

where Δ is the Laplace–Beltrami operator associated to σ . Finally, the tangent space to the orbit $\text{Diff}_0(S) \cdot \sigma$ at σ consists of Lie derivatives of the metric:

$$\delta\sigma_{\alpha\beta} = (L_v\sigma)_{\alpha\beta} = \nabla_\alpha v_\beta + \nabla_\beta v_\alpha \quad (2.14)$$

for smooth vector fields $\{v^\alpha\}$. From (2.12) and (2.14), the L^2 -orthogonal in $T_\sigma \text{Met}(S)$ to $T_\sigma(\text{Diff}_0(S) \cdot \sigma)$ consists of variations satisfying

$$\nabla^\alpha \delta\sigma_{\alpha\beta} = 0. \quad (2.15)$$

Restricting to hyperbolic metrics, it then follows from (2.13) that these variations must also be traceless. Hence, we have an identification of $T_{[\sigma]}^* \mathcal{T}_{\text{hyp}}(S)$ with the space of traceless symmetric 2-tensors satisfying (2.15). Now the bundle of traceless symmetric 2-tensors is real isomorphic to K_S^2 via $2\varphi(z) = \delta\sigma_{11} - i\delta\sigma_{12}$. Moreover, (2.15) is precisely the statement that the corresponding quadratic differential $\varphi = \varphi(z)dz^2$ is holomorphic. Hence, $T_{[\sigma]}^* \mathcal{T}(S) \simeq \text{QD}(S)$.

This description makes contact with the Kodaira–Spencer theory of deformations of a complex structure (cf. [108]). Indeed, infinitesimal deformations of a complex structure are parametrized by smooth sections μ of $\bar{K}_S \otimes K_S^{-1}$. These are just (smooth) Beltrami differentials. Note that there is a natural pairing between Beltrami differentials and holomorphic quadratic differentials on a Riemann surface S obtained by raising indices in (2.12):

$$\langle \mu, \varphi \rangle = \int_S \mu(z)\varphi(z)|dz|^2, \quad (2.16)$$

where $\mu = \mu(z)d\bar{z} \otimes (dz)^{-1}$ and $\varphi = \varphi(z)dz^2$. Let $\mathcal{HB}(S)$ denote the space of *harmonic* Beltrami differentials, i.e.

$$\mathcal{HB}(S) = \{\mu \in \mathcal{B}(S) : \bar{\partial}^* \mu = 0\},$$

where the adjoint $\bar{\partial}^*$ is defined with respect to the hyperbolic metric. For any holomorphic quadratic differential φ , the Beltrami differential $\mu = \sigma^{-1}\bar{\varphi}$ is harmonic. Moreover, $\langle \mu, \varphi \rangle = \|\varphi\|_2^2$, where $\|\cdot\|_2$ denotes the L^2 -norm with respect to the metric σ :

$$\|\varphi\|_2^2 = \int_S |\varphi(z)|^2 \sigma(z)^{-1} dx dy. \quad (2.17)$$

It follows that the pairing

$$\langle \cdot, \cdot \rangle : \mathcal{HB}(S) \times \text{QD}(S) \longrightarrow \mathbb{C} \quad (2.18)$$

is nondegenerate and that the tangent space is given by $T_{[\sigma]} \mathcal{T}(S) = \mathcal{HB}(S)$.

To complete this circle of ideas, one can directly compute the Beltrami differential associated to $\delta\sigma$. Let σ_t be a differentiable family of hyperbolic metrics with $\sigma_0 = \sigma$, $(d\sigma_t/dt)|_{t=0} = \delta\sigma$, and let v_t be the Beltrami differentials associated to the identity map $(S, \sigma) \rightarrow (S, \sigma_t)$, $v_0 = 0$, $(dv_t/dt)|_{t=0} = \mu$. If $w = w_t$, $w_0 = z$ is a family of

conformal coordinates such that

$$ds_{\sigma_t}^2 = \sigma_t(w)|dw|^2 = \sigma_t(w)|w_z|^2|dz + v_t d\bar{z}|^2,$$

then since $\delta\sigma_{\alpha\beta}$ is traceless,

$$\left. \frac{d}{dt} \right|_{t=0} \sigma_t(w)|w_z|^2 = 0.$$

It then follows that $2\sigma\mu = \delta\sigma_{11} + i\delta\sigma_{12}$, in agreement with the correspondence above.

There is a canonical map $\mathfrak{c}: \mathcal{T}_{\text{hyp}}(S) \rightarrow \mathcal{T}(S)$ obtained by associating to an equivalence class of hyperbolic metrics the corresponding equivalence classes of complex structures obtained via Gauss' theorem (see Section 2.1.1). This map is continuous, for if two hyperbolic metrics are close in the smooth topology, then the identity has small dilatation. Furthermore, \mathfrak{c} is a bijection by the uniformization theorem. With this understood, we now see that the two definitions of Teichmüller space are equivalent.

Theorem 2.11. *The canonical map $\mathfrak{c}: \mathcal{T}_{\text{hyp}}(S) \rightarrow \mathcal{T}(S)$ obtained by associating to the hyperbolic metric its conformal class is a homeomorphism.*

Proof. Recall from the proof of Theorem 2.9 that the map $F: \mathcal{T}(S) \rightarrow \mathfrak{F}(S)$ is also a continuous bijection. Since $\mathfrak{F}(S) \subset \mathbb{R}^{6p-6}$, it follows by Proposition 2.10 and Invariance of Domain that the composition

$$F \circ \mathfrak{c}: \mathcal{T}_{\text{hyp}}(S) \rightarrow \mathfrak{F}(S) \hookrightarrow \mathbb{R}^{6p-6}$$

is a homeomorphism; hence, both F and \mathfrak{c} are as well. \square

Remark 2.12. (1) We emphasize that the proof of the homeomorphism $\mathcal{T}_{\text{hyp}}(S) \simeq \mathcal{T}(S)$ given above is independent of the Teichmüller Theorems 2.5, 2.4, and 2.8.

(2) By Theorem 2.11, we may regard the topological space $\mathcal{T}(S)$ either as equivalence classes of marked Riemann surfaces or as the moduli space of hyperbolic metrics. In particular, for the alternative proofs of Theorem 2.8 given below, it suffices to prove that $\mathcal{T}_{\text{hyp}}(S)$ is homeomorphic to a cell.

(3) The L^2 -metric (2.17) is the *Weil–Petersson cometric* on $\mathcal{T}(S)$ (see Section 4.1.1 below). In this description, it is easy to see that the Teichmüller metric (2.9) is a Finsler metric defined by the L^1 -norm (2.6).

2.2 Harmonic maps

This section is a brief summary of the theory of harmonic maps with an emphasis on those aspects that relate to Teichmüller theory. In Section 2.2.1 we give the basic definitions and present the variational formulation along with some examples. In Section 2.2.2 we state the existence and uniqueness theorem of Eells–Sampson–Hartman for nonpositively curved targets, and we indicate the importance of the Bochner formula. In Section 2.2.3 we specialize to the case of surface domains. We discuss

conformal invariance, the Hopf differential, and some applications. In Section 2.2.4 we present another proof that Teichmüller space is a cell using harmonic maps.

2.2.1 Definitions. Let (M, g) and (N, h) be Riemannian manifolds. With respect to coordinates $\{x^\alpha\}$ on M and $\{y^i\}$ on N , write $g = (g_{\alpha\beta})$, $h = (h_{ij})$. Given a smooth map $f: M \rightarrow N$, its differential

$$(df)_\alpha^k = (\partial f^k / \partial x^\alpha) dx^\alpha \otimes (\partial / \partial y^k),$$

is a section of the bundle $T^*M \otimes f^*TN$ with the induced metric and connection. Define the *energy density* and *energy of f* by

$$e(f) = \frac{1}{2} \langle df, df \rangle_{T^*M \otimes f^*TN} = \frac{1}{2} \frac{\partial f^i}{\partial x^\alpha} \frac{\partial f^j}{\partial x^\beta} g^{\alpha\beta} h_{ij} \circ f, \quad (2.19)$$

$$E(f) = \int_M e(f) d \operatorname{vol}_M, \quad (2.20)$$

respectively (repeated indices are summed). The energy can be viewed as a functional on the space of smooth maps between M and N .

The second extremal problem, analogous to the Teichmüller problem in Section 2.1.2, may now be formulated as follows: given a smooth map $f: (M, g) \rightarrow (N, h)$, let

$$E^*[f] = \inf \{E(f') : f' \text{ smooth, } f' \sim f\}. \quad (2.21)$$

Energy extremal problem. Is $E^*[f]$ realized as the energy of a smooth map, and if so, what are the properties of the map?

A smooth map f such that $E(f) = E^*[f]$ is called an *energy minimizer*. Unlike the problem for quasiconformal maps, existence of energy minimizers is not obvious. We shall discuss this at greater length in the next section.

The covariant derivative ∇df is a section of $\operatorname{Sym}^2(T^*M) \otimes f^*TN$, where Sym^2 denotes symmetric 2-tensors. The trace $\tau(f) = \operatorname{Tr}_g \nabla df$ is called the *tension field of f* . Let Δ denote the Laplace–Beltrami operator on (M, g) . Then

$$\tau(f)^k = \Delta f^k + (\Gamma_{ij}^k \circ f) \frac{\partial f^i}{\partial x^\alpha} \frac{\partial f^j}{\partial x^\beta} g^{\alpha\beta}.$$

Here, Γ_{ij}^k denotes the Christoffel symbols of N . A smooth map $f: M \rightarrow N$ is called *harmonic* if $\tau(f) \equiv 0$. Let

$$d_\nabla: \Omega^p(f^*TN) \longrightarrow \Omega^{p+1}(f^*TN)$$

denote the exterior derivative coupled with pulled-back Levi-Civita connection on N . It is easily seen that $d_\nabla(df) = 0$ for all differentiable maps. The equations for harmonic maps are then equivalent to

$$d_\nabla(*df) = 0, \quad (2.22)$$

i.e. df is a harmonic form (cf. [51], [50]). Here are some examples.

- Harmonic maps $S^1 \rightarrow N$ are closed geodesics in N .
- When $N = \mathbb{R}^n$ the harmonic map equations are equivalent to the harmonicity of the coordinate functions.
- Totally geodesic maps satisfy $\nabla df = 0$, and so are harmonic.
- Holomorphic or anti-holomorphic maps between Kähler manifolds are harmonic.
- Minimal isometric immersions are harmonic.

Now let us consider variational formulas for the energy $E(f)$. A smooth vector field v along f , i.e. $v \in C^\infty(f^*TN)$, defines a variation of f by $f_t(x) = \exp_{f(x)}(tv(x))$. Since N is assumed to be complete, this defines a smooth map $M \times \mathbb{R} \rightarrow N$ with $f_0 = f$. The *first variational formula* is

$$\delta_v E(f) = \left. \frac{dE(f_t)}{dt} \right|_{t=0} = - \int_M \langle \tau(f), v \rangle_h d \text{vol}_M. \quad (2.23)$$

It follows that the Euler–Lagrange equations for E are precisely the harmonic map equations (2.22).

In general there is a distinction between energy minimizers, smooth minimizers of E which then necessarily satisfy (2.22), and smooth solutions to (2.22) which may represent higher critical points of the energy functional. We shall see below that this distinction vanishes when the target manifold N has nonpositive curvature. Another case where minimizers can be detected is the following: let S be a compact Riemann surface and N a compact Kähler manifold.

Proposition 2.13. *If $f: S \rightarrow N$ is holomorphic or anti-holomorphic, then for any conformal metric on S , f is harmonic and is energy minimizing in its homotopy class.*

Indeed, a computation in local coordinates as above shows that for any smooth map $f: S \rightarrow N$,

$$E(f) = \int_S f^* \omega + 2 \int_S |\bar{\partial} f|^2 d \text{vol}_S \quad (2.24)$$

$$= - \int_S f^* \omega + 2 \int_S |\partial f|^2 d \text{vol}_S \quad (2.25)$$

where ω is the Kähler form on N . Since the first terms on the right hand sides depend only on the homotopy class of f , the result follows.

Now let $v, w \in C^\infty(f^*TN)$ and $f_{s,t}$ be a two-parameter family of maps such that $f_{0,0} = f$, $v = (\partial f_{s,t} / \partial s)|_{s=t=0}$, $w = (\partial f_{s,t} / \partial t)|_{(s,t)=(0,0)}$, where f is harmonic. Then

$$H_f(v, w) = \left. \frac{\partial^2 E(f_{s,t})}{\partial s \partial t} \right|_{s=t=0} = - \int_M \langle J_f v, w \rangle_h d \text{vol}_M, \quad (2.26)$$

where

$$J_f(v) = \text{Tr}_g(\nabla^2 v + \text{Riem}^N(df, v)df) \quad (2.27)$$

is the Jacobi operator, and Riem^N is the Riemannian curvature of (N, h) . In particular, if N has nonpositive Riemannian sectional curvature, then

$$H_f(v, v) \geq \int_M |\nabla v|^2 d \text{vol}_M \geq 0,$$

and hence every harmonic map is a local minimum of the energy.

Given smooth maps $f: M \rightarrow N$ and $\psi: N \rightarrow P$, one has the composition formula

$$\nabla d(\psi \circ f) = d\psi \circ \nabla df + \nabla d\psi(df, df).$$

Taking traces we obtain the formula for the tension (cf. [51])

$$\tau(\psi \circ f) = d\psi \circ \tau(f) + \text{Tr}_g \nabla d\psi(df, df). \quad (2.28)$$

In particular, if f is harmonic and ψ is totally geodesic then $\psi \circ f$ is also harmonic. If $P = \mathbb{R}$ and f is harmonic, then (2.28) becomes

$$\Delta(\psi \circ f) = \text{Tr}_g \nabla d\psi(df, df),$$

and therefore a harmonic map pulls back germs of convex functions to germs of subharmonic functions. The converse is also true:

Theorem 2.14 (Ishihara [87]). *A map is harmonic if and only if it pulls back germs of convex functions to germs of subharmonic functions.*

2.2.2 Existence and uniqueness. In the case of nonpositively curved targets the energy extremal problem has a solution. The basic existence result is the following

Theorem 2.15 (Eells–Sampson [51]). *Let $f: M \rightarrow N$ be a continuous map between compact Riemannian manifolds, and suppose that N has nonpositive sectional curvature. Then there exists an energy minimizing harmonic map homotopic to f .*

The proof is based on the heat equation method to deform a map to a harmonic one (cf. [78]). Namely, one solves the initial value problem for a nonlinear parabolic equation

$$\frac{\partial f}{\partial t}(x, t) = \tau(f)(x, t), \quad f(x, 0) = f(x). \quad (2.29)$$

Stationary solutions to (2.29) satisfy the harmonic map equations. Furthermore, by taking the inner product on both sides in (2.29) with $\tau(f)(x, t)$ and integrating over M , one observes, using (2.23), that the energy of the map $x \mapsto f(x, t)$ is decreasing in t . Hence, one hopes that as $t \rightarrow \infty$, $f(\cdot, t)$ converges to a harmonic map. Unfortunately, it turns out that this procedure does not always work. In general, even existence of a solution to (2.29) for all $t \geq 0$ is not guaranteed (cf. [23], [31]). However, we have

Theorem 2.16. *Assume M, N are compact Riemannian manifolds and N has nonpositive sectional curvature. Given a smooth map $f: M \rightarrow N$, then the solution to*

(2.29) exists for all $t \in [0, \infty)$ and converges as $t \rightarrow \infty$ uniformly to a harmonic map homotopic to f .

The key to this theorem is following parabolic *Bochner formula*. Suppose $f(x, t)$ is a solution to (2.29) for $0 \leq t < T$, and let $e(f)(x, t)$ denote the energy density of the map $x \mapsto f(x, t)$. Then for any orthonormal frame $\{u_\alpha\}$ at a point $x \in M$ we have the following pointwise identity:

$$-\frac{\partial e(f)}{\partial t} + \Delta e(f) = |\nabla df|^2 + \langle df \operatorname{Ric}^M(u_\alpha), df(u_\alpha) \rangle - \langle \operatorname{Riem}^N(df(u_\alpha), df(u_\beta))df(u_\beta), df(u_\alpha) \rangle. \quad (2.30)$$

where Ric^M is the Ricci curvature of (M, g) . In particular, if M is compact and N is nonpositively curved, then

$$\frac{\partial e(f)}{\partial t} \leq \Delta e(f) + Ce(f) \quad (2.31)$$

for some constant $C \geq 0$. If the solution $f(x, t)$ exists for $0 \leq t < T$, then it follows easily from (2.31) that $e(f)$ is uniformly bounded in x and $t < T$. The bound on the energy density means that the maps $f(\cdot, t)$ form an equicontinuous family from which convergence as $t \rightarrow T$ can be deduced. Resolving the initial problem at $t = T$ then allows one to extend the solution for some small time $t > T$. This is the rough idea behind the existence for all $0 \leq t < \infty$. In fact, more sophisticated methods show that $e(f)$ is bounded for all time (cf. [145]).

The Bochner formula for harmonic maps, i.e. stationary solutions of (2.29), is

$$\Delta e(f) = |\nabla df|^2 + \langle df \operatorname{Ric}^M(u_\alpha), df(u_\alpha) \rangle - \langle \operatorname{Riem}^N(df(u_\alpha), df(u_\beta))df(u_\beta), df(u_\alpha) \rangle. \quad (2.32)$$

As before, this implies

$$\Delta e(f) \geq -Ce(f). \quad (2.33)$$

Inequality (2.33) is the key to regularity of weakly harmonic maps to nonpositively curved spaces (cf. [166]). To state the result precisely, we note that if $\Omega \subset M$ is a domain with smooth boundary, one can solve the Dirichlet problem for an energy minimizing map $f: \Omega \rightarrow N$ with prescribed boundary conditions. If $f: M \rightarrow N$ is energy minimizing then it is automatically energy minimizing with respect to its boundary values for any $\Omega \subset M$. This is what is meant by *locally energy minimizing*. The following Lipschitz bound follows from (2.33) by iterating the Sobolev embedding.

Proposition 2.17. *If $f: \Omega \rightarrow N$ is harmonic with energy $E(f)$ and N has nonpositive curvature, then for any $U \subset\subset \Omega$,*

$$\sup_{x \in U} e(f)(x) \leq C(U)E(f)$$

for some constant $C(U)$ independent of f .

Next, we have the following result on uniqueness.

Theorem 2.18 (Hartman [79]). *Assume M, N are compact Riemannian manifolds and N has nonpositive sectional curvature. Let $f_0, f_1: M \rightarrow N$ be homotopic harmonic maps, and let $f_s: M \rightarrow N$ be a geodesic homotopy where $s \in [0, 1]$ is proportional to arc length. Then:*

- (1) *for every s , f_s is a harmonic map with $E(f_s) = E(f_0) = E(f_1)$; and*
- (2) *the length of the geodesic $s \mapsto f_s(x)$ is independent of x .*

In case N has negative sectional curvature, any nonconstant harmonic map $f: M \rightarrow N$ is unique in its homotopy class unless f maps onto a geodesic, in which case all homotopic harmonic maps are translations of f along the geodesic.

A good reference for these results is Jost's book [93].

The theorems above fail to hold if the curvature assumption on N is relaxed. In this case the analytic complexity increases substantially, and there is no satisfactory existence result in general. There is something of an exception in the case of surface domains (see [95], [162], [163], [169]), where the conformal invariance with respect to the domain metric leads to bubbling phenomena. We shall not attempt to present any results for the case of higher dimensional domains, since the relation with Teichmüller theory is less important.

2.2.3 Two dimensional domains. We now specialize to the case where the domain is a Riemann surface. Here the salient feature, as we have just mentioned above, is that the energy functional is invariant under conformal changes of metric on S , i.e. $g \mapsto e^\phi g$. Hence, the harmonic map equations for surface domains depend only on the complex structure on S .

Let $f: (S, \sigma) \rightarrow (N, h)$ be a smooth map, where N is an arbitrary Riemannian manifold. Then $\varphi = (f^*h)^{2,0} = \text{Hopf}(f)$ is a quadratic differential, called the *Hopf differential of f* . A key fact is that φ is holomorphic if f is harmonic. Indeed, in local coordinates, $\varphi = \varphi(z)dz^2$, where

$$\varphi(z) = \langle f_z, f_z \rangle = \frac{1}{4} (|f_x|^2 - |f_y|^2 - 2i \langle f_x, f_y \rangle). \quad (2.34)$$

Notice that $\varphi \equiv 0$ if and only if f is conformal. In normal coordinates at $f(z)$, the harmonic map equations are $\Delta f^k = 0$, for all k . Together with the vanishing of the derivatives of the metric, this implies

$$\varphi_{\bar{z}}(z) = \langle f_z, f_z \rangle_{\bar{z}} = (h_{ij}(f(z)) f_z^i f_z^j)_{\bar{z}} = 2h_{ij}(f(z)) f_z^i f_{z\bar{z}}^j = 0.$$

Several results in this chapter depend on the holomorphicity of the Hopf differential. In Section 2.3.3 we shall present a different argument due to Schoen [167] which works for a more general class of metric space targets.

To see how holomorphicity can have topological consequences, take for example the case where the target is also a Riemann surface R . Writing the metric h on R in

local conformal coordinates w , the energy of a map f is then

$$E_h(f) = \int_S h(f(z))(|f_z|^2 + |f_{\bar{z}}|^2) dx dy, \quad z = x + iy \quad (2.35)$$

where we have confused the notation f and $w \circ f$. When the metric h is understood, we shall simply write $E(f)$. The harmonic map equations are (cf. [170, Ch. 1])

$$f_{z\bar{z}} + \frac{h_w}{h} f_z f_{\bar{z}} = 0. \quad (2.36)$$

As an immediate application of (2.36) it follows that if $f: S \rightarrow R$ is harmonic, then $|\partial f|$ and $|\bar{\partial} f|$ are either identically zero or have a well-defined order. Indeed, if $H = f_z$ and $G = -(h_w/h)f_{\bar{z}}$, and ζ satisfies the equation $\zeta_{\bar{z}} = -G$, then it is easily checked that He^ζ is holomorphic. By setting $n_p = \text{ord}_p He^\zeta$ we obtain $|\partial f| = |z|^{n_p} k(z)$, where $k(z)$ is a smooth strictly positive function. We call n_p the *order* of $|\partial f|$ at p . This leads to

Theorem 2.19 (Eells–Wood [52]). *Let $f: S \rightarrow R$ be a harmonic map between surfaces. If $|\partial f|$ is not identically zero, then*

$$\sum_{|\partial f|(p)=0} n_p = \text{deg}(f)\chi_R - \chi_S.$$

If $|\bar{\partial} f|$ is not identically zero, then

$$\sum_{|\bar{\partial} f|(p)=0} m_p = -\text{deg}(f)\chi_R - \chi_S.$$

Here n_p and m_p are the orders of $|\partial f|$, $|\bar{\partial} f|$ at p , respectively. An immediate consequence of this is Kneser's theorem:

Corollary 2.20 (Kneser [106]). *Let $f: S \rightarrow R$ be a continuous map between surfaces, $\chi_R < 0$. Then $|\text{deg}(f)|\chi_R \geq \chi_S$.*

Pushing these ideas further, Schoen–Yau and Sampson proved

Theorem 2.21 (Schoen–Yau [168], Sampson [164], see also Jost–Schoen [98]). *Suppose $f: S \rightarrow R$ is a harmonic map between surfaces of the same genus. If $\text{deg } f = 1$ and R has negative curvature, then f is a diffeomorphism.*

Theorems 2.19 and 2.21 depend on the following formulas for a harmonic map between surfaces.

$$\Delta \log |\partial f| = -K_R J(f) + K_S, \quad \Delta \log |\bar{\partial} f| = K_R J(f) + K_S, \quad (2.37)$$

where K_S , K_R are the Gaussian curvatures of S and R , and $J(f) = |\partial f|^2 - |\bar{\partial} f|^2$ is the Jacobian of f . The equations (2.37) are related to the Bochner formula (2.32). The proof is a simple calculation which can be found, for example, in [170, Ch. 1].

As an application, the next theorem regarding the quotient (2.11) is due to Earle and Eells (cf. [47] and also [48]).

Theorem 2.22. *The bundle given by the quotient map $p: \text{Met}_{\text{hyp}}(S) \rightarrow \mathcal{T}_{\text{hyp}}(S)$ is trivial, i.e. there exists a homeomorphism H with the property that the diagram*

$$\begin{array}{ccc} \text{Met}_{\text{hyp}}(S)_p & \xrightarrow{H} & \mathcal{T}_{\text{hyp}}(S) \times \text{Diff}_0(S) \\ & \searrow & \downarrow \pi \\ & & \mathcal{T}_{\text{hyp}}(S) \end{array}$$

commutes, where π is the projection onto the first factor.

The map H can be constructed as follows: fix a metric $\sigma_0 \in \text{Met}_{\text{hyp}}(S)$. For any other $\sigma \in \text{Met}_{\text{hyp}}(S)$, let $f_\sigma: (S, \sigma_0) \rightarrow (S, \sigma)$ be the harmonic diffeomorphism \sim id from Theorem 2.21. Then H is defined by $F(\sigma) = (p(\sigma), f_\sigma^{-1})$.

2.2.4 A second proof of Teichmüller's theorem. We now give a second proof that Teichmüller space is a cell (Theorem 2.8) using harmonic maps and Hopf differentials as opposed to Teichmüller maps. Let $\varphi_\sigma = \text{Hopf}(f_\sigma)$ be the Hopf differential of the map f_σ defined above. By uniqueness of the harmonic diffeomorphism in its homotopy class (Theorem 2.18) we obtain a well-defined map

$$\mathcal{H}: \mathcal{T}_{\text{hyp}}(S) \longrightarrow \text{QD}(S), \quad [\sigma] \mapsto \mathcal{H}[\sigma] = \varphi_\sigma. \quad (2.38)$$

Then we have

Theorem 2.23 (Wolf [196]). *The map \mathcal{H} is a diffeomorphism.*

The fact that \mathcal{H} is 1-1 is due to Sampson [164]. The smooth dependence of \mathcal{H} follows easily as in [51]. This seems to have been first observed also by Sampson. That \mathcal{H} is proper is due to Wolf. The idea is based on the following energy bound (see also [137]):

$$E(f_\sigma) \leq 2 \int_S |\varphi_\sigma| - 2\pi \chi_S. \quad (2.39)$$

To see this, let $f: (S, \sigma_0) \rightarrow (S, \sigma)$ be any quasiconformal map with Beltrami coefficient μ and Hopf differential φ . Then

$$|\bar{\partial}f|^2 d \text{vol} = \sigma f_{\bar{z}} \bar{f}_z |dz|^2 = \sigma f_z \bar{f}_{\bar{z}} \frac{f_{\bar{z}}}{f_z} |dz|^2 = \varphi \mu |dz|^2 \leq |\varphi|,$$

since $|\mu| < 1$. Then since f_σ has degree 1, (2.39) is a consequence of the above inequality and (2.24). Similarly, using (2.25), one has by the same argument

$$2 \int_S |\varphi_\sigma| + 2\pi \chi_S \leq E(f_\sigma). \quad (2.40)$$

Properness of \mathcal{H} now follows from (2.39) and properness of the energy. The latter is due to Schoen and Yau [169]. In Section 4.2.2 we shall sketch the proof. Finally, that \mathcal{H} is onto follows from the properness and the fact that both the domain and target are smooth manifolds of dimension $6p - 6$. Hence, Theorem 2.8 is a consequence of the theorem above, along with Theorem 2.11.

2.3 Singular space targets

Harmonic maps to singular spaces were first introduced in a systematic way in the paper of Gromov and Schoen [75] in connection with arithmetic superrigidity. Since then the subject has played an important role in Teichmüller theory and is one of the main themes of this review. In Section 2.3.1 we shall indicate how singular space targets make a connection between the extremal maps discussed in Section 2.1.4 and the harmonic maps of Section 2.2.1. The highlight is the proof of Teichmüller's existence and uniqueness theorems. The idea, going back to Gerstenhaber and Rauch, provides a clear motivation for the use of singular targets from the point of view of Teichmüller theory. We shall defer the technical aspects of the general theory to Section 2.3.3. In Section 2.3.2 we discuss the notion of \mathbb{R} -trees and their connection to measured foliations and quadratic differentials. We also state the famous Hubbard–Masur theorem. Section 2.3.3 contains all of the technical results on harmonic maps to metric spaces that we shall need. There we give an outline of the main results of [75], [109], [110]. In addition, we describe several results that are special to harmonic maps to trees.

2.3.1 The Gerstenhaber–Rauch approach to Teichmüller's extremal problem. Teichmüller's extremal problem (Section 2.1.2) and the energy extremal problem (Section 2.2.1) bear obvious similarities; hence, the natural

Question. Are Teichmüller maps harmonic for some metric?

This leads to the notion of energy minimizing maps to singular space targets, which is the subject of this section. We begin with a simple example. Given a holomorphic quadratic differential ψ on a Riemann surface R , $|\psi|$ defines a singular flat metric with conical singularities at the zeros of ψ (cf. [179]). Indeed, away from the zeros we may write $|\psi| = |dw|^2$ for some conformal coordinate w , whereas at a zero of order $m \geq 1$, $|\psi| = |w|^m |dw|^2$. Notice that for $h(w) = |w|^m$, the Gauss curvature

$$K = -\frac{1}{2h} \Delta \log h \leq 0, \quad (2.41)$$

in the sense of distributions. We say that S with the metric $h = |\psi|$ is a *nonpositively curved space*.

Let S be another Riemann surface. Given a map $f: S \rightarrow R$ one can define the Sobolev class H^1 and the energy of f with respect to the singular conformal metric

$|\psi|$ on R by (2.35). Following the definitions of Section 2.2.1 we call such a map harmonic if it is an energy minimizer (see also Theorem 2.18). This is a special case of the general theory of Gromov, Korevaar, and Schoen that we shall describe below; in particular, such minimizers always exist and are Lipschitz by Theorem 2.31. The following result builds on earlier, weaker versions due to Miyahara [138] and Leite [115].

Theorem 2.24 (Kuwert [113]). *A Teichmüller map $f_0 : S \rightarrow R$ is the unique harmonic map in its homotopy class when R is endowed with the singular flat metric $h = |\psi|$ defined by the terminal quadratic differential of f_0 .*

Let us show how this gives a

Proof of Teichmüller's Uniqueness Theorem 2.4. Let $f : S \rightarrow R$ be any quasiconformal map with Beltrami differential μ_f . Then by (2.35) we have

$$\begin{aligned} E(f) &= \int_S (|f_z|^2 + |f_{\bar{z}}|^2) |\psi(f(z))| \, dx dy \\ &= \int_S (1 + |\mu_f|^2) |\psi(f(z))| |f_z|^2 \, dx dy \\ &\leq (1 + \|\mu_f\|_\infty^2) \int_S |\partial f|^2 \, d \operatorname{vol}_S. \end{aligned}$$

Now by (2.25), which continues to hold for the singular metric,

$$\begin{aligned} E(f) &\leq \frac{1}{2} (1 + \|\mu_f\|_\infty^2) (E(f) + C[f]), \\ E(f) &\leq \frac{1 + \|\mu_f\|_\infty^2}{1 - \|\mu_f\|_\infty^2} C[f], \end{aligned} \tag{2.42}$$

where $C[f]$ is a constant depending only on the homotopy class of $[f]$ and the area of the metric $|\psi|$. On the other hand, for the Teichmüller map f_0 we have by the same computation

$$E(f_0) = \frac{1 + \|\mu_{f_0}\|_\infty^2}{1 - \|\mu_{f_0}\|_\infty^2} C[f_0]. \tag{2.43}$$

If $f \sim f_0$, then $C[f] = C[f_0]$. By Theorem 2.24, $E(f_0) \leq E(f)$, which by (2.42) and (2.43) implies $\|\mu_{f_0}\|_\infty \leq \|\mu_f\|_\infty$, with equality if and only if $f = f_0$. \square

This result does not answer the question of existence of extremal maps by harmonic map methods. In their 1954 paper, Gerstenhaber and Rauch proposed a minimax method of finding a Teichmüller map [69]. Let $\mathcal{CM}(R)$ denote the space of conformal metrics on R with unit area and with at most conical singularities (see below for more details). For each $h \in \mathcal{CM}(R)$, let $E_h(f)$ be defined as in (2.35), where $f : S \rightarrow R$

is in H^1 . Gerstenhaber–Rauch conjectured that

$$\sup_{h \in \mathcal{C}\mathcal{M}(R)} \inf_{f \sim f_1} E_h(f) = \frac{1}{2} \left(K^*[f_1] + \frac{1}{K^*[f_1]} \right), \quad (2.44)$$

and that the sup-inf in (2.44) is realized by the Teichmüller map homotopic to f_1 . The problem was investigated further by Reich and Reich–Strebel in the case where S, R are both the disk [157], [158]. Kuwert, assuming the existence of the Teichmüller map, proved

Theorem 2.25 (Kuwert [113]). *The Teichmüller map f_0 and the singular metric $h_0 = |\psi|$ defined by its terminal differential realize the sup-inf in (2.44).*

The full Gerstenhaber–Rauch conjecture was recently proved by Mese (cf. [133], [135]) using the harmonic map theory of Gromov, Korevaar, and Schoen. Before we state Mese’s theorem we need to set up some notation and terminology. Let (X, d) be a metric space which is also a *length space*, i.e. for all pairs $p, q \in X$ there exists a rectifiable curve γ_{pq} whose length equals $d(p, q)$ (which we sometimes write d_{pq}). We call γ_{pq} a geodesic from p to q . Then X is *NPC* (= nonpositively curved) if every point of X is contained in a neighborhood U so that for all $p, q, r \in U$,

$$d_{pq_\tau}^2 \leq (1 - \tau)d_{pq}^2 + \tau d_{pr}^2 - \tau(1 - \tau)d_{qr}^2,$$

where q_τ is the point on γ_{qr} so that $d_{qq_\tau} = \tau d_{qr}$. Note that equality is achieved for every triple $p, q, r \in \mathbb{R}^2$. More generally, one defines a length space with curvature bounded above by κ by making comparisons with geodesic triangles in surfaces of constant curvature κ (cf. [132]). It follows from (2.41) that if $h(w)|dw|^2$ is a conformal metric on R with

$$\Delta \log h \geq -2\kappa h \quad (2.45)$$

then the induced metric space has curvature bounded above by κ (cf. [131]). We shall use this fact when we give a harmonic map construction of the Teichmüller map.

Let $\mathcal{C}\mathcal{M}_{\alpha, \kappa}(R)$ denote the set of metrics $h = h(w)|dw|^2$ on R where $h \geq 0$ is bounded of Sobolev class H^1 , satisfies (2.45) weakly, and has area = α . Let d_h denote the distance function associated to the above metric. As we have discussed before it is not hard to see that (R, d_h) has curvature bounded above by κ . The key result is the following

Theorem 2.26 (Mese [135]). *Let $h_i \in \mathcal{C}\mathcal{M}_{\alpha, \kappa}(R)$, $\kappa > 0$, and $f_i: S \rightarrow (R, h_i)$ be such that*

- (1) f_i is harmonic;
- (2) $\lim_{i \rightarrow \infty} E_{h_i}(f_i) = \sup_{h \in \mathcal{C}\mathcal{M}_{\alpha, \kappa}(R)} \inf_{f \sim f_1} E_h(f)$.

Then the f_i converge in the pullback sense to the Teichmüller map f_0 .

Convergence in the pullback sense is essentially Gromov–Hausdorff convergence. This will be explained in greater detail below (see Section 3.2.3). Theorem 2.26, along with earlier work, gives a proof of Teichmüller’s Existence Theorem 2.5.

2.3.2 \mathbb{R} -trees. The use of singular metrics to prove the Teichmüller theorems is motivation to study energy minimizing maps for other metric space targets. Here we discuss another ubiquitous example. An \mathbb{R} -tree is a length space such that any two points can be joined by a *unique* path parametrized by arc length. This path is called *the geodesic* between the points, say p, q , and it is denoted pq . An equivalent definition is that an \mathbb{R} -tree is a simply connected length space with curvature bounded above by κ for any $\kappa \in \mathbb{R}$ (cf. [180]).

Example 2.27. Let T be a simplicial tree, i.e. a simply connected 1-dimensional simplicial complex. Then T can be thought of as an \mathbb{R} -tree by assigning to each edge a unit length. An \mathbb{R} -tree is called simplicial if it is obtained from a simplicial tree in this way. Note that we do not assume the simplicial tree is locally finite, although the set of vertices clearly is.

Example 2.28. Take $T = \mathbb{R}^2$ and define $d(p, q) = |p - q|$ if p, q lie on some ray from the origin, and $d(p, q) = |p| + |q|$, otherwise. Clearly, T with this metric is not locally compact, though it is simplicial.

Example 2.29. A slight modification of the above yields a non-simplicial tree. Again take $T = \mathbb{R}^2$ and define $d(p, q) = |p - q|$ if p and q lie on the same vertical line. In all other cases, let $d(p, q) = d(p, p') + d(p', q') + d(q, q')$, where p', q' are the projections of p, q to the x -axis. Then every point on the x -axis becomes a vertex.

\mathbb{R} -trees appear in Teichmüller theory in several ways. The primary example is the leaf space of the horizontal and vertical foliations of a holomorphic quadratic differential. First recall that a *measured foliation* \mathcal{F} on a surface S with singularities at the points z_1, \dots, z_ℓ and multiplicities k_1, \dots, k_ℓ is described by the following (cf. [57]): an open cover $\{U_i\}$ of $S \setminus \{z_1, \dots, z_\ell\}$ and open sets V_1, \dots, V_ℓ about z_1, \dots, z_ℓ along with smooth real valued functions u_i defined on U_i such that

- (1) $|du_i| = |du_j|$ on $U_i \cap U_j$;
- (2) $|du_i| = |\operatorname{Im}(z - z_j)^{k_j/2} dz|$ on $U_i \cap V_j$.

Clearly, $\ker du_i$ defines a vector field on S which integrates to give a foliation away from $\{z_1, \dots, z_\ell\}$, with $(k_j + 2)$ -pronged singularities at z_j (see Figure 1). A leaf containing a singularity is called a *critical trajectory*, whereas the other leaves are called *noncritical*. An important attribute of measured foliations is that they carry a *transverse measure*. More precisely, if c is a rectifiable path then we denote by $\nu(c)$, the number $\nu(c) = \int_c |du|$, where $|du|$ is defined by $|du|_{U_i} = |du_i|$. An important feature of this measure is its translation invariance along the leaves. Namely, if c_0 is a path transverse to the foliation \mathcal{F} , and if we deform c_0 to c_1 via an isotopy that

maintains the transversality to the foliation at every time, then $\nu(c_1) = \nu(c_0)$. For the free homotopy class $[c]$ of a simple closed curve we define

$$i([c], \mathcal{F}) = \inf\{\nu(c) : c \in [c]\}. \quad (2.46)$$

Two measured foliations (\mathcal{F}, ν) and (\mathcal{F}', ν') are called equivalent if $i([c], \mathcal{F}) = i([c], \mathcal{F}')$ for all free homotopy classes of simple closed curves. We denote the space of equivalence classes of measured foliations on S by $\text{MF}(S)$. Then the collection of intersection numbers (2.46), as c ranges over isotopy classes of simple closed curves, endows $\text{MF}(S)$ with a topology. We call \mathcal{F} and \mathcal{F}' *projectively equivalent* if there is $b > 0$ such that $i([c], \mathcal{F}') = b i([c], \mathcal{F})$ for all free homotopy classes of simple closed curves. In this case, we write $\mathcal{F}' = b\mathcal{F}$. The space of projective equivalence classes will be denoted $\text{PMF}(S)$.

Given a measured foliation (\mathcal{F}, ν) we can associate a *dual tree* $T_{\mathcal{F}}$ to the foliation with an isometric action of $\Gamma = \pi_1(S)$. Explicitly, let $(\tilde{\mathcal{F}}, \tilde{\nu})$ denote the pullback of (\mathcal{F}, ν) to the universal cover \mathbb{H} of S . On \mathbb{H} we define a pseudodistance \tilde{d} via

$$\tilde{d}(p, q) = \inf\{\tilde{\nu}(c) : c \text{ a rectifiable path between } p, q\}.$$

It follows by [18, Corollary 2.6] that the Hausdorffification of (\mathbb{H}, \tilde{d}) is an \mathbb{R} -tree with an isometric action of Γ . Strictly speaking, the setup in [18] works for measured foliations on arbitrary 2-complexes. The approach is useful in that it avoids introducing the notion of a geodesic lamination. For a proof using laminations, see [142], [153].

For a holomorphic quadratic differential $\varphi \neq 0$ on S we have seen in Section 2.1.3 how to define horizontal and vertical foliations. If the φ -coordinate is locally given by $w = u + iv$, then transverse measures may be defined by $|du|$ and $|dv|$, respectively. In other words, a nonzero quadratic differential defines a measured foliation via its horizontal foliation. We denote the corresponding dual tree by T_{φ} .

The following fundamental theorem, due to Hubbard–Masur and also announced by Thurston, asserts that every measured foliation on S arises in this way:

Theorem 2.30 (Hubbard–Masur [84]). *Given a measured foliation (\mathcal{F}, ν) on a closed Riemann surface S of genus $p \geq 2$ there is a unique holomorphic quadratic differential whose horizontal foliation is equivalent to (\mathcal{F}, ν) . In particular, $\text{MF}(S)$ is homeomorphic to $\mathbb{R}^{6p-6} \setminus \{0\}$, and $\text{PMF}(S) \simeq S^{6p-6}$.*

In Section 3.1.2 we shall sketch how we can interpret the Hubbard–Masur theorem via harmonic maps to trees (see [198], [199]).

There is a particular class of quadratic differentials on S called *Jenkins–Strebel differentials* (cf. [179]). They are characterized by the property that the noncritical trajectories are all closed and they partition the complement of the critical trajectories in S into cylinders with the standard foliations (see Figure 2). Notice that in this case the dual tree T_{φ} is a simplicial tree with a Γ action. The quotient T_{φ}/Γ is a graph G_{φ} , and the quotient map $\tilde{p}: \mathbb{H} \rightarrow T_{\varphi}$ descends to a map $p: S \rightarrow G_{\varphi}$, as indicated in Figure 2.

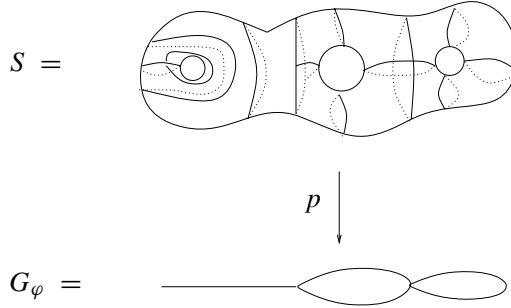


Figure 2

Hence, the intersection number (2.46) number is a generalization of the geometric intersection number of simple closed curves. Let us point out two facts (cf. [57], [179]).

- There are examples of measured foliations where all the noncritical leaves are noncompact. For example, the fixed points in $\text{PMF}(S)$ of pseudo-Anosov mapping classes. These give rise to \mathbb{R} -trees which are not simplicial.
- However, the measured foliations whose associated trees are simplicial are dense in $\text{MF}(S)$. Furthermore, the intersection number (2.46) extends continuously to $\text{MF}(S) \times \text{MF}(S)$.

Prior to a rigorous definition of energy minimizers to NPC spaces, we first introduce the notion of a harmonic map to a tree. This definition is due to Wolf [197] and is motivated by Ishihara's Theorem 2.14. As we shall see in the next subsection, it turns out that for the case of trees it is equivalent to the definition of energy minimizers due to Korevaar–Schoen.

Let S be a Riemann surface and let (T, d) be a minimal \mathbb{R} -tree with an isometric action of $\Gamma = \pi_1(S)$. Let $f: \mathbb{H} \rightarrow T$ be a Γ -equivariant, continuous map. We say that f is *harmonic* if it pulls back germs of convex functions on T to germs of subharmonic functions on \mathbb{H} . Notice that a function $f: U \rightarrow \mathbb{R}$, where U is a convex open subset of an \mathbb{R} -tree, is called convex if for any segment $pq \subset U$ and $r \in pq$ we have

$$(f(r) - f(p))d(q, r) \leq (f(q) - f(r))d(p, r).$$

A basic example of a harmonic map to a tree is the projection

$$p: \mathbb{H} \longrightarrow T_\varphi \tag{2.47}$$

where T_φ is the dual tree to the horizontal foliation of a holomorphic quadratic differential. It is not hard to see by direct observation that p is harmonic (cf. [197]).

2.3.3 Harmonic maps to NPC spaces. For the purpose of this subsection (Ω, g) will be a bounded Riemannian domain of dimension m with Lipschitz boundary and

(X, d) any complete NPC space. References for the following are [75], [96], [109], [110]. The generalization to the case where X is assumed only to have curvature bounded from above can be found in [132].

A Borel measurable map $f: \Omega \rightarrow X$ is said to be in $L^2(\Omega, X)$ if for $p \in X$,

$$\int_{\Omega} d^2(p, f(x)) d \operatorname{vol}_{\Omega}(x) < \infty.$$

By the triangle inequality, the condition is independent of the choice of point p . For $f \in L^2(\Omega, X)$ we construct an ε -approximate energy function $e_{\varepsilon}(f): \Omega_{\varepsilon} \rightarrow \mathbb{R}$, where $\Omega_{\varepsilon} = \{x \in \Omega : d(x, \partial\Omega) > \varepsilon\}$ by

$$e_{\varepsilon}(f)(x) = \frac{1}{2\omega_m} \int_{\partial B_{\varepsilon}(x)} \frac{d^2(f(x), f(y)) d\sigma(y)}{\varepsilon^2 \varepsilon^{m-1}},$$

where ω_m is the volume of the unit sphere in \mathbb{R}^m and $d\sigma$ is the induced volume on the sphere $\partial B_{\varepsilon}(x) \subset \Omega$ of radius ε about x . Setting $e_{\varepsilon}(f)(x) = 0$ for $x \in \Omega \setminus \Omega_{\varepsilon}$, we can consider $e_{\varepsilon}(f)$ to be an L^1 function on Ω . In particular, it defines a linear functional $E_{\varepsilon}: C_c(\Omega) \rightarrow \mathbb{R}$. We say that f has *finite energy* (or that $f \in H^1(\Omega, X)$) if

$$E(f) \equiv \sup_{0 \leq \varphi \leq 1} \limsup_{\varepsilon \rightarrow 0} E_{\varepsilon}(\varphi) < \infty.$$

It can be shown that if f has finite energy, the measures $e_{\varepsilon}(f)(x) d \operatorname{vol}_{\Omega}(x)$ converge weakly to a measure that is absolutely continuous with respect to Lebesgue measure on Ω . Therefore, there is a well-defined integrable function $e(f)(x)$, which we call the *energy density*, so that for each $\varphi \in C_c(\Omega)$,

$$\lim_{\varepsilon \rightarrow 0} \int_{\Omega} e_{\varepsilon}(f)(x) \varphi(x) d \operatorname{vol}_{\Omega}(x) = \int_{\Omega} e(f)(x) \varphi(x) d \operatorname{vol}_{\Omega}(x).$$

By analogy with the case of smooth maps we write $e(f)(x) = \frac{1}{2} |\nabla f|^2(x)$ with total energy

$$E(f) = \frac{1}{2} \int_{\Omega} |\nabla f|^2 d \operatorname{vol}_{\Omega}.$$

Similarly, the directional energy measures $|f_*(Z)|^2 d \operatorname{vol}_{\Omega}$ for $Z \in \Gamma(T\Omega)$ is a Lipschitz tangent vector field can also be defined as a weak-* limit of measures ${}^Z e_{\varepsilon}(f) d \operatorname{vol}_{\Omega}$. Here,

$${}^Z e_{\varepsilon}(f)(x) = \frac{d^2(f(x), f(\bar{x}(x, \varepsilon)))}{\varepsilon^2},$$

where $\bar{x}(x, \varepsilon)$ denotes the flow along Z at time ε , starting at x . For almost all $x \in \Omega$,

$$|\nabla f|^2(x) = \frac{1}{\omega_m} \int_{S^{m-1}} |f_*(v)|^2 d\sigma(v),$$

where S^{m-1} is the unit sphere in $T_x\Omega$. This definition of the Sobolev space $H^1(\Omega, X)$ is consistent with the usual definition when X is a Riemannian manifold.

For any map $f \in H^1(\Omega, X)$ we can also make sense of the notion of the pullback metric

$$\pi : \Gamma(T\Omega) \times \Gamma(T\Omega) \longrightarrow L^1(\Omega) \quad (2.48)$$

defined by

$$\pi(V, W) = \frac{1}{4}|f_*(V+W)|^2 - \frac{1}{4}|f_*(V-W)|^2, \quad V, W \in \Gamma(T\Omega).$$

If the tangent space to (Ω, g) has a local frame (u_1, \dots, u_m) , we write $\pi_{\alpha\beta} = \pi(u_\alpha, u_\beta)$, and

$$e(f) = \frac{1}{2}|\nabla f|^2 = \frac{1}{2}g^{\alpha\beta}\pi_{\alpha\beta}. \quad (2.49)$$

The L^1 -tensor will be used in the next section to define the Hopf differential.

A finite energy map $f : \Omega \rightarrow X$ is said to be *harmonic* if it is locally energy minimizing. In other words, for each point $x \in \Omega$ and each neighborhood of x , all comparison maps agreeing with f outside this neighborhood have total energy no less than f . The following are the basic existence and regularity results. For an alternative approach, see [96].

Theorem 2.31 (Korevaar–Schoen [109], see also [172]). *Let (X, d) be an NPC space. If $f : \Omega \rightarrow X$ is harmonic, then f is locally Lipschitz continuous. The Lipschitz constant on $U \subset\subset \Omega$ is of the form $C(U)\sqrt{E(f)}$, where $C(U)$ is independent of the map f (cf. Proposition 2.17).*

Theorem 2.32 (Korevaar–Schoen [109]). *Let (X, d) be compact and NPC. Let M be a compact Riemannian manifold without boundary, and $f : M \rightarrow X$ a continuous map. Then there exists a Lipschitz harmonic map homotopic to f .*

Note that Theorem 2.32 is a generalization of the Eells–Sampson Theorem 2.15. The uniqueness result in the singular case is due to Mese.

Theorem 2.33 (Mese [134]). *Let M be a compact Riemannian manifold and X a compact metric space with curvature bounded above by a constant $\kappa < 0$. If $f : M \rightarrow X$ is a nonconstant harmonic map, then f is unique in its homotopy class unless it maps onto a geodesic.*

An important tool in understanding the structure of harmonic maps is the *monotonicity formula* for energy minimizers. The idea goes back to Almgren [7]. The statement is that for nonconstant energy minimizers, the quantity

$$e^{C\varepsilon} \frac{\varepsilon \int_{B_\varepsilon(x)} |\nabla f|^2 d \text{vol}_\Omega}{\int_{\partial B_\varepsilon(x)} d^2(f(x), f(y)) ds(y)}, \quad (2.50)$$

is monotone increasing in ε , for some constant C . The extension of this to singular space targets was obtained in [75], [167], and further developed in [132]. The basic idea is that since the derivation of the formula depends only on domain variations, and

not on any differentiability of the target space, it continues to hold for maps to metric space targets.

The monotonicity (2.50) can be used to construct linear approximations to harmonic maps, and in some cases further regularity can be derived. A key quantity is the *order function*. Roughly speaking, the order of a harmonic map $f: \Omega \rightarrow X$ at a point x measures the degree of the dominant homogeneous harmonic polynomial which approximates $f - f(x)$. This is precisely true when X is a smooth manifold. In the general case, it is defined as follows. Define

$$\text{ord}_x(f) = \lim_{\varepsilon \downarrow 0} \frac{\varepsilon \int_{B_\varepsilon(x)} |\nabla f|^2 d \text{vol}_\Omega}{\int_{\partial B_\varepsilon(x)} d^2(f(x), f(y)) ds(y)}. \quad (2.51)$$

It follows from the monotonicity formula (2.50) that the above limit exists and is ≥ 1 for nonconstant maps. We call this limit the *order of f at x* . It is not an integer in general. For example, let $p: \mathbb{H} \rightarrow T_\varphi$ be the projection map (2.47). If x is not a zero of φ , then p is locally a harmonic function and $\text{ord}_x(p)$ is the order of vanishing. If x is a zero of order k , then $\text{ord}_x(p) = (k + 2)/2$. The order is related to the eigenvalues of the Laplacian on subdomains of $\partial B_r(x)$, as explained in [75, Theorem 5.5]. In Figure 3, $\text{ord}_x(p)$ is equal to the first Dirichlet eigenvalue of the domain D_i in the circle around x . It is clear in this case that it is equal to $3/2$.

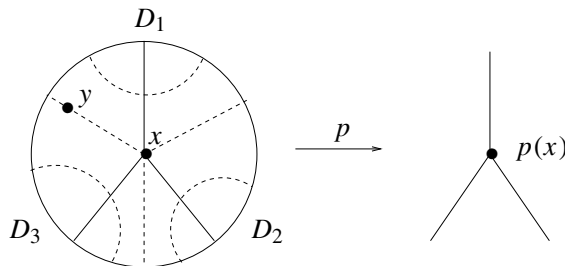


Figure 3

On the other hand, if $y \in p^{-1}(p(x))$ is not a zero, then $\text{ord}_y(p) = 1$, and indeed locally near y , p maps to an interval.

This can be generalized. Let $f: \Omega \rightarrow T$ be a harmonic map to an \mathbb{R} -tree. A point $x \in \Omega$ is called *regular* if there exists $r > 0$ such that $f(B_r(x))$ is an embedded arc. In particular, f restricted to $B_r(x)$ is then a harmonic function. Nonregular points are called *singular*. In the case of two-dimensional domains, the harmonic map $p: \mathbb{H} \rightarrow T_\varphi$ has singularities precisely at the zeros of φ . In particular, they are of codimension 2. The next result was proven in [75] for simplicial trees and in [180] for \mathbb{R} -trees.

Theorem 2.34. *Let $f: \Omega \rightarrow T$ be a harmonic map to an \mathbb{R} -tree. Then $x \in \Omega$ is regular if $\text{ord}_x(f) = 1$. Moreover, the Hausdorff codimension of the singular set is at least 2.*

X. Sun also proved the following useful fact.

Theorem 2.35 (Sun [180]). *Let $f: \Omega \rightarrow T$ be a harmonic map to an \mathbb{R} -tree. Then for any point $x \in \Omega$ there is $r > 0$ such that $f(B_r(x))$ lies in a locally finite subtree.*

3 Harmonic maps and representations

3.1 Equivariant harmonic maps

In this section we describe the equivariant harmonic map problem and its applications. In Section 3.1.1 we introduce the notion of reductivity (or semisimplicity) in different contexts and indicate how it is related to the existence problem for equivariant harmonic maps. In Section 3.1.2 we discuss the holomorphicity of the Hopf differential for harmonic maps and show how it can be used to simplify the proofs of the Hubbard–Masur and Skora theorems. We also give the first variation harmonic maps with respect to the domain metric and apply this to derive Gardiner’s formula.

3.1.1 Reductive representations. Throughout this section, unless otherwise noted, (M, g) is a closed Riemannian manifold with $\Gamma = \pi_1(M)$, and (X, d) is a simply connected NPC space. Let \tilde{M} denote the universal cover of M . We assume that Γ acts on X via isometries, i.e. that there is a homomorphism $\rho: \Gamma \rightarrow \text{Iso}(X)$. Associated to ρ is a *translation length function*

$$\mathfrak{L}_\rho: \Gamma \longrightarrow \mathbb{R}^+, \quad \gamma \mapsto \inf_{x \in X} d(x, \rho(\gamma)x). \quad (3.1)$$

Let $f: \tilde{M} \rightarrow X$ be a ρ -equivariant map. Provided that f is a locally in H^1 , the energy density $|\nabla f|^2$ is Γ -invariant, and therefore we can define the energy by

$$E(f) = \frac{1}{2} \int_{M=\tilde{M}/\Gamma} |\nabla f|^2 d \text{vol}_M. \quad (3.2)$$

Finite energy maps always exist, and indeed energy minimizing sequences can be taken to be uniformly Lipschitz [110]. Under conditions that will be made precise below and which we shall always assume, there exist maps with finite energy. A ρ -equivariant map $f: \tilde{M} \rightarrow X$ which is locally in H^1 is called *harmonic* if it minimizes the energy (3.2) among all other equivariant maps in H_{loc}^1 .

It follows from the trace theory in [109] that equivariant harmonic maps are locally energy minimizers. Therefore, in the case where X is a smooth manifold the first variational formula (2.23) implies that a ρ -equivariant harmonic map is equivalent to

a smooth ρ -equivariant map that satisfies the harmonic map equations (2.22). For general NPC targets it follows from Theorem 2.31 that ρ -equivariant harmonic maps are Lipschitz.

The existence of equivariant harmonic maps is more complicated than in the case of compact targets. The reason for this is that in the process of choosing an energy minimizing sequence, e.g. using the heat flow as in the Eells–Sampson theory, the map can “escape to infinity”, and fail to converge. An example of this phenomenon can be found in [51]. One naturally looks for a condition on the homomorphism ρ which rules out this kind of behavior. For example, it is reasonable to rule out the existence of a sequence of points escaping to infinity whose translates by fixed elements in the image of ρ remain bounded. This is the notion of a *proper* action (see below).

Before making this more precise we introduce the notion of the *ideal boundary* of an NPC space. By a *ray* in X we mean a geodesic α parametrized by arc length on the interval $[0, \infty)$. Two rays α_1, α_2 , are said to be equivalent if the Hausdorff distance between them is finite. Denote by ∂X the set of equivalence classes of rays. Notice that since Γ acts by isometries, Γ also acts on ∂X . We have the following facts:

- (1) (cf. [19]) If X is locally compact then $\bar{X} = X \cup \partial X$ can be topologized so that it becomes a compact metric space.
- (2) (cf. [32]) If (X, d) is an \mathbb{R} -tree (not necessarily locally compact) then two rays α_1 and α_2 are equivalent if and only if $\alpha_1 \cap \alpha_2$ is another ray.
- (3) (cf. [32]) If (X, d) is an \mathbb{R} -tree with Γ action, then Γ fixes a point on ∂X if and only if $\mathcal{L}_\rho(\gamma) = |r(\gamma)|$ where $r: \Gamma \rightarrow \mathbb{R}$ is a homomorphism.

We now state

Theorem 3.1 (Korevaar–Schoen [110]). *Suppose $\rho: \Gamma \rightarrow \text{Iso}(X)$ is a homomorphism that does not fix a point of ∂X . If either (i) X is locally compact, or (ii) X has curvature bounded above by $\kappa < 0$, then there exists a ρ -equivariant harmonic map $f: \tilde{M} \rightarrow X$.*

The equivariant version of Theorem 2.33 also holds:

Theorem 3.2 (Mese [134]). *If X has curvature bounded above by a constant $\kappa < 0$, and if $f: \tilde{M} \rightarrow X$ is a nonconstant equivariant harmonic map, then f is unique in its equivariant homotopy class unless it maps onto a geodesic.*

Special cases of Theorem 3.1 had been proven earlier:

- *The Corlette–Donaldson theorem.*

Theorem 3.3 (cf. [29], [30], [43]). *Let X be a Riemannian symmetric space of non-compact type $X = G/K$, where G is a semisimple Lie group and K a maximal compact subgroup. Let $\rho: \Gamma \rightarrow G$ be a homomorphism with Zariski dense image. Then there is a ρ -equivariant (smooth) harmonic map $f: \tilde{M} \rightarrow X$.*

This theorem is implied by Theorem 3.1, since if $\rho(\Gamma)$ fixes a point $[\alpha] \in \partial X$, then $\overline{\rho(\Gamma)}$ would be closed subgroup contained in the stabilizer of $[\alpha]$, which is a proper subgroup of G . See also [99].

• *Labourie's theorem.* In the Riemannian case, the criterion for existence in terms of fixing a point in the ideal boundary was conjectured in [29] and proved in [114] (see also [97]). A homomorphism $\rho: \Gamma \rightarrow \text{Iso}(X)$ is called *semisimple* (or *reductive*) if either $\rho(\Gamma)$ does not fix a point in ∂X or it fixes a geodesic. Then we have the following

Theorem 3.4 (Labourie [114]). *Let X be a Riemannian manifold with negative sectional curvature. Then there exists a ρ -equivariant harmonic map $f: \tilde{M} \rightarrow X$ if and only if ρ is semisimple.*

• *\mathbb{R} -trees.* Let (T, d) be an \mathbb{R} -tree and $\rho: \Gamma \rightarrow \text{Iso}(T)$ a homomorphism. We assume (without loss of generality) that the action of Γ on T is minimal.

Theorem 3.5 (Culler–Morgan [32]). *Let ρ_1, ρ_2 be nontrivial semisimple actions on \mathbb{R} -trees T_1, T_2 with the same translation length functions. Then there exists an equivariant isometry $T_1 \simeq T_2$. If either action is not isometric to an action on \mathbb{R} , then the equivariant isometry is unique.*

Then we have the following generalization of Theorem 3.4 to trees.

Theorem 3.6. *Let (T, d) be a minimal \mathbb{R} -tree and $\rho: \Gamma \rightarrow \text{Iso}(T)$. Then there exists a ρ -equivariant harmonic map $u: \tilde{M} \rightarrow T$ if and only if ρ is semisimple.*

Proof. The sufficiency follows from Theorem 3.1 (see Section 2.3.2). For the converse, suppose Γ fixes a point in ∂T . If there is a ρ -equivariant harmonic map there would necessarily be a family of distinct such maps (see [34]). By the uniqueness Theorem 3.2 and the minimality of T , it follows that T in this case is equivariantly isometric to \mathbb{R} . \square

In the case where X is not locally compact, the condition of not fixing a point at infinity does not seem to be sufficient to guarantee existence. Korevaar and Schoen developed a slightly stronger condition to cover this case. Let $\rho: \Gamma \rightarrow \text{Iso}(X)$ be a homomorphism. To each set of generators \mathcal{G} of Γ we associate a function on X :

$$D_\rho(x) = \max \{d(x, \rho(\gamma)x) : \gamma \in \mathcal{G}\}.$$

A homomorphism $\rho: \Gamma \rightarrow \text{Iso}(X)$ is called *proper* if for every $B \geq 0$, the set $\{x \in X : D_\rho(x) \leq B\}$ is bounded. Clearly, this condition is independent of the choice of generating set \mathcal{G} . For complete manifolds of nonpositive curvature, the existence of two hyperbolic isometries in the image of ρ with nonasymptotic axes is sufficient to prove properness. More generally, ρ being proper implies that ρ has no fixed end, for if R is a fixed ray then D_ρ is bounded along R .

Theorem 3.7 (Korevaar–Schoen [110]). *Suppose $\rho: \Gamma \rightarrow \text{Iso}(X)$ is proper. Then there exists a ρ -equivariant harmonic map $f: \tilde{M} \rightarrow X$.*

In case X is locally compact this is implied by Theorem 3.1, but for nonlocally compact spaces it is not. Yet another sufficient condition is introduced in [111].

To end this section, we connect the definition of harmonicity given in this section with that at the end of Section 2.3.2.

Theorem 3.8. *Let S be a Riemann surface, (T, d) an \mathbb{R} -tree, and $\rho: \Gamma = \pi_1(S) \rightarrow \text{Iso}(T)$ a reductive action. A ρ -equivariant map $f: \tilde{M} \rightarrow T$ is harmonic if and only if f pulls back germs of convex functions to germs of subharmonic functions.*

Proof. The fact that harmonic maps pull back functions to subharmonic ones is the content of [56, Prop. 3.2] (see also [109]). For the converse, we argue as follows: suppose $f: \mathbb{H} \rightarrow T$ is a ρ -equivariant map that pulls back germs of convex functions to subharmonic ones. Let $f': \mathbb{H} \rightarrow T$ be a ρ -equivariant harmonic map. Since both f, f' pull back germs of convex functions to subharmonic functions, it follows that the same is true for $f \times f': \mathbb{H} \rightarrow T \times T$. Hence, $d(f, f')$ is Γ -equivariant and subharmonic, hence constant. But because of the 1-dimensionality of trees it is easy to see that the energy densities of f and f' must be equal, so that f is energy minimizing. \square

So far as we know, this result for general NPC targets is open.

3.1.2 Measured foliations and Hopf differentials. Recall from Section 2.3.3 that if X is a metric space target and $f: (M, g) \rightarrow (X, d)$ is a finite energy map, then one can associate an integrable symmetric 2-tensor $\pi_{\alpha\beta}$ on S with the property that the energy density $|\nabla f|^2 = g^{\alpha\beta}\pi_{\alpha\beta}$. Hence, while the energy density may not be the square of the norm of a derivative, it is a trace of directional energies. Let us specialize to the case where the domain is a Riemann surface, and let f be an energy minimizer. By varying among finite energy maps obtained from pulling f back by a local diffeomorphism defined by a vector field v , we arrive at

$$0 = \int_M \langle \pi, L_v g - (1/2) \text{Tr}_g(L_v g) \rangle_g d \text{vol}_M.$$

Note that the integrand is well defined since π is integrable. By a particular choice of v , and using Weyl's lemma on integrable weakly holomorphic functions, we obtain [167] that

$$\varphi(z) dz^2 = \frac{1}{4}(\pi_{11} - \pi_{22} - 2i\pi_{12}) dz^2$$

is a holomorphic quadratic differential on S (cf. Section 2.2.2). We call φ the *Hopf differential* of f . Since these computations are local, they apply as well to the case of equivariant harmonic maps.

Thus far we have seen that a measured foliation \mathcal{F} on a surface S gives rise to an \mathbb{R} -tree $T_{\mathcal{F}}$ with an isometric action of Γ . This action has the following properties:

- (1) the action is *minimal* in the sense that no proper subtree is invariant under Γ (cf. [153] – strictly speaking, the proof there uses geodesic laminations but it can be easily adapted to the case of measured foliations);
- (2) the action is *small* in the sense that the edge stabilizer subgroups do not contain free groups on 2-generators (cf. [140] – more precisely, the stabilizers are cyclic, since leaves on the quotient surface are either lines or circles).

Shalen conjectured [173] that every minimal, small action of a surface group on an \mathbb{R} -tree is dual to a measured foliation. This conjecture, which plays an important role in Thurston’s hyperbolization theorem for fibered 3-manifolds (see [153]), was proved by R. Skora, building upon previous work of Morgan–Otal [140].

Theorem 3.9 (Skora [178]). *Let S be a surface of genus at least 2. Then if (T, d) is an \mathbb{R} -tree with a minimal, small isometric action of $\Gamma = \pi_1(S)$, there is a measured foliation \mathcal{F} on S such that (T, d) is equivariantly isometric to $T_{\mathcal{F}}$.*

For example, we have seen that if (T, d) is dual to a measured foliation on S then the action is small. It is also a simple matter to see that a small action is semisimple. Indeed, choose $\gamma_i \in \Gamma$, $i = 1, \dots, 4$ such that the commutators $[\gamma_1, \gamma_2]$ and $[\gamma_3, \gamma_4]$ generate a group G containing a free group on 2-generators. Then if ρ had a fixed end, then $\rho(\gamma_i)$ would act by translations along a common ray. In particular, $\rho(G)$ would stabilize this ray, contradicting the assumption of smallness. Hence, small actions are semisimple, and by Theorem 3.6 there exists a ρ -equivariant harmonic map $f: \mathbb{H} \rightarrow T$.

In general, let $\tilde{\varphi}$ be the Hopf differential of an equivariant harmonic map. Then $\tilde{\varphi}$ is the lift of a holomorphic quadratic differential φ on S . Let T_φ denote the dual tree to the vertical foliation of $\tilde{\varphi}$. It is not hard to see (cf. [35], [56]) that there is a Γ -equivariant map $F: T_\varphi \rightarrow T$ such that the following diagram commutes

$$\begin{array}{ccc}
 \mathbb{H}_f & \xrightarrow{p} & T_\varphi \\
 & \searrow & \downarrow F \\
 & & T
 \end{array}
 \tag{3.3}$$

where $p: \mathbb{H} \rightarrow T_\varphi$ is the natural projection. Moreover, this is a *morphism of trees*, meaning that any segment $xy \in T_\varphi$ decomposes into a finite union of subsegments along which p is an isometry. By [140], it follows that F is either an equivariant isometry, or F *folds* at some point. This means there is an identification of two or more segments $z'y'$ and $z'y''$ in T_φ to a single segment zy in T . An example of a folding is shown in Figure 4.

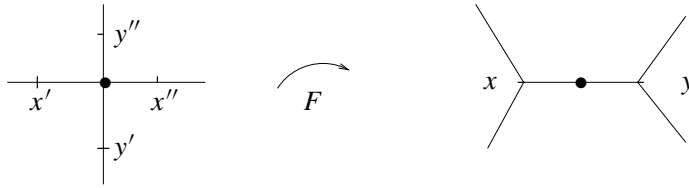


Figure 4

Interestingly, the harmonicity of the map f precludes a whole class of undesirable foldings. For example, the following is a consequence of the maximum principle.

Proposition 3.10 ([35], [56]). *Suppose $T_\varphi \rightarrow T$ arises from an equivariant harmonic map to T , as in (3.3). Then folding occurs only at points in T_φ corresponding to zeros of $\tilde{\varphi}$ of multiplicity at least two. Moreover, adjacent edges may not be identified under such a folding.*

This type of resolution of the tree T by the dual tree T_φ to a measured foliation, with the folding properties of the proposition, had been obtained by Morgan–Otal in [140], and it is the first step in proving Skora’s Theorem 3.9.

By an ingenious counting argument using interval exchanges, Skora went on to show that provided the action of Γ is small, folding at vertices cannot occur either, and in fact F is an isometry. This completes the proof of Theorems 2.30 and 3.9. An alternative source for the counting argument is [153, §8.4]. The reader may also consult [56], [198], [199].

As a second application, consider a measured foliation \mathcal{F} on a Riemann surface (S, σ) . We have seen above that there is a unique holomorphic quadratic differential $\varphi_{\mathcal{F}} = \varphi(\sigma, \mathcal{F})$ whose horizontal foliation is measure equivalent to \mathcal{F} . The *extremal length* of \mathcal{F} is defined by

$$\text{Ext}_{\mathcal{F}}[\sigma] = \int_S |\varphi_{\mathcal{F}}| \quad (3.4)$$

and is a well-defined function on $\mathcal{T}(S)$. It is a generalization of the extremal length of a simple closed curve to the case of arbitrary measured foliations. In [64], Gardiner gave a formula for the first variation of $\text{Ext}_{\mathcal{F}}[\sigma]$. Here we show how this formula arises naturally as the variation of the energy of harmonic maps.

Let X be an NPC space, and suppose $\rho: \Gamma \rightarrow \text{Iso}(X)$ is proper in the sense of Section 3.1.1. For simplicity, assume X has curvature bounded above by some $\kappa < 0$. Then for each complex structure j on S , Theorems 3.7 and 3.2 guarantee the existence of a unique ρ -equivariant harmonic map $f: \tilde{S} \rightarrow X$. The energy of $f = f_{[j], \rho}$ gives a well-defined function depending upon $[j]$ and ρ :

$$\mathcal{E}_\rho^-: \mathcal{T}(S) \rightarrow \mathbb{R}^+, \quad [j] \mapsto E(f_{[j], \rho}). \quad (3.5)$$

Theorem 3.11. *The function \mathfrak{E}_ρ^- is differentiable on $\mathcal{T}(S)$. If σ_t , $-1 \leq t \leq 1$, is a differentiable family of metrics on S with Beltrami differential μ at $t = 0$, and φ is the Hopf differential of a ρ -equivariant energy minimizer $(\tilde{S}, \sigma_0) \rightarrow X$, then*

$$\left. \frac{d}{dt} \right|_{t=0} \mathfrak{E}_\rho^-[\sigma_t] = -4 \operatorname{Re}\langle \mu, \varphi \rangle, \quad (3.6)$$

where the pairing is as in (2.16).

In the case where X is a smooth Riemannian manifold, this formula has some history. Wolf [200] provides a derivation and refers to earlier notes of Schoen, as well as [185], [186], [95]. The earliest computation of this sort may be due to Douglas (cf. [45, eq. (12.29)]). Formally, the proof of (3.6) goes as follows. The total energy is the contraction of the energy density tensor π_{ij} with the metric on S . Hence, the first variation involves varying first π_{ij} , i.e. the harmonic map, and then the metric. But the term associated to the variation of the map is necessarily zero, since the map is energy minimizing. It follows that the only contribution comes from variations with respect to the metric. Formula (3.6) then follows easily. Some care must be taken to justify this in the case of metric space targets (see [194]).

Now consider a measured foliation \mathcal{F} on S with associated dual tree $T_{\mathcal{F}}$. The energy of the unique equivariant harmonic map $f: \tilde{S} \rightarrow T_{\mathcal{F}}$ is precisely the extremal length $\operatorname{Ext}_{\mathcal{F}}$. From (3.6) we have

Theorem 3.12 (Gardiner [64], [65]). *For any measured foliation \mathcal{F} , $\operatorname{Ext}_{\mathcal{F}}$ is differentiable on $\mathcal{T}(S)$ with derivative*

$$\left. \frac{d}{dt} \right|_{t=0} \operatorname{Ext}_{\mathcal{F}}[\sigma_t] = 2 \operatorname{Re}\langle \mu, \varphi_{\mathcal{F}} \rangle.$$

Here, $\varphi_{\mathcal{F}}$ is the Hubbard–Masur differential for \mathcal{F} at σ_0 .

3.2 Higgs bundles and character varieties

This section discusses the relationship between character varieties and certain special metrics on holomorphic vector bundles. The link between these two comes via the equivariant harmonic map problem of the previous section. In Section 3.2.1 we introduce the notion of a Higgs bundle and discuss the correspondence between stable Higgs bundles, the self-duality equations, and flat $\operatorname{SL}(2, \mathbb{C})$ connections. In Section 3.2.2 we give a Higgs bundle interpretation of the Teichmüller space and another proof of Theorem 2.8 using the self-duality equations. Finally, in Section 3.2.3, we discuss the notion of convergence in the pullback sense and give a harmonic maps interpretation of the Morgan–Shalen–Thurston compactification of character varieties.

3.2.1 Stability and the Hitchin–Simpson theorem. By a *Higgs bundle* on a Riemann surface S we mean a pair (V, Φ) , where $V \rightarrow S$ is a holomorphic vector bundle

and Φ is a holomorphic section of the associated bundle $\text{End}(V) \otimes K_S$. Two Higgs bundles (V, Φ) , (V', Φ') are isomorphic if there exists an isomorphism $\iota: V \rightarrow V'$ of holomorphic structures such $\Phi' \circ \iota = \iota \circ \Phi$.

Recall that a complex bundle has a well-defined *degree*,

$$\deg(V) = \int_S c_1(V),$$

where $c_1(V)$ denotes the first Chern class. The *slope* is defined by

$$\text{slope}(V) = \deg(V) / \text{rk}(V),$$

where $\text{rk}(V)$ is the rank of V . A Higgs bundle (V, Φ) is called *stable* if $\text{slope}(V') < \text{slope}(V)$ for all nontrivial Φ -invariant proper subbundles $V' \subset V$, i.e. $V' \neq 0$, V and $\Phi(V') \subset V' \otimes K_S$. A Higgs bundle is called *polystable* if can be written as a direct sum of stable Higgs bundles.

Given a hermitian metric H on a holomorphic bundle V we shall denote the Chern connection by ∇_H , i.e. ∇_H is the unique connection compatible with H and the holomorphic structure (cf. [26]). The curvature F_{∇_H} takes values in $\text{ad}(V) \otimes \Omega^2(S)$, where $\text{ad}(V) \subset \text{End}(V)$ is the bundle of skew-hermitian endomorphisms. Let ω be a Kähler form on S normalized so that $\int_S \omega = 1$. The following result is due to Hitchin, who first introduced Higgs bundles in this form [81], [83]. The result for higher dimensional Kähler manifolds is due to Simpson [174]. The case $\Phi \equiv 0$ corresponds to stable bundles on Riemann surfaces and was proved first by Narasimhan–Seshadri [148] and later, using very different methods, by Donaldson [42]. Higher dimensional versions of the Narasimhan–Seshadri theorem were obtained by Donaldson [43] and Uhlenbeck–Yau [189].

Theorem 3.13 (Hitchin, Simpson). *Let (V, Φ) be a Higgs bundle on a closed Riemann surface S . Then (V, Φ) is polystable if and only if there exists a hermitian metric H on V solving the self-duality equations*

$$\frac{i}{2\pi} F_{\nabla_H} + [\Phi, \Phi^{*H}] = s \mathbf{I} \otimes \omega, \quad (3.7)$$

where Φ^{*H} is the adjoint of Φ with respect to H , and $s = \text{slope}(V)$. Furthermore, H is unique up to scalars.

From both the algebro-geometric and topological points of view, it is preferable to fix determinants. In other words, fix a holomorphic line bundle $L \rightarrow S$ with hermitian metric h such that $\det(V) = L$, and let $\Phi \in H^0(\text{End}_0 V \otimes K_S)$, where $\text{End}_0(V)$ is the bundle of traceless endomorphisms. We shall call (V, Φ) a Higgs bundle of fixed determinant L .

Corollary 3.14. *A Higgs bundle (V, Φ) of fixed determinant L is polystable if and only if there exists a hermitian metric H on V with $\det H = h$ and such that (3.7) holds. In the case such an H exists, it is unique.*

Following Corlette [29] we call a flat $\mathrm{SL}(r, \mathbb{C})$ connection ∇ on the trivial rank r bundle on S *reductive* if any ∇ -invariant subbundle has a ∇ -invariant complement. Clearly, a reductive flat $\mathrm{SL}(r, \mathbb{C})$ connection is a direct sum of irreducible flat $\mathrm{SL}(r', \mathbb{C})$ connections for values $r' < r$.

Define $\mathfrak{M}(S, r)$ to be the moduli space of isomorphism classes of polystable Higgs bundles on S of rank r and fixed trivial determinant. We denote the space of equivalence classes of reductive flat $\mathrm{SL}(r, \mathbb{C})$ connections on the trivial rank r bundle $V \rightarrow S$ by $\chi(\Gamma, r)$. We have the following

Theorem 3.15 (Corlette, Donaldson). *The map*

$$\begin{aligned} \Psi: \mathfrak{M}(S, r) &\rightarrow \chi(\Gamma, r) \\ (V, \Phi) &\mapsto \nabla_H + \Phi + \Phi^{*H}, \end{aligned}$$

is a bijection, where H satisfies (3.7).

That Ψ is well defined follows from Corollary 3.14, and the injectivity is a consequence of the uniqueness of the solution H . The surjectivity part was first conjectured by Hitchin in [81] and was subsequently proven for rank 2 by Donaldson [44] and in general by Corlette [29]. It is equivalent to the Corlette–Donaldson Theorem 3.3 on equivariant harmonic maps discussed above.

Indeed, given a reductive flat connection ∇ , let $\rho: \Gamma \rightarrow \mathrm{SL}(r, \mathbb{C})$ denote its holonomy representation. Since reductive representations split into irreducible factors, we may assume without loss of generality that ρ is irreducible. By Theorem 3.3 there exists an equivariant harmonic map $H: \mathbb{H} \rightarrow \mathrm{SL}(r, \mathbb{C})/\mathrm{SU}(r)$. Equivalently, we can view H as a section of the “twisted bundle”

$$\mathbb{H} \times_{\rho} \mathrm{SL}(r, \mathbb{C})/\mathrm{SU}(r) \longrightarrow S,$$

i.e. H is nothing but the choice of a hermitian metric on V . Therefore, we can split $\nabla = \nabla_H + \Phi + \Phi^{*H}$, where ∇_H is a hermitian connection with respect to H , and Φ is a smooth section of $\mathrm{End}_0(V) \otimes K_S$. Clearly, the flatness of ∇ is equivalent to the equations (3.7) together with the Bianchi identity

$$d_{\nabla_H}(\Phi + \Phi^{*H}) = 0. \quad (3.8)$$

The *harmonicity* of H is equivalent to the condition [29], [44]

$$(d_{\nabla_H})^{*H}(\Phi + \Phi^{*H}) = 0. \quad (3.9)$$

Conditions (3.8) and (3.9) are together equivalent to the holomorphicity of Φ . Hence, V with the induced holomorphic structure from ∇_H and Φ define a Higgs bundle with $\Psi(V, \Phi) = \nabla$.

Notice that in the above argument we indicated that the existence of a ρ -equivariant harmonic map $\mathbb{H} \rightarrow \mathrm{SL}(r, \mathbb{C})/\mathrm{SU}(r)$ was equivalent to the reductivity of the flat connection ∇ . Therefore, by Labourie’s Theorem 3.4, it is also equivalent to the reductivity of the holonomy representation $\rho: \Gamma \rightarrow \mathrm{SL}(2, \mathbb{C})$, in the case of $\mathbb{H}^3 = \mathrm{SL}(2, \mathbb{C})/\mathrm{SU}(2)$.

The question of the complex structure on the spaces $\mathfrak{M}(S, r) \simeq \chi(\Gamma, r)$, originally addressed by Hitchin [81], is an extremely interesting one. As a *character variety*, $\chi(\Gamma, r)$ is an affine algebraic variety. For example, given any $\gamma \in \Gamma$ we define a regular function $\tau_\gamma: \chi(\Gamma, 2) \rightarrow \mathbb{C}$ by $\tau_\gamma[\rho] = \text{Tr } \rho(\gamma)$. Here, $[\rho]$ denotes the conjugacy class of representations containing ρ . By [33] the ring generated by all elements $\tau_\gamma, \gamma \in \Gamma$, is finitely generated. Fix a generating set associated to $\{\gamma_1, \dots, \gamma_m\}$, and define

$$t: \chi(\Gamma, 2) \longrightarrow \mathbb{C}^m, \quad [\rho] \mapsto (\tau_{\gamma_1}(\rho), \dots, \tau_{\gamma_m}(\rho)).$$

Then t is a bijection onto its image and gives $\chi(\Gamma, 2)$ the structure of an affine variety. For higher rank, one needs to consider other invariant polynomials in addition to traces. On the other hand, Nitsure and Simpson have shown that $\mathfrak{M}(S, r)$ with its complex structure induced as a moduli space over the Riemann surface S has the structure of a quasiprojective algebraic scheme [151], [175], [176]. The bijection $\mathfrak{M}(S, r) \simeq \chi(\Gamma, r)$ is *not* complex analytic. On the contrary, Hitchin shows that the two complex structures are part of a hyperkähler family. For more details, we refer to [81], [175].

A consequence of the realization of $\chi(\Gamma, r)$ as a moduli space of Higgs bundles is that there is a natural \mathbb{C}^* -action. Indeed, if (V, Φ) is a polystable Higgs bundle then so is $(V, t\Phi), t \in \mathbb{C}^*$. This defines a holomorphic action on $\mathfrak{M}(S, r)$, and therefore also an action (not holomorphic) on $\chi(\Gamma, r)$. This action depends on the complex structure on S and is not apparent from the point of view of representations. Nevertheless, we shall see in the next section that it has some connection with Teichmüller theory.

3.2.2 Higgs bundle proof of Teichmüller's theorem. For the purposes of this section we specialize to the case $r = 2$ and set $\mathfrak{M}(S) = \mathfrak{M}(S, 2)$ and $\chi(\Gamma) = \chi(\Gamma, 2)$. Define the *Hitchin map*

$$\det: \mathfrak{M}(S) \longrightarrow \text{QD}(S), \quad [V, \Phi] \mapsto \det \Phi = -\frac{1}{2} \text{Tr } \Phi^2. \quad (3.10)$$

Hitchin proved that \det is a proper, surjective map with generic fibers being half-dimensional tori. This last property in fact realizes $\mathfrak{M}(S)$ as a completely integrable system (see [82]). More importantly for us, notice that under the Corlette–Donaldson correspondence $\Psi: \mathfrak{M}(S) \rightarrow \chi(\Gamma)$, $\det \circ \Psi^{-1}$ is just the Hopf differential of the associated harmonic map (cf. [35]). Indeed, for $[\rho] \in \chi(\Gamma)$ with an associated equivariant harmonic map $f_\rho, \varphi_\rho = \text{Hopf}(f_\rho)$ is given by

$$\varphi_\rho = \langle \nabla f_\rho^{1,0}, \nabla f_\rho^{1,0} \rangle = -\text{Tr}(\nabla f_\rho^{1,0})^2 = 2 \det \circ \Psi^{-1}[\rho].$$

In order to realize the Teichmüller space inside $\mathfrak{M}(S)$, let $\iota: \mathfrak{M}(S) \rightarrow \mathfrak{M}(S)$ denote the involution $\iota(V, \Phi) = (V, -\Phi)$. Notice that ι is a restriction of the full \mathbb{C}^* -action on $\mathfrak{M}(S)$ described at the end of the previous section. Also notice that under the Corlette–Donaldson correspondence Ψ , ι corresponds to complex conjugation. Hence, the fixed points of ι are either $\text{SU}(2)$ or $\text{SL}(2, \mathbb{R})$ representations. The former correspond under the Narasimhan–Seshadri theorem to the Higgs pair $(V, 0)$, i.e. $\Phi \equiv 0$. If (V, Φ) is a

fixed point of ι with $\Phi \neq 0$, Hitchin shows that V must be a split holomorphic bundle $L \oplus L^*$, and with respect to this splitting Φ is of the form

$$\Phi = \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix},$$

where $a \in H^0(S, L^2 \otimes K_S)$, and $b \in H^0(S, L^{-2} \otimes K_S)$. Stability implies $b \neq 0$, and hence by vanishing of cohomology, $\deg L \leq p - 1$, where p is the genus of S . This fact, as pointed out in [81], turns out to be equivalent to the Milnor–Wood inequality which states that the Euler class of any $\mathrm{PSL}(2, \mathbb{R})$ bundle on S is $\leq 2p - 2$ (cf. [136], [208]).

We next restrict ourselves to the components of the fixed point set of ι corresponding to line bundles L of maximal degree $p - 1$. In this case, L must be a spin structure, i.e. $L^2 = K_S$, for otherwise $b = 0$, contradicting stability. We denote this moduli space by $\mathcal{N}_L(S)$. After normalizing by automorphisms of $L \oplus L^{-1}$, we can write

$$\Phi = \begin{pmatrix} 0 & a \\ 1 & 0 \end{pmatrix},$$

for some quadratic differential $a \in \mathrm{QD}(S)$. It follows that the restriction of the Hitchin map to $\mathcal{N}_L(S)$ defines a homeomorphism $\det: \mathcal{N}_L(S) \xrightarrow{\sim} \mathrm{QD}(S)$.

The following gives another proof of Theorem 2.8.

Theorem 3.16 (Hitchin). *Given a Higgs bundle*

$$\left(L \oplus L^{-1}, \begin{pmatrix} 0 & a \\ 1 & 0 \end{pmatrix} \right)$$

in $\mathcal{N}_L(S)$, let H denote the metric on $L \oplus L^{-1}$ solving the self-duality equations, and let h be the induced metric on $K_S^{-1} = L^{-2} = T^{1,0}S$. Then

(1) *the tensor*

$$\hat{h} = a + (h + h^{-1}a\bar{a}) + \bar{a} \in \Omega^0(S, \mathrm{Sym}^2(T^*S) \otimes \mathbb{C})$$

is a Riemannian metric on S of constant curvature -4 ;

(2) *any metric of constant curvature -4 on S is isometric to one of this form for some $a \in \mathrm{QD}(S)$.*

The new ingredient in this theorem is the use of the existence of solutions to the self-duality equations (3.7). Notice that in the reducible case described in Theorem 3.13 the self-duality equations reduce to the *abelian vortex equations* $F_h = -2(1 - \|a\|_{L^2}^2)\omega$ (cf. [92]). The relation between the vortex equations and curvature of metrics on surfaces had been noted previously in the work of Kazdan and Warner [103].

Notice that the definition of $\mathcal{N}_L(S)$ depends on a choice of spin structure L , and there are $\#H^1(S, \mathbb{Z}_2) = 2^{2p}$ such choices. This reflects the fact that on $\chi(\Gamma)$ there is an action of \mathbb{Z}_2^{2p} , and the quotient is

$$\chi(\Gamma)/\mathbb{Z}_2^{2p} = \mathrm{Hom}(\Gamma, \mathrm{PSL}(2, \mathbb{C}))//\mathrm{PSL}(2, \mathbb{C}),$$

the character variety of $\mathrm{PSL}(2, \mathbb{C})$, of which the Fricke space $\mathfrak{F}(S)$ is a natural subset. The preimage of $\mathfrak{F}(S)$ in $\chi(\Gamma)$ is the disjoint union of the $\mathcal{N}_L(S)$, and each of these is homeomorphic to Teichmüller space.

3.2.3 The Thurston–Morgan–Shalen compactification. Let us first explain the notion of convergence in the pullback sense, due to Korevaar–Schoen, that appears in the statement of Theorem 2.26. Let Ω be a set and $f: \Omega \rightarrow X$ a map into a simply connected NPC space (X, d) . Use f to define a pseudometric on Ω , $d_f(x, y) = d(f(x), f(y))$, $x, y \in \Omega$. To obtain convergence in an NPC setting, some convexity is needed. This is achieved by enlarging Ω to a space Ω_∞ , defined recursively by:

$$\begin{aligned}\Omega_0 &= \Omega, \\ \Omega_{k+1} &= \Omega_k \times \Omega_k \times [0, 1], \\ \Omega_\infty &= \bigsqcup_{k=0}^{\infty} \Omega_k / \sim,\end{aligned}$$

where the identification \sim is generated by an inclusion $\Omega_k \hookrightarrow \Omega_{k+1}$, $x \mapsto (x, x, 0)$. The map f extends to Ω_∞ recursively by setting $f(x, y, t)$, where $x, y \in \Omega_{k+1}$, equal to the point on the geodesic t of the way from $f(x)$ to $f(y)$. Let d_∞ denote the pullback pseudometric on Ω_∞ . After identifying points of zero pseudodistance in $(\Omega_\infty, d_\infty)$ and completing, one obtains a metric space (Z, d_Z) isometric to the closed convex hull $C(f(\Omega)) \subset X$ (see [110]).

Given a sequence $f_i: \Omega \rightarrow X_i$ of maps into simply connected NPC spaces X_i , we say that $f_i \rightarrow f$ in the pullback sense if the pullback pseudodistances $d_{i,\infty}$ on Ω_∞ converge locally uniformly to a pseudometric d_∞ , and if the map f is the quotient $\Omega \hookrightarrow \Omega_\infty \rightarrow (Z, d_Z)$.

This notion is equivalent to Gromov–Hausdorff convergence (cf. [197]). Indeed, (uniform) Gromov–Hausdorff convergence $(Z_i, d_i) \rightarrow (Z, d_Z)$ means that for any $\varepsilon > 0$ there are relations $R_i \subset Z_i \times Z$ whose projections surject onto Z_i and Z , and such that if $(z_i, z), (z'_i, z') \in R_i$, then

$$|d_i(z_i, z'_i) - d_Z(z, z')| < \varepsilon.$$

Convergence of the maps $f_i: \Omega \rightarrow X_i$ to $f: \Omega \rightarrow Z$ imposes the additional requirement that $(f_i(x), f(x)) \in R_i$ for all $x \in \Omega$. It is easy to see that $f_i: \Omega \rightarrow X_i$ converges in the pullback sense if and only if the convex hulls $Z_i = C(f_i(\Omega))$ and the maps f_i converge in the Gromov–Hausdorff sense. Indeed, pulling everything back to Ω_∞ , the relations R_i can be taken to be the diagonal. We also point out that it is easy to extend these notions equivariantly in the presence of isometric group actions.

We have the following compactness property:

Proposition 3.17. *Let Ω be a metric space, and let $f_i: \Omega \rightarrow X_i$ be a sequence of maps into NPC spaces such there is a uniform modulus of continuity: i.e. for*

each $x \in \Omega$ there is a monotone function ω_x so that $\lim_{R \rightarrow 0} \omega_x(R) = 0$, and $\max_{y \in B(x, R)} d_{f_i}(x, y) \leq \omega_x(R)$. Then f_i converges (after passing to a subsequence) in the pullback sense to a map $f: \Omega \rightarrow Z$, where Z is an NPC space.

We call the NPC space (Z, d_Z) a *Korevaar–Schoen limit*. Strictly speaking, the target surfaces (R, h_i) in Theorem 2.26 are not simply connected and are not NPC. To deal with the former, consider equivariant convergence of the lifts to the universal covers as mentioned above. For the latter, one shows that under the assumption that curvature is bounded from above, geodesics are locally unique, so the construction of Ω_∞ above works at a local level.

As usual, we denote by Γ the fundamental group of a hyperbolic surface. Let $\chi(\Gamma)$ be the $\mathrm{SL}(2, \mathbb{C})$ -character variety of Γ . As we have seen, $\chi(\Gamma)$ is a noncompact algebraic variety. In this section we describe a construction, introduced by Thurston in the case of $\mathrm{SL}(2, \mathbb{R})$ representations, to compactify $\chi(\Gamma)$. It is important to note that this is not a compactification in an algebro-geometric sense, and indeed $\overline{\chi(\Gamma)}$ will not be an complex analytic space.

Let \mathcal{C} denote the set of conjugacy classes of Γ , and let

$$\mathbb{P}(\mathcal{C}) = \{[0, \infty)^{\mathcal{C}} \setminus \{0\}\} / \mathbb{R}^+,$$

where \mathbb{R}^+ acts by homotheties. Topologize $\mathbb{P}(\mathcal{C})$ with the product topology. We define a map

$$\vartheta: \chi(\Gamma) \longrightarrow \mathbb{P}(\mathcal{C}), \quad [\rho] \mapsto \left\{ \log(|\mathrm{Tr} \rho(\gamma)| + 2) \right\}_{\gamma \in \mathcal{C}}$$

The purpose of the “+2” in the formula is to truncate the logarithm so that it goes to infinity only when the trace goes to infinity. It is easy to see (cf. [28]) that $\log(|\mathrm{Tr} \rho(\gamma)| + 2)$ is asymptotic to $\mathcal{L}_\rho(\gamma)$, where \mathcal{L}_ρ is the translation length function of ρ acting on hyperbolic space \mathbb{H}^3 (see (3.1) and recall that $\mathrm{Iso}(\mathbb{H}^3) = \mathrm{PSL}(2, \mathbb{C})$). In case ρ is a discrete faithful $\mathrm{SL}(2, \mathbb{R})$ representation, hence defining an element of Teichmüller space, $\mathcal{L}_\rho(\gamma)$ is just the length of the closed geodesic in the hyperbolic surface $S = \mathbb{H} / \rho(\Gamma)$ in the free homotopy class of γ .

Next, recall from Section 3.2.1 that by definition of the affine variety structure on $\chi(\Gamma)$, coordinate functions are of the form τ_γ , where $\tau_\gamma(\rho) = \mathrm{Tr} \rho(\gamma)$. Hence, τ_γ , $\gamma \in \mathcal{C}$ generate the coordinate ring of $\chi(\Gamma)$ as a \mathbb{C} -algebra, and it follows that ϑ is a continuous injection. Define $\overline{\chi(\Gamma)}$ to be the closure of the image of ϑ as a subset of $\mathbb{P}(\mathcal{C})$. It follows, essentially from the finite generation of the coordinate ring of $\chi(\Gamma)$, that $\overline{\chi(\Gamma)}$ is compact (cf. [141]). We call $\overline{\chi(\Gamma)}$ the *Morgan–Shalen compactification of $\chi(\Gamma)$* , and set $\partial\chi(\Gamma) = \overline{\chi(\Gamma)} \setminus \chi(\Gamma)$ to be the set of ideal points. The really useful ingredient in this construction is that the ideal points are not arbitrary but are translation length functions for isometric actions of Γ on \mathbb{R} -trees. Another important property is the following: the group $\mathrm{Aut}(\Gamma)$ of automorphisms of Γ clearly acts continuously on $\chi(\Gamma)$ and this action admits a continuous extension to $\overline{\chi(\Gamma)}$ (this is essentially the action of the *mapping class group* to be discussed in Section 4.1.2 below).

Let $\mathcal{T}(S) \subset \chi(\Gamma)$ denote the Teichmüller space, viewed as a component of the discrete faithful representations. The closure $\overline{\mathcal{T}}^{\text{th}}(S)$ of $\mathcal{T}(S)$ in $\overline{\chi(\Gamma)}$ is called the *Thurston compactification of $\mathcal{T}(S)$* and $\partial\mathcal{T}^{\text{th}}(S) = \overline{\mathcal{T}}^{\text{th}}(S) \setminus \mathcal{T}(S)$ is called the *Thurston boundary of $\mathcal{T}(S)$* . The action of $\text{Aut}(\Gamma)$ extends continuously to $\mathcal{T}(S) \subset \overline{\mathcal{T}}^{\text{th}}(S)$, and indeed this was part of the motivation for Thurston's compactification.

In terms of a finite set $\gamma_1, \dots, \gamma_m \in \Gamma$, where $\{\tau_{\gamma_i}\}$ generate the coordinate ring of $\chi(\Gamma)$, we can rephrase the compactness of $\overline{\chi(\Gamma)}$ as follows. Given a sequence of representations $\rho_i: \Gamma \rightarrow \text{SL}(2, \mathbb{C})$, only one of the following can occur:

- (1) For some subsequence $\{i'\}$, all traces $\rho_{i'}(\gamma_j)$, $j = 1, \dots, m$, are bounded (in this case, we call the sequence $\rho_{i'}$ *bounded*). Then $[\rho_{i'}]$ converges (after possibly passing to a further subsequence) in $\chi(\Gamma)$.
- (2) For any subsequence $\{i'\}$ there is some $s = 1, \dots, m$ such that $\text{Tr } \rho_{i'}(\gamma_s) \rightarrow \infty$ as $i' \rightarrow \infty$. Then there is a function $\ell: \mathbb{C} \rightarrow \mathbb{R}^+$, $\ell \neq 0$, such that (after possibly passing to a further subsequence) $\ell_{\rho_{i'}} \rightarrow \ell$, projectively.

In terms of the relationship between representations and equivariant harmonic maps we have the following simple but important observation (cf. [34]):

Proposition 3.18. *A sequence of representations $\rho_i: \Gamma \rightarrow \text{SL}(2, \mathbb{C})$ with associated ρ_i -equivariant harmonic maps $f_i: \mathbb{H} \rightarrow \mathbb{H}^3$ is bounded (up to conjugation) if and only if the energy of the harmonic maps f_i is uniformly bounded.*

We now assume that ρ_i is an unbounded sequence of representations with f_i as above. Consider the sequence of ρ_i -equivariant harmonic maps

$$\hat{f}_i: \mathbb{H} \longrightarrow (\mathbb{H}^3, d_i), \quad (3.11)$$

where the hyperbolic metric d on \mathbb{H}^3 is scaled by the square-root of the energy: $d_i(x, y) = d(x, y)/E^{1/2}(f_i)$, and $\hat{f}_i = f_i$. Then because of the scaling the \hat{f}_i have uniform modulus of continuity. Furthermore, by properties of thin triangles in \mathbb{H}^3 and the fact that $E(f_i) \rightarrow \infty$, one can see that geodesic triangles in the convex hull of the image of \hat{f}_i become infinitely thin (cf. [16], [156]).

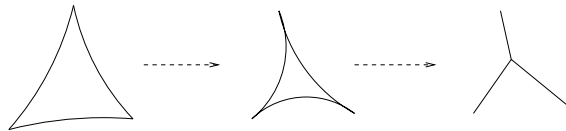


Figure 5

Using these ideas we have

Theorem 3.19 (Daskalopoulos–Dostoglou–Wentworth [34]). *For an unbounded sequence of irreducible $\mathrm{SL}(2, \mathbb{C})$ representations ρ_i the corresponding harmonic maps \hat{f}_i in (3.11) converge (after possibly passing to a subsequence) in the pullback sense to a Γ -equivariant harmonic map $\hat{f}: \mathbb{H} \rightarrow X$, where X is an \mathbb{R} -tree with isometric Γ action such that*

- (1) Γ acts on X without fixed points;
- (2) the length function of the action of Γ on X is in the projective class of the Morgan–Shalen limit of the sequence ρ_i ;
- (3) the image of \hat{f} is a minimal tree.

Let $\chi_{df}(\Gamma) \subset \chi(\Gamma)$ denote the subspace of discrete faithful representations. It is a consequence of Jorgenson type inequalities (cf. [141]) that the Morgan–Shalen limit of a sequence of discrete faithful representations is the length function of a *small* action on an \mathbb{R} -tree. By Skora’s Theorem 3.9, the tree is dual to a measured foliation, and therefore $\partial\chi_{df}(\Gamma) \subset \mathrm{PMF}(S)$. We actually have

$$\partial\chi_{df}(\Gamma) = \mathrm{PMF}(S) \simeq \mathrm{QD}(S), \quad (3.12)$$

$$\partial\mathcal{T}^{\mathrm{th}}(S) = \mathrm{PMF}(S) \simeq \mathrm{QD}(S). \quad (3.13)$$

The second equality (3.13), first proven by Thurston using the density of Jenkins–Strebel differentials, was also proven by Wolf using harmonic maps [196]. We show how this result follows from the discussion above. Recall from Theorem 2.23 that the map $\mathcal{H}: \mathcal{T}_{\mathrm{hyp}}(S) \rightarrow \mathrm{QD}(S)$ defined in (2.38) is a homeomorphism. Choose $t_j \rightarrow \infty$ and a sequence $\{\varphi_j\} \in \mathrm{QD}(S)$, $\|\varphi_j\|_1 = 1$. Without loss of generality, we may assume $\{\varphi_j\}$ converges to some nonzero $\varphi \in \mathrm{QD}(S)$. Let $[\sigma_j] = \mathcal{H}^{-1}(t_j\varphi_j)$, and let f_j be the associated harmonic maps. By definition, the Hopf differentials of the rescaled maps \hat{f}_j converge to $(1/2)\varphi$. Indeed, $\mathrm{Hopf}(\hat{f}_j) = t_j\varphi_j/E(f_j)$, and by (2.39) and (2.40), $E(f_j) \sim 2t_j$. On the other hand, the Hopf differentials of \hat{f}_j converge to the Hopf differential of the limiting equivariant map $\mathbb{H} \rightarrow X$. By the smallness of the action of Γ on X , Skora’s theorem implies that X is dual to a measured foliation. This measured foliation must coincide with the horizontal foliation of $(1/2)\varphi$. Hence, we have shown the equality (3.13) and that the map \mathcal{H} defined in (2.38) extends continuously as a map from $\partial\mathcal{T}(S)$ to the sphere at infinity in $\mathrm{QD}(S)$. Equality (3.12) follows from (3.13) and the fact that $\partial\mathcal{T}^{\mathrm{th}}(S) \subset \partial\chi_{df}(\Gamma) \subset \mathrm{PMF}(S)$.

4 Weil–Petersson geometry and mapping class groups

4.1 Weil–Petersson geodesics and isometries

Teichmüller space has a length space structure given by the Teichmüller distance (2.9). An alternative Riemannian structure arises from the description of Teichmüller space

via hyperbolic metrics presented in Section 2.1.5. This is the Weil–Petersson metric, and its properties continue to be the subject of much research. In this section, we present a short review of some of the aspects of Weil–Petersson geometry that will be relevant later on. The basic definitions as well as properties of the Weil–Petersson completion are discussed in Section 4.1.1. In Section 4.1.2 we introduce the mapping class group, and in Section 4.1.3 we indicate how the classification of individual mapping classes follows from the structure of Weil–Petersson geodesics.

4.1.1 The Weil–Petersson metric and its completion. Recall from Section 2.1.5 that the cotangent space $T_{[\sigma]}^* \mathcal{F}(S)$ is identified with the space of holomorphic quadratic differentials on (S, σ) . The complete hyperbolic metric on (S, σ) can be expressed in local conformal coordinates as $ds^2 = \sigma(z)|dz|^2$. Similarly, a quadratic differential has a local expression $\varphi = \varphi(z)dz^2$. Then for $\varphi \in T_{[\sigma]}^* \mathcal{F}(S)$, the Weil–Petersson cometric is given by $\|\varphi\|_{wp} = \|\varphi\|_2$ (see (2.17)). While there exist a wide variety of invariant metrics, the Weil–Petersson metric is in a real sense the most useful for applications. We refer the reader to Wolpert’s recent survey [206]. The two most important facts for us here are that (1) the Weil–Petersson metric has *negative sectional curvature* and (2) it is *incomplete*.

The curvature properties of Teichmüller space with the Weil–Petersson or Teichmüller metrics have an interesting history. It was long thought that the Teichmüller metric had negative curvature in the sense of triangle comparisons (see [112]). This was disproven by Masur in [123] (see also [119], and more recently [91], [129], [130], [125]). For the Weil–Petersson metric, the first step was taken by Ahlfors [4], who showed that the first variation of the area element induced by the hyperbolic metric vanishes. This implies the kählerity. He also established the negativity of the Ricci and holomorphic sectional curvatures. The following result was established later:

Theorem 4.1 (Tromba [183], Wolpert [203], see also [95], [177].). *The curvature of the Weil–Petersson metric has*

- (1) *holomorphic sectional curvatures and Ricci curvatures bounded above by $-1/2\pi(p-1)$, and*
- (2) *negative sectional curvature.*

Incompleteness is a consequence of the nature of degenerating Riemann surfaces. This was first recognized in the work of Bers, Chu, Wolpert and Masur (cf. [27], [124], [202]). A model for degeneration is given by the “plumbing construction”. Here is a simple version: let S_1 and S_2 be compact surfaces of genera p_1, p_2 . Choose local coordinates z_1, z_2 centered at points $x_1 \in S_1, x_2 \in S_2$. Fix $0 < t < 1$ and construct a new surface from the following three pieces: $S_1 \setminus \{|z_1| \leq 1\}$, $S_2 \setminus \{|z_2| \leq 1\}$, and the annulus $\{(z_1, z_2) : z_1 z_2 = t\}$. The boundary of the annulus is identified in the obvious way with the boundaries of the surfaces with disks deleted. In this way, one obtains a compact Riemann surface S_t of genus $p_1 + p_2$.

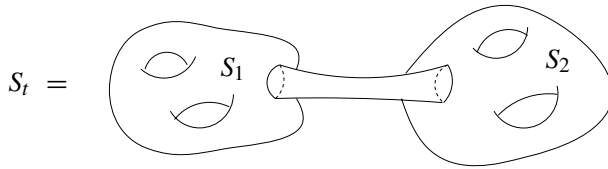


Figure 6

As $t \rightarrow 0$, the points in Teichmüller space corresponding to S_t diverge, because the annulus is begin “pinched”. This can also be seen from the hyperbolic geometry. Using the maximal principle, one can approximate the behavior of the hyperbolic metric on S_t (cf. [205]). In the pinching region, it is roughly approximated by the hyperbolic metric on the annulus given by

$$ds_t^2 = \frac{|dz|^2}{|z|^2 (\log |z|)^2} \frac{\Phi_t^2}{(\sin \Phi_t)^2},$$

where $\Phi_t = \pi \log |z| / \log |t|$, and z is either z_1 or z_2 . As a result the length ℓ of the “waist” of the annulus is shrinking to zero as $t \rightarrow 0$. In fact, the length is of order $\ell \sim 1/\log(1/t)$ (see [205]). Notice that every curve passing through the annulus must then become rather long. This is a general fact in the hyperbolic geometry of surfaces. The following rough statement of the *Collar Lemma* indicates that around short geodesics on a hyperbolic surface one always can find long cylinders. For a more precise statement, see [104].

Lemma 4.2. *Let (S, σ) be a hyperbolic surface and c a simple closed geodesic of length $\ell \neq 0$. Then any simple closed essential curve having nonzero geometric intersection with c has length on the order $\sim \log(1/\ell)$.*

The behavior of the Weil–Petersson metric at points in $\mathcal{T}(S)$ described by these degenerations has the following model due to Masur, Yamada, and Wolpert (for a review, see [206]). Define an incomplete metric space

$$\mathbb{M} = \{(\xi, \theta) \in \mathbb{R}^2 : \xi > 0\}, \quad ds_{\mathbb{M}}^2 = 4d\xi^2 + \xi^6 d\theta^2. \quad (4.1)$$

The metric completion $\overline{\mathbb{M}}$ of \mathbb{M} is obtained by adding a single point $\partial\mathbb{M}$ corresponding to the entire real axis $\xi = 0$. The completion is then an NPC space which is, however, not locally compact. Indeed, an ε -neighborhood of $\partial\mathbb{M}$ contains all points of the form (ξ, θ) , $\xi < \varepsilon$, and θ arbitrary.

The importance of $\overline{\mathbb{M}}$ is that it is a model for the normal space to the boundary strata. Let $\overline{\mathcal{T}}(S)$ denote the metric completion of $\mathcal{T}(S)$. We have the following local description (cf. [124]): $\partial\mathcal{T}(S) = \overline{\mathcal{T}}(S) \setminus \mathcal{T}(S)$ is a disjoint union of smooth connected strata formed by collapsing a collection of disjoint simple closed essential curves on S to points. Associated to the nodal surface is another Teichmüller space which is

by definition the set of equivalence classes of complex structures on the normalized (possibly disconnected) surface, with the preimages of the nodes as additional marked points. It is therefore naturally isomorphic to a product of lower dimensional Teichmüller spaces. A neighborhood of a point in the boundary is then homeomorphic to an open set in the lower dimensional product crossed with as many multiples of \mathbb{M} as there are collapsed curves. Metrically, the statement is that the Weil–Petersson metric in this neighborhood is equal to the product metric up to third order in the ξ variables (see [39], [124], [152], [205], [206], [201], [210]). Moreover, by Wolpert’s theorem (see Theorem 4.16 below) $\mathcal{T}(S)$ with the Weil–Petersson metric is geodesically convex and the boundary strata are totally geodesically embedded.

The following observation is also due to Yamada:

Theorem 4.3. *The completion $\overline{\mathcal{T}}(S)$ of $\mathcal{T}(S)$ is a complete NPC space.*

While this follows on general principles (cf. [19]), the identification of the boundary strata of the completion with lower dimensional Teichmüller spaces (and Weil–Petersson metrics) is especially useful.

Let us point out two properties of the geometry of the Weil–Petersson completion that are consequences of this expansion. These were first stated by Yamada [210]. The first result, dubbed *nonrefraction* by Wolpert, is the statement that geodesics from points in Teichmüller space to the boundary touch the boundary only at their endpoints (see Figure 7 (a)). It is easy to see that this is true for the model space above. Indeed, the equations for a unit speed geodesic $\alpha(t) = (\xi(t), \theta(t))$ in \mathbb{M} are

$$\xi\ddot{\xi} = (3/4)\xi^6\dot{\theta}^2, \quad \xi^6\dot{\theta} = \text{constant}, \quad 2|\dot{\xi}|, \quad \xi^3|\dot{\theta}| \leq 1. \quad (4.2)$$

If $\xi(t) \rightarrow 0$ as $t \rightarrow 1$, say, then the second and third equations imply that the constant above must vanish. In other words, $\theta(t)$ is constant and $\xi(t)$ is linear. The proof of the statement for geodesics in $\overline{\mathcal{T}}(S)$ involves a scaling argument to approximate geodesics in $\overline{\mathcal{T}}(S)$ by corresponding geodesics in the model space. The third order approximation of the Weil–Petersson metric by the model metric is sufficient to show that the approximation of geodesics is also to high order, and the qualitative behavior of geodesics in $\overline{\mathcal{T}}(S)$ is the same as for the model space.

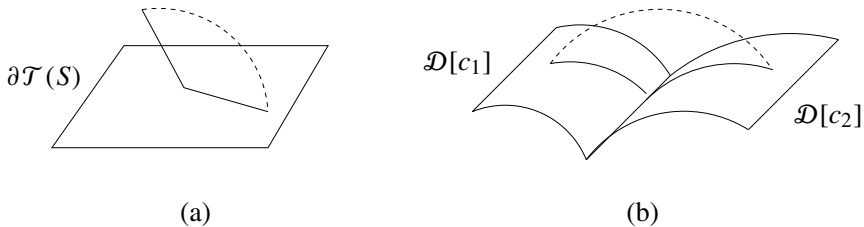


Figure 7

Another application of this approximation gives the second important result: the different strata of the boundary of $\overline{\mathcal{T}}(S)$ intersect transversely. For example, consider disjoint nonisotopic simple closed essential curves $\{c_1, c_2\}$ on a closed compact surface S with isotopy classes $[c_1], [c_2]$. Let $[\sigma_1]$ denote a point in the boundary component $\mathcal{D}[c_1]$ of $\overline{\mathcal{T}}(S)$ corresponding to pinching c_1 . Similarly, let $[\sigma_2]$ denote a point in the boundary component $\mathcal{D}[c_2]$ of $\overline{\mathcal{T}}(S)$ corresponding to pinching c_2 . Since c_1 and c_2 are disjoint, the intersection of the closures $\overline{\mathcal{D}}[c_1] \cap \overline{\mathcal{D}}[c_2]$ is nonempty, and in fact contains $\mathcal{D}([c_1], [c_2])$, the stratum where both c_1 and c_2 are pinched. In particular, there is a path in $\overline{\mathcal{T}}(S)$ from σ_1 to σ_2 , lying completely in the boundary, which corresponds to first pinching c_2 , and then “opening up” c_1 . The theorem states that this path has a “corner” at its intersection with $\mathcal{D}([c_1], [c_2])$, and is therefore not length minimizing. In fact, the geodesic from $[\sigma_1]$ to $[\sigma_2]$ intersects the boundary of $\overline{\mathcal{T}}(S)$ only in its endpoints (see Figure 7 (b)).

4.1.2 The mapping class group. Denote by $\text{Diff}(S)$ (resp. $\text{Diff}^+(S)$) denote the group of smooth diffeomorphisms (resp. orientation preserving diffeomorphisms) of S with the smooth topology. Recall that $\text{Diff}_0(S)$ denotes the identity component of $\text{Diff}(S)$, that is, the group of all diffeomorphisms isotopic to the identity. The *mapping class group* of S is the quotient

$$\text{Mod}(S) = \text{Diff}^+(S)/\text{Diff}_0(S).$$

See [54] for a recent survey on mapping class groups. From any of the several definitions of Teichmüller space given previously, it is clear that $\text{Mod}(S)$ acts on $\mathcal{T}(S)$. The first important result about this action is the following

Theorem 4.4. *The mapping class group acts properly discontinuously on $\mathcal{T}(S)$.*

This result is commonly attributed to Fricke. One method of proof follows from the general fact that the action of $\text{Diff}(S)$ on the space $\text{Met}(S)$ of smooth Riemannian metrics is properly discontinuous (cf. [48], [47]). In particular the restriction to the action of $\text{Diff}_0(S)$ on $\text{Met}_{\text{hyp}}(S)$ is properly discontinuous, and Teichmüller space, which is the quotient $\mathcal{T}_{\text{hyp}}(S) = \text{Met}_{\text{hyp}}(S)/\text{Diff}_0(S)$, inherits such an action of $\text{Mod}(S)$.

Diffeomorphisms of S determine automorphisms of $\Gamma = \pi_1(S)$ as follows. Let $x_0 \in S$ be a fixed basepoint. A diffeomorphism $\phi: S \rightarrow S$ determines an automorphism of the fundamental group $\pi_1(S, x_0)$ if $\phi(x_0) = x_0$. Now any diffeomorphism is isotopic to one which fixes x_0 . Different choices of isotopy define automorphisms of Γ which differ by an inner automorphism. Hence, there is a homomorphism $\text{Diff}(S)/\text{Diff}_0(S) \rightarrow \text{Out}(S)$ where $\text{Out}(S)$ is the outer automorphism group of $\pi_1(S)$.

Theorem 4.5 (Dehn–Nielsen [149]). *The homomorphism described above gives an isomorphism $\text{Diff}(S)/\text{Diff}_0(S) \simeq \text{Out}(S)$.*

The quotient $\mathcal{M}(S) = \mathcal{T}(S)/\text{Mod}(S)$ is the classical Riemann moduli space. Since by Teichmüller's theorem $\mathcal{T}(S)$ is contractible (and in particular, simply connected), $\text{Mod}(S)$ may be regarded as the fundamental group of $\mathcal{M}(S)$. However, $\text{Mod}(S)$ does not quite act freely, so this interpretation holds only in the *orbifold* sense. Indeed, $\mathcal{M}(S)$ is actually simply connected [121]. The compactification

$$\overline{\mathcal{M}}(S) = \overline{\mathcal{T}}(S)/\text{Mod}(S)$$

is homeomorphic to the *Deligne–Mumford compactification* of $\mathcal{M}(S)$, and it is a projective algebraic variety (with orbifold singularities) [40]. An important measure of the interior regions of $\mathcal{M}(S)$ is given by the Mumford–Mahler compactness theorem:

Theorem 4.6 (Mumford [146]). *The set $\mathcal{M}(S)_\varepsilon \subset \mathcal{M}(S)$ consisting of equivalence classes of Riemann surfaces where the hyperbolic lengths of all closed geodesics are bounded below by $\varepsilon > 0$ is compact.*

Note the condition in the theorem is $\text{Diff}(S)$ -invariant and so is valid on the moduli space $\mathcal{M}(S)$. The corresponding result is, of course, not true for $\mathcal{T}(S)$ because of the proper action of the infinite discrete group $\text{Mod}(S)$. For example, the orbit of a point in $\mathcal{T}(S)$ by $\text{Mod}(S)$ is unbounded, but projects to a single point in $\mathcal{M}(S)$. This, however, is the *only* distinction between $\mathcal{T}(S)$ and $\mathcal{M}(S)$, and so the Mumford–Mahler compactness theorem can be used effectively to address convergence questions in $\mathcal{T}(S)$ as well.

We illustrate this by proving a fact that will be useful later on. Given a simple closed curve $c \subset S$, let $\ell_c[\sigma]$ denote the length of the geodesic in the homotopy class of c with respect to the hyperbolic metric σ . Note that this is independent of the choice of σ up to the action of $\text{Diff}_0(S)$. Hence, ℓ_c gives a well-defined function

$$\ell_c: \mathcal{T}(S) \longrightarrow \mathbb{R}^+. \quad (4.3)$$

Then we have the following

Corollary 4.7. *If $[\sigma_j]$ is a sequence in $\mathcal{T}(S)$ contained in no compact subset then there is a simple closed curve $c \subset S$ such that $\ell_c[\sigma_j]$ is unbounded.*

Proof. For a point $[\sigma] \in \mathcal{T}(S)$, let $[[\sigma]] \in \mathcal{M}(S)$ denote the corresponding point in $\mathcal{M}(S)$. Without loss of generality, we may assume $[\sigma_j]$ has no convergent subsequence in $\mathcal{T}(S)$. The same may or may not be true for the sequence $[[\sigma_j]] \subset \mathcal{M}(S)$. Indeed, by Theorem 4.6, there are two cases: (1) there are elements $[\phi_j] \in \text{Mod}(S)$ and a point $[\sigma_\infty] \in \mathcal{T}(S)$ such that $[\phi_j][\sigma_j] \rightarrow [\sigma_\infty]$ (after passing to a subsequence); (2) there are simple closed curves c_j such that $\ell_{c_j}[\sigma_j] \rightarrow 0$ (after passing to a subsequence). In the first case, our assumptions imply that infinitely many $[\phi_j]$ are distinct. It follows that there is a simple closed curve c such that $\ell_{f_j(c)}[\sigma_\infty] \rightarrow \infty$. But then $\ell_c[\sigma_j] \rightarrow \infty$, as desired. In the second case, we may assume c_j converges projectively to a nontrivial measured foliation \mathcal{F} (see Section 2.3.2). If c is any simple closed curve

with $i([c], \mathcal{F}) \neq 0$, then $i(c, c_j) \neq 0$ for j large. But since $\ell_{c_j}[\sigma_j] \rightarrow 0$, $\ell_c[\sigma_j] \rightarrow \infty$ by the Collar Lemma 4.2. \square

Thurston’s classification of surface diffeomorphisms may be described in terms of the natural action of $\text{Mod}(S)$ on $\text{MF}(S)$ and $\text{PMF}(S)$: an element $[\phi] \in \text{Mod}(S)$ is called *reducible* if $[\phi]$ fixes (up to isotopy) some collection of disjoint simple closed essential curves on S . It is called *pseudo-Anosov* if there is $r > 1$ and transverse measured foliations \mathcal{F}_+ , \mathcal{F}_- on S such that $[\phi]\mathcal{F}_+$ is measure equivalent to $r\mathcal{F}_+$, and $[\phi]\mathcal{F}_-$ is measure equivalent to $r^{-1}\mathcal{F}_-$. \mathcal{F}_+ and \mathcal{F}_- are called the *stable* and *unstable* foliations of $[\phi]$, respectively. The classification states that any $[\phi] \in \text{Mod}(S)$ is either *periodic* (i.e. finite order), infinite order and reducible, or pseudo-Anosov. Moreover, these are mutually exclusive possibilities.

4.1.3 Classification of Weil–Petersson isometries. We now indicate how the Thurston classification of mapping classes is mirrored by the Weil–Petersson geometry. The action of $\text{Mod}(S)$ on $\mathcal{T}(S)$ is isometric with respect to the Weil–Petersson metric. Conversely, every Weil–Petersson isometry is essentially given by a mapping class (see [21], [129], [126], [206]). Since the Weil–Petersson metric has negative curvature it is a natural to classify individual mapping classes in a manner similar to isometries of Cartan–Hadamard manifolds.

Theorem 4.8 (Daskalopoulos–Wentworth [39], Wolpert [206]). *If $[\phi] \in \text{Mod}(S)$ is infinite order and irreducible, then there is a unique $[\phi]$ -invariant complete Weil–Petersson geodesic in $\mathcal{T}(S)$.*

Here is a very rough idea of proof of this result. Let $\tilde{\alpha}_j: [0, 1] \rightarrow \mathcal{T}(S)$, $\tilde{\alpha}_j(1) = [\phi]\tilde{\alpha}_j(0)$ be a sequence of curves minimizing the translation length of $[\phi]$, i.e.

$$\lim_{j \rightarrow \infty} \int_0^1 \|\dot{\tilde{\alpha}}_j\|_{wp} dt = \mathcal{L}_{wp}[\phi].$$

Let $\alpha_j: S^1 \rightarrow \bar{\mathcal{M}}(S)$ be the projection of $\tilde{\alpha}_j$. Since $\bar{\mathcal{M}}(S)$ is compact one can show using Ascoli’s theorem that, after passing to a subsequence, α_j converge uniformly to some curve $\alpha: S^1 \rightarrow \bar{\mathcal{M}}(S)$. The trick now is to show that this curve admits a lift $\tilde{\alpha}: [0, 1] \rightarrow \bar{\mathcal{T}}(S)$, $\tilde{\alpha}(1) = [\phi]\tilde{\alpha}(0)$. Then $\tilde{\alpha}$ must be an invariant geodesic. Since $[\phi]$ is irreducible, by the nonrefraction results we know that $\tilde{\alpha}$ must have image in $\mathcal{T}(S)$. The existence of a lift is not obvious, since $\bar{\mathcal{T}}(S) \rightarrow \bar{\mathcal{M}}(S)$ is “branched” to infinite order along the boundary $\partial\mathcal{T}(S)$. One needs to exploit the fact that α is the *limit* of curves that are liftable. We refer to [39] for more details.

The existence of invariant geodesics for infinite order irreducible mapping classes allows for the precise classification of Weil–Petersson isometries in terms of translation length that we have given in Table 1. For $[\phi] \in \text{Mod}(S)$, define the *Weil–Petersson*

translation length by

$$\mathfrak{L}_{wp}[\phi] = \inf_{[\sigma] \in \mathcal{T}(S)} d_{wp}([\sigma], [\phi][\sigma]) . \quad (4.4)$$

Table 1. Classification of Weil–Petersson Isometries.

	semisimple	not semisimple
$\mathfrak{L}_{wp} = 0$	periodic	strictly pseudoperiodic
$\mathfrak{L}_{wp} \neq 0$	infinite order irreducible	reducible but not pseudoperiodic

First, let us clarify the terminology used there: $[\phi] \in \text{Mod}(S)$ is *pseudoperiodic* if it is either periodic, or it is reducible and periodic on the reduced components; it is strictly pseudoperiodic if it is pseudoperiodic but not periodic. Furthermore, we say that $[\phi]$ is *semisimple* if there is $[\sigma] \in \mathcal{T}(S)$ such that $\mathfrak{L}_{wp}[\phi] = d_{wp}([\sigma], [\phi][\sigma])$.

Here is a sketch of the proof: first, note that it is a consequence of Theorem 4.4 that $\mathfrak{L}_{wp}[\phi] = 0$ if and only if $[\phi]$ is pseudoperiodic. The first row of Table 1 then follows from this and the fact that $[\phi]$ has a fixed point in $\mathcal{T}(S)$ if and only if $[\phi]$ is periodic. If $[\phi]$ is infinite order irreducible, then as a consequence of Theorem 4.8, $\mathfrak{L}_{wp}[\phi]$ is attained along an invariant geodesic, so these mapping classes are semisimple. Conversely, suppose the translation length is attained at $[\sigma] \in \mathcal{T}(S)$, but $[\phi][\sigma] \neq [\sigma]$. Then we argue as in Bers [14] (see also, [9, p. 81]) to show that the geodesic from $[\sigma]$ to $[\phi][\sigma]$, which exists by the geodesic convexity of the Weil–Petersson metric, may be extended to a complete $[\phi]$ -invariant geodesic. On the other hand, if there is a complete, nonconstant Weil–Petersson geodesic in $\mathcal{T}(S)$ that is invariant with respect to a mapping class $[\phi] \in \text{Mod}(S)$, the negative curvature implies that $[\phi]$ must be infinite order and irreducible.

It is worth mentioning that no properties of pseudo-Anosov’s other than the fact that they have infinite order and are irreducible were used in the proof above. In particular, the description given in Table 1 is independent of Thurston’s classification.

We point out a further property of the axes of pseudo-Anosov’s.

Theorem 4.9 (Daskalopoulos–Wentworth [39], Wolpert [206]). *Let $A_{[\phi]}$ and $A_{[\phi']}$ be the axes for independent pseudo-Anosov mapping classes $[\phi]$ and $[\phi']$. Then $A_{[\phi]}$ and $A_{[\phi']}$ diverge.*

This result is also not completely obvious because of the noncompleteness of $\mathcal{T}(S)$. More to the point, there exist *flats*, i.e. a totally geodesically embedded copy of $\mathbb{R}^m \hookrightarrow \overline{\mathcal{T}}(S)$, which potentially hinder the divergence. A much more detailed discussion of asymptotics of complete Weil–Petersson geodesics is forthcoming (see [22]).

4.2 Energy of harmonic maps

In this section we return to harmonic maps and show how they can be used to probe the action of the mapping class group on Teichmüller space. In Section 4.2.1 we discuss Nielsen's realization problem for finite subgroups of the mapping class group. In Section 4.2.2 we introduce two classes of functions on Teichmüller space that are constructed using the energy of harmonic maps, and we indicate when these functions are proper. In Section 4.2.3 we discuss the convexity of one of the two classes and show how this resolves the Nielsen conjecture. We also state Wolpert's result on convexity of length functions. Finally, in Section 4.2.4, we indicate some other applications of the energy functionals.

4.2.1 Nielsen realization. Here we discuss the classical question of Nielsen [150]. The exact sequence

$$1 \longrightarrow \text{Diff}_0(S) \longrightarrow \text{Diff}^+(S) \xrightarrow{\pi} \text{Mod}(S) \longrightarrow 1 \quad (4.5)$$

which defines the mapping class group does not split in general (see [122], [143], [144]). The *realization problem* asks for which subgroups $G \subset \text{Mod}(S)$ does there exist a homomorphism $J: G \rightarrow \text{Diff}^+(S)$ such that $\pi \circ J = \text{id}$.

Let S be a closed Riemann surface of negative Euler characteristic. Then we have the following two important facts. First, if ϕ is a holomorphic automorphism of S homotopic to the identity, then ϕ is in fact *equal* to the identity. Indeed, if this were not the case then since complex curves in a complex surface intersect positively the number of fixed points of ϕ , counted with multiplicity, would necessarily be positive. On the other hand, if $\phi \sim \text{id}$, then by the Lefschetz fixed point theorem the total intersection number is just the Euler characteristic of S , which we have assumed is negative.

From this fact we arrive at Fenchel's observation that if a subgroup $G \subset \text{Mod}(S)$ fixes a point $[j] \in \mathcal{T}(S)$, then G can be realized as the automorphism group of a Riemann surface (S, j) with j in the class $[j]$. For if ϕ_1, \dots, ϕ_m are holomorphic lifts to $\text{Diff}^+(S)$ of generators $[\phi_1], \dots, [\phi_m]$ of G , then any relation on the $[\phi_j]$'s, applied to the ϕ_j 's, is a holomorphic map $\sim \text{id}$, and so by the previous paragraph the relations in the group also lift. In particular, (4.5) splits over G .

The second fact is that the automorphism group of a Riemann surface of genus $p \geq 2$ is finite. This is because on the one hand it is the isometry group of the hyperbolic metric, which is compact, and on the other hand it is discrete, since there are no holomorphic vector fields. Hence, any subgroup of the mapping class group which fixes a point in Teichmüller space is finite and (4.5) splits over it. These two facts motivate the following result, which is known as the *Nielsen Realization Theorem*.

Theorem 4.10 (Kerckhoff [105]). *The sequence (4.5) splits over all finite subgroups of $\text{Mod}(S)$.*

From the discussion above, the idea of the proof is to show the following

Theorem 4.11. *Let $G \subset \text{Mod}(S)$ be a finite subgroup of the mapping class group. Then G has a fixed point in $\mathcal{T}(S)$.*

The complete proof of Theorem 4.10 was first obtained by Kerckhoff in [105] and later by Wolpert [204]. Both proofs proceed via Theorem 4.11. Partial results had been found earlier by Fenchel [58], [59] and Zieschang [214]. See also Tromba [187].

4.2.2 Properness of the energy. Let M be an arbitrary compact Riemannian manifold and S a closed hyperbolic surface with negative Euler characteristic. Now if $\rho: \pi_1(M) \rightarrow \pi_1(S)$ is a given homomorphism it follows by Theorem 2.15 that there is a harmonic map $f: M \rightarrow S$ such that the induced action $f_*: \pi_1(M) \rightarrow \pi_1(S)$ coincides with ρ . The energy $E(f)$ then depends only on the equivalence class of hyperbolic metrics $[\sigma] \in \mathcal{T}(S)$ (see Theorem 2.18). In other words, there is a well-defined function

$$\mathcal{E}_\rho^+: \mathcal{T}(S) \longrightarrow \mathbb{R}^+.$$

The existence of a minimum is in turn a reflection of the homomorphism ρ . One way to guarantee a minimum is to show that \mathcal{E}_ρ^+ diverges at infinity. In this context, we have the following

Proposition 4.12. *If ρ is surjective then the associated function \mathcal{E}_ρ^+ is proper.*

Proof. This is easy to see, given the Lipschitz bound Proposition 2.17 and the Mumford–Mahler Compactness Theorem 4.6 (or more precisely, Corollary 4.7). Indeed, if \mathcal{E}_ρ^+ is not proper, there is a sequence $[\sigma_j]$ and harmonic maps $f_j: M \rightarrow (S, \sigma_j)$ in the homotopy class defined by ρ , such that $E[\sigma_j] \leq B$ for some constant B . Furthermore, we may assume there is a simple closed curve c with $\ell_c[\sigma_j] \rightarrow \infty$. Let s be a closed curve in M with $f_j(s)$ homotopic to c . Then since the f_j are uniformly Lipschitz,

$$\ell_c[\sigma_j] \leq \text{length}(f_j(s)) \leq \tilde{B} \text{length}(s).$$

Since the right hand side is fixed independent of j and the left hand side diverges with j , we derive a contradiction. \square

The superscript $+$ on \mathcal{E}_ρ^+ is to remind us that this is a function of the hyperbolic metric on the *target*. It is also interesting to consider the energy as a function of the domain metric (cf. (3.5)). Let M be a compact Riemannian manifold with nonpositive sectional curvature. Let S be a closed surface and let $\rho: \pi_1(S) \rightarrow \pi_1(M)$ is a homomorphism. Then for each complex structure σ on S there is a harmonic map $f: (S, \sigma) \rightarrow M$ whose induced action on π_1 coincides with ρ . The energy of this map gives a well-defined function

$$\mathcal{E}_\rho^- \longrightarrow \mathcal{T}(S) \longrightarrow \mathbb{R}^+.$$

Again, the existence of minima can be deduced from the properness of this functional. The following can be proved using the same ideas as in the proof of Proposition 4.12 and Corollary 4.7.

Proposition 4.13 (see [163], [169]). *If ρ is injective then \mathcal{E}_ρ^- is proper.*

As we have seen in Theorem 3.11, the function \mathcal{E}_ρ^- is differentiable. From the discussion in Section 2.1.5 (see esp. (2.18)), critical points correspond to *conformal* harmonic maps, i.e. those for which the Hopf differential vanishes. According to Sacks–Uhlenbeck [163], these are branched minimal surfaces in M .

There is a remarkable connection between these functionals and the Weil–Petersson metric. If we take $M = (S, \sigma_0)$ for some hyperbolic metric σ_0 and $\rho = \text{id}$, we have defined two functions $\mathcal{E}_{\text{id}}^\pm$ on $\mathcal{T}(S)$, both of which clearly have critical points at $[\sigma_0]$. We have

Theorem 4.14 (Tromba [184], Wolf [196], Jost [94]). *The second variation of either $\mathcal{E}_{\text{id}}^\pm$ at $[\sigma_0]$ is a positive definite hermitian form on $T_{[\sigma_0]}\mathcal{T}(S)$ which coincides with the Weil–Petersson metric.*

A critical point of $\mathcal{E}_{\text{id}}^-$ is a holomorphic map $(S, \sigma) \rightarrow (S, \sigma_0)$ homotopic to the identity. As argued in Section 4.2.1, this must be the identity and $\sigma = \sigma_0$. We conclude that $\mathcal{E}_{\text{id}}^-$ is a proper function on $\mathcal{T}(S)$ with a unique critical point. By Theorem 4.14, it is also nondegenerate. It follows that $\mathcal{T}(S)$ is diffeomorphic to \mathbb{R}^n ; hence, we have a fourth proof of Theorem 2.8 (see [61]).

4.2.3 Convexity of energy and length functionals

Theorem 4.15 (Tromba [187], Yamada [209]). *The energy \mathcal{E}_ρ^+ defined above is strictly convex along Weil–Petersson geodesics.*

This result was first obtained by Tromba in the case where M is homeomorphic to S . It was later generalized to the statement above by Yamada. It follows that the minimum of \mathcal{E}_ρ^+ is unique if it exists.

The conclusion is that there exists an abundance of convex exhaustion functions on Teichmüller space and an explicit method to construct them. Any one of these gives a solution to the Nielsen problem! For the average of such a function over a finite subgroup $G \subset \text{Mod}(S)$ is again strictly convex *and* G -invariant. Hence, its unique minimum is also G -invariant, i.e. a fixed point of G , and Theorem 4.11 is proven. The easiest example is to take $M = (S, \sigma_0)$, for any complex structure σ_0 , and $\rho = \text{id}$, as in the previous section.

It turns out that the analogous statement Theorem 4.15 for \mathcal{E}_ρ^- is false. For example, we could choose M to be a fibered hyperbolic 3-manifold with S a fiber and ρ the homomorphism coming from the inclusion. Then ρ is invariant by conjugation of the monodromy of the fibration, which by a theorem of Thurston is represented by a

pseudo-Anosov diffeomorphism (cf. [153]). In particular, it has infinite order. This fact leads to infinitely many minima of \mathcal{E}_ρ^- , whereas if \mathcal{E}_ρ^- were strictly convex, it would have a unique minimum. It is certainly an interesting question to find conditions where convexity holds.

A very special case of the previous discussion is when M is a circle. Harmonic maps from a circle correspond to geodesics. Historically, geodesic length functions were considered before the energy of harmonic maps from higher dimensional domains. In particular, we have the following important result of Wolpert.

Theorem 4.16 (Wolpert [204], [207]). *For any simple closed curve c , the function $\ell_c: \mathcal{T}(S) \rightarrow \mathbb{R}^+$ defined in (4.3) is strictly convex along Weil–Petersson geodesics. The extension of the length function to geodesic currents is also strictly convex.*

One consequence of this is the geodesic convexity of Teichmüller space, i.e. between any two points in $\mathcal{T}(S)$ there exists a unique Weil–Petersson geodesic. One can also construct convex exhaustion functions, although in a manner slightly different from that of the previous section. If we choose a collection c_1, \dots, c_m of simple closed curves which are *filling* in S in the sense that any other simple closed essential curve has nontrivial intersection with at least one c_j , then the function

$$\beta = \ell_{c_1} + \dots + \ell_{c_m}, \quad (4.6)$$

is an exhaustion function. This again follows by Mumford–Mahler compactness and the Collar Lemma. Since β is also strictly convex, this gives a solution to Nielsen’s problem as above, and indeed this is Wolpert’s method.

Finally, we point out that Kerckhoff’s proof of Theorem 4.10 was the first to lay out this type of argument. The difference is that he proved convexity not with respect to the Weil–Petersson geometry but along Thurston’s earthquake deformations.

4.2.4 Further applications. We now enumerate some other applications of the ideas developed in previous section.

- *Convex cocompact representations.* Note that Proposition 4.13 can also be adapted to the equivariant case and metric space targets. Here, $\rho: \Gamma = \pi_1(S) \rightarrow \text{Iso}(X)$, where X is a simply connected NPC space. Injectivity is replaced by the condition that the translation length of any isometry in the image is bounded below by a uniform constant.

A discrete embedding $\rho: \Gamma \rightarrow \text{Iso}(X)$ is *convex cocompact* if there exists a ρ -invariant closed geodesically convex subset $N \subset X$ such that $N/\rho\Gamma$ is compact.

Theorem 4.17 (Goldman–Wentworth [73]). *Mod(S) acts properly discontinuously on the space of convex cocompact embeddings $\rho: \Gamma \rightarrow \text{Iso}(X)$.*

When $\text{Iso}(X) = \text{PSL}(2, \mathbb{C})$, a convex cocompact representation is *quasi-Fuchsian*, that is, a discrete embedding whose action on $S^2 = \partial\mathbb{H}^3$ is topologically conjugate

to the action of a discrete subgroup of $\mathrm{PSL}(2, \mathbb{C})$. In this case, Theorem 4.17 is just the known fact that $\mathrm{Mod}(S)$ acts properly on the space $\mathcal{QF}(S)$ of quasi-Fuchsian embeddings. Indeed, Bers' simultaneous uniformization theorem [12] provides a $\mathrm{Mod}(S)$ -equivariant homeomorphism

$$\mathcal{QF}(S) \longrightarrow \mathcal{T}(S) \times \mathcal{T}(S).$$

Properness of the action of $\mathrm{Mod}(S)$ on $\mathcal{T}(S)$, Theorem 4.4, implies properness on $\mathcal{QF}(S)$.

The idea of the proof of Theorem 4.17 is to show that if ρ is convex cocompact

- (1) then there exists a ρ -equivariant harmonic map $f: \tilde{S} \rightarrow X$, and
- (2) the corresponding energy functional $\mathcal{E}_\rho^-: \mathcal{T}(S) \rightarrow \mathbb{R}^+$ is proper.

Then one associates to each ρ the compact subset of minima of \mathcal{E}_ρ^- in $\mathcal{T}(S)$, and properness of the action of $\mathrm{Mod}(S)$ on $\mathcal{T}(S)$ implies the result. See [73] for more details.

• *Filling foliations.* Recall from Section 3.1.2 that by the Hubbard–Masur theorem any measured foliation can be realized as the horizontal foliation of a holomorphic quadratic differential. As a second application, consider the problem of realizing a pair of measured foliations as the horizontal and vertical foliations of a single quadratic differential on some Riemann surface. A pair \mathcal{F}_+ , \mathcal{F}_- of measured foliations on S is called *filling* if for any third measured foliation \mathcal{G}

$$i(\mathcal{F}_+, \mathcal{G}) + i(\mathcal{F}_-, \mathcal{G}) \neq 0,$$

where $i(\cdot, \cdot)$ denotes the intersection number (see Section 2.3.2).

Theorem 4.18 (Gardiner–Masur [67]). *\mathcal{F}_+ , \mathcal{F}_- are filling if and only if there is a complex structure j and a holomorphic quadratic differential φ on (S, j) such that \mathcal{F}_+ and \mathcal{F}_- are measure equivalent to the vertical and horizontal foliations of φ , respectively. Moreover, $[j] \in \mathcal{T}(S)$, and φ for each $j \in [j]$, are uniquely determined by \mathcal{F}_\pm .*

It is relatively easy to see that the horizontal and vertical trajectories of a holomorphic quadratic differential are filling (cf. [67, Lemma 5.3]). The proof of the converse follows by showing, using arguments similar to those in the proof of Proposition 4.13, that $\mathrm{Ext}_{\mathcal{F}_+} + \mathrm{Ext}_{\mathcal{F}_-}$ is a proper function on $\mathcal{T}(S)$. The first variational formula Theorem 3.11 shows that a local minimum is a point at which the quadratic differentials for \mathcal{F}_+ and \mathcal{F}_- are related by a minus sign. On the other hand, by the argument in Section 3.1.2, \mathcal{F}_\pm are therefore vertical and horizontal foliations of one and the same differential. Uniqueness can also be proven by analytic methods (see [194] for more details).

• *Holomorphic convexity of $\mathcal{T}(S)$.* The convex exhaustion functions constructed in the previous sections are, in particular, strictly plurisubharmonic (Tromba [188] showed that this is true for $\mathcal{E}_{\mathrm{id}}^-$ as well). This gives a new proof of the following

Theorem 4.19 (Bers–Ehrenpreis [15]). *Teichmüller space is a Stein manifold.*

By a slight modification of length functions, we also have

Theorem 4.20 (Yeung [212]). *$\mathcal{T}(S)$ admits a bounded strictly plurisubharmonic function.*

Explicitly, one may take $-\beta^{-\varepsilon}$, where β is the function in (4.6) and $0 < \varepsilon < 1$. The existence of a bounded plurisubharmonic function has important implications for the equivalence of invariant metrics on Teichmüller space (see [24], [120], [212]).

5 Harmonic maps to Teichmüller space

5.1 Existence of equivariant harmonic maps

In many ways this last chapter combines ideas from all of the previous ones. Because of the nonpositive curvature of the Weil–Petersson metric, harmonic maps with Teichmüller space as a target have good regularity properties. The isometry group is the mapping class group, so the equivariant problem gives a way to study representations of fundamental groups to $\text{Mod}(S)$. Since the Weil–Petersson metric is not complete, we need to pass to the completion $\overline{\mathcal{T}}(S)$ and use the theory of singular space targets of Gromov–Korevaar–Schoen. In Section 5.1.1 we show how the results of Section 4.1.3 can be used to prove existence of equivariant harmonic maps to $\overline{\mathcal{T}}(S)$, and in Section 5.1.2 we state a result on the regularity of energy minimizing maps for surface domains. Finally, in Section 5.1.3, we discuss the special case of holomorphic maps from surfaces to Teichmüller space. An a priori bound on the energy of such maps gives rise to the Arakelov–Paršin finiteness result (see Theorem 5.9).

5.1.1 Maps to the completion. As an application of the previous results, we consider the problem of finding energy minimizing equivariant maps to Teichmüller space with the Weil–Petersson metric. Recall the set-up: let M be a compact Riemannian manifold with universal cover \tilde{M} , and let $\rho: \Gamma = \pi_1(M) \rightarrow \text{Mod}(S)$ be a homomorphism. Since $\text{Mod}(S)$ acts on $\overline{\mathcal{T}}(S)$ by isometries, we may ask under what conditions does there exist a ρ -equivariant energy minimizing map $f: \tilde{M} \rightarrow \overline{\mathcal{T}}(S)$.

Note that these may be regarded as harmonic maps $M \rightarrow \overline{\mathcal{M}}(S)$, although there are two points of caution. The first is that strictly speaking $\overline{\mathcal{M}}(S)$ is not a manifold, but has orbifold singularities at those points corresponding to Riemann surfaces with automorphisms. Hence, the smoothness of the map, and the harmonic map equations, should be understood on a smooth finite (local) cover of $\overline{\mathcal{M}}(S)$. The second (more important) point is that the homotopy class of a map $M \rightarrow \mathcal{M}(S)$ should be taken in the orbifold sense (i.e. equivariantly with respect to a homomorphism $\Gamma \rightarrow \text{Mod}(S)$).

Indeed, by the simple connectivity of $\mathcal{M}(S)$ remarked on in Section 4.1.2, homotopy classes of maps to $\mathcal{M}(S)$ are very different from equivariant homotopy classes of maps to $\mathcal{T}(S)$.

As in Section 3.1.1, the answer to the existence question depends on the asymptotic dynamics of the image subgroup $\rho(\Gamma) \subset \text{Mod}(S)$. In general, the asymptotic behavior of Weil–Petersson geodesics is quite complicated (see [20], [22]). As an approximation, one can consider the action on the Thurston boundary $\text{PMF}(S)$ of projective measured foliations. From this point of view there derives a complete classification, analogous to the Thurston classification, of subgroups of the mapping class group.

Theorem 5.1 (McCarthy–Papadopoulos [128]). *A subgroup of $\text{Mod}(S)$ is exactly one of the following types:*

- (1) *finite;*
- (2) *infinite irreducible and virtually cyclic;*
- (3) *infinite reducible;*
- (4) *sufficiently large.*

By *sufficiently large* we mean that the subgroup contains two pseudo-Anosov’s with distinct fixed point sets in $\text{PMF}(S)$. These groups contain free groups on two generators.

We apply this theorem to the image $G = \rho(\Gamma)$ of the homomorphism ρ . By the Nielsen Realization Theorem 4.10, if G is finite then it fixes a point $[\sigma]$ in Teichmüller space. Hence, the constant map $f(x) = [\sigma]$ is equivariant and clearly harmonic.

Case (2) arises when G has a finite index subgroup $\langle[\phi]\rangle \simeq \mathbb{Z}$ generated by a pseudo-Anosov $[\phi]$. By Theorem 4.8 this stabilizes a complete Weil–Petersson geodesic $A_{[\phi]} \subset \mathcal{T}(S)$. The corresponding finite index subgroup $\hat{\Gamma} \subset \Gamma$ defines a finite cover $M_{\hat{\Gamma}} \rightarrow M$, and the group of deck transformations then acts on S^1 . Hence, it suffices to find an equivariant harmonic map $M_{\hat{\Gamma}} \rightarrow S^1 \hookrightarrow \mathcal{T}(S)/\langle[\phi]\rangle$. This can be done using the heat equation approach, since equivariance is preserved under the flow (2.29).

In Case (3), G fixes a stratum in the boundary $\partial\mathcal{T}(S)$ isomorphic to a product of lower dimensional Teichmüller spaces. Since the boundary strata are totally geodesically embedded, the problem of finding an energy minimizer to $\overline{\mathcal{T}}(S)$ is reduced to Cases (1), (2), and (4) for lower dimensions.

Finally, we come to Case (4).

Theorem 5.2 (Daskalopoulos–Wentworth [39]). *If $\rho: \Gamma \rightarrow \text{Mod}(S)$ is sufficiently large then it is proper in the sense of Korevaar–Schoen (see Section 3.1.1).*

This is a consequence of Theorem 4.9. Using Theorem 3.7, it follows that there exist equivariant harmonic maps in this case as well. Putting all of these considerations together, we have

Corollary 5.3. *Let $\rho: \pi_1(M) \rightarrow \text{Mod}(S)$ be a homomorphism, where M is a compact Riemannian manifold. Then there exists a finite energy ρ -equivariant harmonic map $f: \tilde{M} \rightarrow \overline{\mathcal{T}}(S)$. Moreover, f is uniformly Lipschitz.*

We note that this statement, apparently stronger than the one appearing in [39], follows from considering the possibilities in Theorem 5.1.

It is certainly expected that uniqueness holds in the corollary under certain assumptions. Some generalization of Theorem 2.33 is needed. Roughly speaking, one expects uniqueness to fail only if the image of f lies in a flat. Alternatively, one should be able to prove a priori that if ρ is sufficiently large then some point of the image of f lies in the interior $\mathcal{T}(S)$. Then the strictly negative curvature implies that the image is a geodesic, which again contradicts the assumption of sufficiently large.

5.1.2 Surface domains. A natural question arises from the statement of Corollary 5.3. Under what conditions does the image of a harmonic map to $\overline{\mathcal{T}}(S)$ actually lie in $\mathcal{T}(S)$? This is an important issue, since if $f(x) \in \mathcal{T}(S)$, then since $\mathcal{T}(S)$ is a manifold f is smooth near the point x . More generally, one would at least like to have control over the size of the singular set (cf. Theorem 2.35).

The first result in this direction is the following

Theorem 5.4 (Wentworth [193]). *Let $\Omega \subset \mathbb{R}^2$ be a bounded domain, and suppose $f: \Omega \rightarrow \overline{\mathcal{T}}(S)$ is energy minimizing with respect to its boundary conditions. If $f(x) \in \mathcal{T}(S)$ for some $x \in \Omega$, then $f(\Omega) \subset \mathcal{T}(S)$.*

To give a rough idea of why this should be the case, we again consider the model for the Weil–Petersson geometry near the boundary $\partial\mathcal{T}(S)$ discussed in Section 4.1.1. Let $f: B_1(0) \rightarrow \overline{\mathbb{M}}$ be a finite energy harmonic map. By Theorem 2.31, f is uniformly Lipschitz. The generalization of (4.2) are the equations

$$\xi \Delta \xi = \frac{3}{4} \xi^6 |\nabla \theta|^2, \quad \text{div}(\xi^6 \nabla \theta) = 0, \quad |\nabla \xi|, \quad \xi^3 |\nabla \theta| \text{ are locally bounded.}$$

Because of the singularities, $\xi(x, y)$ and $\theta(x, y)$ are only weak solutions of these equations. We may assume that f is nonconstant with $f(0) \in \partial\mathbb{M}$. Furthermore, suppose the origin is not a zero of the Hopf differential φ . It is not hard to show that the singular set, i.e. $f^{-1}(\partial\mathbb{M})$ is a leaf of the horizontal foliation of φ . In φ -coordinates (x, y) , one shows with some more analysis of the situation that $\xi(x, y) \sim y$. Then the second equation above becomes essentially $\text{div}(y^6 \nabla \theta) = 0$. This kind of degenerate equation appears in the study of the porous medium equation [107], and one can show that $\theta(x, y)$ itself is locally bounded. Then a scaling argument using the monotonicity formula (2.50) can be used to derive a contradiction.

To go from a regularity result for harmonic maps to $\overline{\mathbb{M}}$ to a result for maps to $\overline{\mathcal{T}}(S)$ requires an approximation of harmonic maps to targets with asymptotically product metrics. This is similar to the discussion of geodesics above. For more details we refer to [193].

The local regularity implies

Corollary 5.5. *Let $\rho: \pi_1(B) \rightarrow \text{Mod}(S)$ be irreducible, where B is a compact Riemann surface. Then there exists a smooth ρ -equivariant harmonic map $f: \tilde{B} \rightarrow \mathcal{T}(S)$. Moreover, if ρ is sufficiently large, then f is unique.*

An interesting potential application of this result pertains to the following

Question. Let B be a closed surface. Does there exist an injective homomorphism $\rho: \pi_1(B) \rightarrow \text{Mod}(S)$ such that the image of ρ consists entirely of pseudo-Anosov's?

Examples of all pseudo-Anosov subgroups of $\text{Mod}(S)$ have been constructed in [195], but these are not surface groups. Such groups, should they exist, would admit minimal surface representations in $\text{Mod}(S)$:

Corollary 5.6. *Let B be a closed surface and $\rho: \pi_1(B) \rightarrow \text{Mod}(S)$. In addition, we assume that for every simple closed essential curve in B , the image by ρ of the associated conjugacy class in $\pi_1(B)$ is pseudo-Anosov. Then there is a conformal harmonic ρ -equivariant map $f: (\tilde{B}, j) \rightarrow \mathcal{T}(S)$ for some complex structure j on B .*

The argument proceeds as in the proof of Proposition 4.13. Note that there is a lower bound, depending only on the genus, of the Weil–Petersson translation length of any pseudo-Anosov (see [39]).

5.1.3 Holomorphic maps from Riemann surfaces. By Proposition 2.13 (see esp. (2.25)), since the Weil–Petersson metric is Kähler, equivariant holomorphic maps from surfaces to $\mathcal{T}(S)$ are examples of energy minimizers; in particular, harmonic maps. These are given by holomorphic curves in $\mathcal{M}(S)$ that are locally liftable to $\mathcal{T}(S)$. Alternatively, consider a family $X \rightarrow B$, where B is a compact Riemann surface, and X is a locally liftable holomorphic fibration of genus p Riemann surfaces. Associated to this is a *monodromy* homomorphism $\rho: \pi_1(B) \rightarrow \text{Mod}(S)$. By Corollary 5.5, if ρ is irreducible there is a ρ -equivariant harmonic map $f: \tilde{B} \rightarrow \mathcal{T}(S)$. In general, this will *not* be holomorphic for any choice of complex structure on B . By the essential uniqueness of the harmonic map, we see that the issue of holomorphicity is a property of the (conjugacy class) of the monodromy representation ρ . Let us call a homomorphism $\rho: \pi_1(B) \rightarrow \text{Mod}(S)$ *holomorphic* if there exists a ρ -equivariant holomorphic map $\tilde{B} \rightarrow \mathcal{T}(S)$.

A simple example occurs when the monodromy has finite image. Then by the Nielsen realization theorem, ρ fixes a point in $\mathcal{T}(S)$. In particular, there is a (constant) holomorphic map. In terms of the family $X \rightarrow B$, this is precisely the case where the lift of the fibration $p^*X \rightarrow \hat{B}$ to some finite cover $p: \hat{B} \rightarrow B$ is trivial. Such a fibration is called *isotrivial*.

The harmonic map point of view provides a tool to study holomorphic families. Here is one property:

Theorem 5.7. *If $\rho: \pi_1(B) \rightarrow \text{Mod}(S)$ is holomorphic and nonisotrivial, then ρ is sufficiently large.*

Proof. Suppose not. By the classification of subgroups of the mapping class group Theorem 5.1, ρ is either reducible or virtually cyclic. In the former case, there is a proper totally geodesic stratum $S \subset \partial\mathcal{T}(S)$ that is invariant under ρ . Since projection to S from the interior $\mathcal{T}(S)$ is strictly distance decreasing, the geodesic homotopy of f to S is both ρ -equivariant and strictly energy decreasing. This contradicts the fact that u is the energy minimizer. If ρ is virtually cyclic, then the energy minimizer maps onto a geodesic. Since the image is one dimensional, this contradicts holomorphicity. \square

The following is also a consequence of the uniqueness of harmonic maps to $\overline{\mathcal{T}}(S)$ discussed in the proof above. This is sometimes called the *rigidity* theorem.

Theorem 5.8. *Holomorphic families with the same monodromy (up to conjugation) are equivalent.*

The main finiteness result is the following

Theorem 5.9 (Arakelov [8], Paršin [155]). *Fix a closed Riemann surface B , and let $\text{Mod}(S)$ denote the mapping class group of a compact surface of genus $p \geq 2$. Then there are at most finitely many conjugacy classes of non-isotrivial holomorphic homomorphisms $\rho: \pi_1(B) \rightarrow \text{Mod}(S)$.*

We note that this can be extended to the case where B is a Riemann surface with punctures. The punctures correspond to singularities in the surface fibration, and in the holomorphic case the local monodromy around the punctures is pseudoperiodic. Finite energy maps always exist in this case (see [36]).

The key to Theorem 5.9 is a uniform bound on the energy. Since $\mathcal{T}(S)$ has holomorphic sectional curvature bounded above by a negative constant (see Theorem 4.1), Royden's version of the Yau-Schwartz lemma implies that if $f: \tilde{B} \rightarrow \mathcal{T}(S)$ is holomorphic, then

$$f^* ds_{wp}^2 \leq C ds_{\tilde{B}}^2,$$

for a uniform constant C (see [160]). In particular, by (2.25), the energy of a holomorphic map is uniformly bounded. Since by Proposition 2.17 the Lipschitz constant of harmonic maps is bounded by the total energy, a sequence of holomorphic maps to $\mathcal{M}(S)$ is necessarily equicontinuous (see also [74]). This allows one to construct convergent subsequences for the maps $\mathcal{M}(S)$. As in the argument in Section 4.1.3 there is the issue of lifting the limiting map. In this way, one derives a contradiction to the existence of infinitely many distinct conjugacy classes of holomorphic ρ . For a fuller account of this approach to the Arakelov–Paršin theorem, we refer to [100] and [85].

5.2 Superrigidity

In this final section we briefly describe how equivariant harmonic map theory can be used to study homomorphisms of fundamental groups of compact manifolds to the mapping class group. The link between superrigidity and harmonic maps uses a technique which can be traced back to Bochner and Calabi–Weil and was first fully utilized in connection with the Margulis superrigidity theorem. In fact, as mentioned earlier, many of the ideas presented here were inspired by the attempt to give a harmonic maps proof of superrigidity. In Section 5.2.1 we state the Ivanov–Farb–Kaimanovich–Masur theorem for homomorphisms of superrigid lattices into mapping class groups. In Section 5.2.2 we describe two approaches in generalizing harmonic maps by allowing the domain to be singular as well. The first is the analytic approach along the lines for smooth domains, the second is the combinatorial approach. As an application one can prove a statement on the non-Archimedean superirigidity of lattices in mapping class groups.

5.2.1 The Ivanov–Farb–Kaimanovich–Masur theorem. Harvey originally asked whether the mapping class group could be isomorphic to a lattice in a symmetric space [80]. This was shown not to be the case by Ivanov [88], [89]. For some of the similarities and differences between $\text{Mod}(S)$ and arithmetic lattices, see [53], [88], [127] and Ivanov’s survey article [90]. Indeed, a stronger statement is true:

Theorem 5.10. *Let Γ be a cocompact lattice in any symmetric space with nonpositive curvature other than the real or complex hyperbolic spaces. Then any homomorphism $\Gamma \rightarrow \text{Mod}(S)$ has finite image.*

For symmetric spaces of rank ≥ 2 this result is due to Farb–Masur [55], following earlier work of Kaimanovich–Masur [102]. Ivanov has announced an independent proof. Bestvina–Fujiwara [17] gave a proof using bounded cohomology, and for hermitian symmetric spaces an independent proof can be found in Hain [76]. Using the method of [30], [101], [139] the remaining rank 1 cases were proven by S.-K. Yeung [213].

Geometric superrigidity uses harmonic maps to prove results of this type. The basic philosophy is to show that equivariant harmonic maps $f: G/K \rightarrow N$, where G/K is a symmetric space of higher rank and N has nonpositive curvature, would necessarily be totally geodesic. Recall from Section 2.2.1 that the harmonic map equations are of the form $\text{Tr } \nabla df = 0$, whereas the equations for a totally geodesic map are $\nabla df = 0$. Curvature conditions must be used to show that the stronger (overdetermined) set of equations are automatically satisfied. One then attempts to use geometric considerations to rule out the existence of nonconstant totally geodesic maps.

To give a simple example of how this might come about, consider the following result.

Theorem 5.11 (Eells–Sampson [51]). *If $f: \tilde{M} \rightarrow N$ is an equivariant harmonic map, N is a Riemannian manifold of nonpositive curvature, and M is closed compact with non-negative Ricci curvature, then f is totally geodesic. If the Ricci curvature of M is positive at one point, f is constant. If the sectional curvature of N is negative then f is either constant or maps to a geodesic.*

Indeed, the statement easily follows by integrating both sides of the Bochner formula (2.32) and using the divergence theorem. When the domain does not satisfy this curvature restriction, the proof fails. Nevertheless, more sophisticated forms of the Bochner formulas have been derived in the case of domains with Einstein metrics, or more generally, certain parallel tensors. For more details, we refer to [30], [101], [139].

In light of Corollary 5.3, one is tempted to prove Theorem 5.10 using harmonic maps to $\tilde{\mathcal{F}}(S)$. The difficulty is in the singular nature of the NPC space $\tilde{\mathcal{F}}(S)$. However, the idea that these techniques could be generalized to singular space targets is one of the major contributions of [75]. The argument based on the Bochner formula given above continues to be valid, so long as the singular set of f is relatively small, e.g. has codimension at least 2, so that the integration by parts needed to apply the divergence theorem holds. All of this is motivation to extend the regularity result of Theorem 5.4 to higher dimensional domains.

5.2.2 Harmonic maps from singular domains. Thus far we have discussed the theory of harmonic maps from smooth domains into (possibly singular) metric space targets. These included singular surfaces, \mathbb{R} -trees, and the Weil–Petersson completion of Teichmüller space. In this section we sketch two generalizations of this study to the case where the domain is also allowed to be singular.

We start with an analytic approach closely related to the techniques discussed above. Let Σ be a finite 2-dimensional simplicial complex. The restriction to two dimensions is not essential and most of the following results hold in general. It is important, however, to assume that Σ is *admissible* (cf. [25], [49]), meaning that it satisfies the following conditions:

- (1) Every simplex is contained in a face (i.e. a 2-simplex).
- (2) Every pair of faces can be joined by a sequence of pairwise adjacent faces.
- (3) Σ has no boundary, i.e. every edge is contained in at least two faces.
- (4) Σ is flat in that every open face is isometric to an equilateral triangle in \mathbb{R}^2 .

We also allow ourselves a choice w of weights $w(F) > 0$ for each face. This is an important technical point. Given an NPC space (X, d) and a map $f: \Sigma \rightarrow X$, define the w -energy

$$E_w(f) = \frac{1}{2} \sum_F w(F) \int_F |\nabla f|^2(x) dx,$$

where the sum is over all faces F of Σ . A map f is called w -harmonic if it is locally energy minimizing among all maps of finite w -energy. As before, we also consider the equivariant theory, where f is a map from the universal cover $\tilde{\Sigma}$ of Σ that is equivariant with respect to a homomorphism $\rho: \Gamma = \pi_1(\Sigma) \rightarrow \text{Iso}(X)$. The existence Theorem 3.1 then holds for domains Σ as well (cf. [49], [37]).

In the following, we shall assume a fixed choice of weights and omit w from the notation. Perhaps the most interesting feature of harmonic maps from simplicial domains is the Hölder continuity. This was first proven by J. Chen [25] for flat metrics and in a more general context by Eells–Fuglede [49]. The following stronger version describes the singular behavior near the vertices.

Theorem 5.12 (Daskalopoulos–Mese [37]). *Let $f: \Sigma \rightarrow X$ be harmonic. Then for domains $U \subset\subset \Omega \subset \Sigma$ the following holds:*

- (1) *f is Lipschitz continuous on U away from the vertices of Σ , where the Lipschitz constant depends only on U , the total energy on Ω , and the distance to the vertex set.*
- (2) *Let v be a vertex with $\alpha = \text{ord}_v(f)$, where the order is defined as in (2.51). Then there exists $r_0 > 0$ and C depending only on the energy of f such that*

$$\sup_{x \in B_r(v)} |\nabla f|^2(x) \leq Cr^{2\alpha-2}$$

for all $0 < r \leq r_0$.

The important point here is that, unlike the case of smooth domains, α need not be ≥ 1 . One application of Theorem 5.12 is the compactification of character varieties for arbitrarily finitely presented groups along the lines of Theorem 3.19. Indeed, one can always realize such a group as the fundamental group of an admissible 2-complex. Other potential consequences use the notion of a Hopf differential. Clearly, for energy minimizers, $\varphi = \text{Hopf}(f)$ is a holomorphic quadratic differential on the interior of each face (cf. Section 3.1.2). For points x on an edge e , we have the following *balancing condition*:

$$\text{Im} \sum_F \varphi_F(x) = 0,$$

where the sum is over all faces F adjacent to e at x . An important open question is whether zeros of φ can accumulate along the edges. If not, then the Hopf differentials of w -harmonic maps define *geometric* or *track foliations* on Σ (cf. [18], [46], [116]). Another important issue is the asymptotic behavior of the induced foliation on $\tilde{\Sigma}$. More generally, one might ask under what conditions one can generalize to this setting the results for surface groups discussed previously in this chapter.

We now return to the relationship between regularity and rigidity. We have the following

Theorem 5.13 (Daskalopoulos–Mese [37]). *Let $f: \Omega \subset \Sigma \rightarrow X$ be energy minimizing, where X is a smooth manifold of nonpositive curvature. For any $x \in \Omega$ which is not a vertex, then there is a neighborhood U of x such that for any face F the restriction of f to $\bar{F} \cap U$ is smooth.*

Using this, one has a nontrivial generalization of Theorem 5.11 to the case of singular domains:

Theorem 5.14 (Daskalopoulos–Mese [38]). *Suppose Σ is an admissible 2-complex and X is a complete Riemannian manifold of nonpositive curvature. If $f: \Sigma \rightarrow X$ is harmonic and $|\nabla f|^2$ bounded, then f is totally geodesic on each simplex of X . If the sectional curvature of X is strictly negative, then either f is constant or it maps each simplex to a geodesic.*

This is a kind of rigidity result for the group $\Gamma = \pi_1(\Sigma)$, and a combinatorial version was first proven by M.-T. Wang (see below). The result follows by the Bochner formula (2.32), the vanishing of the Ricci curvature on the domain, and the fact that $|\nabla f|^2$ allows us to integrate by parts. Global boundedness of the energy density is guaranteed by a combinatorial condition on Σ . Namely, the first eigenvalue of the discrete Laplacian on the link of every vertex with the induced weights should be $\geq 1/2$ (see [38]). This condition is a generalization of the notion of p-adic curvature that first appeared in the work of Garland (cf. [68]).

The second approach discretizes the notion of an energy minimizer. Let Σ be an admissible 2-complex and X an NPC space as above and $\rho: \Gamma \rightarrow \text{Iso}(X)$ a homomorphism. Given a system of weights on the faces of Σ there is a standard way to induce weights on the lower dimensional simplices. For example, the weight of an edge is the sum of the weights of adjacent faces. Let $\Sigma^i, \tilde{\Sigma}^i$ denote the i -skeletons. Given a ρ -equivariant map $f: \tilde{\Sigma}^0 \rightarrow X$ define its energy by

$$E_{\text{comb}}(f) = \frac{1}{2} \sum_{e_{xy} \in \Sigma^1} w(e_{xy}) d^2(f(\tilde{x}), f(\tilde{y})),$$

where e_{xy} denotes an edge with adjacent vertices x and y , and \tilde{x}, \tilde{y} are adjacent vertices of a lift of e_{xy} to $\tilde{\Sigma}$. We say that f is a ρ -equivariant combinatorial harmonic map if it minimizes $E_{\text{comb}}(f)$. Under the assumption that X is locally compact and that $\rho(\Gamma)$ does not fix a point in ∂X one can prove the existence of combinatorial harmonic maps (see Wang [191], [192]). Furthermore, assuming the first eigenvalue of the combinatorial Laplacian of the link of every vertex with the induced weights is $> 1/2$, one can deduce rigidity results as in the first approach. This can be used to deduce non-Archimedean generalizations of Theorem 5.10 [171].

References

- [1] W. Abikoff, *The real analytic theory of Teichmüller space*. Lecture Notes in Math. 820, Springer-Verlag, Berlin 1980. [35](#), [38](#), [39](#), [42](#)
- [2] W. Abikoff, Oswald Teichmüller. *Math. Intelligencer* 8 (1986), 8–16, 33. [35](#)
- [3] L. V. Ahlfors, On quasiconformal mappings. *J. Analyse Math.* 3 (1954), 1–58; correction, 207–208. [42](#)
- [4] L. V. Ahlfors, Curvature properties of Teichmüller’s space. *J. Analyse Math.* 9 (1961/1962), 161–176. [78](#)
- [5] L. V. Ahlfors, *Lectures on quasiconformal mappings*. Van Nostrand, Inc., New York 1966. [40](#)
- [6] L. V. Ahlfors and L. Sario, *Riemann surfaces*. Princeton Math. Ser. 26, Princeton University Press, Princeton, N.J., 1960.
- [7] F. J. Almgren, Jr., *Almgren’s big regularity paper*. World Sci. Monogr. Ser. Math. 1, World Scientific, Inc., River Edge, N.J., 2000. [61](#)
- [8] S. J. Arakelov, Families of algebraic curves with fixed degeneracies. *Izv. Akad. Nauk SSSR Ser. Mat.* 35 (1971), 1269–1293; English transl. *Math. USSR, Izv.* 5 (1971), 1277–1302. [94](#)
- [9] W. Ballmann, M. Gromov, and V. Schroeder, *Manifolds of nonpositive curvature*. Progr. Math. 61, Birkhäuser, Boston 1985. [84](#)
- [10] W. Ballmann and J. Świątkowski, On L^2 -cohomology and property (T) for automorphism groups of polyhedral cell complexes. *Geom. Funct. Anal.* 7 (1997), 615–645.
- [11] A. F. Beardon, *The geometry of discrete groups*. Grad. Texts in Math. 91, Springer-Verlag, New York 1983. [43](#)
- [12] L. Bers, Simultaneous uniformization. *Bull. Amer. Math. Soc.* 66 (1960), 94–97. [89](#)
- [13] L. Bers, Quasiconformal mappings and Teichmüller’s theorem. In *Analytic functions*, Princeton University Press, Princeton, N.J., 1960, 89–119. [42](#)
- [14] L. Bers, An extremal problem for quasiconformal mappings and a theorem by Thurston. *Acta Math.* 141 (1978), 73–98. [84](#)
- [15] L. Bers and L. Ehrenpreis, Holomorphic convexity of Teichmüller spaces. *Bull. Amer. Math. Soc.* 70 (1964), 761–764. [90](#)
- [16] M. Bestvina, Degenerations of the hyperbolic space. *Duke Math. J.* 56 (1988), 143–161. [76](#)
- [17] M. Bestvina and K. Fujiwara, Bounded cohomology of subgroups of mapping class groups. *Geom. Topol.* 6 (2002), 69–89. [95](#)
- [18] B. H. Bowditch, Group actions on trees and dendrons. *Topology* 37 (1998), 1275–1298. [58](#), [97](#)
- [19] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*. Grundlehren Math. Wiss. 319, Springer-Verlag, Berlin 1999. [64](#), [80](#)
- [20] J. Brock, The Weil-Petersson visual sphere. *Geom. Dedicata* 115 (2005), 1–18. [91](#)
- [21] J. Brock and D. Margalit, Weil–Petersson isometries via the pants complex. *Proc. Amer. Math. Soc.* 135 (2007), 795–803. [83](#)

- [22] J. Brock, H. Masur, and Y. Minsky. Asymptotics of Weil–Petersson geodesics. In preparation. [84](#), [91](#)
- [23] K.-C. Chang, W. Y. Ding, and R. Ye. Finite-time blow-up of the heat flow of harmonic maps from surfaces. *J. Differential Geom.* 36 (1992), 507–515. [49](#)
- [24] B.-Y. Chen, Equivalence of the Bergman and Teichmüller metrics on Teichmüller spaces. Preprint, 2004. [90](#)
- [25] J. Chen, On energy minimizing mappings between and into singular spaces. *Duke Math. J.* 79 (1995), 77–99. [96](#), [97](#)
- [26] S. S. Chern, *Complex manifolds without potential theory*. 2nd edn., Universitext, Springer-Verlag, New York 1979. [70](#)
- [27] T. Chu, The Weil-Petersson metric in the moduli space. *Chinese J. Math.* 4 (1976), 29–51. [78](#)
- [28] D. Cooper. Degenerations of representations into $SL_2(\mathbb{C})$. Preprint 1994. [75](#)
- [29] K. Corlette, Flat G -bundles with canonical metrics. *J. Differential Geom.* 28 (1988), 361–382. [64](#), [65](#), [71](#)
- [30] K. Corlette, Archimedean superrigidity and hyperbolic geometry. *Ann. of Math.* (2) 135 (1992), 165–182. [64](#), [95](#), [96](#)
- [31] J.-M. Coron and J.-M. Ghidaglia, Explosion en temps fini pour le flot des applications harmoniques. *C. R. Acad. Sci. Paris Sér. I Math.* 308 (1989), 339–344. [49](#)
- [32] M. Culler and J. W. Morgan, Group actions on \mathbb{R} -trees. *Proc. London Math. Soc.* (3) 55 (1987), 571–604. [64](#), [65](#)
- [33] M. Culler and P. B. Shalen, Varieties of group representations and splittings of 3-manifolds. *Ann. of Math.* (2) 117 (1983), 109–146. [38](#), [72](#)
- [34] G. Daskalopoulos, S. Dostoglou, and R. Wentworth, Character varieties and harmonic maps to \mathbb{R} -trees. *Math. Res. Lett.* 5 (1998), 523–533. [65](#), [76](#), [77](#)
- [35] G. Daskalopoulos, S. Dostoglou, and R. Wentworth, On the Morgan-Shalen compactification of the $SL(2, \mathbb{C})$ character varieties of surface groups. *Duke Math. J.* 101 (2000), 189–207. [67](#), [68](#), [72](#)
- [36] G. Daskalopoulos, L. Katzarkov, and R. Wentworth, Harmonic maps to Teichmüller space. *Math. Res. Lett.* 7 (2000), 133–146. [94](#)
- [37] G. Daskalopoulos and C. Mese. Harmonic maps from two-complexes. *Comm. Anal. Geom.* 14 (3) (2006), 497–549. [97](#), [98](#)
- [38] G. Daskalopoulos and C. Mese. Harmonic maps from a simplicial complexes and geometric rigidity. Preprint. [98](#)
- [39] G. Daskalopoulos and R. Wentworth, Classification of Weil-Petersson isometries. *Amer. J. Math.* 125 (2003), 941–975. [80](#), [83](#), [84](#), [91](#), [92](#), [93](#)
- [40] P. Deligne and D. Mumford, The irreducibility of the space of curves of given genus. *Inst. Hautes Études Sci. Publ. Math.* (1969), 75–109. [82](#)
- [41] E. D’Hoker and D. H. Phong, Multiloop amplitudes for the bosonic Polyakov string. *Nuclear Phys. B* 269 (1986), 205–234. [44](#)
- [42] S. K. Donaldson, A new proof of a theorem of Narasimhan and Seshadri. *J. Differential Geom.* 18 (1983), 269–277. [70](#)

- [43] S. K. Donaldson, Anti self-dual Yang-Mills connections over complex algebraic surfaces and stable vector bundles. *Proc. London Math. Soc.* (3) 50 (1985), 1–26. [64](#), [70](#)
- [44] S. K. Donaldson, Twisted harmonic maps and the self-duality equations. *Proc. London Math. Soc.* (3) 55 (1987), 127–131. [71](#)
- [45] J. Douglas, Minimal surfaces of higher topological structure. *Ann. of Math.* (2) 40 (1939), 205–298. [69](#)
- [46] M. J. Dunwoody, Groups acting on \mathbb{R} -trees. *Comm. Algebra* 19 (1991), 2125–2136. [97](#)
- [47] C. J. Earle and J. Eells, A fibre bundle description of Teichmüller theory. *J. Differential Geometry* 3 (1969), 19–43. [53](#), [81](#)
- [48] D. G. Ebin, The manifold of Riemannian metrics. In *Global Analysis* (Berkeley, Calif., 1968), Proc. Sympos. Pure Math. 15, Amer. Math. Soc., Providence, R.I., 1970, 11–40. [44](#), [53](#), [81](#)
- [49] J. Eells and B. Fuglede, *Harmonic maps between Riemannian polyhedra*. Cambridge Tracts in Math. 142, Cambridge University Press, Cambridge 2001. [96](#), [97](#)
- [50] J. Eells and L. Lemaire, *Two reports on harmonic maps*. World Scientific, River Edge, N.J., 1995. [35](#), [48](#)
- [51] J. Eells and J. H. Sampson, Harmonic mappings of Riemannian manifolds. *Amer. J. Math.* 86 (1964), 109–160. [48](#), [49](#), [53](#), [64](#), [96](#)
- [52] J. Eells and J. C. Wood, Restrictions on harmonic maps of surfaces. *Topology* 15 (1976), 263–266. [52](#)
- [53] B. Farb, A. Lubotzky, and Y. Minsky, Rank-1 phenomena for mapping class groups. *Duke Math. J.* 106 (2001), 581–597. [95](#)
- [54] B. Farb and D. Margalit, A primer on mapping class groups. In preparation. [81](#)
- [55] B. Farb and H. Masur, Superrigidity and mapping class groups. *Topology* 37 (1998), 1169–1176. [95](#)
- [56] B. Farb and M. Wolf, Harmonic splittings of surfaces. *Topology* 40 (2001), 1395–1414. [66](#), [67](#), [68](#)
- [57] A. Fathi, F. Laudenbach and V. Poénaru (eds.), *Travaux de Thurston sur les surfaces*. *Astérisque* 66–67 (1979). [57](#), [59](#)
- [58] W. Fenchel, Estensioni di gruppi discontinui e trasformazioni periodiche delle superficie. *Atti Accad. Naz. Lincei. Rend. Cl. Sci. Fis. Mat. Nat.* (8) 5 (1948), 326–329. [86](#)
- [59] W. Fenchel, Remarks on finite groups of mapping classes. *Mat. Tidsskr. B.* 1950 (1950), 90–95 (in Danish). [86](#)
- [60] A. E. Fischer and A. J. Tromba, On a purely “Riemannian” proof of the structure and dimension of the unramified moduli space of a compact Riemann surface. *Math. Ann.* 267 (1984), 311–345. [44](#)
- [61] A. E. Fischer and A. J. Tromba, A new proof that Teichmüller space is a cell. *Trans. Amer. Math. Soc.* 303 (1987), 257–262. [87](#)
- [62] A. Fletcher and V. Markovic, Infinite dimensional Teichmüller spaces. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2008. [35](#)
- [63] O. Forster, *Lectures on Riemann surfaces*. Grad. Texts in Math. 81, Springer-Verlag, New York 1991. [38](#)

- [64] F. P. Gardiner, Measured foliations and the minimal norm property for quadratic differentials. *Acta Math.* 152 (1984), 57–76. [68](#), [69](#)
- [65] F. P. Gardiner, *Teichmüller theory and quadratic differentials*. John Wiley & Sons, New York 1987. [69](#)
- [66] F. P. Gardiner and N. Lakic, *Quasiconformal Teichmüller theory*. Math. Surveys Monogr. 76, Amer. Math. Soc., Providence, R.I., 2000. [42](#)
- [67] F. P. Gardiner and H. Masur, Extremal length geometry of Teichmüller space. *Complex Variables Theory Appl.* 16 (1991), 209–237. [89](#)
- [68] H. Garland, p -adic curvature and the cohomology of discrete subgroups of p -adic groups. *Ann. of Math. (2)* 97 (1973), 375–423. [98](#)
- [69] M. Gerstenhaber and H. E. Rauch, On extremal quasi-conformal mappings. I; II. *Proc. Nat. Acad. Sci. U.S.A.* 40 (1954), 808–812; 991–994. [55](#)
- [70] W. M. Goldman, Representations of fundamental groups of surfaces. In *Geometry and topology* (College Park, Md., 1983/84), Lecture Notes in Math. 1167, Springer-Verlag, Berlin 1985, 95–117. [38](#)
- [71] W. M. Goldman and J. J. Millson, The deformation theory of representations of fundamental groups of compact Kähler manifolds. *Inst. Hautes Études Sci. Publ. Math.* (1988), 43–96. [38](#)
- [72] W. M. Goldman, Trace coordinates on Fricke space of some simple hyperbolic surfaces. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2008. [38](#)
- [73] W. M. Goldman and R. Wentworth, Energy of twisted harmonic maps of Riemann surfaces. In *Proceedings of the Ahlfors-Bers Colloquium* (Ann Arbor, Michigan, 2005), to appear. [88](#), [89](#)
- [74] H. Grauert and H. Reckziegel, Hermitesche Metriken und normale Familien holomorpher Abbildungen. *Math. Z.* 89 (1965), 108–125. [94](#)
- [75] M. Gromov and R. Schoen, Harmonic maps into singular spaces and p -adic superrigidity for lattices in groups of rank 1. *Inst. Hautes Études Sci. Publ. Math.* 76 (1992), 165–246. [54](#), [60](#), [61](#), [62](#), [96](#)
- [76] R. Hain, Locally symmetric families of curves and Jacobians. In *Moduli of curves and abelian varieties*, Aspects Math. E33, Vieweg, Braunschweig 1999, 91–108. [95](#)
- [77] R. S. Hamilton, Extremal quasiconformal mappings with prescribed boundary values. *Trans. Amer. Math. Soc.* 138 (1969), 399–406. [42](#)
- [78] R. S. Hamilton, *Harmonic maps of manifolds with boundary*. Lecture Notes in Math. 471, Springer-Verlag, Berlin 1975. [49](#)
- [79] P. Hartman, On homotopic harmonic maps. *Canad. J. Math.* 19 (1967), 673–687. [51](#)
- [80] 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. [95](#)
- [81] N. J. Hitchin, The self-duality equations on a Riemann surface. *Proc. London Math. Soc.* (3) 55 (1987), 59–126. [70](#), [71](#), [72](#), [73](#)
- [82] N. Hitchin, Stable bundles and integrable systems. *Duke Math. J.* 54 (1987), 91–114. [72](#)
- [83] N. J. Hitchin, Lie groups and Teichmüller space. *Topology* 31 (1992), 449–473. [70](#)

- [84] J. Hubbard and H. Masur, Quadratic differentials and foliations. *Acta Math.* 142 (1979), 221–274. [58](#)
- [85] Y. Iwayoshi and H. Shiga, A finiteness theorem for holomorphic families of Riemann surfaces. In *Holomorphic functions and moduli*, vol. II (Berkeley, CA, 1986), Math. Sci. Res. Inst. Publ. 11, Springer-Verlag, New York 1988, 207–219. [94](#)
- [86] Y. Iwayoshi and M. Taniguchi, *An introduction to Teichmüller spaces*. Springer-Verlag, Tokyo 1992. [35](#), [38](#), [40](#), [42](#), [43](#)
- [87] T. Ishihara, A mapping of Riemannian manifolds which preserves harmonic functions. *J. Math. Kyoto Univ.* 19 (1979), 215–229. [49](#)
- [88] N. V. Ivanov, Algebraic properties of the Teichmüller modular group. *Dokl. Akad. Nauk SSSR* 275 (1984), 786–789; English transl. *Soviet Math. Dokl.* 29 (1984), 288–291. [95](#)
- [89] N. V. Ivanov, Teichmüller modular groups and arithmetic groups. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* 167 (1988), 95–110, 190–191; English transl. *J. Soviet Math.* 52 (1990), 2809–2818. [95](#)
- [90] N. V. Ivanov, Mapping class groups. In *Handbook of geometric topology*, North-Holland, Amsterdam 2002, 523–633. [95](#)
- [91] N. V. Ivanov, A short proof of non-Gromov hyperbolicity of Teichmüller spaces. *Ann. Acad. Sci. Fenn. Math.* 27 (2002), 3–5. [78](#)
- [92] A. Jaffe and C. Taubes, *Vortices and monopoles*. Progr. Phys. 2, Birkhäuser, Boston 1980. [73](#)
- [93] J. Jost, *Nonlinear methods in Riemannian and Kählerian geometry*. DMV Seminar 10, Birkhäuser, Basel 1988. [51](#)
- [94] J. Jost, Harmonic maps and curvature computations in Teichmüller theory. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 16 (1991), 13–46. [87](#)
- [95] J. Jost, *Two-dimensional geometric variational problems*. Pure Appl. Math., John Wiley & Sons, Chichester 1991. [51](#), [69](#), [78](#)
- [96] J. Jost, Equilibrium maps between metric spaces. *Calc. Var. Partial Differential Equations* 2 (1994), 173–204. [60](#), [61](#)
- [97] J. Jost, *Nonpositive curvature: geometric and analytic aspects*. Lectures Math. ETH Zürich, Birkhäuser, Basel 1997. [65](#)
- [98] J. Jost and R. Schoen, On the existence of harmonic diffeomorphisms. *Invent. Math.* 66 (1982), 353–359. [52](#)
- [99] J. Jost and S.-T. Yau, Harmonic maps and group representations. In *Differential geometry*, Pitman Monogr. Surveys Pure Appl. Math. 52, Longman Sci. Tech., Harlow 1991, 241–259. [65](#)
- [100] J. Jost and S.-T. Yau, Harmonic mappings and algebraic varieties over function fields. *Amer. J. Math.* 115 (1993), 1197–1227. [94](#)
- [101] J. Jost and S.-T. Yau, Harmonic maps and superrigidity. In *Tsing Hua lectures on geometry & analysis* (Hsinchu, 1990–1991), International Press, Cambridge, MA, 1997, 213–246. [95](#), [96](#)
- [102] V. A. Kaimanovich and H. Masur, The Poisson boundary of the mapping class group. *Invent. Math.* 125 (1996), 221–264. [95](#)

- [103] J. L. Kazdan and F. W. Warner, Curvature functions for open 2-manifolds. *Ann. of Math.* (2) 99 (1974), 203–219. [73](#)
- [104] L. Keen, Collars on 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, 263–268. [79](#)
- [105] S. P. Kerckhoff, The Nielsen realization problem. *Ann. of Math.* (2) 117 (1983), 235–265. [85](#), [86](#)
- [106] H. Kneser, Die kleinste Bedeckungszahl innerhalb einer Klasse von Flächenabbildungen. *Math. Ann.* 103 (1930), 347–358. [52](#)
- [107] H. Koch, Non-euclidean singular integrals and the porous medium equation. Habilitation thesis, 1999. [92](#)
- [108] K. Kodaira, *Complex manifolds and deformation of complex structures*. Grundlehren Math. Wiss. 283, Springer-Verlag, Berlin 1986. [45](#)
- [109] N. J. Korevaar and R. M. Schoen, Sobolev spaces and harmonic maps for metric space targets. *Comm. Anal. Geom.* 1 (1993), 561–659. [54](#), [60](#), [61](#), [63](#), [66](#)
- [110] N. J. Korevaar and R. M. Schoen, Global existence theorems for harmonic maps to non-locally compact spaces. *Comm. Anal. Geom.* 5 (1997), 333–387. [54](#), [60](#), [63](#), [64](#), [66](#), [74](#)
- [111] N. J. Korevaar and R. M. Schoen, Global existence theorems for harmonic maps: finite rank spaces and an approach to rigidity for smooth actions. Preprint. [66](#)
- [112] S. Kravetz, On the geometry of Teichmüller spaces and the structure of their modular groups. *Ann. Acad. Sci. Fenn. Ser. A I No.* 278 (1959). [78](#)
- [113] E. Kuwert, Harmonic maps between flat surfaces with conical singularities. *Math. Z.* 221 (1996), 421–436. [55](#), [56](#)
- [114] F. Labourie, Existence d’applications harmoniques tordues à valeurs dans les variétés à courbure négative. *Proc. Amer. Math. Soc.* 111 (1991), 877–882. [65](#)
- [115] M. L. Leite, Harmonic mappings of surfaces with respect to degenerate metrics. *Amer. J. Math.* 110 (1988), 399–412. [55](#)
- [116] G. Levitt and F. Paulin, Geometric group actions on trees. *Amer. J. Math.* 119 (1997), 83–102. [97](#)
- [117] P. Li and L.-F. Tam, Uniqueness and regularity of proper harmonic maps. *Ann. of Math.* (2) 137 (1993), 167–201. [34](#)
- [118] P. Li and L.-F. Tam, Uniqueness and regularity of proper harmonic maps. II. *Indiana Univ. Math. J.* 42 (1993), 591–635. [34](#)
- [119] M. Linch, On metrics in Teichmüller spaces. Ph.D. Thesis, Columbia University, 1971. [78](#)
- [120] K. Liu, X. Sun, and S.-T. Yau, Canonical metrics on the moduli space of Riemann surfaces. I; II. *J. Differential Geom.* 68 (2004), 571–637; 69 (2005), 163–216. [90](#)
- [121] C. Maclachlan, Modulus space is simply-connected. *Proc. Amer. Math. Soc.* 29 (1971), 85–86. [82](#)
- [122] V. Markovic, Realization of the mapping class group by homeomorphisms. Preprint 2005. [85](#)

- [123] H. Masur, On a class of geodesics in Teichmüller space. *Ann. of Math.* (2) 102 (1975), 205–221. [78](#)
- [124] H. Masur, Extension of the Weil-Petersson metric to the boundary of Teichmüller space. *Duke Math. J.* 43 (1976), 623–635. [78](#), [79](#), [80](#)
- [125] H. A. Masur and M. Wolf, Teichmüller space is not Gromov hyperbolic. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 20 (1995), 259–267. [78](#)
- [126] H. Masur and M. Wolf, The Weil-Petersson isometry group. *Geom. Dedicata* 93 (2002), 177–190. [83](#)
- [127] J. McCarthy, A “Tits-alternative” for subgroups of surface mapping class groups. *Trans. Amer. Math. Soc.* 291 (1985), 583–612. [95](#)
- [128] J. McCarthy and A. Papadopoulos, Dynamics on Thurston’s sphere of projective measured foliations. *Comment. Math. Helv.* 64 (1989), 133–166. [91](#)
- [129] J. D. McCarthy and A. Papadopoulos, The visual sphere of Teichmüller space and a theorem of Masur-Wolf. *Ann. Acad. Sci. Fenn. Math.* 24 (1999), 147–154. [78](#), [83](#)
- [130] J. D. McCarthy and A. Papadopoulos, The mapping class group and a theorem of Masur-Wolf. *Topology Appl.* 96 (1999), 75–84. [78](#)
- [131] C. Mese, The curvature of minimal surfaces in singular spaces. *Comm. Anal. Geom.* 9 (2001), 3–34. [56](#)
- [132] C. Mese, Harmonic maps between surfaces and Teichmüller spaces. *Amer. J. Math.* 124 (2002), 451–481. [56](#), [60](#), [61](#)
- [133] C. Mese, Harmonic maps into spaces with an upper curvature bound in the sense of Alexandrov. *Math. Z.* 242 (2002), 633–661. [56](#)
- [134] C. Mese, Uniqueness theorems for harmonic maps into metric spaces. *Commun. Contemp. Math.* 4 (2002), 725–750. [61](#), [64](#)
- [135] C. Mese, A variational construction of the Teichmüller map. *Calc. Var. Partial Differential Equations* 21 (2004), 15–46. [56](#)
- [136] J. Milnor, On the existence of a connection with curvature zero. *Comment. Math. Helv.* 32 (1958), 215–223. [73](#)
- [137] Y. N. Minsky, Harmonic maps, length, and energy in Teichmüller space. *J. Differential Geom.* 35 (1992), 151–217. [53](#)
- [138] Y. Miyahara, On some properties of a Teichmüller mapping. *TRU Math.* 4 (1968), 36–43. [55](#)
- [139] N. Mok, Y. T. Siu, and S.-K. Yeung, Geometric superrigidity. *Invent. Math.* 113 (1993), 57–83. [95](#), [96](#)
- [140] J. W. Morgan and J.-P. Otal, Relative growth rates of closed geodesics on a surface under varying hyperbolic structures. *Comment. Math. Helv.* 68 (1993), 171–208. [67](#), [68](#)
- [141] J. W. Morgan and P. B. Shalen, Valuations, trees, and degenerations of hyperbolic structures. I. *Ann. of Math.* (2) 120 (1984), 401–476. [75](#), [77](#)
- [142] J. W. Morgan and P. B. Shalen, Free actions of surface groups on \mathbb{R} -trees. *Topology* 30 (1991), 143–154. [58](#)
- [143] S. Morita, Characteristic classes of surface bundles. *Invent. Math.* 90 (1987), 551–577. [85](#)

- [144] S. Morita, *Geometry of characteristic classes*. Transl. Math. Monogr. 199, Amer. Math. Soc., Providence, R.I., 2001. 85
- [145] J. Moser, A Harnack inequality for parabolic differential equations. *Comm. Pure Appl. Math.* 17 (1964), 101–134. 50
- [146] D. Mumford, A remark on Mahler’s compactness theorem. *Proc. Amer. Math. Soc.* 28 (1971), 289–294. 82
- [147] S. Nag, *The complex analytic theory of Teichmüller spaces*. Canad. Math. Soc. Ser. Monogr. Adv. Texts, John Wiley & Sons, New York 1988. 35
- [148] M. S. Narasimhan and C. S. Seshadri, Stable and unitary vector bundles on a compact Riemann surface. *Ann. of Math. (2)* 82 (1965), 540–567. 70
- [149] J. Nielsen, Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen. I. *Acta Math.* 50 (1927), 189–358. 81
- [150] J. Nielsen, Abbildungsklassen endlicher Ordnung. *Acta Math.* 75 (1943), 23–115. 85
- [151] N. Nitsure, Moduli space of semistable pairs on a curve. *Proc. London Math. Soc.* (3) 62 (1991), 275–300. 72
- [152] K. Obitsu and S. Wolpert, in preparation. 80
- [153] J.-P. Otal, Le théorème d’hyperbolisation pour les variétés fibrées de dimension 3. *Astérisque* 235 (1996). 58, 67, 68, 88
- [154] R. Palais, unpublished. 44
- [155] A. N. Paršin, Algebraic curves over function fields. *Dokl. Akad. Nauk SSSR* 183 (1968), 524–526; English transl. *Soviet Math. Dokl.* 9 (1968), 1419–1422. 94
- [156] F. Paulin, Topologie de Gromov équivariante, structures hyperboliques et arbres réels. *Invent. Math.* 94 (1988), 53–80. 76
- [157] E. Reich, On the variational principle of Gerstenhaber and Rauch. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 10 (1985), 469–475. 56
- [158] E. Reich and K. Strebel, On the Gerstenhaber-Rauch principle. *Israel J. Math.* 57 (1987), 89–100. 56
- [159] B. Rivet, Difféomorphismes harmoniques du plan hyperbolique. Preprint 2001. 34
- [160] H. L. Royden, The Ahlfors-Schwarz lemma in several complex variables. *Comment. Math. Helv.* 55 (1980), 547–558. 94
- [161] H. Royden, unpublished.
- [162] J. Sacks and K. Uhlenbeck, The existence of minimal immersions of 2-spheres. *Ann. of Math. (2)* 113 (1981), 1–24. 51
- [163] J. Sacks and K. Uhlenbeck, Minimal immersions of closed Riemann surfaces. *Trans. Amer. Math. Soc.* 271 (1982), 639–652. 51, 87
- [164] J. H. Sampson, Some properties and applications of harmonic mappings. *Ann. Sci. École Norm. Sup.* (4) 11 (1978), 211–228. 52, 53
- [165] J. H. Sampson, Applications of harmonic maps to Kähler geometry. In *Complex differential geometry and nonlinear differential equations* (Brunswick, Maine, 1984), Contemp. Math. 49, Amer. Math. Soc., Providence, R.I. 1986, 125–134.

- [166] 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, New York 1984, 321–358. [35](#), [50](#)
- [167] R. Schoen, The effect of curvature on the behavior of harmonic functions and mappings. In *Nonlinear partial differential equations in differential geometry* (Park City, UT, 1992), IAS/Park City Math. Ser. 2, Amer. Math. Soc., Providence, R.I. 1996, 127–184. [51](#), [61](#), [66](#)
- [168] R. Schoen and S. T. Yau, On univalent harmonic maps between surfaces. *Invent. Math.* 44 (1978), 265–278. [52](#)
- [169] R. Schoen and S. T. Yau, Existence of incompressible minimal surfaces and the topology of three-dimensional manifolds with nonnegative scalar curvature. *Ann. of Math.* (2) 110 (1979), 127–142. [51](#), [54](#), [87](#)
- [170] R. Schoen and S. T. Yau, *Lectures on harmonic maps*. Conference Proceedings and Lecture Notes in Geometry and Topology, II, International Press, Cambridge, MA, 1997. [52](#)
- [171] R. Schoen and M.-T. Wang, in preparation. [98](#)
- [172] T. Serbinowski, Boundary regularity of harmonic maps to nonpositively curved metric spaces. *Comm. Anal. Geom.* 2 (1994), 139–153. [61](#)
- [173] P. B. Shalen, Dendrology of groups: an introduction. In *Essays in group theory*, Math. Sci. Res. Inst. Publ. 8, Springer-Verlag, New York 1987, 265–319. [67](#)
- [174] C. T. Simpson, Constructing variations of Hodge structure using Yang-Mills theory and applications to uniformization. *J. Amer. Math. Soc.* 1 (1988), 867–918. [70](#)
- [175] C. T. Simpson, Higgs bundles and local systems. *Inst. Hautes Études Sci. Publ. Math.* (1992), 5–95. [72](#)
- [176] C. T. Simpson, Moduli of representations of the fundamental group of a smooth projective variety. I. *Inst. Hautes Études Sci. Publ. Math.* (1994), 47–129. [72](#)
- [177] Y. T. Siu, Curvature of the Weil-Petersson metric in the moduli space of compact Kähler-Einstein manifolds of negative first Chern class. In *Contributions to several complex variables*, Aspects Math. E9, Vieweg, Braunschweig 1986, 261–298. [78](#)
- [178] R. K. Skora, Splittings of surfaces. *J. Amer. Math. Soc.* 9 (1996), 605–616. [67](#)
- [179] K. Strebel, *Quadratic differentials*. *Ergeb. Math. Grenzgeb.* (3) 5, Springer-Verlag, Berlin 1984. [41](#), [54](#), [58](#), [59](#)
- [180] X. Sun, Regularity of harmonic maps to trees. *Amer. J. Math.* 125 (2003), 737–771. [57](#), [62](#), [63](#)
- [181] L.-F. Tam and T. Y. H. Wan, Quasi-conformal harmonic diffeomorphism and the universal Teichmüller space. *J. Differential Geom.* 42 (1995), 368–410. [34](#)
- [182] O. Teichmüller, *Gesammelte Abhandlungen*. Springer-Verlag, Berlin 1982. [42](#)
- [183] A. J. Tromba, On a natural algebraic affine connection on the space of almost complex structures and the curvature of Teichmüller space with respect to its Weil-Petersson metric. *Manuscripta Math.* 56 (1986), 475–497. [78](#)
- [184] A. J. Tromba, On an energy function for the Weil-Petersson metric on Teichmüller space. *Manuscripta Math.* 59 (1987), 249–260. [87](#)

- [185] A. J. Tromba, A new proof that Teichmüller space is a cell. *Trans. Amer. Math. Soc.* 303 (1990), 257–262. [69](#)
- [186] A. J. Tromba, *Teichmüller theory in Riemannian geometry*. Lectures Math. ETH Zürich, Birkhäuser, Basel 1992. [35](#), [44](#), [69](#)
- [187] A. J. Tromba, Dirichlet’s energy on Teichmüller’s moduli space and the Nielsen realization problem. *Math. Z.* 222 (1996), 451–464. [86](#), [87](#)
- [188] A. J. Tromba, Dirichlet’s energy of Teichmüller moduli space is strictly pluri-subharmonic. In *Geometric analysis and the calculus of variations*, International Press, Cambridge, MA, 1996, 315–341. [89](#)
- [189] K. Uhlenbeck and S.-T. Yau, On the existence of Hermitian-Yang-Mills connections in stable vector bundles. *Comm. Pure Appl. Math.* 39 (1986), S257–S293. [70](#)
- [190] T. Y.-H. Wan, Constant mean curvature surface, harmonic maps, and universal Teichmüller space. *J. Differential Geom.* 35 (1992), 643–657. [34](#)
- [191] M.-T. Wang, A fixed point theorem of discrete group actions on Riemannian manifolds. *J. Differential Geom.* 50 (1998), 249–267. [98](#)
- [192] M.-T. Wang, Generalized harmonic maps and representations of discrete groups. *Comm. Anal. Geom.* 8 (2000), 545–563. [98](#)
- [193] R. Wentworth, Regularity of harmonic maps from Riemann surfaces to the Weil–Petersson completion of Teichmüller space. Preprint 2004. [92](#)
- [194] R. Wentworth, Energy of harmonic maps and Gardiner’s formula. In *Proceedings of the Ahlfors-Bers Colloquium* (Ann Arbor, Michigan, 2005), to appear. [69](#), [89](#)
- [195] K. Whittlesey, Normal all pseudo-Anosov subgroups of mapping class groups. *Geom. Topol.* 4 (2000), 293–307. [93](#)
- [196] M. Wolf, The Teichmüller theory of harmonic maps. *J. Differential Geom.* 29 (1989), 449–479. [53](#), [77](#), [87](#)
- [197] M. Wolf, Harmonic maps from surfaces to \mathbb{R} -trees. *Math. Z.* 218 (1995), 577–593. [59](#), [74](#)
- [198] M. Wolf, On the existence of Jenkins-Strebel differentials using harmonic maps from surfaces to graphs. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 20 (1995), 269–278. [58](#), [68](#)
- [199] M. Wolf, On realizing measured foliations via quadratic differentials of harmonic maps to \mathbb{R} -trees. *J. Anal. Math.* 68 (1996), 107–120. [58](#), [68](#)
- [200] M. Wolf, Measured foliations and harmonic maps of surfaces. *J. Differential Geom.* 49 (1998), 437–467. [69](#)
- [201] M. Wolf and S. Wolpert, Real analytic structures on the moduli space of curves. *Amer. J. Math.* 114 (1992), 1079–1102. [80](#)
- [202] S. Wolpert, Noncompleteness of the Weil-Petersson metric for Teichmüller space. *Pacific J. Math.* 61 (1975), 573–577. [78](#)
- [203] S. A. Wolpert, Chern forms and the Riemann tensor for the moduli space of curves. *Invent. Math.* 85 (1986), 119–145. [78](#)
- [204] S. A. Wolpert, Geodesic length functions and the Nielsen problem. *J. Differential Geom.* 25 (1987), 275–296. [86](#), [88](#)

- [205] S. A. Wolpert, The hyperbolic metric and the geometry of the universal curve. *J. Differential Geom.* 31 (1990), 417–472. [79](#), [80](#)
- [206] S. A. Wolpert, Geometry of the Weil-Petersson completion of Teichmüller space. In *Surveys in differential geometry* (Boston, MA, 2002), Surv. Differ. Geom. 8, International Press, Somerville, MA, 2003, 357–393. [35](#), [78](#), [79](#), [80](#), [83](#), [84](#)
- [207] S. A. Wolpert, Convexity of geodesic-length functions: a reprise. In *Spaces of Kleinian groups* (Cambridge, UK, 2003), London Math. Soc. Lecture Note Ser. 329, Cambridge University Press, Cambridge 2006, 233–245. [88](#)
- [208] J. W. Wood, Bundles with totally disconnected structure group. *Comment. Math. Helv.* 46 (1971), 257–273. [73](#)
- [209] S. Yamada, Weil-Petersson convexity of the energy functional on classical and universal Teichmüller spaces. *J. Differential Geom.* 51 (1999), 35–96. [87](#)
- [210] S. Yamada, On the geometry of Weil-Petersson completion of Teichmüller spaces. *Math. Res. Lett.* 11 (2004), 327–344. [80](#)
- [211] S. T. Yau, A general Schwarz lemma for Kähler manifolds. *Amer. J. Math.* 100 (1978), 197–203.
- [212] S.-K. Yeung, Bounded smooth strictly plurisubharmonic exhaustion functions on Teichmüller spaces. *Math. Res. Lett.* 10 (2003), 391–400. [90](#)
- [213] S.-K. Yeung, Representations of semisimple lattices in mapping class groups. *Internat. Math. Res. Notices* (2003), 1677–1686. [95](#)
- [214] H. Zieschang, *Finite groups of mapping classes of surfaces*. Lecture Notes in Math. 875, Springer-Verlag, Berlin 1981. [86](#)

Chapter 2

On Teichmüller's metric and Thurston's asymmetric metric on Teichmüller space

Athanase Papadopoulos and Guillaume Th  ret

Contents

1	Introduction	112
2	Teichm��ller's metric	116
	2.1 Measured foliations	116
	2.2 Conformal structures	120
	2.3 Moduli and extremal lengths	124
	2.4 The Teichm��ller metric	133
	2.5 Teichm��ller maps and Teichm��ller geodesics	134
	2.6 On the asymptotic behaviour of Teichm��ller rays	137
	2.7 The Teichm��ller boundary and convergent rays	138
	2.8 Convergence to Thurston's boundary	139
	2.9 On the visual boundary of the Teichm��ller metric	141
3	Thurston's asymmetric metric	142
	3.1 Hyperbolic structures	142
	3.2 Decompositions by generalized pairs of pants and by ideal triangles	143
	3.3 Geodesic laminations	147
	3.4 Horocyclic foliations	152
	3.5 Thurston's cataclysm coordinates	154
	3.6 Stretch lines, stretch rays, anti-stretch rays and stretch maps	164
	3.7 Thurston's asymmetric metric	165
	3.8 Comparison with other metrics	172
	3.9 Example I of a stretch line: the four-punctured sphere	173
	3.10 Example II of a stretch line: the punctured torus	175
	3.11 Example III of a stretch line: the closed surface of genus 2	178
	3.12 The behaviour of the lengths of measured geodesic laminations along stretch and anti-stretch lines	191
	3.13 Thurston's boundary	192
	3.14 Converging to Thurston's boundary	193
	3.15 Stretch lines and earthquakes	195
4	Problems	196
	References	201

1 Introduction

In this chapter, g and n are two nonnegative integers and $S = S_{g,n}$ is a connected oriented surface obtained from a closed surface of genus g (which we shall denote by \hat{S}) by removing n points called the punctures.¹ We assume that the Euler characteristic of S is negative. $\mathcal{T} = \mathcal{T}_{g,n}$ denotes the Teichm  ller space of S . This space carries several interesting metrics. In this chapter, we shall consider two of them, namely, Teichm  ller's metric and Thurston's asymmetric metric. We shall study these metrics respectively in Sections 2 and 3 below. These two metrics are Finsler metrics, that is, the distance between two points in Teichm  ller space can be defined by minimizing the lengths of paths joining these points, and lengths of paths are computed by using a norm defined on the tangent space at each point of \mathcal{T} . In the case of Thurston's asymmetric metric, the norms on the tangent spaces are not symmetric.

Teichm  ller space can be defined either as a space of equivalence classes of conformal structures on S or as a space of equivalence classes of hyperbolic structures on S . Each of the two metrics considered here is natural from one of these points of view: Teichm  ller's metric from the point of view of conformal geometry and Thurston's asymmetric metric from the point of view of hyperbolic geometry. Indeed, Teichm  ller's metric is natural as a measure of distances between (equivalence classes of) conformal structures, since it is defined as the logarithm of the least quasiconformal dilatation of a homeomorphism isotopic to the identity between these conformal structures, whereas Thurston's asymmetric metric is natural from the point of view of measuring distances between (equivalence classes of) hyperbolic structures, since it is defined as the logarithm of the smallest Lipschitz constant of homeomorphisms isotopic to the identity from one hyperbolic structure to the other one. (The order in which we take the hyperbolic surfaces is important in the last definition because the smallest Lipschitz constant in one direction is generally different from the smallest Lipschitz constant in the other direction.) The difference between the conformal and the hyperbolic points of view makes the techniques used in the study of these two metrics on Teichm  ller space of different natures: on the one hand, we use complex analysis (quasiconformal mappings, quadratic differentials, extremal length and so on), and, on the other hand, we use two-dimensional hyperbolic geometry. But the problems that we try to solve are formally the same: the global behaviour of geodesic lines (that is, isometric images of \mathbb{R}) in Teichm  ller space, e.g. the question of whether they are properly embedded or not, the convergence of geodesic rays to a point on Thurston's boundary, the study of visual boundaries, and other related problems. There are also several analogies between the results obtained so far for the two metrics. Finally, there are natural questions about the comparison between the two metrics. These are reasons for which we present the two metrics in parallel.

Let us start with a few words about some analogies and some differences between the general features of the two metrics.

¹Such a surface S is said to be of *finite type*, in reference to the fact that its fundamental group is of finite type.

We recall two descriptions of the Teichmüller distance between two conformal structures. Consider two conformal structures on a surface. On the one hand, the Teichmüller distance is the logarithm of the infimum of quasiconformal dilatations of homeomorphisms isotopic to the identity between the two conformal structures. On the other hand, this distance is the logarithm of the supremum of quotients of extremal lengths of closed curves with respect to these structures. Likewise, there are two descriptions of Thurston's asymmetric distance from one hyperbolic surface to another. On the one hand, this distance is the logarithm of the infimum of Lipschitz constants of homeomorphisms isotopic to the identity between the two hyperbolic surfaces, and on the other hand it is the logarithm of the supremum of quotients of hyperbolic lengths of closed geodesics, with respect to the two hyperbolic surfaces.

Teichmüller proved that given two conformal structures on S , there exists, in each isotopy class of homeomorphisms, a “best quasiconformal stretch homeomorphism”, that is, a homeomorphism for which the infimum of the quasiconformal dilatation in the definition of the Teichmüller distance between the two conformal structures is attained. Likewise, Thurston proved that given two hyperbolic structures on S , there exists, in each isotopy class of homeomorphisms, a “best Lipschitz stretch homeomorphism”, that is, a homeomorphism for which the infimum of the Lipschitz constant in the definition of Thurston's asymmetric distance is attained.

Teichmüller space, equipped with the Teichmüller metric, is a straight G-space in the sense of Busemann (cf. Kravetz [33]). This means that any two distinct points in that space lie on a unique geodesic line. Thurston's asymmetric metric has a different character: it is not symmetric (as its name indicates). Furthermore, any two distinct points in the space lie on a geodesic line, but this geodesic line is not necessarily unique.

Let us look more closely at the geodesic lines for the two metrics.

Geodesics for Teichmüller's metric are usually described (since the work of Teichmüller himself) in terms of quadratic differentials. In this chapter, we have avoided talking about quadratic differentials, but we have used instead (the equivalent point of view of) pairs of transverse measured foliations. This is not because we do not like quadratic differentials, but it is for the sake of stressing a further analogy between Teichmüller's metric and Thurston's asymmetric metric. To describe this analogy, let us be more precise.

To characterize a geodesic for Teichmüller's metric, we represent each point in Teichmüller space by a conformal structure defined by a pair (F_1, F_2) of transverse measured foliations on the surface S . The pair of measured foliations defines a “grid” on the surface, and there is a natural notion of distance measured along the leaves of each of these foliations, coming from the transverse measure of the other foliation. Thus, to a pair of transverse measured foliations, it is easy to associate holomorphic local parameters $z = x + iy$, the x -direction defined by the leaves of the first foliation and the y -direction by those of the other one. A geodesic line for Teichmüller's metric is then described as a family of surfaces $S^t = (e^{-t}F_1, e^tF_2)$, $t \in \mathbb{R}$. The measured foliations F_1 and F_2 are called respectively the horizontal and vertical foliations as-

sociated to the geodesic line $(S^t)_{t \in \mathbb{R}}$. This description makes it easy to visualize the holomorphic coordinates of the surface S^t when this surface varies along a geodesic line. In this way, each conformal structure S^t is represented by a Euclidean structure with isolated conical singular points on the surface, where the transverse measure of the foliation $e^{-t} F_1$ (respectively $e^t F_2$) determines the Euclidean length on the leaves of the transverse foliation $e^t F_2$ (respectively $e^{-t} F_1$).

In the case of Thurston's asymmetric metric, a distinguished class of geodesics can be described using an object which is less symmetrical than a pair of measured foliations. This object is a pair consisting of a complete (not necessarily measured) geodesic lamination, and a measured foliation which is transverse to it. More precisely, Thurston showed that any two points in Teichm  ller space can be joined by a geodesic (for his asymmetric metric) made up of a concatenation of pieces of "stretch lines", a stretch line being a parametrized family of hyperbolic structures $S^t = (\mu, e^t F)$, $t \in \mathbb{R}$, where μ is a complete geodesic lamination on S^t and F a measured foliation on $S = S^0$ which is transverse to μ . Here also, the pair $(\mu, e^t F)$ defines a privileged set of directions on the hyperbolic surface S^t , viz the geodesic directions of the leaves of μ , and the perpendicular directions which are the leaves of $e^t F$ and which are made out of pieces of horocycles, with the transverse measure of $e^t F$ coinciding on the leaves of μ with hyperbolic length. Thus, the measured foliation $e^t F$ plays here the role of a "vertical foliation", and the complete lamination μ plays the role of a "horizontal foliation", associated to the stretch line. Varying the parameter t describes the stretch line.

These descriptions of geodesic lines for the two metrics on Teichm  ller space lead naturally to similar questions concerning both metrics, as well as to questions concerning the comparison between them. We now enumerate some of these questions.

- The descriptions of the geodesic rays for Teichm  ller's metric and for Thurston's asymmetric metric lead to two distinct natural parametrizations of Teichm  ller space equipped with actions of the group \mathbb{R}_+^* of positive reals on the parameter spaces. A first question related to these parametrizations concerns the study of the extension of the parameters by adjoining limit points to the orbits of the \mathbb{R}_+^* -actions. This leads to defining boundaries to Teichm  ller space, and one natural question is about the dependence of such a boundary on the chosen parametrization in each case. Other questions concern the comparison of these boundaries among themselves and with other geometrically defined boundaries. More precisely, for each of the two metrics, there is a collection of parameter spaces, each of which being a set of equivalence classes of measured foliations. In the case of Teichm  ller's metric, a parameter space is a space of equivalence classes of measured foliations that are transverse to a fixed measured foliation, and in the case of Thurston's asymmetric metric, a parameter space is a space of equivalence classes of measured foliations that are transverse to a fixed complete geodesic lamination. As such, the two parameter spaces admit natural \mathbb{R}_+^* -action (induced by the action of \mathbb{R}_+^* on measures), and the orbits of these actions correspond to geodesic lines in Teichm  ller space, for each of these metrics respectively. The questions about the extension of the parameters to the boundary

involve the study of the asymptotic behaviour of geodesic rays for each of these metrics. In each case, the \mathbb{R}_+^* -orbits are properly embedded in the parameter space. In these notes, we present these facts in some detail, as well as results on the following questions.

- *Convergence of geodesic rays.* Some of the convergence results are formulated in terms of a boundary of \mathcal{T} which Steve Kerckhoff called Teichmüller boundary, others in terms of Thurston's boundary, and others in terms of the visual boundaries of Teichmüller space. In the early 1980s, Howard Masur obtained results on the convergence of some particular classes of geodesic rays for Teichmüller's metric to points on Thurston's boundary, and Kerckhoff obtained results on the relative behaviour of pairs of Teichmüller geodesic rays. We present these results below. It seems that there were no other significant results of that type until a recent work by Anna Lenzhen, in which she produces a geodesic ray that does not converge to a point on Thurston's boundary. The question of the description of the behaviour of an arbitrary geodesic ray for Teichmüller's metric with respect to Thurston's boundary is still open. Concerning the convergence of geodesic rays for Thurston's asymmetric metric to a point on Thurston's boundary, we shall present some recent results.

- *The asymptotic behaviour of "anti-stretch" rays.* An anti-stretch ray is the negative part of a stretch line, oriented in the direction opposite to the one given by the parametrization of the stretch line. Due to the fact that Thurston's metric is not symmetric, an anti-stretch ray is in general not a stretch ray (even after reparametrization). We note that an anti-stretch ray is (up to reparametrization) a geodesic ray for the asymmetric metric on Teichmüller space which is "dual" to Thurston's asymmetric metric. Here, the dual K^* of an asymmetric metric K is defined by the formula $K^*(x, y) = K(y, x)$.

- *The asymptotic behaviour of the length of an arbitrary measured geodesic lamination of compact support under a stretch or an anti-stretch ray.* More precisely, for a given family of hyperbolic surfaces $(S^t)_{t \in \mathbb{R}}$ parametrized by a stretch line and for any compactly supported measured geodesic lamination α , we are interested in the existence of the limits $\lim_{t \rightarrow \infty} l_{S^t}(\alpha)$ and $\lim_{t \rightarrow -\infty} l_{S^t}(\alpha)$, and whether these limits are finite or infinite.

We note that in the case of the Teichmüller metric, we have stated some of the results for the restricted case of closed surfaces, because the written sources for these results exist only in that special case (although most of these results are certainly valid in the larger context of surfaces of finite type). For Thurston's asymmetric metric, we present the results in the case of surfaces with or without punctures.

The results on the Teichmüller metric that we give are classical, but some of those on Thurston's asymmetric metric are new.

At the end of this chapter, we formulate some open problems which concern especially Thurston's asymmetric metric.

2 Teichmüller's metric

References for Teichmüller's metric can be found in Teichmüller's collected papers [62], as well as in those of Ahlfors and of Bers [3], [12], who rewrote part of the theory and developed it. There are several introductory books on the subject of Teichmüller's theory, for instance the books by Abikoff [1] and by Iwayoshi and Taniguchi [27]. We also recommend the recent books by Hubbard [25] and by Fletcher and Markovic [21].

2.1 Measured foliations

We start by recalling a few facts about measured foliations that are used in the theory of deformation of conformal structures which we present below.

Definition 2.1 (Measured foliation). A *measured foliation* on S is a foliation with isolated singularities, equipped with a positive measure on each transverse arc that is equivalent to the Lebesgue measure of a closed interval of \mathbb{R} . (An arc in a surface is, by definition, a homeomorphic image of the interval $[0, 1]$.) These measures are invariant by isotopies of the transverse arcs during which each point stays on the same leaf. The isolated singularities are of the type suggested in Figure 1, and we call them s -prong singularities, where s can be any integer ≥ 3 . We require that at the punctures of S , the foliation extends as a measured foliation of the unpunctured surface \hat{S} in such a way that each puncture becomes either a nonsingular point, or an s -prong singular point with s being here any integer ≥ 1 (see Figure 2). Note that in the case $s = 2$, the foliation extends as a nonsingular point at the puncture.

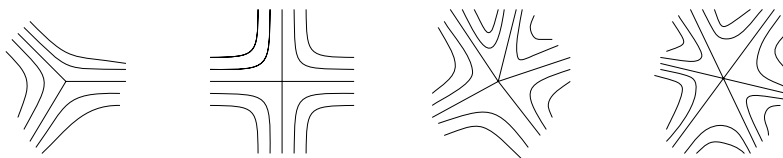


Figure 1. The four pictures represent s -prong singular points with $s = 3, 4, 5, 6$ respectively.

There is an equivalence relation between measured foliations, called *Whitehead-equivalence*. It is generated by the following transformations:

- Homeomorphisms of the surface which are isotopic to the identity sending one foliation to the other and preserving the transverse measures. By abuse of language, we shall sometimes call a homeomorphism isotopic to the identity an *isotopy*. We recall that two homeomorphisms of a surface are isotopic if and only if they are homotopic. This is a result of Baer [6]; see also Mangler [39] and Epstein [19].

- Whitehead moves: These are deformations of the surface that take place in a neighborhood of arcs that join two singular points and whose effect is to collapse such

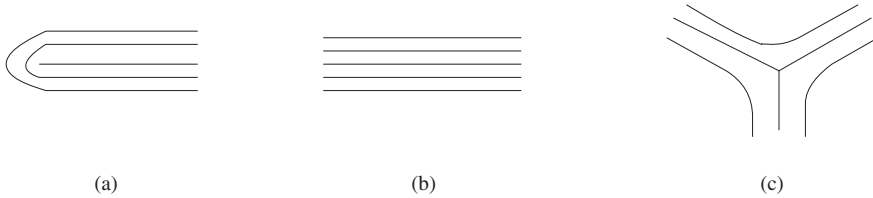


Figure 2. The three pictures represent the local model for a measured foliation at a puncture of S . The foliation extends in three possible ways: either as a 1-prong singular point (case (a)), or as a regular point (case (b)), or as a singular point with at least 3 prongs (case (c)), that is, like the singular points at interior points of the surface.

an arc to a point. (Remember that by our definition, an arc is embedded, which implies that after such a collapse, the surface remains a surface.) An example of a Whitehead move is given in Figure 3. Again, these moves are required to respect the transverse measures. The inverse move of a Whitehead move is also called a Whitehead move.

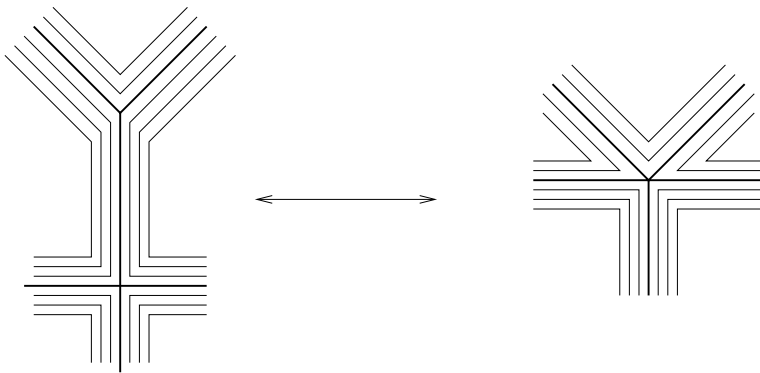


Figure 3. Whitehead move: collapsing or creating an arc joining two singular points.

Note that a singular point involved in a Whitehead move can be at a puncture of S , and that it is sometimes possible to eliminate a 1-prong singularity at a puncture by using a Whitehead move; see for instance Figure 4.

We let $\mathcal{MF}(S) = \mathcal{MF}$ denote the set of equivalence classes of measured foliations on S . An element of \mathcal{MF} is called a *measured foliation class*.

If x is a positive real number and if F is a measured foliation, then xF denotes the foliation F (as a topological object), equipped with the transverse measure obtained by multiplying the original transverse measure of F by the factor x . This action is compatible with the Whitehead equivalence relation, and it induces an action of \mathbb{R}_+^* on the set \mathcal{MF} . The quotient of \mathcal{MF} by this action is denoted by $\mathcal{PMF}(S)$ or \mathcal{PMF} . An element of \mathcal{PMF} is called a *projective measured foliation class*.

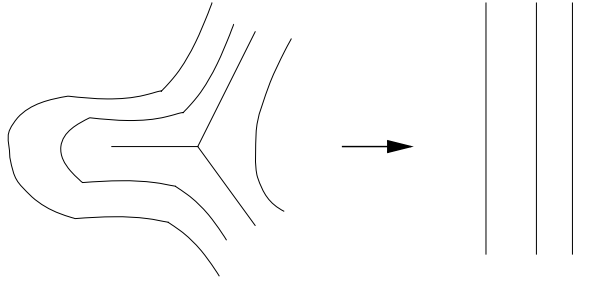


Figure 4. Eliminating a 1-prong singularity at a puncture of S by applying a Whitehead move.

If F is a measured foliation, then $[F]$ will usually denote its equivalence class in \mathcal{MF} . We shall also sometimes use the same notation, $[F]$, for the corresponding element in \mathcal{PMF} .

Measured foliations already appear in Teichmüller's work as horizontal and vertical line fields associated to holomorphic quadratic differentials on Riemann surfaces, but it was Thurston who initiated their systematic study, and defined the space of equivalence classes of measured foliations, in his paper *On the geometry and dynamics of diffeomorphisms of surfaces* ([69], published several years after it has been written). For a complete presentation of these results in the case of closed surfaces, we refer the reader to [20].

A closed curve in S is called *essential* if it is not homotopic to a point or to a puncture, and it is called *simple* if it has no self-intersection, that is, if it is an embedded image of a circle.

Let \mathcal{S} be the set of homotopy classes of unoriented essential simple closed curves in S .

There is a natural embedding

$$\mathcal{S} \rightarrow \mathcal{MF} \tag{2.1}$$

defined as follows. For any element γ in \mathcal{S} , we take a foliated cylinder C embedded in S , whose leaves are simple closed curves that are in the homotopy class γ . This foliated cylinder C defines a *partial* foliation on S . Here, the adjective *partial* means that the support of the foliation is a subset of the surface S . We choose an arc c that joins the two boundary components of C and which is transverse to the foliations, and a homeomorphism between this arc and the interval $[0, 1] \subset \mathbb{R}$, and we equip the arc c with the pull-back of the Lebesgue measure of $[0, 1]$. We then take the only invariant transverse measure for the foliation on C that induces the given Lebesgue measure on the arc c . We shall sometimes say that C , equipped with this measured foliation, is a foliated cylinder of height one with core curve in γ . To get a measured foliation on S , we collapse the closure of each connected component of $S \setminus C$ onto a spine. We recall that a spine of a compact surface with boundary is a one-dimensional complex on which the surface collapses by a homotopy equivalence (see [20], p. 90).

Here, the spine is chosen so that the singular points of the resulting foliation on S are of the allowed type. Note that the exposition in [20] works for the case where S is compact. To deal with the case where S is noncompact (that is, the case where S has punctures), we can use the spine of a compact surface with boundary obtained by replacing each puncture by a boundary closed curve and then collapsing that boundary curve to a point. We do this in such a way that the local model of the resulting measured foliation on the original surface S is of the allowed type, at the interior singularities and at the punctures. The various choices of spines for the complementary components of the cylinder C differ precisely by Whitehead moves performed on spines. Using this fact, to any element of \mathcal{F} , we associate a measured foliation on S which is well defined up to isotopy and Whitehead moves. It is a fact of the classical theory of measured foliations that the map defined in (2.1) is injective (see [20], p. 89).

Note that in the case where the surface S is the three-punctured sphere, the set \mathcal{F} is empty. As is well known, the Teichmüller space in that case consists of a single point, and there is not a lot more to say. Therefore, we discard this case in what follows.

Definition 2.2 (Measured foliation with one cylinder). A measured foliation (respectively a measured foliation class) which up to a constant factor is obtained from an element of \mathcal{F} by the map described in (2.1) will be called a *measured foliation* (respectively a *measured foliation class*) with one cylinder.

Equivalently, a measured foliation with one cylinder is a measured foliation such that when extended to the unpunctured surface \hat{S} , the union of the leaves starting at the singular points, with the singular points included, is a compact graph whose complement is connected. This graph is called the *critical graph* of the foliation.

The *height* of a measured foliation with one cylinder is defined as the total mass of a transverse arc joining the two boundary components of that cylinder.

We need to recall another description of the equivalence relation between measured foliations.

To each measured foliation F , we associate a map $i(F, \cdot): \mathcal{F} \rightarrow \mathbb{R}_+$ defined for each γ in \mathcal{F} by the formula

$$i(F, \gamma) = \inf_{c \in \gamma} I(F, c) \quad (2.2)$$

where the infimum is taken over all closed curves c that are in the homotopy class γ and that are made up of a finite concatenation of segments which are either contained in leaves of F or are transverse to F , and where $I(F, c)$ denotes the total mass (with respect to the transverse measure of F) of all the sub-segments of c that are transverse to F .

In this way, a measured foliation defines an element of $\mathbb{R}_+^{\mathcal{F}}$, the set of functions from \mathcal{F} to \mathbb{R}_+ . Two measured foliations are said to be *measure-equivalent* if they have the same image in $\mathbb{R}_+^{\mathcal{F}}$.

The next two theorems summarize some basic results of Thurston that we shall refer to in the rest of the chapter.

Theorem 2.3 (Thurston, see [20], p. 110). *Two measured foliations are Whitehead-equivalent if and only if they are measure-equivalent.*

Theorem 2.3 gives an embedding of the space \mathcal{MF} in the function space $\mathbb{R}_+^{\mathcal{S}}$, and with this embedding we have the following:

Theorem 2.4 (Thurston, see [20]), p. 117 and 150). *With the topology induced on the space \mathcal{MF} by its embedding in the function space $\mathbb{R}_+^{\mathcal{S}}$ equipped with the weak topology, \mathcal{MF} is homeomorphic to $\mathbb{R}^{6g-6+2n} \setminus \{0\}$. The set of elements in \mathcal{MF} that are of the form $x\gamma$ with $x \in \mathbb{R}_+^*$ and $\gamma \in \mathcal{S}$ (where γ is regarded as an element of \mathcal{MF} by the embedding defined in (2.1) above) is a dense subset of \mathcal{MF} . The projectivized space \mathcal{PMF} , equipped with the quotient topology, is homeomorphic to a $(6g - 7 + 2n)$ -dimensional sphere $S^{6g-7+2n}$, in which the natural image of \mathcal{S} is dense.*

2.2 Conformal structures

We start with a word about atlases.

We shall define a conformal structure and, later on, a hyperbolic structure on S as an atlas. An atlas is a certain collection of local charts satisfying a certain property, and it is possible to form unions of collections of local charts. We shall say that two atlases are *compatible* if their union is an atlas satisfying the required properties. A *maximal* atlas is then a maximal union of compatible atlases.

Definition 2.5 (Conformal structure). A *conformal structure* (which we shall also call a *conformal atlas*) on S is a maximal atlas $\{(U_i, \phi_i)\}_{i \in \mathcal{I}}$ of local charts where for each $i \in \mathcal{I}$, U_i is an open subset of S and ϕ_i is a homeomorphism from U_i onto an open subset of the complex plane \mathbb{C} satisfying $\bigcup_{i \in \mathcal{I}} U_i = S$ and such that for all i and j in \mathcal{I} , the map $\phi_i \circ \phi_j^{-1}$, which is called a coordinate change map or transition map, and which is defined on $\phi_j(U_i \cap U_j)$, is conformal. Furthermore, we suppose that each puncture of S has a neighborhood which is conformally equivalent to a punctured disk in \mathbb{C} . (Without this condition, the neighborhood of a puncture could also be conformally a cylinder.) We shall sometimes use the word holomorphic instead of conformal. Each pair (U_i, ϕ_i) is called a *holomorphic chart*, and (by abuse of language) the variable $z \in \phi_i(U_i)$ is called a *holomorphic local coordinate* for the structure. On each domain U_i of a holomorphic chart, there is a natural orientation induced by the map ϕ_i from the usual orientation of \mathbb{C} . The transition functions of a conformal atlas, being holomorphic, are orientation preserving. Thus, a conformal structure on S equips S with a canonical orientation. We shall always assume that this orientation coincides with the orientation on S that we started with.

A surface equipped with a conformal structure is also called a *Riemann surface*.

Example 2.6 (Conformal structure induced by a Riemannian metric). Any Riemannian metric on S has an underlying conformal structure. This is usually deduced from the existence of the so-called isothermal coordinates, that is, local coordinates z in which the Riemannian metric can be written as $ds^2 = \lambda(z)|dz|^2$, with $\lambda(z) > 0$, see [5], p. 124–126, and the references therein. Isothermal coordinates have been investigated by Gauss, who proved their existence under some restrictive conditions which have been weakened later on. Conversely, every conformal structure on S is induced by some Riemannian metric. Furthermore, under our requirement that the Euler characteristic of S is negative, such a metric can be chosen to be hyperbolic, that is, of constant Gaussian curvature -1 . The existence of such a metric is the “uniformization theorem” for surfaces of negative Euler characteristic, attributed to Klein, Poincaré and Koebe, which is also based on the existence of isothermal parameters, cf. [57] and the references therein. More precisely, the uniformization theorem states that each conformal surface of finite type and of negative Euler characteristic can be realized as a quotient of the hyperbolic plane \mathbb{H}^2 by a discrete group Γ of isometries of \mathbb{H}^2 . The conformal surface is then induced by a hyperbolic metric. The hyperbolic metric is unique up to isotopy, and this makes hyperbolic metrics play a particularly important role in the theory of deformations of conformal structures. We shall use the hyperbolic point of view in Section 3 below.

The connection between a hyperbolic metric and the underlying conformal structure is not easy to handle, and classical Teichmüller theory (that is, the theory based on the techniques developed by Teichmüller) makes little use of hyperbolic geometry. However, there is a class of metrics which is more useful in the conformal theory, because the conformal structures that underly them are convenient to manipulate; these are the singular flat metrics that are defined by pairs of transverse measured foliations on the surface, of which we now recall the definition.

Example 2.7 (Conformal structure defined by a pair of transverse measured foliations). Let F_1 and F_2 be two transverse measured foliations on S . Recall from Definition 2.1 that each measured foliation extends to a measured foliation on the closed surface \hat{S} obtained from S by filling in the punctures. We require here that at each point of $\hat{S} \setminus S$, both F_1 and F_2 extend in the same manner, that is, both of them extend as s -prong singularities with the same $s \geq 1$. The local model for two transverse foliations F_1 and F_2 at a puncture of S is represented in Figure 6.

Such a pair of transverse measured foliations defines a conformal structure on S , and in fact, a distinguished class of holomorphic parameters, in the following way. We identify the neighborhood of each point on S which is a nonsingular point of the foliations to a subset of \mathbb{C} by using a parameter $z = x + iy \in \mathbb{C}$, where x is a variable along the leaves of F_1 and y is a variable along the leaves of F_2 , with the distance along a leaf of each foliation being measured using the transverse measure of the other foliation. Furthermore, we suppose that the positive orientation of x followed by the positive orientation of y coincides with the orientation of the surface S . The local parameter z (called a *distinguished* parameter) is defined up to the transformation

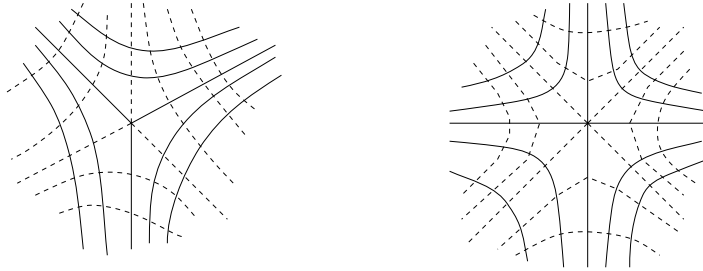


Figure 5. Each picture represents two transverse foliations near a singular point in S .

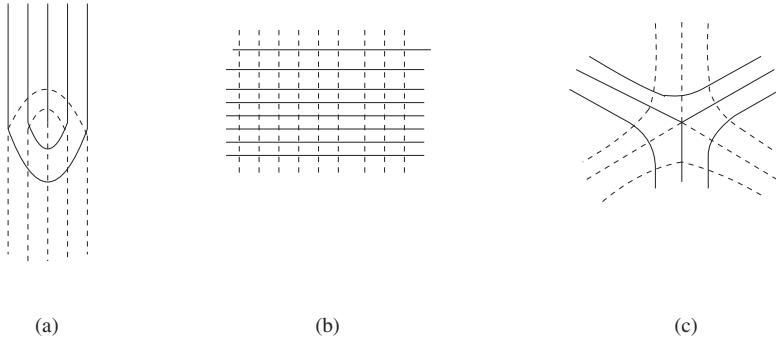


Figure 6. Each picture represents an extension of two transverse foliations at a puncture of S : in case (a) each foliation extends as a 1-prong singular point, in case (b) it extends as a nonsingular foliation and in case (c) it extends as a 3-prong singular point.

$z \mapsto -z$ and up to a translation in \mathbb{C} . In that way, the leaves of F_1 (respectively F_2) are locally defined by the equation $y = \text{constant}$ (respectively $x = \text{constant}$). The fact that the local parameters $z = x + iy$ are well defined at each point up to sign and up to the addition of a complex number implies that the local parameters associated to the various nonsingular points are compatible with each other from the holomorphic point of view, and they define a conformal structure in the complement of the singular points. Now we have to see that this conformal structure extends to the whole surface. Let p be a singular point of F_1 (or, equivalently, F_2) and let s be the number of prongs at that point. On each small enough disk neighborhood $V(p)$ of p in S , we can define a map $\phi_p: V(p) \rightarrow \mathbb{C}$ which sends p to 0 and which at each point of $V(p) \setminus \{p\}$ coincides with a branch of the map $z \mapsto z^{2/s}$, z being the distinguished parameter. This map is well defined up to composition in the range by a rotation of \mathbb{C} that fixes the origin. The maximal atlas generated by the collection of distinguished local charts at the nonsingular points of S , together with the maps $(V(p), \phi_p)$ associated to the various singular points, defines a conformal structure on S . (We note that the

condition we imposed on the way the foliations extend to the punctures of S ensures that the requirement in Definition 2.5 that each puncture has a neighborhood which is conformally equivalent to a punctured disk in \mathbb{C} is satisfied.)

Since in each small enough neighborhood of each point in S , the distinguished parameter in \mathbb{C} is unique up to addition of a complex constant and to multiplication by -1 , such a parameter defines a metric on that neighborhood, which makes the neighborhood isometric to an open subset of \mathbb{C} equipped with its Euclidean metric. Thus, if Z denotes the set of singular points of F_1 and F_2 , the surface $S \setminus Z$ is equipped with a canonical Euclidean metric. The leaves of the foliations F_1 and F_2 are geodesic with respect to that metric, and the foliations are mutually orthogonal. Such a metric is called a singular flat metric or a Euclidean metric with cone singularities, the singular points being the cone points.

Let $\text{Hom}(S)$ be the group of orientation-preserving homeomorphisms of S . We consider the following action of $\text{Hom}(S)$ on the set of conformal structures on S : for any maximal atlas $G = \{(U_i, \phi_i)\}_{i \in \mathcal{I}}$ and for any orientation-preserving homeomorphism $f: S \rightarrow S$, f^*G is the maximal atlas $\{(U'_i, \phi'_i)\}_{i \in \mathcal{I}}$ where for each i in \mathcal{I} , $U'_i = f^{-1}(U_i)$ and $\phi'_i = \phi_i \circ f|_{U'_i}$. The maximal atlas f^*G is then a conformal atlas on S .

If G and H are any two conformal structures on S , we shall sometimes denote (by abuse of language) $f: G \rightarrow H$ as a map $f: S \rightarrow S$ in which the domain space (respectively the target space) is the surface S equipped with the conformal structure G (respectively H).

We shall say that two conformal structures G and H are *equivalent* if there exists a homeomorphism $f: G \rightarrow H$ isotopic to the identity satisfying $f^*H = G$.

Definition 2.8 (Teichmüller space). The *Teichmüller space* of S , which we shall denote by $\mathcal{T}_{g,n}$, $\mathcal{T}_{g,n}(S)$, \mathcal{T} or $\mathcal{T}(S)$, is the space of equivalence classes of conformal structures on S .

The subgroup $\text{Hom}_0(S)$ of $\text{Hom}(S)$ consisting of the homeomorphisms that are isotopic to the identity is a normal subgroup, and the quotient group $\mathcal{MCG}(S) = \text{Hom}(S)/\text{Hom}_0(S)$ is called the *mapping class group* of S . It has a natural action on the Teichmüller space $\mathcal{T}_{g,n}(S)$, which is the quotient action of that of $\text{Hom}(S)$ on the set of maximal conformal atlases defined above.

The set of conformal structures on S is equipped with a natural topology in which two conformal structures G and H are close if we can find two coverings $\{(U_i, \phi_i)\}_{i \in \mathcal{I}}$ and $\{(V_j, \psi_j)\}_{j \in \mathcal{J}}$, where for each $i \in \mathcal{I}$ (respectively $j \in \mathcal{J}$), (U_i, ϕ_i) (respectively (V_j, ψ_j)) is a holomorphic chart for G (respectively H), such that $\bigcup_{i \in \mathcal{I}} U_i = \bigcup_{j \in \mathcal{J}} V_j = S$ and such that any map of the form $\phi_i \circ \psi_j^{-1}$ (with the usual convention for its domain of definition) is \mathcal{C}^∞ -close to the identity. (We are using the fact that a conformal structure on S defines a canonical \mathcal{C}^∞ -structure on that surface.) Teichmüller space $\mathcal{T}_{g,n}(S)$ is equipped with the quotient topology. There are

many nice ways of describing that topology, which show that this space is homeomorphic to $\mathbb{R}^{6g-6+2n}$. In particular, this topology is induced by the Teichmüller metric which is one of the main subject matters of this chapter (Definition 2.23 below), but we shall also see several other descriptions of this topology.

2.3 Moduli and extremal lengths

We shall talk about quadrilaterals and their moduli in Riemann surfaces, but it is natural to start with a few words about triangles, which are simpler objects.

A *topological triangle* T in a Riemann surface S is an embedded closed disk with three distinguished pairwise distinct points on its boundary. Such an object is equipped with a conformal structure, inherited from that of S . Strictly speaking, T is equipped with a *conformal structure with boundary*, which is defined as in Definition 2.5 above, except that instead of requiring the range of each homeomorphism ϕ_i to be an open subset of \mathbb{C} , we require it to be a relatively open subset of the closed half-plane $\{z \in \mathbb{C} \mid \text{Im}(z) \geq 0\}$. To simplify the exposition, we shall suppose that each time we consider a closed disk equipped with a conformal structure and a set of distinguished points on its boundary, these points are taken in a cyclic order which is compatible with the usual orientation on the boundary of the disk that is induced from the orientation of the disk induced from the orientation of the surface. Now any two topological triangles equipped with conformal structures are conformally equivalent, that is, there exists a conformal homeomorphism between the two disks that respects the distinguished points. This follows from the Riemann mapping theorem, which says that any simply connected open subset of the plane bounded by a Jordan curve can be mapped conformally onto the unit disk, that this mapping can be extended to a homeomorphism between the closures of the domains, that the extended homeomorphism is also conformal and that if we choose arbitrarily three distinguished pairwise distinct points in the boundary of the two disks, then the conformal map between the two closed disks can be taken so as to respect the distinguished points, and, finally, that this condition completely determines the map (see [34], Chapter 1).

Thus, topological triangles cannot be used to distinguish different Riemann surfaces, and the next objects of study along that line are topological disks with four distinguished points on their boundary instead of three. These objects are not all conformally equivalent, and they are classified by a conformal invariant called the modulus. In fact, it is sometimes more convenient to deal with closed disks with two disjoint distinguished closed arcs (instead of four distinguished points) on their boundary. The four boundary points of the two arcs will be the distinguished points. Thus, we call a *quadrilateral* in a Riemann surface S an embedded closed disk with two distinguished disjoint closed arcs in its boundary. We shall call the distinguished arcs the *vertical sides* of the quadrilateral. In the same way as for topological triangles, a quadrilateral in S is equipped with a conformal structure with boundary induced from that of S .

Definition 2.9 (Modulus of a quadrilateral). By the Riemann mapping Theorem, for any quadrilateral Q in a Riemann surface, there exists a unique positive real number $\text{Mod}(Q)$ together with a conformal homeomorphism ϕ from Q to the rectangle R in the Euclidean plane \mathbb{R}^2 with vertices at $(0, 0)$, $(\text{Mod}(Q), 0)$, $(\text{Mod}(Q), 1)$ and $(0, 1)$, such that ϕ sends the vertical sides of Q to the vertical sides of R (that is, the sides of length 1, see Figure 7). The value $\text{Mod}(Q)$ is called the *modulus* of Q (and of R).

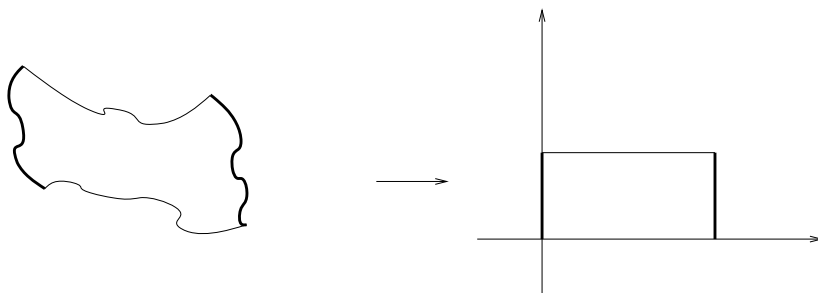


Figure 7. There is a conformal map from the quadrilateral Q on the left to the Euclidean rectangle on the right, sending the vertical sides of Q (which are drawn in bold lines) to the vertical sides of the Euclidean rectangle. The Euclidean rectangle is unique up to isometry, if we take the lengths of its vertical sides to be equal to 1.

Remarks. 1) Explicit formulae for conformal mappings between quadrilaterals having the same modulus are usually given by means of elliptic integrals (see [34]).

2) Another invariant of quadrilaterals in Riemann surfaces is the cross ratio of the four distinguished points. More precisely, one starts by mapping conformally the quadrilateral on the closed unit disk in the complex plane (by using the Riemann mapping associated to the interior of the quadrilateral which, as we recalled, extends to the boundary) and then taking the cross ratio of the images of the four distinguished points that we obtain on the boundary of that disk (see [4], p. 343).

3) One can define the modulus of a quadrilateral without using the Riemann mapping theorem (see [34]).

4) After moduli of topological triangles and quadrilaterals, one can study moduli of n -gons in Riemann surfaces. Here, an n -gon is defined as a topological disk with n distinguished points on its boundary. The dimension of the space of moduli (real parameters) of an n -gon is $n - 3$. More generally, Ahlfors and Beurling initiated the study of conformal invariants (or moduli) of arbitrary domains in the plane which are bounded by finitely many closed curves, with a finite number of distinguished points on their boundary and in their interior. The dimension of the moduli space of such a domain is $3n + 2p + q - 6$, where n is the number of boundary components, p the number of distinguished points in the interior and q the number of distinguished points on the boundary (see [4], p. 342).

Now back to quadrilaterals.

If G and H are Riemann surfaces and if $f: G \rightarrow H$ is a homeomorphism, then f transforms any quadrilateral in G into a quadrilateral in H . If f is conformal, it preserves the moduli of quadrilaterals. If $f: G \rightarrow H$ is a general homeomorphism, then the defect in conformality of f is measured by a quantity which is called the quasiconformal dilatation of f , defined as follows.

Definition 2.10 (Quasiconformal homeomorphism and quasiconformal dilatation). Let G and H be two Riemann surfaces and let $f: G \rightarrow H$ be an orientation-preserving homeomorphism. Then f is said to be *quasiconformal* if we have

$$K(f) = \sup_Q \frac{\text{Mod}(f(Q))}{\text{Mod}(Q)} < \infty,$$

where the supremum is taken over all quadrilaterals Q in G . The value $K(f)$ is called the *quasiconformal dilatation* of f . For every $K \geq K(f)$, f is said to be K -quasiconformal homeomorphism.

Notice that if Q is a given quadrilateral and if Q' is the quadrilateral obtained from Q by keeping the same topological disk and the same set of distinguished points, but applying to these points an order-one translation of the indices, then we have $\text{Mod}(Q) = 1/\text{Mod}(Q')$. With this remark, we can see that the value of $K(f)$ in Definition 2.10 is always ≥ 1 .

Definition 2.10 says in a geometric manner that a map is quasiconformal if it has uniformly bounded distortion. To see that this definition is equivalent to other (more commonly used) analytic definitions of quasiconformal homeomorphisms, we refer the reader to the paper [10] by Bers.

Remark 2.11. In the case where f is a \mathcal{C}^1 -diffeomorphism and where the conformal structures G and H are induced by Riemannian metrics, then, for each z in G , the differential of f at z , being an \mathbb{R} -linear map, takes a circle centered at the origin in the tangent space $T_z G$ of G at z to an ellipse centered at the origin in the tangent space $T_{f(z)} H$ of H at $f(z)$. The *(local) quasiconformal dilatation of f at z* , denoted by $K_z(f)$, is defined as the ratio of the major axis to the minor axis of that image ellipse. This ratio does not depend on the choice of the circle in $T_z G$ centered at the origin that we started with. The quasiconformal dilatation of f is then equal to

$$K(f) = \sup_{z \in G} K_z(f).$$

Finally, we note that the quasiconformal dilatation of f at z is also given by the formula

$$K_z(f) = \frac{\sup\{\|df_z(u)\| \text{ such that } u \in T_z \mathcal{S}, \|u\| = 1\}}{\inf\{\|df_z(u)\| \text{ such that } u \in T_z, \|u\| = 1\}}.$$

In this formula, the norm of the tangent vector u (respectively $df_z(u)$) is measured with respect to the Riemannian metric defining G (respectively H).

The study of quasiconformal dilatations of homeomorphisms between Euclidean rectangles was initiated by H. Grötzsch around 1928, who wrote several papers on that subject. Grötzsch showed that the (real-) affine homeomorphism between two arbitrary Euclidean rectangles realizes the minimum of the quasiconformal dilatation among all homeomorphisms that respect the sides of these rectangles (see for instance [23]). This result was one of the starting points of the theory of extremal quasiconformal mappings between general Riemann surfaces and in fact the affine map between rectangles is a building block for that general theory. Indeed, it follows from Teichmüller's results (which we shall recall in §2.5 below) that for any two conformal structures G and H on S , one can find two decompositions $\{R_1, \dots, R_k\}$ and $\{R_1^*, \dots, R_k^*\}$ of that surface and a homeomorphism $f: G \rightarrow H$ (which is called the Teichmüller map) such that the following conditions hold:

- (1) f realizes the minimum of the quasiconformal dilatation in the isotopy class of the identity, and it is the unique map that has this property;
- (2) for each $i = 1, \dots, k$, R_i and R_i^* are quadrilaterals satisfying $\bigcup_{i=1}^k R_i = \bigcup_{i=1}^k R_i^* = S$;
- (3) for each $i \neq j$, the interiors of R_i and R_j and of R_i^* and R_j^* are disjoint;
- (4) for all $i = 1, \dots, k$, the map f sends the quadrilateral R_i to the quadrilateral R_i^* , and if $f_i: R_i \rightarrow R_i^*$ (respectively $f_i^*: R_i^* \rightarrow R_i$) is the G - (respectively the H -) conformal homeomorphism sending R_i (respectively R_i^*) onto a Euclidean rectangle as in Definition 2.9, then, the map $f_i^* \circ f|_{R_i} \circ f_i^{-1}$ is an affine map;
- (5) the quasiconformal dilatations of the affine maps $f_i^* \circ f|_{R_i} \circ f_i^{-1}$ are equal for all $i = 1, \dots, k$.

It is in this sense that the affine map between two rectangles is a basic model for the Teichmüller map between arbitrary Riemann surfaces.

Notice that the local quasiconformal dilatation $K_z(f)$ of an affine homeomorphism f between Euclidean rectangles is constant (independent of z), and that the local quasiconformal dilatation on each of the quadrilaterals of the decomposition $\{R_1, \dots, R_k\}$ is the same (independently of the choice of the quadrilateral). This is a basic property of the Teichmüller maps between Riemann surfaces.

Examples. 1) Consider the surface obtained by doubling a Euclidean rectangle along its open edges (that is, without the vertices). This surface is the four-punctured sphere $S = S_{0,4}$, and it is equipped with two transverse measured foliations obtained by gluing the linear vertical (respectively horizontal) foliations of the two rectangles we started with, equipped with the transverse measures that induce Lebesgue measure on the edges of the rectangles. Each of these foliations extends as a one-prong singular point at each of the four punctures. We choose a common x and y -coordinate on the sides of the two rectangles. For each $\lambda > 0$ the real-affine map defined by $(x, y) \mapsto ((1/\lambda)x, \lambda y)$ is an extremal quasiconformal map on each rectangle, and the two maps fit together well and define a map on $S = S_{0,4}$ which is a Teichmüller map.

2) More generally, we can take an arbitrary (simply connected or not) closed subset of the Euclidean plane forming a subsurface with boundary with each boundary curve being made of a finite number of vertical and horizontal segments (see Figure 8 for an example). Doubling this surface, excluding the vertices, as in Example 1 above, produces a Riemann surface equipped with a canonical product foliation structure (see Definition 2.16), the one induced by the vertical and horizontal measured foliations of the plane. As in the particular case considered above, for each $\lambda > 0$, one can easily visualize a Teichmüller map on that surface, as a map induced by an affine transformation of the Euclidean plane. Notice that in all these examples, a singular point of each of the horizontal and vertical measured foliation is either a 1-prong or 3-prong singularity. At a 3-prong singularity, we can close the puncture, that is, we can include in the gluing of the two planar surfaces that we started with the endpoints of the edges that abut on that singular point. However, there are necessarily 1-prong singularities left, and we cannot obtain closed surfaces with this kind of construction. But it is easy to construct closed surfaces equipped with measured foliation pairs (see for instance the examples in the chapter by Herrlich and Schmithüsen in this Handbook [24]).

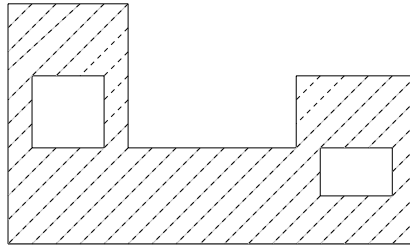


Figure 8. Doubling two such shaded regions along the open edges gives examples of surfaces with punctures equipped with measured foliation pairs. (The foliations we are talking about are those induced by the horizontal and vertical foliations of the plane.)

Definition 2.12 (Modulus of a topological cylinder). Let S be a Riemann surface and let C be a *topological cylinder in S* , that is, a surface homeomorphic to $S^1 \times I$ immersed in S , with its interior embedded. Such a cylinder is equipped with an induced conformal structure with boundary, and it is conformally equivalent to a Euclidean cylinder C^* which is unique up to scaling (see [2]). The *modulus* of C is the height of the Euclidean cylinder C^* divided by its circumference. It is denoted by $\text{Mod}_S(C)$ or $\text{Mod}(C)$.

Definition 2.13 (Modulus and extremal length of a homotopy class of curves). Let S be a Riemann surface and let γ be a homotopy class of essential simple closed curves in S . The *modulus* of γ , denoted by $\text{Mod}_S(\gamma)$ or $\text{Mod}(\gamma)$, is the supremum of the moduli of topological cylinders in S with core curve in the class γ . The *extremal length* of γ , denoted by $\text{Ext}_S(\gamma)$ or $\text{Ext}(\gamma)$, is defined as $1/\text{Mod}(\gamma)$.

The notion of extremal length of a family of curves in a Riemann surface was first defined by Beurling and it was developed later on by Beurling and Ahlfors, see [4]. The following result provides examples of Riemann surfaces in which one can see explicitly the maximal modulus cylinder in a given homotopy class.

Theorem 2.14. *Let S be a closed surface equipped with a conformal structure defined by two transverse measured foliations F_1 and F_2 as in Example 2.7. Suppose that F_1 is a foliation with one cylinder, and let γ be the homotopy class of the core curve of that cylinder. Then, the cylinder C in S whose interior is the complement of the critical graph of F_1 has the largest modulus among all topological cylinders in S with core curve in the homotopy class γ . Thus, in this situation, we have $\text{Mod}(\gamma) = \text{Mod}(C)$ or, equivalently, $\text{Ext}(\gamma) = 1/\text{Mod}(C)$.*

Theorem 2.14 is stated (in an equivalent form) by Kerckhoff in [29] (Theorem 3.1) and it is attributed there to J. A. Jenkins and K. Strebel.

There is an extension of this result to the extremal length of an arbitrary measured foliation class instead of a homotopy class of a simple closed curve. This is Theorem 2.21 below and, before stating it, we need to recall Kerckhoff's extension of the notion of extremal length from the set \mathcal{S} of homotopy classes of simple closed curves to the space \mathcal{MF} of equivalence classes of measured foliations. (We are using here the inclusion $\mathcal{S} \subset \mathcal{MF}$ defined in (2.1).)

Theorem 2.15 (Kerckhoff [29]). *Let S be a closed surface. The extremal length function defined on the set \mathcal{S} of homotopy classes of essential simple closed curves in S extends in a unique way to a continuous function $\text{ext}_S = \text{ext}: \mathcal{MF} \rightarrow \mathbb{R}_+^*$ satisfying $\text{ext}(xF) = x^2\text{ext}(F)$ for every x in \mathbb{R}_+^* and for every F in \mathcal{MF} .*

Definition 2.16 (Product foliation structure). Let $\mathcal{Q} = \mathcal{Q}(S)$ be the subset of $\mathcal{MF} \times \mathcal{MF}$ consisting of pairs of measured foliation classes ($[F_1], [F_2]$) such that $[F_1]$ and $[F_2]$ are representable by two foliations F_1 and F_2 that are transverse. We shall call an element of \mathcal{Q} a *product foliation structure*. The space \mathcal{Q} is equipped with the topology induced from the weak topology on $\mathcal{MF} \times \mathcal{MF}$.

The following lemma will be useful in considerations about product foliation structures.

Lemma 2.17 (Masur [41]). *Let S be a closed surface and let $[F_1]$ and $[F_2]$ be two measured foliation classes on S that can be represented by two transverse foliations F_1 and F_2 . Then the pair (F_1, F_2) is unique up to isotopy.*

In particular, we cannot perform Whitehead moves on F_1 or on F_2 while keeping the transversality.

Remark 2.18 (Product foliation structures and quadratic differentials). We must mention that a product foliation structure can be regarded as a Riemann surface equipped

with a holomorphic quadratic differential, although we do not make use of this fact in this chapter. We recall that a *holomorphic quadratic differential* q on a Riemann surface S is an invariant object that has an expression $q_i(z)dz^2$ in each holomorphic chart (U_i, ϕ_i) , where $q_i(z)$ is a holomorphic function of z , the holomorphic local coordinate in $\phi_i(U_i)$. Invariance of q means that if (U_j, ϕ_j) is another holomorphic chart, if w is the holomorphic local coordinate in $\phi_j(U_j)$, and if a local expression of q in that chart is $q_j(w)dw^2$, then at the overlap between the two charts we have $q_i(z)dz^2 = q_j(w)dw^2$ or, equivalently, $q_i(z)(dz/dw)^2 = q_j(w)$. This invariance property implies in particular that the zeroes of q and their orders are well defined. (In other words, the zeroes of the functions $q_i(z)$ and their orders are independent of the choice of the local chart, see [59], p. 18). Now to each pair (F_1, F_2) of transverse measured foliations on S is associated a conformal structure on that surface, as in Example 2.7, with local coordinates $z = x + iy$ which are well defined up to sign and up to composition by a translation in \mathbb{C} . From that we deduce that the local holomorphic quadratic differential form dz^2 is the local expression of a well-defined holomorphic quadratic differential form on S (that is, the invariance property is satisfied). The *norm* of a quadratic differential is defined as $\|q\| = \iint |q(z)| dx dy$ (this uses the fact that the area element $dx dy$ is independent of the choice of the coordinate chart), and the condition that we imposed on the foliations F_1 and F_2 at the punctures of S (see Definition 2.1) insures that the associated quadratic differential has finite norm. Conversely, to each conformal structure on S equipped with a holomorphic quadratic differential of finite norm, one can naturally associate a Euclidean metric with cone singularities, and a *horizontal* and a *vertical* foliation defining an element of the space $\mathcal{Q}(S)$, these two foliations being orthogonal and their leaves being local geodesics with respect to that Euclidean metric. Let us briefly recall the definitions of these foliations and of the Euclidean structure. Let q be a nonzero holomorphic quadratic differential on S , let (U_i, ϕ_i) be a holomorphic chart in S , and let $q_i(z)dz^2$ be an expression of q in that chart. If $z_0 = \phi_i(x)$ is the image by ϕ_i of a point x which is a nonzero point of q , then, since $q_i(z_0) \neq 0$, taking if necessary a smaller neighborhood U_i of z_0 , we may define a branch $q_i^{\frac{1}{2}}$ of the square root of q_i on that neighborhood. The integral $Q_i(z) = \int_{z_0}^z q_i(v)^{\frac{1}{2}} dv$ is then a holomorphic function in z and it determines a new holomorphic chart for S at the point x . The parameter $w = Q_i(z)$ is called a *distinguished* parameter for q at the nonzero point x . In terms of that parameter, the expression of q is dw^2 . (Remember that the differential dw is only defined up to sign, but that dw^2 is well defined.) The distinguished parameter w is unique up to addition of a complex constant and multiplication by -1 . Therefore, it establishes an isometry between its domain in S and an open subset of \mathbb{C} equipped with its Euclidean metric. Thus, if $Z(q) \subset S$ denotes the set of zeroes of the quadratic differential q , the surface $S \setminus Z(q)$ is equipped with a canonical Euclidean structure induced by q . At a zero of q of order p , the Euclidean structure has a cone singularity of cone angle $(p + 2)\pi$. On $S \setminus Z(q)$, the vertical (respectively horizontal) foliation is defined locally by taking a distinguished parameter $w = u + iv$ and transporting on

the domain of this local chart the foliation by vertical (respectively horizontal) lines $u = \text{constant}$ (respectively $v = \text{constant}$) in \mathbb{C} . Since the distinguished parameters are unique up to sign and up to translation in \mathcal{C} , these foliations on the various chart domains match up and give a well-defined vertical (respectively horizontal) foliation on $S \setminus Z(q)$. It also follows from this construction that the leaves of these foliations are local geodesics and are orthogonal with respect to the canonical Euclidean structure on $S \setminus Z(q)$ induced by q . Of course, this Euclidean structure is the same as the Euclidean structure associated to the pair of transverse measured foliations defined in Example 2.7. For more details on quadratic differentials, we refer the reader to Strebel's book [59] and to other chapters of this Handbook (see [18] and [24]).

Now we can define a map

$$p: \mathcal{Q}(S) \rightarrow \mathcal{T}(S) \tag{2.3}$$

by associating to each product foliation structure $([F_1], [F_2]) \in \mathcal{Q}(S)$ the equivalence class of conformal structures associated as in Example 2.7 to a pair (F_1, F_2) representing $([F_1], [F_2])$. By Lemma 2.17, the pair (F_1, F_2) is uniquely defined up to isotopy. Therefore, the corresponding element of Teichmüller space is well defined. The foliation F_1 (respectively F_2) is called the *horizontal* (respectively *vertical*) foliation of q . We shall say indifferently that the conformal structure is represented by the pair (F_1, F_2) or by the pair $([F_1], [F_2])$.

Jenkins and Strebel studied conformal structures defined by pairs of measured foliations, where the vertical foliation has a special property that we state in the following definition:

Definition 2.19 (Jenkins–Strebel structure). A *Jenkins–Strebel structure* is a product foliation structure $([F_1], [F_2])$ such that $[F_2]$ is the equivalence class of a measured foliation whose leaves are all compact.²

Notice that a measured foliation has all of its leaves compact if and only if the subset of the surface consisting of the union of the leaves that start at singular points is a compact subset of the surface. This subset is then a graph, called the *critical graph* of the foliation. The complement of this graph is a finite union of cylinders foliated by parallel leaves. The core curves of these cylinders are essential and pairwise non-homotopic closed curves. The property for a measured foliation to have all its leaves compact, if it holds for a given measured foliation, holds for any equivalent foliation. A particular class of Jenkins–Strebel structures is the class of product foliation structures whose vertical measured foliations are foliations with one cylinder, that we already considered above (Definition 2.2).

Jenkins studied in [28] (what we now call) Jenkins–Strebel structures as solutions of certain extremal problems concerning maps between Riemann surfaces. It follows

²Note that the fact that in this definition we talk about the vertical (rather than the horizontal) foliation is just a matter of convention. In comparing the statements here with statements in some of the papers that we refer to, one has to be careful about the conventions used in those papers.

from his work and from the work of Strebel in [58] that for any Riemann surface S and for any measured foliation F on S whose leaves are all compact, there exists a unique Jenkins–Strebel structure whose vertical foliation is equivalent to F and whose underlying conformal structure is the given one on S . The following more general result was obtained later on by Hubbard and Masur:

Theorem 2.20 (Hubbard & Masur [26]). *For any Riemann surface S and for any measured foliation F on S , there exists a unique product foliation structure whose underlying conformal structure is the one of S and whose vertical foliation is equivalent to F . Moreover, the map $\psi : \mathcal{MF}(S) \times \mathcal{T}(S) \rightarrow \mathcal{Q}(S)$ that associates to any Riemann surface S and to any measured foliation class $[F]$ this uniquely defined element of $\mathcal{Q}(S)$ is a homeomorphism.*

This implies in particular that the map $p : \mathcal{Q}(S) \rightarrow \mathcal{T}(S)$ defined in (2.3) is surjective.

Theorem 2.20 was proved by Hubbard and Masur in the case where S is a closed surface, and another proof was given by Kerckhoff [29]. An adaptation for the case of non-closed surfaces is contained in [22].

Theorem 2.20 gives the following parametrization of Teichmüller space:

Given a measured foliation F on S , let $\mathcal{MF}(F)$ be the space of measured foliation classes that are representable by measured foliations transverse to a measured foliation equivalent to F . For each $[F]$ in \mathcal{MF} , we have a map

$$\psi_{[F]} : \mathcal{T} \rightarrow \mathcal{MF}(F) \tag{2.4}$$

obtained by restricting the map ψ of Theorem 2.20 to the subset $\{[F]\} \times \mathcal{T}$ of $\mathcal{MF} \times \mathcal{T}$. By Theorem 2.20, the map $\psi_{[F]}$ is a homeomorphism. We regard this map as a parametrization of Teichmüller space by the subset $\mathcal{MF}(F)$ of measured foliation space. The flowlines of the natural action of \mathbb{R}_+^* on \mathcal{MF} correspond by this parametrization to images of geodesic lines in Teichmüller space, for the Teichmüller metric that we recall below.

Using Theorem 2.20, it is possible to compute explicitly the extremal length of some particular measured foliation classes, for Kerckhoff’s extension of the extremal length function $\text{ext} : \mathcal{MF} \rightarrow \mathbb{R}_+$ (Theorem 2.15).

Theorem 2.21 (Kerckhoff [29], p. 34). *Let S be a closed Riemann surface defined by two transverse measured foliations F_1 and F_2 as in Example 2.7. Then, $\text{ext}([F_1])$ and $\text{ext}([F_2])$ are equal to the total area of the singular flat metric associated to the pair (F_1, F_2) .*

Remember that a homotopy class of essential simple closed curves considered as an element of \mathcal{MF} is represented by a measured foliation with one cylinder whose height is equal to 1 (see (2.1) above). Therefore, if the foliation F_1 in the statement of Theorem 2.21 is a foliation with one cylinder whose height is h , then we have $\text{ext}([F_1]) = h^2 \text{Ext}_S(\gamma)$, which gives by Theorem 2.14 $\text{ext}([F_1]) = h^2 \frac{\ell}{h} = \ell h$, where

ℓ is the length of a closed leaf of the foliation of the cylinder, the length being computed in the singular flat metric defined by the two transverse foliations F_1 and F_2 . This value ℓh is the total area of that singular flat metric. Thus, the proof of Theorem 2.21 for $\text{ext}([F_1])$ follows from Theorem 2.14 which contains that result in the particular case where F_1 is a foliation with one cylinder, and from the continuity of the map ψ of Theorem 2.20. By symmetry, the same result holds for $\text{ext}([F_2])$.

Note that in view of Theorem 2.21 and of the quadratic behaviour in Theorem 2.15, extremal length should be considered as an area rather than as a length.

Remark 2.22 (Area and intersection number). The *geometric intersection function* $i: \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}_+$ is defined by

$$i(\gamma_1, \gamma_2) = \min \text{Card}\{C_1 \cap C_2\},$$

where the minimum is taken over all simple closed curves C_1 and C_2 in the homotopy classes γ_1 and γ_2 respectively. Thurston showed that this function i extends as a continuous function $i: \mathcal{MF} \times \mathcal{MF} \rightarrow \mathbb{R}_+$, called the geometric intersection function for measured foliations, which extends the map $i(F, \gamma)$ defined in (2.2) above. If F_1 and F_2 are two transverse measured foliations, then the intersection $i([F_1], [F_2])$ is equal to the area of the singular flat metric that this pair defines.

2.4 The Teichmüller metric

Definition 2.23 (The Teichmüller metric). Let G and H be two conformal structures on S . The *Teichmüller distance* between G and H is given by

$$d_T(G, H) = \frac{1}{2} \inf_f \log K(f)$$

where the infimum is taken over all quasiconformal homeomorphisms $f: G \rightarrow H$ that are isotopic to the identity. Since the value $d_T(G, H)$ remains unchanged if we replace G or H by an isotopic conformal structure, the map d_T induces a map on $\mathcal{T}_{g,n} \times \mathcal{T}_{g,n}$. This map is a metric, and it is called the *Teichmüller metric*. We shall denote it by d_T .

The symmetry of the map d_T follows from the fact that the inverse of a quasiconformal homeomorphism is a quasiconformal homeomorphism with the same quasiconformal dilatation. The triangle inequality follows from the fact that the composition of a K_1 -quasiconformal homeomorphism with a K_2 -quasiconformal homeomorphism is a $K_1 K_2$ -quasiconformal homeomorphism.

From the way we introduced the quasiconformal dilatation of a homeomorphism, Definition 2.23 gives the Teichmüller distance between two Riemann surfaces as a comparison between moduli of quadrilaterals in these surfaces. The following theorem gives a characterization of the Teichmüller distance as a comparison between extremal lengths.

Theorem 2.24 (Kerckhoff [29], p. 36). *Let S be a closed surface and let G and H be two conformal structures on S . The Teichmüller distance between G and H is equal to*

$$d_T(G, H) = \frac{1}{2} \log \sup_{\gamma \in \mathcal{B}} \frac{\text{Ext}_H(\gamma)}{\text{Ext}_G(\gamma)}.$$

Remarks 2.25. 1) By Kerckhoff's extension of the extremal length function to \mathcal{MF} (Theorem 2.15) and the fact that the set $\{x\gamma \mid x \in \mathbb{R}_+^*, \gamma \in \mathcal{B}\}$ is dense in \mathcal{MF} (Theorem 2.4), the Teichmüller distance is also given by

$$d_T(G, H) = \frac{1}{2} \log \sup_{[F] \in \mathcal{MF}} \frac{\text{ext}_H([F])}{\text{ext}_G([F])}, \quad (2.5)$$

and by

$$d_T(G, H) = \frac{1}{2} \log \sup_{[F] \in \mathcal{PMF}} \frac{\text{ext}_H([F])}{\text{ext}_G([F])}. \quad (2.6)$$

Note that to simplify notation, we are using in 2.6 the same notation, $[F]$, to denote the equivalence class of F in \mathcal{MF} and that in \mathcal{PMF} .

In order to see that the last expression is meaningful (even though the value $\text{ext}_H([F])$ is not defined for $[F]$ in \mathcal{PMF}), notice that the ratio $\frac{\text{ext}_H([F])}{\text{ext}_G([F])}$ in (2.6) is defined by choosing a representative in \mathcal{MF} of the element $[F]$ in \mathcal{PMF} and that the value of this ratio is independent of the choice that we make.

Formula (2.6) is more useful than Formula (2.5) because the space \mathcal{PMF} is compact, and therefore the supremum is attained.

2) In particular, we have $\sup_{[F] \in \mathcal{MF}} \frac{\text{ext}_H([F])}{\text{ext}_G([F])} = \sup_{[F] \in \mathcal{MF}} \frac{\text{ext}_G([F])}{\text{ext}_H([F])}$, and we note right away that the corresponding quotients that define Thurston's asymmetric metric (see Section 3 below) are not equal.

3) Teichmüller's metric is a Finsler metric.

2.5 Teichmüller maps and Teichmüller geodesics

Definition 2.26 (Teichmüller map). A *Teichmüller map* is the identity map $f: (S, G) \rightarrow (S, H)$, where G and H are two conformal structures on S that are defined as follows: there exist two transverse measured foliations F_1 and F_2 and a positive real λ , such that G is the conformal structure associated to the pair (F_1, F_2) as in Example 2.7 and H is the structure associated to the pair $((1/\lambda)F_1, \lambda F_2)$.

Remark 2.27. In terms of the distinguished local coordinate $z = x + iy$ associated to the conformal structure determined by the pair (F_1, F_2) , the homeomorphism f of Theorem 2.26 is defined by

$$(x, y) \mapsto ((1/\lambda)x, \lambda y). \quad (2.7)$$

The local quasiconformal dilatation $K_z(f)$ of f is constant on S , and it is equal to $\max(\lambda^2, 1/\lambda^2)$.

If we start with a topological surface S equipped with a pair (F_1, F_2) of transverse measured foliations, and a positive real number λ , then we can define an associated Teichmüller map. This map is the identity map on S equipped with the conformal structure (F_1, F_2) on the domain and $((1/\lambda)F_1, \lambda F_2)$ on the target. In other words, we can consider a Teichmüller map as defining a new conformal atlas on the surface S , obtained by composing each local chart of the conformal structure defined by (F_1, F_2) with the real-affine map defined, using the distinguished local coordinates, by Formula (2.7) (see Figure 9).

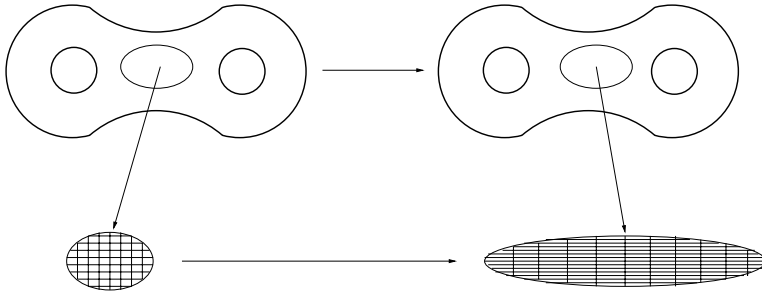


Figure 9. The effect of a Teichmüller map in local coordinates, with $\lambda > 1$ (λ is the parameter in Definition 2.26).

Theorem 2.28 (Teichmüller [60], [61]). *For any two conformal structures G and H on S , we can find two transverse measured foliations F_1 and F_2 and a positive real λ such that (F_1, F_2) represents the structure G and $((1/\lambda)F_1, \lambda F_2)$ represents the structure H . Furthermore, the Teichmüller map associated to these conformal structures G and H is the unique homeomorphism that has the least quasiconformal dilatation among the homeomorphisms between G and H that are isotopic to the identity.*

Remark 2.29. By Theorem 2.28, the least quasiconformal dilatation depends only on the equivalence classes g and h of G and H respectively.

Definition 2.30 (Teichmüller line and Teichmüller ray). A *Teichmüller line* (respectively *Teichmüller ray*) is a map $\ell: \mathbb{R} \rightarrow \mathcal{T}(S)$ (respectively $\ell: [0, \infty[\rightarrow \mathcal{T}(S)$) defined by $t \mapsto \ell(t) = (e^{-t}F_1, e^tF_2)$ where F_1 and F_2 are transverse measured foliations on S . (Here, as usual, we are identifying each pair of transverse measured foliations with the equivalence class of conformal structures that it determines, as in Example 2.7.) We say that the ray ℓ starts at the (equivalence class of the) conformal structure determined by the pair (F_1, F_2) . By abuse of language, we shall say that F_1

(respectively the measured foliation class, the projective class of F_1) is the *horizontal* measured foliation (respectively the measured foliation class, the projective class) of the geodesic line (or ray), and that F_2 is its *vertical* measured foliation (respectively measured foliation class, projective class). (In fact, the pair, and therefore, the foliations are defined up to a constant factor.)

Recall that if X is a metric space, then a *geodesic* (respectively a *geodesic line*, *geodesic ray*) in X is a distance-preserving map from a compact interval of \mathbb{R} (respectively from \mathbb{R} , from an interval of the form $[a, \infty)$ with $a \in \mathbb{R}$) into X . If $\gamma : [a, b] \rightarrow X$ (respectively $\gamma : [a, \infty) \rightarrow X$) is a geodesic (respectively a geodesic ray) in X , then we say that γ *starts* at the point $\gamma(a) \in X$, and that γ *connects* the points $\gamma(a)$ and $\gamma(b)$ in X . A *local geodesic* in X is a map from an interval I of \mathbb{R} into X such that each point in the interior of I has a neighborhood such that the restriction of the map to that neighborhood is a geodesic.

We also recall that a metric space X is said to be *geodesic* if any two points in X can be connected by a geodesic, and that a metric space is said to be *proper* if its closed balls are compact.

The following is a basic result on the metric structure of Teichm  ller space (cf. Kravetz [33]).

Theorem 2.31 (Teichm  ller geodesics). *Teichm  ller space $\mathcal{T}_{g,n}$, equipped with the Teichm  ller metric, is complete, proper and geodesic. Furthermore, each pair of distinct points in $\mathcal{T}_{g,n}$ is contained in the image of a geodesic line which is unique up to reparametrization. Teichm  ller lines are geodesic lines for the Teichm  ller metric, that is, for each Teichm  ller line $\ell : \mathbb{R} \rightarrow \mathcal{T}_{g,n}$, we have, for each t_1 and t_2 in \mathbb{R} , $d_T(\ell(t_1), \ell(t_2)) = |t_2 - t_1|$. Furthermore, all the bi-infinite geodesic lines for that metric are of this form.*

The following theorem is also well known. We provide a proof because it follows directly from previously mentioned results.

Theorem 2.32. *Let S be a closed surface, let $t \in \mathbb{R}$, let F_1 and F_2 be two transverse measured foliations on S and let G and H be the conformal structures associated respectively to the pair (F_1, F_2) and $(e^{-t}F_1, e^tF_2)$. Then the Teichm  ller distance between G and H is equal to $|t|$. Furthermore, for $t > 0$ (respectively $t < 0$), the measured foliation F_1 (respectively F_2) realizes the supremum in Formula (2.5) for the Teichm  ller distance between the two conformal structures defined by G and by H .*

Proof. By Remark 2.27, the quasiconformal dilatation of the Teichm  ller map between the conformal structures defined by (F_1, F_2) and $(e^{-t}F_1, e^tF_2)$ in the domain and target respectively is equal to $\max(e^{-2t}, e^{2t}) = e^{2|t|}$. This gives

$$d_T(G, H) = \frac{1}{2} \log e^{2|t|} = |t|.$$

Suppose now that $t > 0$. By Theorem 2.21 and Remark 2.22, we have

$$\text{ext}_G(F_1) = i(F_1, F_2)$$

and

$$\text{ext}_H(e^{-t} F_1) = i(e^{-t} F_1, e^t F_2) = i(F_1, F_2).$$

Therefore, we have, by Theorem 2.15

$$\frac{\text{ext}_H(F_1)}{\text{ext}_G(F_1)} = \frac{e^{2t} \text{ext}_H(e^{-t} F_1)}{\text{ext}_G(F_1)} = \frac{e^{2t} i(F_1, F_2)}{i(F_1, F_2)} = e^{2t}.$$

Thus,

$$d_T(G, H) = \frac{1}{2} \log \frac{\text{ext}_H(F_1)}{\text{ext}_G(F_1)} = t.$$

For $t < 0$, we can obtain the required result by the same reasoning applied to the Teichmüller map from the structure to the structure $(e^{-t} F_1, e^t F_2)$ to the structure (F_1, F_2) . \square

2.6 On the asymptotic behaviour of Teichmüller rays

Definition 2.33 (Jenkins–Strebel geodesic). A *Jenkins–Strebel geodesic line* (respectively a *Jenkins–Strebel geodesic ray*) is a Teichmüller geodesic line (respectively geodesic ray) defined by $t \mapsto (e^{-t} F_1, e^t F_2)$ where the vertical measured foliation F_2 has all of its leaves compact.

In other words, a Jenkins–Strebel geodesic ray is a geodesic ray that starts at a Jenkins–Strebel structure (see Definition 2.19).

Recall that if a measured foliation has all its leaves compact, then it consists of a finite union of maximal foliated cylinders with disjoint interiors, each cylinder foliated by homotopic leaves. The homotopy classes of leaves associated to different maximal cylinders are distinct. Thus, to each Jenkins–Strebel geodesic line or ray we can associate a system of homotopy classes of disjoint and pairwise nonhomotopic essential simple closed curves on the surface, which we shall henceforth call the *vertical* system of homotopy classes. In other words, these are the homotopy classes of the nonsingular closed leaves of the associated vertical measured foliation. We shall use the following definition:

Definition 2.34 (Similar Jenkins–Strebel rays, cf. Masur [40], p. 211). Two Jenkins–Strebel geodesic rays are said to be *similar* if the associated vertical system of homotopy classes of closed curves are equal.

Theorem 2.35 (Masur [40]). Let S be a closed surface and let $r_1: [0, \infty) \rightarrow \mathcal{T}$ and $r_2: [0, \infty) \rightarrow \mathcal{T}$ be two similar Jenkins–Strebel rays starting at the same point. Then there exists $N > 0$ such that $d_T(r_1(t), r_2(t)) \leq N$ for all $t \in [0, \infty)$.

In his paper [29], Kerckhoff studied the relative asymptotic behaviour of pairs of Jenkins–Strebel rays that do not necessarily start at the same point. To state a result, we need the following

Definition 2.36 (Modularly equivalent Jenkins–Strebel rays, cf. Kerckhoff [29], p. 29). Suppose that $r_1: [0, \infty) \rightarrow \mathcal{T}$ and $r_2: [0, \infty) \rightarrow \mathcal{T}$ are two similar Jenkins–Strebel geodesic rays and let $\gamma_1, \dots, \gamma_n$ be their associated vertical system of homotopy classes of simple closed curves. For every $t \in [0, \infty)$, we consider the sequences of moduli m_1, \dots, m_n (respectively m'_1, \dots, m'_n) of the homotopy classes $\gamma_1, \dots, \gamma_n$, defined with respect to the Riemann surface $r_1(t)$ (respectively $r_2(t)$). We say that the two rays r_1 and r_2 are *modularly equivalent* if these two sequences of moduli are the same up to a multiplicative constant. (Note that if this condition holds for some $t > 0$, then it holds for any $t > 0$.)

Unlike the property of similarity between pairs of Jenkins–Strebel rays (Definition 2.34), modular equivalence is a property that involves both the vertical and horizontal foliations of the rays.

Theorem 2.37 (Kerckhoff, [29], p. 29). *Let S be a closed surface and let $r_1: [0, \infty) \rightarrow \mathcal{T}$ and $r_2: [0, \infty) \rightarrow \mathcal{T}$ be two similar Jenkins–Strebel rays (that do not necessarily start at the same point) whose vertical system of homotopy classes of simple closed curves have $3g - 3$ elements (which is the largest possible number of elements such a system can have). Then the following are equivalent:*

- (1) r_1 and r_2 are modularly equivalent;
- (2) $\lim_{t \rightarrow \infty} \inf_{M = \text{Im}(r_2(t))} d(M, \text{Im}(r_1)) = 0$. (Here, Im denotes the image set of a map.)

2.7 The Teichmüller boundary and convergent rays

The following definition is contained in Kerckhoff’s paper [29].

Definition 2.38 (The Teichmüller boundary). Let M be a point of \mathcal{T} . The *Teichmüller boundary of \mathcal{T} relative to the basepoint M* , denoted by $\partial_M \mathcal{T}$, is the set of geodesic rays (for the Teichmüller metric) starting at M , equipped with the topology of uniform convergence on compact sets.

Kerckhoff studied a compactification of Teichmüller space obtained by taking a basepoint M and putting an endpoint at each Teichmüller ray starting at that point. The result is a topological space $\bar{\mathcal{T}}_M = \mathcal{T} \cup \partial_M \mathcal{T}$ which is homeomorphic to a closed $(6g - 6 + 2n)$ -dimensional ball, equipped with a natural center (the point M) and a ray structure (i.e. an action of \mathbb{R}_+^*), in which the closure of the image of each ray starting at M is the union of that image with a single point in $\partial_M \mathcal{T}$ (see [29], p. 31). Kerckhoff

calls such a compactification of \mathcal{T} a *Teichmüller compactification* (the name referring to Teichmüller's metric and not only to Teichmüller space), and one of the main objects of the paper [29] is the study of the dependence of such a compactification on the basepoint. For instance, a natural question which Kerckhoff addresses is whether a Teichmüller ray, starting at some point other than the basepoint, has also a unique limit point in $\partial_M \mathcal{T}$. He makes the following definition:

Definition 2.39 (Convergent rays, cf. Kerckhoff [29], p. 28). Let M be a basepoint in Teichmüller space \mathcal{T} and consider the corresponding Teichmüller boundary $\partial_M \mathcal{T}$. Let $r': [0, \infty) \rightarrow \mathcal{T}$ be a geodesic ray starting at some other point M' in \mathcal{T} , and let $\overline{\text{Im}(r')}$ be its closure in $\overline{\mathcal{T}}_M$. Then r' is said to be *convergent* in $\overline{\mathcal{T}}_M$ if $\overline{\text{Im}(r')} \setminus \text{Im}(r')$ consists of a single point. In that case, $\overline{\text{Im}(r')} \setminus \text{Im}(r')$ is the endpoint of some geodesic ray r starting at M , and the ray r' is said to be *convergent* to the ray r .

For any element γ in \mathcal{S} and for any M in \mathcal{T} , we know, by Theorem 2.20, that there exists a unique Jenkins–Strebel ray $r_{M,\gamma}$ that starts at M and whose vertical foliation has one cylinder, with core curve in the homotopy class γ .

Theorem 2.40 (Kerckhoff [29], p. 29). *Let S be a closed surface. Then, for any homotopy class γ of essential simple closed curves and for any M and M' in \mathcal{T} , the Jenkins–Strebel ray $r_{M,\gamma}$ is convergent to the Jenkins–Strebel ray $r_{M',\gamma}$.*

Theorem 2.41 (Kerckhoff, [29], p. 29). *Let S be a closed surface and let $r_1: [0, \infty) \rightarrow \mathcal{T}$ and $r_2: [0, \infty) \rightarrow \mathcal{T}$ be two similar Jenkins–Strebel rays with vertical system of homotopy classes of curves having the largest possible number of elements (that is, $3g - 3$ elements). Then r_1 and r_2 are convergent if and only if they are modularly equivalent.*

2.8 Convergence to Thurston's boundary

A natural question to consider is whether a Teichmüller ray converges to a point on Thurston's boundary \mathcal{PMF} of Teichmüller space. We have postponed the definition of Thurston's boundary to Section 3 below (§3.13) since its natural setting is hyperbolic geometry, but we nonetheless state here a result of Masur about convergence of Teichmüller rays to points on that boundary:

Theorem 2.42 (Limits of Jenkins–Strebel rays, Masur [43]). *Let S be a closed surface, let $r: [0, \infty) \rightarrow \mathcal{T}$ be a Jenkins–Strebel ray and let $\alpha_1, \dots, \alpha_n$ be the associated vertical system of homotopy classes. Let $F = F(\alpha_1, \dots, \alpha_n)$ be a measured foliation whose nonsingular leaves are all compact, such that the complement of the critical graph of F consists of n cylinders whose core curves are in the classes $\alpha_1, \dots, \alpha_n$ and such that the height of each of these cylinders is equal to one. Then the ray r converges to the projective class $[F]$, considered as a point in Thurston's boundary of Teichmüller space.*

It follows from Theorem 2.42 that there are pairs of points on Thurston's boundary that are limits of a large family of Teichm  ller geodesics. Indeed, let F_1 and F_2 be two transverse measured foliations whose nonsingular leaves are all compact and such that for each of these foliations, the height of each foliated cylinder in the complement of the critical graph is equal to 1. One can see that such pairs exist from the fact that for each element $[F]$ in \mathcal{MF} , the subset $\mathcal{MF}(F)$ of equivalence classes representable by measured foliations transverse to a representative of $[F]$ is open, therefore, by the minimality of the action of the mapping class group on \mathcal{PMF} ([20], p. 117), this open set contains a conjugate of the equivalence class of any given measured foliation. By the same argument, for any $1 \leq m \leq 3g - 3$ and $1 \leq n \leq 3g - 3$, we can also choose F_1 (respectively F_2) to have its nonsingular closed leaves in m (respectively n) distinct homotopy classes. Let c_1, \dots, c_m (respectively c'_1, \dots, c'_n) be these homotopy classes. Let \bar{F}_1 (respectively \bar{F}_2) be now an arbitrary measured foliation having all of its nonsingular leaves compact and with associated homotopy classes c_1, \dots, c_m (respectively c'_1, \dots, c'_n). By Theorem 2.42, the Teichm  ller geodesic line $t \mapsto (e^{-t}\bar{F}_1, e^t\bar{F}_2)$ has $[F_1]$ and $[F_2]$ as its limit points. Note that if the projective equivalence classes of two such pairs (\bar{F}_1, \bar{F}_2) and (\bar{F}'_1, \bar{F}'_2) are distinct, then, by Teichm  ller's uniqueness theorem, the images of the two geodesics $t \mapsto (e^{-t}\bar{F}_1, e^t\bar{F}_2)$ and $t \mapsto (e^{-t}\bar{F}'_1, e^t\bar{F}'_2)$ have at most one common point. Thus, we have a large family of Teichm  ller geodesics whose limit points are $[F_1]$ and $[F_2]$.

We recall that a measured foliation F is said to be *uniquely ergodic* if F , as a topological foliation, carries a unique transverse measure up to a multiplicative factor. The property of being uniquely ergodic is invariant by Whitehead-equivalence and by homothety. Therefore one can talk about uniquely ergodic projective classes of measured foliations.

We shall also use the following terminology from [20]: a measured foliation F is said to be *arational* if it does not contain any closed curve made up of segments connecting singular points. Equivalently, up to performing Whitehead moves on F , every leaf of F is dense in the surface.

Theorem 2.43 (Limits of Teichm  ller rays with uniquely ergodic vertical foliations, Masur [43]). *Let S be a closed surface. Then any Teichm  ller ray whose associated vertical foliation is arational and uniquely ergodic converges in the positive direction to the projective equivalence class of that vertical foliation, considered as an element of Thurston's boundary of Teichm  ller space.*

There is a natural Lebesgue measure on the space \mathcal{MF} , which is provided, for instance, by train track coordinates (see [65] and [54]). By a result which was proved independently by Masur and by Veech (see [42] and [71]), the subspace of \mathcal{MF} consisting of equivalence classes of uniquely ergodic measured foliations is of full measure with respect to that measure. This measure on \mathcal{MF} defines a measure class on the set of geodesic rays starting at any given point M of Teichm  ller space. This can

be seen by referring to the result of Hubbard and Masur stated in Theorem 2.20 which says that for any element $M \in \mathcal{T}$ and for any measured foliation class $[F_1] \in \mathcal{MF}$, there exists a unique measured foliation class $[F_2] \in \mathcal{MF}$ such that the conformal structure defined by the pair (F_1, F_2) is M . Thus, the pair (F_1, F_2) can be regarded as a direction in Teichmüller space, at the point M , that is, the direction of the ray defined by $t \mapsto (e^{-t} F_1, e^t F_2)$ starting M . In this sense, Theorem 2.43 says that for any point M in Teichmüller space, almost any geodesic starting at M converges to a point in Thurston's boundary, which is the equivalence class of the vertical foliation of that ray.

In a recent preprint, A. Lenzhen gave an explicit example of a Teichmüller ray that does not have a limit in \mathcal{PMF} . We state this as

Theorem 2.44 (Teichmüller rays with no limit in \mathcal{PMF} , Lenzhen [35]). *On the closed surface of genus 2, there exists a Teichmüller ray that does not converge to any point in Thurston's boundary \mathcal{PMF} .*

The construction of that ray is based on an arithmetic property of the measured foliations associated to the geodesic ray considered.

The asymptotic behaviour of an arbitrary Teichmüller geodesic ray in terms of the dynamical/topological properties of its vertical foliation is still an open question.

2.9 On the visual boundary of the Teichmüller metric

We start by recalling a general definition.

Definition 2.45 (Asymptotic geodesic rays). Let (X, d) be a metric space. Two geodesic rays $r_1: [a_1, \infty) \rightarrow X$ and $r_2: [a_2, \infty) \rightarrow X$ are said to be *asymptotic* if the function defined on $[0, \infty)$ by $t \mapsto d(r_1(t + a_1), r_2(t + a_2))$ is bounded.

It is clear that the relation of being asymptotic is an equivalence relation on the set of geodesic rays in a metric space. We shall write $r_1 \sim r_2$ to say that the two geodesic rays r_1 and r_2 are asymptotic.

Definition 2.46 (The visual boundary). Let X be a proper geodesic metric space and let x be a point in X . The *visual boundary of X at x* , denoted by $\partial_{\text{vis},x} X$, is the set of equivalence classes (for the relation \sim) of geodesic rays starting at x , equipped with the quotient of the topology of uniform convergence on compact sets.

Note that the visual boundary at a point in Teichmüller space equipped with the Teichmüller metric is a quotient of the Teichmüller boundary relative to that point.

In the paper [46], based on Masur's result stated as Theorem 2.35 above, the following is proved

Theorem 2.47 (McCarthy–Papadopoulos [46]). *Let S be a closed surface. For any point x in the Teichmüller space $\mathcal{T}(S)$ equipped with the Teichmüller metric, the visual boundary $\partial_{\text{vis},x}\mathcal{T}(S)$ is not Hausdorff.*

This result was used to give a new proof of the fact due to Masur and Wolf that Teichmüller space equipped with the Teichmüller metric is not Gromov hyperbolic cf. [44].

3 Thurston’s asymmetric metric

3.1 Hyperbolic structures

Definition 3.1 (Hyperbolic structure). *A hyperbolic structure on S is a maximal atlas $\{(U_i, \phi_i)\}_{i \in \mathcal{I}}$ where for each $i \in \mathcal{I}$, U_i is an open subset of S and ϕ_i is a homeomorphism from U_i onto an open subset of the hyperbolic plane \mathbb{H}^2 , satisfying $\bigcup_{i \in \mathcal{I}} U_i = S$ and such that any map of the form $\phi_i \circ \phi_j^{-1}$ is, on each connected component of $\phi_j(U_i \cap U_j)$, the restriction of an orientation-preserving isometry of \mathbb{H}^2 .*

A surface equipped with a hyperbolic structure is called a *hyperbolic surface*. It carries a length metric defined as follows. On each chart domain U_i , we pull-back by the map ϕ_i the metric on $\phi_i(U_i)$ induced from its inclusion in \mathbb{H}^2 . These metrics on the various sets U_i give a consistent way of measuring lengths of paths in S , and the metric we consider on S is the associated length metric. It is a Riemannian metric of constant curvature -1 . The metric universal covering of S is a subset of hyperbolic plane \mathbb{H}^2 .

If we regard the various ϕ_i ’s as diffeomorphisms (instead of homeomorphisms), then we can pull-back the Riemannian metrics on $\phi_i(U_i)$ to the various sets U_i . This is another way of defining the Riemannian metric on S . In any case, a hyperbolic structure is also regarded as a metric on S , called a *hyperbolic metric*. We shall only consider hyperbolic metrics that are complete and of finite area. Equivalently, these are the metrics that have the property that each puncture of S has a neighborhood which is isometric to a *cusp*, that is, the quotient of a subset $\{x + iy \mid y > a\}$ ($a > 0$) of the upper half-plane model of \mathbb{H}^2 by the map $z \mapsto z + 1$. Equivalently, these are the metrics for which the metric universal cover is \mathbb{H}^2 (see [70]).

The group $\text{Hom}(S)$ of orientation preserving homeomorphisms of S acts on the set of hyperbolic structures in the same way as it acts on the set of conformal structures (cf. Definition 2.8), and, again, we say that two hyperbolic structures are *equivalent* if they are related by a homeomorphism of S which is isotopic to the identity.

There is a natural one-to-one correspondence between the set of equivalence classes of hyperbolic structures and the set of equivalence classes of conformal structures on S . This correspondence is given by the quotient of the map which assigns to each hyperbolic structure its underlying conformal structure (Example 2.6 above). The fact that the hyperbolic structure is complete and of finite area implies that the neighborhood

of each puncture is conformally a punctured disk (with respect to the conformal structure associated to that hyperbolic metric). Using this natural correspondence, one can make the following definition, which is an equivalent form of Definition 2.8 above:

Definition 3.2 (Teichmüller space). The *Teichmüller space* of S , denoted by $\mathcal{T}_{g,n}(S)$ or $\mathcal{T}(S)$, is the space of equivalence classes of complete finite area hyperbolic structures on S .

By abuse of language, we shall often call an element of Teichmüller space a “hyperbolic structure” instead of an equivalence class of hyperbolic structures. We shall also use interchangeably the terms “hyperbolic structure” and “hyperbolic metric”.

The topology of $\mathcal{T}(S)$ can be defined using several metrics. For instance, it is induced by the “quasi-isometry” metric d_{qi} given by

$$d_{qi}(g, h) = \frac{1}{2} \log \inf\{K\} \quad (3.1)$$

where the infimum is taken over the set of real numbers $K \geq 1$ such that there exists a homeomorphism $f: S \rightarrow S$ satisfying

$$\frac{1}{K} d_g(x, y) \leq d_h(f(x), f(y)) \leq K d_g(x, y)$$

(see Thurston [70], p. 266).

We can also describe the topology of $\mathcal{T}(S)$ by means of the *length functional* $l: \mathcal{T}(S) \rightarrow \mathbb{R}_+^{\mathcal{A}}$ defined by

$$g \mapsto l_g(\cdot): \alpha \mapsto l_g(\alpha) \quad \text{for all } \alpha \in \mathcal{A}, \quad (3.2)$$

with $l_g(\alpha)$ being the length of the unique g -geodesic in the homotopy class α . This map l is an embedding, and the topology of $\mathcal{T}(S)$ induced by the metric defined in (3.1) coincides with the one induced on the image of l by the weak topology on the space $\mathbb{R}_+^{\mathcal{A}}$. Thus, a sequence $(g_n)_{n \geq 0}$ in $\mathcal{T}(S)$ converges as $n \rightarrow \infty$ to an element $g \in \mathcal{T}(S)$ if and only if for every α in \mathcal{A} , we have $l_{g_n}(\alpha) \rightarrow l_g(\alpha)$ as $n \rightarrow \infty$. In fact, more is true: there exists a finite set $\{\alpha_1, \dots, \alpha_k\}$ of homotopy classes of simple closed curves in S such that an arbitrary sequence $(g_n)_{n \geq 0}$ in $\mathcal{T}(S)$ converges to an element g in that space if and only if for every $i = 1, \dots, k$ we have $l_{g_n}(\alpha_i) \rightarrow l_g(\alpha_i)$ as $n \rightarrow \infty$. (For a proof in the case of closed surfaces, see [20].) Finally, let us recall that this topology on $\mathcal{T}(S)$ is the one induced by Teichmüller's metric, and also by Thurston's asymmetric metric, which will be the main object of study in this section (see §3.7 below).

3.2 Decompositions by generalized pairs of pants and by ideal triangles

By studying the representation of conformal structures on a surface $S_{g,n}$ by hyperbolic structures one can easily get an intuition of the fact that its Teichmüller space is homeomorphic to an open ball of dimension $6g - 6 + 2n$, a fact that is harder to conceive

when one starts with the purely conformal point of view (that is, Definition 2.8). The statement that the space of deformations of equivalence classes of conformal structures on a closed orientable surface of genus $g \geq 2$ is of complex dimension $3g - 3$ was made by G. F. B. Riemann (without giving a formal proof) in his paper *Theorie der Abel'schen Functionen*, Crelle's J. f. M., B. 54 (1857). A homeomorphism $\mathcal{T}_{g,n} \simeq \mathbb{R}^{6g-6+2n}$, if one does not use hyperbolic geometry, can be obtained using Teichmüller's results. But when the surface $S_{g,n}$ is equipped with a hyperbolic structure, it is easy to produce such a homeomorphism using a decomposition of the surface into objects we call "generalized hyperbolic pairs of pants" with geodesic boundary. Another way to realize that homeomorphism, in the case where the set of punctures of S is not empty, can be achieved by using ideal triangulations. Our aim in this section is to discuss these two sorts of decompositions.

We recall that a *hyperbolic pair of pants* is a sphere with three open disks removed, equipped with a hyperbolic metric for which the three boundary components are geodesic.³ A *degenerate hyperbolic pair of pants* is a complete finite area hyperbolic surface which is either a sphere with three cusps, or a sphere with two cusps and one open disk removed, or a sphere with one cusp and two open disks removed, and where the boundary components, whenever they exist, are closed geodesics (see Figure 10). There is a precise sense in which any degenerate hyperbolic pair of pants is a limit of a family of hyperbolic pairs of pants. This can be seen from the construction of hyperbolic pairs of pants using right-angled hexagons, which we now briefly recall.

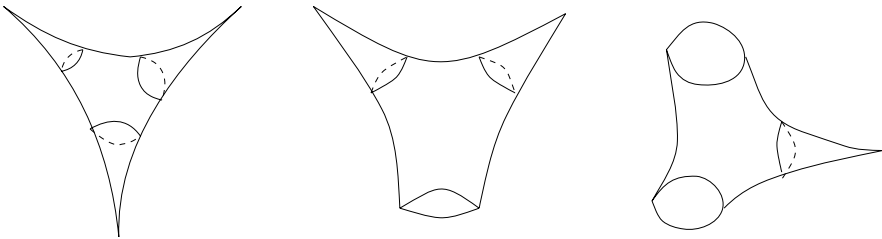


Figure 10. Degenerate hyperbolic pairs of pants.

The classification of hyperbolic structures on a pair of pants with geodesic boundary is usually done by decomposing the pair of pants into two isometric right-angled hexagons. The isometry class of a right-angled hexagon is completely determined by the lengths of any three alternating edges. For an exposition of this fact, we refer the reader to Thurston's book [70] (Exercise 2.4.11, Example 4.6.7 and Figure 4.15).

Consider a right-angled hexagon in hyperbolic space \mathbb{H}^2 . If we make the length of one of its edges tend to zero while keeping fixed the other two which form with the degenerating edge an alternating triple of edges, and keeping all the angles to be right

³In talking here about closed curves in the surface S , it should be noted that the term *geodesic* is used in the sense of Riemannian geometry, that is, in the sense of (the image of) a locally isometric map, as opposed to the other possible use of that word, denoting an isometric map.

angles, the vertices of the degenerating edge converge to a single point in the boundary of hyperbolic space, and the right-angled hexagon becomes, at the limit, an object we call a *degenerate right-angled hexagon* (see Figure 11 for a picture of such a hexagon).

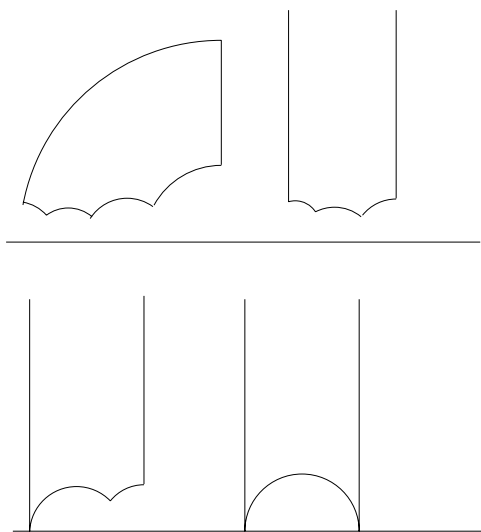


Figure 11. These four generalized polygons represent respectively a right-angled hexagon and three degenerate right-angled hexagons in the upper half-plane model of the hyperbolic plane.

Likewise, we can make (successively) the lengths of two or of three alternating edges tend to zero. We obtain in this manner degenerate right-angled hexagons with one, two or three edges at infinity, called *spikes*. Note that a degenerate right-angled hexagon with three edges at infinity is a hyperbolic ideal triangle (that is, the convex hull, in the hyperbolic plane \mathbb{H}^2 , of three distinct points in the boundary of that plane). A degenerate right-angled hexagon with one, two or three edges at infinity is completely determined up to isometry by the lengths of the edges that make with the one (or the ones) at infinity a triple of alternating edges. Since a hyperbolic pair of pants is obtained by gluing two isometric right-angled hexagons along three alternating edges, it is easy to see that the isometry type of the pair of pants is completely determined by the lengths of its three boundary geodesics. Likewise, a hyperbolic structure on a degenerate pair of pants is completely determined by the lengths of its boundary geodesic curves, a cusp being considered as a boundary curve of length zero. In this sense, the hyperbolic structure is rigid at the cusps. For instance, a degenerate right-angled hexagon with three alternating edges at infinity is a hyperbolic ideal triangle, and, as is well known, there is a unique hyperbolic ideal triangle up to isometry. In the same way, any two degenerate hyperbolic pairs of pants that are spheres with three cusps are isometric. We shall call a *generalized hyperbolic pair of pants* either

a hyperbolic pair of pants in the usual sense, or a degenerate hyperbolic pair of pants with one, two or three cusps.

A closed surface of genus $g \geq 2$ is decomposable into $2g - 2$ pairs of pants, by using $3g - 3$ simple closed curves (see [20]). Using this fact, we can count the number of curves in a generalized pair of pants decomposition of a surface of genus g with n cusps by first replacing each cusp (if such a cusp exists) by a boundary closed curve, obtaining a surface with boundary that is homeomorphic to a closed surface of genus g with n open disks deleted. Doubling that surface along its boundary curves, we obtain a closed surface of genus $2g + n - 1$, which we can decompose into pairs of pants by using $6g - 6 + 3n$ closed geodesics. By symmetry, we conclude that the number of closed curves needed to decompose the surface $S_{g,n}$ into generalized pairs of pants is $3g - 3 + n$.

Now we can count the parameters for the Teichmüller space of a surface of genus g with n punctures. Consider a set of closed geodesics decomposing that surface into generalized hyperbolic pairs of pants. There are $3g - 3 + n$ such geodesics. There are two parameters associated to each geodesic, one parameter (in \mathbb{R}_+^*) being the length of the geodesic itself, and the other parameter (in \mathbb{R}) describing the twist along that curve, when we glue together its two sides. (The twist parameter is defined up to the choice of an origin corresponding to the zero twist, and up to an orientation of the boundary component.) This gives two sets of $3g - 3 + n$ real parameters each. Thus, we have $6g - 6 + 2n$ parameters, which is an indication of the fact that Teichmüller space is homeomorphic to $\mathbb{R}^{6g-6+2n}$. For a geometric and concise proof in the case where $n = 0$, we refer the reader to Thurston [70], p. 271. There are also proofs of that fact in [1], [8], [20], [27]. These length-twist parameters associated to a (generalized) pair of pants decomposition are called (generalized) *Fenchel–Nielsen parameters*.

Another way of obtaining a homeomorphism $\mathcal{T}_{g,n} \simeq \mathbb{R}^{6g-6+2n}$ uses a decomposition of $S_{g,n}$ into ideal triangles, and we now describe it in the case where $n \geq 1$. Recall that any hyperbolic ideal triangle has a well-defined center (its center of gravity), and one distinguished point on each of its edges, which is the orthogonal projection on that edge of the center of gravity, or, equivalently, the orthogonal projection on that edge of the ideal vertex that is opposite to that edge. An *ideal triangulation* λ of $S = S_{g,n}$ is a decomposition of the associated closed surface \hat{S} into triangular cells whose vertices are all at the punctures. If S is equipped with a hyperbolic structure, then one can make the edges of λ to be embedded bi-infinite local geodesics with limit points at the punctures. (The bi-infinite local geodesics are, by definition, embedded images of the bi-infinite geodesic lines in \mathbb{H}^2 realized as the metric universal covering of the surface.) In this way, each face of λ is isometric to the interior of a hyperbolic ideal triangle, and the hyperbolic structure on S is determined by the gluing maps between the edges of these ideal triangles. In this gluing, there is one parameter associated to each edge of λ , which we can take to be the algebraic distance, called the *shift parameter*, between the two distinguished points on that edge, each distinguished point being associated to one of the triangles adjacent to that edge. (Recall that using the orientation on the surface S , one can define a notion of a left and of a right shift along the edges of μ .

The sign of the distance between the distinguished points is then defined accordingly.) Thus, the Teichmüller space $\mathcal{T}_{g,n}$ of $S_{g,n}$ is seen from that point of view as the set of shift parameters on the edges of an ideal triangulation λ of that surface, and a rough computation of the dimension of $\mathcal{T}_{g,n}$ can be done in the following way. Let E denote the number of edges of λ and let T be the number of its faces (triangles). Since each triangle has 3 boundary edges and since each edge is adjacent to 2 triangles, we have $E = 3T/2$. The Euler characteristic of $S_{g,n}$ is $\chi = 2 - 2g - n$. We construct an auxiliary foliation F on the surface \hat{S} obtained from $S_{g,n}$ by filling in the punctures, such that F is transverse to the edges of the triangulation λ , with a 3-prong singularity in the interior of each triangle, and with a center-type singularity at each puncture. There is an index formula relating the sum of the indices at the singular points of such a foliation to the Euler characteristic χ of the surface (see for instance [20], p. 75). The formula says that 2χ is equal to the sum of the indices of the singular points of the foliation. The index of a 3-prong singularity is -1 and the index of a center-type singularity is 0 . Therefore, -2χ is equal to the number of 3-prong singularities, which is the number of triangles. In other words, we have $T = -2\chi$. Thus, we obtain $E = \frac{3}{2}(-2)(2 - 2g - n) = 6g - 6 + 3n$. Now for the dimension of Teichmüller space, there is one shift parameter associated to each edge and one equation associated to each puncture; this is the equation which ensures that the hyperbolic structure near that puncture is complete, which is equivalent to the fact that around each puncture of $S_{g,n}$, there is a foliation by closed leaves made up of pieces of horocycles that are perpendicular to the edges that abut on that puncture, see Definition 3.6 below. Therefore, the number of parameters is $E - n = 6g - 6 + 3n - n = 6g - 6 + 2n$, as expected.

3.3 Geodesic laminations

Definition 3.3 (Geodesic lamination). Let S be a hyperbolic surface and let $p: \mathbb{H}^2 \simeq \tilde{S} \rightarrow S$ be its metric universal covering. A *geodesic lamination* on S is a closed subset of S which is the union of disjoint images of bi-infinite geodesics of \mathbb{H}^2 by the covering map p such that if $\ell \subset \tilde{S}$ is such a bi-infinite geodesic, then, either

- (1) the restriction of p to ℓ is injective, or
- (2) $p(\ell)$ is a simple closed geodesic on S .

The decomposition of a geodesic lamination as a union of such images of bi-infinite geodesics of \mathbb{H}^2 is unique, and each such image is called a *leaf* of λ . There are several good references on geodesic laminations. We refer the reader to Thurston's original notes [65], to the book by Penner and Harer [54] or to the book by Casson and Bleiler [15].

By abuse of language, we shall sometimes call the leaves of λ geodesics (rather than local geodesics). We shall talk about *bi-infinite leaves* of λ to denote leaves of kind (1) above, and about *closed leaves* to denote leaves of kind (2).

It is well known that there is a natural one-to-one correspondence between geodesic laminations associated to any two hyperbolic structures on a given surface (see [36]). This makes it possible to talk about a geodesic lamination on a surface without reference to any particular hyperbolic structure. In fact, we shall sometimes call a geodesic lamination on a surface S an object which is a geodesic lamination for some hyperbolic structure on S . Equivalently, a geodesic lamination on S will be an object which is isotopic to a geodesic lamination for any hyperbolic structure on S .

A geodesic lamination is said to be *complete* if there is no geodesic lamination that strictly contains it. Equivalently, a geodesic lamination λ is complete if each connected component of $S \setminus \lambda$, equipped with its intrinsic metric, is isometric to the interior of a hyperbolic ideal triangle.

An ideal triangulation of a surface $S = S_{g,n}$ with $n \geq 1$ is an example of a complete geodesic lamination on S .

If λ is a geodesic lamination and if μ is any complete geodesic lamination containing it, then we shall call μ a *completion* of λ .

Any non-complete geodesic lamination can be completed in several ways, and it is interesting to understand these different ways because they will correspond to various ways of constructing geodesics for Thurston's asymmetric metric which we shall study below. It is possible to complete any geodesic lamination λ by inserting a finite number of bi-infinite geodesics in its complement. The surface $S \setminus \lambda$ has a finite number of connected components, and the metric completion of each such connected component is a hyperbolic surface with geodesic boundary (see [15]). In particular, if λ contains closed leaves, then, in order to obtain a completion of λ , we can add geodesics that spiral around the closed leaves. For instance, in any hyperbolic pair of pants (with geodesic boundary), we can draw three bi-infinite geodesics that spiral around the boundary components so that the complement of these bi-infinite geodesics in the pair of pants is a union of the interiors of two hyperbolic ideal triangles (see Figure 12). In fact, there are 32 distinct ways of completing a hyperbolic pair of pants.

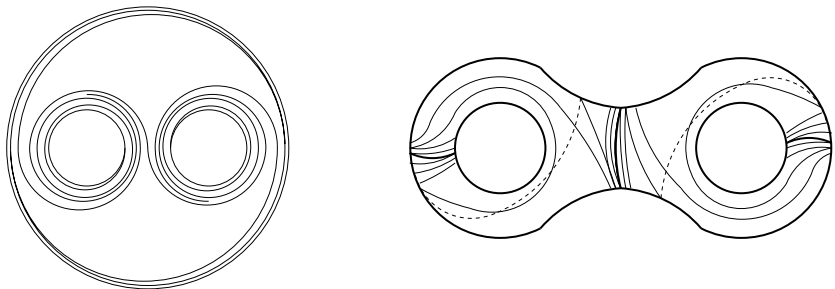


Figure 12. One way of obtaining complete geodesic laminations on a closed surface is to divide that surface into hyperbolic pairs of pants, and then decompose each pair of pants by bi-infinite leaves that spiral around the boundary components, as in the example drawn here. The stump of the complete lamination that we obtain is the union of the closed geodesics defining the pair of pants decomposition.

If the lamination λ that we started with does not have closed leaves, then we can still complete it by adding a finite number of bi-infinite geodesics that “spiral along bi-infinite leaves of λ ”. Spiraling (along closed geodesics or along bi-infinite leaves) can be defined as an operation in the universal covering of S , that consists in adding in an equivariant manner to the lift of the lamination λ a certain number of geodesic lines that converge in each direction to the limit point of the lift of some (finite or infinite) leaf of λ . To be more precise, let us first suppose that all the leaves of λ are closed. Let α and β be two closed leaves in λ and let us consider a segment c having one endpoint on α and one endpoint on β , and whose interior is disjoint from λ . Starting from the segment c , we define a continuous family of segments $(c_t)_{t \geq 0}$ in S , with $c_0 = c$, and such that for any $t \geq 0$, the endpoints of c_t are on α and on β , and spin around these curves as t increases to ∞ with constant speed. For each $t \geq 0$, we can take the segment c_t to be a local geodesic, its interior being disjoint from the lamination λ . Let $p: \tilde{S} \rightarrow S$ be again the metric universal covering of S , with \tilde{S} identified with the upper half-plane \mathbb{H}^2 . Consider a continuous family of segments $(\tilde{c}_t)_{t \geq 0}$ in \mathbb{H}^2 , which lifts the family $(c_t)_{t \geq 0}$ in S . For each $t \geq 0$, the endpoints of the segment \tilde{c}_t are on lifts $\tilde{\alpha}$ and $\tilde{\beta}$ of α and β . The family of geodesic segments $(\tilde{c}_t)_{t \geq 0}$ converges in an obvious sense to the image of a bi-infinite geodesic line $\tilde{c}: \mathbb{R} \rightarrow \mathbb{H}^2$ whose endpoints are endpoints of $\tilde{\alpha}$ and $\tilde{\beta}$. The distance from the point $\tilde{c}(t)$ to the geodesic lines $\tilde{\alpha}$ and $\tilde{\beta}$ tends to 0 as $t \rightarrow \infty$ or $t \rightarrow -\infty$ (and the convergence is of the order of e^{-t}). The covering map $p: \tilde{S} \rightarrow S$ restricted to \tilde{c} is injective, and the image of \tilde{c} by this map is a bi-infinite geodesic line in S which spirals at one end around α and at the other end around β . We take this geodesic as a new leaf in the completion of λ .

To continue completing the geodesic lamination λ by filling-in the connected components of $S \setminus \lambda$ with bi-infinite geodesics, we must consider the general case (that is, the case where leaves of λ are not all closed). Each completed connected component of the surface S cut along λ is a surface with boundary, with each boundary component being either a closed geodesic or a union of bi-infinite geodesics, as in Figure 13. We can decompose each connected component of $S \setminus \lambda$ into ideal triangles by succes-

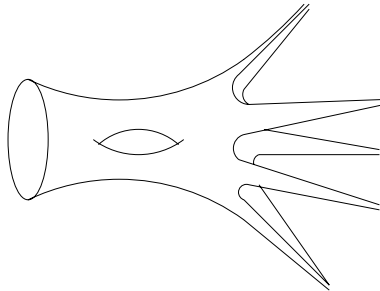


Figure 13. A possible component of the surface S cut along the lamination μ .

sively inserting a finite number of bi-infinite geodesics which at each end either spiral along a closed leaf or converge to one of the ends of a bi-infinite leaf or converge to

a cusp. The number of leaves that we can add is bounded in terms of the topology of the surface since we eventually decompose the surface into ideal triangles, each of them having a fixed area (equal to π), and the area of the hyperbolic surface S is a topological invariant.

Definition 3.4 (Measured geodesic lamination). *A measured geodesic lamination λ on a hyperbolic surface is a geodesic lamination which is equipped with a nonnegative Radon measure on each transverse arc. Furthermore, the measure is invariant under homotopies of the arc respecting the lamination λ and the support of the measure is equal to the intersection of the arc with the support of λ .*

To each measured geodesic lamination λ is associated an element of the function space \mathbb{R}_+^δ , by a rule analogous to the one that associates to each measured foliation an element of \mathbb{R}_+^δ (see (2.2) above). On a given hyperbolic surface, two distinct measured geodesic laminations define distinct elements of \mathbb{R}_+^δ , unlike pairs of measured foliations which are more flexible objects. (There is no need to talk about isotopy or Whitehead-equivalence relation between geodesic laminations on the same hyperbolic surface.) The two subsets of \mathbb{R}_+^δ consisting respectively of images of equivalence classes of measured foliations and images of measured geodesic laminations coincide, and this fact can be used to define a one-to-one correspondence between spaces of equivalence classes of measured foliations on a surface S and spaces of measured geodesic laminations on the surface S equipped with a hyperbolic structure. Note that there is an alternative and more geometric procedure to associate to each measured foliation class a measured geodesic lamination, which gives the same correspondence between the two sets (see for instance [36]).

We already mentioned that there is a natural one-to-one correspondence between sets of geodesic laminations on the surface S equipped with two distinct hyperbolic structures. This correspondence leads to a metric-independent definition of a geodesic lamination. In fact, Thurston shows in his Notes [65] that two geodesic laminations which correspond by that natural correspondence are isotopic as topological objects, by a global isotopy of S that fixes the cusps. Therefore, on a topological surface S , a geodesic lamination can also be thought of as an object defined up to isotopy.

The *support* of a lamination is the union of its leaves. By abuse of notation, we shall sometimes denote the support of a measured geodesic lamination λ by the same letter λ .

We denote by $\mathcal{ML}(S)$ the set of measured geodesic laminations on S , and by $\mathcal{ML}_0(S) \subset \mathcal{ML}(S)$ the subset consisting of measured geodesic laminations whose support is compact. Note that a compactly supported measured geodesic lamination cannot be complete if the surface has punctures, since by the definition of the transverse measure for a geodesic lamination, the support of the measure is equal to the support of the lamination, and therefore the lamination cannot have leaves which converge to punctures (the closure of such a leaf cannot be compact, and a complete geodesic lamination has necessarily such a leaf).

There is a natural action of the group \mathbb{R}_+^* of positive reals on the space $\mathcal{ML}(S)$, and the quotient of $\mathcal{ML}(S)$ by this action is the set $\mathcal{PML}(S)$ of projective classes of measured geodesic laminations, whose elements are called *projective measured geodesic laminations*. The subset of $\mathcal{PML}(S)$ consisting of the projective classes of compactly supported measured geodesic laminations is denoted by $\mathcal{PML}_0(S)$. One can define topologies on the spaces $\mathcal{ML}(S)$ and $\mathcal{ML}_0(S)$ by using the inclusion maps of these spaces in the space \mathbb{R}_+^δ and taking the restriction of the weak topology. In this way, the correspondences between spaces of measured geodesic laminations and spaces of equivalence classes of measured foliations become homeomorphisms.

Definition 3.5 (The stump of a geodesic lamination, [63]). The *stump* of a geodesic lamination μ is the support of any maximal (with respect to inclusion) compactly supported measured geodesic sublamination λ of μ .

The stump of a geodesic lamination is itself a geodesic lamination, and it is empty if and only if each leaf of μ converge at each end towards a cusp of the surface. For instance, if μ is complete, then the stump of μ is empty if and only if μ is an ideal triangulation.

Completing a compactly supported measured geodesic lamination λ by adding spiraling leaves gives a complete geodesic lamination whose stump is λ .

We shall sometimes consider the stump of a lamination μ as being equipped with some transverse measure, and we shall also call this measured lamination the *stump* of μ , although the transverse measure it carries is not unique.

We shall use the notion of length of a (compactly supported) measured geodesic lamination and we now recall the definition.

There is a function defined on $\mathcal{T}(S) \times \mathcal{ML}_0(S)$ which to every hyperbolic structure g and every compactly supported measured geodesic lamination λ on S associates a quantity denoted by $l_g(\lambda)$ and called the *length* of λ with respect to g . This function is a continuous extension of the notion of length of a simple closed geodesic, the set of simple closed geodesics being considered as a subset of the space of measured geodesic laminations, each simple closed geodesic being equipped with the Dirac measure of mass one. The definition of $l_g(\lambda)$ can be made by covering the geodesic lamination λ by a finite collection of quadrilaterals R_1, \dots, R_n with disjoint interiors, such that the intersection of the leaves of λ with each quadrilateral is a union of segments that crosses it from one side to the opposite side. We shall call these two sides the “vertical” sides of R_i . We choose, for each quadrilateral R_i , one vertical side, and we call it $\partial_v R_i$. The length $l_g(\lambda)$ is then the sum of the integrals, over all the quadrilaterals R_i , of the lengths of the geodesic segments of λ that cross them, with respect to the transverse measure $d\lambda$ of λ induced on the vertical sides $\partial_v R_i$. In formulae, if we denote by α_x , $x \in \lambda \cap \partial_v R_i$, the collection of geodesic sub-segments of λ that traverse the quadrilateral R_i , we have

$$l_g(\lambda) = \sum_{i=1}^n \int_{\partial_v R_i \cap \lambda} l_g(\alpha_x) d\lambda(x).$$

3.4 Horocyclic foliations

We need to consider measured foliations on S which are slightly different from those we defined in Section 2.1. These foliations naturally appear in the study of hyperbolic surfaces with cusps.

Definition 3.6 (Measured foliation trivial around the punctures). A measured foliation F on S is said to be *trivial around the punctures* if F is a measured foliation in the sense of Definition 2.1 except for the condition at the punctures, which is replaced by the following: each puncture has a neighborhood on which the induced foliation is a cylinder foliated by homotopic closed leaves, and any segment transverse to F and converging to a puncture has infinite total mass with respect to the transverse measure of F .

In what follows, we shall sometimes omit the adjective ‘‘trivial around the punctures’’ when the context is clear. We can pass from a measured foliation trivial around the punctures to a measured foliation in the sense of Definition 2.1 by deleting, for each puncture of S , the maximal annulus foliated by closed leaves parallel to that puncture, obtaining thus a partial measured foliation on the surface, and then collapsing the complementary components of the support of that partial foliation in order to get a foliation whose support is the entire surface S . The resulting measured foliation is well defined up to equivalence, except that it may be empty. Thus, to each measured foliation which is trivial around the punctures we can naturally associate an object which is either a well-defined element in \mathcal{MF} or the empty foliation.

For any fixed hyperbolic structure on S , we next define an object which is ‘‘dual’’ to a complete geodesic lamination and which we call the associated ‘‘horocyclic foliation’’. We start with a construction in an ideal triangle. Any hyperbolic ideal triangle is equipped with a canonical partial measured foliation which is called the *horocyclic foliation* of the triangle. This partial foliation is characterized by the following three properties:

- (1) the leaves are pieces of horocycles that are perpendicular to the edges;
- (2) the non-foliated region is a triangle bordered by three of these pieces of horocycles (see Figure 14);
- (3) the transverse measure assigned to any arc which is contained in an edge of the ideal triangle coincides with the Lebesgue measure induced from the hyperbolic metric.

Note that the non-foliated triangular region intersects each edge of the ideal triangle at its distinguished point.

Definition 3.7 (Horocyclic measured foliation). Let S be a hyperbolic surface and let μ be a complete geodesic lamination on S . The completion of each connected component of $S \setminus \mu$ being an ideal triangle, we can equip it with its horocyclic measured foliation.

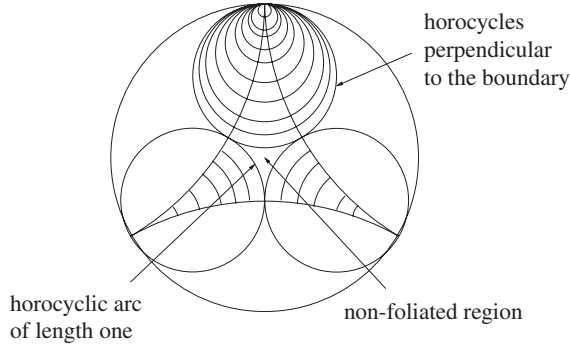


Figure 14. The horocyclic foliation of an ideal triangle.

The union of the horocyclic measured foliations associated to the various ideal triangles fit together smoothly, since they are perpendicular to the edges of the ideal triangles and therefore they form a Lipschitz-line field on the surface. They define a partial measured foliation of S , which we call the *horocyclic measured foliation associated to μ and S* , and which we denote by $F_\mu(S)$. Collapsing each nonfoliated triangular piece onto a tripod, we obtain a genuine (i.e. not partial) measured foliation on the surface, which is well defined up to isotopy, and which we also call the horocyclic measured foliation associated to μ .

The fact that the hyperbolic structures that we consider are complete and have finite area is equivalent to the fact that for any complete geodesic lamination μ , the associated horocyclic foliation $F_\mu(S)$ is trivial around the punctures.

The isotopy class of $F_\mu(S)$ depends only on the isotopy class of the hyperbolic structure and on the complete geodesic lamination μ that we started with. Therefore, the equivalence class of $F_\mu(S)$ is associated to the hyperbolic structure S viewed as an element of Teichmüller space. As usual, $[F_\mu(S)] \in \mathcal{MF}$ denotes the equivalence class of the horocyclic foliation $F_\mu(S)$ after removal of the closed leaves that are parallel to punctures (provided $F_\mu(S)$ is not the empty foliation).

Definition 3.8 (Horocyclic measured geodesic lamination). Let S be a hyperbolic surface and let μ be a complete geodesic lamination on S . The measured geodesic lamination that represents the (partial) measured foliation $F_\mu(S)$ (after removal of the leaves that are parallel to the punctures, and provided the remaining measured foliation is not the empty foliation) will be called the *horocyclic measured geodesic lamination associated to μ* , and it will be denoted by $\lambda_\mu(S)$.

The horocyclic measured geodesic lamination $\lambda_\mu(S)$ is the empty lamination if and only if the horocyclic foliation $F_\mu(S)$ is a union of cylinders foliated by leaves that are all parallel to punctures. It is easy to construct examples of hyperbolic structures whose associated horocyclic foliation has this property. For instance, we can start, as

in Figure 15, with an ideal quadrilateral in \mathbb{H}^2 which is the union of two ideal triangles glued along two edges in such a way that the distinguished points on the common edges agree. We then glue the remaining four edges by pairs, respecting the labels a and b in Figure 15, again in such a way that the distinguished points on these edges coincide after the gluing. The resulting hyperbolic surface is a once-punctured torus, and the resulting lamination μ on that surface is the union of the images of the edges of the two ideal triangles that we started with.

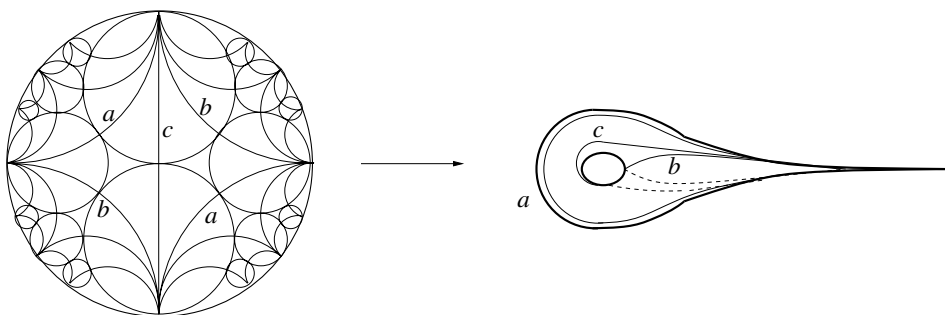


Figure 15. An example of a hyperbolic structure on the punctured torus equipped with a complete geodesic lamination μ which is an ideal triangulation. The gluing of the various ideal triangles in the universal covering is performed in such a way that for each edge of the ideal triangulation μ , the two distinguished points corresponding to the two triangles adjacent to it coincide. In this case, the associated horocyclic measured geodesic lamination is empty.

3.5 Thurston's cataclysm coordinates

In this section, we review coordinates for Teichmüller space that were introduced by Thurston, who called them “cataclysm coordinates” and which are essential in the study of Thurston’s asymmetric metric. The parameter space is a space of measured foliations that are transverse to a complete geodesic lamination. We start with the following:

Definition 3.9 (Measured foliation totally transverse to a lamination). Let μ be a geodesic lamination on S . A measured foliation F on S is said to be *totally transverse* to μ if it satisfies the following:

- (1) F is transverse to μ ;
- (2) F is trivial around the punctures.

In this definition, we have tacitly chosen an auxiliary hyperbolic structure on S in order to talk about geodesic laminations, but recall that we can make a definition which

is independent of that choice. In other words, we say that μ is a geodesic lamination if μ is a geodesic lamination for some hyperbolic structure on S .

A measured foliation class is said to be *totally transverse* to μ if it can be represented by a measured foliation that is totally transverse to μ .

For any geodesic lamination μ on S , we denote by $\mathcal{MF}(\mu) \subset \mathcal{MF}$ the set of measured foliation classes that are totally transverse to μ . This subset $\mathcal{MF}(\mu)$ is open in \mathcal{MF} .

If μ is complete, then, for any choice of a hyperbolic metric g on S , the associated horocyclic measured foliation $F_\mu(g)$ is totally transverse to μ . If h is a hyperbolic metric which is isotopic to g , then the horocyclic foliation $F_\mu(h)$ is isotopic to $F_\mu(g)$. This enables us to associate to each element g in $\mathcal{T}(S)$ a well-defined element in $\mathcal{MF}(\mu)$. We denote by $\phi_\mu: \mathcal{T}(S) \rightarrow \mathcal{MF}(\mu)$ the resulting map.

Theorem 3.10 (Thurston, [67], §9). *For any complete geodesic lamination μ on S , the map $\phi_\mu: \mathcal{T}(S) \rightarrow \mathcal{MF}(\mu)$ is a homeomorphism.*

The global coordinates for Teichmüller space that are provided by this map ϕ_μ are called by Thurston *cataclysm coordinates*.

The proof of Theorem 3.10 is given by Thurston in §4 and §9 of the paper [67]. This proof involves a lot of interesting details which we shall review below. In particular, we shall discuss in length the proof of the surjectivity of the map ϕ_μ , that is, the fact that for any element $[F]$ of $\mathcal{MF}(\mu)$, there exists a hyperbolic structure $g \in \mathcal{T}(S)$ such that $[F] = [F_\mu(g)]$. This surjectivity is contained in Proposition 9.2 of [67].

Let us first make a few remarks about the proof of the surjectivity of ϕ_μ . Given an element $[F] \in \mathcal{MF}(\mu)$, the transverse measure of a representative F of $[F]$ determines a measure on each leaf of μ , which we think of as the one-dimensional Lebesgue measure induced on the leaves of μ by the hyperbolic metric that we seek. Of course this is not enough to determine the hyperbolic metric; to do this, one needs to define an isometry between each connected component of $S \setminus \mu$ and a hyperbolic ideal triangle, and then define in a consistent way a gluing of the various ideal triangles among themselves. In the case where μ is an ideal triangulation, that is, if every leaf of μ is isolated, then the metric g can easily be defined by gluing edge-to-edge the ideal triangles (which are finite in number). This gluing is determined by the measured foliation by saying that the algebraic distances between the distinguished points on each edge of μ that are associated to the hyperbolic metric and to the measured foliation F coincide. Here, the distinguished points on an edge of μ that are associated to a measured foliation F are the hitting points of the singular leaves starting at the singular points of F that are contained in the ideal triangles that are on each side of that edge. But in the case where μ has non-isolated leaves, the gluing has to be defined by an infinite process, and the main point of the proof is the convergence of that process.

Let us fix a geodesic lamination μ on S . We start by proving some useful facts about measured foliations totally transverse to μ .

Lemma 3.11. *Let β be an infinite half-leaf of μ . Then exactly one of the following occurs:*

- (1) β converges to a cusp;
- (2) β has a recurrence point contained in the stump of μ (that is, there exists a point x in the stump of μ such that β returns infinitely often in any neighborhood of x).

Proof. Suppose that β does not converge to a cusp. Then there is a nonempty compactly supported geodesic sub-lamination in its closure. Since each compactly supported geodesic lamination admits a nonzero transverse measure (which may be of smaller support), we obtain a compactly supported measured geodesic sublamination μ_0 of μ in the closure of β . The sub-lamination μ_0 is contained in the stump of μ . The lemma follows then from the fact that each point in a measured geodesic lamination of compact support is a recurrence point. \square

Lemma 3.12 (Infinite measure for infinite half-leaves of μ). *Let F be a measured foliation totally transverse to μ . Then the F -transverse measure of any infinite half-leaf of μ is infinite.*

Proof. We use Lemma 3.11. If β converges to a cusp, then it has infinite transverse measure, by assumption on F (Definitions 3.9 and 3.6). Suppose now that β does not converge to a cusp and let x be a recurrence point of β in the stump γ of μ . We choose a homeomorphism $\varphi: U \rightarrow [0, 1]^2$ such that $U \subset S$ contains x in its interior, such that the image of $F|_U$ by φ is made of vertical segments (which we call plaques) defined by the equations $u = \text{constant}$, where u is the coordinate of the first factor of $[0, 1]^2$, and such that the connected components of the intersection of μ with U are segments joining the ‘‘vertical’’ sides $\{u = 0\}$ and $\{u = 1\}$ of $[0, 1]^2$. In particular, there is a segment of $\gamma \cap U$ containing x and crossing the plaques transversely from one vertical side to the other vertical side. Since x is a recurrence point of β , there are infinitely many segments in $\beta \cap U$ whose images by φ cross the square $[0, 1]^2$ from one vertical side to the other one. All these segments have the same transverse measure. Therefore, the transverse measure of β is infinite. \square

In the rest of this section, the geodesic lamination μ is complete.

Let $[F]$ be an element of $\mathcal{MF}(\mu)$ and let us choose a representative F of $[F]$ that is totally transverse to μ .

From the transversality of F to μ and by an Euler characteristic argument, we can see that F has exactly one singular point in each connected component of $S \setminus \mu$ or on the boundary of that component, and that this singular point is a 3-prong singularity. Up to a small homotopy of F , we can suppose that all the singular points of F lie in the interior of the components of $S \setminus \mu$.

We first show that F is homotopic (in a sense that will be made precise) in each ideal triangle to the horocyclic foliation of some hyperbolic structure on S .

We fix an ideal triangle T^* in \mathbb{H}^2 , equipped with its horocyclic foliation, modified by collapsing the non-foliated region onto a spine (that is, a tripod). We call F_{T^*} this

measured foliation. We perform the collapse in such a way that it induces the identity map on the boundary of T^* .

Let T be a connected component (including its boundary) of $S \setminus \mu$ equipped with the measured foliation induced by F . Notice that T has the structure of a topological ideal triangle. (In fact, it becomes a genuine ideal triangle as soon as we equip the surface S with a hyperbolic metric, but we do not need to introduce arbitrarily such a metric right now.) We have the following:

Lemma 3.13 (Equivalence to horocyclic foliation). *There exists a homeomorphism $\varphi_T: T \rightarrow T^*$ respecting the foliations of these triangles (that is, sending leaves to leaves and preserving the transverse measures) which is unique up to the choice of a one-to-one correspondence between the ideal vertices of T and those of T^* , and up to post-composition of φ_T with a homeomorphism of T^* that preserves its horocyclic foliation and that induces the identity map on the boundary of that triangle.*

Proof. Let F_T denote the measured foliation induced by F on T . Choose a one-to-one correspondence between the ideal vertices of the triangles T and T^* . There is a homeomorphism $\varphi_T: T \rightarrow T^*$ preserving the singular points and mapping each leaf of F_T to a leaf of F_{T^*} , preserving the transverse measures. To define φ_T , it suffices to start by mapping the three singular leaves of F_T to the three singular leaves of F_{T^*} (respecting the correspondence between these leaves induced by the correspondence between the ideal vertices) and then continue mapping homeomorphically each leaf of F_T onto a leaf of F_{T^*} preserving the transverse measure. This is possible because the transverse measure induced by F_T on each half-edge of T is infinite (Lemma 3.12). We thereby get a map $\varphi_T: T \rightarrow T^*$ which is well defined up to the deformations mentioned in the statement of the lemma. \square

A *spike* of a hyperbolic ideal triangle of \mathbb{H}^2 is a connected component of the complement of the closure of the unfoliated region of its horocyclic measured foliation (or a connected component of the complement of the singular graph, if the unfoliated region has been collapsed onto a spine). Likewise, given a complete geodesic lamination μ on S together with a totally transverse foliation F , a *spike* of $S \setminus \mu$ is a connected component of the complement of the singular leaves of the restriction of F to a connected component of $S \setminus \mu$.

Using Lemma 3.13, we now define a function on each spike of $S \setminus \mu$. Let q be such a spike and let T be the triangle of $S \setminus \mu$ that contains it. We choose a homeomorphism φ_T between the triangle T and a fixed hyperbolic triangle T^* in \mathbb{H}^2 that preserves the measured foliations of these triangles, as in Lemma 3.13. The *sharpness function* $f_q: q \rightarrow [0, \infty)$ is defined by

$$f_q(x) = -\log \text{length}(\alpha(\varphi_T(x)))$$

where, for each $x \in q$, $\alpha(\varphi_T(x))$ denotes the leaf of the horocyclic foliation corresponding to the leaf of F_{T^*} passing through the point $\varphi_T(x) \in T^*$. In other words, $\alpha(\varphi_T(x))$ is the horocyclic arc that passes through the point of T^* that is mapped to

the point $\varphi_T(x)$ by the map $T^* \rightarrow T^*$ that we used to collapse the unfoliated region of that ideal triangle onto a tripod. It is easy to check that the function f_q does not depend on the choices involved in the definition of φ_T . Since the length of any arc of the horocyclic foliation of an ideal triangle is bounded from above by 1 (which is the length of the horocyclic arc that is on the boundary of the unfoliated region), $f_q(x)$ is nonnegative for all x in q . Furthermore, this function converges to ∞ as the point x converges to the cusp of q . (In fact, f_q is a linear function with respect to the ‘‘distance’’ from x to the unfoliated region, where this distance is defined by the transverse measure of any transverse arc joining x to a point in the unfoliated region.)

Thus, associated to the measured foliation F , we have a family $\{f_q\}$ of functions, one for each spike q of $S \setminus \mu$. We shall use these functions to construct the hyperbolic structure g satisfying $[F] = [F_\mu(g)]$. Following Thurston, to construct g , we construct a map from the universal covering of S to \mathbb{H}^2 . This map will turn out to be the developing map of that structure.

Let us choose a basepoint s on S and let $\pi : \tilde{S} \rightarrow S$ be the topological universal covering of S . We regard \tilde{S} as the set of homotopy classes of paths with fixed endpoints $\alpha : [0, 1] \rightarrow S$ with $\alpha(0) = s$; the projection map $\pi : \tilde{S} \rightarrow S$ is given by $[\alpha] \mapsto \alpha(1)$.

Given such a path α , we can replace it by a path α^* which is homotopic to α by a homotopy with fixed endpoints, which is made of a finite concatenation of segments, each such segment being contained either in a leaf of F or in a leaf of μ . We call such a path α^* a *horo-geodesic* path. (Of course, we are using this terminology because we imagine S as being equipped with a hyperbolic structure g for which μ is geodesic, with F being the corresponding horocyclic measured foliation; in fact, g will be the structure that we are seeking.) The construction of α^* may be done as follows. We start by taking a train track approximation τ of μ , and for this we can use an auxiliary hyperbolic metric for which μ is geodesic, and take the leaves of τ so that they are nearly parallel to the geodesic leaves of μ , in the metric sense, as in [65], §9.5.

We can cover μ by a finite collection of rectangles obtained by thickening the edges of the train track τ . The horizontal sides of these rectangles are parallel to the edges of τ , and the leaves of μ cross each such rectangle from a vertical side to a vertical side. We also take τ to be close enough to μ in such a way that F is transverse to τ . Moreover, up to taking smaller rectangles, we can assume that the foliation induced by F on each rectangle consists in segments joining its horizontal sides. Performing a homotopy with fixed endpoints on the path α , we can replace it by a path whose intersection with each rectangle is a finite number of segments joining its horizontal sides. Finally, replacing each such segment (by using again a homotopy with fixed endpoints) by a concatenation of (at most three) segments in leaves of μ and in leaves of F , we obtain the desired horo-geodesic path α^* . From this construction, we can assume without loss of generality that α^* satisfies the following two properties:

- (1) each subsegment of α^* contained in a leaf of F or in a leaf of μ does not backtrack on that leaf;

- (2) there does not exist any closed immersed disk in S whose interior is embedded in that surface and whose boundary is the union of a subsegment of α^* with a segment which is either in a leaf of F or in a leaf of μ .

A horogeodesic path satisfying these two properties will be called a *minimal horogeodesic path*.

From now on, we regard \tilde{S} as the set of homotopy classes with fixed endpoints of minimal horogeodesic paths $\alpha^*: [0, 1] \rightarrow S$ satisfying $\alpha^*(0) = s$.

We note that in the construction that follows, a change in the auxiliary hyperbolic metric will only affect the developing map of the hyperbolic structure that we are seeking by conjugating it by an orientation-preserving isometry of \mathbb{H}^2 . Precisely, this hyperbolic structure can be seen as the plane \mathbb{H}^2 mod out by a subgroup of the group of isometries of \mathbb{H}^2 acting properly discontinuously on that space and defined up to conjugation, and therefore the choice of the auxiliary hyperbolic metric will not affect the final result.

We now fix a minimal horogeodesic path α^* . We shall say that a subpath of α^* is *maximal* if its image is either contained in a leaf of F or in a leaf of μ , and if this subpath is maximal (with respect to inclusion) for that property. We shall call a maximal subpath of α^* , if it is contained in a leaf of F (respectively of μ), a *maximal vertical* (respectively *horizontal*) subpath. We shall consider the image of α^* as being equipped with the natural orientation induced from that of the interval $[0, 1]$.

Let c be a maximal vertical subpath of α^* . The connected components of the complement of c with respect to μ form a linearly ordered set of open intervals (a_j, b_j) , $j \in J$, each interval (a_j, b_j) joining opposite edges of a topological ideal triangle T_j of $S \setminus \mu$. To each such interval (a_j, b_j) , we now associate a parabolic isometry of \mathbb{H}^2 . Recall that to the triangle T_j containing (a_j, b_j) , there is an associated homeomorphism φ_{T_j} which is defined up to an isotopy that induces the identity map on the boundary of that triangle and up to a permutation of the ideal vertices, such that φ_{T_j} maps T_j onto some fixed ideal triangle T^* of \mathbb{H}^2 equipped with its horocyclic foliation, respecting the foliations and the transverse measures. We also have an associated family of sharpness functions $\{f_q\}$, one function for each spike q of $S \setminus \mu$. Let us fix an oriented geodesic line γ in \mathbb{H}^2 which we shall henceforth refer to as the *base geodesic*, and let us choose a basepoint x_0 on (the image of) that geodesic. For each j in J and for each choice of a hyperbolic ideal triangle T^* having the base geodesic γ as one of its edges and situated to the right of γ (we are using an orientation of the hyperbolic plane), we consider the associated orientation-preserving map φ_{T_j} between T_j and T^* that sends the edge of T_j containing a_j to the base geodesic γ . (The fact that φ_{T_j} preserves orientations implies that the images of the vertices of T_j are well defined.) As the ideal vertex of T^* that is not an endpoint of γ varies on the circle at infinity, the image $\varphi_{T_j}(a_j)$ varies on the base geodesic γ , and we can manage so that $\varphi_{T_j}(a_j)$ is the point x_0 . Now let P_j be the parabolic isometry of \mathbb{H}^2 that fixes the endpoint of the spike of T^* that contains the segment $\varphi_{T_j}((a_j, b_j))$ and that sends $\varphi_{T_j}(a_j)$ to $\varphi_{T_j}(b_j)$. Note that the length of the horocyclic arc joining the points $\varphi_{T_j}(a_j)$ and $\varphi_{T_j}(b_j)$ equals $e^{-f_q(a_j)}$. After associating a parabolic isometry P_j to each open interval (a_j, b_j) , we

associate to the subpath c of α^* the (possibly infinite) product of isometries $\prod_{j \in J} P_j$ (see Figure 16). Note that the association of the product of isometries is in the reverse order of the one used for composition of maps. Thus, for instance, if α crosses the ideal triangles T_1, T_2, T_3 in that order with associated parabolic isometries P_1, P_2, P_3 respectively, then the product will be $P_1 P_2 P_3$.

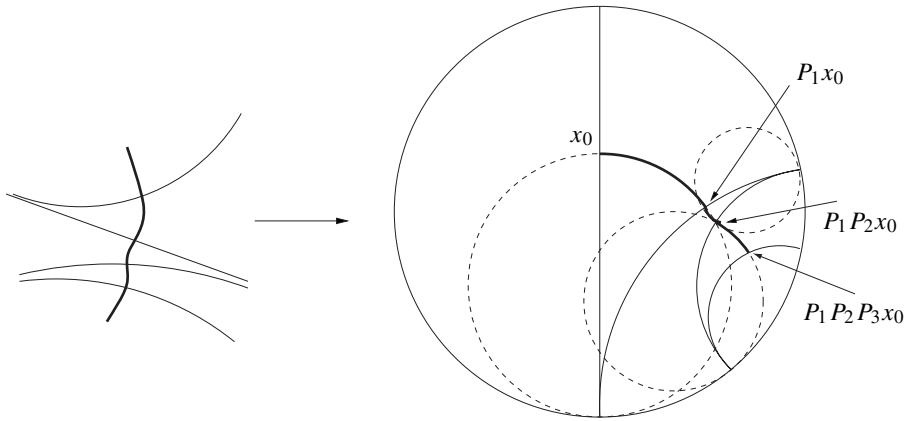


Figure 16

Let $\text{Isom}_+(\mathbb{H}^2)$ be the group of orientation-preserving isometries of \mathbb{H}^2 . Given a basepoint x_0 in \mathbb{H}^2 , we endow this group with the following complete left-invariant metric

$$d_{x_0}(A, B) = \sup_{x \in \mathbb{H}^2} |A(x) - B(x)| e^{-|x_0 - x|} \quad \text{for all } A, B \in \text{Isom}_+(\mathbb{H}^2),$$

where for every x and y in \mathbb{H}^2 , $|x - y|$ denotes their hyperbolic distance. It is easy to see that if $x_1 \in \mathbb{H}^2$ is another basepoint, then we have

$$d_{x_1}(A, B) \leq d_{x_0}(A, B) e^{|x_0 - x_1|}. \tag{3.3}$$

The distance between $A \in \text{Isom}_+(\mathbb{H}^2)$ and the identity I is denoted by $\|A\|_{x_0} = \|A\|$. Using the left-invariance of the metric d_{x_0} , it is easy to see that we have $d_{x_0}(A, B) = \|B^{-1}A\|$ for all A and B in $\text{Isom}_+(\mathbb{H}^2)$. We need to estimate the norm of a parabolic element.

Lemma 3.14. *Let P be a parabolic element fixing a point p in S_∞^1 , the boundary at infinity of \mathbb{H}^2 , and let $l(0)$ denote the length of the horocyclic arc joining x_0 and $P(x_0)$ and contained in the horocycle centered at p . Then $\|P\| \leq l(0)$.*

Proof. Let $\gamma : \mathbb{R} \rightarrow \mathbb{H}^2$ be the geodesic line in \mathbb{H}^2 satisfying $\gamma(0) = x_0$, having p as one of its endpoints and oriented so that $\lim_{t \rightarrow -\infty} \gamma(t) = p$. Consider the foliation of \mathbb{H}^2 by the family $\{H_t : t \in \mathbb{R}\}$ of horocycles centered at p , where for each t in \mathbb{R} ,

H_t denotes the horocycle through $\gamma(t)$, and let $l(t)$ denote the length of the arc in H_t joining $\gamma(t)$ to $P(\gamma(t))$.

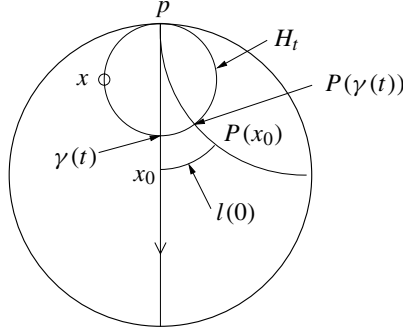


Figure 17. Notation used in Lemma 3.14.

Let $x \in \mathbb{H}^2$. This point is on a unique horocycle H_t , $t \in \mathbb{R}$. We have $|P(x) - x| = |P(\gamma(t)) - \gamma(t)|$. Moreover, a geometric argument shows that the projection along horocycles is distance-decreasing, that is, $|x_0 - x| \geq |x_0 - \gamma(t)| = |t|$. Therefore, $\|P\| = \sup_{t \in \mathbb{R}} |P(\gamma(t)) - \gamma(t)|e^{-|t|}$.

A computation (for instance, using Figure 36 below) gives $|P(\gamma(t)) - \gamma(t)| = 2 \sinh^{-1}(l(t)/2)$, where $l(t)$ denotes the length of the horocyclic arc joining $\gamma(t)$ and $P(\gamma(t))$. Using the fact that $l(t) = e^{-t}l(0)$ and using the inequality $\sinh^{-1}(x) \leq x$ for every $x \geq 0$, we obtain

$$\begin{aligned} \|P\| &= \sup_{t \in \mathbb{R}} |P(\gamma(t)) - \gamma(t)| e^{-|t|} \\ &= \sup_{t \in \mathbb{R}} 2 \sinh^{-1} \left(\frac{l(t)}{2} \right) e^{-|t|} \\ &\leq l(0) e^{-t-|t|} \leq l(0). \end{aligned}$$

This proves Lemma 3.14. □

Proposition 3.15. *The product of isometries $\prod_{j \in J} P_j$ that is associated to the maximal vertical subpath c of α^* converges in $\text{Isom}_+(\mathbb{H}^2)$.*

Proof. In order to show that the product $\prod_{j \in J} P_j$ converges, we shall use the fact that $\text{Isom}_+(\mathbb{H}^2)$ is complete. Let $J_0 \subset J_1 \subset \dots \subset J_n \subset \dots$ be a sequence of finite subsets of J converging to J , that is, such that $\cup_{n \in \mathbb{N}} J_n = J$. Without loss of generality, we can assume that $\text{Card}(J_n) = n$. For each $n \in \mathbb{N}$, the finite product $\prod_{j \in J_{n+1}} P_j$ is obtained from the previous finite product $\prod_{j \in J_n} P_j$ by inserting one element e_n . More precisely, we can find elements $A_n, B_n, e_n \in \text{Isom}_+(\mathbb{H}^2)$ such that $\prod_{j \in J_n} P_j = A_n B_n$

and $\prod_{j \in J_{n+1}} P_j = A_n e_n B_n$. We have

$$\begin{aligned}
 d_{x_0}(A_n e_n B_n, A_n B_n) &= \|B_n^{-1} e_n B_n\| \\
 &= \sup_{x \in \mathbb{H}^2} |B_n^{-1} e_n B_n(x) - x| e^{-|x_0 - x|} \\
 &= \sup_{y \in \mathbb{H}^2} |B_n^{-1} e_n(y) - B_n^{-1}(y)| e^{-|x_0 - B_n^{-1}(y)|} \\
 &= \sup_{y \in \mathbb{H}^2} |e_n(y) - y| e^{-|B_n(x_0) - y|} \\
 &= d_{B_n(x_0)}(e_n, I) \\
 &\leq d_{x_0}(e_n, I) e^{|B_n(x_0) - x_0|} \quad (\text{using 3.3}) \\
 &\leq \|e_n\| e^{\|B_n\|}.
 \end{aligned}$$

Let Π_n denote $\prod_{j \in J_n} P_j$. We have

$$\begin{aligned}
 d_{x_0}(\Pi_{n+p}, \Pi_n) &\leq \sum_{k=1}^p d_{x_0}(\Pi_{n+k}, \Pi_{n+k-1}) \\
 &\leq \sum_{k=1}^p \|e_{n+k-1}\| e^{\|B_{n+k-1}\|}.
 \end{aligned}$$

Let q be a spike of $S \setminus \mu$. The intersection of q with α is a countable family of horocyclic arcs which, once reordered, have decreasing lengths l_n , $n \in \mathbb{N}$. There is a positive lower bound $M(F, c)$, depending on the compact segment c , to the set of numbers $\{l_n - l_{n+1}\}$. We have $l_n \leq l_0 e^{-nM(F, c)}$. Moreover, $l(0) \leq 1$. Therefore, for every spike q , we have $l_n \leq e^{-nM(F, c)}$. Since there are finitely many spikes in $S \setminus \mu$, then, using Lemma using 3.14, we can see that there exists an integer $m = m(n)$ such that $\|e_n\| \leq l_m$ where l_m is the length of the m -th horocyclic arc of $q \cap c$, for every spike q . Therefore, for all $\varepsilon > 0$, there exists $N = N(F, c) \geq 0$ such that for all $n \geq N$, we have $\|e_n\| \leq \varepsilon$. Thus, the infinite product $\prod_{j \in J} P_j$ is bounded from above by a convergent geometric series, whence the norm of any finite subproduct is uniformly bounded from above. This shows that the sequence $\{\Pi_n : n \in \mathbb{N}\}$ is Cauchy, hence converges. \square

Let c be now a maximal horizontal subsegment of α^* . We wish to associate to c a hyperbolic isometry of \mathbb{H}^2 . We do this by induction, and for this, we first consider the isometries that were already associated to the sequence of maximal subsegments of α^* that precede c . (In the case where c is the initial maximal subsegment of α^* , the preceding isometry is taken to be the identity.) We let C be the product of all these isometries, this time the product being taken in the usual order of composition of maps. In other words, if c_1, \dots, c_n is the ordered set of maximal subsegments of α^* that precede c (that is, c_1 starts at the point s and the endpoint of c_n is the starting point of c), and if C_1, \dots, C_n is the ordered set of isometries that we already associated to

c_1, \dots, c_n , then we let $C = C_n \dots C_1$ (and if c is the first maximal subsegment of α^* , we take C to be the identity). Now if we denote by γ' the image of the base geodesic γ by the isometry C , then, we associate to c the hyperbolic isometry $H \in \text{Isom}_+(\mathbb{H}^2)$ that fixes γ' , whose translation length is equal to the transverse measure of c with respect to F and such that the orientation of the action of H on γ' coincides with the natural orientation associated with the parametrization of α^* .

Finally, we associate to α^* the product of the finite sequence of isometries associated to the finite sequence of maximal subsegments of α^* . (Remember that the product is taken here in the order of composition of maps, as already mentioned.)

Lemma 3.16. *The isometry associated to the minimal horogeodesic path α^* does not depend on the choice of α^* in its homotopy class with fixed endpoints.*

Proof. Any two horogeodesic paths belonging to the same class $[\alpha]$ are related to each other by a finite number of moves of the type described in Figure 18.



Figure 18. The dotted lines represent leaves of the foliation F , the horizontal lines are leaves of μ and the bold lines represent the maximal subsegments of the path α^* . (There may be other leaves of μ that are inside the rectangles shown and that are not drawn.)

Therefore, it suffices to examine the invariance of the isometry associated to α^* by such a move. The product of isometries associated to the left-hand rectangle of α^* in Figure 18 is of the form PH , where H is the hyperbolic isometry associated to the maximal horizontal subsegment of α^* that is drawn in that figure and P is the product of parabolic isometries that are associated to the maximal vertical segment. Likewise, in the right-hand rectangle, the product is the form $H'P'$, where H' and P' are defined in an analogous way. From the definition of the products associated to the subsegments of α^* that are in F , it easily follows that $P = P'$, since these are products of a possibly infinite family of parabolic isometries indexed by the same set and which are all equal. From the invariance of the transverse measure of F , the translation distances of the two hyperbolic isometries H and H' are equal. From this fact and from the way these isometries are defined, it follows that $H = P^{-1}H'P'$, that is, $PH = H'P'$, which implies the invariance by the move considered. This completes the proof of the lemma. \square

Thus, we have a map $I_{s,F}$ from \tilde{S} to $\text{Isom}_+(\mathbb{H}^2)$, which associates to each homotopy class of curves with fixed endpoints the isometry associated to a minimal horogeodesic path representative, and which gives a map

$$D_{s,F}: \tilde{S} \longrightarrow \mathbb{H}^2$$

$$[\alpha] \mapsto I_{s,F}([\alpha]) x_0.$$

This map is a local isometry. Changes in the choice involved in any of the parameters s and F have the effect of conjugating the map $D_{s,F}$ by an isometry (see [70]).

This map induces a representation of the fundamental group in $\text{Isom}_+(\mathbb{H}^2)$, which is free and discrete, since the quotient of \mathbb{H}^2 by the image group $\Gamma_{[F]}$ is the surface S . We obtain in this way a hyperbolic structure g on S whose developing map is precisely the map $D_{s,F}$. The hyperbolic structure g is complete because each cusp of S equipped with that structure has a neighborhood equipped with a foliation by closed leaves that are quotients by $\Gamma_{[F]}$ of horocycles in \mathbb{H}^2 , which follows from the fact that the measured foliation F that we started with is trivial around the punctures. This measured foliation is isotopic to the horocyclic measured foliation $F_\mu(g)$. This completes the proof of the surjectivity of the map ϕ_μ . This map is also injective, since it admits an inverse, which is precisely the map which we produced in the proof of the surjectivity. It is clear from the construction of the horocyclic foliation that the map ϕ_μ is continuous. By invariance of domain, this map is a homeomorphism. This completes the proof of Theorem 3.10. \square

3.6 Stretch lines, stretch rays, anti-stretch rays and stretch maps

In the next definition, we use Thurston's cataclysm coordinates ϕ_μ introduced in §3.5.

Definition 3.17 (Stretch lines, stretch rays, etc.). Let μ be a complete geodesic lamination on S . For any element F in $\mathcal{MF}(\mu)$, the map from \mathbb{R} to \mathcal{T} (respectively from $[0, \infty)$ to \mathcal{T}) defined by $t \mapsto \phi_\mu^{-1}(e^t F)$ is called a *stretch line directed by μ and starting at $g = \phi_\mu^{-1}(F)$* . A stretch line travelled up backwards is called an *anti-stretch line*. More precisely, given a stretch line $t \mapsto \phi_\mu^{-1}(e^t F)$, the map from \mathbb{R} to \mathcal{T} defined by $t \mapsto \phi_\mu^{-1}(e^{-t} F)$ is called an *anti-stretch line* directed by μ and starting at $g = \phi_\mu^{-1}(F)$. A *stretch ray* (respectively *anti-stretch ray*) directed by μ and starting at $g = \phi_\mu^{-1}(F)$ is the restriction of a stretch line (respectively an anti-stretch line) to the half-line \mathbb{R}_+ . Given a stretch (respectively anti-stretch) ray $t \mapsto \phi_\mu^{-1}(e^t F)$ (respectively $t \mapsto \phi_\mu^{-1}(e^{-t} F)$) that starts at a point g in \mathcal{T} , we shall denote by g^t the hyperbolic surface $\phi_\mu^{-1}(e^t F)$ (respectively $\phi_\mu^{-1}(e^{-t} F)$). The hyperbolic structure g^0 is the structure g that we started with. Given a stretch ray starting at g , then, for each $t \geq 0$, the identity map, considered as a map from the surface S equipped with the hyperbolic structure g^0 to the same surface equipped with the hyperbolic structure g^t , is called the *stretch map directed by μ* .

We already mentioned that there exist punctured hyperbolic surfaces g equipped with complete geodesic laminations for which the associated horocyclic foliation is a union of cylinders that are all parallel to punctures. A stretch ray (or an anti-stretch ray) starting at a hyperbolic structure g with a horocyclic foliation satisfying this property is a constant ray, that is, we have $g^t = g$ for all $t \geq 0$. For instance, in the example of Figure 15, stretching along μ does not change the hyperbolic structure. One way

of seeing this is the fact that the relative distances between the distinguished points remain constant (equal to zero). On the other hand, any complete lamination μ which is not an ideal triangulation (that is, which is not a finite union of bi-infinite leaves converging at both ends to punctures) has a nonempty stump. Therefore, in this case, for all $t \neq 0$, the hyperbolic structure g^t is distinct from g (and for the same reason, for all $0 \leq t_1 < t_2$, g^{t_1} is distinct from g^{t_2}). To see this, it suffices to consider the effect of the stretch map on any nonempty sublamination of μ that can be equipped with an invariant measure (for instance, the stump of μ). In the surface g^t , the length of that measured sub-lamination is multiplied by the factor e^t , which implies that the structure g^t is distinct from g .

3.7 Thurston's asymmetric metric

An *asymmetric metric* on a set X is a map $L: X \times X \rightarrow \mathbb{R}_+$ that satisfies all the axioms of a metric except the symmetry axiom, and such that the symmetry axiom is not satisfied, i.e. there exist x and y in X such that $L(x, y) \neq L(y, x)$. Teichmüller space is equipped with an asymmetric metric that was defined by Thurston. The stretch lines that we considered in Section 3.6 are geodesics for that metric, provided the stump of the complete lamination μ is not empty. We now recall the definition of this metric.

Let g and h be two hyperbolic structures on S and let $\varphi: S \rightarrow S$ be a diffeomorphism of S which is isotopic to the identity. The *Lipschitz constant* $\text{Lip}(\varphi)$ of φ is defined as

$$\text{Lip}(\varphi) = \sup_{x \neq y \in S} \frac{d_h(\varphi(x), \varphi(y))}{d_g(x, y)}.$$

The logarithm of the infimum of these Lipschitz constants over all diffeomorphisms φ in the isotopy class of the identity is denoted by

$$L(g, h) = \log \inf_{\varphi \sim \text{Id}_S} \text{Lip}(\varphi). \quad (3.4)$$

It is obvious that L satisfies the triangle inequality. It is less obvious that $L(g, h) \geq 0$ for all g and h (this uses the fact that any two hyperbolic structures on S have the same area, see [67], Proposition 2.1).

Making g and h vary in their respective homotopy classes does not change $L(g, h)$ and thus we obtain a function which is well defined on $\mathcal{T}(S) \times \mathcal{T}(S)$, which is Thurston's asymmetric metric. We shall denote it by the same letter:

$$L: \mathcal{T}(S) \times \mathcal{T}(S) \rightarrow \mathbb{R}_+.$$

Thurston showed that the quantity $L(g, h)$ can also be computed by comparing lengths of homotopic closed geodesics for the metrics g and h . More precisely, for any homotopy class α of essential simple closed curves on S , we consider the quantity

$$r_{g,h}(\alpha) = \frac{l_h(\alpha)}{l_g(\alpha)}$$

and we set

$$K(g, h) = \log \sup_{\alpha \in \mathcal{S}} r_{g,h}(\alpha). \quad (3.5)$$

Again, it is easy to see that the function K satisfies the triangle inequality. The fact that we have $K(g, h) > 0$ for all $g \neq h$ is not trivial, and it is proved by Thurston in [67], Theorem 3.1.

Since each homeomorphism φ considered in Formula (3.4) is isotopic to the identity, it preserves each homotopy class of simple closed curves in S . From that observation, it is easy to see that $K \leq L$. In his paper [67], Thurston proves the following result:

Theorem 3.18 (Thurston [67], Theorem 8.5). $K = L$.

Note that in the same way as in the definition of the Teichm  ller metric (see Remark 2.25 above), by using the extension of the hyperbolic length function to the space of compactly supported measured geodesic laminations, we can replace the supremum in 3.5 by the supremum over the set of compactly supported measured geodesic laminations, that is, we have

$$K(g, h) = \log \sup_{\alpha \in \mathcal{M}\mathcal{L}_0} \frac{l_h(\alpha)}{l_g(\alpha)} = \log \sup_{\alpha \in \mathcal{P}\mathcal{L}_0} r_{g,h}(\alpha). \quad (3.6)$$

By compactness of the space $\mathcal{P}\mathcal{L}_0(S)$, the supremum in (3.6) is attained at some measured geodesic lamination.

Along a stretch line directed by a complete geodesic lamination μ whose stump γ is not empty, the stretch map between g and g^t is e^t -Lipschitz and $r_{g,g^t}(\gamma) = e^t$. This gives $L(g, g^t) = K(g, g^t) = t$ for all $t \geq 0$ (and justifies the dilatation factor e^t). A measured lamination α attaining the supremum in (3.6) is generally not unique, even up to a scalar factor. For instance, any nonempty sub-lamination of such a lamination has the same property. There exists a geodesic lamination that is canonically associated to an ordered pair of hyperbolic structures (g, h) , but this lamination is not necessarily equipped with a transverse measure. Thurston introduced it in the course of proving Theorem 3.18. He first defined the notion of a *maximally stretched geodesic lamination* from a hyperbolic structure g to a hyperbolic structure h . This is a (chain-recurrent) geodesic lamination λ for which there exists an $e^{K(g,h)}$ -Lipschitz map from a neighborhood of λ in the surface S equipped with the metric g to a neighborhood of λ in the surface S equipped with the metric h . Thurston proved that the union of all maximally stretched geodesic laminations from g to h is a geodesic lamination. This geodesic lamination is denoted by $\mu(g, h)$.

Remarks. 1) For any η in $\mathcal{M}\mathcal{L}_0(S)$, we can find two hyperbolic structures g and h such that the supremum in $r_{g,h}(\alpha)$ in (3.6) is attained for $\alpha = \eta$. This follows from

Thurston’s construction in [67] §4, taking in that construction a complete geodesic lamination μ that contains η and taking g and h (in the right order) on a stretch line directed by μ .

2) If an element η in $\mathcal{ML}_0(S)$ realizes the maximum of $r_{g,h}(\alpha)$ as above, then any other element of $\mathcal{ML}_0(S)$ which is topologically equal to η (that is, with a possibly different transverse measure) also realizes the maximum. This follows from Thurston’s definition of $\mu(g, h)$ in [67], §8, which is independent of the transverse measure of the lamination (see Theorem 8.1 of [67]).

One can see that the function L is generally not symmetric by looking at the following example.

Example 3.19 (The metric L is not symmetric). In Figure 19, the surface $S_{2,0}$ is equipped with two hyperbolic structures g and h for which the left and right handles are almost isometric. One can do explicit computations of lengths of closed geodesics, by decomposing the two hyperbolic surfaces into hyperbolic pairs of pants, with the central curve being an element of the decomposition, and with the two other closed curves being the core curves of the two handles represented in Figure 19. Consider the central cylinder that joins the handles. The height of this cylinder, measured in the structures g and h respectively, is approximately equal to c and d , with $d > c$. Applying the formulae that give the distance between two boundary components of a hyperbolic pair of pants in terms of the length of the boundary components (see e.g. [20], p. 151), we obtain $l_g(\alpha) \simeq Ae^{-c}$ and $l_h(\alpha) \simeq Ae^{-d}$, where A is a constant that depends on the size of the handles. Thus, the smallest Lipschitz constant of a map from g to h is approximately equal to d/c , whereas the smallest Lipschitz constant of a map from h to g is bounded below by $e^{-c}/e^{-d} = e^d/e^c$. The latter is very large compared to d/c when d is large compared to c . Thus, we obtain examples of hyperbolic structures g and h satisfying $L(g, h) \neq L(h, g)$.

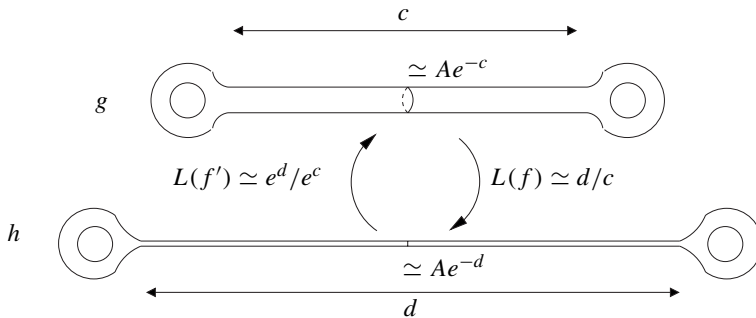


Figure 19. An example (due to Thurston) of hyperbolic structures g and h satisfying $L(g, h) \neq L(h, g)$.

On the other hand, there are instances of distinct hyperbolic structures g and h satisfying $K(g, h) = K(h, g)$ (hence $L(g, h) = L(h, g)$). An example of such a pair is given in Figure 20.

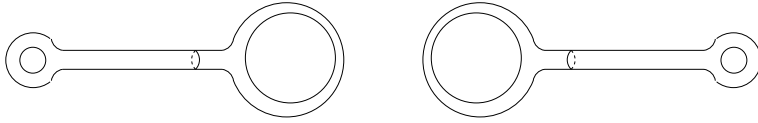


Figure 20. An example of distinct hyperbolic structures g and h satisfying $K(g, h) = K(h, g)$: h is the image of g by an order-two isometry which is not isotopic to the identity.

The notion of a geodesic (respectively of a geodesic line, or a geodesic ray) for an asymmetric metric space can be defined as in the case of a genuine metric, except that one has to be careful about the order of the variables in the case of an asymmetric metric. More precisely, we shall say that a map $\gamma: I \rightarrow \mathcal{T}(S)$, (where I is a closed interval of \mathbb{R} and where $\mathcal{T}(S)$ is equipped with the asymmetric metric L) is geodesic if for all $t \leq t'$ in I we have $t' - t = L(\gamma(t), \gamma(t'))$. Thurston proved the following

Theorem 3.20 (Stretch lines are geodesics, Thurston [67], Theorem 8.5 and Corollary 4.2). *Let μ be a complete geodesic lamination on S . If the stump of μ is not the empty lamination, then any stretch line in $\mathcal{T}(S)$ directed by μ is a geodesic line for Thurston's asymmetric metric. In other words, we have, for every $0 \leq t \leq t'$,*

$$L(\phi_\mu^{-1}(e^t F), \phi_\mu^{-1}(e^{t'} F)) = t' - t.$$

Theorem 3.21 (Concatenations of stretches, Thurston [67], Theorem 8.5). *Let g and h be two hyperbolic structures. Then we can pass from g to h by a geodesic which is a finite concatenation of pieces of stretch lines along complete geodesic laminations μ_1, \dots, μ_k , all of them containing $\mu(g, h)$. Furthermore, the number of such pieces needed to go from g to h is bounded by a constant which depends only on the topological type of the surface (that is, on the genus and the number of punctures). Such a geodesic path is obtained in the following way. We first choose an arbitrary completion μ_1 of $\mu(g, h)$ and we stretch the structure g along μ_1 until we reach a first point g' satisfying $\mu(g', h) \neq \mu(g, h)$. We necessarily have $\mu(g', h) \supset \mu(g, h)$. Starting now with g' instead of g and continuing in the same manner, we reach h after a finite number of steps.*

Remarks. 1) That the number of pieces in this statement is bounded from above follows from the fact that there is a bound (that depends only on the topological type of the surface) for the length of a strictly increasing (with respect to inclusion) sequence of geodesic laminations.

2) The geodesic path leading from g to h is in general not unique, since it depends on the choice of the completion at each step.

3) The bound in 1) does not imply that the number of distinct geodesics from g to h is finite, since at each step there are in general infinitely many ways of completing the given geodesic lamination.

4) An anti-stretch ray is (up to reparametrization) a geodesic ray for the asymmetric metric on Teichmüller space which is dual to Thurston's asymmetric metric, that is, the asymmetric metric defined by

$$K^*(g, h) = \log \sup_{\alpha \in \mathcal{S}} \frac{l_g(\alpha)}{l_h(\alpha)} \quad (3.7)$$

for any hyperbolic metrics g and h .

Here are some questions related to the example described in Figure 20.

- What does a typical geodesic from g to h look like? Is it a segment of a stretch line? Is it a concatenation of at least two such segments? Is it something else?
- Describe a geodesic from g to h that (up to reparametrization) is also a geodesic from h to g when traversed in the opposite direction.

We now discuss a few properties of Thurston's asymmetric metric L with respect to sequences in $\mathcal{T}(S)$ that tend to infinity. Then, we shall discuss the topologies that this asymmetric metric induces on $\mathcal{T}(S)$. Here, as usual, we say that a sequence (g_n) in $\mathcal{T}(S)$ tends to infinity (and write $g_n \rightarrow \infty$) if for any compact subset \mathcal{K} of $\mathcal{T}(S)$, we have $g_n \in \mathcal{T}(S) \setminus \mathcal{K}$ for all n large enough.

It is known that Teichmüller's metric d_T is proper (that is, closed balls are compact), which implies that a sequence of points (g_n) in \mathcal{T} tends to infinity if and only if for every g in \mathcal{T} (or, equivalently, for some g in \mathcal{T}), we have $d_T(g, g_n) \rightarrow \infty$ as $n \rightarrow \infty$. In [53], we prove the following analogous result for the asymmetric metric L .

Theorem 3.22. *For any sequence (g_n) in \mathcal{T} and for any $h \in \mathcal{T}$, we have the following equivalences:*

$$g_n \rightarrow \infty \iff L(h, g_n) \rightarrow \infty \iff L(g_n, h) \rightarrow \infty.$$

We deduce the following

Corollary 3.23. *Let $t \mapsto g^t$ be a stretch ray starting at a hyperbolic structure g and directed by a complete geodesic lamination μ , such that either the stump of μ is not empty (or, equivalently, μ is not an ideal triangulation) or the horocyclic measured lamination $\lambda_\mu(g)$ is not empty. Then the map $t \mapsto g^t$ is proper. (In other words, as t tends to infinity, the point g^t tends to infinity in Teichmüller space.)*

Proof. If the stump of μ is not empty, then we have, by Theorem 3.20, $L(g, g^t) = t$, which, by Proposition 3.22, implies that g^t tends to infinity. If the horocyclic measured lamination $\lambda_\mu(g)$ is not empty, then, as $t \rightarrow \infty$, we have $l_{g^t}(\lambda_\mu(g)) \rightarrow 0$. Indeed, by Proposition 3.3 and Lemma 3.9 of [52], there exists a constant $C > 0$ that depends only on the topological type of the surface S such that $l_{g^t}(\lambda_\mu(g^t)) \leq C$. (In fact, by

Lemma 3.9 of [52], we can take $C = -\chi$ where χ is the Euler characteristic of S . Note that the setting in [52] is that of closed surfaces, but the arguments work for general surfaces of finite type. We have $\lambda_\mu(g^t) = e^t(\lambda_\mu(g))$, therefore $l_{g^t}(\lambda_\mu(g)) \leq Ce^{-t}$, which tends to zero as t tends to infinity. This proves the claim. Now if $l_{g^t}(\lambda_\mu(g)) \rightarrow 0$, then $r_{g^t, g}(\lambda_\mu(g)) = \frac{l_g(\lambda_\mu(g))}{l_{g^t}(\lambda_\mu(g))} \rightarrow \infty$, which gives $L(g^t, g) \rightarrow \infty$, which, again by Proposition 3.22, implies that g^t tends to infinity. \square

The following was obtained by L. X. Liu in [38]. It is also proved in [53].

Proposition 3.24. *For any sequence (g_n) in $\mathcal{T}(S)$ and for any element g in $\mathcal{T}(S)$, we have the following equivalences:*

$$g_n \rightarrow g \iff L(g, g_n) \rightarrow 0 \iff L(g_n, g) \rightarrow 0.$$

Associated to the asymmetric metric L , for each g in \mathcal{T} and for each $R > 0$, there are two different notions of closed balls, which we call the *left closed ball* centered at $g \in \mathcal{T}(S)$ of radius R and the *right closed ball* centered at $g \in \mathcal{T}(S)$ of radius R , and which we denote respectively by ${}_gB(R)$ and $B_g(R)$. These are defined respectively by

$${}_gB(R) = \{h \in \mathcal{T}(S) \mid L(g, h) \leq R\} \quad (3.8)$$

and

$$B_g(R) = \{h \in \mathcal{T}(S) \mid L(h, g) \leq R\}. \quad (3.9)$$

In [53], we also prove the following

Proposition 3.25. *For any g in $\mathcal{T}(S)$ and for any $R > 0$, the closed balls ${}_gB(R)$ and $B_g(R)$, centered at $g \in \mathcal{T}(S)$ and of radius R , are compact for the usual topology.*

One can similarly define left (respectively right) open balls in \mathcal{T} , by taking the same definitions in (3.8) (respectively (3.9)) above, except that the large inequality is replaced by a strict inequality.

Since the metric on Teichmüller space that we are studying here is asymmetric, it is natural to consider two topologies on that space, the one generated by the collection of right open balls, and the one generated by left open balls. We shall call the first of these topologies the topology associated to Thurston's asymmetric metric K and the second one the topology associated to its dual asymmetric K^* . It is easy to see that a sequence (g_n) of points in Teichmüller space converges to a point g in that space for the topology generated by the left (respectively right) open balls if and only if we have $K(g, g_n) \rightarrow 0$ as $n \rightarrow \infty$ (respectively $K^*(g, g_n) \rightarrow 0$) (that is, we have the same convergence criteria than for genuine metrics). Proposition 3.24 is equivalent to the following

Corollary 3.26. *The topologies on Teichmüller space generated by Thurston's asymmetric metric and by its dual asymmetric metric coincide with the usual topology of that space.*

From Theorem 3.24, Proposition 3.25 and from Corollary 3.26, we obtain the following

Proposition 3.27. *Teichmüller space equipped with Thurston's asymmetric metric is proper. More precisely, left and right closed balls are compact for the topology generated by the asymmetric metric.*

Equivalently, we have the following

Proposition 3.28. *For every point g in Teichmüller space, the functions $h \mapsto L(g, h)$ and $h \mapsto L(h, g)$ are proper.*

Herbert Busemann developed a theory of spaces (X, δ) satisfying all the axioms of a metric space except the symmetry axiom, and satisfying the following additional axiom:

$$\delta(x, x_n) \rightarrow 0 \iff \delta(x_n, x) \rightarrow 0 \text{ for any } x \text{ and for any sequence } (x_n) \text{ in } X.$$

For such spaces, there is a well-defined topology on X , the one associated to the bona fide metric $\max\{\delta(x, y), \delta(y, x)\}$. Theorem 3.24 insures that Thurston's asymmetric metric fits into that theory, and that this topology is the same as the one generated by the collection of left (or right) open balls. Given such a space (X, δ) , Busemann introduced the following notion for such spaces (see [14], Chapter 1): (X, δ) is *complete* if and only if for every sequence $(x_n)_{n \geq 0}$ in X satisfying $\delta(x_n, x_{n+m}) \rightarrow 0$ as n and $m \rightarrow \infty$, the sequence converges to a point in X .

In that setting, the following generalization of a classical theorem of Hopf and Rinow holds (see [14], Theorem 8, p. 4). If a locally compact space (X, δ) satisfies Busemann's axioms stated above, and if this "generalized metric" δ is intrinsic in the sense that for any x and y in X , the value $\delta(x, y)$ is equal to the infimum of the lengths of all curves joining x and y , then the following two properties are equivalent:

- (1) left closed balls in X are compact;
- (2) X is complete.

Here, the length of a curve is defined in a similar way as in the case of a genuine metric space. We do not need to enter here into the details of this definition because, by a result of Thurston (see [67], §6) that we already mentioned, Thurston's asymmetric metric is geodesic (that is, any two points can be joined by a geodesic), and a generalized geodesic metric in this sense is intrinsic (see [14], p. 3).

Thus, we obtain the following

Corollary 3.29. *Thurston's asymmetric metric is complete.*

3.8 Comparison with other metrics

Other results on the comparison between the Teichmüller metric d_T and Thurston's asymmetric metric L on Teichmüller space have been obtained by L. X. Liu [37] and in joint work by Y. Choi and K. Rafi [16]. Let us start by stating the following result by Liu:

Theorem 3.30 (Liu [37]). *Let \mathcal{T} be the Teichmüller space of a closed surface of negative Euler characteristic. Then,*

- (1) *there does not exist any constant $k > 0$ such that $d_T(g, h) \leq kL(g, h)$ for every g and h in \mathcal{T} ;*
- (2) *there does not exist any constant $k > 0$ such that $d_T(g, h) \leq kL(h, g)$ for every g and h in \mathcal{T} .*

Choi and Rafi consider the following symmetrization d_L of Thurston's asymmetric metric, which they call the *Lipschitz distance* on Teichmüller space:

$$d_L(g, h) = \max\{L(g, h) + L(h, g)\}$$

for g and h in \mathcal{T} .

They obtain results which compare this metric with Teichmüller's metric. These results are based on a geometric study which they make, which is inspired by work of Yair Minsky, involving a Margulis constant for Teichmüller space, and the existence of regions in that space that have a product structure. (Whereas Minsky's work involves estimates of ratios of extremal lengths of curves, the work of Choi and Rafi involves estimates of ratios of hyperbolic lengths of curves.)

More precisely, Choi and Rafi take a number $\varepsilon_0 > 0$ which is less than the Margulis constant of Teichmüller space, and they define the *thick part* of Teichmüller space to be the set of points in that space which are represented by hyperbolic metrics whose injectivity radius is larger than ε_0 . Then, they say that a simple closed curve α which is not homotopic to a point is *short* with respect to a hyperbolic surface g if the hyperbolic length of α in g , $l_g(\alpha)$, is less than ε_0 . They prove the following:

Theorem 3.31 (Choi and Rafi, [16]). *Let \mathcal{T} be the Teichmüller space of a surface of finite type and of negative Euler characteristic. Then,*

- (1) *If g and h are in the thick part of \mathcal{T} , the distances $d_T(g, h)$ and $d_L(g, h)$ are equal up to an additive constant which is independent of g and h . More precisely, there exists $c > 0$ such that for all g and h in the thick part of \mathcal{T} , we have*

$$d_L(g, h) - c \leq d_T(g, h) \leq d_L(g, h) + C.$$

- (2) *If g and h are not in the thick part of \mathcal{T} and do not have a common short curve, the distances $d_T(g, h)$ and $d_L(g, h)$ are equal up to a multiplicative factor which is independent of g and h . More precisely, there exists $k > 0$ such that for all g*

and h in the thick part of \mathcal{T} , we have

$$\frac{1}{k}d_L(g, h) \leq d_T(g, h) \leq kd_L(g, h).$$

(3) In the general situation (that is, if g and h are allowed to have common short curves), the two distances $d_T(g, h)$ and $d_L(g, h)$ are not comparable. Precisely, there are sequences g_n and h_n in \mathcal{T} satisfying

$$d_L(g_n, h_n) \rightarrow 0 \text{ and } d_T(g_n, h_n) \rightarrow \infty.$$

In the same paper, Choi and Rafi also show that the thin part of Teichmüller space with the metric d_L has a product structure.

More precisely, let Γ be a collection of k disjoint and pairwise homotopically distinct simple closed curves on S , and consider the set

$$\text{Thin}_\varepsilon(S, \Gamma) = \{g \in \mathcal{T}(S) \mid l_g(\gamma) \leq \varepsilon \forall \gamma \in \Gamma\}.$$

Let

$$\mathcal{T}_\Gamma = \mathcal{T}(S \setminus \Gamma) \times U_1 \times \cdots \times U_k,$$

where $S \setminus \Gamma$ is the surface obtained from S by pinching all the curves in Γ and where for all $i = 1, \dots, k$, U_i is the subset $\{(x, y) \mid y \geq 1/\varepsilon\}$ of the upper-half plane \mathbb{H}^2 .

Finally, Choi and Rafi define the *sup metric*

$$d_{L_\Gamma} = \sup\{d_{L(S \setminus \Gamma)}, d_{L(A_1)}, \dots, d_{L(A_k)}\}$$

on \mathcal{T}_Γ , where $d_{L(S \setminus \Gamma)}$ is the Lipschitz metric on $\mathcal{T}(S \setminus \Gamma)$ and where for $i = 1, \dots, k$, $d_{L(A_i)}$ is a modification of the hyperbolic metric on U_i . Then, the Fenchel–Nielsen coordinates on \mathcal{T} give rise to a natural homeomorphism

$$\Pi : \text{Thin}_\varepsilon(S, \Gamma) \rightarrow \mathcal{T}_\Gamma,$$

and Choi and Rafi prove the following

Theorem 3.32 (Choi and Rafi, [16]). *For g and h in $\text{Thin}_\varepsilon(S, \Gamma)$, we have*

$$|d_L(g, h) - d_{L_\Gamma}(\Pi(g), \Pi(h))| = O(1).$$

In the next three sections, we study explicit examples of stretch lines that are simple enough to make computations.

3.9 Example I of a stretch line: the four-punctured sphere

In this section, $S = S_{0,4}$ is the four-punctured sphere. We consider the simple case of a hyperbolic structure on S that is obtained by gluing two ideal quadrilaterals that can be exchanged by an order-two isometry of S that fixes pointwise each of the four edges of the quadrilaterals. Let us call such a hyperbolic structure a *symmetric* hyperbolic structure on the four-punctured sphere. Figure 21 represents the surface S as a union of

two isometric quadrilaterals, with vertices A, B, C, D . Each quadrilateral is equipped with a diagonal in such a way that the two diagonals join the same cusps of the surface S , say D and B , as represented in Figure 21.

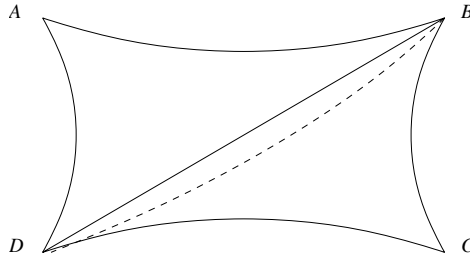


Figure 21. The four-punctured sphere with its ideal triangulation μ .

The union of the edges of the quadrilaterals (that are glued by pairs in the surface), together with the two diagonals that join D and B , constitute an ideal triangulation of S which we denote by μ . In the case considered of a symmetric structure, the fact that the hyperbolic structure is complete is equivalent to the fact that for each of the edges AB, BC, CD and DA , the two distinguished points (in the sense defined in §3.2) corresponding to the two ideal triangles that are adjacent to that edge coincide. Furthermore, since the two quadrilaterals we started with are isometric, the algebraic distance d between the distinguished points on each of the two diagonals joining the vertices D and B coincide. As a matter of fact, the set of symmetric hyperbolic structures on the four-punctured sphere is parametrized by \mathbb{R} , since such a hyperbolic structure is completely determined by the algebraic distance d .

The horocyclic foliation associated to a symmetric hyperbolic structure g together with the ideal triangulation μ is a foliation which has one cylinder whose height is equal to $|d|$ (which could be equal to 0) and four other cylinders forming neighborhoods of cusps. In other words, the associated horocyclic measured geodesic lamination $\lambda_\mu(g)$ is a simple closed geodesic α equipped with a Dirac transverse measure of mass $|d|$ (and it is the empty foliation if $d = 0$). Stretching the hyperbolic structure g in the direction μ always produces a symmetric hyperbolic structure. If $t \mapsto g^t, t \in \mathbb{R}$, is the stretch line starting at g and directed by μ , then, provided $d \neq 0$, we have the following:

- (1) $l_{g^t}(\alpha) \rightarrow 0$ as $t \rightarrow \infty$;
- (2) the distance between the distinguished points on the diagonal BD for the hyperbolic structure g^t equals e^t times that distance for the hyperbolic structure $g = g^0$;
- (3) if ν is any compactly supported measured geodesic lamination on S with $\nu \neq \alpha$, then, we have $i(\nu, \alpha) \neq 0$ and $l_{g^t}(\nu) \rightarrow \infty$ as $t \rightarrow \infty$;
- (4) $\lim_{t \rightarrow \infty} g^t = [\alpha]$ as a point on Thurston's boundary \mathcal{PMF}_0 ;

- (5) as $t \rightarrow -\infty$, g^t does not go to infinity but converges to the hyperbolic structure that is obtained by the completely symmetric gluing of the four ideal triangles that we started with, that is, the gluing where the distances between the pairs of distinguished points on the six edges of μ are all equal to zero.

3.10 Example II of a stretch line: the punctured torus

In this section, we consider the surface $S = S_{1,1}$, that is, the punctured torus, equipped with a hyperbolic metric g satisfying some properties which we now state. Let μ be a complete geodesic lamination on S whose leaves consist in a simple closed geodesic γ , together with a bi-infinite geodesic δ converging at both ends to the cusp, and in two other bi-infinite geodesics, each having one end converging to the cusp and the other end spiraling around γ , in such a way that the spiraling of the two geodesics takes place in the same direction, as in Figure 22. The surface $S \setminus \mu$ has two connected components, each of them, with its intrinsic metric, being the interior of a hyperbolic

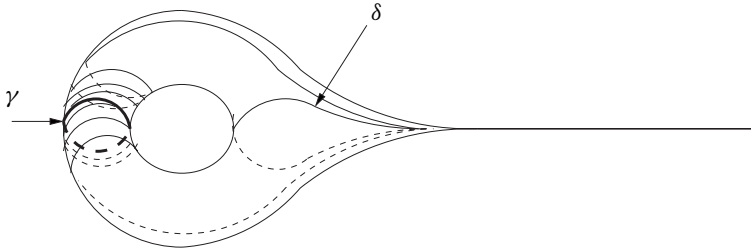


Figure 22. The complete geodesic lamination μ is the union of the closed geodesic γ , the bi-infinite geodesic δ and two other bi-infinite geodesics that spiral around γ at one end and that converge to the cusp of S at their other end.

ideal triangle. We choose the metric g so that its associated horocyclic lamination $\lambda_\mu(g)$ is a closed geodesic α satisfying $i(\alpha, \gamma) = 1$, as represented in Figure 23.

In Figure 23 are represented the stump γ of μ , the closed geodesic α and the simple closed geodesic β obtained from α by a left Dehn twist along γ .

We shall study the behaviour of the lengths of these particular closed geodesics under the stretch and anti-stretch rays directed by μ and starting at $g = g^0$. The behaviour of the lengths of these geodesics will permit us to understand the change in geometry of the surface g under the stretch map.

The closed geodesics α and β satisfy $\text{Card}(\gamma \cap \beta) = \text{Card}(\alpha \cap \gamma) = 1$. The reason why we consider these geodesics is that eventually we want to get an idea of the behaviour of the lengths of an arbitrary simple closed geodesic (and more generally of an arbitrary compactly supported measured geodesic lamination) on S under the stretch (respectively the anti-stretch) ray g^t as $t \rightarrow \infty$, and we expect this behaviour to depend on the intersection pattern of that lamination with the stump of μ and with the horocyclic measured geodesic lamination $\lambda_\mu(g)$. The general results are given

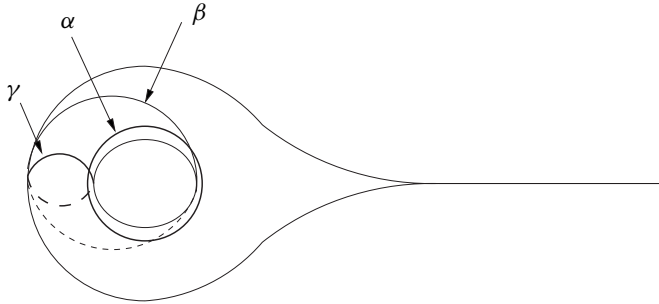


Figure 23. α is the horocyclic measured geodesic lamination associated to g and μ . We are interested in the behaviour of the lengths of the three curves α , β and γ along the stretch and the anti-stretch ray directed by μ and starting at g .

in Theorems 3.34 and 3.35 below, and the results that we present in the example considered here are illustrations of those general results. What makes things work easily in this example is that we can determine the exact positions of these geodesics for the structure g^t as t varies.

The horocyclic foliation $F_\mu(g)$ is the union of a cylinder C with core curve α and of another cylinder foliated by leaves that are parallel to the puncture of S , the two cylinders being glued along the critical graph of $F_\mu(g)$. This critical graph is represented in Figure 24. Figure 25 represents the cylinder C to which are attached the nonfoliated parts of $F_\mu(g)$. In that figure, the pairs of regions with the same name

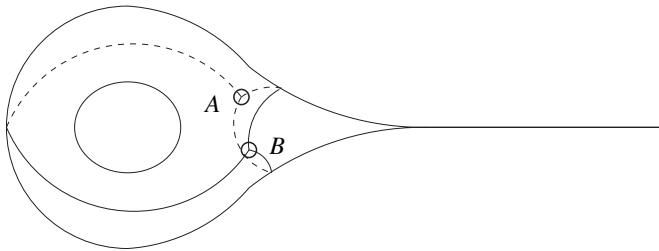


Figure 24. The critical graph of the foliation $F_\mu(g)$ with its two singular points A and B .

are identified in the surface S . In the same figure, we have drawn in bold lines the segments induced by γ and δ on the cylinder C . This cylinder has a symmetry which is probably more apparent if we cut the cylinder along the arc induced on it by δ . After this cutting, we obtain Figure 26, that is, a large quadrilateral R which is the union of two smaller isometric quadrilaterals R_1 and R_2 . (The internal face of the cylinder C of Figure 25 is the quadrilateral R of Figure 26.) The quadrilateral R has an order-two symmetry of center w , where w is the midpoint of the segment induced by γ on the cylinder C . Let s denote the midpoint of the arc induced by δ on that cylinder.

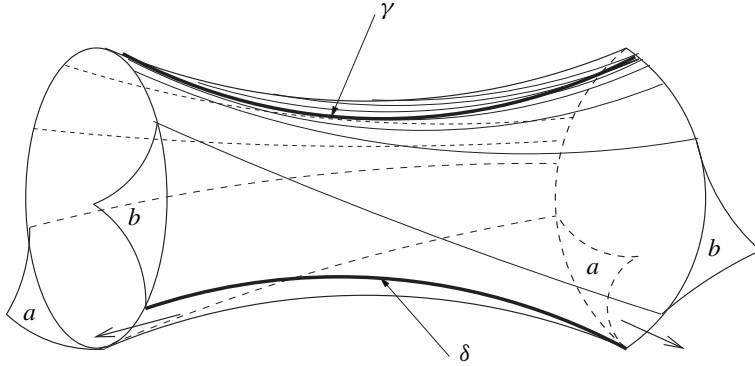


Figure 25. The cylinder C and on its boundary the nonfoliated regions of $F_\mu(g)$. These nonfoliated regions are labeled a and b , these letters corresponding to the points A and B respectively on the critical graph of Figure 24. The two arrows indicate that the corresponding spikes converge to the cusp.

By the uniqueness of the closed geodesic in each homotopy class of closed curves on S , the closed geodesic α is preserved by this order-two symmetry of R . Therefore this geodesic passes by the points s and w . The same holds for the closed geodesic β , which passes by w (see Figure 26).

All these results on the relative positions of α and β remain true when the metric g is replaced by the stretched (respectively anti-stretched) metric g^t .

To do computations, we can use Figure 27 in which the quadrilateral R_1 is represented in the upper half-plane model of hyperbolic space. In that figure, l_t denotes the length of γ for the structure g^t . However, it is possible to determine the asymptotic behaviour of the lengths of the closed curves α , β and γ along the stretch (respectively anti-stretch) ray without computations. It suffices for that to determine the behaviour of the lengths of the sides of R_1 along these rays. We refer to Figure 26, in which the common length of the horizontal sides of R is equal to the length of γ , and the height of R is defined to be the length of the leaf of $F_{g^t}(\mu)$ that is equidistant from the boundaries of C . This last quantity is an upper bound for the length of α . Now we observe the following facts:

- (1) The distances between the non-foliated regions grow linearly with e^t along the stretch ray, and they decrease to zero along the anti-stretch ray.
- (2) In each spike, the lengths of the horocyclic arcs that are not on the boundary of a non-foliated region and that are leaves of the horocyclic foliation decrease to zero along the stretch ray and grow to 1 along the anti-stretch ray.

Using these facts, it is easy to see that along a stretch ray, the height of R tends to zero, whereas the common length of the horizontal sides of R tends to infinity. Along an anti-stretch line, the height of R tends to infinity whereas the length of a horizontal side tends to zero. Now since we know the exact positions of the various curves, we obtain the following, along the stretch ray g^t :

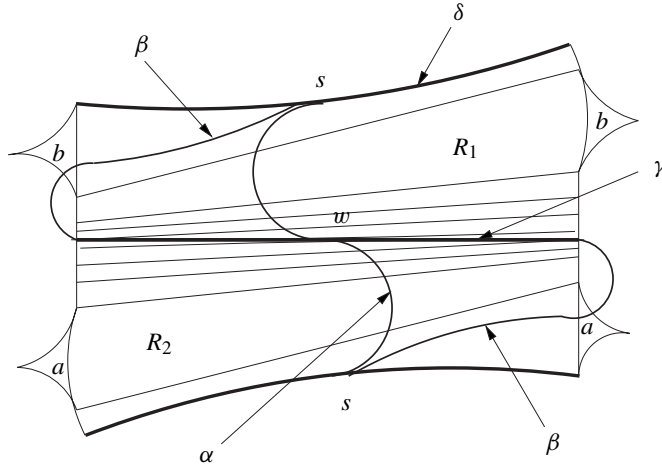


Figure 26. The quadrilateral obtained by cutting the cylinder C along δ .

- (1) $\lim_{t \rightarrow \infty} l_{g^t}(\gamma) = \infty$;
- (2) $\lim_{t \rightarrow \infty} l_{g^t}(\alpha) = 0$;
- (3) $\lim_{t \rightarrow \infty} l_{g^t}(\beta) = \infty$.

Likewise, along the anti-stretch ray g^t , we have

- (1) $\lim_{t \rightarrow \infty} l_{g^t}(\gamma) = 0$;
- (2) $\lim_{t \rightarrow \infty} l_{g^t}(\alpha) = \infty$;
- (3) $\lim_{t \rightarrow \infty} l_{g^t}(\beta) = \infty$.

This example is rather limited in scope because the surface considered is not large enough so as to contain closed curves that are disjoint from the stump, or from the horocyclic lamination, or from both. This will be possible in the example which we consider next.

3.11 Example III of a stretch line: the closed surface of genus 2

In this example, $S = S_{2,0}$ is a closed surface of genus 2 equipped with a complete geodesic lamination μ whose stump consists of two nonseparating closed geodesics γ_1 and γ_2 , represented in Figure 28, with the other leaves of μ spiraling around $\gamma_1 \cup \gamma_2$ in the same direction, as represented in Figure 29. We choose a hyperbolic structure g on S whose associated horocyclic measured geodesic lamination $\lambda_\mu(g)$ consists of a simple closed geodesic α , satisfying $\text{Card}\{\gamma_1 \cap \alpha\} = \text{Card}\{\gamma_2 \cap \alpha\} = 1$. The topological pattern made by the curves γ_1 , γ_2 and α is represented in Figure 28.

As in the preceding examples, we study the behaviour of the lengths of some particular simple closed geodesics under a stretch or an anti-stretch ray directed by μ

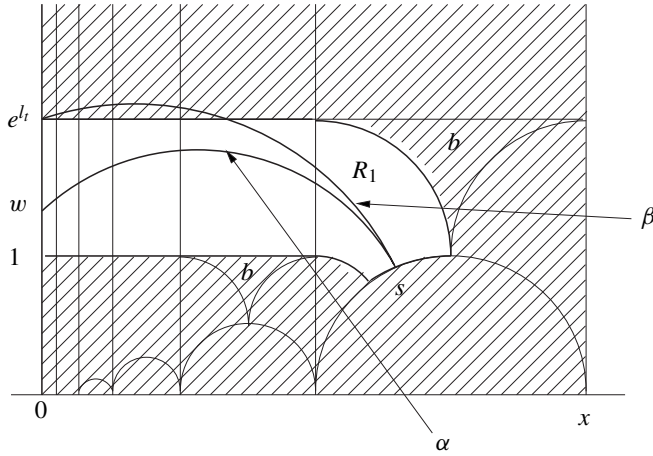


Figure 27. The quadrilateral R_1 in the upper half-plane model of \mathbb{H}^2 is the non-shaded region.

and starting at g . The closed geodesics considered here are γ_1 and γ_2 (chosen because they are contained in the stump of μ), the geodesic α , and the two geodesics β_1 and β_2 that are represented in Figure 30, satisfying $(\gamma_1 \cup \gamma_2) \cap \beta_1 \neq \emptyset$, $(\gamma_1 \cup \gamma_2) \cap \beta_2 = \emptyset$ and $\alpha \cap \beta_1 = \alpha \cap \beta_2 = \emptyset$. Again, the explicit computations that we do in this example are illustrations of Theorems 3.34 and 3.35 below.

The horocyclic foliation $F_\mu(g)$ is a foliation with one cylinder with core curve in the class of α . In Figure 31 we have represented the critical graph of $F_\mu(g)$ (after collapsing the non-foliated regions). Let C denote the cylinder obtained by cutting the surface along the critical graph of $F_\mu(g)$. This cylinder is represented in Figure 33 (together with the non-foliated regions on its boundary, as in the preceding example). Each connected component of the intersection of the leaves of μ with C is a segment

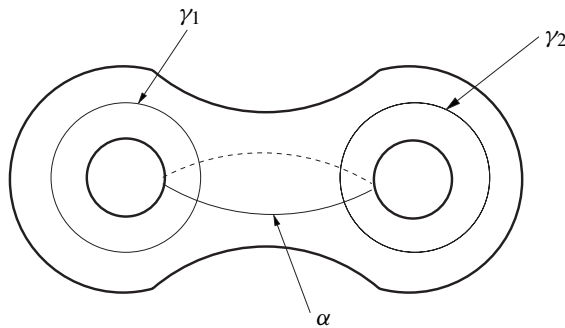


Figure 28. The union $\gamma_1 \cup \gamma_2$ represents the stump of the complete geodesic lamination μ , and α represents the horocyclic measured geodesic lamination $\lambda_\mu(g)$ of the surface $S = S_{2,0}$.

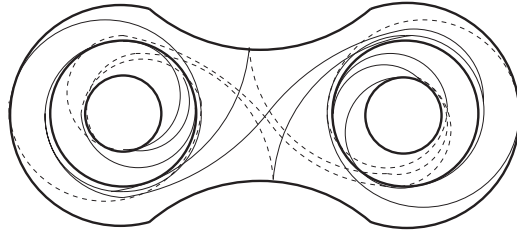


Figure 29. The geodesic lamination μ , with stump $\gamma_1 \cup \gamma_2$. There are six leaves of μ which are not contained in the stump. Each of these leaves is a bi-infinite geodesic that spirals around γ_1 or γ_2 . The spiraling around each curve γ_1 and γ_2 is in the same direction.

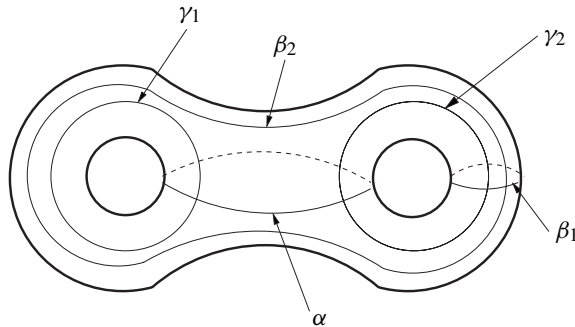


Figure 30. We are interested in the behaviour of the lengths of the five closed geodesics represented here under the stretch line directed by μ (of stump $\gamma_1 \cup \gamma_2$) and passing by the hyperbolic structure whose horocyclic measured geodesic lamination is α .

that joins the two boundary components of that cylinder, as shown in Figure 33. From the definition of the transverse measure of $F_\mu(g)$, it follows that all these connected components have the same length. In particular, we have $l_g(\gamma_1) = l_g(\gamma_2)$.

We consider the cylinder C as the union of four quadrilaterals R_1, \dots, R_4 . This decomposition into quadrilaterals is induced by γ_1 and γ_2 and by two other geodesic segments that join distinguished points on the leaves of μ , whose trace on the cylinder C cut up along γ_1 is shown in Figure 34. It is easy to see that the quadrilaterals R_1, \dots, R_4 are congruent. For each of these quadrilaterals, two of the opposite sides are geodesics contained in leaves of μ , and the remaining two sides are made out of horocycles contained in leaves of $F_\mu(g)$. (Note that each such side is a segment of a horocycle and not just a union of segments of horocycles. This can be seen by examining Figure 35 which represents one of the quadrilaterals drawn in the upper half-plane.) The cylinder C has an order-two symmetry with respect to the midpoint of each geodesic side of any of the quadrilaterals R_i . To see this, let us consider the quadrilateral R obtained by cutting the cylinder C along the segment γ_1 . This quadrilateral is represented in Figure 34, with its induced decomposition into the four

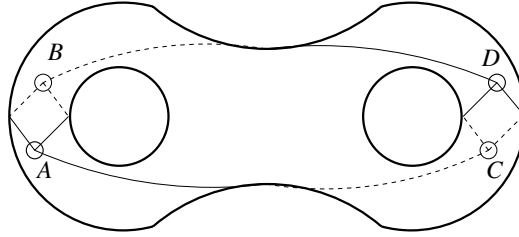


Figure 31. The critical graph of the foliation obtained from $F_\mu(g)$ by collapsing the non-foliated regions on the points A, B, C, D .

quadrilaterals R_1, \dots, R_4 . The four points s, w, u, v are the intersection points of the leaf of $F_\mu(g)$ that is equidistant from the boundary of C (distances being measured with respect to the transverse measure of that foliation) with the geodesic sides of the quadrilaterals R_1, \dots, R_4 . The point u is the intersection point of that leaf with γ_2 . We claim that u is a center of symmetry for the segments induced on R by each of the curves α, β_1 and β_2 . To see this, first of all, note that these segments are situated as shown in Figure 34 because of the combinatorics of the gluing of the boundary components of R . Now, if we call $\alpha', \beta'_1, \beta'_2$ the images of the curves α, β_1, β_2 by the symmetry of center u , then $\alpha', \beta'_1, \beta'_2$ are also simple and closed since the interior of R is symmetric with respect to u . Since these curves are closed geodesics that are homotopic to the closed geodesics α, β_1 and β_2 respectively, they coincide with them. In particular, α passes through the four points s, w, u, v .

For later use, we compute the value x , shown in Figure 35, which represents the quadrilateral R_1 in the upper half-plane. This value x is a Euclidean distance, measured on the boundary of the upper half-plane. It is also equal to the hyperbolic length of the horocyclic edge of the rectangle R_1 lying at height 1, in that same figure. The Euclidean distance y represented in the same figure is equal to 3, because it is equal to the length of three horocyclic segments that are boundaries of non-foliated triangles in the standard horocyclic foliation of an ideal triangle, and that are at height 1 in the upper half-plane. Therefore, we have the following (see also the caption of Figure 35):

$$x = y(1 + e^{-l} + e^{-2l} + \dots) = \frac{3}{1 - e^{-l}} = \frac{3e^{l/2}}{2 \sinh(l/2)},$$

where $l = l_g(\gamma_1) = l_g(\gamma_2)$.

Now we compute the lengths of the segments labeled p, q and r in Figure 37. These segments are induced by α, β_1 and β_2 respectively on the quadrilateral represented. We use the formula of Figure 36 giving a relation between the length and the slope.

For the segment p , the slope is $\frac{2e^{l/2}}{x}$, therefore the length is

$$2 \sinh^{-1} \left(\frac{x}{2e^{l/2}} \right) = 2 \sinh^{-1} \left(\frac{3}{4(l/2)} \right).$$

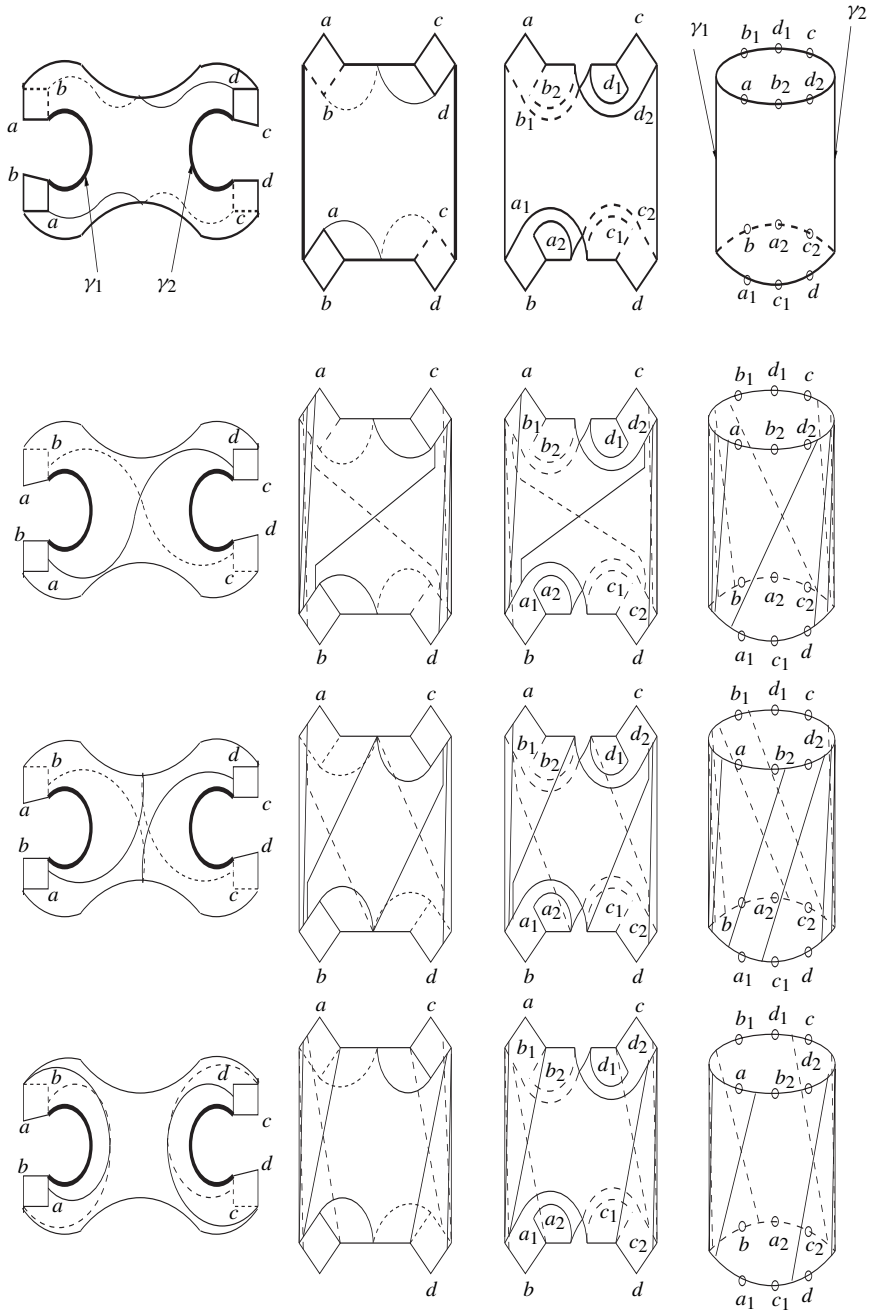


Figure 32. Line 1 represents the passage from the singular graph to the cylinder C obtained by cutting the surface along the singular graph (see Figure 33). On lines 2, 3 and 4 we have repeated the cutting, showing at each line the two closed curves γ_1 and γ_2 in bold lines, together with two other leaves of μ that spiral along these two closed curves.

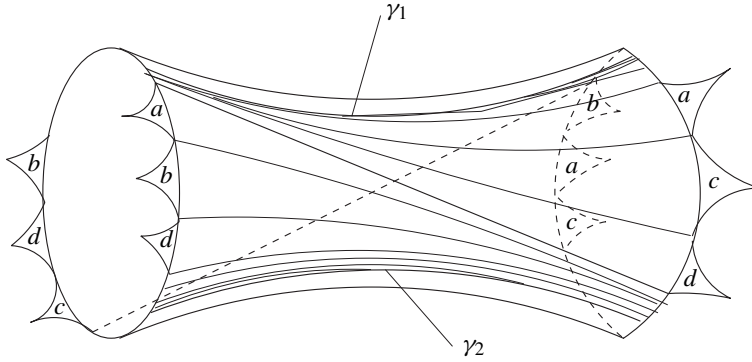


Figure 33. The cylinder C , together with the non-foliated regions of ideal triangles on its boundary. Non-foliated regions with the same names are identified in the surface S .

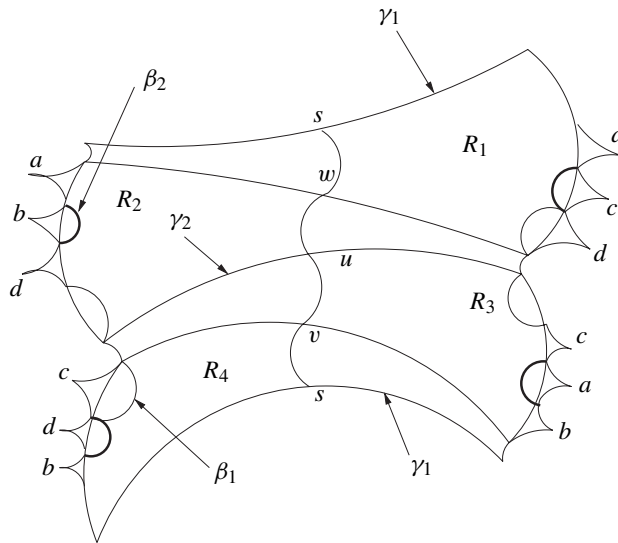


Figure 34. The quadrilateral R representing the cylinder C cut along the geodesic γ_1 . This quadrilateral is the union of the four quadrilaterals R_1, \dots, R_4 that are drawn. The two horizontal sides of these quadrilaterals that are not labeled are geodesic segments in μ that join distinguished points on the edges of that lamination. We have also represented the leaf of $F_\mu(g)$ that is equidistant from the two boundary components of C , together with its intersection points s, w, u, v with the geodesic edges of the quadrilaterals R_1, \dots, R_4 . Finally, we have represented the geodesic segments induced by β_1 and β_2 on the quadrilateral R . Each of these segments has two boundary points on the same vertical side of R . The trace of β_2 is drawn in bold lines.

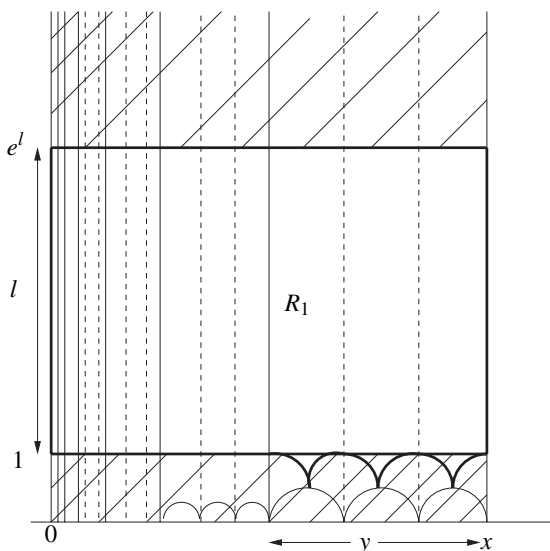


Figure 35. The quadrilateral R_1 , represented in the upper half-plane model of the hyperbolic plane, is the region between the two shaded regions. It is bounded by two vertical geodesic segments, both of length $l = l_g(\gamma_1) = l_g(\gamma_2)$, and two horizontal horocyclic segments, the lower one having length x and the upper one having length $x/h = xe^{-l}$. The geodesic segment corresponding to γ_1 is the segment contained in the vertical ray starting at the origin, joining the point at height 1 to the point at height e^l . Both the continuous and the dotted vertical lines in the picture are induced by the leaves of μ that cross R_1 and that spiral along γ_1 . This spiraling is obtained by iterating, under the map $z \mapsto e^{-l}z$, the right hand block delimited by the geodesics that are based at the abscissas x and $x - y$. This block is made out of the three ideal triangles suggested by the dotted lines. The (Euclidean) distance y is equal to 3.

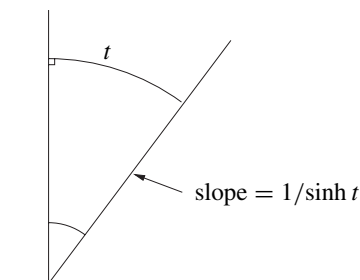


Figure 36. This is a useful ingredient for computing $l_{g^t}(\alpha)$, $l_{g^t}(\beta_1)$ and $l_{g^t}(\beta_2)$. In this figure, the vertical line is a geodesic in the upper half-space model of \mathbb{H}^2 , and the oblique line is a hypercycle (a line consisting of points at hyperbolic distance t from the geodesic line). The value t is equal to the (hyperbolic) length of any arc of circle joining these two lines perpendicularly. The slope of the oblique line is then equal to $1/\sinh t$.

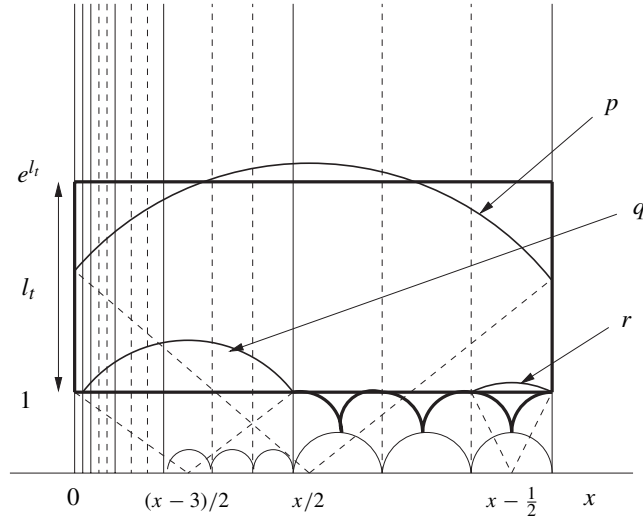


Figure 37. One of the four quadrilaterals of Figure 34, drawn in the upper half-space. The segments labeled p, q, r are the segments induced on that quadrilateral by the geodesics α, β_1 and β_2 respectively. The segment p is only contained in α , the segment q is only contained in β_2 , whereas the segments r is contained in β_1 and β_2 . The computation of the lengths of these segments uses the coordinates that are indicated here and the ingredient in Figure 36. The labels $(x - 3)/2, x/2$ and $x - \frac{1}{2}$ are the abscissas of the corners in dotted lines that are above these labels.

For the segment q , the slope is $\frac{2}{x-3}$, therefore the length is

$$2 \sinh^{-1} \left(\frac{x-3}{2} \right) = 2 \sinh^{-1} \left(\frac{3e^{-l/2}}{4(l/2)} \right).$$

For the segment r , the slope is 2, therefore

$$2 \sinh^{-1} \left(\frac{1}{2} \right) = 2 \log \left(\frac{1 + \sqrt{5}}{2} \right).$$

Now we stretch the structure g along μ . Recall that g^t is the hyperbolic structure defined by the equality $F_\mu(g^t) = e^t F_\mu(g)$. The distances between the ideal triangles of $S \setminus \mu$ are uniformly stretched. All the remarks that we made about the quadrilaterals R_1, \dots, R_4 with respect to the metric g are valid for g^t . The lengths of the boundary sides of these quadrilaterals depend on t , and again, using the invariance of the transverse measure of $F_\mu(g^t)$, we have

$$l_{g^t}(\gamma_1) = l_{g^t}(\gamma_2).$$

We set

$$l_t = l_{g^t}(\gamma_1) = l_{g^t}(\gamma_2).$$

Note that $l_t = e^t l_g(\gamma_1) = e^t l_g(\gamma_2)$, since γ_1 and γ_2 are in the stump of μ .

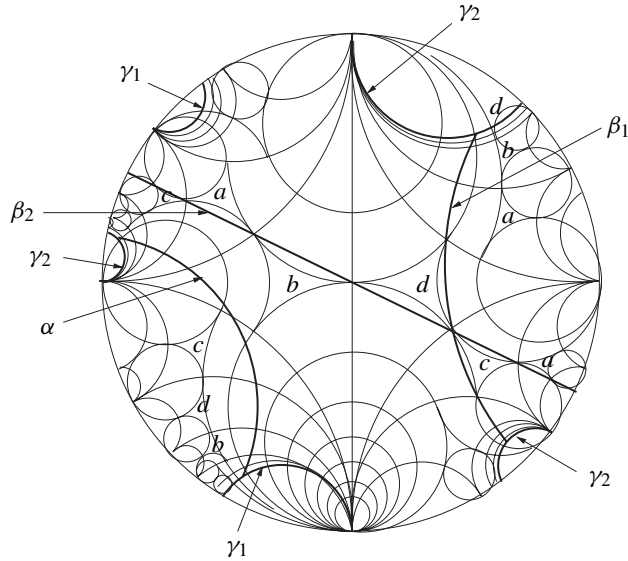


Figure 38. The universal covering of the structure g . In this picture, we can see the annuli of the horocyclic foliation by examining the non-foliated triangular regions of the ideal triangulation. Lifts of α , β_1 and β_2 are also represented.

The curve α consists of four segments of the type labeled by p in Figure 37. The curve β_1 is composed of two segments of the type r and two segments of the type q , and the curve β_2 is composed of four segments of the type r . From this we deduce the following:

$$l_{g^t}(\alpha) = 8 \sinh^{-1} \left(\frac{3}{4 \sinh(l_t/2)} \right),$$

$$l_{g^t}(\beta_1) = 4 \sinh^{-1} \left(\frac{3e^{-l_t/2}}{4 \sinh(l_t/2)} \right) + 4 \log \left(1 + \frac{\sqrt{5}}{2} \right)$$

and

$$l_{g^t}(\beta_2) = 8 \log \left(\frac{1 + \sqrt{5}}{2} \right).$$

Summing up, if $t \mapsto g^t$ is the stretch ray directed by μ and starting at g , we have

- (1) $\lim_{t \rightarrow \infty} l_{g^t}(\alpha) = 0$. More precisely, we have $l_{g^t} \sim_{t \rightarrow \infty} 12e^{-l_t/2} = 12e^{-e^t/2}$;
- (2) $\{l_{g^t}(\beta_1) \mid t \geq 0\}$ and $\{l_{g^t}(\beta_2) \mid t \geq 0\}$ are bounded subsets of \mathbb{R}_+^* . In fact, $l_{g^t}(\beta_2)$ is constant and $l_{g^t}(\beta_1)$ decreases towards $4 \log \left(\frac{1 + \sqrt{5}}{2} \right)$.

Now if $t \mapsto g^t$ is the anti-stretch ray directed by μ and starting at g , we have

- (1) $\lim_{t \rightarrow \infty} l_{g^t}(\alpha) = \infty$;

- (2) $\lim_{t \rightarrow \infty} l_{g^t}(\gamma_i) = 0$ for $i = 1, 2$;
- (3) $\{l_{g^t}(\beta_2) \mid t \leq 0\}$ is a bounded subset of \mathbb{R}_+^* (and in fact it is constant);
- (4) $\lim_{t \rightarrow \infty} l_{g^t}(\beta_1) = \infty$.

From these computations, one gets the feeling that the following facts hold:

– Changing the orientation of a stretch line seems to interchange the roles of the stump and of the horocyclic measured geodesic lamination.

– The length of any simple closed geodesic with empty intersection with the stump and with the horocyclic measured geodesic lamination is almost constant along a stretch line.

– The length of a simple closed geodesic intersecting the horocyclic lamination (respectively the stump) tends to infinity under the stretch ray (respectively anti-stretch ray).

The results stated in Section 3.12 below (Theorems 3.34 and 3.35) show that this is indeed the case, and not only for lengths of simple closed geodesics, but more generally for lengths of arbitrary measured geodesic laminations.

Figure 39 gives another view on the dynamical behaviour of the stretch line considered in the example that we are studying in this section. In that figure, we have

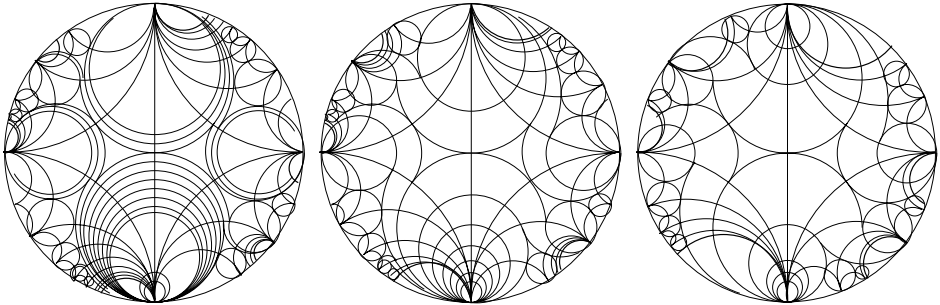


Figure 39. Each of these three pictures represents the universal covering \mathbb{H}^2 of the surface $S_{2,0}$ studied in Section 3.11 successively equipped with the structures g^{-t} , $g = g^0$ and g^t for some $t > 0$. We have represented a part of the preimage of $\gamma_1 \cup \gamma_2$ by geodesics in bold lines, as well as the preimage of μ that decomposes in a neat way \mathbb{H}^2 into ideal triangles. In each ideal triangle, we have represented the small non-foliated triangular region, which permits the visualization of the horocyclic foliation. In this way, we can see that the width of the cylinder C increases whereas its circumference decreases.

drawn the Poincaré disk realized as the universal covering of the surface S equipped respectively with hyperbolic structures f , g and h that belong to a stretch line directed by μ , and appearing in that order. The central disk represents the structure g that we started with, the left-hand disk corresponds to a structure f that lies before g on the stretch line (that is, we can obtain g by stretching f along the same complete lamination μ) and the right-hand figure represents the structure h obtained by stretching g along μ .

From this example, we deduce the following result which says that Thurston’s asymmetric metric and its dual are not Lipschitz-equivalent.

Proposition 3.33. *There does not exist any constant C satisfying $K(g, h) \leq CK(h, g)$ for all g and h in \mathcal{T} .*

Proof. We consider the stretch line studied above, directed by μ and starting at g , and the closed geodesic α . Then we have, for every $t \geq 0$,

$$r_{g^t, g}(\alpha) \sim_{t \rightarrow \infty} \frac{l_g(\alpha)}{12e^{-\frac{e^t}{2}}} = A(g)e^{\frac{e^t}{2}}$$

where $A(g)$ is a constant that depends on the metric g . Thus we obtain $K(g^t, g) \geq \log r_{g^t, g}(\alpha) \sim e^t$, whereas $K(g, g^t) = t$. This proves the proposition. \square

We can use the explicit formulae that we produced in this example to plot the graphs of the functions $t \mapsto r_{g, g^t}(\lambda)$ for $\lambda \in \{\alpha, \beta_1, \beta_2, \gamma_1, \gamma_2\}$. These graphs give us an idea of the behaviour of the hyperbolic surfaces g^t as t varies. It is also interesting to draw the graph of the function $t \mapsto r_{g, g^t}(\delta)$ where δ is the closed geodesic represented in Figure 40, because the intersection pattern of δ with α satisfies $\delta \cap \alpha \neq \emptyset$ and $\delta \cap \gamma = \emptyset$. Figures 41 and 42 represent respectively the intersection of δ with the

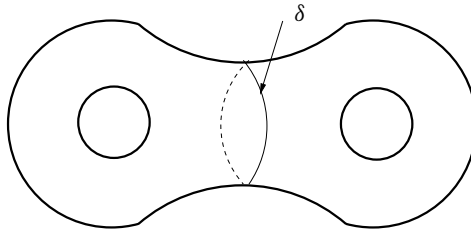


Figure 40

cylinder R of Figures 34 and 37.

The computations give

$$l_{g^t}(\delta) = \log \left(\frac{e^{t/2} + \frac{1}{16}(e^{t/2} - 5 - \sqrt{(e^{t/2} - 5)^2 + 16e^{t/2}})^2}{e^{t/4}(1 + \frac{1}{16}(e^{t/2} + 3 - \sqrt{(e^{t/2} - 5)^2 + 16e^{t/2}})^2)} \right)$$

Therefore,

$$\lim_{t \rightarrow \infty} l_{g^t}(\delta) = \infty$$

and

$$\lim_{t \rightarrow -\infty} l_{g^t}(\delta) = 4 \log(1 + \sqrt{2}).$$

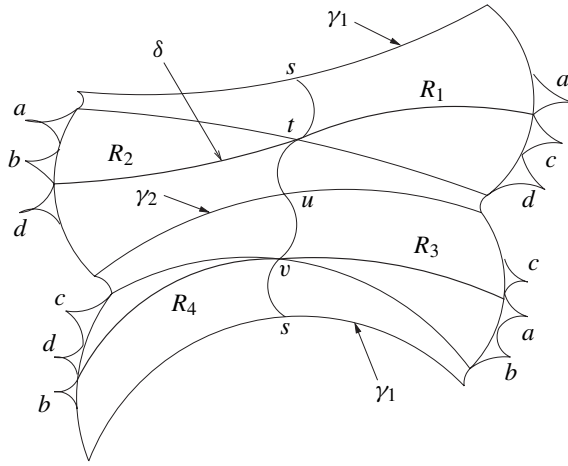


Figure 41

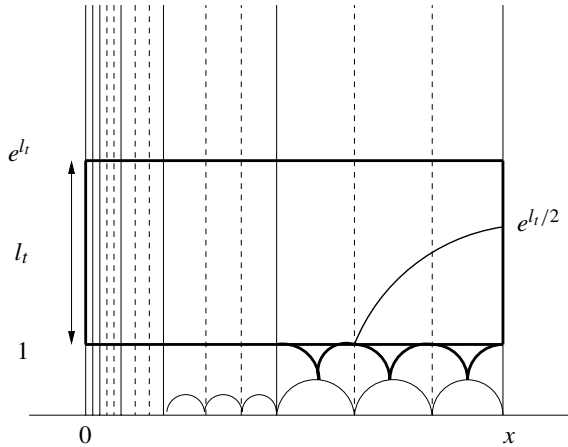


Figure 42

We have drawn below the graphs of the functions $t \mapsto r_{g, g^t}(\lambda)$ for $\lambda \in \{\alpha, \beta_1, \delta\}$.

It is interesting to note that the length function $t \mapsto l_{g^t}(\alpha)$ of the horocyclic measured geodesic lamination is strictly decreasing and convex. We can see on the other figures that in the general case, the length function along a stretch line is neither monotonic nor convex. But the figures indicate that the function might be peakless in the sense of Busemann. We recall that a function f defined on a closed interval $I \subset \mathbb{R}$ is said to be *peakless* if I can be decomposed into three subintervals I_l , I_0 and I_r whose interiors are pairwise disjoint and following each other in the order indicated, such that f is strictly decreasing on I_l , constant on I_0 and strictly increasing on I_r .

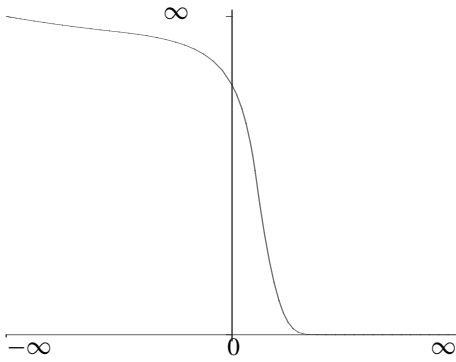


Figure 43. $t \mapsto l_{g^t}(\alpha)$.

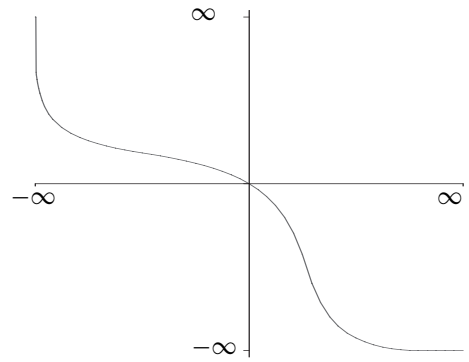


Figure 44. $t \mapsto \log(r_{g, g^t}(\alpha))$.

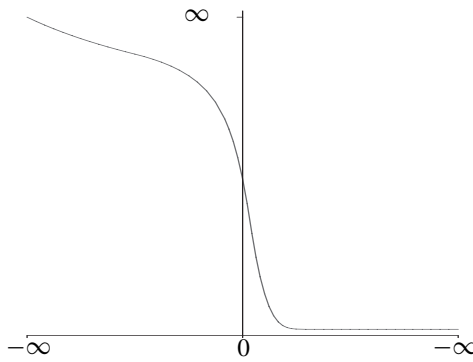


Figure 45. $t \mapsto l_{g^t}(\beta_1)$.

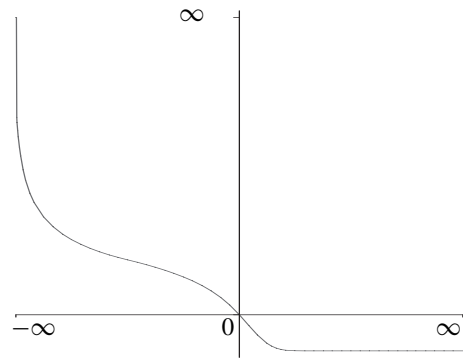


Figure 46. $t \mapsto \log(r_{g, g^t}(\beta_1))$.

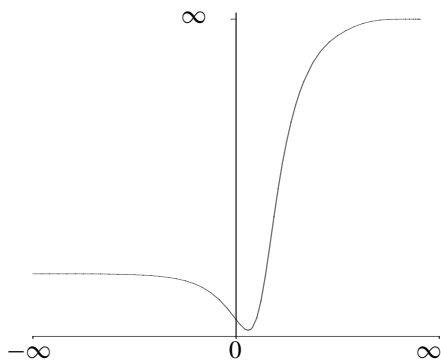


Figure 47. $t \mapsto l_{g^t}(\delta)$.

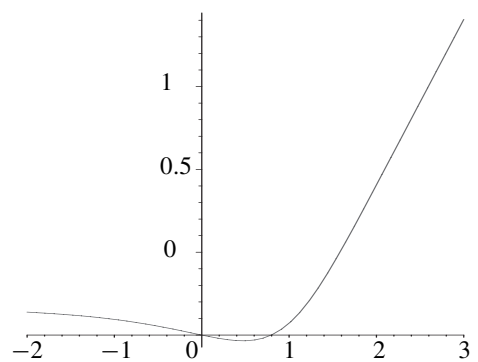


Figure 48. $t \mapsto \log(r_{g, g^t}(\delta))$.

(Some of the intervals I_l , I_0 and I_r might be empty). In the case where I_0 is empty, the function f is said to be *strictly peakless*. This notion was introduced by Busemann as a generalization of convexity (see [13], p. 109). Regarding the closed geodesics γ_1 and γ_2 , which are in the stump, we recall that their g^t -lengths (which are equal) are strictly convex and increasing.

In the sections that follow, we shall study the asymptotic behaviour of general stretch and anti-stretch lines in Teichmüller space.

3.12 The behaviour of the lengths of measured geodesic laminations along stretch and anti-stretch lines

In this section, we present some results that show that the estimates made in the examples studied in Sections 3.9, 3.10 and 3.11 about the lengths of measured geodesic laminations along stretch and anti-stretch lines, are valid in a general setting. We shall give estimates on the behaviour of lengths of geodesic laminations under stretch and anti-stretch rays.

The proofs are contained in [64] and [63].

In what follows, we shall use the term *topological measured geodesic lamination* to denote a geodesic lamination that admits a transverse measure in the usual sense, but which is not equipped with any particular such transverse measure. In other words, a topological measured geodesic lamination is the support of some measured geodesic lamination.

We first deal with the case of a stretch ray $\{g^t \mid t \geq 0\}$, and then we shall consider the case of an anti-stretch ray $\{g^t \mid t \geq 0\}$. Our aim is to determine, in each case, the limit, if it exists, of the length $l_{g^t}(\alpha)$ of a measured geodesic lamination α , as $t \rightarrow \infty$.

Theorem 3.34 (Théret [64]). *Let g be a hyperbolic structure on S , let μ be a complete geodesic lamination, let $\{g^t \mid t \geq 0\}$ be a stretch ray directed by μ and starting at $g = g^0$ and let $\lambda = \lambda_\mu(g)$ be the associated horocyclic measured geodesic lamination. Let α be an arbitrary compactly supported measured geodesic lamination on S . We have the following:*

- (1) *If the support of α is contained in the support of λ , then $\lim_{t \rightarrow \infty} l_{g^t}(\alpha) = 0$.*
- (2) *If $i(\alpha, \lambda) \neq 0$, then $\lim_{t \rightarrow \infty} l_{g^t}(\alpha) = \infty$.*
- (3) *If the support of α is disjoint from the support of λ , then $\{l_{g^t}(\alpha) \mid t \geq 0\}$ is a bounded subset of \mathbb{R}_+^* .*

The next result concerns the limits of the same quantities, but this time along anti-stretch rays.

Theorem 3.35 (Théret [63]). *Let μ a complete geodesic lamination on S and let $\{g^t \mid t \geq 0\}$ be an anti-stretch ray starting at a hyperbolic structure $g = g^0$ and directed by μ . Assume the stump of μ is nonempty, and call it γ . Let α be an arbitrary compactly supported measured geodesic lamination on S . We have the following:*

- (1) *If the support of α is contained in the support of γ , then $\lim_{t \rightarrow \infty} l_{g^t}(\alpha) = 0$.*
- (2) *If $i(\alpha, \gamma) \neq 0$, then $\lim_{t \rightarrow \infty} l_{g^t}(\alpha) = \infty$. Note that in order to talk about $i(\alpha, \gamma)$, one has to put a transverse measure on γ . The condition $i(\alpha, \gamma) \neq 0$ is then independent of the choice of the transverse measure on γ .*
- (3) *If the support of α is disjoint from the support of γ , then $\{l_{g^t}(\alpha) \mid t \geq 0\}$ is a bounded subset of \mathbb{R}_+^* .*

These two theorems suggest that the two measured geodesic laminations γ and λ play in some cases symmetric roles.

The two theorems show in particular that the length of a measured geodesic lamination that is disjoint from γ and from λ remains bounded along a stretch line.

A consequence of this fact is that subsurfaces of S that have empty intersection with the stump and the horocyclic measured geodesic lamination are distorted by a uniformly bounded amount as one follows a stretch line. Let us be more precise. Consider a stretch line directed by a complete geodesic lamination μ whose stump γ is non-empty. Let λ be the support of any horocyclic measured geodesic lamination associated to this stretch line. Now assume that there exists a subsurface S' of S with nonempty boundary satisfying the following conditions:

- the Euler characteristic of S' is negative;
- S' contains at least one simple closed curve which is essential in S ;
- $S' \cap \lambda = S' \cap \gamma = \emptyset$.

Note that these requirements imply that any hyperbolic structure on S induces a hyperbolic structure with geodesic boundary on S' and that S' , equipped with such a structure, is isometrically embedded in S .

Let us say that two hyperbolic structures g and g' on S' are K -quasi-isometric if there exists a homeomorphism f of S' sending g to g' , which is isotopic to the identity of S' , and a number $K > 0$ such that $L(f) \leq K$ and $L(f^{-1}) \leq K$.

If g and g' on S' are K -quasi-isometric, then the ratios $r_{g,g'}(\alpha)$ are bounded from above by K and from below by $1/K$. Conversely, one can show that if the ratios $r_{g,g'}(\alpha)$, as α varies over the set of essential simple closed curves of S' , are bounded, then there exists some K for which the two structures g and g' are K -quasi-isometric. Therefore, we have the following:

Corollary 3.36 (Th  ret [63]). *Let $t \mapsto g^t$ be a stretch line directed by μ , with stump γ . Assume that there exists a subsurface S' of S as above. Then there exists a positive number K such that for every t and t' in \mathbb{R} , the hyperbolic structures induced on S' by g^t and $g^{t'}$ are K -quasi-isometric.*

3.13 Thurston's boundary

Teichm  ller space $\mathcal{T}(S)$ is embedded in the function space $\mathbb{R}_+^{\mathcal{S}}$ by the length functional $l: \mathcal{T} \rightarrow \mathbb{R}_+^{\mathcal{S}}$ defined in (3.2) above. This embedding, composed with the natural map

from \mathbb{R}_+^δ onto its projectivization $P\mathbb{R}_+^\delta$, gives an embedding (see [20])

$$\mathcal{T}(S) \rightarrow P\mathbb{R}_+^\delta. \quad (3.10)$$

We also recall that the intersection number functional associates to each $\lambda \in \mathcal{ML}(S)$ the element $i(\lambda, \cdot) \in \mathbb{R}_+^\delta$, and that this defines a map from $\mathcal{ML}(S)$ into \mathbb{R}_+^δ . At the level of projectivizations, we obtain a map

$$\mathcal{PL}(S) \rightarrow P\mathbb{R}_+^\delta$$

which restricts to an embedding

$$\mathcal{PL}_0(S) \rightarrow P\mathbb{R}_+^\delta, \quad (3.11)$$

which induces a topology on the space $\mathcal{PL}_0(S)$, by restriction of the projectivization of the weak topology on $P\mathbb{R}_+^\delta$.

We have the following result of Thurston (cf. [20], where the result is described using measured foliations instead of measured laminations).

Theorem 3.37 (Thurston). *The images of the spaces $\mathcal{T}(S)$ and $\mathcal{PL}_0(S)$ in $P\mathbb{R}_+^\delta$ by the embeddings (3.10) and (3.11) are disjoint. With the space $P\mathbb{R}_+^\delta$ being equipped with the quotient of the weak topology, the closure of the image of $\mathcal{T}(S)$ in $P\mathbb{R}_+^\delta$ is compact, and the complement of this image in the closure coincides with the image of $\mathcal{PL}_0(S)$. Equipped with the induced topology, the union $\overline{\mathcal{T}(S)} = \mathcal{T}(S) \cup \mathcal{PL}_0(S)$ is homeomorphic to a closed ball of dimension $6g - 6 + 2n$, whose boundary is the image of $\mathcal{PL}_0(S)$.*

For this reason, the space $\mathcal{PL}_0(S)$ is also called Thurston's boundary of Teichmüller space.

The following convergence criterion is useful:

Let λ be an element of $\mathcal{PL}_0(S)$. A sequence (g_n) in $\mathcal{T}(S)$ converges to the projective class $[\lambda] \in \mathcal{PL}_0(S)$ if and only if there exists a sequence $x_n \in \mathbb{R}_+^$ such that for all $\alpha \in \mathcal{S}$, we have*

$$\lim_{n \rightarrow \infty} x_n l_{g_n}(\alpha) = i(\lambda, \alpha).$$

3.14 Converging to Thurston's boundary

Questions about the convergence of stretch and anti-stretch rays were already considered in [52] where the following is proved:

Theorem 3.38 (Papadopoulos [52], Theorem 5.1 and the remark following it). *Let g be an element of Teichmüller space $\mathcal{T}(S)$ and let μ be a complete geodesic lamination in S . Then the stretch ray directed by μ and starting at g converges to the projective*

class of the associated horocyclic measured geodesic lamination $\lambda_\mu(g)$, considered as an element of Thurston's boundary $\mathcal{P}\mathcal{L}_0(S)$ of $\mathcal{T}(S)$.

We shall say that a measured geodesic lamination μ is *uniquely ergodic* if as a topological lamination, μ admits a unique transverse measure up to a scalar multiple. In particular, if μ consists of a weighted simple closed geodesic, then it is uniquely ergodic. A uniquely ergodic measured geodesic lamination is *minimal* in the sense that every leaf is dense in its support.

Note that our definitions of unique ergodicity for measured geodesic laminations and for measured foliations (Definition 2.8 above) do not coincide under the natural correspondence between the spaces $\mathcal{M}\mathcal{L}_0$ and $\mathcal{M}\mathcal{F}_0$. (This is so because of the case of foliations which contain cylindrical components.)

As in the case of measured foliations, one can talk about uniquely ergodic projective classes of measured laminations.

In the paper [52], it is shown (Proposition 5.2) that if the complete geodesic lamination μ is measured and uniquely ergodic, then the anti-stretch ray directed by μ converges to the projective class of μ . A more general result has been obtained by Th  ret in [63], where the following is proved:

Theorem 3.39 (Th  ret [63]). *Let μ be a complete geodesic lamination on S whose stump γ is nonempty and is uniquely ergodic. Then any anti-stretch ray directed by μ converges to the projective class of γ .*

Note that there are instances where the anti-stretch line converges to a point in Teichm  ller space (and not on its boundary). This occurs for any anti-stretch line directed by a complete geodesic lamination μ which has finitely many leaves, and where all of these leaves converge to punctures at both ends, and it occurs only for such complete geodesic laminations. (In other words, μ is an ideal triangulation.) Note that μ is an ideal triangulation if and only if the stump of μ is empty. The limiting hyperbolic structure g is the one for which all the distances between the distinguished points on that lamination are zero, that is, $\lambda_\mu(g) = \emptyset$.

To state the next result, we say that a measured geodesic lamination μ' is *totally transverse* to the complete geodesic lamination μ if the measure-equivalence class of a measured foliation representing μ' is totally transverse to μ in the sense of Definition 3.9.

One application of Theorems 3.38 and 3.39 is the following

Corollary 3.40. *Let α and β be two points in the boundary $\mathcal{P}\mathcal{L}_0(S)$ of $\mathcal{T}(S)$ that can be represented by totally transverse measured geodesic laminations, and suppose that α is uniquely ergodic. Then there exists a line in Teichm  ller space which is geodesic for Thurston's asymmetric metric and which converges to α in the negative direction and to β in the positive direction.*

In general, this geodesic is not unique: it suffices to consider a uniquely ergodic stump which possesses several completions (see the discussion in Section 3.3). The

stretch lines that are directed by these various complete geodesic laminations, passing through hyperbolic structures which have the same associated horocyclic measured geodesic laminations, converge in the negative direction and in the positive direction to the same points, α and β respectively, in the boundary $\mathcal{P}\mathcal{L}_0(S)$ of $\mathcal{T}(S)$. (To see that it is possible to find such hyperbolic structures, one can use Thurston’s parametrization ϕ_μ of Teichmüller space described in Section 3.5 above.)

It is possible to permute the points α and β , obtaining geodesics that converge to α in the positive direction and to β in the negative direction (at least if β is uniquely ergodic). Note that *a priori* these geodesics will be distinct from the preceding ones.

3.15 Stretch lines and earthquakes

There is no attempt here to present any of the basic results on earthquakes. We consider earthquakes only in their relation to stretch maps. We start by recalling the definition.

Definition 3.41 (Fenchel–Nielsen flow). This is a flow on Teichmüller space which is defined as follows. Let S be a hyperbolic surface and let α be a homotopy class of an essential simple closed curve on S . The *normalized left Fenchel–Nielsen flowline* through S is the one-parameter family $(\mathcal{E}_\alpha^t(S))_{t \in \mathbb{R}}$ of hyperbolic surfaces such that for each $t \geq 0$ (respectively $t \leq 0$), $\mathcal{E}_\alpha^t(S)$ is the hyperbolic surface obtained by cutting the hyperbolic surface S along the closed geodesic in the class α and gluing back the two boundary geodesics of the resulting surface after a left (respectively right) twist of magnitude $|t|l_g(\alpha)$. We say that the hyperbolic surface $\mathcal{E}_\alpha^t(S)$ is obtained from S by a time- t normalized Fenchel–Nielsen twist along α .

We note that this “normalized” parametrization of the Fenchel–Nielsen flow is different from the parametrization used by Kerckhoff in [30]; it is the one used in Papadopoulos [50], because it admits a non-trivial extension to Thurston’s boundary.

Definition 3.42 (Earthquake flow, see [30] and [66]). Let S be a hyperbolic surface, let γ be a compactly supported measured geodesic lamination on S and let α_n be a sequence of elements in $\mathbb{R}_+^* \times \mathcal{S}$ converging to γ in the topology of $\mathcal{M}\mathcal{L}_0(S)$. Then, for each $t \in \mathbb{R}$, the sequence of hyperbolic structures $\mathcal{E}_{\alpha_n}^t(S)$ converges to a hyperbolic structure $\mathcal{E}_\gamma^t(S)$ that does not depend on the choice of the sequence α_n converging to γ . We say that the hyperbolic structure $\mathcal{E}_\gamma^t(S)$ is obtained from S by a time- t normalized earthquake along γ . The earthquake is said to be a *left* (respectively *right*) earthquake if $t \geq 0$ (respectively $t \leq 0$).

Theorem 3.43 (Earthquake and stretch commute, Théret [63]). *Let μ be a complete geodesic lamination on S and let γ be a sublamination of μ equipped with a transverse measure. Let \mathcal{E}_γ^t denote as above the normalized earthquake flow along γ and let \mathcal{S}_μ^t denote the stretch flow directed by μ . Then, for every t and s in \mathbb{R} , we have $\mathcal{E}_\gamma^t \circ \mathcal{S}_\mu^s = \mathcal{S}_\mu^s \circ \mathcal{E}_\gamma^t$.*

In the paper [52], there is a definition and a study of the extension of the normalized earthquake flow to Thurston's boundary $\mathcal{PL}_0(S)$ which is a quotient flow of a flow defined on the unprojectivized space $\mathcal{ML}_0(S)$. In fact, in the paper [52], in the case where S is closed, these flows are defined on the space $\mathcal{MF}_0(S)$ of equivalence classes of measured foliation, and on the space $\mathcal{PMF}_0(S)$ of projective classes of measured foliations on S . We can define the flows on $\mathcal{ML}_0(S)$ and $\mathcal{PL}_0(S)$ by using the natural correspondence between foliations and laminations. As an application of Theorem 3.43, we have the following:

Corollary 3.44 (Théret [63]). *Assume that the surface S is closed. Let γ be a measured geodesic lamination on S and let μ be a completion of γ with stump γ . Then the time- t normalized earthquake (defined on $\mathcal{ML}_0(S)$) along γ of the horocyclic geodesic lamination $\lambda_\mu(S)$ is the horocyclic geodesic lamination $\lambda_\mu(\mathcal{E}_t^\alpha(S))$. In other words, we have, for all $t \in \mathbb{R}$, $\lambda_\mu(\mathcal{E}_t^\alpha(S)) = \mathcal{E}_t^\alpha(\lambda_\mu(S))$.*

Theorem 3.43 on the commutativity of the earthquake flow along components of the stump with the stretch flow (along completions of the stump) can provide a class of examples of converging anti-stretch lines. Indeed, as soon as we know that some anti-stretch line directed by a complete geodesic lamination μ converges to a point on Thurston's boundary, all the other anti-stretch lines obtained by the action of an earthquake along a component of the stump also converge to the same point.

4 Problems

In this last section, we have collected a few problems, which concern mainly Thurston's asymmetric metric. Some of them may be easy.

Problem I. On the non-symmetry of Thurston's metric. There are several natural questions that arise directly from the fact that Thurston's metric K is non-symmetric. For instance: characterize the pairs of hyperbolic structures g and h that satisfy $K(g, h) = K(h, g)$. In other words, study the locus in $\mathcal{T}(S) \times \mathcal{T}(S)$ defined by the equation $K(g, h) = K(h, g)$. Since K is a geodesic metric, another natural question is: give necessary and/or sufficient conditions on the hyperbolic surfaces g and h under which there exists a geodesic segment from g to h which is also (up to reparametrization) a geodesic segment from h to g .

Problem II. On the symmetrization of Thurston's asymmetric metric. There are several definitions for the symmetrization of an asymmetric metric K , none of them being more natural than the others. Two such options are

$$\sigma K(g, h) = \max\{K(g, h), K(h, g)\}$$

and

$$SK(g, h) = \frac{1}{2}(K(g, h) + K(h, g)).$$

Each of these symmetrizations is a genuine metric. In the case where K is Thurston's asymmetric metric, it is natural to try to relate these two associated symmetrizations to other known metrics on that space.

In this respect, we mention that one can adapt Thurston's definition of his asymmetric metric to the context of the Teichmüller space $\mathcal{T}_{1,0}$ of $S_{1,0}$, that is, of the torus. $\mathcal{T}_{1,0}$ is the space of flat metrics on the torus up to isotopy and homothety. *A priori*, we have two distinct asymmetric metrics, L and K , defined as in (3.4) and (3.5) above, on $\mathcal{T}_{1,0}$, hyperbolic length being replaced here by Euclidean length, with a suitable normalization that takes care of homothety. Such a study has been carried out in the paper [7], where it is shown that $K = L$, as in the case of surfaces of negative Euler characteristic, but with different techniques of proof. Recall that there is a natural identification between $\mathcal{T}_{1,0}$ and \mathbb{H}^2 . With this identification, we obtain a nonseparating and nonsymmetric metric $\delta = K = L$ on \mathbb{H}^2 . An explicit formula for δ is given in [7], and it is shown there that its symmetrization $S\delta$ is the Poincaré metric of \mathbb{H}^2 , which, as is well known, is also the Teichmüller metric on $\mathcal{T}_{1,0}$.

It is unlikely that for surfaces of negative Euler characteristic, some particular symmetrization of Thurston's asymmetric metric is Teichmüller's metric, but one can ask the reverse question, that is, to find an interesting asymmetric metric on Teichmüller space whose symmetrization is Teichmüller's metric. In fact, one can ask the same question for the other known metrics on Teichmüller space.

Problem III. Comparing Thurston's asymmetric metric with other metrics on Teichmüller space. In a naive approach, one can think that Thurston's asymmetric metric is very different from Teichmüller's metric, because the former is defined using hyperbolic geometry, whereas the definition of the latter is based on complex analysis. But, as is well known, there are many tools that make the relation between conformal and hyperbolic geometry. One of the basic tools is a result of Wolpert ([72], p. 326) stating that given any two hyperbolic structures g and h on S , then, for any quasiconformal homeomorphism $f: g \rightarrow h$ with quasiconformal dilatation $K(f)$ and for any homotopy class γ of essential simple closed curves on S , we have

$$l_h(f(\gamma)) \leq K(f)l_g(\gamma).$$

This gives the following inequality between Thurston's asymmetric metric L and the Teichmüller metric:

$$L(g, h) \leq 2d_T(g, h).$$

Indeed, let $f: (S, g) \rightarrow (S, h)$ be the Teichmüller map, where g and h are considered as conformal structures. Wolpert's result implies

$$\frac{l_h(\gamma)}{l_g(\gamma)} \leq K(f) = e^{2d_T(g, h)}$$

which gives

$$\log \frac{l_h(\gamma)}{l_g(\gamma)} \leq \log K(f) = 2d_T(g, h),$$

hence $L(g, h) \leq 2d_T(g, h)$.

We already mentioned a result by L. X. Liu (see [37]) showing that there is no Lipschitz comparison in the reverse sense.

In this respect, we note that in [7], it is shown that in the case of the Teichm  ller space of the torus, Thurston's asymmetric metric and Teichm  ller's metric are not quasi-isometric. (In that paper, there is a discussion of the notion of quasi-isometry in the context of non-symmetric metrics.)

It may also be possible to relate Thurston's asymmetric metric to the Weil–Petersson metric on Teichm  ller space $\mathcal{T}(S)$. This question is motivated by a characterization due to Thurston and Wolpert of the Weil–Petersson metric that is based, like Thurston's asymmetric metric, on the length-spectrum of closed curves. Indeed, Thurston introduced a Riemannian metric on $\mathcal{T}(S)$ where the scalar product of two tangent vectors at some hyperbolic surface (considered as an element of $\mathcal{T}(S)$) is defined as the second derivative with respect to the earthquake flows along these vectors of the length of a uniformly distributed sequence of closed geodesics on the given hyperbolic surface, and Wolpert showed that this metric coincides with the Weil–Petersson metric; see [73]. Note that the Weil–Petersson geodesics that consist in pinching a closed curve to a point have a certain resemblance to anti-stretch lines, without the parametrization.

We note that if d_{qi} is the metric introduced in [70] by Thurston on Teichm  ller space, whose definition we recalled in (3.1) above, and if d_T denotes as before Teichm  ller's metric, then there exists a constant $C > 0$ such that

$$d_T \leq d_{qi} \leq Cd_T$$

(see [70], p. 268, where Thurston attributes this result to Douady and Earle).

We note finally that it is easy to make definitions of (symmetric or asymmetric) metrics on Teichm  ller space that are based on the comparison of lengths of closed geodesics between hyperbolic surfaces, but the interesting goal is to define metrics that have nice geometrical properties. For instance, one can take any finite collection $\{\gamma_1, \dots, \gamma_n\}$ of simple closed curves on the surface whose lengths for an arbitrary hyperbolic metric completely determine that metric, and using these curves, one can define a distance between two hyperbolic metrics g and h by taking

$$\log \sup_{i=1, \dots, n} \left\{ \frac{l_g(\gamma_i)}{l_h(\gamma_i)}, \frac{l_h(\gamma_i)}{l_g(\gamma_i)} \right\}.$$

Clearly, this defines a metric on Teichm  ller space, but does it have interesting geometric properties?

Problem IV. Isometries. Isometries between spaces equipped with asymmetric metrics can be defined in the same way as between usual metric spaces, except that one has to be careful about the order of the variables in the case of an asymmetric metric. A well-known result of H. L. Royden [55] states that the group of isometries of the Teichm  ller metric is the mapping class group. Masur and Wolf proved an analogous result for the Weil–Petersson metric, see [45]. Is the same statement true for Thurston's asymmetric metric? It is easy to see that the elements of the mapping

class group are isometries for Thurston's metric. Therefore, the question is about the reverse inclusion.

We also recall that Bers obtained in [11] a classification of the isometries of the Teichmüller metric in terms of the properties of the displacement function and of the intersection pattern of the minimal displacement sets of the isometries. (Bers' classification is based on Thurston's classification of mapping classes into pseudo-Anosov, reducible and parabolic, but it also constitutes an independent approach to the classification.) Likewise, Daskalopoulos and Wentworth obtained in [17] a classification of isometries of the Weil–Petersson metric, again in terms of the displacement function and the displacement sets. This result is also described in Chapter 1 of this Handbook, see [18]. It is natural to ask for an analogous classification for the isometries of Thurston's asymmetric metric. Note that this is not equivalent to the problem of showing that the isometry group of Thurston's asymmetric metric is the mapping class group.

Problem V. Geodesics. Thurston showed that stretch lines are geodesics for Thurston's asymmetric metric and that any two points in Teichmüller space can be joined by a geodesic path that is a concatenation of stretch segments (see Theorem 8.5 of [67] or Theorem 3.21 above for a precise statement). However, there are other types of geodesics. Thus, an interesting problem is to describe an arbitrary geodesic. It seems unlikely that any geodesic is a limit of a concatenations of stretch segments. Thurston also proved that a geodesic path joining two points in Teichmüller space is in general not unique. Therefore, another natural question is to characterize the set of ordered pairs of points such that the geodesic joining them is unique.

Problem VI. The dual Thurston asymmetric metric. Work out an asymptotic formula linking $K(g, h)$ and $K(h, g)$. To find a precise formula is probably not a reasonable problem. Along a stretch line, we suspect a formula reminding the collar formula $\sinh(aK(g, h)) \sinh(bK(h, g)) \simeq c$ with some constants a, b, c depending on the genus and on the number and punctures of the surface S , and on the "complexity" of the complete lamination directing the stretch line. We already know that Thurston's asymmetric metric and its dual are not Lipschitz equivalent. In fact, there are no constants C_1 and C_2 such that for all g and h in \mathcal{T} , we have $K(g, h) \leq C_1 K(h, g) + C_2$ (Proposition 3.33 above).

Problem VII. Anti-stretch lines. Find conditions on a hyperbolic metric equipped with a complete geodesic lamination so that the anti-stretch line starting at that hyperbolic metric and directed by that lamination is also a stretch line (up to reparametrization).

Problem VIII. Curvature. Study the various existing notions of curvature of Teichmüller space equipped with Thurston's asymmetric metric (Finsler curvature, an asymmetric version of Gromov hyperbolicity, boundedness of curvature in the sense of Alexandroff and so on). Note that this metric is not nonpositively curved in the sense of Busemann since there may be several geodesic segments joining two points.

Problem IX. Convergence of anti-stretch rays. For a given complete geodesic lamination μ with nonempty stump γ , find weaker hypotheses than those of Theorem 3.39 that imply the convergence of an anti-stretch ray directed by μ to the projective class of γ . The convergence result in Theorem 3.39 may suggest that if the measured geodesic lamination γ is not uniquely ergodic, then either an anti-stretch ray directed by μ does not converge to a definite point on Thurston’s boundary, or it converges to the projective class of the geodesic lamination γ equipped with some transverse measure which would be a “barycenter” of all transverse measures carried by γ . Thus, a natural question would be to find conditions under which one of these two cases occurs. One can reason by analogy with Masur’s result regarding Teichm  ller’s metric stated as Theorem 2.42 above, where a whole family of rays corresponding to different transverse measures on a given foliation converge to a single point, corresponding to a special transverse measure on that foliation.

Problem X. Convexity. Kerckhoff showed in [30] that the geodesic-length function on Teichm  ller space associated to a simple closed geodesic is convex along earthquake paths and that the sum of geodesic-length functions associated to a finite family of geodesics that fills up the surface is strictly convex. Wolpert showed in [74] that geodesic-length functions are strictly convex along Weil–Peterson geodesics. Such convexity results have been used by several authors to solve various problems. For instance, Kerckhoff used the convexity of geodesic-length functions along earthquake paths to obtain a solution of the Nielsen realization problem [30]. Wolpert, in his paper [74], used the strict convexity of geodesic-length functions along Weil–Peterson geodesics to obtain a new proof of the fact that Teichm  ller space is Stein, and a new solution of the Nielsen realization problem. Kerckhoff, in his paper [32] used the convexity properties that he obtained in [30] to develop his theory of *lines of minima* in Teichm  ller space, which opened a new geometrical point of view on that space. Thurston, in his preprint [68], constructed a mapping class group-equivariant spine for the Teichm  ller space of a closed surface based on the convexity of the geodesic-length function. P. Schmutz further developed such a theory in his paper [56], where he also used the convexity of the geodesic-length function to study what he called a *systole function* on Teichm  ller space. In view of all that, it is natural to study convexity properties of length functions along stretch lines. The graphs that we plotted in 3.11 show that this function is not convex in general, but one can conjecture that it is peakless in the sense of Busemann. At the infinitesimal level, Thurston’s asymmetric metric is convex in the sense that it is a Finsler metric, that is, it is defined by a length structure induced by a norm on each tangent space whose closed ball is a convex body (which is not strictly convex).

Problem XI. The visual boundary. Describe the visual boundary at any point (or at some class of points) of Teichm  ller space equipped with Thurston’s asymmetric metric. For that, one needs first to understand all the geodesic rays starting at a point (cf. Problem V above). Does the visual boundary depend on the choice of the basepoint? Does the action of the mapping class group extend to the space union its

visual boundary? One can ask similar questions about the visual boundary of the dual asymmetric metric $K^*(x, y) = K(y, x)$.

Problem XII. Stretch maps between general metric spaces. Work out a theory of stretch maps between general (i.e. not necessarily hyperbolic) metrics on a surface. This problem is mentioned by Thurston in his paper [67]. A particularly interesting class of metrics on surfaces is the class of Euclidean metrics with cone singularities. It is also an interesting problem to study stretch maps between higher-dimensional manifolds equipped with metrics of constant curvature, or between singular spaces (graphs, two-complexes and so on).

Problem XIII. Moduli space. Study the behaviour of stretch and anti-stretch lines in moduli space.

References

- [1] W. Abikoff, *The real analytic theory of Teichmüller space*. Lecture Notes in Math. 820, Springer-Verlag, Berlin 1980. 116, 146
- [2] L. V. Ahlfors, *Lectures on quasiconformal mappings*. Van Nostrand Math. Stud. 10, D. Van Nostrand Company, Princeton, N.J., 1966. 128
- [3] L. V. Ahlfors, *Collected Papers*. Vol. 1, 2. Birkhäuser, Basel 1982. 116
- [4] L. Ahlfors and A. Beurling, Invariants conformes et problèmes extrémaux. In *Den 10. Skandinaviske Matematiker Kongres* (København, 26.–30. August 1946), Jul. Gjellerups Forlag, København 1947, 341–351; reprinted in *Collected Works of Arne Beurling*, Birkhäuser, Basel 1989, 135–145. 125, 129
- [5] L. V. Ahlfors and L. Sario, *Riemann surfaces*. Princeton Math. Ser. 26, Princeton University Press, Princeton, N.J., 1960. 121
- [6] R. Baer, Isotopien von Kurven auf orientierbaren, geschlossenen Flächen. *J. Reine Angew. Math.* 159 (1928), 101–116. 116
- [7] A. Belkhirat, A. Papadopoulos, and M. Troyanov, Thurston's weak metric on the Teichmüller space of the torus. *Trans. Amer. Math. Soc.* 357 (2005), 3311–3324. 197, 198
- [8] R. Benedetti and C. Petronio, *Lectures on hyperbolic geometry*. Universitext, Springer-Verlag, Berlin 1992. 146
- [9] L. Bers, Quasiconformal mappings and Teichmüller's theorem. In *Analytic functions*, Princeton University Press, Princeton, N.J., 1960, 89–119.
- [10] L. Bers, The equivalence of two definitions of quasiconformal mappings. *Comment. Math. Helv.* 37 (1962/1963), 148–154. 126
- [11] L. Bers, An extremal problem for quasiconformal mappings and a theorem by Thurston. *Acta Math.* 141 (1978), 73–98. 199
- [12] L. Bers, *Selected works of Lipman Bers: papers on complex analysis*. Edited by Irwin Kra and Bernard Maskit, Amer. Math. Soc., Providence, R.I., 1998. 116
- [13] H. Busemann, *The geometry of geodesics*. Pure Appl. Math. 6, Academic Press, New York, N.Y. 1955. 191

- [14] H. Busemann, *Recent synthetic differential geometry*. Ergeb. Math. Grenzgeb. 54, Springer-Verlag, New York 1970. [171](#)
- [15] A. J. Casson and S. A. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*. London Math. Soc. Stud. Texts 9, Cambridge University Press, Cambridge 1988. [147](#), [148](#)
- [16] Y-E. Choi, K. Rafi, Comparison Between Teichmüller and Lipschitz Metrics. Preprint, 2005 ;arXiv:math.GT/0510136. [172](#), [173](#)
- [17] G. Daskalopoulos and R. Wentworth, Classification of Weil-Petersson isometries. *Amer. J. Math.* 125 (2003), 941–975. [199](#)
- [18] 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. [131](#), [199](#)
- [19] D. B. A. Epstein, Curves on 2-manifolds and isotopies. *Acta Math.* 115 (1966), 83–107. [116](#)
- [20] A. Fathi, F. Laudenbach, and V. Poénaru (eds.), Travaux de Thurston sur les surfaces. *Astérisque* 66–67 (1979). [118](#), [119](#), [120](#), [140](#), [143](#), [146](#), [147](#), [167](#), [193](#)
- [21] A. Fletcher and V. Markovic, *Quasiconformal maps and Teichmüller theory*. Oxford Grad. Texts Math. 11, Oxford University Press, Oxford 2007.
- [22] F. P. Gardiner, *Teichmüller theory and quadratic differentials*. John Wiley & Sons, New York 1987. [116](#)
- [23] H. Grötzsch, Über die Verzerrung bei schlichten nichtkonformen Abbildungen und über eine damit zusammenhängende Erweiterung des Picardschen Satzes. *Ber. Verh. Sächs. Akad. Wiss. Leipzig* 80 (1928), 367–376. [132](#)
- [24] 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. [127](#)
- [25] J. H. Hubbard, *Teichmüller theory and applications to geometry, topology, and dynamics*. Vol. 1, Matrix Editions, Ithaca, N.Y., 2006. [128](#), [131](#)
- [26] J. Hubbard and H. Masur, Quadratic differentials and foliations. *Acta Math.* 142 (1979), 221–274. [116](#)
- [27] Y. Imayoshi and M. Taniguchi, *An introduction to Teichmüller spaces*. Springer-Verlag, Tokyo 1992. [132](#)
- [28] J. A. Jenkins, On the existence of certain general extremal metrics. *Ann. of Math.* (2) 66 (1957), 440–453. [116](#), [146](#)
- [29] S. P. Kerckhoff, The asymptotic geometry of Teichmüller space. *Topology* 19 (1980), 23–41. [131](#)
- [30] S. P. Kerckhoff, The Nielsen realization problem. *Ann. of Math.* (2) 117 (1983), 235–265. [129](#), [132](#), [134](#), [138](#), [139](#)
- [31] S. P. Kerckhoff, Earthquakes are analytic. *Comment. Math. Helv.* 60 (1985), 17–30. [195](#), [200](#)
- [32] S. P. Kerckhoff, Lines of minima in Teichmüller space. *Duke Math. J.* 65 (1992), 187–213.
- [33] S. Kravetz, On the geometry of Teichmüller spaces and the structure of their modular groups. *Ann. Acad. Sci. Fenn. Ser. A I* 278 (1959), 35. [200](#)
- [34] O. Lehto and K. I. Virtanen, *Quasiconformal mappings in the plane*. Grundlehren Math. Wiss. 126, 2nd ed., Springer-Verlag, Berlin 1973. [113](#), [136](#)

- [35] A. Lenzhen, Teichmüller geodesics that don't have a limit in \mathcal{PMF} . Preprint, University of Illinois at Chicago, Chicago 2005. [124](#), [125](#)
- [36] G. Levitt, Foliations and laminations on hyperbolic surfaces. *Topology* 22 (1983), 119–135. [141](#)
- [37] L. X. Liu, On non-quasiisometry of the Teichmüller metric and Thurston's pseudometric. *Chinese J. Contemp. Math.* 20 (1) (1999), 31–36. [148](#), [150](#)
[172](#), [198](#)
- [38] L. X. Liu, On the metrics of length spectrum in Teichmüller space. *Chinese J. Contemp. Math.* 22 (1) (2001), 23–34. [170](#)
- [39] W. Mangler, Die Klassen von topologischen Abbildungen einer geschlossenen Fläche auf sich. *Math. Z.* 44 (1939), 541–554. [116](#)
- [40] H. Masur, On a class of geodesics in Teichmüller space. *Ann. of Math.* (2) 102 (1975), 205–221. [137](#)
- [41] H. Masur, Dense geodesics in moduli space. In *Riemann surfaces and related topics* (State Univ. New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud. 97, Princeton Univ. Press, Princeton, N.J., 1981, 417–438. [129](#)
- [42] H. Masur, Interval exchange transformations and measured foliations. *Ann. of Math.* (2) 115 (1982), 169–200. [140](#)
- [43] H. Masur, Two boundaries of Teichmüller space. *Duke Math. J.* 49 (1982), 183–190. [139](#), [140](#)
- [44] H. A. Masur and M. Wolf, Teichmüller space is not Gromov hyperbolic. *Ann. Acad. Sci. Fenn. Ser. A I Math.* 20 (1995), 259–267. [142](#)
- [45] H. Masur and M. Wolf, The Weil-Petersson isometry group. *Geom. Dedicata* 93 (2002), 177–190. [198](#)
- [46] J. D. McCarthy and A. Papadopoulos, The visual sphere of Teichmüller space and a theorem of Masur-Wolf. *Ann. Acad. Sci. Fenn. Math. A I Math.* 24 (1999), 147–154. [141](#), [142](#)
- [47] Y. Minsky, Extremal length estimates and product regions in Teichmüller space. *Duke Math. J.* 83 (1996), 249–286.
- [48] J. Nielsen, Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen. *Acta Math.* 50 (1927), 189–358.
- [49] A. Papadopoulos, Geometric intersection functions and Hamiltonian flows on the space of measured foliations on a surface. *Pacific J. Math.* 124 (1986), 375–402.
- [50] A. Papadopoulos, L'extension du flot de Fenchel-Nielsen au bord de Thurston de l'espace de Teichmüller. *C. R. Acad. Sci. Paris Sér. I Math.* 302 (1986), 325–327. [195](#)
- [51] A. Papadopoulos, Sur le bord de Thurston de l'espace de Teichmüller d'une surface non compacte. *Math. Ann.* 282 (1988), 353–359.
- [52] A. Papadopoulos, On Thurston's boundary of Teichmüller space and the extension of earthquakes. *Topology Appl.* 41 (1991), 147–177. [169](#), [170](#), [193](#), [194](#), [196](#)
- [53] A. Papadopoulos and G. Théret, On the topology defined by Thurston's asymmetric metric. *Math. Proc. Cambridge Philos. Soc.* (2007), to appear. [169](#), [170](#)
- [54] R. C. Penner and J. L. Harer, *Combinatorics of train tracks*. Ann. of Math. Stud. 125, Princeton University Press, Princeton, N.J., 1992. [140](#), [147](#)

- [55] H. L. Royden, Automorphisms and isometries of Teichm  ller space. 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, 369–383. [198](#)
- [56] P. Schmutz Schaller, A cell decomposition of Teichm  ller space based on geodesic length functions. *Geom. Funct. Anal.* 11 (2001), 142–174. [200](#)
- [57] G. Springer, *Introduction to Riemann surfaces*. Addison-Wesley, Reading, Mass., 1957. [121](#)
- [58] K. Strebel,   ber quadratische Differentiale mit geschlossenen Trajektorien und extremale quasikonforme Abbildungen. In *Festband zum 70. Geburtstag von Rolf Nevanlinna* (Z  rich, 4.–6. November 1965), edited by H. P. K  nzi and A. Pfluger, Springer-Verlag, Berlin 1966, 105–127. [132](#)
- [59] K. Strebel, *Quadratic differentials*. *Ergeb. Math. Grenzgeb.* (3) 5, Springer-Verlag, Berlin 1984. [130](#), [131](#)
- [60] O. Teichm  ller, Extremale quasikonforme Abbildungen und quadratische Differentiale. *Abh. Preuss. Akad. Wiss. Math.-Nat. Kl.* 1939 (1940). [135](#)
- [61] O. Teichm  ller, Bestimmung der extremalen quasikonformen Abbildungen bei geschlossenen orientierten Riemannschen Fl  chen. *Abh. Preuss. Akad. Wiss. Math.-Nat. Kl.* 1943 (1943). [135](#)
- [62] O. Teichm  ller, *Gesammelte Abhandlungen*. Springer-Verlag, Berlin 1982. [116](#)
- [63] G. Th  ret, On the negative convergence of Thurston’s negative stretch lines towards the boundary of teichm  ller space. *Ann. Acad. Sci. Fenn. Math.*, to appear. [151](#), [191](#), [192](#), [194](#), [195](#), [196](#)
- [64] G. Th  ret, On Thurston’s stretch lines in Teichm  ller space. Preprint, Universit   Louis Pasteur, Strasbourg 2005. [191](#)
- [65] W. P. Thurston, The geometry and topology of three-manifolds. Mimeographed notes, Princeton University, Princeton 1976. [140](#), [147](#), [150](#), [158](#)
- [66] W. P. 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. [195](#)
- [67] W. P. Thurston, Minimal stretch maps between hyperbolic surfaces. Preprint 1986. [155](#), [165](#), [166](#), [167](#), [168](#), [171](#), [199](#), [201](#)
- [68] W. P. Thurston, A spine for Teichm  ller space. Preprint 1986. [200](#)
- [69] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)* 19 (1988), 417–431. [118](#)
- [70] W. P. Thurston, *Three-dimensional geometry and topology*. Vol. 1, edited by Silvio Levy, Princeton Math. Ser. 35, Princeton University Press, Princeton, N.J., 1997. [142](#), [143](#), [144](#), [146](#), [164](#), [198](#)
- [71] W. A. Veech, Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math.* (2) 115 (1982), 201–242. [140](#)
- [72] S. Wolpert, The length spectra as moduli for compact Riemann surfaces. *Ann. of Math.* (2) 109 (1979), 323–351. [197](#)
- [73] S. A. Wolpert, Thurston’s Riemannian metric for Teichm  ller space. *J. Differential Geom.* 23 (1986), 143–174. [198](#)
- [74] S. A. Wolpert, Geodesic length functions and the Nielsen problem. *J. Differential Geom.* 25 (1987), 275–296. [200](#)

Chapter 3

Surfaces, circles, and solenoids

Robert C. Penner

Contents

1	Introduction	205
2	Background	206
3	Punctured surfaces	209
4	Coordinates for circle homeomorphisms	211
5	Coordinates for the solenoid	214
6	Concluding remarks	220
	References	220

1 Introduction

The “lambda length” of a pair of disjoint horocycles in upper halfspace centered at $u, v \in \mathbb{R}$ of respective diameters c, d is defined to be $\sqrt{\frac{2}{cd}} |u - v|$ and is roughly the exponential of the hyperbolic distance between them. (See §1 for more precision.) These invariants can be used to devise coordinates in several different guises: for the Teichmüller space of punctured surfaces [7]; for the space of cosets of the Möbius group Möb of real fractional linear transformations in the topological group of all orientation-preserving homeomorphisms of the circle, which forms a generalized universal Teichmüller space [8]; and for the Teichmüller space of the “punctured solenoid”, which is the punctured analogue introduced in [10] (and defined in §4) of the space studied by Sullivan [12] to analyze dynamical properties of the mapping class group actions on the Teichmüller spaces for closed surfaces. In fact in each case, lambda lengths give coordinates for the “decorated” Teichmüller space rather than the Teichmüller space. (The respective notions of decoration are defined in §§ 2, 3, 4.) Furthermore, the manifestation of lambda lengths as coordinates on the decorated Teichmüller space of the punctured solenoid is the first step of a larger ongoing program with Šarić [10] to extend the decorated Teichmüller theory [7]–[9] to the solenoid.

To define the punctured solenoid \mathcal{H} as a topological space, for definiteness fix the “modular” group $G = \text{PSL}_2(\mathbb{Z})$ of integral fractional linear transformations, let \hat{G} denote its pro-finite completion (whose definition is recalled in §4), let \mathbb{D} denote the open unit disk in the complex plane, and define $\mathcal{H} = (\mathbb{D} \times \hat{G})/G$, where $\gamma \in G$ acts on $(z, t) \in \mathbb{D} \times \hat{G}$ by $\gamma(z, t) = (\gamma z, t\gamma^{-1})$. In analogy to the case of punctured

surfaces, we may produce appropriate geometric structures on \mathcal{H} by taking suitable quotients $(\mathbb{D} \times \hat{G})/G$ by other actions of G on $\mathbb{D} \times \hat{G}$. As a pro-finite completion, the punctured solenoid itself is defined essentially number theoretically in terms of finite-index subgroups of the modular group, and aspects of its Teichmüller theory bear close relation to classical questions in number theory (as mentioned at the end of §5, which also contains other concluding and speculative remarks).

We take this opportunity to correct Theorem 6.4 from [8]. See the remarks following Theorem 8 for the correction to the universal Teichmüller theory and Proposition 12 for the corresponding affirmative statement for the solenoid. Volume II of this Handbook contains a chapter on the Teichmüller theory of the solenoid written by D. Šarić [12].

Acknowledgement. It was discussions with Mahmoud Zeinalian that led to the original idea of employing lambda lengths for solenoids, and is a pleasure to thank him, as well as Bob Guralnick for useful comments on classical number theory. The new material in §4 on the punctured solenoid is joint work [10] with Dragomir Šarić, who patiently explained his earlier work [11], and it is also a pleasure to thank him for many stimulating conversations.

2 Background

Define the Minkowski inner product $\langle \cdot, \cdot \rangle$ on \mathbb{R}^3 whose quadratic form is given by $x^2 + y^2 - z^2$ in the usual coordinates. The upper sheet

$$\mathbb{H} = \{u = (x, y, z) \in \mathbb{R}^3 : \langle u, u \rangle = -1 \text{ and } z > 0\}$$

of the two-sheeted hyperboloid is isometric to the hyperbolic plane. Indeed, identifying the Poincaré disk \mathbb{D} with the open unit disk at height zero about the origin in \mathbb{R}^3 , central projection $\mathbb{H} \rightarrow \mathbb{D}$ from $(0, 0, -1) \in \mathbb{R}^3$ establishes an isometry. Moreover, the open positive light cone

$$L^+ = \{u = (x, y, z) \in \mathbb{R}^3 : \langle u, u \rangle = 0 \text{ and } z > 0\}$$

is identified with the collection of all horocycles in \mathbb{H} via the affine duality $u \mapsto h(u) = \{w \in \mathbb{H} : \langle w, u \rangle = -1\}$. Identifying the unit circle S^1 with the boundary of \mathbb{D} , the central projection extends continuously to the projection $\Pi: L^+ \rightarrow S^1$ which maps a horocycle in L^+ to its center in S^1 .

Define a “decorated geodesic” to be an unordered pair $\{h_0, h_1\}$ of horocycles with distinct centers in the hyperbolic plane, so there is a well-defined geodesic connecting the centers of h_0 and h_1 ; the two horocycles may or may not be disjoint, and there is a well-defined signed hyperbolic distance δ between them (taken to be positive if and only if $h_0 \cap h_1 = \emptyset$) as illustrated in the two cases of Figure 1. The *lambda length* of the decorated geodesic $\{h_0, h_1\}$ is defined to be the transform $\lambda(h_0, h_1) = \sqrt{2 \exp \delta}$. Taking this particular transform renders the identification h geometrically natural in the sense that $\lambda(h(u_0), h(u_1)) = \sqrt{-\langle u_0, u_1 \rangle}$, for $u_0, u_1 \in L^+$ as one can check.

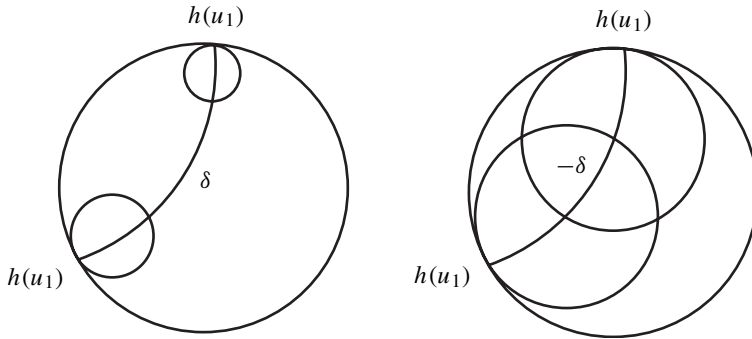


Figure 1. Decorated geodesics.

Three useful lemmas (with computational proofs, which we do not reproduce here) are as follows:

Lemma 1 ([7], Lemma 2.4). *Given three rays $\vec{r}_0, \vec{r}_1, \vec{r}_2 \subseteq L^+$ from the origin which contain linearly independent vectors and given three numbers $\lambda_0, \lambda_1, \lambda_2 \in \mathbb{R}_{>0}$, there are unique points $u_i \in \vec{r}_i$, for $i = 0, 1, 2$ so that $\lambda(h(u_i), h(u_j)) = \lambda_k$, where $\{i, j, k\} = \{0, 1, 2\}$. The points u_0, u_1, u_2 depend continuously on $\lambda_0, \lambda_1, \lambda_2$ and on $\vec{r}_0, \vec{r}_1, \vec{r}_2$.*

Lemma 2 ([7], Lemma 2.3). *Given two points $u_0, u_1 \in L^+$, which do not lie on a common ray through the origin, and given two numbers $\lambda_0, \lambda_1 \in \mathbb{R}_{>0}$, there is a unique point $v \in L^+$ on either side of the plane through the origin containing u_0, u_1 satisfying $\lambda(h(v), h(u_i)) = \lambda_i$, for $i = 0, 1$. The point v depends continuously on u_0, u_1 and on λ_0, λ_1 .*

Lemma 3 ([7], Proposition 2.8). *Suppose that $u_0, u_1, u_2 \in L^+$ are linearly independent, let $\gamma(u_i, u_j)$ denote the geodesic in \mathbb{H} with ideal vertices given by the centers of $h(u_i)$ and $h(u_j)$, for $i \neq j$, and define*

$$-\lambda_i^2 = \langle u_j, u_k \rangle, \quad \alpha_i = \frac{\lambda_i}{\lambda_j \lambda_k} \quad \text{for } \{i, j, k\} = \{0, 1, 2\}.$$

Then $\sqrt{2}\alpha_i$ is the hyperbolic length along the horocycle $h(u_i)$ between $\gamma(u_i, u_j)$ and $\gamma(u_i, u_k)$, for $\{i, j, k\} = \{0, 1, 2\}$.

Remark 1. Consider an ideal quadrilateral Q in \mathbb{D} decorated so as to give four points in L^+ . The edges of Q have well defined lambda lengths, say a, b, c, d in correct cyclic (clockwise) order about the boundary of Q . Choose a diagonal of Q , where the diagonal has lambda length e and separates edges with lambda lengths a, b from edges with lambda lengths c, d . The other diagonal of Q has its lambda length f given by

$$ef = ac + bd,$$

and we say that f arises from e by a *Ptolemy transformation* on the lambda lengths. To see this, note that the formula for a Ptolemy transformation is independent under scaling any of the four points in L^+ , so we may alter the decoration and assume that the four points lie in a common horizontal plane. In this plane, the Minkowski inner product induces a multiple of the usual Euclidean metric, and the intersection of L^+ with this plane is a round circle. The formula for the Ptolemy transformation thus follows from Ptolemy's classical formula on Euclidean lengths of quadrilaterals that inscribe in a circle.

Remark 2. Consider a decorated triangle, say with lambda lengths x, y, z in the cyclic order about the boundary of the triangle determined by an orientation, and define a 2-form

$$\omega(x, y, z) = d\ln x \wedge d\ln y + d\ln y \wedge d\ln z + d\ln z \wedge d\ln x.$$

A calculation shows that $\omega(a, b, e) + \omega(c, d, e) = \omega(b, c, f) + \omega(d, a, f)$, thus assigning a well-defined Ptolemy-invariant 2-form to an oriented decorated quadrilateral. Regard the Poincaré disk as the open unit disk \mathbb{D} in the complex plane in the usual way so that the unit circle S^1 is identified with the circle at infinity, and let Δ denote the ideal hyperbolic triangle with vertices $+1, -1, -\sqrt{-1} \in S^1$ as in Figure 2.

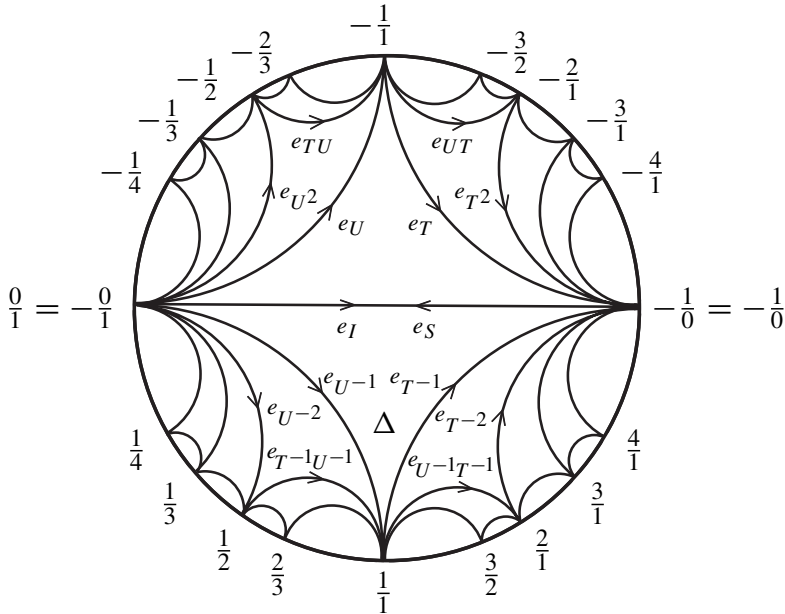


Figure 2

Let Γ denote the group generated by reflections in the sides of Δ , and define the *Farey tessellation* τ_* to be the full Γ -orbit of the frontier of Δ . We refer to geodesics in τ_* as *edges* of τ_* and think of τ_* itself as a set of edges. The ideal vertices of the edges

of τ_* are naturally identified with the set \mathbb{Q} of all rational numbers including infinity, where for instance $+1, -1, -\sqrt{-1} \in S^1$ correspond respectively to $\infty = \frac{1}{0}, 0 = \frac{0}{1}, 1 = \frac{1}{1}$ as illustrated in Figure 2. Let $\mathbb{Q} \subseteq S^1$ denote the corresponding countable dense subset of S^1 which we refer to simply as the set of *rational points* of S^1 . Define the *distinguished oriented edge* or *doe* of the Farey tessellation to be the oriented edge from $\frac{0}{1}$ to $\frac{1}{0}$.

The *modular group* $\mathrm{PSL}_2 = \mathrm{PSL}_2(\mathbb{Z})$ of integral fractional linear transformations is the subgroup of Γ consisting of compositions of an even number of reflections, and PSL_2 acts simply transitively on the set of orientations on the edges of τ_* . The assignment

$$e_A = (\mathrm{doe})A \quad \text{for } A \in \mathrm{PSL}_2$$

establishes a bijection between PSL_2 and the set of oriented edges of τ_* as illustrated in Figure 2. In particular, the doe of τ_* is e_I , where I denotes the identity of PSL_2 .

We adopt the standard notation

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

for certain elements of PSL_2 , where S is involutive and fixes the unoriented edge of τ_* underlying the doe while changing its orientation, and U (respectively T) is the parabolic transformation with fixed point $\frac{0}{1}$ (respectively $\frac{1}{0}$) which cyclically permutes the incident edges of τ_* in the counter-clockwise sense about $\frac{0}{1}$ (respectively the clockwise sense about $\frac{1}{0}$). In fact, $U^{-1} = STS$, $T^{-1} = SUS$, and any two of S, T, U generate PSL_2 .

We shall also require the full *Möbius group* $\mathrm{Möb} = \mathrm{PSL}_2(\mathbb{R}) \supseteq \mathrm{PSL}_2(\mathbb{Z}) = \mathrm{PSL}_2$ consisting of all real fractional linear transformations.

3 Punctured surfaces

Let $F = F_g^s$ denote a fixed smooth surface of genus g with $s \geq 1$ punctures, where $2 - 2g - s < 0$.

Choose any base-point to determine the fundamental group G of F , and consider the space $\mathrm{Hom}'(G, \mathrm{Möb})$ of all discrete and faithful representations $\rho: G \rightarrow \mathrm{Möb}$ so that no holonomy $\rho(\gamma)$ is elliptic for $\gamma \in G$, and the holonomies around the punctures of F are parabolic. Define the *Teichmüller space*

$$\mathcal{T}(F) = \mathrm{Hom}'(G, \mathrm{Möb}) / \mathrm{Möb},$$

where $\mathrm{Möb}$ acts by conjugacy.

If $\rho \in \mathrm{Hom}'(G, \mathrm{Möb})$, then $\mathbb{D}/\rho(G)$ induces a complete finite-area hyperbolic structure on F , whose punctures are in one-to-one correspondence with the $\rho(G)$ -orbits of the set of fixed points of parabolic elements of $\rho(G)$.

Define the *decorated Teichmüller space* $\tilde{\mathcal{T}}(F) \rightarrow \mathcal{T}(F)$ of F to be the trivial $\mathbb{R}_{>0}^s$ -bundle, where the fiber over a point is the set of all s -tuples of horocycles, one horocycle about each puncture of F , say parametrized by hyperbolic length.

By an *arc family* in F , we mean the isotopy class of a family of essential arcs disjointly embedded in F connecting punctures, where no two arcs in a family may be homotopic rel punctures. If α is a maximal arc family, so that each component of $F - \bigcup \alpha$ is a triangle, then we say that α is an *ideal triangulation* of F .

Theorem 4 ([7], Theorem 3.1). *Fix an ideal triangulation τ of F . Then the assignment of λ -lengths $\tilde{\mathcal{T}}(F) \rightarrow \mathbb{R}_{>0}^s$ is a homeomorphism onto.*

Proof. We must describe an inverse to the mapping and thus give the construction of a decorated hyperbolic structure from an assignment of putative λ -lengths. To this end, consider the topological universal cover \tilde{F} of F and the lift $\tilde{\tau}$ of τ to \tilde{F} ; to each component arc of $\tilde{\tau}$ is associated the lambda length of its projection.

The proof proceeds by induction, and for the basis step, choose any triangle Δ'_0 of $\tilde{\tau}$ and any ideal triangle Δ_0 in \mathbb{H} . The ideal vertices of Δ_0 determine three rays in L^+ , so by Lemma 1, there are three well-defined points in L^+ realizing the putative λ -lengths on the edges of Δ'_0 . (In effect, this basis step of normalizing a triangle “kills” the conjugacy by the Möbius group in the definition of Teichmüller space.) Of course, the triple of points in L^+ corresponds by affine duality to a triple of horocycles, one centered at each ideal vertex of Δ_0 , i.e., a “decoration” on Δ_0 .

To begin the induction step, consider a triangle Δ'_1 adjacent to Δ'_0 across an arc in $\tilde{\tau}$. The two ideal points which Δ'_0 and Δ'_1 share have been lifted to $u, v \in L^+$ in the basis step, and we let $w \in L^+$ denote the lift of the third ideal point of Δ'_0 and consider the plane through the origin determined by u, v . According to Lemma 2, there is a unique lift $z \in L^+$ of the third ideal point of Δ'_1 on the side of this plane not containing the lift of w , where z realizes the putative λ -lengths. Again, u, v, z gives rise via affine duality to another decorated triangle Δ_1 in \mathbb{H} sharing one edge and two horocycles with Δ_0 .

One continues in this manner serially applying Lemma 2 to produce a collection of decorated triangles pairwise sharing edges in \mathbb{H} , where any two triangles have disjoint interiors (because of our choice of the side of the plane in Lemma 2). Thus, the construction gives an injection $\tilde{F} \rightarrow \mathbb{H}$, and we next show that in fact this mapping is also a surjection. To this end, note first that the inductive construction has an image which is open in \mathbb{H} by construction. According to Lemma 3, there is some $\varepsilon > 0$ so that each horocyclic arc inside of each triangle has length at least ε ; indeed, there are only finitely many values for such lengths because the surface is comprised of finitely many triangles. Thus, each application of the inductive step moves a definite amount along each horocycle, and it follows that the construction has an image which is closed as well. It follows from connectivity of \mathbb{H} that $\tilde{F} \rightarrow \mathbb{H}$ is surjective, and furthermore, and can see that $\tilde{\tau}$ is a tessellation of \mathbb{H} , i.e., a locally finite collection of geodesics decomposing \mathbb{H} into ideal triangles.

Following Poincaré, the hyperbolic symmetry group of this tessellation is the required (normalized) Fuchsian group G giving a point of Teichmüller space, and the construction likewise provides a decoration on the quotient \mathbb{H}/G as required. \square

Remark 3. One thinks of the choice of ideal triangulation as a choice of “basis” for the lambda length coordinates. Formulas for the change of basis are given by Ptolemy transformations, and this leads [7], [8] to a faithful representation of the mapping class group of F as well as its universal generalization to \mathcal{Tess} , which is defined in the next section. Furthermore as in Remark 2, the 2-form $\omega = -2 \sum \omega(a, b, c)$ is invariant under this action and descends to the Weil–Petersson form on Riemann’s moduli space, where the sum is over all triangles complementary to any fixed ideal triangulation and the edges of the triangle have lambda lengths a, b, c in correct cyclic order determined by an orientation of F . These two ingredients lead to natural quantizations [3], [4] of Teichmüller space.

4 Coordinates for circle homeomorphisms

Define a *tessellation* τ of the Poincaré disk \mathbb{D} to be a countable locally finite collection of hyperbolic geodesics in \mathbb{D} each of whose complementary regions is an ideal triangle. A *distinguished oriented edge* or *doe* of τ is the specification of an orientation on one of the geodesics in τ . Each geodesic in τ has a pair of asymptotes in S^1 , and we let $\tau^0 \subseteq S^1$ denote the collection of all such asymptotes of geodesics in τ and τ^2 denote the collection of all triangles complementary to $\bigcup \tau$.

Tessellations with doe are “combinatorially rigid” in the following sense. Suppose that τ_1, τ_2 are each a tessellation with doe, say the initial and terminal points of the doe in τ_i are $x_i \in \tau_i^0$ and $y_i \in \tau_i^0$, respectively, for $i = 1, 2$. There is a unique bijection $f: \tau_1^0 \rightarrow \tau_2^0$ so that $f(x_1) = x_2, f(y_1) = y_2$, and whenever x, y, z in correct cyclic order span an oriented triangle in τ_1^2 , then $f(x), f(y), f(z)$ in correct cyclic order also span an oriented triangle in τ_2^2 . This mapping $f: \tau_1^0 \rightarrow \tau_2^0$ is called the *characteristic mapping* of the pair of tessellations with doe. In particular, we may fix $\tau_1 = \tau_*$ to be the Farey tessellation with doe defined in §1, so $\tau_*^0 = \mathbb{Q}$. We may thus define the characteristic mapping $f_\tau: \mathbb{Q} \rightarrow \tau^0$ of the tessellation $\tau = \tau_2$ with doe.

Define the set

$$\mathcal{Tess}' = \{\text{tessellations with doe of } \mathbb{D}\}.$$

To define a topology on \mathcal{Tess}' , if τ is a tessellation with doe, then we may extend the range of the characteristic mapping $f_\tau: \mathbb{Q} \rightarrow \tau^0 \subseteq S^1$ to S^1 . The assignment $\tau \mapsto f_\tau$ determines an embedding of \mathcal{Tess}' into the function space $(S^1)^\mathbb{Q}$ with the compact-open topology (where \mathbb{Q} is given its discrete topology), and we endow \mathcal{Tess}' with the subspace topology.

Define the topological group $\text{Homeo}_+ = \text{Homeo}_+(S^1)$ to be the group of all orientation-preserving homeomorphisms of the circle taken with the compact-open

topology. If $f \in \text{Homeo}_+$ and e is any geodesic in \mathbb{D} , say with ideal points $x, y \in S^1$, then define $f(e)$ to be the geodesic in \mathbb{D} with ideal points $f(x), f(y) \in S^1$. It is not difficult to see that if τ is a tessellation and $f \in \text{Homeo}_+$, then $f(\tau) = \{f(e) : e \in \tau\}$ is also a tessellation. Since a doe on τ determines a doe on $f(\tau)$ in the natural way, there is thus an action of Homeo_+ on \mathcal{Tess} .

Theorem 5 ([8], Theorem 2.3). *The mapping $\text{Homeo}_+ \rightarrow \mathcal{Tess}'$ given by $f \mapsto f(\tau_*)$ is a homeomorphism onto.*

Proof. Injectivity follows from the fact that a homeomorphism is uniquely determined by its values on a dense set. For surjectivity, consider any tessellation with doe τ . Using the fact that \mathbb{Q} and τ^0 are dense in S^1 and the characteristic mapping f_τ is order-preserving by construction, a standard point-set topology argument show that there is a unique orientation-preserving homeomorphism $f_\tau : S^1 \rightarrow S^1$ which restricts to the characteristic mapping. Both mappings $f \mapsto f(\tau)$ and $\tau \mapsto f_\tau$ are continuous by construction. □

There is the natural diagonal left action of the group Möb on $(S^1)^\mathbb{Q}$, which induces a left action of Möb on the subspace \mathcal{Tess}' , and we finally define the *universal Teichmüller space*

$$\mathcal{Tess} = \text{Möb} \backslash \mathcal{Tess}' \approx \text{Möb} \backslash \text{Homeo}_+$$

to be the orbit space with the quotient topology.

A *decoration* on a tessellation τ is the specification of horocycles in \mathbb{D} , one horocycle centered at each point of τ^0 . Via the affine duality discussed in §1, the characteristic mapping $f_\tau : \mathbb{Q} \rightarrow S^1$ on a decorated tessellation τ with doe extends to a mapping $g_\tau : \mathbb{Q} \rightarrow L^+$. The image $g_\tau(\mathbb{Q})$ is automatically “radially dense” in L^+ in the sense that $\Pi(g_\tau(\mathbb{Q}))$ is a dense subset of S^1 , where $\Pi : L^+ \rightarrow S^1$ is the natural projection. Define

$$\widetilde{\mathcal{Tess}}' = \{\text{decorated tessellations } \tau \text{ with doe} : g_\tau(\mathbb{Q}) \text{ is discrete in } L^+\}.$$

The Hausdorff topology on the set of all closed subsets of L^+ induces a subspace topology on the set of all discrete subsets of L^+ , and this in turn induces a compact-open topology on $\widetilde{\mathcal{Tess}}'$. There is again a diagonal left action of Möb by Minkowski isometries on $\widetilde{\mathcal{Tess}}'$, and the *decorated universal Teichmüller space* is finally defined to be the topological quotient

$$\widetilde{\mathcal{Tess}} = \text{Möb} \backslash \widetilde{\mathcal{Tess}}'.$$

There is the natural forgetful map $\widetilde{\mathcal{Tess}} \rightarrow \mathcal{Tess}$, which is evidently continuous.

Given a decorated tessellation $\tilde{\tau}$ with doe and $e \in \tau_*$, there is the corresponding lambda length of the decorated geodesic in $\tilde{\tau}$ with underlying geodesic $f_\tau(e)$, where f_τ is the characteristic mapping of τ . Thus, lambda lengths naturally determine an element of $\mathbb{R}_{>0}^{\tau_*}$.

Theorem 6 ([8], Theorem 3.1). *The assignment of lambda lengths determines an embedding*

$$\widetilde{\mathcal{T}ess} \rightarrow \mathbb{R}_{>0}^{\tau_*}$$

onto an open set, where $\mathbb{R}_{>0}^{\tau_*}$ is given the weak topology (compact-open on $\mathbb{R}_{>0}^{\tau_*}$ with τ_* discrete). Thus, $\widetilde{\mathcal{T}ess}$ inherits the structure of a Fréchet manifold.

Proof. We say that an element $\tau \in \mathcal{T}ess'$ is *normalized* provided that $\{\pm 1\} \subseteq \tau^0$, the doe of τ runs from -1 to $+1$, and the triangle in τ^2 lying to the right of the doe coincides with the triangle spanned by $-1, +1, -\sqrt{-1} \in S^1$. Since Möb acts three-effectively on S^1 and the value of a Möbius transformation at three points of S^1 determines it uniquely, each Möb-orbit on $\mathcal{T}ess'$ admits a unique normalized representative. $\mathcal{T}ess$ is thus canonically identified with the collection of all normalized tessellations. (Again, we have “killed” the Möbius group by normalization.)

To define a left inverse to the assignment $\lambda \in \mathbb{R}_{>0}^{\tau_*}$ of lambda lengths, use Lemma 1 to uniquely lift the vertices $\pm 1, -\sqrt{-1}$ of the triangle of τ_* to the right of the doe to points in the rays in L^+ lying over these vertices realizing the lambda lengths. As in the proof of Theorem 4, we may then uniquely extend using Lemma 2 to a function $g: \mathbb{Q} \rightarrow L^+$ realizing the lambda lengths.

If $g(\mathbb{Q}) \subseteq L^+$ is radially dense, then the order-preserving mapping $\mathbb{Q} \rightarrow L^+ \rightarrow S^1$ interpolates a unique homeomorphism $f: S^1 \rightarrow S^1$ as before. One can always produce a discrete decoration, say by taking the point $f(p) \in L^+$ to have height i in \mathbb{R}^3 if $p \in \mathbb{Q}$ is of Farey generation i .

It follows that the mapping $\widetilde{\mathcal{T}ess} \rightarrow \mathbb{R}_{>0}^{\tau_*}$ is indeed injective. Continuity follows from the definition of the topologies, and openness of the image follows from the construction. □

Recall [1] that a *quasisymmetric* homeomorphism of S^1 is the restriction to S^1 of a quasiconformal homeomorphism of \mathbb{D} , and let $\text{Homeo}_{\text{qs}} \subseteq \text{Homeo}_+$ denote the subspace of all quasisymmetric homeomorphisms of the circle. *Bers’ universal Teichmüller space* [2] is the quotient

$$\text{Möb} \setminus \text{Homeo}_{\text{qs}} \subseteq \text{Möb} \setminus \text{Homeo}_+ \approx \mathcal{T}ess$$

and is highly studied.

As is usual in these circumstances, it is difficult to explicitly characterize the image $\widetilde{\mathcal{T}ess} \subseteq \mathbb{R}_{>0}^{\tau_*}$. On the other hand, there are the following useful characterizations of subsets of $\widetilde{\mathcal{T}ess} \subseteq \mathbb{R}_{>0}^{\tau_*}$.

We say that $\lambda \in \mathbb{R}_{>0}^{\tau_*}$ is *pinched* provided there is some real number $K > 1$ so that

$$K^{-1} \leq \lambda(e) \leq K \quad \text{for each } e \in \tau_*.$$

Theorem 7 ([8], Theorem 6.3). *If $\lambda \in \mathbb{R}_{>0}^{\tau_*}$ is pinched, then there is a decorated tessellation whose lambda lengths are given by λ , and the corresponding subset of L^+ is discrete and radially dense.*

Theorem 8 ([8], Theorem 6.4 (joint with Sullivan)). *If $\lambda \in \mathbb{R}_{>0}^{\mathbb{T}^*}$ is pinched, then the corresponding homeomorphism of the circle is quasisymmetric.*

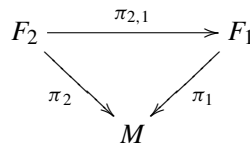
In particular, consider any decoration on a marked punctured Riemann surface F uniformized by $G < \text{Möb}$. Choose an ideal triangulation of \mathbb{D}/G , and lift it to a tessellation τ of \mathbb{D} which inherits a G -invariant decoration. Choose a doe of τ to determine a point of $\widetilde{\mathcal{T}ess}$. The lambda lengths in F lift to G -invariant lambda lengths on τ , and they are pinched since they take only finitely many values. Furthermore, if $G < \text{PSL}_2$ is finite-index and free of elliptics and $\phi: \mathbb{D} \rightarrow \mathbb{D}$ is any G -invariant quasiconformal map conjugating G to an isomorphic group, then we claim that the boundary values of ϕ satisfy the smoothness conditions above. To see this, conjugate in domain and range so that corresponding parabolic covering transformations are each given in upper halfspace by $z \mapsto z + 1$ (thus, destroying the normalization in $\widetilde{\mathcal{T}ess}$), so the monotone function $\phi(t)$ nearly agrees with the integral part of t . It follows directly that $\phi(t)$ is differentiable at each point of \mathbb{Q} , and the derivatives at points of \mathbb{Q} are uniformly near unity. (Compare with [13].)

In contrast, a quasisymmetric map $\phi: S^1 \rightarrow S^1$ arising from pinched lambda lengths need not have these differentiability properties at \mathbb{Q} . To see this, use the formula in the Introduction for lambda lengths in the upper halfspace model to produce pinched lambda lengths so that the two one-sided derivatives at infinity disagree.

This corrects the second part of Theorem 6.4 from [8]. For the corresponding result in the setting of the solenoid, see Proposition 12.

5 Coordinates for the solenoid

Let $G < \text{PSL}_2$ be any finite-index subgroup, and choose a base-point in the quotient surface or orbifold $M = \mathbb{D}/G$; in particular, for $G = \text{PSL}_2$, M is the orbifold modular curve. Consider the category \mathcal{C}_M of all finite-sheeted orbifold covers $\pi: F \rightarrow M$, where F is a punctured Riemann surface. \mathcal{C}_M is a directed set, where $\pi_1 \leq \pi_2$ if there is a finite-sheeted unbranched cover $\pi_{2,1}: F_2 \rightarrow F_1$ of Riemann surfaces so that the following diagram commutes:



In other words by covering space theory, if $\Gamma_i < G < \text{PSL}_2$ uniformizes F_i for $i = 1, 2$, then $\pi_1 \leq \pi_2$ if and only if Γ_1 is a finite-index subgroup of Γ_2 .

A topological space, the *punctured solenoid*, is defined in analogy to [12] to be the inverse limit

$$\mathcal{H}_M = \lim_{\leftarrow} \mathcal{C}_M;$$

a point of \mathcal{H}_M is thus determined by choices of points $y_i \in F_i$ for each cover $\pi_i: F_i \rightarrow M$, where the choices are “compatible” in the sense that if $\pi_1 \leq \pi_2$, then we have in the notation above $\pi_{2,1}(y_2) = y_1$.

Since punctured surface groups are cofinal in the set of punctured orbifold groups, we could have equivalently considered the category of orbifold covers of M in the definition of \mathcal{H}_M . Furthermore, if $\Gamma < G$ is of finite-index, then \mathcal{H}_Γ is naturally homeomorphic to \mathcal{H}_G , and we may thus think of *the* punctured solenoid $\mathcal{H} = \mathcal{H}_{\text{PSL}_2}$

One can from first principals develop the Teichmüller theory of \mathcal{H} along classical lines [10] as has been done for the solenoid of closed surfaces [12], [5], [11]. Instead, we next introduce an explicit space homeomorphic to \mathcal{H} following [5], and we shall then *define* the Teichmüller space representation theoretically in analogy to punctured surfaces in §1.

G has characteristic subgroups

$$G_N = \bigcap \{ \Gamma < G : [\Gamma : G] \leq N \},$$

for each $N \geq 1$, and these are nested $G_{N+1} < G_N$. Define a metric $G \times G \rightarrow \mathbb{R}$ by

$$\gamma \times \delta \mapsto \min \{ N^{-1} : \gamma \delta^{-1} \in G_N \},$$

and define the *pro-finite completion* \hat{G} of G as a space to be the metric completion of G , i.e., suitable equivalence classes of Cauchy sequences in G . Termwise multiplication of Cauchy sequences gives \hat{G} the structure of a topological group, and termwise multiplication by G gives a continuous action of G on \hat{G} .

For any sub-group $G < \text{PSL}_2$ of finite-index, we may define the quotient

$$\mathcal{H}_G = \mathbb{D} \times_G \hat{G} = (\mathbb{D} \times \hat{G}) / G,$$

where $\gamma \in G$ acts by

$$\begin{aligned} \gamma : \mathbb{D} \times \hat{G} &\rightarrow \mathbb{D} \times \hat{G} \\ (z, t) &\mapsto (\gamma z, t\gamma^{-1}). \end{aligned}$$

Lemma 9 ([5]). *\mathcal{H} is homeomorphic to \mathcal{H}_G for any $G < \text{PSL}_2$ of finite-index.*

In particular, each path component, or “leaf”, of \mathcal{H} is homeomorphic to a disk (by residual finiteness of G), and each leaf is dense in \mathcal{H} (since G is dense in \hat{G}).

Let us for definiteness simply fix $G = \text{PSL}_2$ and consider the collection $\text{Hom}'(G \times \hat{G}, \text{Möb})$ of all functions $\rho: G \times \hat{G} \rightarrow \text{Möb}$ satisfying the following three properties.

Property 1. ρ is continuous.

Property 2. 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 $\phi_t : \mathbb{D} \rightarrow \mathbb{D}$ so that for every $\gamma \in G$, the following diagram commutes:

$$\begin{array}{ccc}
 \mathbb{D} \times \hat{G} & \xrightarrow{(z,t) \mapsto (\gamma z, t\gamma^{-1})} & \mathbb{D} \times \hat{G} \\
 \phi_t \times \text{id} \downarrow & & \downarrow \phi_{t\gamma^{-1}} \times \text{id} \\
 \mathbb{D} \times \hat{G} & \xrightarrow{(\phi_t(z), t) \mapsto (\phi_{t\gamma^{-1}} \circ \gamma(z), t\gamma^{-1})} & \mathbb{D} \times \hat{G}
 \end{array}$$

Furthermore, ϕ_t varies continuously in $t \in \hat{G}$ for the common refinement of the C^∞ topology of uniform convergence on compacta in \mathbb{D} and the usual topology on Bers' universal Teichmüller space $\text{Möb} \setminus \text{Homeo}_{\text{qs}}$ of the extension of ϕ_t to the circle at infinity.

As to Property 1, notice that since G is discrete, ρ is continuous if and only if it is so in its second variable only. Property 2 is a kind of homomorphism property of ρ mixing the leaves; notice in particular that taking $\gamma_2 = I$ shows that $\rho(I, t) = I$ for all $t \in \hat{G}$. Property 3 mandates that for each $t \in \hat{G}$, ϕ_t conjugates the standard action of 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 $S^1 \times \hat{G}$. We finally define the solenoid (with marked hyperbolic structure)

$$\mathcal{H}_\rho = (\mathbb{D} \times_\rho \hat{G}) = (\mathbb{D} \times \hat{G}) / G_\rho.$$

The collection ϕ_t , for $t \in \hat{G}$, thus induces a homeomorphism $\mathcal{H} \rightarrow \mathcal{H}_\rho$.

Define the group $\text{Cont}(\hat{G}, \text{Möb})$ to be the collection of all continuous maps $\alpha : \hat{G} \rightarrow \text{Möb}$, where the product of two $\alpha, \beta \in \text{Cont}(\hat{G}, \text{Möb})$ is taken pointwise $(\alpha\beta)(t) = \alpha(t) \circ \beta(t)$ in Möb . $\alpha \in \text{Cont}(\hat{G}, \text{Möb})$ acts continuously on $\rho \in \text{Hom}'(G \times \hat{G}, \text{Möb})$ according to

$$(\alpha\rho)(\gamma, t) = \alpha^{-1}(t\gamma^{-1}) \circ \rho(\gamma, t) \circ \alpha(t).$$

Theorem 10 ([10]). *There is a natural homeomorphism of the Teichmüller space of the solenoid \mathcal{H} with*

$$\mathcal{T}(\mathcal{H}) = \text{Hom}'(G \times \hat{G}, \text{Möb}) / \text{Cont}(\hat{G}, \text{Möb}).$$

Rather than describe the proof here, we shall for simplicity simply take this quotient as the definition of the Teichmüller space $\mathcal{T}(\mathcal{H})$. Again with an eye towards simplicity here, rather than define punctures of solenoids intrinsically (as suitable equivalence classes of ends of escaping rays), we can more simply proceed as follows. Each $\phi_t : \mathbb{D} \rightarrow \mathbb{D}$ extends continuously to a quasisymmetric mapping $\phi_t : S^1 \rightarrow S^1$. We

say that a point $(p, t) \in S^1 \times \hat{G}$ is a ρ -puncture if $\phi_t^{-1}(p) \in \mathbb{Q}$, and a puncture of \mathcal{H}_ρ itself is a G_ρ -orbit of ρ -punctures. A ρ -horocycle at a ρ -puncture (p, t) is the image under ϕ_t of a horocycle in \mathbb{D} centered at $\phi_t^{-1}(p)$.

A *decoration* on \mathcal{H}_ρ , or a “decorated hyperbolic structure” on \mathcal{H} , is a function $\tilde{\rho}: G \times \hat{G} \times \mathbb{Q} \rightarrow \text{Möb} \times L^+$, where

$$\tilde{\rho}(\gamma, t, q) = \rho(\gamma, t) \times h(t, q)$$

with $\rho(\gamma, t) \in \text{Hom}'(G \times \hat{G}, \text{Möb})$, which satisfies the following conditions.

Property 4. For each $t \in \hat{G}$, the image $h(t, \mathbb{Q}) \subseteq L^+$ is discrete and radially dense.

Property 5. For each $q \in \mathbb{Q}$, the restriction $h(\cdot, q): \hat{G} \rightarrow L^+$ is continuous.

Property 6. $\tilde{\rho}$ is G -invariant in the sense that if $\delta \in G$, then

$$\delta \circ \tilde{\rho}(\gamma, t, q) = \tilde{\rho}(\delta\gamma, t\delta^{-1}, \delta q),$$

where δ acts diagonally $\delta: (\gamma, q) \mapsto (\delta\gamma, \delta q)$ on $\text{Möb} \times L^+$ with δq the natural action of $G = \text{PSL}_2$ on L^+ .

Finally, let $\text{Hom}'(G \times \hat{G} \times \mathbb{Q}, \text{Möb} \times L^+)$ denote the space of all decorated hyperbolic structures satisfying the properties above, and define the *decorated Teichmüller space* as the quotient

$$\tilde{\mathcal{T}}(\mathcal{H}) = \text{Hom}'(G \times \hat{G} \times \mathbb{Q}, \text{Möb} \times L^+) / \text{Cont}(\hat{G}, \text{Möb}),$$

where $\alpha: \hat{G} \rightarrow \text{Möb}$ acts on $\tilde{\rho}$ by

$$(\alpha\tilde{\rho})(\gamma, t, q) = (\alpha^{-1}(t\gamma^{-1}) \circ \rho(\gamma, t) \circ \alpha(t)) \times (h(t, \alpha(t)q)).$$

It is clear that forgetting decoration induces a continuous surjection $\tilde{\mathcal{T}}(\mathcal{H}) \rightarrow \mathcal{T}(\mathcal{H})$.

There is a natural mapping $\lambda: \tilde{\mathcal{T}}(\mathcal{H}) \rightarrow (\mathbb{R}_{>0}^{\tau_*})^{\hat{G}}$ which assigns to a function $\tilde{\rho}: G \times \hat{G} \times \mathbb{Q} \rightarrow \text{Möb} \times L^+$ the lambda length for the G_ρ metric of the ρ -horocycles determined by h at the endpoints of the geodesic in \mathcal{H}_ρ labeled by γ .

Theorem 11 ([10]). *The assignment of lambda lengths determines an embedding*

$$\tilde{\mathcal{T}}(\mathcal{H}) \rightarrow \text{Cont}(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$$

onto an open set, where we take the strong topology on $\mathbb{R}_{>0}^{\tau_}$ and on $\text{Cont}(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$.*

Sketch of Proof. To prove the mapping is injective, we must again define the construction of decorated hyperbolic structure from a continuous $\lambda_t \in (\mathbb{R}_{>0}^{\tau_*})^{\hat{G}}$. To this end, begin the definition of $\tilde{\rho} = \rho \times h$ on the triangle to the right of the doe in τ_* with lambda lengths given by λ_t . As usual according to Lemma 1, we can uniquely lift to a triple of points in L^+ lying over $\pm 1, -\sqrt{-1}$.

It is easily seen from the identification of $G = \text{PSL}_2$ with the oriented edges of τ_* that any $\gamma \in \text{PSL}_2$ can be written uniquely in the one of the following forms:

- i) $\gamma = I$;

- ii) γ lies in the free semi-group generated by either U, T or U^{-1}, T^{-1} ;
- iii) γ arises from either i) or ii) by addition of the prefix S .

To define $\rho(\gamma, t) \in \text{Möb}$, we shall specify an ideal triangle-doe pair (Δ', e') in \mathbb{D} , where e' is oriented with Δ' to its right. There is then a unique $\rho \in \text{Möb}$ mapping to the vertices of Δ' the vertices $\pm 1, -\sqrt{-1}$ of the triangle Δ to the right of the doe e_I in τ_* and mapping e_I to e' .

Let us write $\gamma \in G$ in one of the forms i-iii) above. Of course, if $\gamma = I$, then $\rho(\gamma, t) = I$ as follows from the functional equation in Property 2, and we take $(\Delta', e') = (\Delta, e_I)$.

If $\gamma = S$, then let us begin with the lambda lengths $\lambda_t \in \mathbb{R}_{>0}^{\tau_*}$ on the edges of Δ and employ Lemma 1 to uniquely realize a lift to L^+ of the vertices of this decoration on the triangle Δ . On the triangle to the left of the doe, consider the lambda lengths $\lambda_{tS^{-1}} = \lambda_{tS}$. It need not be that $\lambda_t(e_I) = \lambda_{tS}(e_I)$, and we re-scale, taking lambda lengths

$$\frac{\lambda_t(e_I)}{\lambda_{tS}(e_I)} \lambda_{tS}(\cdot)$$

on the edges e_U, e_T . This defines lambda lengths on the edges of the quadrilateral in τ_* triangulated by the doe. Again using Lemma 2, there is a unique lift of $\sqrt{-1}$ to L^+ realizing the lambda lengths, and the projection of this point to S^1 is one of the vertices of Δ' . The other two vertices of Δ' are ± 1 , and the doe is $e' = e_S$, completing the definition in this case. Notice that this element of Möb that maps $(\Delta, e_I) \rightarrow (\Delta', e_S)$ is necessarily involutive.

The case of any word in one of the semi-groups in ii) is handled by induction on the length, where for instance for any such γ that has a prefix U , one begins from the lambda lengths λ_t on Δ and uses the re-scaled lambda lengths $\lambda_{tU^{-1}}$ on the edges e_U, e_T ; the edge corresponding to e_U is the doe of the first step.

The remaining case iii) of a word from one of the semi-groups with prefix S is handled in exactly the same manner completing the definition of the construction of $\rho: G \times \hat{G} \rightarrow \text{Möb}$. The functional equation in Property 2 on ρ follows by definition. Furthermore, since $\lambda_t \in \mathbb{R}_{>0}^{\tau_*}$ depends continuously on t (because of the definition of the topology), $\rho(\gamma, t)$ is also continuous in t as required in Property 1; indeed, the entries of $\rho(\gamma, t) \in \text{Möb}$ are algebraic function of finitely many lambda lengths.

As for Property 3 in the definition of $\mathcal{T}(\mathcal{H})$, we claim that for each $t \in \hat{G}, \lambda_t \in \mathbb{R}_{>0}^{\tau_*}$ is pinched if $\lambda \in \text{Cont}(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$.

To see this, note that the very definition of a continuous function $\lambda: \hat{G} \rightarrow \mathbb{R}_{>0}^{\tau_*}$ means that for all K there exists an N such that we have

$$1 + K^{-1} \leq \frac{\lambda_t(e)}{\lambda_t(\gamma e)} \leq 1 + K \quad \text{for all } e \in \tau_*, \gamma \in G_N.$$

Take say $K = \frac{1}{2}$ and its corresponding N . A fundamental domain for G_N has only a finite collection of values of lambda lengths, and any other lambda length is at most

a factor of 3/2 times a lambda length this finite set, and at least a factor of 1/2 times a lambda length in this finite set. λ_t is therefore pinched, proving the claim.

It follows that for each $t \in \hat{G}$, $\lambda_t: \tau_* \rightarrow \mathbb{R}_{>0}$ is necessarily pinched. By Theorem 8, there is a corresponding quasiconformal homeomorphism $\phi_t: \mathbb{D} \rightarrow \mathbb{D}$. Commutativity of the diagram and continuity in Property 3 follow by construction, and this completes the proof that the function ρ constructed above lies in $\text{Hom}'(G \times \hat{G}, \text{Möb})$ and hence determines a point of $\mathcal{T}(\mathcal{H})$.

To define a decoration on the ρ -punctures, each λ_t determines a decoration on $\tau_* \times \{t\} \subseteq \mathbb{D} \times \hat{G}$, as required. Property 4 is guaranteed by the claim and Theorem 7. Property 5 holds as before in the strong sense that the Euclidean coordinates of each $h(t, q)$ are algebraic functions of finitely many lambda lengths, and Property 6 holds by invariance of lambda lengths under Möb. \square

Proposition 12. *Suppose that $\lambda \in \text{Cont}(\hat{G}, \mathbb{R}_{>0}^{\tau_*})$. Then for each $t \in \hat{G}$, $\lambda_t: \tau_* \rightarrow \mathbb{R}_{>0}$ corresponds to a quasiconformal homeomorphism $\phi_t: \mathbb{D} \rightarrow \mathbb{D}$ whose quasisymmetric extension $\phi_t: S^1 \rightarrow S^1$ is differentiable at each point of \mathbb{Q} with derivative uniformly near unity.*

Proof. As above, continuity of λ_t in t implies that each λ_t is pinched, which gives a quasisymmetric map ϕ_t , for $t \in \hat{G}$, by Theorem 8. Using the upper halfspace model, normalize ϕ_t such that it fixes 0 and ∞ , whence the geodesics of τ_* that limit to ∞ are mapped by ϕ_t onto geodesics which likewise limit to ∞ . Again by continuity, we conclude that for each $\varepsilon > 0$, $e \in \tau_*$ and $\gamma \in \text{PSL}_2$ fixing ∞ , we have $|\lambda_t(e) - \lambda_t(\gamma^n e)| < \varepsilon$ for n sufficiently large. The differences $a_n = \phi_t(n) - \phi_t(n-1)$ are then ε_1 close using continuity of the assignment of decorated ideal triangles given lambda lengths.

We show that $\lim_{n \rightarrow \infty} \frac{1}{n} \phi_t(n)$ exists and is bounded, which proves the proposition. To this end since $|a_i - a_{nk+i}| < \varepsilon_1$ for all i, k , we find

$$\begin{aligned} & \left| \frac{1}{n}(a_1 + \cdots + a_n) - \frac{1}{nk}(a_1 + \cdots + a_{nk}) \right| \\ & \leq 1/n \sum_{i=1}^n |a_i - \frac{1}{k}(a_i + a_{i+n} + \cdots + a_{i+n(k-1)})| \leq \varepsilon_1, \end{aligned}$$

and it follows that $\frac{1}{n} \phi_t(n)$ is a Cauchy sequence, as desired. \square

Differentiability at the rational points, which holds in the case of punctured surfaces thus also holds for the solenoid according to Proposition 12 but does *not* hold for general decorated tessellations with pinched lambda lengths however (cf. the discussion following Theorem 8).

6 Concluding remarks

In addition to the function $h: \hat{G} \times \mathbb{Q} \rightarrow L^+$ constructed in the proof of Theorem 11, there is another natural function $h_1: \hat{G} \times \mathbb{Q} \rightarrow L^+$ defined as follows. Begin with the “standard” decoration on τ_* where all lambda lengths are $\sqrt{2}$. By Proposition 12, each ϕ_t is smooth with bounded derivative $\phi'_t(q)$ at each point of $\mathbb{Q} \subseteq S^1$. In an upper halfspace model of \mathbb{H} with the endpoint of the doe at infinity, scale the Euclidean diameter of the horocycle centered at $q \in \mathbb{Q}$ by the absolute value of the derivative $|\phi'_t(q)|$ to determine the diameter of the horocycle at $\phi_t(q)$. This defines the function $h_1(t, q)$. Notice that $h_1(t, \mathbb{Q})$ is again discrete and radially dense in L^+ , but there seems to be no guarantee that the function h_1 discussed above satisfies Property 5.

The representation-theoretic treatment of the Teichmüller theory of the punctured solenoid seems to us quite appealing. For example, the strong topology for the solenoid (in Theorem 11) in contrast to the weak topology for circle homeomorphisms (in Theorem 6) is interesting. Furthermore, one can naturally induce stronger or weaker transverse structures in the \hat{G} direction in \mathcal{H} by imposing conditions other than continuity on the lambda length functions in Theorem 11, and we wonder in particular what is the transverse regularity of h_1 .

As was mentioned in the Introduction, Theorem 11 is the first step of an ongoing program [10] to develop the decorated Teichmüller theory of the punctured solenoid. Though there is an alternative to the construction of $\rho: G \times \hat{G} \rightarrow \text{Möb}$ in the proof of Theorem 11, the argument given here bears a close relation to the treatment of broken hyperbolic structures in [6].

Lambda lengths enjoy the simple transformation property of Ptolemy transformations (cf. Remark 1), and furthermore, there is a simple invariant two-form (cf. Remark 2). These ingredients have been used [3], [4] to give a quantization of classical Teichmüller theory (cf. Remark 3). These same two ingredients persist for the universal Teichmüller theory as well as for the punctured solenoid and might be used to quantize these Teichmüller theories as well.

It is an open (but not centrally important) problem in number theory to calculate the index of G_N in $G = \text{PSL}_2$, and an algorithm for its calculation devolves to the “cubic fatgraph” enumeration problem” for surfaces of fixed Euler characteristic arising as total spaces of degree- N covers of \mathbb{D}/G . More speculatively, there is a natural group [10] generated by the G_N -equivariant Ptolemy moves for some N , which seems closely related to the completions [9] of the universal Ptolemy group studied in the context of Grothendieck absolute Galois theory.

References

- [1] L. V. Ahlfors, *Lectures on quasiconformal mappings*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1987; reprint of the 1966 original edition.
- [2] L. Bers, Universal Teichmüller space. In *Analytic methods in mathematical physics* (Sympos., Indiana Univ., Bloomington, Ind., 1968), Gordon and Breach, New York 1970, 65–83.

- [3] V. V. Fock and L. O. Chekhov, A quantum Teichmüller space. *Teoret. Mat. Fiz.* 120 (1999), 511–528; English transl. *Theoret. and Math. Phys.* 120 (1999), 1245–1259. Quantum mapping class group, pentagon relation, and geodesics. *Tr. Mat. Inst. Steklova* 226 (1999), 163–179; English transl. *Proc. Steklov Math. Inst.* 226 (1999), 149–163.
- [4] R. M. Kashaev, Quantization of Teichmüller spaces and the quantum dilogarithm. *Lett. Math. Phys.* 43 (2) (1998), 105–115.
- [5] C. Odden, Virtual automorphism group of the fundamental group of a closed surface. Ph.D. Thesis, Duke University, 1997; The baseleaf preserving mapping class group of the universal hyperbolic solenoid. *Trans. Amer. Math. Soc.* 357 (2004), 1829–1858.
- [6] A. Papadopoulos and R. C. Penner, The Weil–Petersson Kähler form and affine foliations on surfaces. *Ann. Global Anal. Geom.* 27 (2005), 53–77.
- [7] R. C. Penner, The decorated Teichmüller space of punctured surfaces. *Comm. Math. Phys.* 113 (1987), 299–339.
- [8] R. C. Penner, Universal constructions in Teichmüller theory. *Adv. Math.* 98 (1993), 143–215.
- [9] R. C. Penner, The universal Ptolemy group and its completions. In *Geometric galois actions II* (edited by L. Schneps and P. Lochak), London Math. Soc. Lecture Note Ser. 243, Cambridge University Press, Cambridge 1997, 293–312.
- [10] R. C. Penner and D. Šarić, Teichmüller theory of the punctured solenoid. Preprint, 2005; arXiv.math.DS/0508476.
- [11] D. Šarić, On quasiconformal deformations of the universal hyperbolic solenoid. Preprint, 2003; www.math.sunysb.edu/~saric/research/solenoid2.pdf.
- [12] D. Šarić, The Teichmüller theory of the solenoid. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2008. 206
- [13] D. Sullivan, Linking the universalities of Milnor–Thurston, Feigenbaum and Ahlfors–Bers. In *Topological methods in modern mathematics* (Stony Brook, NY, 1991), edited by L. R. Goldberg and A. V. Phillips, Publish or Perish, Inc., Houston, TX, 1993, 543–564.
- [14] P. Tukia, Differentiability and rigidity of Möbius groups. *Invent. Math.* 82 (1985), 557–578.

Chapter 4

About the embedding of Teichmüller space in the space of geodesic Hölder distributions

Jean-Pierre Otal

Contents

1	Introduction	223
2	The Teichmüller space of a Fuchsian group	224
	2.1 The quasiconformal homeomorphisms	224
	2.2 The complex structure on Teichmüller space	226
3	Geodesic currents and Hölder distributions	229
	3.1 The space of geodesics of \mathbb{H}^2	230
	3.2 Geodesic currents	230
	3.3 The Liouville measure	231
	3.4 Geodesic Hölder distributions for a cocompact Fuchsian group	232
	3.5 Geodesic Hölder distributions for a general Fuchsian group	233
4	Analytic properties of the Liouville map	237
	4.1 Hölder regularity of holomorphic motions	239
	4.2 Continuity properties of derivatives of Hölder functions	241
	4.3 Proof of Theorems 4.3, 4.1 and 4.2	245
	References	248

1 Introduction

Thanks to a construction due to F. Bonahon, the Teichmüller space $\mathcal{T}(\Gamma)$ of a cofinite volume Fuchsian group Γ can be viewed as a subspace of a Banach space: indeed the Liouville map \mathcal{L} is an embedding of $\mathcal{T}(\Gamma)$ into the space of geodesic currents of Γ . This construction allowed Bonahon to recover one important result of Thurston, namely, that the Teichmüller space $\mathcal{T}(\Gamma)$ can be compactified by the space of projective measured laminations. This embedding however is only topological and the space of geodesic currents needs to be enlarged in order for the Liouville map to become differentiable. This larger space is the (Fréchet) space of geodesic Hölder distributions $\mathcal{H}(\Gamma)$; it was introduced by Bonahon, who showed with Sözen, that the Liouville map $\mathcal{L}: \mathcal{T}(\Gamma) \rightarrow \mathcal{H}(\Gamma)$, is differentiable along paths. This theorem was recently generalized by D. Šarić to the setting of an arbitrary Fuchsian group (with

maybe infinite covolume). Our main result in this chapter is that the Liouville map $\mathcal{L}: \mathcal{T}(\Gamma) \rightarrow \mathcal{H}(\Gamma)$ is analytic; the analytic structure on $\mathcal{T}(\Gamma)$ is the one induced by the complex structure, the analytic structure on $\mathcal{H}(\Gamma)$ is the one induced from the linear (Fréchet) structure.

In the first section, we review the construction due to L. Bers of the complex structure on $\mathcal{T}(\Gamma)$. In Section 3, we recall the definition of the space of Hölder distributions on the space of geodesics of \mathbb{H}^2 , introduced by F. Bonahon for cocompact Fuchsian groups and its generalization to arbitrary Fuchsian groups by D. Šarić. We show that \mathcal{L} is analytic in the last section.

Acknowledgement. Part of this work was written at the Bernoulli Center in Lausanne during the Workshop “Spaces of Negative Curvature”.

2 The Teichmüller space of a Fuchsian group

The group $\mathrm{PSL}(2, \mathbb{C})$ acts by homographies on the Riemann sphere $\overline{\mathbb{C}}$. Any discrete subgroup $G \subset \mathrm{PSL}(2, \mathbb{C})$ leaves invariant a smallest non-empty closed subset $L(G) \subset \overline{\mathbb{C}}$ called the *limit set of G* .

We denote by Γ a *Fuchsian group*, i.e. a discrete and torsion-free subgroup of $\mathrm{PSL}(2, \mathbb{R})$.

The most efficient way to define the Teichmüller space $\mathcal{T}(\Gamma)$ is within the theory of quasiconformal homeomorphisms.

2.1 The quasiconformal homeomorphisms

Definition 2.1. A *quasiconformal homeomorphism* f of the Riemann sphere $\overline{\mathbb{C}}$ is an orientation-preserving homeomorphism such that

- (1) the distributional derivatives $\frac{\partial f}{\partial z}$ and $\frac{\partial f}{\partial \bar{z}}$ are measurable functions which are locally square integrable, and
- (2) for some measurable function $\mu \in L^\infty(\mathbb{C})$ with $\|\mu\|_\infty < 1$, one has, for almost all $z \in \mathbb{C}$,

$$\frac{\partial f}{\partial \bar{z}} = \mu \frac{\partial f}{\partial z}. \quad (2.1)$$

The function $\mu \in L^\infty(U)$ with this property is called the *Beltrami coefficient of f* . The *dilatation of f* is the quantity

$$K(f) = \frac{1 + \|\mu\|_\infty}{1 - \|\mu\|_\infty}.$$

Equation 2.1 is called the *Beltrami equation*.

The first main property of quasiconformal homeomorphisms is the following compactness property [1].

Proposition 2.2. *Let $K > 0$. The set of K -quasiconformal homeomorphisms of $\overline{\mathbb{C}}$ which fix 3 distinct points is compact for the topology of uniform convergence on $\overline{\mathbb{C}}$.*

Another main property of quasiconformal homeomorphisms is that they are tractable by analytic tools. The importance of the following theorem for the study of Teichmüller space was realized by L. Ahlfors and L. Bers (see the commentary of Ahlfors in [2, p. 226]).

Theorem 2.3 (Existence of solutions to the Beltrami equation [3], [14], [13], [15]). *Let $\mu \in L^\infty(\mathbb{C})$, with $\|\mu\|_\infty < 1$. There is a unique quasiconformal homeomorphism $f^\mu : \mathbb{C} \rightarrow \mathbb{C}$ such that*

- (i) $f^\mu(0) = 0$ and $f^\mu(1) = 1$;
- (ii) the Beltrami coefficient of f^μ is equal to μ .

The uniqueness part of Theorem 2.3 and Proposition 2.2 have the following consequence.

Corollary 2.4. *Let $k < 1$. Let (μ_i) be a sequence in $L^\infty(\mathbb{C})$ such that $\|\mu_i\|_\infty \leq k < 1$ for all i and such that (μ_i) tends to μ in $L^\infty(\mathbb{C})$. Then the sequence (f^{μ_i}) converges to f^μ uniformly over $\overline{\mathbb{C}}$*

Definition 2.5 (The Teichmüller space and the quasi-Fuchsian space). *A quasi-Fuchsian deformation of Γ (resp. a Fuchsian deformation of Γ) is a pair (ρ, f) where ρ is a representation of Γ into $\text{PSL}(2, \mathbb{C})$ (resp. into $\text{PSL}(2, \mathbb{R})$) and f is a quasiconformal homeomorphism of \mathbb{C} (resp. a quasiconformal homeomorphism of \mathbb{C} leaving \mathbb{R} invariant) normalized by $f(0) = 0$ and $f(1) = 1$ and which conjugates Γ and $\rho(\Gamma)$:*

$$\rho(\gamma) = f \circ \gamma \circ f^{-1} \quad \text{for all } \gamma \in \Gamma.$$

One defines an equivalence relation \simeq on Fuchsian or quasi-Fuchsian deformations by setting $(\rho_1, f_1) \simeq (\rho_2, f_2)$ if and only if $f_1|_{\mathbb{R}} = f_2|_{\mathbb{R}}$. When the limit set of Γ is all $\overline{\mathbb{R}}$, one has $(\rho_1, f_1) \simeq (\rho_2, f_2)$ if and only if $\rho_1 = \rho_2$.

The Teichmüller space $\mathcal{T}(\Gamma)$ (resp. the quasi-Fuchsian space $\mathcal{QF}(\Gamma)$) is the quotient of the space of Fuchsian deformations of Γ (resp. the space of quasi-Fuchsian deformations of Γ) by the equivalence relation \simeq (cf. [11], [12], [13]). The spaces $\mathcal{T}(\Gamma)$ and $\mathcal{QF}(\Gamma)$ are complete metric spaces for the distance which is the quotient distance of a distance on the space of Fuchsian or quasi-Fuchsian deformations, namely

$$d((\rho_1, f_1), (\rho_2, f_2)) = \frac{1}{2} \log K(f_2 \circ f_1^{-1}).$$

The theorem of existence of solutions to the Beltrami equation allows to parameterize $\mathcal{T}(\Gamma)$ and $\mathcal{QF}(\Gamma)$.

Definition 2.6 (The Beltrami differentials). A *Beltrami differential* for Γ is a measurable function $\mu \in L^\infty(\mathbb{C}, \mathbb{C})$ such that for all $\gamma \in \Gamma$ and for almost all $z \in \mathbb{C}$,

$$\mu(\gamma z) = \frac{\gamma'(z)}{\gamma'(z)} \mu(z).$$

The space of Beltrami differentials for Γ is denoted by $B(\Gamma)$. With the L^∞ -norm, it is a Banach space; we denote by $B^1(\Gamma)$ its open unit ball.

Let $\mu \in B^1(\Gamma)$, let $f^\mu : \mathbb{C} \rightarrow \mathbb{C}$ be the quasiconformal homeomorphism provided by Theorem 2.3. The transformation rule of μ under the action of Γ implies that the Beltrami coefficient of the quasiconformal homeomorphism $f^\mu \circ \gamma \circ (f^\mu)^{-1}$ is 0; therefore $\rho^\mu(\gamma) = f^\mu \circ \gamma \circ (f^\mu)^{-1} \in \text{PSL}(2, \mathbb{C})$ and the map $\gamma \mapsto \rho^\mu(\gamma)$ is a representation of Γ into $\text{PSL}(2, \mathbb{C})$. Thus $\mu \mapsto (\rho^\mu, f^\mu)$ defines a map from $B^1(\Gamma)$ to the space of quasi-Fuchsian deformations of Γ . This map is a homeomorphism. After post-composing this map with the equivalence relation \simeq which defines $\mathcal{QF}(\Gamma)$ as a quotient of the space of quasi-Fuchsian deformations, we obtain a map denoted by $\mu \mapsto [\mu]$ from $B^1(\Gamma)$ onto $\mathcal{QF}(\Gamma)$; so the notation $[\mu]$ represents the equivalence class of (ρ^μ, f^μ) .

For $\mu \in B^1(\Gamma)$, the quasiconformal homeomorphism f^μ leaves \mathbb{R} invariant if and only if the Beltrami differential $\mu \in B^1(\Gamma)$ commutes with the complex involution, i.e. if and only if $\mu(z) = \overline{\mu(\bar{z})}$ for almost all z . We shall denote by $B_{\mathcal{T}}(\Gamma) \subset B(\Gamma)$ the subspace of Beltrami differentials μ with this symmetry property. Then $\mu \mapsto [\mu]$ maps $B_{\mathcal{T}}^1(\Gamma)$ onto $\mathcal{T}(\Gamma)$.

Definition 2.7 (The change of basepoint map). Let $\mu_0 \in B_{\mathcal{T}}^1(\Gamma)$; denote by Γ^{μ_0} the Fuchsian group $\rho^{\mu_0}(\Gamma)$. The Teichmüller spaces $\mathcal{T}(\Gamma)$ and $\mathcal{T}(\Gamma^{\mu_0})$ are homeomorphic: the map $(\rho, f) \mapsto (\rho \circ (\rho^{\mu_0})^{-1}, f \circ (f^{\mu_0})^{-1})$ at the level of (quasi-)Fuchsian deformations induces a homeomorphism from $\mathcal{T}(\Gamma)$ to $\mathcal{T}(\Gamma^{\mu_0})$ which sends $[\mu]$ to $[0]$. This homeomorphism is the *change of basepoint map*. It is induced by a homeomorphism $\nu : B^1(\Gamma) \rightarrow B^1(\Gamma^{\mu_0})$,

$$\mu \mapsto \nu(\mu) \left(\frac{\mu - \mu_0}{1 - \bar{\mu}_0 \mu} \frac{f_z^{\mu_0}}{f_z^{\mu_0}} \right) \circ (f^{\mu_0})^{-1}.$$

The homeomorphism ν restricts to a homeomorphism between $B_{\mathcal{T}}^1(\Gamma)$ and $B_{\mathcal{T}}^1(\Gamma^{\mu_0})$.

2.2 The complex structure on Teichmüller space

Definition 2.8 (Holomorphic family of Beltrami differentials). A *holomorphic family of Beltrami differentials parametrized by the unit disc* $\mathbb{D} \subset \mathbb{C}$ is a map $w \rightarrow \mu_w$ from \mathbb{D} to $L^\infty(\mathbb{C})$ which satisfies

- (i) there exists a $k < 1$ such that $\|\mu_w\| < k$ for all $w \in \mathbb{D}$, and
- (ii) for almost all $z \in \mathbb{C}$, the map $w \mapsto \mu_w(z)$ is holomorphic.

The following theorem of Ahlfors and Bers [3] is the basic tool for the construction of a complex structure on $\mathcal{T}(\Gamma)$.

Theorem 2.9 (Holomorphic dependence of solutions to the Beltrami equation). *Let $w \rightarrow \mu_w \in L^\infty(\mathbb{C})$ be a holomorphic family of Beltrami differentials parametrized by the unit disc $\mathbb{D} \subset \mathbb{C}$. Then, for any $z \in \mathbb{C}$ the map $w \mapsto f^{\mu_w}(z)$ is holomorphic.*

The existence of a complex structure on $\mathcal{T}(\Gamma)$ was established by L. Bers [5]. We explain now this construction. Denote by \mathbb{L} the lower half-plane $\{z \in \mathbb{C} \mid \Im z < 0\}$.

Definition 2.10 (Holomorphic quadratic differentials). *A holomorphic quadratic differential for Γ is (here) a holomorphic function $\phi: \mathbb{L} \rightarrow \mathbb{C}$ such that for all $\gamma \in \Gamma$, $\phi(\gamma(z))(\gamma'(z))^2 = \phi(z)$. For such a ϕ , $z \rightarrow y^2|\phi(z)|$ defines a Γ -invariant function on \mathbb{L} . We say that the holomorphic quadratic differential ϕ is *bounded* when the norm*

$$\|\phi\|_b = \sup_{z \in \mathbb{H}} \{y^2|\phi(z)|\}$$

is finite. The space of bounded holomorphic quadratic differentials for Γ with the norm $\|\cdot\|_b$ is a Banach space denoted by $\mathcal{Q}_b(\Gamma)$. When the Riemann surface \mathbb{H}/Γ is isomorphic to a compact surface of genus g with n points deleted, then $\mathcal{Q}_b(\Gamma)$ is a finite dimensional space of complex dimension $3g - 3 + n$.

The existence of the complex structure on $\mathcal{T}(\Gamma)$ is based on the following construction. For $\mu \in B^1_{\mathcal{T}}(\Gamma)$, let $\tilde{\mu}$ be the differential in $B^1(\Gamma)$ such that $\tilde{\mu}(z) = \mu(z)$ for $z \in \mathbb{H}$ and $\mu(z) = 0$ for $z \in \mathbb{L}$. Then, $f^{\tilde{\mu}}$ is holomorphic when restricted to \mathbb{L} : it is a univalent map from \mathbb{L} to \mathbb{C} . A basic invariant of a (locally) univalent map f is its Schwarzian derivative

$$\mathcal{S}(f) = \frac{f'''}{f'} - \frac{3}{2} \left(\frac{f''}{f'} \right)^2.$$

Since $f^{\tilde{\mu}}$ conjugates the action of Γ to the action of a subgroup of $\text{PSL}(2, \mathbb{C})$, the Schwarzian $\mathcal{S}(f^{\tilde{\mu}})$ transforms like a holomorphic quadratic differential for Γ . Since $f^{\tilde{\mu}}$ is globally univalent, the Nehari–Kraus Lemma says that $\mathcal{S}(f^{\tilde{\mu}})$ is a bounded holomorphic quadratic differential and that it belongs to the ball of radius $3/2$ of $\mathcal{Q}_b(\Gamma)$. Thus $\mu \mapsto \mathcal{S}(f^{\tilde{\mu}})$ is a map from $B^1_{\mathcal{T}}(\Gamma)$ to $\mathcal{Q}_b(\Gamma)$ with image contained in the ball of radius $3/2$.

One uses now the following observation (cf. [11, p. 98], [12, p. 133]).

Claim 2.11. *Let μ_1 and $\mu_2 \in B^1(\Gamma)$. Then $\mu_1 \simeq \mu_2$ if and only if $\tilde{\mu}_1 \simeq \tilde{\mu}_2$.*

Clearly $\tilde{\mu}_1 \simeq \tilde{\mu}_2$ is equivalent to say that the two holomorphic maps $f^{\tilde{\mu}_1}|_{\mathbb{L}}$ and $f^{\tilde{\mu}_2}|_{\mathbb{L}}$ are equal. Therefore, by the direct part of Claim 2.11, if the two differentials μ_1 and μ_2 are equivalent, then the Schwarzian $\mathcal{S}(f^{\tilde{\mu}_1})$ and $\mathcal{S}(f^{\tilde{\mu}_2})$ are equal. Thus the map $\mu \mapsto \mathcal{S}(f^{\tilde{\mu}})$ factorizes through a map Φ from $\mathcal{T}(\Gamma)$ to $\mathcal{Q}_b(\Gamma)$: $\mathcal{S}(f^{\tilde{\mu}}) = \Phi([\mu])$. This map Φ is injective. To see that, recall that if two univalent maps have the same

Schwarzian, they differ by post-composition with an element of $\text{PSL}(2, \mathbb{C})$. Therefore, since $f^{\tilde{\mu}_1}$ and $f^{\tilde{\mu}_2}$ fix 0, 1 and ∞ , $\mathcal{S}(f^{\tilde{\mu}_1}) = \mathcal{S}(f^{\tilde{\mu}_2})$ implies that the restrictions $f^{\tilde{\mu}_1}|_{\mathbb{L}}$ and $f^{\tilde{\mu}_2}|_{\mathbb{L}}$ are equal. By the reciprocal part of Claim 2.11, this holds only when the two differentials μ_1 and μ_2 are equivalent. Therefore Φ induces a bijection between $\mathcal{T}(\Gamma)$ and $\Phi(\mathcal{T}(\Gamma))$ which is indeed a homeomorphism (cf. [12], [11]; note that the continuity follows from Corollary 2.4. We shall identify $\mathcal{T}(\Gamma)$ (as a topological space) and $\Phi(\mathcal{T}(\Gamma))$.

The next tool for proving the existence of the complex structure is a construction due to L. Ahlfors and G. Weill which provides local sections to Φ .

Definition 2.12 (Harmonic Beltrami differentials). Let $\phi \in \mathcal{Q}_b(\Gamma)$. Setting $\mu_\phi(z) = -2y^2\phi(\bar{z})$ for $z \in \mathbb{H}$ and $\mu_\phi(z) = -2y^2\overline{\phi(z)}$ for $z \in \mathbb{L}$ defines a Beltrami differential, element of $B(\Gamma)$. A Beltrami differential of the form μ_ϕ is said a *harmonic Beltrami differential*. The space of harmonic Beltrami differentials is denoted by $\mathbb{B}_{\mathcal{T}}(\Gamma)$: it is a closed subspace of $B(\Gamma)$. If $\phi \in \mathcal{Q}_b^{\frac{1}{2}}(\Gamma)$, the open ball of radius $\frac{1}{2}$ in $\mathcal{Q}_b(\Gamma)$, then $\mu_\phi \in B^1(\Gamma)$. It is now a theorem of L. Ahlfors and G. Weill that the Schwarzian derivative of $f^{\tilde{\mu}_\phi}|_{\mathbb{L}}$ is equal to ϕ [4]. This means that $\Phi([\mu_\phi]) = \phi$; in other words, $\phi \mapsto [\mu_\phi]$ is a section to Φ . In particular $\Phi(\mathcal{T}(\Gamma))$ contains the ball of radius $\mathcal{Q}_b^{\frac{1}{2}}(\Gamma)$. The map $\phi \mapsto [\mu_\phi]$ from the neighborhood $\mathcal{Q}_b^{\frac{1}{2}}(\Gamma)$ of $[0] \in \mathcal{T}(\Gamma)$ to $\mathbb{B}_{\mathcal{T}}^1(\Gamma)$ is a chart for $\mathcal{T}(\Gamma)$ around $[0]$ called a *harmonic chart*.

The same property holds for all other points [12, p.136]. Let $\mu_0 \in B_{\mathcal{T}}^1(\Gamma)$, and let $(\rho^{\mu_0}, f^{\mu_0})$ be the Fuchsian deformation of Γ representing the point $[\mu_0] \in \mathcal{T}(\Gamma)$. The change of base point map $[v]$ (cf. Definition 2.7) maps neighborhoods of $[\mu_0]$ in $\mathcal{T}(\Gamma)$ to neighborhoods of $[0]$ in $\mathcal{T}(\Gamma^{\mu_0})$. The preceding construction provided a chart for a neighborhood of $[0]$ in $\mathcal{T}(\Gamma^{\mu_0})$; precomposed with $[v]$, it gives a chart which models a neighborhood of $[\mu_0]$ in $\mathcal{T}(\Gamma)$ with a neighborhood of 0 in $\mathbb{B}_{\mathcal{T}}(\Gamma^{\mu_0})$. The transition charts are holomorphic: this follows from the holomorphic dependence of the solutions to the Beltrami equation. The complex structure on Teichmüller space is obtained now since the Banach spaces $\mathbb{B}_{\mathcal{T}}(\Gamma^\mu)$ are isomorphic. This is clear when \mathbb{H}^2/Γ is a surface of finite type since $\mathbb{B}_{\mathcal{T}}(\Gamma^\mu)$ is then isomorphic to \mathbb{C}^{3g+n} . This is true also in general (cf. [12, p. 136]).

Definition 2.13 (The Ahlfors–Bers map). Let $[\mu] \in \mathcal{Q}\mathcal{F}(\Gamma)$. By the Riemann uniformisation theorem, there exist two Fuchsian groups $\rho^1(\Gamma)$, and $\rho^2(\Gamma)$, two univalent maps $f^1: f^\mu(\mathbb{H}) \rightarrow \mathbb{C}$ and $f^2: f^\mu(\mathbb{L}) \rightarrow \mathbb{C}$ such that f^1 (resp. f^2) conjugates the action of $\rho^\mu(\Gamma)$ on $f^\mu(\mathbb{H})$ (resp. on $f^\mu(\mathbb{L})$) to the action of $\rho_1(\Gamma)$ on \mathbb{H} (resp. to the action of $\rho_2(\Gamma)$ on \mathbb{L}). Since $f^\mu(\mathbb{R})$ is locally connected, Carathéodory’s theorem implies that the homeomorphism f^1 extends to a homeomorphism from $f^\mu(\overline{\mathbb{H}})$ to $\overline{\mathbb{H}}$. Using this, it is not difficult to see that the homeomorphism defined by $\tilde{f}^1(z) = f^1 \circ f^\mu(z)$ for $z \in \mathbb{H}$ and $\tilde{f}^1(z) = \overline{f^1 \circ f^\mu(\bar{z})}$ for $z \in \mathbb{L}$ is a quasiconformal homeomorphism of \mathbb{C} which conjugates the actions of Γ and $\rho^1(\Gamma)$ on \mathbb{H} . The same procedure gives

a quasiconformal homeomorphism f^2 of \mathbb{C} which conjugates the actions of Γ and $\rho^2(\Gamma)$ on \mathbb{L} .

In this way, one obtains a map from the space of quasi-Fuchsian deformations of Γ to the product of two copies of the space of Fuchsian deformations of Γ . This map can be expressed as follows at the level of the Beltrami differentials. Let $\mu = (\mu_1, \mu_2) \in B(\Gamma)$, where $\mu_1 = \mu|_{\mathbb{H}}$ and $\mu_2 = \mu|_{\mathbb{L}}$; then the map assigns to μ the pair of Beltrami differentials $(\tilde{\mu}_1, \tilde{\mu}_2)$ in $B_{\mathcal{T}}(\Gamma) \times B_{\mathcal{T}}(\Gamma)$ where $\tilde{\mu}_1(z) = \mu_1(z)$ if $z \in \mathbb{H}$, $\tilde{\mu}_1(z) = \overline{\mu_1(\bar{z})}$ if $z \in \mathbb{L}$, and $\tilde{\mu}_2(z) = \mu_2(z)$ if $z \in \mathbb{L}$, $\tilde{\mu}_2(z) = \mu_2(\bar{z})$ if $z \in \mathbb{H}$. From the construction, this map commutes with $[\cdot]$, giving a homeomorphism, called the *Ahlfors–Bers map*, from $\mathcal{QF}(\Gamma)$ to $\mathcal{T}(\Gamma) \times \mathcal{T}(\Gamma)$.

Definition 2.14 (Complex harmonic charts for $\mathcal{QF}(\Gamma)$). Using the Ahlfors–Bers map, $\mathcal{QF}(\Gamma)$ becomes a complex manifold whose charts are the products of the harmonic charts for $\mathcal{T}(\Gamma)$: these charts are called the *complex harmonic charts*. Such a chart identifies the neighborhood $\mathcal{Q}_b^{\frac{1}{2}}(\Gamma) \times \mathcal{Q}_b^{\frac{1}{2}}(\Gamma)$ of $[0] \in \mathcal{QF}(\Gamma)$ with $\mathbb{B}_{\mathcal{T}}^1(\Gamma) \times \mathbb{B}_{\mathcal{T}}^1(\Gamma)$ assigning to $\varphi = (\phi_1, \phi_2) \in \mathcal{Q}_b^{\frac{1}{2}}(\Gamma) \times \mathcal{Q}_b^{\frac{1}{2}}(\Gamma)$ the point $[v_\varphi] \in \mathcal{QF}(\Gamma)$ where $v_\phi \in B(\Gamma)$ is defined by $v_\phi(z) = \mu_{\phi_1}(z)$ for $z \in \mathbb{H}$ and $v_\phi(z) = \mu_{\phi_2}(z)$ for $z \in \mathbb{L}$. The Beltrami differentials of the form v_ϕ will be called *complex harmonic*, in order to differentiate them from the harmonic Beltrami differentials, elements of $\mathbb{B}_{\mathcal{T}}(\Gamma)$. The complex harmonic differentials form a complex subspace $\mathbb{B}(\Gamma)$ of $B(\Gamma)$. The space $\mathbb{B}_{\mathcal{T}}(\Gamma)$ is a totally real subspace of $\mathbb{B}(\Gamma)$, namely $\mathbb{B}_{\mathcal{T}}(\Gamma) = \{(\mu_1, \mu_2) \mid \mu_1 = \overline{\mu_2}\}$. Under the complex harmonic chart, $\mathcal{T}(\Gamma)$ maps to $\mathbb{B}_{\mathcal{T}}(\Gamma)$. The same charts can be constructed near any point $[\mu] \in \mathcal{T}(\Gamma)$ showing that $\mathcal{T}(\Gamma)$ is a totally real submanifold of the complex manifold $\mathcal{QF}(\Gamma)$.

3 Geodesic currents and Hölder distributions

In this section we associate to a Fuchsian group Γ the *space of geodesic currents* $\mathcal{M}(\Gamma)$ and the *space of Hölder distributions* $\mathcal{H}(\Gamma)$. For a group Γ which is cocompact, the space $\mathcal{M}(\Gamma)$ was introduced in [6]. In [7], Francis Bonahon defined also a map $\mathcal{L}: \mathcal{T}(\Gamma) \rightarrow \mathcal{M}(\Gamma)$, the *Liouville map*. One main feature of this map is that it is a topological embedding and that it allows to recover the Thurston compactification of the Teichmüller space by the *space of measured geodesic laminations*. In [8], Bonahon and Sözen defined, for a cocompact Fuchsian group Γ , a new space, larger than $\mathcal{M}(\Gamma)$: the *space of Hölder distributions* $\mathcal{H}(\Gamma)$. They showed that \mathcal{L} was differentiable in a certain sense (Theorem 3.6).

The definition of the space of Hölder distributions was generalized for a general Fuchsian groups by Dragomir Šarić: he showed that \mathcal{L} is a topological embedding and that the same theorem as Theorem 3.6 did hold in this context too [17], [18].

3.1 The space of geodesics of \mathbb{H}^2

An unparametrized geodesic g of \mathbb{H}^2 is entirely determined by its two endpoints (ξ, ζ) which are two distinct points in $\partial\mathbb{H}^2 = \overline{\mathbb{R}}$. The space of geodesics of \mathbb{H}^2 , \mathcal{G} , is the complement of the diagonal Δ in $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$. The action of $\mathrm{PSL}(2, \mathbb{R})$ on $\overline{\mathbb{R}}$ induces an action on \mathcal{G} which is the product action on $\overline{\mathbb{R}} \times \overline{\mathbb{R}} - \Delta$. The space \mathcal{G} is a C^∞ -manifold covered by the open sets $\gamma\mathcal{U}$ for $\gamma \in \mathrm{PSL}(2, \mathbb{R})$ where $\mathcal{U} = \{(\xi, \zeta) \in \mathbb{R}^2 \mid \xi \neq \zeta\}$. We denote by d a distance on \mathcal{G} corresponding to this smooth structure. The boundary $\overline{\mathbb{R}}$ being identified with the circle \mathbb{R}/\mathbb{Z} , we shall take the distance induced on $\overline{\mathbb{R}}$ from the angular distance on the circle that we shall denote by $|\xi - \xi'|$ and then the distance induced from the product distance on $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$: $d((\xi, \zeta), (\xi', \zeta')) = \sup(|\xi - \xi'|, |\zeta - \zeta'|)$.

Let $\mathcal{C}(\mathcal{G})$ be the space of continuous functions $h: \mathcal{G} \rightarrow \mathbb{C}$ which have compact support, with the norm $\|h\|_\infty = \sup_{g \in \mathcal{G}} |h(g)|$. If K is a compact set contained in \mathcal{G} , we shall denote by $C_K(\mathcal{G})$ the subspace of functions in $\mathcal{C}(\mathcal{G})$ whose support is contained in K . It is a Banach space for the norm $\|h\|_\infty = \sup_{g \in K} |h(g)|$. The dual of $C_K(\mathcal{G})$ is a Banach space for the norm $\|T\|_K = \sup |T(h)|$, the supremum being taken over all the functions $h \in C_K(\mathcal{G})$ with $\|h\|_\infty \leq 1$.

If f is a homeomorphism of $\overline{\mathbb{R}}$, we denote by F the homeomorphism of \mathcal{G} induced by the diagonal action of f

$$(\xi, \zeta) \mapsto (f(\xi), f(\zeta)) = F(\xi, \zeta).$$

This homeomorphism F induces an isomorphism of $C(\mathcal{G})$ by sending h to $f^*(h) = h \circ F^{-1}$.

3.2 Geodesic currents

A linear form on $C(\mathcal{G})$ is *continuous* if, for each compact set $K \subset \mathcal{G}$, its restriction to $C_K(\mathcal{G})$ is continuous. We denote by \mathcal{M} the space of continuous linear forms on $C(\mathcal{G})$.

Definition 3.1 (The geodesic currents). The action of Γ on $\mathcal{C}(\mathcal{G})$ induces by duality an action of Γ on \mathcal{M} . A *geodesic current* is a functional $T \in \mathcal{M}$ which is Γ -invariant, i.e. which satisfies, for any $\gamma \in \Gamma$, and for any $h \in \mathcal{C}(\mathcal{G})$, $T(\gamma^*(h)) = T(h)$. The space of the geodesic currents is denoted by $\mathcal{M}(\Gamma)$.

By the Riesz representation theorem [16], if $T \in \mathcal{M}$, there exists a complex Borel measure λ on \mathcal{G} such that, for all $h \in \mathcal{C}(\mathcal{G})$, one has $T(h) = \int h d\lambda$. Furthermore, the measure λ is positive if and only if the functional T is. For any compact set $K \subset \mathcal{G}$ such that $\lambda(\partial K) = 0$, one has

$$\lambda(K) = \|T\|_K = \sup\{T(h) \mid h \in C_K(\mathcal{G}), \|h\|_\infty \leq 1\}.$$

One of the features of the space $\mathcal{M}(\Gamma)$ is that it contains the Teichmüller space $\mathcal{T}(\Gamma)$. The natural map which realizes this embedding is the *Liouville map*.

3.3 The Liouville measure

We recall first some equivalent ways to define the *Liouville measure* on \mathcal{G} .

- The group $\text{PSL}(2, \mathbb{R})$ acts on \mathcal{G} preserving a Radon measure, the *Liouville measure*. In the charts $\mathcal{U} = \mathbb{R} \times \mathbb{R} - \Delta$, this measure has the expression $d\lambda = \frac{d\xi d\zeta}{|\xi - \zeta|^2}$. In the charts $\gamma\mathcal{U}$, it is given by exactly the same formula. The fact that these local expressions patch together is a reformulation of the basic functional equation satisfied by all elements of the Möbius group $\text{PSL}(2, \mathbb{C})$: for all ξ, ζ in \mathbb{C}

$$(\gamma(\xi) - \gamma(\zeta))^2 = \gamma'(\xi)\gamma'(\zeta)(\xi - \zeta)^2.$$

- The space \mathcal{G} is also a pseudo-Riemannian manifold: the metric

$$ds^2 = \frac{1}{|\xi - \zeta|^2} d\xi \otimes d\zeta$$

is invariant under the action of $\text{PSL}(2, \mathbb{R})$ and the Liouville measure λ is the (pseudo-Riemannian)-measure associated to this metric.

- Consider the differential 2-form ω on $\overline{\mathbb{C}} \times \overline{\mathbb{C}} - \Delta$, defined in the chart $\mathbb{C} \times \mathbb{C}$ by

$$\omega = \frac{\partial^2}{\partial \xi \partial \zeta} \log |\xi - \zeta| d\xi \wedge d\zeta.$$

In restriction to $\overline{\mathbb{R}} \times \overline{\mathbb{R}} - \Delta$, ω is a volume form whose absolute value is λ .

- The measure λ is related to the cross-ratio. Let a, b, c and d be four points on $\overline{\mathbb{R}}$ be such that the intervals $[a, b]$ and $[c, d]$ are disjoint. Then $[a, b] \times [c, d]$ is a rectangle in \mathcal{G} and

$$\int_{[a,b] \times [c,d]} \omega = \log |a - c| + \log |b - d| - \log |a - d| - \log |b - c|.$$

Then $\lambda([ab] \times [cd]) = |\log |(a, b, c, d)||$, where $(a, b, c, d) = \frac{(a-c)(b-d)}{(a-d)(b-c)}$ is the cross-ratio of the four points a, b, c, d . If these points occur in this order on $\overline{\mathbb{R}}$, then $\log |(a, b, c, d)| > 0$.

Definition 3.2 (The Liouville map). Let $[\mu] \in \mathcal{T}(\Gamma)$ and let F^μ denotes the homeomorphism of \mathcal{G} which is induced by the diagonal action of f^μ , i.e. $F^\mu(\xi, \zeta) = (f^\mu(\xi), f^\mu(\zeta))$. The group $\rho^\mu(\Gamma)$ preserves the Liouville measure λ . Thus, since the actions on \mathcal{G} of the groups Γ and $\rho^\mu(\Gamma)$ are conjugated by F^μ , the group Γ preserves the *pullback measure* $(F^\mu)^*(\lambda)$. Recall that the pullback measure $(F^\mu)^*(\lambda)$ is characterized by the property that for any Borel set A , one has $(F^\mu)^*(\lambda)(A) = \lambda(F^\mu(A))$. Since F^μ only depends on the point $[\mu]$ in $\mathcal{T}(\Gamma)$, one defines a map $\mathcal{L}: \mathcal{T}(\Gamma) \rightarrow \mathcal{M}(\Gamma)$ by setting $\mathcal{L}([\mu]) = (F^\mu)^*(\lambda)$. This is the *Liouville map*.

As a map from $\mathcal{T}(\Gamma)$ to $\mathcal{M}(\Gamma)$, \mathcal{L} has no smoothness properties. However, if one replaces $\mathcal{M}(\Gamma)$ by the larger space $\mathcal{H}(\Gamma)$ of geodesic Hölder distributions, Bonahon

and Sözen showed that the new map $\mathcal{L}: \mathcal{T}(\Gamma) \rightarrow \mathcal{H}(\Gamma)$ becomes C^1 along paths. We introduce now the space $\mathcal{H}(\Gamma)$, first when Γ is cocompact like in [8], then when Γ is a general Fuchsian group [17], [18].

3.4 Geodesic Hölder distributions for a cocompact Fuchsian group

In this section we shall suppose that Γ is cocompact.

Continuous linear forms on $C(\mathcal{G})$ are particular examples of distributions in the sense of Laurent Schwartz. However, if one intends to generalize the notion of geodesic currents invariant under a Fuchsian group to the context of distributions, one is faced with the difficulty that \mathcal{G} , although it is a smooth manifold, is not smooth in a *natural way*. Indeed let $[\mu] \in \mathcal{T}(\Gamma)$; then the homeomorphism $f^\mu: \overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}$ which conjugates the actions of Γ and of Γ^μ on $\overline{\mathbb{R}}$ is not smooth in general and the same holds a fortiori for F^μ . It follows that the map that F^μ induces on $C(\mathcal{G})$ does not preserve the spaces of differentiable functions of any differentiability class. In particular, there are no natural isomorphisms between the spaces of distributions which are invariant under Γ and of those invariant under Γ^μ . Francis Bonahon solved this difficulty by introducing the appropriate functional space, the *space of Hölder geodesic distributions*.

Definition 3.3 (Distributions of order α). Let (K, d) a metric space. Let $\alpha \in]0, 1[$. A function $h: K \rightarrow \mathbb{C}$ is α -Hölder if there is a constant C such that $|h(x) - h(x')| \leq Cd(x, x')^\alpha$ for all $x, x' \in K$. For such a function h , one defines the α -Hölder norm by

$$\|h\|_\alpha = \sup_{x \in K} |h(x)| + \sup_{x \neq x'} \frac{|h(x) - h(x')|}{d(x, x')^\alpha}.$$

The space of α -Hölder functions $h: K \rightarrow \mathbb{C}$ which have compact support is denoted $\text{Lip}^\alpha(K)$. For K a compact subset of \mathcal{G} , we denote by $\text{Lip}_K^\alpha(\mathcal{G})$ the subspace of $\text{Lip}^\alpha(\mathcal{G})$ which consists of the functions whose support is contained in K . With the norm $\|\cdot\|_\alpha$, $\text{Lip}_K^\alpha(\mathcal{G})$ is a Banach space. The dual of $\text{Lip}_K^\alpha(\mathcal{G})$ is called the *space of distributions of order α on K* . It is a Banach space for the norm $\|\cdot\|^\alpha$ which is the dual norm to the norm $\|\cdot\|_\alpha$; it will be denoted by $\mathcal{H}^\alpha(K)$.

When K is a compact submanifold of \mathcal{G} , it is a classical fact that the subspace $C^\infty(\mathcal{G}) \cap \text{Lip}_K^\alpha(\mathcal{G})$ is dense in $\text{Lip}_K^\alpha(\mathcal{G})$ for the α -Hölder norm.

Definition 3.4 (Geodesic Hölder distribution). A *geodesic Hölder distribution of order α* is a linear functional $T: C^\infty(\mathcal{G}) \rightarrow \mathbb{C}$ which, for each compact submanifold K of \mathcal{G} , can be extended continuously to $\text{Lip}_K^\alpha(\mathcal{G})$ as an element of $\mathcal{H}^\alpha(K)$. The space of geodesic Hölder distributions of order α is denoted by \mathcal{H}^α . If $\alpha' \leq \alpha$, one has $\mathcal{H}^{\alpha'} \subset \mathcal{H}^\alpha$.

A *geodesic Hölder distribution* is a linear functional $T: C^\infty(\mathcal{G}) \rightarrow \mathbb{C}$ which belongs to \mathcal{H}^α for all $\alpha < 1$. The space of geodesic Hölder distributions is denoted by \mathcal{H} . One has $\mathcal{H} = \bigcap_\alpha \mathcal{H}^\alpha$.

A *geodesic Hölder distribution* for Γ is a functional $T \in \mathcal{H}$ which is Γ -invariant; these distributions form a closed subspace $\mathcal{H}(\Gamma)$ of \mathcal{H} . One has $\mathcal{H}(\Gamma) = \bigcap_{\alpha} \mathcal{H}^{\alpha}(\Gamma)$.

Since Γ is cocompact, there is a compact submanifold $K \subset \mathcal{G}$ such that \mathcal{G} is covered by the translates of K by the elements of Γ : choose for instance a submanifold K which contains in its interior the set of the geodesics of \mathcal{G} which pass through a fundamental domain for the action of Γ on \mathbb{H} . Define now a norm on $\mathcal{H}^{\alpha}(\Gamma)$ by $\|T\|_K^{\alpha}$. This norm turns \mathcal{H}^{α} into a Banach space and the family of norms $\{\|\cdot\|_K^{\alpha}\}$, as $\alpha \in]0, 1[$ turns \mathcal{H} into a Fréchet space. It is not difficult to see that this topology can also be defined by the family of norms $\{\|\cdot\|_K^{1/n}\}$. Clearly the choice of another compact K defines the same topology as long as K keeps the property that the union of its translates equals \mathcal{G} .

Remark 3.5. For $[\mu] \in \mathcal{T}(\Gamma)$, it is known that f^{μ} is Hölder. Let $\delta > 0$ be such that $(f^{\mu})^{-1}$ is Hölder with exponent δ . Then $F^{\mu}(\xi, \zeta) = (f^{\mu}(\xi), f^{\mu}(\zeta))$ is δ -Hölder also. In particular, for each compact set $K \subset \mathcal{G}$ and for each $\alpha < 1$, F^{μ} acting on $C^c(\mathcal{G})$ by $h \mapsto h \circ (F^{\mu})^{-1}$ maps continuously $\text{Lip}_K^{\alpha}(\mathcal{G})$ to $\text{Lip}_{F^{\mu}(K)}^{\delta\alpha}(\mathcal{G})$. By duality, this map induces a linear map $\mathcal{H}(f^{\mu}) : \mathcal{H}^{\alpha\delta}(\Gamma) \rightarrow \mathcal{H}^{\alpha}(\Gamma)$ which is continuous. Since for $[\mu] \in \mathcal{T}(\Gamma)$, $(f^{\mu})^{-1}$ is also Hölder, $\mathcal{H}(f^{\mu})$ is a linear isomorphism.

The space $\mathcal{M}(\Gamma)$ is contained in $\mathcal{H}(\Gamma)$: for ν an element in $\mathcal{M}(\Gamma)$, we shall use the notation $\langle \nu, h \rangle = \int_{\mathcal{G}} h d\nu$. The Liouville map can therefore be viewed as a map from $\mathcal{T}(\Gamma)$ to $\mathcal{H}(\Gamma)$. The following theorem is due to Bonahon and Sözen [8].

Theorem 3.6. *Let $t \mapsto c(t) \in \mathcal{T}(\Gamma)$ be a smooth path defined on the interval $|t| \leq \varepsilon$. Then the map $t \mapsto \mathcal{L}(c(t))$ is differentiable: for any $\alpha < 1$, and for any function $h \in \text{Lip}_K^{\alpha}(\mathcal{G})$, the function $t \mapsto \langle \mathcal{L}(c(t)), h \rangle$ is differentiable.*

3.5 Geodesic Hölder distributions for a general Fuchsian group

For a Fuchsian group which is not cocompact, the above definition of the space $\mathcal{H}(\Gamma)$ is not sufficient in order for the map \mathcal{L} to be even a topological embedding. An extra boundedness property for the geodesic Hölder distributions is required. The definition which follows was introduced by D. Šarić [17], [18].

To begin with, let \mathcal{R} be the set of rectangles $[a, b] \times [c, d] \subset \mathcal{G}$ for which the cross-ratio (a, b, c, d) equals 2: this means that the geodesics ac and bd make an angle equal to $\pi/2$ at their intersection point. Fix an element of \mathcal{R} , $R_0 = [a_0, b_0] \times [c_0, d_0]$ such that the intervals $[a_0, b_0]$ and $[c_0, d_0]$ are contained in \mathbb{R} . For any rectangle $R = [a, b] \times [c, d]$ in \mathcal{R} , let q be the element of $\text{PSL}(2, \mathbb{R})$ which maps (a_0, b_0, c_0, d_0) to (a, b, c, d) . If $h \in \text{Lip}_R^{\alpha}(\mathcal{G})$, then $h \circ Q \in \text{Lip}_{R_0}^{\alpha}(\mathcal{G})$ (recall the notation $Q(\xi, \zeta) = (q(\xi), q(\zeta))$).

Definition 3.7 (Geodesic Hölder distribution). Let $\alpha < 1$. A *bounded geodesic Hölder distribution of order α* is a linear functional $T : C^{\infty}(\mathcal{G}) \rightarrow \mathbb{C}$ such that

- (1) for each rectangle $R \in \mathcal{R}$, T extends continuously to $\text{Lip}_R^\alpha(\mathcal{G})$ as an element of $\mathcal{H}^\alpha(R)$, and
- (2) the quantity $\|T\|^\alpha = \sup\{|T(\phi)|\}$ is finite, where the supremum is taken over all rectangles $R \in \mathcal{R}$ and all $\phi \in \text{Lip}_R^\alpha(\mathcal{G})$ such that $\|\phi \circ Q\|_\alpha \leq 1$.

The space of bounded geodesic Hölder distributions of order α is denoted by \mathcal{H}^α . It is a Banach space for the norm $\|\cdot\|^\alpha$.

Let Γ be a Fuchsian group. The space of bounded geodesic Hölder distributions of order α which are invariant under the action of Γ is a closed subspace of \mathcal{H}^α , denoted by $\mathcal{H}^\alpha(\Gamma)$. For $\alpha' \leq \alpha$, one has $\mathcal{H}^{\alpha'}(\Gamma) \subset \mathcal{H}^\alpha(\Gamma)$.

A *geodesic Hölder distribution* for Γ is a linear functional $T : C^\infty(\mathcal{G}) \rightarrow \mathbb{C}$ which belongs to $\mathcal{H}^\alpha(\Gamma)$ for each α . They form a space \mathcal{H} , equal to the intersection $\bigcap_\alpha \mathcal{H}^\alpha$; \mathcal{H} is a Fréchet space whose topology is defined by the family of semi-norms $\|\cdot\|^\alpha$.

Remark 3.8. When Γ is cocompact, the above definition agrees with the one given in the preceding subsection.

Remark 3.9. If one replaces, in the definition of bounded geodesic Hölder distribution, the set of rectangles \mathcal{R} by the set \mathcal{R}_β of those rectangles $[a, b] \times [c, d]$ with (a, b, c, d) equal to β , one obtains a set of bounded Hölder geodesic distributions which is equal to $\mathcal{H}(\Gamma)$. For each $\alpha < 1$, the norm on $\mathcal{H}^\alpha(\Gamma)$ defined as before, after the choice of a rectangle $R'_0 \in \mathcal{R}_\beta$ playing the same role as R_0 , is equivalent to the norm $\|\cdot\|^\alpha$.

Remark 3.10. Remark 3.5 can be made also in the present situation. Let $[\mu] \in \mathcal{T}(\Gamma)$, and suppose that $(f^\mu)^{-1}$ is δ -Hölder. Then the map $h \rightarrow h \circ (F^\mu)^{-1}$ induces by duality a map $\mathcal{H}(f^\mu) : \mathcal{H}^{\alpha\delta}(\Gamma^\mu) \rightarrow \mathcal{H}^\alpha(\Gamma)$ which is continuous for each $\alpha < 1$. Also $\mathcal{H}(f^\mu) : \mathcal{H}(\Gamma^\mu) \rightarrow \mathcal{H}(\Gamma)$ is an isomorphism.

Claim 3.11. For $[\mu] \in \mathcal{T}(\Gamma)$, the Liouville measure $\mathcal{L}([\mu])$ belongs to $\mathcal{H}(\Gamma)$.

Proof. A positive Radon measure ν on \mathcal{G} defines a distribution by integration $h \mapsto \int_{\mathcal{G}} h d\nu$. For each rectangle $R \in \mathcal{R}$, this distribution is clearly continuous on $\text{Lip}_R^\alpha(\mathcal{G})$ with norm less than $\nu(K)$. To prove Claim 3.11, we need to show that for each $[\mu] \in \mathcal{T}(\Gamma)$, and for each $R \in \mathcal{R}$, the measures $\mathcal{L}([\mu])(R)$ are bounded independently of R . Let R be a rectangle in \mathcal{R} : $R = [a, b] \times [c, d]$ with $(a, b, c, d) = 2$. Let $q \in \text{PSL}(2, \mathbb{R})$ be the element sending R_0 to R and let $r \in \text{PSL}(2, \mathbb{R})$ be the element such that $r \circ f^\mu \circ q$ fixes three of the vertices of the rectangle R_0 . Then the Beltrami differential of the quasiconformal homeomorphism $r \circ f^\mu \circ q$ equals $\mu(q(z)) \frac{q'(z)}{q'(z)}$; in particular, it has the same L^∞ -norm as μ . It follows from Proposition 2.2 that $\lambda([f^\mu(a), f^\mu(b)] \times [f^\mu(c), f^\mu(d)])$ is bounded from above and from below by two positive constants depending only on $\|\mu\|_\infty$ and not on R . Saying that these numbers are bounded from above implies that $\mathcal{L}([\mu])$ belongs to $\mathcal{H}(\Gamma)$. □

This remark allows us to consider the Liouville map \mathcal{L} as a map from $\mathcal{T}(\Gamma)$ to $\mathcal{H}(\Gamma)$. The following theorem is due to D. Šarić [18].

Theorem 3.12. *The Liouville map $\mathcal{L}: \mathcal{T}(\Gamma) \rightarrow \mathcal{H}(\Gamma)$ is a proper topological embedding.*

Proof. *The map \mathcal{L} is continuous.* We begin with the continuity at $[0] \in \mathcal{T}(\Gamma)$. Let $\alpha > 1$. Let R be a rectangle in \mathcal{R} and $h \in \text{Lip}_R^\alpha(\mathcal{G})$. Let $q \in \text{PSL}(2, \mathbb{R})$ be the Möbius map sending R_0 to R . We need to show that $|\mathcal{L}([\mu])(h) - \mathcal{L}([0])(h)| \leq C(\mu) \|h \circ Q\|_\alpha$ for a constant $C(\mu)$ which is independent of R and h and which tends to 0 with $\|\mu\|_\infty$.

By definition of the Liouville map, one has

$$\langle \mathcal{L}([\mu]), h \rangle = \int_{\mathcal{G}} h \circ (F^\mu)^{-1} d\lambda = \int_{\mathcal{G}} (h \circ Q) \circ (Q^{-1} \circ (F^\mu)^{-1}) d\lambda.$$

Let $q^\mu \in \text{PSL}(2, \mathbb{R})$ be the Möbius map such that the quasiconformal homeomorphism

$$\phi^\mu = q^\mu \circ f^\mu \circ q$$

fixes 0, 1 and ∞ ; let $\Phi^\mu(\xi, \zeta) = (\phi^\mu(\xi), \phi^\mu(\zeta))$. Then, since the homeomorphism Q^μ , where $Q^\mu(\xi, \zeta) = (q^\mu(\xi), q^\mu(\zeta))$ preserves λ , one has

$$\langle \mathcal{L}([\mu]), h \rangle = \int_{\mathcal{G}} (h \circ Q) \circ (\Phi^\mu)^{-1} d\lambda = \langle (\Phi^\mu)^* \lambda, h \circ Q \rangle. \tag{3.1}$$

Since $h \circ Q \in \text{Lip}_{R_0}^\alpha(\mathcal{G})$, one has

$$|\langle \mathcal{L}([\mu]), h \rangle - \langle \mathcal{L}([0]), h \rangle| \leq \lambda(R_0) \|h \circ Q\|_\alpha \sup_{g \in R_0} d((\Phi^\mu)^{-1}(g), g)^\alpha.$$

Observe that the Beltrami coefficient of ϕ^μ is $q^*(\mu) = (\mu \circ q) \frac{\bar{q}}{q}$. In particular, $\|q^*(\mu)\|_\infty = \|\mu\|_\infty$. Corollary 2.4 implies that $\sup_{\xi \in \mathbb{R}} d((\phi^\mu)^{-1}(\xi), \xi) \leq C(\|\mu\|_\infty)$ where $C(\|\mu\|_\infty)$ tends to 0 with $\|\mu\|_\infty$; the same property is therefore also satisfied by $\sup_{g \in R_0} d((\Phi^\mu)^{-1}(g), g)$. Thus, one has, for all $h \in \text{Lip}_R^\alpha(\mathcal{G})$

$$|\langle \mathcal{L}([\mu]), h \rangle - \langle \mathcal{L}([0]), h \rangle| \leq C(\|\mu\|_\infty) \|h\|_\alpha$$

where $C(\|\mu\|_\infty)$ tends to 0 with $\|\mu\|_\infty$ and does not depend on R . This means that \mathcal{L} is continuous at $[0]$.

The proof of the continuity at an arbitrary point $[\mu_0] \in \mathcal{T}(\Gamma)$ reduces to the above argument using the change of basepoint map $[v]$ (cf. Definition 2.7). For any $\mu \in B^1(\Gamma)$, one has $f^\mu = f^{v(\mu)} \circ f^{\mu_0}$. So, if $h \in \mathcal{C}(\mathcal{G})$, one has

$$\langle \mathcal{L}([\mu]), h \rangle = \int h \circ (F^{\mu_0})^{-1} (F^{v(\mu)})^* (d\lambda) = \langle \mathcal{L}([v(\mu)]), h \circ (F^{\mu_0})^{-1} \rangle. \tag{3.2}$$

Equation 3.2 says that $\mathcal{L} = \mathcal{H}(f^{\mu_0}) \circ \mathcal{L}^{\mu_0} \circ [v]$, where $\mathcal{H}(f^{\mu_0}): \mathcal{H}(\Gamma^{\mu_0}) \rightarrow \mathcal{H}(\Gamma)$ is the map induced by f^{μ_0} (cf. Remark 3.5), $\mathcal{L}^{\mu_0}: \mathcal{T}(\Gamma^{\mu_0}) \rightarrow \mathcal{H}(\Gamma^{\mu_0})$ is the Liouville map for the group Γ^{μ_0} , and $[v]$ is the change of base point map. In the remark 3.10, we noticed that $\mathcal{H}(f^{\mu_0})$ is a linear isomorphism between the Fréchet spaces $\mathcal{H}(\Gamma^{\mu_0})$ and $\mathcal{H}(\Gamma)$. By the first part of the proof, \mathcal{L} is continuous at $[0]$. Since $[v]$ is a homeomorphism, it follows that \mathcal{L} is continuous at $[\mu_0]$.

\mathcal{L} is injective. If $(F^{\mu_1})^*(\lambda) = (F^{\mu_2})^*(\lambda)$, one obtains, by evaluating these measures on rectangles $[a, b] \times [c, d] \subset \overline{\mathbb{R}} \times \overline{\mathbb{R}}$ that $f^{\mu_2} \circ (f^{\mu_1})^{-1}$ preserves the cross-ratio of any 4-tuples in $\overline{\mathbb{R}}$. Since the homeomorphisms f^{μ_i} are normalized, this implies that $f^{\mu_1}|_{\mathbb{R}} = f^{\mu_2}|_{\mathbb{R}}$, and so $[\mu_1] = [\mu_2]$.

\mathcal{L} is an embedding. This means that \mathcal{L} is a homeomorphism from $\mathcal{T}(\Gamma)$ to its image with the topology induced from that of $\mathcal{H}(\Gamma)$. To see that, it remains to prove that the inverse map \mathcal{L}^{-1} is continuous.

Suppose that $([\mu_i])$ is a sequence in $\mathcal{T}(\Gamma)$ such that $\mathcal{L}([\mu_i])$ tends to $\mathcal{L}([v])$ in $\mathcal{H}(\Gamma)$. We want to prove that $[\mu_i]$ tends to $[v]$ in $\mathcal{T}(\Gamma)$. Extend the characteristic function χ_{R_0} of R_0 to a positive and α -Hölder function $h \leq 1$ supported on a rectangle $R'_0 = [a, b] \times [c, d]$ with $(a, b, c, d) = \beta$. Let R be a rectangle in \mathcal{R} , and let $q \in \text{PSL}(2, \mathbb{R})$ the element sending R_0 to R . Then $h \circ Q^{-1}$ is a positive and α -Hölder function which extends the characteristic function of R ; it is bounded from above by 1 and supported in the rectangle $Q(R'_0)$, which is an element of \mathcal{R}_β . By Remark 3.9 and by the definition of the topology of $\mathcal{H}(\Gamma)$, for all ε , there is an i_0 such that for $i \geq i_0$, the inequality

$$|\langle \mathcal{L}([\mu_i]), h \circ Q^{-1} \rangle - \langle \mathcal{L}([v]), h \circ Q^{-1} \rangle| \leq \varepsilon$$

holds for any rectangle $R \in \mathcal{R}$. By the proof of claim 3.11,

$$\langle \mathcal{L}([v]), h \circ Q^{-1} \rangle \leq \langle \mathcal{L}([v]), \chi_{qR'} \rangle \leq C,$$

where the constant C only depends on $\|v\|_\infty$ and on β . Also,

$$\langle \mathcal{L}([\mu_i]), h \circ Q^{-1} \rangle \geq \langle \mathcal{L}([\mu_i]), \chi_R \rangle.$$

Therefore, for all $i \geq i_0$, and for all $R \in \mathcal{R}$, one has

$$\lambda(F^{\mu_i}(R)) = \langle \mathcal{L}([\mu_i]), \chi_R \rangle \leq C + \varepsilon. \tag{3.3}$$

We now have

Claim 3.13. *The sequence $([\mu_i])$ is bounded in $\mathcal{T}(\Gamma)$.*

Proof. The ‘‘complementary’’ rectangle to a rectangle $[a, b] \times [c, d] \in \mathcal{R}$, i.e. the rectangle $[b, c] \times [d, a]$ is also an element of \mathcal{R} . Therefore Equation 3.3 implies that the measures $\lambda(F^{\mu_i}(R))$ are also bounded from below by a non-zero constant independent of i and of $R \in \mathcal{R}$. From the definition of the Liouville measure, we deduce that for all rectangles $R = [a, b] \times [c, d]$ in \mathcal{R} , the cross-ratios $(f^{\mu_i}(a), f^{\mu_i}(b), f^{\mu_i}(c), f^{\mu_i}(d))$ are bounded between two positive constants independent of i . This implies that f^{μ_i} is S -quasisymmetric, with a constant S independent of i . Indeed, for $x \in \mathbb{R}$ and $t \in \mathbb{R}$, the rectangle $[x - t, x] \times [x + t, \infty]$ belongs to \mathcal{R} since $(x - t, x, x + t, \infty) = 2$; therefore

$$\frac{1}{S} |f^{\mu_i}(x) - f^{\mu_i}(x - t)| \leq |f^{\mu_i}(x + t) - f^{\mu_i}(x)| \leq S |f^{\mu_i}(x) - f^{\mu_i}(x - t)|,$$

for a constant S independent of x , t and i . It is known that any S -quasisymmetric homeomorphism f of $\overline{\mathbb{R}}$ extends to a $K(S)$ -quasiconformal homeomorphism of \mathbb{H} . The extension operator constructed by Douady–Earle [10] has the property of commuting with the action of $\mathrm{PSL}(2, \mathbb{R})$: it provides therefore an extension of $f^{\mu_i}|_{\mathbb{R}}$ which is a $K(S)$ -quasiconformal homeomorphism and which is of the form f^{v_i} , for $v_i \in B(\Gamma)$. By definition, $[\mu_i] = [v_i]$: the sequence $([\mu_i])$ is therefore bounded in $\mathcal{T}(\Gamma)$. □

By Claim 3.13, the sequence $([\mu_i])$ converges to $[\mu] \in \mathcal{T}(\Gamma)$ up to possibly extracting a subsequence. By the continuity of \mathcal{L} , $\mathcal{L}([\mu]) = \mathcal{L}([v])$. By the injectivity of \mathcal{L} , $[\mu] = [v]$. Therefore \mathcal{L}^{-1} is continuous.

\mathcal{L} is proper. Let $([\mu_i])_{i \in \mathbb{N}}$ be an unbounded sequence in $\mathcal{T}(\Gamma)$. Then, by the proof of Claim 3.13, for some rectangle $R \in \mathcal{R}$ the sequence $(\mathcal{L}([\mu_i])(R))$ is unbounded as $i \rightarrow \infty$. It follows that $(\mathcal{L}([\mu_i]))$ is unbounded in $\mathcal{H}(\Gamma)$. □

4 Analytic properties of the Liouville map

This section is devoted to the study of regularity properties of the map \mathcal{L} with respect to the analytic structure on $\mathcal{T}(\Gamma)$. By definition, the space $\mathcal{H}(\Gamma)$ of the geodesic Hölder distributions for Γ is the intersection of the spaces $\mathcal{H}^\alpha(\Gamma)$ as α decreases to 0. In particular, for each $\alpha < 1$, \mathcal{L} is a map from $\mathcal{T}(\Gamma)$ to $\mathcal{H}^\alpha(\Gamma)$.

Theorem 4.1. *For each $0 < \alpha < 1$, there exists an open neighborhood \mathcal{N}^α of $\mathcal{T}(\Gamma)$ in $\mathcal{QF}(\Gamma)$ such that \mathcal{L} extends to a holomorphic map $\tilde{\mathcal{L}}: \mathcal{N}^\alpha \rightarrow \mathcal{H}^\alpha(\Gamma)$.*

Recall that a map between complex Banach manifolds is holomorphic if and only if, when it is expressed in local charts, its restriction to each affine complex disc is holomorphic [9, p. 28]. Saying that a map $F: \mathbb{D} \rightarrow E$ from the unit disc to a Banach manifold E is holomorphic can be formulated in any of the following equivalent ways:

- (1) the map $F(w)$ can be expanded as a series $\sum_{n \in \mathbb{N}} w^n F_n$ which is uniformly convergent over compact sets in \mathbb{D} ;
- (2) the map F is continuous and the Cauchy formula holds: for all $r < 1$ and for all $|w| \leq r$, one has

$$F(w) = \frac{1}{2i\pi} \int_{|z|=r} \frac{F(z)}{z-w} dz.$$

To prove Theorem 4.1 we shall construct a neighborhood \mathcal{N}^α of $\mathcal{T}(\Gamma)$ in $\mathcal{QF}(\Gamma)$, and a map $\tilde{\mathcal{L}}: \mathcal{N}^\alpha \rightarrow \mathcal{H}^\alpha(\Gamma)$ which extends \mathcal{L} and whose restriction to any affine holomorphic disc can be expanded as a series $\sum_{n \in \mathbb{N}} w^n \tilde{\mathcal{L}}_n$, absolutely convergent in $\mathcal{H}^\alpha(\Gamma)$ over compact sets. This neighborhood \mathcal{N}^α depends on α and it seems that this is *necessary*, i.e. it seems impossible that $\mathcal{L}: \mathcal{T}(\Gamma) \rightarrow \mathcal{H}(\Gamma)$ extends to

a holomorphic map defined on some neighborhood of $\mathcal{T}(\Gamma)$ in $\mathcal{QF}(\Gamma)$. However, Theorem 4.1 has the following corollary, which we state as a theorem, although it follows directly from 4.1.

Theorem 4.2. *The Liouville map $\mathcal{L} : \mathcal{T}(\Gamma) \rightarrow \mathcal{H}(\Gamma)$ is analytic.*

We shall first prove the following “local” version of Theorem 4.1. Denote by $\mathcal{G}^f \subset \mathcal{G}$, the “finite part of \mathcal{G} ”, i.e. the set $\{(\xi, \zeta) \in \mathbb{R}^2 \mid \xi < \zeta\}$. For technical reasons – namely the expression for the Liouville measure is simpler in \mathcal{G}^f – we first consider rectangles contained in \mathcal{G}^f . By restriction to functions supported in a rectangle R , \mathcal{L} can be viewed as a map taking values in $\mathcal{H}^\alpha(R)$; we continue to denote this map by \mathcal{L} .

Theorem 4.3. *Let R be a rectangle contained in \mathcal{G}^f . Then, there exists a neighborhood \mathcal{N}^α of $\mathcal{T}(\Gamma)$ in $\mathcal{QF}(\Gamma)$ such that $\mathcal{L} : \mathcal{T}(\Gamma) \rightarrow \mathcal{H}^\alpha(R)$ extends to a holomorphic map $\tilde{\mathcal{L}} : \mathcal{N}^\alpha \rightarrow \mathcal{H}^\alpha(R)$.*

The neighborhood \mathcal{N}^α provided by the proof will depend on R and on α .

To prove Theorem 4.3, we give another definition of the Liouville map.

Claim 4.4. *Let $[\mu] \in \mathcal{T}(\Gamma)$. Then in restriction to \mathcal{G}^f , the measure $\mathcal{L}([\mu])$ is equal to the distribution $\frac{\partial^2}{\partial \xi \partial \zeta} \log(f^\mu(\zeta) - f^\mu(\xi))$, i.e. for each function $h : \mathcal{G}^f \rightarrow \mathbb{C}$ of class C^2 and with compact support, one has*

$$\langle \mathcal{L}([\mu]), h \rangle = \int_{\mathcal{G}} \frac{\partial^2 h}{\partial \xi \partial \zeta} \cdot \log(f^\mu(\zeta) - f^\mu(\xi)) d\xi d\zeta.$$

Proof. One can approximate f^μ by a sequence $(f_i)_{i \in \mathbb{N}}$ of diffeomorphisms $f_i : \overline{\mathbb{R}} \rightarrow \overline{\mathbb{R}}$ such that, when $i \rightarrow \infty$, (f_i) converges to f^μ and (f_i^{-1}) converges $(f^\mu)^{-1}$ uniformly over $\overline{\mathbb{R}}$. Denote by F_i the diffeomorphism $(\xi, \zeta) \mapsto (f_i(\xi), f_i(\zeta))$ of \mathcal{G} . Then the measure $F_i^*(\lambda)$ is absolutely continuous with respect to the Lebesgue measure $d\xi d\zeta$ on \mathcal{G}^f with density equal to $\frac{\partial^2}{\partial \xi \partial \zeta} \log(f_i(\zeta) - f_i(\xi))$. Integrating by parts we deduce that for each function $h : \mathcal{G}^f \rightarrow \mathbb{C}$ of class C^2 and with compact support

$$\langle F_i^*(\lambda), h \rangle = \int_{\mathcal{G}} \frac{\partial^2 h}{\partial \xi \partial \zeta} \log(f_i(\zeta) - f_i(\xi)) d\xi d\zeta. \quad (4.1)$$

As i tends to ∞ , the uniform convergence of F_i^{-1} to $(F^\mu)^{-1}$ implies that $\langle F_i^*(\lambda), h \rangle = \langle \lambda, h \circ F_i^{-1} \rangle$ tends to $\langle \lambda, h \circ (F^\mu)^{-1} \rangle$ that is $\langle \mathcal{L}([\mu]), h \rangle$. The right term of 4.1 converges to $\int_{\mathcal{G}} \frac{\partial^2 h}{\partial \xi \partial \zeta} \log(f_i(\zeta) - f_i(\xi)) d\xi d\zeta$. This proves Claim 4.4. \square

Definition 4.5. Denote by $C^2(R)$ the space of the functions $h \in \mathcal{G}$ which are of class C^2 and whose support is contained in R . Observe for later use, that for each

$0 < \alpha < 1$, $\mathcal{C}^2(R)$ is dense in $\text{Lip}_K^\alpha(\mathcal{G})$. Denote by $\mathcal{D}(R)$ the dual of $C^2(R)$; with the norm dual to the C^2 -norm, $\mathcal{D}(R)$ is a Banach space.

Since $\mathcal{M}(\Gamma)$ is contained in $\mathcal{D}(R)$, one can consider \mathcal{L} as a map taking values in $\mathcal{D}(R)$.

The following corollary is an easy consequence of Claim 4.4.

Corollary 4.6. *The map $\mathcal{L}: \mathcal{T}(\Gamma) \rightarrow \mathcal{D}(R)$ extends to a holomorphic map*

$$\tilde{\mathcal{L}}: \mathcal{QF}(\Gamma) \rightarrow \mathcal{D}(R).$$

Proof. It suffices to see that the map $B_{\mathcal{T}}^1(\Gamma) \rightarrow \mathcal{D}(R)$ which assigns to $\mu \in B_{\mathcal{T}}^1(\Gamma)$ the Liouville measure $\mathcal{L}([\mu])$ extends to a map $\tilde{\mathcal{L}}: B^1(\Gamma) \rightarrow \mathcal{D}(R)$ which is holomorphic and only depends on $[\mu]$ (and so defines a map on $\mathcal{QF}(\Gamma)$). Recall that the rectangle R is contained in \mathcal{G}^f ; the expression for $\tilde{\mathcal{L}}$ is suggested by Claim 4.4. Let us define, for $\mu \in B^1(\Gamma)$, $\tilde{\mathcal{L}}(\mu) \in \mathcal{D}(R)$ as the distribution

$$\frac{\partial^2}{\partial \xi \partial \zeta} \log(f^\mu(\zeta) - f^\mu(\xi)).$$

This distribution certainly only depends on the equivalence class of $[\mu]$. Also, by Claim 4.4, for $\mu \in B_{\mathcal{T}}^1(\Gamma)$, one has $\tilde{\mathcal{L}}([\mu]) = \mathcal{L}([\mu])$. By Theorem 2.9, $\log(f^\mu(\zeta) - f^\mu(\xi))$ is a holomorphic function of μ , for each (ξ, ζ) . Since this function is also continuous in the three variables, the measure $\log(f^\mu(\zeta) - f^\mu(\xi))d\xi d\zeta$ is holomorphic in μ . The corollary follows. □

For proving Theorem 4.3, we shall construct for each $\alpha < 1$, a neighborhood \mathcal{N}^α of $\mathcal{T}(\Gamma)$ in $\mathcal{QF}(\Gamma)$ on which the map $\tilde{\mathcal{L}}$ restricts to a holomorphic map with values in $\mathcal{H}^\alpha(R)$. The next two subsections develop the tools necessary for doing this.

4.1 Hölder regularity of holomorphic motions

Let $w \mapsto \mu_w \in L^\infty(\mathbb{C})$ be a holomorphic family of Beltrami differentials parametrized by \mathbb{D} . Denote by f^w the normalized solution to the Beltrami equation with Beltrami coefficient μ_w . The map $(w, z) \mapsto f^w(z)$ from $\mathbb{D} \times \mathbb{C}$ to \mathbb{C} has the following properties:

- (i) for all $w \in \mathbb{D}$, f^w fixes 0, 1 and ∞ ,
- (ii) for all $z \in \mathbb{C}$, $w \mapsto f^w(z)$ is holomorphic (by Theorem 2.9), and
- (iii) for all $w \in \mathbb{D}$, $z \mapsto f^w(z)$ is injective.

Notice that when $\mu_0 = 0 \in L^\infty(\mathbb{C})$, then $f^0(z) = z$ for all $z \in \mathbb{C}$: in that case, the map $(z, w) \rightarrow f^w(z)$ is called a *holomorphic motion of \mathbb{C}* . In our general case, we shall call $f^w(z)$ the *holomorphic motion associated to the family $w \mapsto \mu_w$* . It is classical that the restriction to bounded sets of \mathbb{C} of a holomorphic motion parametrized by the

unit disc is Hölder with exponent $\frac{1-|w|}{1+|w|}$ (cf. for instance [12, p. 13]). We give a proof of this statement in our situation.

For $r > 0$, denote by D_r the euclidean disc of radius r : $D_r = \{z, |z| \leq r\}$.

Proposition 4.7. *Let $w \mapsto \mu_w$ be a holomorphic family of Beltrami differentials such that $\|\mu_w\|_\infty \leq k < 1$ for all $w \in \mathbb{D}$ and let f^w be the holomorphic motion associated to this family. Let $r > 0$. Then there exists a constant $C = C(r, k)$ such that if z_1 and $z_2 \in D_r$, then for all $w \in \mathbb{D}$, one has*

$$|f^w(z_1) - f^w(z_2)| \leq C |f^0(z_1) - f^0(z_2)|^{\frac{1-|w|}{1+|w|}}.$$

Proof. It follows from the holomorphic dependence in w of the map $w \mapsto f^w(z)$, that for any distinct points z_1 and z_2 , the function $w \mapsto \log |f^w(z_1) - f^w(z_2)|$ is harmonic. Suppose $|z_i| \leq r$, for $i = 1, 2$. Since $\|\mu_w\|_\infty \leq k$, the continuity of the solutions to the Beltrami equation (Proposition 2.2) implies that there is a constant $D = D(r, k)$ such that $|f^w(z_1) - f^w(z_2)| \leq D$. Therefore the function

$$w \mapsto -\log \left| \frac{f^w(z_1) - f^w(z_2)}{D} \right|$$

is a positive harmonic function on \mathbb{D} . By the Harnack inequality [16], the values of a positive harmonic function on \mathbb{D} are controlled by its value at 0. In our situation, this gives

$$-\log \left| \frac{f^w(z_1) - f^w(z_2)}{D} \right| \geq -\frac{1-|w|}{1+|w|} \log \left| \frac{f^0(z_1) - f^0(z_2)}{D} \right|.$$

Applying the exponential map to both sides of this inequality gives the required result where the constant $C(r, k)$ is the supremum of $D^{\frac{2|w|}{1+|w|}}$, that is $\sup(D, 1)$. \square

Remark 4.8. If $\mu_0 = 0$, then this is a classical result about holomorphic motions. Notice however that if one does not assume that the differentials μ_w satisfy $\|\mu_w\| \leq k$ for some $k < 1$, then it could happen that the diameter of the image $f^w(D_r)$ is not bounded independently of r : this would prevent the constant C to be independent of w . For instance, the holomorphic motion $(w, z) \mapsto \frac{z+w\bar{z}}{1+w}$ does not satisfy the conclusions of the Proposition 4.7 when w tends to -1 (compare with [12, p. 13]).

Here is now a useful consequence of the Proposition 4.7. Recall the notation: for $\varepsilon > 0$, $B^\varepsilon = \{\mu \in L^\infty(\mathbb{C}) \mid \|\mu\| < \varepsilon\}$.

Proposition 4.9. *Let $r > 0$. For any $0 < \beta < 1$, there exists $\varepsilon(\beta) > 0$ such that for $\mu \in B^{\varepsilon(\beta)}$ one has*

- (i) *the restriction $f^\mu|_{D_r}$ belongs to $\text{Lip}^\beta(D_r)$, and its β -Hölder norm is bounded in terms of r and β ;*
- (ii) *the map $\mu \mapsto f^\mu|_{D_r}$ from B^ε to the Banach space $\text{Lip}^\beta(D_r)$ is holomorphic.*

Proof. Set $\varepsilon(\beta) = \frac{1-\beta}{2(1+\beta)}$. Applying Proposition 4.7 to the (linear) holomorphic family $w \mapsto w \frac{\mu}{2\|\mu\|}$, we deduce that for $\|\mu\| \leq \varepsilon(\beta)$, the restriction $f^\mu|_{D_r}$ belongs to $\text{Lip}^\beta(D_r)$ and its β -Hölder norm is smaller than a constant C which depends only on r and β . This proves (i).

In order to prove (ii), we consider an affine complex disc $w \mapsto \mu_w = \mu_0 + w\mu_1$ defined for $w \in \mathbb{D}$ and with image contained in $B^{\varepsilon(\beta)}$. We denote f^w the holomorphic motion which corresponds to μ_w . For $z \in D_r$, $w \mapsto f^w(z)$ is holomorphic. Therefore one can write $f^w(z) = \sum F_n(z)w^n$, the convergence being uniform in w over the disc $|w| \leq \rho < 1$ for all $\rho < 1$. The Cauchy formula for the n -th derivative of $f^w(z_1) - f^w(z_2)$ at $w = 0$ gives

$$F_n(z_1) - F_n(z_2) = \frac{1}{2i\pi} \int_{|w|=\rho} \frac{(f^w(z_1) - f^w(z_2))}{w^{n+1}} dw.$$

Recall that for all $w \in \mathbb{D}$, the restriction $f^w|_{D_r}$ belongs to $\text{Lip}^\beta(D_r)$, with Hölder norm at most C . The Cauchy formula implies therefore that the functions F_n are β -Hölder over D_r with norm at most $\frac{C}{\rho^n}$. Letting ρ tend to 1, we get that $F_n \in \text{Lip}^\beta(D_r)$ with norm at most C . Therefore, the series $\sum_{\mathbb{N}} F_n(\cdot)w^n$ converges absolutely in $\text{Lip}^\beta(D_r)$ uniformly on compact sets in \mathbb{D} . This proves that the map $\mu \mapsto f^\mu|_{D_r}$ is holomorphic in restriction to any affine complex disc contained in B^ε . Therefore $\mu \mapsto f^\mu|_{D_r}$ is holomorphic on $B^{\varepsilon(\beta)}$. □

Remark 4.10. For the linear family of Beltrami differentials $w \mapsto \mu_w = w.\mu$, the derivative at $w = 0$ of the map $w \mapsto f^w(z)$ is given by

$$\frac{d}{dw} f^w(z)|_0 = -\frac{1}{\pi} \int_{\mathbb{C}} \mu(\xi) \frac{z(z-1)}{\xi(\xi-1)(\xi-z)} dv_{eu}(\xi).$$

It is also a classical fact (cf. [11, p. 73]) that on any compact set of \mathbb{C} , the integral on the right is a continuous function of z with modulus of continuity $t \log \frac{1}{t}$. This is consistent with Proposition 4.9 (ii), which implies that the restriction of $z \mapsto \frac{d}{dw} f^w(z)|_0$ to any disc D_r – as well as the restrictions of all the derivatives of higher order – belongs to $\text{Lip}^\beta(D_r)$, for all $\beta < 1$.

4.2 Continuity properties of derivatives of Hölder functions

The second ingredient in the proof of Theorems 4.3 and 4.1 is the next result of Real Analysis. It is the key ingredient.

Recall that the rectangle R is the product $I \times J$ of two disjoint intervals of \mathbb{R} . Let $r > 0$ be such that I and J are contained in the interior of the euclidean disc D_r .

Proposition 4.11. *Let $0 < \alpha < 1$ and let $1 - \alpha/2 < \beta < 1$. Let $\varepsilon(\beta)$ be the constant provided by Proposition 4.9. Then, if $\mu \in B^{\varepsilon(\beta)}$, one has*

- (1) *the distribution $\tilde{\mathcal{L}}(\mu)$ belongs to $\mathcal{H}^\alpha(R)$, and*

(2) $\|\tilde{\mathcal{L}}(\mu) - \tilde{\mathcal{L}}(0)\| \leq C\|\mu\|_\infty$ for a constant C independent of μ .

Proof. To prove (i), we need to show that if μ satisfies the hypothesis and if $h \in C^2(\mathbb{R})$, then

$$\left| \int_{\mathbb{R}} \frac{\partial^2 h}{\partial \xi \partial \zeta} \log(f^\mu(\zeta) - f^\mu(\xi)) d\xi d\zeta \right| \leq C\|h\|_\alpha$$

for a constant C which does not depend on h .

By Proposition 4.9, the restriction $f^\mu|_{D_r}$ belongs to $\text{Lip}^\beta(D_r)$. One can approximate f^μ by a sequence of C^2 -maps $f_i: \mathbb{R} \rightarrow \mathbb{C}$ such that the restrictions $f_i|_{D_r}$ converge to $f^\mu|_{D_r}$ in $\text{Lip}^\beta(D_r)$. One has

$$\begin{aligned} \int_{\mathbb{R}} \frac{\partial^2 h}{\partial \xi \partial \zeta} \log(f^\mu(\zeta) - f^\mu(\xi)) d\xi d\zeta &= \lim_{i \rightarrow \infty} \int_{\mathbb{R}} \frac{\partial^2 h}{\partial \xi \partial \zeta} \log(f_i(\zeta) - f_i(\xi)) d\xi d\zeta \\ &= \lim_{i \rightarrow \infty} \int_{\mathbb{R}} \frac{h(\xi, \zeta)}{(f_i(\zeta) - f_i(\xi))^2} f'_i(\xi) f'_i(\zeta) d\xi d\zeta. \end{aligned}$$

We can suppose $\alpha < \frac{2}{3}$, so that in particular, $\beta < \alpha$. Since the β -Hölder norm of $f^\mu|_{D_r}$ is bounded independently of μ , due to Proposition 4.9, the same is also true for the α -Hölder norm of $f^\mu|_{D_r}$. Therefore, the α -Hölder norm of $k(\xi, \zeta) = \frac{h(\xi, \zeta)}{(f_i(\xi) - f_i(\zeta))^2}$ is bounded from above in terms of the α -Hölder norm of h : $\|k\|_\alpha \leq C \cdot \|h\|_\alpha$ for a constant C which is independent of h . Thus (i) will be a consequence of the following result. Denote by I_r the interval $] -r, r[$, which certainly contains $I \cup J$.

Proposition 4.12. *Let $0 < \alpha < 1$ and let $1 - \alpha/2 < \beta < 1$. Let $f \in \text{Lip}^\beta(I_r)$. Then there is a constant C which depends only on α and β such that for $k \in C^2(\mathbb{R})$, one has*

$$\left| \int_{\mathbb{R}} \frac{\partial^2 k}{\partial \xi \partial \zeta} f(\xi) f(\zeta) d\xi d\zeta \right| \leq C\|k\|_\alpha \|f\|_\beta.$$

This proposition means that if $f \in \text{Lip}^\beta(I_r)$, the distribution $\frac{\partial^2(f(\xi)f(\zeta))}{\partial \xi \partial \zeta}$ then belongs to $\mathcal{H}^\alpha(\mathbb{R})$ and that its norm is bounded by $C\|f\|_\beta$. In order to prove this proposition, we begin with a 1-dimensional version of it. Denote by $C^1(I)$ the set of functions $g: \mathbb{R} \rightarrow \mathbb{C}$ which are C^2 and with support contained in I .

Proposition 4.13. *Let $\alpha' < 1$ and $1 - \alpha' < \beta < 1$. Let $f \in \text{Lip}^\beta(I_r)$. Then there is a constant C which depends only on α' and β such that for $k \in C^1(I)$, one has*

$$\left| \int_{\mathbb{R}} k'(\xi) f(\xi) d\xi \right| \leq C\|k\|_{\alpha'} \|f\|_\beta.$$

Proof. After adding to k a constant function, we obtain a C^1 -function g on $\overline{\mathbb{R}}$ whose α -Hölder norm is smaller than $2\|k\|_\alpha$ and such that $\int g(\xi) d\xi = 0$. We are led to prove the inequality with g instead of k . Since $\int g(\xi) d\xi = 0$, there is a function $G: \overline{\mathbb{R}} \rightarrow \mathbb{C}$ such that $G'(\xi) = g(\xi)$ (the circle $\overline{\mathbb{R}}$ is identified with \mathbb{R}/\mathbb{Z}). We can

normalize it so that $\int G(\xi) d\xi$. Choose also a C^∞ -function $\chi : I_r \rightarrow I_r$ which has compact support and is equal to 1 on I . Replacing f by χf does not change the value of $\int_{\mathbb{R}} g'(\xi) f(\xi) d\xi$. Also $\|\chi f\|_\beta \leq C \|f\|_\beta$ for a constant C depending only on r .

Recall now the definition of the Zygmund spaces Λ^t for $t \in]0, 2[$ (these spaces can be defined for any real value of t but we shall use them only for $t \in]0, 2[$).

Definition 4.14 (Zygmund spaces; [19], [20]). Let $t \in]0, 2[$. A function $H : \overline{\mathbb{R}} \rightarrow \mathbb{C}$ is t -Zygmund if it is continuous and satisfies for all ξ and $\theta \in \overline{\mathbb{R}}$:

$$|H(\xi + \theta) + H(\xi - \theta) - 2H(\xi)| \leq A|\theta|^t.$$

The space of the t -Zygmund functions which satisfy $\int H(\xi) d\xi = 0$ is denoted $\Lambda^t(\overline{\mathbb{R}})$; it is a Banach space for the Zygmund norm

$$\sup_{\xi} |H(\xi)| + \sup_{\xi, \theta} \left| \frac{H(\xi + \theta) + H(\xi - \theta) - 2H(\xi)}{|\theta|^t} \right|.$$

When $t \in]0, 1[$, the space $\Lambda^t(\overline{\mathbb{R}})$ equals $\text{Lip}^t(\overline{\mathbb{R}})$, and the two norms, Hölder and Zygmund, are equivalent: we shall use the same notation $\|\cdot\|_t$ for the Zygmund norm.

When $t > 1$, it is classical that $G \in \Lambda^t(\overline{\mathbb{R}})$ if and only if G is differentiable and its derivative G' belongs to $\Lambda^{t-1}(\overline{\mathbb{R}}) = \text{Lip}^{t-1}(\overline{\mathbb{R}})$.

Since $g \in \text{Lip}^{\alpha'}(\overline{\mathbb{R}})$, $G \in \Lambda^{1+\alpha'}(\overline{\mathbb{R}})$ and $\|G\|_{1+\alpha'} \leq C \|g\|_\alpha$ for a constant C which depends only on α' .

One has

$$\int_{\mathbb{R}} g'(\xi) f(\xi) d\xi = \int_{\mathbb{R}} G''(\xi) f(\xi) d\xi = - \int_{\mathbb{R}} \Delta G(\xi) f(\xi) d\xi,$$

where Δ denotes the Laplace–Beltrami operator on $\overline{\mathbb{R}}$, identified with \mathbb{R}/\mathbb{Z} , i.e. simply $\Delta G = -G''$. Like on any compact manifold, the operator Δ is a self-adjoint operator densely defined on $L^2(\overline{\mathbb{R}})$ and positive ($\int G(\xi) \Delta G(\xi) d\xi > 0$ for any C^∞ function G). This allows us to define, for any real number γ , the fractional power Δ^γ , as a positive self-adjoint operator. It is a classical theorem of Real Analysis that the fractional power Δ^γ acts on the Zygmund spaces of exponent t by shifting the exponent t : precisely if $t - 2\gamma > 0$, then Δ^γ maps isomorphically $\Lambda^t(\overline{\mathbb{R}})$ to $\Lambda^{t-2\gamma}(\overline{\mathbb{R}})$ (cf. [19, p. 145], [20]).

The hypothesis on α' and β implies the existence of a γ such that $1 + \alpha' - 2\gamma > 0$ and $\beta - 2(1 - \gamma) > 0$. Since $\Delta^{1-\gamma}$ is self-adjoint

$$\int_{\mathbb{R}} \Delta G(\xi) f(\xi) d\xi = \int_{\mathbb{R}} \Delta^\gamma G(\xi) \Delta^{1-\gamma} f(\xi) d\xi.$$

As

$$\|\Delta^\gamma G\|_\infty \leq \|\Delta^\gamma G\|_{1+\alpha'-2\gamma} \leq C_\gamma \|g\|_{\alpha'}$$

and

$$\|\Delta^{1-\gamma} f\|_\infty \leq \|\Delta^{1-\gamma} f\|_{\beta-2(1-\gamma)} \leq C_{1-\gamma} \|f\|_\beta,$$

one has $\left| \int_{\mathbb{R}} g'(\xi) f(\xi) d\xi \right| \leq C \cdot \|g\|_{\alpha'} \|f\|_{\beta}$ for a constant C which is independent of f and g . \square

To prove Proposition 4.12, we write

$$\int_{\mathbb{R}} \frac{\partial^2 k(\xi, \zeta)}{\partial \xi \partial \zeta} f(\xi) f(\zeta) d\xi d\zeta = \int_I \frac{\partial}{\partial \xi} \left(\int_J \frac{\partial k}{\partial \zeta} f(\zeta) d\zeta \right) f(\xi) d\xi. \quad (4.2)$$

For any $\xi \in J$, the function $k(\xi, \cdot)$ is in $C^2(J)$ and its α -Hölder norm is bounded by $\|k\|_{\alpha}$. The β -Hölder norm of $f(\xi, \cdot)$ is also bounded by $\|f\|_{\beta}$. Therefore, applying Proposition 4.13 with $\alpha' = \alpha$ and using that $\beta > 1 - \frac{\alpha}{2} > 1 - \alpha$, one obtains, for all $\xi \in I$

$$\left| \int_J \frac{\partial k(\xi, \zeta)}{\partial \zeta} f(\zeta) d\zeta \right| \leq C \|k\|_{\alpha} \|f\|_{\beta}.$$

To be able to apply Proposition 4.13, but to the variable ξ , we shall use the following

Claim 4.15. *Let $k \in \text{Lip}^{\alpha}(I \times J)$. Then for any $\xi \in I$, the restriction $k(\xi, \cdot)$ belongs to $\text{Lip}^{\frac{\alpha}{2}}(J)$ and the map $\xi \mapsto k(\xi, \cdot)$, from I to $\text{Lip}^{\frac{\alpha}{2}}(J)$, is $\frac{\alpha}{2}$ -Hölder with norm smaller than $2\|k\|_{\alpha}$.*

Proof. Recall that the distance on $I \times J$ we work with is $d((\xi_1, \zeta_1), (\xi_2, \zeta_2)) = \sup(|\xi_2 - \xi_1|, |\zeta_2 - \zeta_1|)$. We need to show that for any two points (ξ_1, ζ_1) and $(\xi_2, \zeta_2) \in I \times J$ with $\zeta_1 \neq \zeta_2$, one has

$$\left| \frac{(k(\xi_2, \zeta_2) - k(\xi_1, \zeta_2)) - (k(\xi_2, \zeta_1) - k(\xi_1, \zeta_1))}{|\zeta_2 - \zeta_1|^{\frac{\alpha}{2}}} \right| \leq 2\|k\|_{\alpha} |\xi_2 - \xi_1|^{\frac{\alpha}{2}}. \quad (4.3)$$

If $|\xi_2 - \xi_1| \leq |\zeta_2 - \zeta_1|$, then

$$\begin{aligned} |(k(\xi_2, \zeta_2) - k(\xi_1, \zeta_2)) - (k(\xi_2, \zeta_1) - k(\xi_1, \zeta_1))| &\leq 2\|k\|_{\alpha} |\xi_2 - \xi_1|^{\alpha} \\ &\leq 2\|k\|_{\alpha} |\zeta_2 - \zeta_1|^{\frac{\alpha}{2}} |\xi_2 - \xi_1|^{\frac{\alpha}{2}}, \end{aligned}$$

from which Inequality 4.3 follows.

If $|\zeta_2 - \zeta_1| \leq |\xi_2 - \xi_1|$, then

$$\begin{aligned} &|(k(\xi_2, \zeta_2) - k(\xi_1, \zeta_2)) - (k(\xi_2, \zeta_1) - k(\xi_1, \zeta_1))| \\ &= |(k(\xi_2, \zeta_2) - k(\xi_2, \zeta_1)) - (k(\xi_1, \zeta_2) - k(\xi_1, \zeta_1))| \\ &\leq 2\|k\|_{\alpha} |\zeta_2 - \zeta_1|^{\alpha} \\ &\leq 2\|k\|_{\alpha} |\xi_2 - \xi_1|^{\frac{\alpha}{2}} |\zeta_2 - \zeta_1|^{\frac{\alpha}{2}}, \end{aligned}$$

proving the claim in that case also. \square

Now, Claim 4.15 and Proposition 4.13 applied with $\alpha' = \frac{\alpha}{2}$ imply Proposition 4.12. \square

As we said before, Proposition 4.12 implies Proposition 4.11 (i). It implies also (ii), using the same argument and since for $\mu \in B^{\varepsilon(\beta)}$, one has $\|(f^\mu - \text{Id})|_{D_r}\|_\beta \leq C\|\mu\|_\infty$ (Proposition 4.9).

4.3 Proof of Theorems 4.3, 4.1 and 4.2

Proof of Theorem 4.3. To prove Theorem 4.3 we shall first construct, for any point $[\mu_0] \in \mathcal{T}(\Gamma)$ a neighborhood $\mathcal{N}_{[\mu_0]} \subset \mathcal{QF}(\Gamma)$ such that, if $\tilde{\mathcal{L}}$ denotes the map constructed in Corollary 4.6, the restriction $\tilde{\mathcal{L}}|_{\mathcal{N}_{[\mu_0]}}$ has image contained in $\mathcal{H}^\alpha(R)$ and is a holomorphic map from $\mathcal{N}_{[\mu_0]}$ to $\mathcal{H}^\alpha(R)$.

We first consider the case when $[\mu_0] = [0]$. The proof in that case uses the following result.

Proposition 4.16. *Let R be a rectangle contained in \mathfrak{g}^f . Let $\alpha < 1$ and $\beta > 1 - \frac{\alpha}{2}$; let $\varepsilon(\beta)$ be the constant provided by Proposition 4.9. Then the restriction of \mathcal{L} to $B^{\varepsilon(\beta)} \cap \mathcal{T}(\Gamma)$ extends to a holomorphic map $\tilde{\mathcal{L}}: B^{\varepsilon(\beta)} \rightarrow \mathcal{H}^\alpha(R)$.*

Proof. Corollary 4.6 provides an extension of \mathcal{L} . It is the map $\tilde{\mathcal{L}}: B^1(\Gamma) \rightarrow \mathcal{D}(R)$ defined by

$$\tilde{\mathcal{L}}([\mu]) = \frac{\partial^2 \log(f^\mu(\zeta) - f^\mu(\xi))}{\partial \xi \partial \zeta},$$

where the derivative is taken in the sense of distributions. By Proposition 4.11, if $\mu \in B^{\varepsilon(\beta)}$, then $\mathcal{L}([\mu]) \in \mathcal{H}^\alpha(R)$. We are going to show that this map $\tilde{\mathcal{L}}$ is holomorphic. Let $w \mapsto \mu_w$ be an affine disc $w \mapsto \mu_0 + w\mu_1$ parameterized by the disc \mathbb{D} with image contained in $B^{\varepsilon(\beta)}$. Let $h \in C^2(R)$. For all $w \in \mathbb{D}$ and r such that $|w| < r < 1$, one can write:

$$\begin{aligned} \langle \tilde{\mathcal{L}}([\mu_w]), h \rangle &= \int_R \frac{\partial^2 h}{\partial \xi \partial \zeta} \log(f^w(\zeta) - f^w(\xi)) \, d\xi d\zeta \\ &= \int_R \frac{\partial^2 h}{\partial \xi \partial \zeta} \left(\frac{1}{2i\pi} \int_{|z|=r} \frac{1}{z-w} \log(f^z(\zeta) - f^z(\xi)) \, dz \right) \, d\xi d\zeta \\ &= \frac{1}{2i\pi} \int_{|z|=r} \frac{1}{z-w} \left(\int_R \frac{\partial^2 h}{\partial \xi \partial \zeta} \log(f^z(\zeta) - f^z(\xi)) \, d\xi d\zeta \right) \, dz \\ &= \frac{1}{2i\pi} \int_{|z|=r} \frac{1}{z-w} \langle \tilde{\mathcal{L}}([\mu_z]), h \rangle \, dz, \end{aligned}$$

where the second equality comes from the fact that $\log(f^w(\zeta) - f^w(\xi))$ is holomorphic in w and the third from the fact that it is a continuous function of the three variables w, ξ and ζ . By Proposition 4.11, the norm in $\mathcal{H}^\alpha(R)$ of $\tilde{\mathcal{L}}([\mu_z])$ is bounded independently of $z \in \mathbb{D}$. Therefore one has

$$|\langle \tilde{\mathcal{L}}([\mu_w]) - \tilde{\mathcal{L}}([\mu_{w'}]), h \rangle| \leq C|w - w'| \|h\|_\alpha$$

when $|w|$ and $|w'|$ are smaller than $r_0 < r < 1$; here C is a constant which only depends on r_0, r and β . It follows that $w \mapsto \tilde{\mathcal{L}}([\mu_w])$ is continuous as a map from \mathbb{D} to $\mathcal{H}^\alpha(R)$. In particular, the integral $\int_{|z|=r} \frac{1}{z-w} \tilde{\mathcal{L}}([\mu_z]) dz$ is well defined as an integral in $\mathcal{H}^\alpha(R)$ and one has $\tilde{\mathcal{L}}([\mu_w]) = \int_{|z|=r} \frac{1}{z-w} \tilde{\mathcal{L}}([\mu_z]) dz$. It follows that $w \rightarrow \tilde{\mathcal{L}}([\mu_w])$ is holomorphic on \mathbb{D} . Therefore the map $\tilde{\mathcal{L}}: B^{\varepsilon(\beta)} \rightarrow \mathcal{H}^\alpha(R)$ is holomorphic. \square

Fix β with $1 - \frac{\alpha}{2} < \beta < 1$.

The harmonic chart identifies a neighborhood of $[0]$ in $\mathcal{QF}(\Gamma)$ with $\mathbb{B}^1(\Gamma)$, the unit ball in the space of the harmonic Beltrami differentials for Γ . We denote by B^r and \mathbb{B}^r the balls of radius r in $B(\Gamma)$ and $\mathbb{B}(\Gamma)$ respectively. Using Proposition 4.16, one can take as neighborhood of $[0]$, $\mathcal{N}_{[0]} = \mathbb{B}^{\varepsilon(\beta)}(\Gamma)$, where $\varepsilon(\beta)$ is the constant provided by Proposition 4.9. To construct the neighborhood $\mathcal{N}_{[\mu_0]}$ of an arbitrary point $[\mu_0] \in \mathcal{T}(\Gamma)$, we use the change of basepoint map (cf. Definition 2.7) as in the proof of Theorem 3.12. This map is a homeomorphism $[\nu]: \mathcal{T}(\Gamma) \rightarrow \mathcal{T}(\Gamma^{\mu_0})$ which maps $[\mu_0] \in \mathcal{T}(\Gamma)$ to $[0] \in \mathcal{T}(\Gamma^{\mu_0})$. We noticed during the proof of Theorem 3.12 the relation $\mathcal{L} = \mathcal{H}(f^{\mu_0}) \circ \mathcal{L}^{\mu_0} \circ [\nu]$. Let $\delta > 0$ be such that $(f^{\mu_0})^{-1}$ is δ -Hölder: this implies that $\mathcal{H}(f^{\mu_0})$ is a continuous linear map from $\mathcal{H}^{\alpha\delta}(F^{\mu_0}(R))$ to $\mathcal{H}^\alpha(R)$. Let $\beta' > 1 - \frac{\alpha\delta}{2}$. By Proposition 4.16, $\mathcal{L}^{\mu_0}: B^{\varepsilon(\beta')} \rightarrow \mathcal{H}^{\alpha\delta}(F^{\mu_0}(R))$ is holomorphic. The formula for ν given in Definition 2.7 shows that $[\nu]$ is holomorphic. It follows that if one sets $\mathcal{N}_{[\mu_0]} = [\nu]^{-1}(\mathbb{B}^{\varepsilon(\beta')})$, then $\tilde{\mathcal{L}}: \mathcal{N}_{[\mu_0]} \rightarrow \mathcal{H}^\alpha(R)$ is holomorphic.

To finish the proof of Theorem 4.3, it suffices to observe that if two holomorphic extensions $\tilde{\mathcal{L}}_1$ and $\tilde{\mathcal{L}}_2$ of \mathcal{L} are defined on a connected open neighborhood \mathcal{U} of $[\mu_0] \in \mathcal{T}(\Gamma)$, then they are equal: $\tilde{\mathcal{L}}_1 = \tilde{\mathcal{L}}_2$. This fact, which is obvious when Γ is cocompact – since then $\mathbb{B}(\Gamma)$ is a complex space of dimension n and $\mathcal{T}(\Gamma)$ totally real of real dimension n – can be seen in general using the local charts for the complex structure of $\mathcal{QF}(\Gamma)$. Recall the isomorphism $\mathbb{B}^1(\Gamma^{\mu_0}) \simeq \mathbb{B}_{\mathcal{T}}^1(\Gamma^{\mu_0}) \times \mathbb{B}_{\mathcal{T}}^1(\Gamma^{\mu_0})$ (cf. Definition 2.14). Therefore we can suppose that \mathcal{U} is a connected open neighborhood of $(0, 0) \in \mathbb{B}^1(\Gamma^{\mu_0}) \times \mathbb{B}^1(\Gamma^{\mu_0})$; the intersection $\mathcal{U} \cap \mathcal{T}(\Gamma)$ is then the set of pairs $(\mu_1, \mu_2) \in \mathcal{U}$ with $\mu_2 = \bar{\mu}_1$. Any point $(\mu_1, \mu_2) \in \mathbb{B}_{\mathcal{T}}^1(\Gamma^{\mu_0}) \times \mathbb{B}_{\mathcal{T}}^1(\Gamma^{\mu_0})$ belongs to a \mathbb{C} -affine plane which is the complexification of an \mathbb{R} -affine line contained in $\mathcal{T}(\Gamma^{\mu_0})$: precisely (μ_1, μ_2) sits in the affine plane which is the complexification of the real line passing through $\frac{(\mu_1 + \bar{\mu}_2, \mu_2 + \bar{\mu}_1)}{2}$, directed by the vector $i(\mu_1 - \bar{\mu}_2, \mu_2 - \bar{\mu}_1)$. In particular, for $\varepsilon > 0$ sufficiently small, any pair $(\mu_1, \mu_2) \in \mathcal{U}$ with $\|\mu_i\|_\infty \leq \varepsilon$ belongs to such an affine disc which is contained in \mathcal{U} . The definition of a holomorphic map implies that $\tilde{\mathcal{L}}_{\mathcal{F}} = \tilde{\mathcal{L}}_G$ on the open set $\{(\mu_1, \mu_2) \mid \|\mu_i\|_\infty < \varepsilon\}$. Since \mathcal{U} is connected, $\tilde{\mathcal{L}}_{\mathcal{F}} = \tilde{\mathcal{L}}_G$ on \mathcal{U} .

The open neighborhood \mathcal{N}^α can now be constructed easily. \square

Remark 4.17. It follows from the proof that the image $\tilde{\mathcal{L}}(B^{\varepsilon(\beta)})$ is bounded in $\mathcal{H}^\alpha(R)$: its diameter is bounded in terms of α, β , and of the size of the rectangle $R = I \times J$, i.e. by an upper bound on the lengths of the intervals I, J and upper and lower bounds on the distance between them.

Proof of Theorem 4.1. Recall the notations introduced during the proof of Theorem 3.12. Given a rectangle $R \in \mathcal{R}$, not necessarily contained in \mathcal{G}^f , $q \in \text{PSL}(2, \mathbb{R})$ is the map sending the reference rectangle R_0 to R . Recall that R_0 is a rectangle contained in \mathcal{G}^f , so that Theorem 4.3 applies to R_0 . For $\mu \in B^1(\Gamma)$, $q^\mu \in \text{PSL}(2, \mathbb{R})$ is the element such that the quasiconformal homeomorphism $\phi^\mu = q^\mu \circ f^\mu \circ q$ is normalized. Equation 3.1, for any $h \in \text{Lip}_R^\alpha(\mathcal{G})$, one has

$$\langle \mathcal{L}([\mu]), h \rangle = \int_{\mathcal{G}} (h \circ Q) \circ (\Phi^\mu)^{-1} d\lambda = \langle \mathcal{L}([q^*(\mu)]), h \circ Q \rangle.$$

By the definition of $\tilde{\mathcal{L}}$, one also has $\langle \tilde{\mathcal{L}}(\mu), h \rangle = \langle \tilde{\mathcal{L}}(q^*(\mu)), h \circ Q \rangle$. Since the neighborhood $\mathcal{N}_{[0]}$ constructed in the proof of Theorem 4.3 is invariant under the map $\mu \mapsto q^*(\mu)$ for all $q \in \text{PSL}(2, \mathbb{R})$, one has $|\langle \tilde{\mathcal{L}}([\mu]), h \rangle| \leq C \|h \circ Q\|_\alpha$ the constant C being independent of h and $R \in \mathcal{R}$. By definition of the space of geodesic Hölder distributions, this means that $\tilde{\mathcal{L}}(\mathcal{N}_{[0]})$ is contained in $\mathcal{H}^\alpha(\Gamma)$. By the remark following the proof of Theorem 4.3, this image is also bounded. Therefore, the argument used for proving Proposition 4.16 shows that $\tilde{\mathcal{L}}: \mathcal{N}_{[0]} \rightarrow \mathcal{H}^\alpha(\Gamma)$ is holomorphic.

For a general point $[\mu_0] \in \mathcal{T}(\Gamma)$, one writes: $\phi^\mu = q^\mu \circ f^\mu \circ q = f^{v_q(\mu)} \circ f^{\mu_0(q)}$, with $f^{v_q(\mu)} = q^\mu \circ f^{v(\mu)} \circ (q^{\mu_0})^{-1}$ and $f^{\mu_0(q)} = q^{\mu_0} \circ f^{\mu_0} \circ q$. Therefore

$$\langle \tilde{\mathcal{L}}([q^*(\mu)]), h \circ Q \rangle = \langle \tilde{\mathcal{L}}[v_q(\mu)], h \circ Q \circ (F^{\mu_0(q)})^{-1} \rangle. \tag{4.4}$$

Now, since $\|q^*(\mu)\|_\infty = \|\mu\|_\infty$ the homeomorphisms $(f^{\mu_0(q)})^{-1}$ are δ -Hölder with the same exponent δ and with a δ -Hölder norm which is bounded independently of q . This implies that the norms of the linear maps $\mathcal{H}((f^{\mu_0(q)})^{-1}): \mathcal{H}^\alpha(F^{\mu_0(q)}(R_0)) \rightarrow \mathcal{H}^{\alpha\delta}(R_0)$ are bounded independently of q . Since the homeomorphisms $f^{\mu_0(q)}$ form an equicontinuous family, the shapes of the rectangles $F^{\mu_0(q)}(R_0)$ are also bounded independently of q . Let $\beta' > 1 - \frac{\alpha\delta}{2}$. The Beltrami coefficient $v_q(\mu)$ can be computed using the formula from Definition 2.7. It comes out that there is $\eta(\beta') > 0$ which depends only on β and on $\|\mu_0\|_\infty$ such that if $\|\mu\|_\infty \leq \eta(\beta')$, then $\|v_q(\mu)\|_\infty \leq \varepsilon(\beta')$ for all q . Therefore, using Theorem 4.3 and the remark following it, we obtain that if $\|\mu - \mu_0\|_\infty \leq \eta(\beta')$, then the term on the right in the equation 4.4 is bounded in absolute value by $C \|h \circ Q\|_\alpha$, where the constant C does not depend on q nor on h . Let $\mathcal{N}_{[\mu_0]}$ be the neighborhood of $[\mu_0] \in \mathcal{Q}\mathcal{F}(\Gamma)$ which is $\{[\mu] \mid \|\mu\|_\infty \leq \eta(\beta')\}$. Then $\tilde{\mathcal{L}}(\mathcal{N}_{[\mu_0]})$ is bounded in $\mathcal{H}^\alpha(\Gamma)$. Like above, one shows that $\tilde{\mathcal{L}}$ is holomorphic on $\mathcal{N}_{[\mu_0]}$. The neighborhood \mathcal{N}^α can now be constructed as in Theorem 4.3. \square

Proof of Theorem 4.2. By Theorem 4.1, \mathcal{L} is the restriction of a holomorphic map from a neighborhood of $\mathcal{T}(\Gamma)$. Its restriction to $\mathcal{T}(\Gamma)$ is therefore analytic [9]. \square

References

- [1] L. V. Ahlfors, *Lectures on quasiconformal mappings*. Van Nostrand Math. Stud. 10, D. Van Nostrand Company, Inc., Princeton 1966. [225](#)
- [2] L. V. Ahlfors, *Collected papers*. Vol. 2. Birkhäuser, Boston 1982. [225](#)
- [3] L. Ahlfors and L. Bers, Riemann's mapping theorem for variable metrics. *Ann. of Math.* (2) 72 (1960), 385–404. [225](#), [227](#)
- [4] L. Ahlfors and G. Weill, A uniqueness theorem for Beltrami equations. *Proc. Amer. Math. Soc.* 13 (1962), 975–978. [228](#)
- [5] L. Bers, Correction to “Spaces of Riemann surfaces as bounded domains”. *Bull. Amer. Math. Soc.* 67 (1961), 465–466. [227](#)
- [6] F. Bonahon, Bouts des variétés hyperboliques de dimension 3. *Ann. of Math.* (2) 124 (1986), 71–158. [229](#)
- [7] F. Bonahon, The geometry of Teichmüller space via geodesic currents. *Invent. Math.* 92 (1988), 139–162. [229](#)
- [8] F. Bonahon and Y. Sözen, Variation of the Liouville measure of a hyperbolic surface. *Ergodic Theory Dynam. Systems* 23 (2003), 729–758. [229](#), [232](#), [233](#)
- [9] N. Bourbaki, *Topologie différentielle*. Fascicule de résultats, Hermann, Paris 1967. [237](#), [247](#)
- [10] A. Douady and C. J. Earle, Conformally natural extension of homeomorphisms of the circle. *Acta Math.* 157 (1986), 23–48. [237](#)
- [11] F. P. Gardiner, *Teichmüller theory and quadratic differentials*. John Wiley & Sons, New York 1987. [225](#), [227](#), [228](#), [241](#)
- [12] F. P. Gardiner and N. Lakic, *Quasiconformal Teichmüller theory*. Math. Surveys. Monogr. 76, American Mathematical Society, Providence, R.I., 2000. [225](#), [227](#), [228](#), [240](#)
- [13] Y. Imayoshi and M. Taniguchi, *An introduction to Teichmüller spaces*. Springer-Verlag, Tokyo 1992. [225](#)
- [14] T. Iwaniec and G. Martin, *Geometric function theory and non-linear analysis*. Oxford Math. Monogr., Oxford University Press, Oxford 2001. [225](#)
- [15] J.-P. Otal, Riemann surfaces and Theta operators, *Quaderns* 25; <http://www.crm.es/Publications/Quaderns/Quadern25-2.pdf>. [225](#)
- [16] W. Rudin, *Real and complex analysis*. 3rd ed., McGraw-Hill Book Co., New York 1987. [230](#), [240](#)
- [17] D. Šarić, Infinitesimal Liouville distributions for Teichmüller space. *Proc. London Math. Soc.* (3) 88 (2004), 436–454. [229](#), [232](#), [233](#)
- [18] D. Šarić, Geodesic currents and Teichmüller space. *Topology* 44 (2005), 99–130. [229](#), [232](#), [233](#), [234](#)
- [19] E. M. Stein, *Singular integrals and differentiability properties of functions*. Princeton Math. Ser. 30, Princeton University Press, Princeton, N.J., 1970. [243](#)
- [20] H. Triebel, *Theory of function spaces. II*. Monographs Math. 84, Birkhäuser, Basel 1992. [243](#)

Chapter 5

**Teichmüller spaces, triangle groups and
Grothendieck dessins**

William J. Harvey

Contents

1	Introduction	250
2	Historical overview	251
3	Grothendieck dessins and Thurston's examples	255
	3.1 Dessins	255
	3.2 Clean dessins and Galois covers of \mathbb{P}^1	255
	3.3 Thurston's examples	256
4	Teichmüller theory	257
	4.1 Deformations of complex structure on a surface	258
	4.2 Teichmüller spaces and modular groups	260
	4.3 Families of Riemann surfaces	263
	4.4 Hyperbolic surfaces inside moduli space	265
	4.5 Example: a T-disc parametrising tori	267
	4.6 Deformation T-discs embed	268
	4.7 Veech's examples	268
	4.8 Hecke triangle groups and polygonal surface tilings	270
	4.9 Fermat curves represented by T-discs	272
	4.10 Regular tessellations and smooth coverings	273
	4.11 Return to Thurston's examples	274
5	Triangle groups and modular families	275
	5.1 Tessellations and triangle groups	275
	5.2 Associated Teichmüller families	276
	5.3 Recent work on Veech curves in \mathcal{M}_g	281
	5.4 Completion of a T-disc with large stabiliser	282
	5.5 Consequences; complete curves in modular varieties	283
6	Towards a space of all \mathcal{Q} -curves: a universal modular family	284
	6.1 A moduli space of Fuchsian groups	284
	6.2 A representation of the absolute Galois group	285
	6.3 Modular forms with closed trajectories	286
	6.4 Arithmetically defined curves in moduli space	288
	References	288

1 Introduction

The theory of moduli for algebraic curves began with Riemann's famous papers on abelian functions, but it took many years before a further ingredient, Teichmüller's ideas on how to deform the underlying topological surface so as to effect a change in the complex analytic structure, produced a rigorous formulation and construction of the *variety of moduli* \mathcal{M}_g , whose points represent bijectively each isomorphism class of non-singular curve of the given genus.

We begin by describing in outline the theory of moduli for complex algebraic curves. The special case of genus 1 is discussed briefly, both as an avatar and in order to describe some special types of object to be used later, classical modular forms and Riemann surfaces, but we concentrate on the curves of genus at least 2, whose Riemann surface is uniformised by the hyperbolic plane, using the elaborate machinery of Teichmüller spaces introduced by Ahlfors and Bers. A method of construction is then given for explicit complex analytic models within the metric structure of these moduli spaces, which exhibits through the medium of hyperbolic plane geometry a large class of curves including those with defining equation over some algebraic number field. These *Teichmüller disc* models of surfaces have achieved much prominence recently through work of W. Veech and others on dynamical properties exhibited by the set of closed geodesics on the surface, relating to interval exchange transformations and billiard trajectories on Euclidean polygons. Here we focus on the symmetry properties of the complex curves so defined and the existence of Belyi structures on them; in a later section we show that there is a collection \mathcal{B} which comprises a union of complex affine algebraic curves contained in the moduli space $\mathcal{M} = \bigcup \mathcal{M}_g$ of all (conformal classes of) compact Riemann surfaces (viewed as hyperbolic 2-dimensional orbifolds) which has an intrinsic arithmetical aspect. The space \mathcal{B} is a (reducible) analytic variety on which the absolute Galois group of the algebraic numbers $\text{Gal}(\overline{\mathbb{Q}})$ acts in a natural way.

To be more precise about the framework of the family $\mathcal{B}_g = \mathcal{B} \cap \mathcal{M}_g$, each member (irreducible component) is an immersed complex affine curve in some modular variety \mathcal{M}_g of nonsingular curves of genus $g > 1$, indexed by a central base point and unit tangent deformation vector, and the whole collection forms part of the tautological geometric structure inherited from the covering Teichmüller space and the holomorphic vector bundle $\Omega^2(\mathcal{V}_g)$ of holomorphic quadratic forms on the universal Teichmüller-family \mathcal{V}_g of Riemann surfaces. Veech refers to this structure, or its unit sphere subbundle, as the Teichmüller geodesic flow. The precise relationship between the space \mathcal{B}_g and the $\overline{\mathbb{Q}}$ -rational points of the moduli spaces \mathcal{M}_g , \mathcal{C}_g is unknown, but we show in Section 6 that for each curve in \mathcal{B}_g , after completion within the Deligne–Mumford compactification $\hat{\mathcal{M}}_g$, there is a ramified covering (with degree bounded above by a linear function of g) which belongs to the conformal isomorphism class of compact Riemann surface defined by the central point.

This survey is based on a revised version of the preprint [26]. There is little that is original here, beyond the selection of material and point of view taken, which

emphasizes wherever possible the hyperbolic geometry and group theory. However, later results presented without attribution are the author's responsibility.

2 Historical overview

The study of algebraic curves and their deformations has a venerable history, beginning with specific investigations of simple types of curve, and progressing to very sophisticated methods of projective geometry. With Riemann's introduction of the underlying topological surface and his spectacular results on theta functions and periods of abelian integrals (Abel–Jacobi theory) ([57]), it became clear that to answer the most fundamental questions about a given projective curve, it is necessary to consider variations of the curve, and in order to maintain the link with abelian integrals one must address the problem of how to vary the curve in all possible ways which preserve the particular topological structure. Riemann showed that the topological genus, the number of handles of the Riemann surface of the curve, dictates the dimension (1 if $g = 1$, $3g - 3$ when $g \geq 2$) of the *moduli space* of local deformations of the curve within the (complex) projective space in which it lies. However, the precise structure of the space of moduli remained an enigma until a more comprehensive account of the topological nature of surfaces became available, including the introduction of homotopy as a refinement of the relation of homology for integration theory along paths. Despite much progress on the analytic side with the proof of uniformisation and the classification of Fuchsian groups, the efforts of Klein, Fricke and (independently) Poincaré left the global moduli problem unresolved, and it was not until the introduction of quasiconformal mappings and their relation to quadratic differential forms by Teichmüller during the 1930s that a genuine theory of deformations of Riemann surfaces became possible. This was developed during the period after 1950 by Ahlfors and by Gerstenhaber and Rauch, but it was only after the landmark Princeton Conference on Analytic Functions of 1957 [56] that the complex deformation theory of Riemann surfaces took flight, together almost simultaneously with the Kodaira–Spencer theory for compact complex manifolds in general.

Grothendieck's introduction [22] of the notion of dessin d'enfant (defined below) brought into sharper focus the interaction between combinatorial surface theory, the complex geometry of algebraic curves and their algebraic number coefficient fields: it implies that the arithmetic field of definition for a curve is in some manner an intrinsic feature of the hyperbolic geometry underlying the transcendental uniformisation of the curve. A primary goal in this chapter is to illuminate some aspects of that perception by systematic use of the theory of Teichmüller spaces, as developed by Ahlfors and Bers; we shall highlight the relationship between a certain class of holomorphic quadratic differential forms on the Riemann surface of an algebraic curve (the Jenkins–Strebel forms) and the existence of a model of the surface (as a quotient of the hyperbolic plane) within the corresponding modular variety. The totality of these (immersed)

model surfaces is a set on which the absolute Galois group of the algebraic numbers acts, a pattern which is in accord with the conjectural picture of [22]. It may equally be viewed as an organised combinatorial framework for the natural actions of the various mapping class groups and subgroups which permute markings on the various levels of ramified surface coverings that constitute the branching locus, or the orbifold singularities, in each \mathcal{M}_g .

A *Grothendieck dessin* or *line drawing in a surface* is a connected graph drawn in a compact surface with the property that each complementary piece of surface is a polygonal disc. During the late 1970s, Grothendieck proposed that there was an intimate link between this purely combinatorial object and arithmetic geometry of curves, and pointed out the relationship between dessins and two distinct, previously unrelated notions: representations of a certain discrete (extended triangle) group on the one hand and the absolute Galois group \mathcal{G} of the field of algebraic numbers on the other.

In 1979, this insight received clear confirmation when G. V. Belyi proved a fascinating theorem [5] bringing out the full significance of this idea: *If a projective algebraic curve X is defined by a set of polynomial equations with coefficients in a number field, an algebraic extension of the rational field \mathbb{Q} , then there is a meromorphic function on X whose only singular values are the three points 0, 1 and ∞ .* The converse is also true, following indirectly from older work of A. Weil on the field of definition for an algebraic variety; for this aspect of the Belyi theorem, the reader may consult recent articles by J. Wolfart [74] and G. Gonzalez-Díez [18].

Grothendieck later wrote [22] an extended account of his ideas as part of a CNRS research proposal, placing it in an elaborate conjectural theory of ‘anabelian’ fundamental groups in geometry. This celebrated manuscript has received widespread unofficial distribution; subsequently, thanks partly to an influential article by G. Shabat and V. A. Voevodsky [64] in the Grothendieck Festschrift, the notion of dessin gained a wider audience and established itself as an independent entity on the fringe of three broad estates – algebraic geometry, number theory and complex analytic geometry. Each of these areas demands heavy investment in basic language and background theory as prerequisites for understanding and progress, and so this insertion of a new elementary structural connection between them has been widely welcomed.

To display the link, we focus on the relationship between a subclass of dessins and a special kind of deformation of the Riemann surface structure. The latter arises from the initial data of a complete complex algebraic curve, carrying the dessin as part of its uniformisation in the hyperbolic plane, and an additional choice of prescribed flat geometric pattern coming via the Belyi ramified covering from the Euclidean plane. These two pieces of data determine a complex affine deformation curve (Riemann surface with punctures), immersed in the corresponding modular variety and passing through a central point which represents the base point of the deformation, the compact Riemann surface associated with the complete curve. It will follow from the basic link between automorphic forms and holomorphic deformations that any point of a modular variety which represents a curve definable over $\overline{\mathbb{Q}}$ belongs in an appropriate

sense to at least one such family. As seen from the perspective of Belyi's theorem, a primary source of this type of structure comes from curves with sufficiently large automorphism group. Simple examples of familiar curves falling into this category include Wiman's class of hyperelliptic curves with affine equation $y^2 = 1 - x^n$, for $n \geq 5$, the famous Klein quartic given (in projective form) by $x^3y + y^3z + z^3x = 0$ and the Fermat curves $x^n + y^n = z^n$, for $n \geq 4$.

The theory of holomorphic families of compact Riemann surfaces is founded on the analytic properties of the Teichmüller spaces $\mathcal{T}_{g,n}$ established by Lars Ahlfors and Lipman Bers during the decade following 1960; these spaces are a countable set of connected complex manifolds, one for each pair of non-negative integers g, n with $2g - 2 + n > 0$, whose points represent marked complex structures on an n -pointed genus g surface up to holomorphic equivalence. There is a small list of special low values of g, n which represent surfaces and deformation spaces familiar from classical work: $\mathcal{T}_{0,n}$ is a single point for $n = 1, 2, 3$ and $\mathcal{T}_{0,4} = \mathcal{U} = \mathcal{T}_{1,0}$ is the upper half plane. In this way, the totality of compact Riemann surfaces (non-singular curves) with n -point subsets is organised into a union of disjoint spaces representing distinct topological types. Each Teichmüller space carries intrinsically a structure of complex manifold of dimension $3g - 3 + n$ and a complete global metric, defined by Teichmüller as the logarithm of the least overall conformal distortion (measured in the sup norm) involved in deforming one complex structure on a surface to another. If the surface under consideration has genus 1, the deformation space is equivalent to the classical identification of (conformal classes of) marked tori, $X = \mathbb{C}/L$, with points of the upper half plane, $\mathcal{U} = \mathcal{T}_1$, on which the modular group $\Gamma(1) = \mathrm{SL}(2, \mathbb{Z})$ acts by fractional-linear transformations, in effect changing the basis elements of the lattice $L \cong \pi_1(X)$, and preserving the Poincaré metric. Thus, in this special case, the deformation space carries an intrinsically richer geometric structure of hyperbolic plane geometry and we sometimes denote it by \mathcal{H}^2 when focussing on this viewpoint. In general, the *mapping class group* $\mathrm{Mod}_{g,n} = \pi_0 \mathrm{Diff}(X)$ of the (n -punctured) surface X acts as the *Teichmüller modular group* by changing the marking on a reference surface. When $n = 0$, the quotient of Teichmüller space \mathcal{T}_g by this action of Mod_g is *Riemann's moduli space* \mathcal{M}_g , which parametrises biholomorphic equivalence classes of (closed, genus g) Riemann surfaces, while if $n = 1$ the resulting quotient is the *modular curve* \mathcal{C}_g , the total space of the fibred *modular family* of genus g surfaces.

Teichmüller also defined a special type of deformation within the spaces \mathcal{T}_g , which is now called a *Teichmüller (geodesic) disc*; we sometimes refer to it as a *T-disc*. They are described in more detail in Section 4: for the purposes of this outline, we mention only that inside any $\mathcal{T}_{g,n}$ they form a natural class of complete complex submanifold of dimension 1, isometric to the Poincaré metric model of the hyperbolic plane \mathcal{H}^2 , and that they are in plentiful supply, through each point and in any given direction, forming an integrated part of the Teichmüller geodesic flow.

It is natural to consider the action of the mapping class group $\Gamma_{g,n}$ on the set of all T-discs in $\mathcal{T}_{g,n}$. Because the modular action is biholomorphic and discontinuous, it preserves the set of T-discs and the stability subgroup preserving a given disc is then

isomorphic to a Fuchsian discrete group of hyperbolic isometries. It is not hard to show (see [13] for instance) that a sufficiently general T-disc will have trivial stabiliser; but how large can the stability group of a T-disc be? One might naturally look first for examples with a cocompact Fuchsian group as stabiliser, since that type of family would be of interest for both the topology of the modular variety and the algebra of the mapping class group, and it would provide a valuable tool for constructing complete families of Riemann surfaces. The search is fruitless, however, since any T-disc contains geodesic rays whose projection is divergent in the moduli space [47], which implies at once that the quotient of a T-disc by its stabiliser cannot be compact. Thus, *there is no totally geodesic, complete, complex suborbifold in any of the higher dimensional moduli spaces $\mathcal{M}_{g,n}$* , and one must use quite different methods to produce complete curves in the modular varieties of closed surfaces \mathcal{M}_g (cf. [38], [19]).

The situation with stabilisers of T-discs becomes more interesting if the compact quotient condition is relaxed. In fact, as we explain in Section 5, in any given \mathcal{T}_g there is a large set of marked surfaces $[S_n]$ which define centre points of T-discs stabilised by (noncompact) *finite volume* Fuchsian groups, so that in this case the quotient in moduli space \mathcal{M}_g is a finite-area immersed Riemann surface, isomorphic to a quotient of the base surface S_n by some subgroup of its automorphism group. They are examples of a more general class of affine algebraic curve in moduli space known as a *Veech curve* to be discussed briefly later; all of the examples we describe are given by algebraic equations defined over some number field, which suggests that the geometric action of the Teichmüller modular group on the unit tangent bundle over the Teichmüller space will form a distinctive part of the relationship conjectured in [22] between the absolute Galois group $\mathcal{G}(\overline{\mathbb{Q}})$ and the Grothendieck–Teichmüller tower, once the meaning of that object has been clarified.

For higher dimensional families of algebraic curves, there are interesting but sporadic results deriving from work of many authors, among them E. Picard (and his students), G. Shimura (and his students), R. Holzapfel, P. Deligne and G. D. Mostow, Paula B. Cohen, J. Wolfart and H. Shiga; often these employ some version of the period mapping for holomorphic 1-forms and its monodromy and have some relation to the present work, which we shall not be able to explore here.

In the next two sections, we shall describe essential background and the examples which motivate our viewpoint. The fourth describes a procedure for constructing one type of Veech–Teichmüller curve in moduli space. In the last chapter, we consider the arithmetic problem of characterising the curves of given genus g that are definable over $\overline{\mathbb{Q}}$. We work within the convenient topological framework of a single space of moduli \mathcal{D}_0 , the space of all cocompact Fuchsian groups, viewed modulo conjugacy within the Lie group $G = \mathrm{PSL}(2, \mathbb{R})$. There is a natural structure of metric space on \mathcal{D}_0 , using the Hausdorff metric on the space of closed subgroups of the Lie group G , and by the uniformisation theorem it contains copies of every modular variety $\mathcal{M}_{g,n}$ with $3g - 3 + n > 0$. A related question asks for a precise description of all $\overline{\mathbb{Q}}$ -curves in a suitable compactification of the space of stable genus g curves $\hat{\mathcal{M}}_g$. In emphasising the totality of hyperbolic discs with crystallographic (i.e. discrete co-finite) Fuchsian

stabilisers, we hope to bring out a general perspective which places dessins within the tautological geometric framework underpinning the patchwork collection of Teichmüller spaces which cover the various modular families and which in the process generate the space \mathcal{D}_0 . At the same time one needs the entire collection of lattices (finite co-volume Fuchsian groups) and Teichmüller space inclusions to reflect the wide range of possible Belyi representations of complex curves defined over $\overline{\mathbb{Q}}$: every meromorphic function which has 3 ramification values determines a covering of the complex projective line with the requisite properties and each *admissible cusp form* of weight 4 produces a Teichmüller disc and, thereby, candidate members of the set of $\overline{\mathbb{Q}}$ -points of the space of moduli. This space may provide an appropriate starting point for the (still incomplete) process of constructing the ultimate arithmetic modular family described in the Grothendieck Esquisse [22].

3 Grothendieck dessins and Thurston's examples

3.1 Dessins

A *dessin d'enfant* or simply *dessin* is defined to be a connected (finite) graph in a compact surface, whose complement is a union of cells and which has a bipartite structure on the vertices. It is convenient to denote the labels for vertices of the graph with \bullet and \star .

To keep the exposition simple, we restrict attention to orientable surfaces and to a special type of dessin, concentrating on the subclass of dessins which arise from pulling back the standard triangulation of $\mathbb{C}P^1$ through a *Belyi function*, a holomorphic branched covering mapping $\beta: X \rightarrow \mathbb{C}P^1$ with three critical values $0 (= \bullet)$, $1 (= \star)$ and $\infty (= \circ)$. This consists of a topological decomposition of the Riemann sphere into two triangles U and L , with interiors the upper and lower halves of the complex plane respectively, with vertices given by the above three symbols and disjoint edges joining them pairwise along the real axis. In addition, we shall often assume some regularity property for the function β , to be spelled out in the next subsection; in an appropriate sense, the restricted class of dessins to be used forms a natural cofinal subclass of the general pattern. For a broader study, the reader may consult a range of recent survey articles ([33], [61]) and the chapter by Herrlich and Schmithüsen in Volume II of this Handbook [29].

3.2 Clean dessins and Galois covers of \mathbb{P}^1

A *Belyi function* $f: X \rightarrow \mathbb{C}P^1$ is called *clean* if every point in the inverse image of \star has ramification of order 2. The corresponding dessin in X , defined as

$$K_f = f^{-1}\{\bullet \rightarrow \star\},$$

is then called *clean* if each vertex mapping to (and labelled) \star has valency 4 in the graph K_f .

It is not difficult to see that every dessin has a standard subdivision for which this property holds: for instance, one can take the triangular subdivision (see [64]) which arises by replacing the function f by $\tilde{f} = 4(f - f^2)$.

The transition from combinatorial surface topology to complex analysis and the modular group is made by the classical representation of the punctured plane $\mathbb{C} - \{0, 1\}$ as the congruence modular surface of level 2, $X(2) = \mathcal{U}/\Gamma(2)$. One finds with the help of the monodromy theorem that the unbranched covering mapping f^* , obtained by puncturing X at all labelled vertices, is determined by a finite index subgroup Γ of $\Gamma(2)$, for which f is just the projection of each Γ -orbit to the corresponding $\Gamma(2)$ -orbit. Consequently there is a uniformised picture of the dessin in the upper half plane which arises as a quotient under the group Γ of the cell subdivision of \mathcal{U} into ideal fundamental triangles for $\Gamma(2)$ – for more details here, consult [33] for instance.

To gain insight into the picture in the hyperbolic plane, it is helpful to begin from the standard triangular fundamental domain for the full modular group, which determines the barycentric subdivision of the $\Gamma(2)$ -invariant triangulation.

In the context of the analysis of general (orientable) dessins in [22], one considers instead permutation representations of the *cartographic group* \mathcal{C}_2^+ , which is isomorphic to the congruence subgroup $\Gamma_0(2)$ containing $\Gamma(2)$ with index 2. This group consists of all projective transformations

$$\gamma: z \mapsto \frac{az + b}{cz + d} \in \Gamma(1) = \text{PSL}(2, \mathbb{Z}) \quad \text{with } c \equiv 0 \pmod{2}.$$

It is not hard to see, using a standard choice of fundamental domains for the groups concerned, that passage from f to \tilde{f} is defined by extending the classifying inclusion $\Gamma < \Gamma(2)$ to $\Gamma < \Gamma_0(2)$, so that there is again a tessellation of \mathcal{U} by hyperbolic triangles which determines a (clean) dessin on the surface X .

A dessin is called *balanced* when both (a) every vertex has the same valency and (b) each cell has the same number of edges. A *Galois dessin* is a graph-and-cell decomposition obtained by lifting the standard division of $\mathbb{C}P^1$ through a Belyi mapping which is also a Galois covering.

It is a simple exercise in finite group actions to see that Galois dessins are balanced, but the converse is not necessarily true. These matters are discussed at greater length in the collection of articles [61].

3.3 Thurston's examples

We use the term *Thurston decomposition* of a compact surface S to refer to a cell decomposition of S determined by two loops which *fill it up*, partitioning the surface into polygonal cells with an even number of edges, which are segments coming from each loop alternately.

Thurston's original 1975 construction [69] was motivated by his work on the classification of surface mappings up to free homotopy, and in particular by the search for a counterexample to an old conjecture of Jakob Nielsen that an algebraically finite mapping class, i.e. one induced by a surface mapping for which all homological eigenvalues are roots of unity, must be periodic. To construct examples of the kind of map which contradicts this, he took a certain product of twists along a pair of loops α, β that *fill up* the surface: this means that the loops, taken to be in general position with no triple intersection points, are such that the complementary regions of S are all simply connected cells with at least four edges. This type of decomposition by two loops determines a dessin on S with the property that all vertices of type \star have valency 4: this is achieved by placing a vertex \circ in each cell and an edge to it from each vertex of the graph. In fact this procedure also yields a dual partition of the surface into quadrangles, each with two vertices labelled \star , which is a subdivision of the cell partition of S dual to the original one; omitting the \bullet vertices gives the dual quadrilateral structure.

The method used by Thurston to produce pseudo-Anosov homeomorphisms from this combinatorial structure is simple but very ingenious: because the surface S has been decomposed into quadrilaterals, it has a branched piecewise Euclidean structure given by the requirement that each cell is to be a unit square. This determines a structure of compact Riemann surface too, by the standard removal of singularities at corners, but the important point is that in the Euclidean structure the Dehn twist mappings along each loop act as *translation by an integer*. Indeed, one can approach the construction of the surface in reverse from a set of square tile pieces by first joining them in line following the path α and then making the necessary identifications of edge segments on the resulting cylinder; in the local coordinate of the resulting geometric structure, a twist along the core of the cylinder is clearly an integral translation. For the β loop, one follows a similar path of identifications and again we obtain as monodromy of the geometric structure around the loop an integer translation in the orthogonal direction. Now the group generated by the two twists is a subgroup, often of finite index, in the modular group $\mathrm{SL}(2, \mathbb{Z})$. Furthermore, it is clear that any word in the two twist mappings that gives a modular group element with two real eigenvalues (a hyperbolic element) will define a pseudo-Anosov mapping class for the surface; with a more careful choice of word, the mapping can be arranged to act trivially on homology (see [69] for more details).

In this construction we see a first glimpse of how dessins and the mapping class group interact, although the complex geometry of moduli space is not yet in evidence.

4 Teichmüller theory

This section presents the background theory of complex deformations of curves, mostly without proofs but with some effort to give at least an intuitive grasp of the results

which are important. To this end, we also include some specific examples of the kind of extremal deformation which is to be highlighted.

4.1 Deformations of complex structure on a surface

A convenient framework for studying analytic moduli of surfaces with all shapes and sizes in an integrated manner is provided by the universal Teichmüller space developed by Ahlfors and Bers; we provide a brief sketch only, sufficient to make it possible to give a unified approach to the kind of deformation which we need in the last section. Detailed accounts may be found in several textbooks [16], [54], [23].

The first fundamental problem encountered in deformation theory is the question of specifying precisely what kind of structure can reasonably be used. We require our object (which is to be deformed) to have no local symmetry, in the sense that there is no object nearby that is isomorphic to it: this is *local rigidity*. In the case of Riemann surfaces, it turns out that one needs, in addition to the categorical object concerned, the additional data of a topological *marking*, which is defined to be a homeomorphism $f: X_0 \rightarrow X$, from a fixed reference surface X_0 , viewed up to free homotopy. This is the appropriate condition to deploy, because *rigidity* of a Riemann surface X marked in this sense follows from the fact that holomorphic self-mappings are represented faithfully in the action induced on the homotopy (or homology) of the surface: a conformal self-mapping (distinct from the identity) cannot be trivial on homotopy. One prefers to employ a homotopy marking rather than a homologically defined one because of the more direct link with covering spaces: also, with hindsight, it turns out that this provides a clearer view of the often intricate interactions of the groups involved.

It is a consequence of the theory of Beltrami equations, to be outlined next, coupled to the uniformisation theorem of Riemann and Koebe, that for every (finite volume) Riemann surface X with universal covering the upper half plane \mathcal{U} , provided with a *marking* $f: X_0 \rightarrow X$, there is a representation of (X, f) as a normalised homeomorphism $h: \mathbb{R} \rightarrow \mathbb{R}$ with $h(0) = 0, h(1) = 1$. Here $\mathbb{R} \cup \{\infty\} \sim S^1$ is the boundary of the upper half plane \mathcal{U} and these maps h , normalised to fix ∞ , are in fact *quasisymmetric* homeomorphisms of the boundary real line, arising as boundary values of quasiconformal self mappings $f: \mathcal{U} \rightarrow \mathcal{U}$ which cover the marking of X . All the essential ingredients of real and complex analysis on the Riemann surface are transferred to analogues defined on the universal cover $p: \mathcal{U} \rightarrow X$ which transform suitably under the discrete (Fuchsian) group of deck transformations of the covering, which is a subgroup of the direct isometry group of the hyperbolic plane. This process delivers within a single universal space all deformations of all Riemann surfaces.

To explain this in more detail, we make the following definitions.

Definition. A *Beltrami coefficient* μ on a hyperbolic Riemann surface X , is a measurable complex function on the universal covering \mathcal{U} transforming under the fundamental (covering) group Γ as a tensor of type $(-1, 1)$, with sup norm less than 1.

Thus μ is an element of $B_1(X)$, the open unit ball in

$$L^\infty_{-1,1}(X) \cong \{\mu \in L^\infty(\mathcal{U}, \mathbb{C}) \mid \mu(\gamma(z))\overline{\gamma'(z)}/\gamma'(z) = \mu(z) \text{ for all } \gamma \in \Gamma\}.$$

The corresponding *Beltrami equation* is a partial differential equation for a locally homeomorphic complex-valued coordinate function $w: X \rightarrow \mathbb{C}$:

$$w_{\bar{z}}(z) = \mu(z)w_z. \tag{1}$$

The Ahlfors–Bers theory (see [3]) provides a unique normalised solution homeomorphism w_μ to this equation in \mathcal{U} , called *quasiconformal* or simply ‘qc’, which extends to a homeomorphism h of the boundary $\partial\mathcal{U}$, compatible with the given Fuchsian group Γ . Here, the process of normalisation amounts to factoring out the natural action on the set of solutions by post-composition with the group $G = \text{PSL}_2(\mathbb{R})$ of all real Möbius transformations, the direct hyperbolic isometries of \mathcal{U} : if w is a solution to the equation, then so is the composition $\gamma \circ w$ with $\gamma \in G$. We shall usually employ the standard normalisation, which requires w to fix 0, 1 and ∞ . Using the above invariance property of the Beltrami form, a computation with partial derivatives shows that each composition of the form

$$\gamma_\mu = w_\mu \circ \gamma \circ w_\mu^{-1}, \quad \text{with } \gamma \in \Gamma,$$

has zero conformal distortion, and it follows that there is a deformed covering group $\Gamma_\mu = w_\mu \Gamma w_\mu^{-1}$ of real Möbius transformations, quasiconformally conjugate to Γ . This determines a geometric isomorphism between the two Fuchsian groups, induced by this qc-homeomorphism between the surfaces, which gives a genuine distortion of the original complex structure if the map is not conformal.

The crucial notion of Teichmüller equivalence for two Beltrami coefficients is introduced next; this reduces the infinite dimensional spaces of Beltrami coefficients to the (finite dimensional) deformation spaces appropriate for these marked structures.

Definition. Two Beltrami coefficients μ and μ' are called *Teichmüller-equivalent* if the corresponding marked groups Γ_μ and $\Gamma_{\mu'}$ are conformally equivalent, that is, conjugate by some real Möbius transformation in G .

This condition, sometimes abbreviated to T-equivalence, implies that the corresponding *normalised homeomorphisms* $h = h_\mu$ of \mathbb{R} , with $h(0) = 0$ and $h(1) = 1$, depend only on the Teichmüller class of μ . The set of all the normalised maps h_μ for arbitrary measurable $\mu \in B_1(L^\infty(\mathcal{U}))$, without any group compatibility condition, is the *universal Teichmüller space*, denoted $\mathcal{T}(1)$. It transpires that there is a purely intrinsic analytic criterion which characterises the homeomorphisms of the boundary circle $\mathbb{R} \cong S^1$ that arise in this way: they are called *quasisymmetric*. See for instance [16]. The detailed analysis of boundary behaviour need not concern us here, but becomes highly important for infinite dimensional Teichmüller spaces, arising from surfaces with more complicated boundary, including curves or infinitely many punctures, where attention switches to asymptotic behaviour. See [17] for details of this very different aspect of the theory.

Definition. The *Teichmüller space* of the Fuchsian group Γ (and by abuse of notation, of the Riemann surface $X = \mathcal{U}/\Gamma$) is the set of T -equivalence classes of Beltrami coefficients for the group Γ .

An inclusion of Fuchsian groups $\Gamma_1 \leq \Gamma_2$ induces a natural relationship between spaces of Beltrami forms $B_1(X_1) \supseteq B_1(X_2)$, for the surfaces $X_1 = \mathcal{U}/\Gamma_1$ and $X_2 = \mathcal{U}/\Gamma_2$, where $B_1(X_1) \cong B_1(\mathcal{U}, \Gamma_1) \supseteq B_1(\mathcal{U}, \Gamma_2) \cong B_1(X_2)$. This projects to an inclusion $\mathcal{T}(X_1) \supseteq \mathcal{T}(X_2)$. Thus, in particular, $\mathcal{T}(1)$ has the significant *universal property* that it contains copies of every Teichmüller space $\mathcal{T}(\Gamma)$. For many purposes, however, this space is too vast and has too great a diversity of subspaces, and one has to focus on the individual spaces to gain real insight into the geometry of the modular families.

4.2 Teichmüller spaces and modular groups

One strongpoint of the Ahlfors–Bers formalism lies in the coordinated description it gives for the action of the *modular groups* of the various Fuchsian groups and surfaces.

Recall first the standard (topological) definition of the *mapping class group* of a surface (or possibly orbifold) $X = \mathcal{U}/\Gamma$ as the group of isotopy classes of homeomorphisms of X to itself preserving any punctures or cone points; equivalently one may use homotopy classes instead, since these notions coincide in two dimensions. For compact surfaces, it is easily shown that each homotopy class of homeomorphism contains quasiconformal representatives and this does extend to orbifolds and surfaces with finitely many punctures. We may apply the method of lifting in the theory of branched coverings to represent each homeomorphism of X as the projection from a Γ -compatible homeomorphism f of \mathcal{U} which induces a *geometric automorphism* α of the covering group Γ by conjugation:

$$\gamma \mapsto f \circ \gamma \circ f^{-1} = \alpha(\gamma).$$

Of course, one must verify that distinct choices in lifting, or a homotopic variation, lead to the replacement of f by $\gamma_0 \circ f$ where $\gamma_0 \in \Gamma$, and that the effect on the automorphism α is to compose it with an inner automorphism of Γ .

If $\text{mod}(\Gamma)$ denotes the group of (geometric) automorphisms of Γ and $QC(X)$ denotes the group of quasiconformal (qc) self-homeomorphisms of the compact (possibly orbifold) Riemann surface X , with $QC_0(X)$ the normal subgroup of null-homotopic ones, then from the above procedure we obtain a homomorphism from $QC(X)$ to $\text{Aut}(\pi_1(X))$, which projects to a homomorphism Φ from the *geometric mapping class group* to the algebraic one: the subgroup Inn of inner automorphisms of $\Gamma = \pi_1(X)$ arise from distinct liftings of a homeomorphism homotopic to the identity and writing $\text{Mod}(X)$ for the quotient $\text{Aut}(\pi_1(X))/\text{Inn}$, we have

$$\Phi: QC(X)/QC_0(X) \cong \pi_0 \text{Diff}^+(X) \longrightarrow \text{Mod}(X) \leq \text{Out}(\pi_1(X)).$$

The presence of either torsion or punctures in the Fuchsian group Γ renders the proofs more difficult but does not change their validity. Furthermore, *the morphism Φ is surjective: that is, we can realise each individual automorphism $\alpha: \Gamma \rightarrow \Gamma$ by a quasiconformal homeomorphism $f = f_\alpha$ of \mathcal{U}/Γ* – this is a classical theorem due to Nielsen and Fenchel. For more details of that result, see for instance [23], Chapters 1 and 9. A more detailed discussion of the algebraic and geometric versions of modular group for orbifold surfaces which includes the general isomorphism result is given in [27].

Employing the geometric definition of mapping class group, we write the fundamental action on Teichmüller space as

$$\rho_\alpha([f]) = [f \circ f_\alpha^{-1}],$$

with $[f]$ denoting Teichmüller-class, which defines a group action on $\mathcal{T}(\Gamma)$ with kernel the subgroup $\text{Inn}(\Gamma)$ of inner automorphisms. In this way, we obtain the important representation of the quotient $\text{Mod}(X) = \text{mod}(\Gamma)/\text{Inn}(\Gamma)$, as the *Teichmüller modular group* $\text{Mod}(\Gamma)$ of (biholomorphic) automorphisms of $\mathcal{T}(\Gamma)$; if Γ represents a compact surface with genus g and n punctures, this group is often denoted $\text{Mod}_{g,n}$ or Mod_g when $n = 0$.

A fundamental paper of Royden [58] analyses the metric structure on $\mathcal{T}(\Gamma)$ which is induced by using the supremum of the distortion function $\mu \in B_1$ to measure the distance between marked compact surfaces, $[X_j, f_j]$ with $f_1 = w_\mu \circ f_0$: write $k = \|\mu\|_\infty = \text{ess. sup}\{|\mu(z)| \mid z \in X_0\}$ and set

$$d([X_0, f_0], [X_1, f_1]) = \frac{1}{2} \ln \left(\frac{1+k}{1-k} \right). \tag{2}$$

This is *Teichmüller’s metric* on $\mathcal{T}(\Gamma)$ – see for instance [16] for a detailed account – and it is not hard to see that it is invariant under the action of the modular group. Indeed, one needs only to redraft the definition of the metric in terms of the qc mappings which interconnect the two marked surfaces:

$$d([X_0, f_0], [X_1, f_1]) = \frac{1}{2} \ln K((f_1 \circ f_0^{-1})). \tag{3}$$

Putting in the action of $\alpha \in \text{Mod}(X)$ leaves this expression unchanged.

Royden proved the following important result.

Theorem 1. *The Teichmüller modular group Mod_g is the full group of Teichmüller-metric isometries of \mathcal{T}_g .*

The result was extended to all finite dimensional Teichmüller spaces by Earle and Kra: the notation g, n refers to the topological surface data of genus g with $n \geq 0$ marked special points and is explained later on.

$$\text{Isom}(\mathcal{T}_{g,n}) \cong \mathcal{M}_{g,n}. \tag{4}$$

This group action is of pivotal importance in the Teichmüller theory, because the action is properly discontinuous on $\mathcal{T}(\Gamma) = \mathcal{T}_{g,n}$ with discrete orbits (see for instance

[16], [54]), producing as quotient the desired *moduli space* $\mathcal{M}_{g,n}$ of conformal classes of marked genus g surfaces.

It can be proved (with some work) that the natural complex structures on the Banach spaces of Beltrami forms project to make the Teichmüller spaces into complex manifolds; one shows that the tangent space at a marked point $[X_0, f] \in \mathcal{T}(X)$ is dual to the complex vector space of holomorphic quadratic differential forms on X_0 in the case of a closed surface, with extra vanishing conditions at any cone points or punctures. From the Riemann–Roch theorem, it is known that this space has finite dimension $3g - 3 + n$, which tallies with the dimensions of the spaces $\mathcal{T}(\Gamma)$. One then shows that this almost complex tangential structure is integrable. Furthermore, there is a Hermitian metric on the tangent bundle which derives from the Petersson inner product on the quadratic differential forms and it was shown by Weil, and independently by Ahlfors, that this metric is Kähler. In fact, Wolpert was able to prove that every moduli space has a projective embedding, using Kodaira’s embedding theorem [76].

Note. A full account of the complex analytic theory of these spaces is best given by considering the Bers embedding, which requires more time and space than we can afford: the reader should consult the standard texts cited earlier for this important work.

A compact genus g Riemann surface with a non-trivial group G of biholomorphic automorphisms will give rise to a marked point in $\mathcal{T}(X) = \mathcal{T}_g$ which has G as the subgroup of $\text{Mod}(X)$ stabilising the marked point; similarly, a finite group of homeomorphisms of a compact surface always produces a marked Riemannian metric on the surface, hence a Teichmüller point with this as its automorphism group. The famous *Realisation Problem of Hurwitz–Nielsen* asked whether, given a finite subgroup H of $\text{Mod}(X)$, there is a choice of marked point $[\mu] \in \mathcal{T}(X)$ for which the stability group is H ; this may be rephrased as follows.

Is there a finite group of homeomorphisms H_1 which realises a given finite group of mapping classes H , in the sense that $H_1 \cong H$ with $H_1 \leq QC(X)$ and $\Phi(H_1) = H$?

This problem was studied, and partially solved, by Nielsen and Fenchel and various special cases were proved by later authors over more than 30 years until it was finally resolved in general by S. P. Kerckhoff [35] using Thurston’s powerful theory of earthquakes. Different proofs have been given by Wolpert (using convexity properties of the hyperbolic length functions on Teichmüller space) and more recently in an interesting paper by D. Gabai [15], which uses methods reminiscent of Nielsen’s and also explains how to relate it to fundamental problems in the geometry and topology of 3-manifolds.

Notice, however that there is no natural overall choice of a realisation, in the following sense.

Theorem 2. *No homomorphism exists from $\text{mod}(\Gamma)$ into the group $QC(\mathcal{U})$ which would split the epimorphism Φ .*

For inverse homomorphisms into the subset of C^∞ diffeomorphisms of X , the impossibility of such a splitting was proved by S. Morita [53], using his theory of characteristic classes for surface bundles and moduli spaces. This restriction has been removed recently by work of D. B. A. Epstein and V. Markovic.

4.3 Families of Riemann surfaces

At this point, it is appropriate to indicate how to construct the corresponding *universal holomorphic family of Riemann surfaces* modelled topologically on $X = \mathcal{U}/\Gamma$, with the space $\mathcal{T}(\Gamma)$ as base. This is a tautological *holomorphic fibre space* $V(\Gamma)$ over the Teichmüller space, obtained as follows.

- (1) Place over each point $[\mu]$ of $\mathcal{T}(\Gamma)$ the corresponding h_μ -deformed hyperbolic disc \mathcal{U}^μ , on which the deformed group Γ_μ operates.
- (2) Take the quotient of this space $F(\Gamma) \cong \mathcal{T}(\Gamma) \times \mathcal{U}$ by the discrete action of Γ on the fibres as Γ_μ on \mathcal{U}^μ .

Note that although we use the notation h_μ , both this normalised map and the deformed disc depend only on the T-class $[\mu]$. The resulting space $V(\Gamma)$ is a complex analytic manifold which is locally a topological product of the fibre orbifold and an open disc in the base $\mathcal{T}(\Gamma)$. However the product structure is holomorphically nontrivial because of the nature of these families of quasiconformal mappings $h_{\mu\nu}$ inter-relating the nearby fibre surfaces. This type of deformation is trivial only when there is a conformal map of \mathcal{U} inducing it.

The same fibre space structure is also realisable real-analytically by a standard bundle-theoretic procedure known as homogeneous reduction, from the *representation space* $R_0(\Gamma, G)$ of all discrete injective homomorphisms ρ of the Fuchsian group Γ into the direct-isometry group $G \cong \mathrm{PSL}_2(\mathbb{R})$ of the hyperbolic plane \mathcal{U} . We assume for simplicity that the quotient surface $X = \mathcal{U}/\Gamma$ is compact or possibly finite area, with finitely many cone points or punctures; the *topological type* (g, n) records the genus g and the total of $n \geq 0$ special points. Since the spaces $\mathcal{T}(\Gamma)$ depend as complex manifolds only on the topological data and not on the cone/cusp nature of the marked points, we usually refer to the moduli spaces as $\mathcal{T}_{g,n}$ and $\mathcal{M}_{g,n}$; modular groups are a little less flexible, and $\mathrm{Mod}_{g,n}$ will refer to the case of n punctures.

The target Lie group G operates on the space R_0 by conjugation: each element $\rho(\gamma)$ in the image of the given representation is sent by $t \in G$ to the conjugate $t\rho(\gamma)t^{-1}$ by the fixed element t in G . In this way, each representation $\rho \in R_0(\Gamma, G)$ determines a conjugacy class of G -similar ones.

Now this action on R_0 is *proper*, and each G -orbit, a collection of marked groups conformally equivalent to the group $\rho(\Gamma)$, may be viewed as the fibre of a map to the quotient space of conjugacy classes, which is none other than our Teichmüller space $\mathcal{T}(\Gamma)$ of normalised marked Fuchsian groups – a comparison of these two perspectives may be found in papers of Kerckhoff [34], [35]; see also recent work of many authors

(Fock, Goncharov, Kashaev, Penner and others) on a quantisation of the theory of moduli which exploits these two approaches to geometric structure both on surfaces and on the Teichmüller spaces.

To convert the fibration $R_0(\Gamma, G) \rightarrow \mathcal{T}(\Gamma)$ into a family of Riemann surfaces, we note that the group $G = \mathrm{PSL}_2(\mathbb{R})$ may be viewed as the unit tangent bundle of the hyperbolic plane, a principal PSO_2 -bundle; this may be converted into a holomorphic family of hyperbolic planes by taking the quotient modulo the rotation subgroup $\mathrm{PSO}(2)$ of G . This space is denoted $F(\Gamma)$. Finally one obtains a fibration with Riemann surfaces as fibres by taking the quotient by the action of Γ via the representations $\rho \in R_0(\Gamma)$ on each fibre disc, as a subgroup of G : to see this in more detail, the reader may consult [45] or [54].

There is another description of the holomorphic fibre space $F(\Gamma)$ as the space of all pairs

$$(t, z) \quad \text{with } t = [w_\mu] \in \mathcal{T}(\Gamma) \text{ and } z \in w_\mu(\mathcal{U});$$

for details of this aspect of the theory, which involves the analytic properties of the Schwarzian derivative and the Bers embedding of each Teichmüller space in the Banach space of holomorphic functions on a half plane (with the hyperbolic L^∞ norm), the reader may consult [7] or [16].

Following on from the fibre space property, the action of the Teichmüller-modular group on the base space – as the group $\mathrm{Mod}(\Gamma)$ (of mapping classes for the surface or orbifold $X = \mathcal{U}/\Gamma$) on $\mathcal{T}(\Gamma)$ – lifts to an action on $F(\Gamma)$ which is effective (except when $g \leq 2$), and the quotient space is a family of hyperbolic discs \mathcal{U}_μ over $\mathcal{M}_{g,n}$. The discrete group Γ acts on this family via quasiconformal deformations Γ_μ on \mathcal{U}_μ as fibre-preserving maps to produce the *modular family* $\mathcal{C}_{g,n}$ of Riemann surfaces of (generic) genus g with $n \geq 0$ special points. The quotient fibration

$$\pi : \mathcal{C}_{g,n} \mapsto \mathcal{M}_{g,n}$$

has *orbifold singularities* which stem from symmetry properties of certain (non-generic) Riemann surfaces. These arise in the following way.

Proposition 1. *At a point x of $\mathcal{M}_{g,n}$ which represents a Riemann surface S_x with direct symmetry group H , the fibre $\pi^{-1}(x) \subset \mathcal{C}_{g,n}$ is the quotient surface S_x/H .*

This fact is easily seen from the covering space properties of the hyperbolic uniformisation: Riemann surface symmetries lift to hyperbolic disc isometries, and these contribute to the quotient action producing the fibre surface in the construction using $R_0(\Gamma, G)$ outlined above; see [45] for more details.

Note. It is an important ingredient of the Fuchsian group approach to moduli that the marking procedure for the group representing a surface with distinguished points provides a canonical way to incorporate the *forgetful map* which forgets one chosen distinguished point (or a designated set of points).

To explain this process, we recall the detailed geometric information present in the datum of a marked Fuchsian group Γ with quotient surface $X = \mathcal{U}/\Gamma$ of finite type: there is a finite presentation for Γ involving the standard commutator relator arising from the canonical surface symbol and further algebraic relations, one for each conjugacy class of torsion or boundary parabolic element. We have by implication an ordered finite set of special points on X , with interior cone points of angle $\frac{2\pi}{n}$ for order n torsion generators and a puncture for each parabolic generator, manifested as a *cusp* of Γ , a pointlike boundary component of X which may be filled in by a standard procedure to yield a hyperbolic Riemann surface X^* with one less boundary point. This process makes it possible to formulate precisely (and prove) the following *Bers fiber-space theorem*, which identifies the fiber space $F(\Gamma)$ as a Teichmüller space in its own right:

Theorem 3. *Let $X = \mathcal{U}/\Gamma$ be a compact Riemann surface uniformised by a (finite co-area) Fuchsian group Γ . Then*

$$F(\Gamma) \cong \mathcal{T}(\Gamma'),$$

where Γ' is a Fuchsian group uniformising the punctured surface $X' = X - \{x_0\}$, where $X = \mathcal{U}/\Gamma$.

It leads also to an identification between Teichmüller modular groups. Under the same hypotheses, we have

Theorem 4. *The modular group $\text{mod}(\Gamma)$ is a finite index subgroup of $\text{Mod}(\Gamma')$.*

In particular, $\text{mod}(\Gamma) \cong \text{Mod}(\Gamma')$ when Γ is a closed surface group, i.e. torsion free and co-compact.

The latter result is the counterpart of the final terms in the topologist's long exact sequence for the homotopy groups of a fibration for the holomorphic fiber spaces $F_{g,n} \rightarrow \mathcal{M}_{g,n}$. In general, these two modular groups are only commensurable, that is, they possess isomorphic 'pure' subgroups which correspond to mapping classes preserving all of the marked points.

4.4 Hyperbolic surfaces inside moduli space

One of our aims is to exhibit and study finite area Riemann surfaces which are totally geodesic suborbifolds of moduli space. We saw earlier the examples discovered by Thurston; a generalisation of this construction was obtained later by Veech and others. In order to describe some of these results, we need a preliminary account of complex Teichmüller geometry.

Definition. A *Teichmüller geodesic disc* is a set of marked complex structures on a Riemann surface X_0 arising from the following construction: fix a quadratic differen-

tial form ϕ on X_0 . This determines a complex 1-parameter family of deformations of X_0 with several alternative descriptions:

(i) A characteristic type of parametrisation of X_0 by *complex affine charts* is determined by ϕ : locally away from the zeros of ϕ , write $\phi = dw^2$ to get local parameters w up to transition functions of the form $w \mapsto \pm w + c$. Equivalently, set

$$w = \int_{z_0}^z \sqrt{\phi(t)}.$$

Following Veech [70], one may call such an atlas an \mathcal{F} -structure on X_0 .

Now for each ε with $|\varepsilon| < 1$, define a new \mathcal{F} -structure on the underlying topological surface S_0 by rotating each chart through $\arg \varepsilon$ and expanding the real foliation of $\mathbb{R}^2 = \mathbb{C}$, while contracting the imaginary (vertical) one, by the mapping

$$z = x + iy \mapsto w = K_\varepsilon^{\frac{1}{2}}x + iK_\varepsilon^{-\frac{1}{2}}y \quad \text{where } K_\varepsilon = \frac{1 + |\varepsilon|}{1 - |\varepsilon|}.$$

If $\arg \varepsilon = 0$, so that $0 < \varepsilon < 1$, the family of \mathcal{F} -structures so defined is called the *Teichmüller ray* at X_0 in the direction ϕ .

(ii) Write $v_\varepsilon(z) = \varepsilon \bar{\phi}(z)/|\phi(z)|$ and solve the Beltrami equation (1)

$$\frac{\partial w}{\partial \bar{z}} = v_\varepsilon \frac{\partial w}{\partial z}$$

for this family of complex dilatations; conjugating the marking of X_0 by the family ω_ε of solution homeomorphisms (suitably normalised to remove the G ambiguity) yields a one-dimensional holomorphic family of Riemann surfaces over the unit disc \mathbb{D} .

It is an exercise in the basic workings of quasiconformal mappings to see that these two descriptions determine the same complex 1-parameter deformation of the central surface X_0 .

There are at least two further valuable ways to view this deformation. One stems from the affine Euclidean geometry of the system of local charts determined by a quadratic differential; this is the point of view employed by Veech and also by Kerckhoff, Masur and Smillie [36]; it is described very neatly in the survey article [13] and we describe one of Veech's examples in these terms later in this section. The second, perhaps more topological, view is older and was developed originally by Jenkins and Strebel, and later by Gardiner (for details, see for instance [32], [67], [16]); it makes use of the concept of *critical trajectories* for a quadratic differential. These are the ϕ -horizontal curves on X_0 through the zeros of ϕ in the initial system \mathcal{F}_0 of Euclidean charts: the set of all (horizontal) critical trajectories determines a decomposition of the surface into a union of annuli and spiral regions, each foliated by simple horizontal trajectories. One may then pose an *extremal stretching* problem on the Riemann surface S_0 , determined by the topological structure of this singular foliation arising from the quadratic differential ϕ : a *weight* is assigned to each component of the noncritical set and one seeks a metric on the surface which minimises the distortion overall in a sense determined by the set of weights.

This last point of view is valuable because it links the classical complex analytic methods with the ideas of Thurston on deformations of hyperbolic structures, which bring into play the geometry of foliations of a surface with permitted ‘finitely pronged’ singularities to set up the space of measured foliations which determine a natural real spherical boundary of Teichmüller space. This *Thurston boundary* does of course include points representing the endpoints of the T-rays emanating from the central point $[X_0]$ as described above, but differs in an essential way from the *Teichmüller boundary*, also a sphere, which is obtained by considering the set of all the boundary points of T-rays from a fixed central base point, for reasons arising from the subtle dependence of the latter sphere on the choice of base point.

The interested reader should consult Kerckhoff’s paper [34] for details of this fundamental distinction: he establishes the crucial fact that whereas the Teichmüller modular group action extends naturally to the topologically-based Thurston boundary with highly significant consequences (see [69]), it cannot extend continuously to the more rigidly constructed Teichmüller boundary.

4.5 Example: a T-disc parametrising tori

The motivating example of this basic construction comes from the deformations of a torus: recall that marked tori X_τ correspond bijectively to $\tau \in \mathcal{U}$. We need to relate this to the T-disc construction, in which we employed the conformally equivalent unit disc as parameter space. To define the deformation homeomorphisms from a reference point, say $i \in \mathcal{U}$, to $\tau \in \mathcal{U}$, we use the real affine mapping $\tilde{f}_\tau: x + yi \mapsto x + y\tau$

$$\begin{array}{ccc} \tilde{f}_\tau: & \mathbb{C} & \longrightarrow & \mathbb{C} \\ & \downarrow & & \downarrow \\ \text{which induces } f_\tau: & \mathbb{C}/\Lambda_i & \longrightarrow & \mathbb{C}/\Lambda_\tau \end{array}$$

and this homeomorphism is *extremal* within its homotopy class in Teichmüller’s sense, that is, it has the least overall distortion measured by making the supremum of the local stretching function on X as small as possible. To obtain the local distortion, one calculates

$$\frac{\bar{\partial} f_\tau}{\partial f_\tau} = \frac{(1 + i\tau) d\bar{z}}{(1 - i\tau) dz}, \quad \text{where } \partial = \frac{\partial}{\partial z}, \quad \bar{\partial} = \frac{\partial}{\partial \bar{z}}$$

and z is the local parameter ($z = x + iy$) induced on the complex torus by the quadratic differential form dz^2 on \mathbb{C} . This can be rewritten as $\varepsilon\phi/|\phi|$ where $\varepsilon = \frac{1+i\tau}{1-i\tau} \in \mathbb{D}$, the unit disc, and ϕ is the projected holomorphic form dz^2 on \mathbb{C}/Λ_i .

In the present case, the Teichmüller distance between i and τ as defined earlier turns out to be (see for instance [2])

$$d(i, \tau) = \frac{1}{2} \log \left(\frac{1 + |\varepsilon|}{1 - |\varepsilon|} \right).$$

This is of course Poincaré's metric on \mathcal{U} , giving it the structure of real hyperbolic plane.

4.6 Deformation T-discs embed

In general, let X be a Riemann surface, and let $\phi \in \Omega^2(X)$, $\phi \neq 0$. The Teichmüller deformation specified by ϕ defines a holomorphic mapping e_ϕ of \mathcal{U} (or \mathbb{D}) into $\mathcal{T}(X)$ via the following map into $B_1(X) \subseteq L_{-1,1}^\infty(X)$

$$e_\phi(\varepsilon) = \varepsilon \frac{\overline{\phi}}{|\phi|} \quad \text{for } |\varepsilon| < 1.$$

Proposition 2. *The mapping e_ϕ is a proper holomorphic injection of the disc \mathbb{D} into $\mathcal{T}(X)$.*

Proof. The map e_ϕ is proper because as $|\varepsilon| \rightarrow 1$, the norm $\|e_\phi(\varepsilon)\| \rightarrow 1$, which implies that the distortion is becoming arbitrarily large and so $e_\phi(\varepsilon)$ tends to the boundary $\partial\mathcal{T}(X)$. The map is injective, either by Teichmüller's uniqueness theorem [6] or by the uniqueness of solutions to Beltrami's equation, and it is holomorphic because the complex structure on $\mathcal{T}(X)$ is inherited from that of the Banach space of distortion measures $B_1(X)$. \square

The global metric between points in Teichmüller space, as defined earlier, is in fact realised by the Poincaré metric for points in a T-disc. This is part of the original approach introduced by Teichmüller who proved that any two points are joined by a Teichmüller ray, and used the formula 2 to define it. As a consequence, every embedded Teichmüller disc is totally geodesic and this leads to another result of Royden (op. cit.) which identifies Teichmüller's metric with the canonical *Kobayashi metric* on this complex manifold. In the present context, we can re-state this classic result, generally known as 'Teichmüller's Theorem', as follows.

Theorem 5. *For any two distinct marked Riemann surfaces in \mathcal{T}_g , there is an embedded T-disc which contains them, unique up to an isometry of the disc, and the distance between them is given by Poincaré's hyperbolic metric.*

For a proof of this theorem, which inaugurated the global theory of moduli, see [1], [6].

4.7 Veech's examples

A recipe using Euclidean geometry for explicit construction of a geometric structure on a surface using holomorphic differential forms is given next; this is part of a general procedure for endowing a surface with a covering by Euclidean polygonal charts,

which Veech has called \mathcal{F} -structures. They are also called *translation surfaces* by some authors, since the surfaces concerned are determined by identifications of edges of polygons using Euclidean translations.

The examples we describe involve surfaces which may be represented explicitly as ramified coverings over $\mathbb{C}P^1$: they are taken from a large class discovered by Veech ([70], [72]).

Let X_n , with $n \geq 5$ and odd, be the compact hyperelliptic surface of genus $g = (n - 1)/2$ with affine equation $y^2 = 1 - x^n$ and consider the holomorphic 1-form $\omega = dx/y$. It has a zero of order $2g - 2$ at the single point where $x = \infty$. The quadratic form $q = \omega^2$ defines a Teichmüller disc in \mathcal{T}_g , which turns out to have particular significance.

Veech considered these structures in connection with the dynamical study of billiard trajectories in a Euclidean isosceles triangle T with angles $(\frac{\pi}{n}, \frac{\pi}{n}, \frac{(n-2)\pi}{n})$. The relationship with X_n is based on the following construction, involving an associated triangle T_ζ with vertices at 0, 1 and $\zeta = e^{2\pi i/n}$.

Let $P = \bigcup_{\ell=0}^{n-1} \zeta^\ell T_\zeta$; multiplication by ζ determines a cyclic Euclidean rotation symmetry of P such that the quotient mapping is an n -fold covering of the plane near 0. Let Q be the equivalent polygon formed with $\tilde{T}_\zeta = -T_\zeta$, and then identify pointwise in $P \cup Q$ the pairs of outside edges in corresponding triangles $\zeta^\ell \tilde{T}_\zeta, \zeta^\ell T_\zeta$ using a translation (the corresponding edges are parallel). The result is a closed surface with a local complex analytic structure except at the projection of the corners, where the union of all the corner angles adds up to more than 2π . After extending the complex structure by removing these isolated singularities at the corners – using Riemann’s classical approach to resolving isolated singularities – this process determines a closed Riemann surface X_ζ (with nontrivial conformal symmetry group) having a local structure over \mathbb{P}^1 given by the following composite map $p_2(z) := f \circ p_1$:

$$\begin{array}{ccc}
 X_\zeta & & \\
 \downarrow p_1 & \searrow p_2(z) := f \circ p_1 & \\
 \mathbb{C}P^1 = X_\zeta / \langle z \mapsto \pm z \rangle & \xrightarrow{f} & X_\zeta / \langle z \mapsto \pm \zeta z \rangle \quad \text{in } P \cup Q.
 \end{array}$$

The arrow labelled f is the standard mapping $z \mapsto z^n$, after moving the origin of z -coordinates to the centre of the polygon P . It is not difficult to infer from this geometric data that X_ζ is the compact Riemann surface X_n defined above: for instance, observe that there is an order n symmetry of X induced by lifting $z \mapsto \zeta z$ which fixes three points, the two centre points and the single orbit of corner points. Also it follows that the 1-form dz in $P \cup Q$ lifts to ω on X_ζ . Veech proves the following result, a striking link between the hyperbolic geometry of the disc and the complex geometry of Teichmüller space and the Teichmüller modular group.

Theorem 6. *The stabiliser in the modular group Mod_g of the Teichmüller disc determined by the differential $q = \omega^2$ is a Fuchsian triangle group*

$$H_n = \left\langle \begin{pmatrix} 1 & 2 \cot \frac{\pi}{n} \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} \cos \frac{2\pi}{n} & -\sin \frac{2\pi}{n} \\ \sin \frac{2\pi}{n} & \cos \frac{2\pi}{n} \end{pmatrix} \right\rangle = \langle \sigma, \beta \rangle$$

which is isomorphic to $\langle x_1^n = x_2^2 = 1 \rangle$ via $x_1^{g+1} x_2 = \sigma$, $x_1 = \beta$.

4.8 Hecke triangle groups and polygonal surface tilings

In this section, we explore some of the hyperbolic geometry underlying Theorem 6. The triangle group H_n obtained in the theorem is known as a *Hecke group* because of Hecke's classic work [28] on this kind of Fuchsian group in connection with Dirichlet series satisfying a functional equation.

A fundamental region for the action of H_n on the hyperbolic plane \mathcal{U} may be obtained by a standard geometric construction: take a quadrilateral F with a reflection symmetry which fixes the diagonal joining one ideal vertex at ∞ and a second vertex having interior angle $2\pi/n$. It is simplest to view F as the union of two abutting hyperbolic triangles T, \bar{T} , each with angles $\pi/n, 0$ and $\pi/2$. Place the vertices of T with one of angle π/n at i , the zero angle at ∞ and the third at the foot $\pm\xi$ of the hyperbolic perpendicular from i to the line $\text{Re } z = \cot(\pi/n)$ and let the triangle \bar{T} be the reflection of T in the imaginary axis. Just as the action of an order n rotation divides a Euclidean regular n -gon into n congruent isosceles triangles or, equivalently, into n quadrangles with angles $\{2\pi/n, \pi/2, (n-2)\pi/n, \pi/2\}$, there is a hyperbolic regular ideal n -gon P' with all vertices at infinity, which is the result of applying to F the order n hyperbolic rotation x_1 fixing i .

Consider now the union of P' and the (hyperbolic) translate $\sigma(P')$, which is a congruent polygon sharing a vertical geodesic side. It is not hard to show that the two operations x_1 and σ generate a discrete group of isometries of \mathcal{H} which turns out to be G_n : the union of all translates of the quadrilateral F tessellate \mathcal{H} and these two elements are side-pairing elements of F which fit into the framework of the classical theorem of Poincaré on fundamental polygons for discrete hyperbolic isometry groups; for more details of this result, see [43], [45], or [52].

We can apply Proposition 2 and the homogeneity of the hyperbolic isometry group G acting on \mathcal{U} to arrange the present pattern neatly into the (local and global) complex geometric structure of the Teichmüller space $\mathcal{T}(X_n)$, by placing the reference point i at the marked surface X_n and, furthermore, fixing the T-ray from i in the direction of q to be the upper part of the imaginary axis.

It is important to maintain contact here with the symmetric properties of the Euclidean tessellation picture imposed on X_n by the form q . Note that the element x_2 of order 2 (which may be chosen to fix the $\pi/2$ -corner point of T) corresponds to interchanging the Euclidean polygons P and Q via the halfturn $z \mapsto -z$, whereas it is represented on the quadrangle domain F by two conjugate involutions, each fixing

one of the $\pi/2$ -corners $\pm\xi$. Also, the isometry x_1 of order n fixing i corresponds to the lifting of $z \mapsto \zeta z$ on each polygon. It is not hard to show using the geometry of the hyperbolic plane that within the tessellation of \mathcal{U} by ideal quadrangles $\gamma \cdot F$, $\gamma \in H_n$, we recover the original decomposition of X_n into a pair of (hyperbolic) regular ideal polygons P' , $Q' = x_2(P')$ with n vertices.

To produce an algebraic description of the hyperbolic surface X_n , it is convenient to consider the quotient space X of \mathcal{U} by the commutator subgroup $C_n = [H_n, H_n]$. To see that this is indeed the same Riemann surface, one needs to note that the commutator quotient group $\mathbb{Z}_2 \oplus \mathbb{Z}_n$ acts as a group of biholomorphic automorphisms of X , with precisely the same symmetry pattern (fixed points etc) as that of $X_n = \{P \cup Q\} / \sim$ taken modulo the symmetries of the dihedral n -gon. Of course, the ideal (zero-angle) vertices, which determine either one or two points at ∞ on the quotient surface depending on the parity of n , must be added to create a compact surface. By adjoining cusps to X , we obtain a compact surface. A fundamental domain for C_n may be formed by first taking the union of n abutting (horizontal) translations of the domain F for G_n to produce a lift of the punctured polygon $P - 0$ and then adjoining a translate of this by the involution x_2 . The result is now recognisable as X_n , at least when n is odd: observe that it has genus $g = (n - 1)/2$ (via the Riemann–Hurwitz formula), an order n symmetry with three fixed points (at the C_n orbits $[i]$, $[0]$ and ∞) and an involution symmetry induced by conjugation with x_2 on \mathcal{U}/C_n , which must be hyperelliptic because it fixes the $n + 1 = 2g + 2$ points on the quotient surface arising from the n right-angle corner points of P' and the (single) orbit of cusp vertices.

In the piecewise-Euclidean model X_ζ of the surface, the critical graph of the holomorphic 1-form q consists of a union of vertical line segments through the vertices of the polygons P and Q , all of which are identified to a single point of X_ζ ; the complement is a set of $k = (n - 1)/2$ disjoint cylinders. When viewed in the hyperbolic model as the surface X_n , the cylinders lift to strips which comprise the connected components of the complement of the critical graph \mathcal{K} . In this way, it follows that a set of disjoint geodesic curves tending to the single cusp at ∞ may be taken as representative liftings of the critical trajectories of the holomorphic form φ of weight 2 on \mathcal{U} whose square defines $q = \omega^2$ on X_n . It also follows that the k -th power of the parabolic element σ of H_n determines a covering map which preserves each cylinder, and so defines a product of Dehn twists about the core curves of the cylinders.

These facts can also be verified directly by matrix calculations within the \mathcal{F} -structure, which are similar to the reasoning given earlier in Thurston’s original examples; see for instance [13] for a careful analysis along these lines.

Because this T-disc $\mathbb{D} = \mathbb{D}_q$ has a lattice group stabiliser in the modular group Mod_g , any sequence of points in \mathbb{D} is modular-equivalent (by elements stabilising the T-disc) to either a convergent sequence or a sequence tending to the cusp of H_n and so divergent in \mathcal{M}_g : it follows that the image in the quotient moduli space is a closed (but non-compact) sub-orbifold of the complex V-manifold \mathcal{M}_g . This proves (in outline) the following theorem, also due essentially to Veech [70].

Theorem 7. *The image in \mathcal{M}_g of the Teichmüller disc through \bar{X}_ζ corresponding to the form q is a complex analytic subspace isomorphic to the affine complex line, $\mathcal{U}/H_n = X_\zeta/\text{Aut}(X_\zeta)$.*

This result motivates the following definition.

Definition. A *Teichmüller–Veech disc* is a Teichmüller disc in a Teichmüller space \mathcal{T}_g for which the stabiliser in the Teichmüller modular group is a lattice subgroup of the hyperbolic isometry group G .

The image in \mathcal{M}_g is then a finite area hyperbolic surface, called a *Veech curve* or sometimes a Teichmüller curve.

Notes. 1. Concealed here within the orbifold structure of the moduli space is a paradoxical appearance by the *hyperelliptic involution* J of the central surface \bar{X}_ζ : as a modular automorphism of Teichmüller space, J fixes the point i of \mathcal{U} corresponding to \bar{X}_ζ and sends the 1-form φ to its negative. However this implies that it must fix *every point* of the Teichmüller–Veech disc we have constructed. Thus the stabiliser of the point i in the hyperbolic plane (representing the T-disc) \mathcal{U} is the cyclic group $\langle x_1 \rangle$ and not the full automorphism group of X_ζ , the surface which underlies this point of \mathcal{T}_g . In fact the element x_2 is a representation of the hyperelliptic involution of X_ζ (on the universal covering space \mathcal{U}) and lies in the kernel of the holonomy homomorphism $f \mapsto Df$ from the group of affine selfmaps of the \mathcal{F} -structure to $\text{GL}(2, \mathbb{R})$. Therefore x_2 is killed by the restriction homomorphism from the modular group stabiliser of the T-disc \mathcal{U} to the isometry group of the T-disc. On the other hand, the hyperelliptic involution j is induced by this order 2 generating element x_2 , when viewed as acting on the universal covering of the central fibre of the T-disc, but this is illusory. The point of our T-disc actually fixed by x_2 represents a different Riemann surface from X_ζ , having an additional involution automorphism; Veech [70] calls this involution a hidden symmetry.

2. The case of regular n -gons with n even follows a related but slightly different pattern which is detailed in [71]. In fact, using the discussion of Veech's examples given by [13] it follows that for every $n \geq 5$, the hyperelliptic curve X_n , of genus $g = [n - 1/2]$, enjoy the property that there are two independent Teichmüller discs centred on the corresponding point in \mathcal{M}_g , determined by the 1-forms $\omega_1 = dx/y$ and $\omega_2 = xdx/y$, the second corresponding to a construction of the same surface as a translation surface quotient from a single regular polygon with $2n$ sides; in the latter case the T-disc has as stabiliser a $\{n, \infty, \infty\}$ triangle group.

4.9 Fermat curves represented by T-discs

The *Fermat curves* Y_n are determined by an affine equation

$$x^n + y^n = 1$$

and have genus $g = (n - 1)(n - 2)/2$, and so when $n \geq 4$ they have hyperbolic uniformisation and fit into a similar picture with respect to the same Hecke triangle groups H_n as above. Note that the corresponding compact Riemann surface is obtained by adding just one or two points at infinity depending on parity.

The tessellation of Y_n by regular n -gons may be seen in several ways, for instance by using a similar description of a tailoring pattern for Y_n using $2n$ copies of the polygon P , as in [71]. A second method, less reliant on diagrams and explicit pastings, comes from consideration of the collection of finite index subgroups of a given Hecke group H_n and the corresponding quotient surfaces covering the affine plane \mathcal{U}/H_n [25]. One sees by either method that the automorphic form of weight 4 for $C_n = [H_n, H_n]$ on \mathcal{U} which determines the quadratic differential form $q = \omega^2$ on X_n also induces the form $q_1 = dx^2/y^{2n-2}$ on the covering surface Y_n ; this may be viewed as arising by pullback on forms via the holomorphic covering. In the case of the Fermat curve, however, the involution x_2 in H_n does not preserve the polygonal tessellation of the surface, because the hyperelliptic involution does not lift to Y_n . The (lattice group) stabiliser in the modular group of the surface \bar{Y}_n for this Teichmüller disc is in fact precisely the index-2 subgroup H'_n that omits this element. This subgroup has fundamental set a union of two right triangles interchanged by x_2 , and hence is easily seen to be a triangle group of type (n, n, ∞) .

Theorem 8. *The image in \mathcal{M}_g of the Teichmüller disc through the complex Fermat curve \bar{Y}_n corresponding to q_1 is a complex curve isomorphic to the affine line, \mathcal{U}/H'_n .*

See [72] for detailed proofs of this and other precise stability results.

4.10 Regular tessellations and smooth coverings

In this section, we show how to produce many other examples of hyperbolic discs in \mathcal{T}_g with large stabiliser, by an immediate inference from the argument used for the Fermat surfaces.

Let \tilde{X} be any smooth genus h covering surface of the (compactified) commutator surface X_n for the Hecke group H_n with $n \geq 3$. Then application of the same reasoning as above with the pullback of the form ω shows that there is a corresponding T-disc in the Teichmüller space \mathcal{T}_h centred at $[\tilde{X}]$ with stabiliser of finite index in H_n . In some sense, it is the same data that is involved here: when lifted to a system of covering surfaces, the T-disc given by ω lives in some inverse limit of Teichmüller modular objects which identifies commensurable structures occurring among all finite coverings of X_n . Dennis Sullivan has called this type of limit object a *solenoidal surface* or *Riemann surface lamination*; our examples are individual levels in the profinite completion of the (complex) algebraic curve X_n .

Of course, one can proceed in analogy with this process for any example of *Veech curve*, by which we mean a Teichmüller disc with stabiliser a lattice Fuchsian group, and achieve further examples with the same property.

Note. A point which merits further exploration is the need to correlate the various modular groups arising in this situation. The result, in the restricted case where the coverings of moduli spaces are required to be finite cyclic and smooth, is a new type of generalised modular group which is called the *commensurability modular group* by Nag and Sullivan [55].

4.11 Return to Thurston's examples

In this section we explain how to generalise the examples discussed earlier.

Take two distinct finite systems of loops which together *fill up a compact surface*, that is, a pair $\alpha = \{\alpha_1, \dots, \alpha_m\}$, $\beta = \{\beta_1, \dots, \beta_n\}$, such that loops in the same set are pairwise disjoint and homotopically distinct, and with α -loops intersecting β -loops transversely in such a way that the complement is a union of topological discs, each with boundary containing an even number of segments of alternately α and β curves and at least four in number, so as to ensure that the loops have no inessential intersections. Choose centres for these disc components of $S - \{\alpha \cup \beta\}$; there is a dual cell decomposition of S into quadrilaterals which arises from joining centres of adjacent components by edges labelled as $v =$ vertical for α , $h =$ horizontal for β .

Now an \mathcal{F} -structure is defined on the surface in the same way as before, by viewing the surface as covered by unit square in the Euclidean plane, with the induced orientation and $\alpha =$ vertical, $\beta =$ horizontal labelling. This again defines an \mathcal{F} -structure on S . Furthermore, a twist about any α or β curve acts as an integer translation when viewed (in the holonomy representation) in this affine structure on the real plane below; the group generated by them is often a finite index subgroup of $SL_2(\mathbb{Z})$.

We note, however, that the subgroup of the stability group of a T-disc generated by twists may have infinite index, as in the example of a surface with genus 2 described by Earle and Gardiner [13]. It is constructed from a 6-by-1 rectangle – six squares in a line – with edge identifications given by a translation by 6 from one end to the other and four involutions, operating in pairs on top and bottom sets of edges, with two edges labelled a_j of length 1 and two labelled b_j of length 2, to produce a handle from each, to give a surface symbol $a_1 b_1 a_1^{-1} b_1^{-1} c a_2 b_2 a_2^{-1} b_2^{-1} c^{-1}$. It is not difficult to see that this structure has twist subgroup generated by the matrices

$$\begin{pmatrix} 1 & 6 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 6 & 1 \end{pmatrix},$$

which has infinite index in the modular group $\Gamma(1)$, while the full stabiliser is commensurable with $\Gamma(1)$.

It follows from Veech's work on the Teichmüller geodesic flow on the bundle of quadratic differentials Q_g over \mathcal{T}_g (see [71]) that the points of \mathcal{T}_g which possess such a representation are dense; we shall make use of this fact in the next chapter. A rather different density property for these structures was proved in a paper of H. Masur [46]: the set of Teichmüller-geodesics given by the original Thurston prescription (using single α and β curves) projects to a dense subset of the unit tangent bundle to moduli

space. As a consequence, there exist Teichmüller geodesic rays whose projection is dense in \mathcal{M}_g .

5 Triangle groups and modular families

The Hecke groups H_n and the more general triangle (p, q, ∞) -groups appear as a crucial ingredient in several areas of mathematics. We shall see in this section that they are naturally related to the construction of many Teichmüller disc families of Riemann surfaces and in particular to the examples discovered by Thurston and Veech.

5.1 Tessellations and triangle groups

Let X be a compact Riemann surface which admits a regular tessellation by hyperbolic n -gons. We describe how such a surface admits reconstruction from a torsion-free subgroup Γ of finite index in a Hecke group H_n ; this follows from a fundamental theorem about maps on surfaces (see for instance [66]). In particular, the case $n = 3$ relates surfaces X with a decomposition into equilateral triangles to the classical Farey tessellation of the hyperbolic plane associated with the classical modular group, $H_3 = \Gamma(1) = \mathrm{PSL}_2(\mathbb{Z})$. The compact surface $X = X_\Gamma$ obtained from the (finite co-volume) Fuchsian group Γ has finitely many *cusps* added to the (finite area) surface \mathcal{U}/Γ in the usual way, one for each Γ -orbit of boundary points at which the stabiliser is non-trivial. The holomorphic projection $\pi: \mathcal{U}/\Gamma \rightarrow \mathcal{U}/H_n \cong \mathbb{C}$, induced from the identity map on the universal covering \mathcal{U} , extends to a Belyi map $f: X \rightarrow \mathbb{P}^1$ from the compactification, with each cusp of X mapped by f to the single cusp $\{\infty\}$ of H_n which compactifies the plane \mathcal{U}/H_n to $\mathbb{C} \cup \infty = \mathbb{P}^1$. The tessellation of X is then determined by projecting part of the standard H_n -invariant subdivision of \mathcal{U} into $\{2, n, \infty\}$ triangles. Such a triangle arose in the preceding section as half of a fundamental domain F : in that picture, the G_n -orbit of the hyperbolic line $L = \{\Re z = \cot(\pi/n)\}$ determines a tiling of \mathcal{U} by ideal n -gons, projecting to a finite tessellation of X whose vertices correspond to the cusps of Γ . To say that the tessellation of X by n -gons is *regular* means that a fixed number m of polygons surround each vertex, which is equivalent to the statement that the total width of each cusp is the same fixed number.

Notes. 1. The data of a regular m -valent tiling by n -gons of a compact surface can be specified combinatorially by prescribing a torsion-free subgroup of a $\{2, n, m\}$ -triangle group, but it is the cusped version of the picture that provides the link with Teichmüller surface families.

2. In [66] a convenient method is given to generate *all* hyperbolic surface tessellations in similar fashion, by a universal construction in \mathcal{U} governed by representative subgroups of the modular (cartographic) subgroup $\Gamma_0(2)$. This will be employed

to give a succinct description of a wide class of modular curve families in the next subsection.

3. The equivalent combinatorial concept for more general triangle $\{p, q, \infty\}$ -groups is a *hypermap* (see [33]): here the corresponding universal construction involves the principal congruence subgroup $\Gamma(2)$ (which has index 2 in $\Gamma_0(2)$) and subgroups Γ of finite index in that group.

5.2 Associated Teichmüller families

We wish to produce, from the covering data $\{X, f\}$ of the preceding section, a corresponding curve in the moduli space of X . This will apply not only to the Riemann surfaces tiled by regular n -gons, which correspond to ‘balanced dessins’, but also to the broader class of complex orbifold structures given as *clean Belyi coverings* of \mathbb{P}^1 ; in this case the surface is the quotient of the disc by a finite index subgroup of a (p, q, ∞) -triangle group $\Gamma_{p,q}$. The collection of pairs (algebraic curve & map) so generated are part of the totality of dessins and of course, by the theorem of Belyi, each is defined over a number field – in fact, they must be part of the $\overline{\mathbb{Q}}$ -structure of the modular varieties themselves by Weil’s results [73] on the field of definition for a variety. Further discussion of this may be found in [74] and in [18].

By the classification of surface maps and hypermaps, we may assume that our covering is given by a Fuchsian group Γ of finite index in one or another of the groups $\Gamma_0(2)$, $\Gamma(2)$ or $\Gamma_{p,q}$. As before, X will denote the compact surface obtained from filling in the cusps on \mathcal{U}/Γ .

Take a tessellation of the hyperbolic plane \mathcal{U} by a standard fundamental set Q for $\Gamma_0(2)$, which is a hyperbolic quadrangle with ideal vertices at $-1, 0, 1$ and ∞ bisected into triangles by the imaginary axis \mathcal{I} . Now we indicate two further ways to approach the construction of quadratic differential forms for Γ , the first one similar to an observation of [13].

Let $\omega(z) = dz^2/(z^3 - z)$ denote the form on $Y = \mathbb{C}P^1$ with simple poles at the point set $B = \{0, \pm 1, \infty\}$. We consider the pullback of ω to a compact surface X through some Belyi function $F: X \rightarrow \mathbb{P}^1$, which we assume factors through another function $f: X \rightarrow Y$, i.e. $F = f^2$. The pullback to X , $f_*(\omega)$ will have a framed horizontal/vertical structure inherited from that of ω on Y , and if f is totally branched over B the form will be holomorphic on X .

For instance, this structure lifts to a geometrically similar \mathcal{F} -structure on the elliptic curve E given by the Weierstrass equation $w^2 = z^3 - z$, which is a double covering of Y branched over the set B , and which has a precise description in terms of the standard \wp -function for the lattice $\mathbb{Z} + \mathbb{Z}i$ corresponding to E .

It is not difficult to verify [13] (cf. [70]) that the affine stabiliser of the induced \mathcal{F} -structure on Y is the congruence subgroup $\Gamma(2)$. Also, one sees from this specific model that for all such pairs X, F the stabiliser of the T-disc for the form $F^*(\omega)$ has finite index in $\Gamma(1)$, by a simple topological argument. Of course, when lifted to

general coverings of Y , the form ω_f will have simple poles at any unramified points of $f^{-1}(B)$.

The *second method* works only for surfaces X covering the modular surface $X_0(2) = \mathcal{U}/\Gamma_0(2)$. It produces forms of an arithmetic nature on X from familiar modular forms and depends on a simple fact: the group $\Gamma_0(2)$ possesses a holomorphic cusp form ω_1 of weight 8 that is real on the imaginary axis \mathcal{I} with a zero of order 3 at each of the two cusps. This form is defined to be

$$\eta(\tau)\eta(2\tau)^4$$

where $\eta(\tau)$ is Dedekind’s well-known (character-automorphic) form for $\Gamma(1)$,

$$\eta(\tau) = e^{2\pi i\tau/24} \prod_{n=1}^{\infty} (1 - q^n), \quad \text{for } \tau \in \mathcal{H} \text{ and } q = e^{2\pi i\tau}.$$

For any subgroup $\Gamma < \Gamma(1)$ of finite index with all cusp widths even, one verifies easily that the lifted form to X_Γ has only even order zeros at the cusps, hence has a square root ω_1 which belongs to the space $S(\Gamma, 4)$ of weight 4 *cusp forms*. Furthermore, the form ω_1 has critical trajectories which interconnect the cusps of the surface X in some finite graph, which includes among its edges the projection of the axis \mathcal{I} and whose pattern depends only on the structure of the covering. In addition, for any conformal automorphism α of X which fixes the cusps, one can average the form over the elements α^ν to obtain a cusp form ω_2 with (at least) two independent directions in which all critical trajectories interconnect cusps and all noncritical trajectories are closed. This property, which is shared by the form ω defined earlier on Y , turns out to be crucial in ensuring a lattice group as Teichmüller-modular stabiliser.

We consider T-discs \mathbb{D}_ω for such forms. In order to proceed as canonically as possible, and to avoid the confusion arising from the plethora of inclusion relations between the various Teichmüller spaces associated to surface coverings, in what follows we place modular group activities within the universal group action of $\text{mod}(1)$, the group $QS(S^1)$ of all quasisymmetric homeomorphisms of the boundary circle, on Teichmüller space $\mathcal{T}(1)$, universal Teichmüller space, the normalised quotient by the subgroup (isomorphic to $G = \text{PSL}_2(\mathbb{R})$) of Möbius mappings of the circle.

Let $\Gamma \leq \Gamma_0(2)$ be a finite index normal subgroup contained in the kernel of the finite character $\chi_\omega: \Gamma_0(2) \rightarrow S^1$, and representing a surface $X = \mathcal{U}/\Gamma$, so that $\omega(\gamma(z)) = \chi_\omega(\gamma)\gamma'(z)^{-2}\omega(z)$ for all $\gamma \in \Gamma_0(2)$. If the form ω is Γ -invariant, we obtain an induced T-disc \mathbb{D}_ω in $\mathcal{T}(\Gamma) \cong \mathcal{T}_{g,n}$, where n is the total number of cusp and elliptic torsion points. This will then be mapped *injectively*, via the (holomorphic) mapping $\pi: \mathcal{T}(\Gamma) \rightarrow \mathcal{T}_g$ which forgets the data of all torsion and cusp points, landing in the Teichmüller space of the compactified surface X_0 .

Theorem 9. *The centre point of this composite embedding is a representative marked surface $[X_0]$ with a quadratic differential ω_0 such that the T-disc defined by ω_0 on X_0 realises the embedding $\pi \circ e_\omega$.*

Proof. The compact marked surface $[X_0]$ will certainly have a quadratic form ω_0 given by the obvious extension to the cusps from ω on \mathcal{U}/Γ . In fact, because we have a holomorphic image of the original T-disc, this disc is also a T-disc, since the distortion at each point is constant, with finitely many exceptional critical points. \square

We would like to characterise if possible the Veech T-discs presented in this manner, which have lattice modular stability group. From the earlier study of examples within Hecke groups H_n , it is clear that one way to do this is to restrict attention to particular covering surfaces X obtained from the commutator subgroup of the triangle group H , where there is always a good supply of (meromorphic) automorphic forms which have the following crucial property:

Definition. An automorphic form φ of weight 4 (or quadratic differential) for a finite co-volume Fuchsian group Γ is of *Jenkins–Strebel type* if the set of critical real trajectories of φ is compact in X_Γ .

For such forms, the complementary noncritical set – which has full measure in X – is a union of cylinders, each comprising a free homotopy class of closed trajectories. Incidentally, the forms occurring in Veech’s examples clearly have the J–S property in the vertical direction: one changes the direction of the trajectories from horizontal to vertical by a quarter-turn rotation of the Euclidean structure which means switching to the form $-\varphi$. It is natural to widen the class of J–S differentials to include all forms of finite L^1 norm which have *some* direction in which the trajectory pattern has this property.

We note that forms of J–S type are dense in the space $A_4^1(X)$ for *every* compact Riemann surface [12]; their construction is usually effected by either analytic or combinatorial means and no clear arithmetic connection has yet been made. For more details, including the essential part played by J–S forms in the solution of a host of extremal problems in conformal mapping and analysis, see for instance [13], [16], [67].

In all the examples described in Section 4, where Γ is a finite index subgroup of a Hecke triangle group, it was possible to use the structure of the covering (or equivalently a fundamental domain for the group) triangulated in canonical fashion, to produce J–S forms from the hyperbolic geometry; our next result follows this line of approach. We shall later (in Section 6) restrict attention to subgroups of the modular group $\mathrm{PSL}_2(\mathbb{Z})(= H_3)$, where a more direct approach is possible.

Theorem 10. *Let H be a Fuchsian triangle group, with (at least) one torsion point and one cusp; assume that there is a reflection symmetry Σ fixing a cusp and a torsion point. Let ω be an invariant holomorphic quadratic differential on some smooth Belyi cover $X_1 = \mathcal{U}/\Gamma_1$ of the compact orbifold Riemann sphere X_H , which is Σ -symmetric and corresponds to some character of H . Then the corresponding Teichmüller disc \mathbb{D}_ω in $\mathcal{T}(X_1) \cong \mathcal{T}_{g,k}$ has stabiliser in the corresponding modular group $\mathrm{Mod}_{g,k}$ isomorphic to a subgroup of H of finite index containing Γ_1 .*

Proof. We move to a conjugate of H in the unit disc so as to locate the fixed point of one torsion generator x at the centre 0 of the disc with the real ray from 0 to 1 fixed by the symmetry Σ and horizontal for the form ω . In the representation $\mathbb{D}_\omega = \pi \circ e_\omega(\mathbb{D})$, this implies that there is a finite order rotational symmetry for the Euclidean structure on the central reference surface of this disc, which is a compact abelian covering surface X_1 on which the form ω is defined, and with all cusps filled in. By the assumptions made on the group H and the form, the critical trajectories of ω emanate from cusps of Γ_1 and are contained in the H -orbit of the edge subset of some Σ -symmetric fundamental polygonal set $P \cup \bar{P}$, with $\bar{P} = \Sigma(P)$ and P containing the ray \mathcal{I} . It follows that the critical graph of ω is compact in X_1 , which implies that ω is of Jenkins–Strebel type, with all non-critical real trajectories closed.

Now consider the variation of complex structure on the surface as we follow the Teichmüller ray from the origin to $1 \in \partial\mathbb{D}$: 1 is a cusp of Γ_1 at which ω vanishes and this ray determines a deformation of X_1 by shrinking the lengths of all closed vertical trajectories to zero. We claim that the disc \mathbb{D} is invariant under a reducible (unipotent) mapping class which effects some composition of twists about a representative set of core loops, one from each cylinder in the non-critical trajectory pattern of the form ω on X_1 . To verify this, let t be the primitive parabolic element of H fixing the cusp 1, and consider the effect of applying conjugation ad_t to the (marked) subgroup Γ_1 : the rule $\text{ad}_t(k) = tkt^{-1}$ defines an element of $\text{mod } \Gamma_1 \cong \text{Mod}(\Gamma'_1)$, the mapping class group of the pointed surface X_1^* . This element is reducible in the Thurston classification of mapping classes [69], since it preserves the set of vertical core curves, but it acts as some cyclic permutation both on the set of cusps and also on the set of trajectory cylinders of ω . A mapping class of the required kind is induced on the compact surface by applying such an action at the level of Γ_1 using the smallest power t^ν of the element t that preserves each trajectory cylinder. If we denote by d the degree at the cusp 1 of the covering $X_0 \rightarrow \mathbb{P}_1 \cong X_H$, then $1 \leq \nu \leq d$. It follows that the modular stabiliser of \mathbb{D}_ω contains at least the nonelementary subgroup $\langle x, t^\nu \rangle$ of H .

To see that the modular stabiliser contains the subgroup Γ_1 , we recall the discussion of Section 4 where the structure of the Bers–Teichmüller fibre spaces was sketched out: in the current situation, we have a T-disc embedded in the universal Teichmüller space $\mathcal{T}(1)$ on which the group H acts as a subgroup of $\text{Mod}(1)$. To descend to $\mathcal{T}(X_1)$ and its fibre space $F(X_1)$ involves passing to the classes of qc mapping invariant under Γ_1 , isomorphic to the iterated fibre space $F^{(k)}(X_1)$ of moduli for k -pointed surfaces. The forgetful map π which fills in all cusps then defines a projection to the genus g Teichmüller space \mathcal{T}_g .

It follows from the definition of the successive Bers fibrations, beginning with the surface X_1 as fibre in $F(\Gamma_1)$ over the point $0 = [X_1] \in \mathcal{T}(X_1)$, and the Bers fiber space theorem (discussed in Section 4.3) that the stabiliser in $\text{mod}_{g,k} \cong \text{Mod}_{g,k+1}$ of the fibre space over the disc \mathbb{D}_ω contains the group Γ_1 . In fact this stabiliser is a subgroup of the k -strand braid group of the surface X_1 , and one has an extension of

groups within the modular group Mod_g of the smooth compact surface X_1 :

$$1 \longrightarrow \pi_1(X_1) \rightarrow E \rightarrow \pi_1(X_1^*) = \Gamma_1 \rightarrow 1.$$

The action of Γ_1 on the kernel (isomorphic to the fundamental group of the compact fibre surface) is by conjugation on E as a subgroup of $\text{mod}(X_1)$. After some careful working out, following the discussion in [13] of the relationship with the real group of automorphisms of the Euclidean planar structure on X_1 , $\text{GL}(2, \mathbb{R})$, this also turns out, to identify Γ_1 with the subgroup which Veech in [70] calls $\text{Aff}(u)$, the stabiliser of the \mathcal{F} -structure of the form ω .

Now there is a complex analytic subspace of the tangent bundle to $F^{(k)}(X_1)$ that is preserved under the action of E ; it constitutes one sheet of the Teichmüller geodesic flow on $\mathcal{T}(X_1)$ lifted to the fibre space. Alternatively, one may view it as an example of the type of construction of analytic complex surface given in [21]. It follows that the modular stabiliser of \mathbb{D}_ω has finite index in H . \square

We describe one way to prove existence of the forms required in this theorem. Work with an inclusion of co-compact Fuchsian groups $\Gamma_0 \leq H_0$ which provides a model of the ramified surface covering which completes the Hecke group picture: one follows the method detailed in [52] to construct generators of fractional weight for the ring of all holomorphic character-automorphic forms on the co-compact genus 0 Fuchsian group H_0 ; each of the generating forms has a simple zero at just one vertex. It is then possible to produce (by transferring to the punctured surface and lifting to the universal cover) weight 4 multiplicative holomorphic forms ω of the same type on the group H with zeros only at designated cusp vertex points of the hyperbolic tessellation of X determined by the group inclusion $\Gamma_0 < H_0$.

Corollary 1. *Let H be a triangle group of type $\{2, q, \infty\}$. Then there are holomorphic forms for the commutator surface of H whose modular stability group has finite index in H .*

The cusp forms of Jenkins–Strebel type for the groups arising in the Veech examples emerge directly from the Euclidean geometry on a given polygonal model surface, as Veech observes in [70]. The commutator surface has $\varepsilon = \text{gcd}(2, q)$ cusps, and the subgroup of mapping classes produced by the theorem at a torsion point of the disc coincides in this case with either the whole group H or a triangular subgroup of index 2, generated in each case by the two elements x^ε and t .

For further details and developments from the above result, the reader may consult [41], which discusses Teichmüller discs stabilised by a pseudo-Anosov mapping class, and [13], where a clear description is given of some further illuminating examples of polygonal type with detailed formulae for the relation between the T-disc deformation and the affine polygonal \mathcal{F} -structure from which it is defined.

For Riemann surfaces X given by finite index subgroups Γ of general triangle groups, one may construct cusp forms of J–S type using the methods indicated above, but a more natural way follows a (less direct) path through the tessellations by ideal

quadrangles induced from the standard fundamental region for $\Gamma(2)$; again one sees from the topology that the stabiliser in the Teichmüller modular group of the cusped surface X contains the subgroup Γ of $\Gamma(2)$ which classifies X and so is of finite index, by examination of the structural properties of the geometric fibre space over the Teichmüller disc.

5.3 Recent work on Veech curves in \mathcal{M}_g

It is unclear at present how to characterise the types of Fuchsian group that occur as the Veech (modular) stability group of a Teichmüller disc in given genus, despite the great activity in this area in recent years. For the case of genus 2, however, the picture is now completely understood, thanks largely to striking recent work by C. McMullen and others, which has produced a classification partly based on an explicit type of Euclidean decomposition of a compact (genus 2) surface X as a translation surface with rectangular L-shape. With specific side-lengths, these structures will have lattice Fuchsian stabiliser, and by bringing to the fore the dynamical structure of the Teichmüller flow on such an X , he provides a striking trichotomy for the case of abelian T-discs, where the form $q = \omega^2$ with ω an abelian 1-form. The papers [48] and [50] bring out the presence of hidden complex rank 2 symmetry in \mathcal{M}_2 by establishing the existence of T-discs \mathbb{D}_q with $q = \omega^2$, $\omega \in \Omega^1(X)$ with rectangular L-shape and side lengths in a quadratic number field $\mathbb{Q}(\sqrt{d})$, such that the projected image of the complex curve in the Siegel moduli space of dimension 2 abelian varieties has closure a *Hilbert modular surface*. This is a homogeneous algebraic surface, the quotient of a product of upper half planes $\mathcal{U} \times \mathcal{U}$ by an (indiscrete) image in $\mathrm{PSL}_2(\mathcal{O}_d)$, where \mathcal{O}_d denotes the ring of integers in the real quadratic field tied to the Euclidean structure on X , and it is a clear sign of the importance of Teichmüller–Veech curves that this study has uncovered such rich structures within these familiar varieties of dimension 3 – it is well known that the complete varieties $\hat{\mathcal{M}}_3$ and $\hat{\mathcal{A}}_3$ are birationally equal to $\mathbb{C}\mathbb{P}^3$. Hubert and Lelièvre [30] have produced a highly refined analysis of the special structures occurring for square-tiled surfaces in \mathcal{M}_2 , including asymptotic estimates of the growth of high genus curves of this type.

Other recent work by Hubert and Schmidt [31] and by McMullen [49] provides interesting geometrically defined examples of T-discs in low genus ($g \leq 4$) with infinitely generated Fuchsian stabilisers.

In a different but related direction concerning families of algebraic curves, work of Cohen and Wolfart [9] showed that uniformising hyperbolic discs for a large class of algebraic curves of genus $g \geq 2$ admit *modular embeddings* into Shimura varieties. This means that there is an embedding into a Siegel variety \mathcal{A}_g parametrising polarised abelian varieties of genus g , using constructions involving the hyperbolic plane representation of the curve. One of their construction methods is related to the special examples of Veech’s T-disc structures we examined where the quadratic differential is the square of a 1-form, and by this (completely independent) process it

follows that, for instance, many finite index subgroups of $\Gamma(1)$ embed in the Siegel modular group $\mathrm{Sp}(2g, \mathbb{Z})$ of the relevant genus as the symplectic monodromy group of the closed 1-form defining the branched Euclidean structure on the surface. An essentially equivalent construction of embedded T-discs in the Torelli–Siegel space of a Riemann surface was given much earlier by [40]; also compare with the related results mentioned above in the discussion of [48] on genus 2.

At the same time there have been important applications to the dynamical study of the bundle $\Omega^1(V_g)$ of holomorphic 1-forms on the Teichmüller curve V_g and the various subspaces into which it divides under the action of $\mathrm{SL}_2(\mathbb{R})$ on the \mathcal{F} -structure coordinates which each point of the fibre space $\Omega^1(V_g)$ represents. The (restricted) Teichmüller geodesic flow is determined by this action and specifying the partition of $2g - 2$ defined by the set of orders occurring in the divisor of zeros of the 1-form produces a connected component. Eskin and Okounkov [14] have computed the volumes of some of these spaces by linking them to the numbers of distinct branched coverings of tori of given degree. See also [39] for more precise geometric analysis of the dynamics of this restricted geodesic flow generated on the moduli space of abelian differentials.

5.4 Completion of a T-disc with large stabiliser

We continue with the assumptions of Section 5.2: Γ is a group of finite index in a group H of signature type $\{p, q, \infty\}$. Because the stability subgroup obtained in Theorem 10 is itself a lattice group, it makes sense in the light of the discussion in Section 4.8 to ask whether the modular image of the T-disc \mathbb{D} in \mathcal{M}_g has a natural completion. This follows naturally in fact: the process of adjoining a point at each cusp of the immersed surface occurs in the projective embedding of the *moduli space* $\hat{\mathcal{M}}_g$ of stable genus g algebraic curves constructed by Deligne and Mumford, [11], which is a projective (ergo compact Kähler) complex V-manifold containing \mathcal{M}_g as a Zariski open subset. For background on this completion of the space of non-singular curves the reader may consult [42], [23], [24] and [75], [76].

We refer again to the discussion in Section 5.2 before Theorem 10 as prelude to the next result, first stated in this form in [25], although it involves mainly a reinterpretation of parts of Veech’s work.

Theorem 11. *For each point $\{X\}$ of the moduli space \mathcal{M}_g represented by a compact surface X which is a clean Galois Belyi covering of $\mathbb{C}P^1$, there is a complete algebraic curve C in the stable compactification $\hat{\mathcal{M}}_g$ passing through $\{X\}$ isomorphic to some quotient surface of X by an abelian subgroup of $\mathrm{Aut}(X)$.*

Proof. The existence of the affine T-curves was proved in Theorem 10. Examples such as the original Veech–Wiman hyperelliptic curves and Fermat curves were met earlier. We consider the natural completion of such a curve in the projectively embedded variety $\hat{\mathcal{M}}_g$ of stable curves. The points of a completed curve C which lie in the

boundary divisor of $\hat{\mathcal{M}}_g$ then correspond, not necessarily bijectively, to surfaces with nodes obtained by passing to the cusps of the stability subgroup, which produces a noded Riemann surface by collapsing the vertical trajectories of the form. The curve C is an abelian quotient because of the construction we used.

There is a co-compact triangle group with signature $\{2, m, n\}$, where n is the l.c.m. of the ramification degrees at the points lying over ∞ , and a corresponding finite index subgroup representing the same curve with cyclic symmetry which has identical ramification structure with this Belyi covering and determines a smooth surface with the same underlying conformal structure as X . Since this is a smooth surface, the periods will in fact satisfy the condition that m divides the l.c.m. of 2 and n while n divides l.c.m. $\{2, m\}$. □

5.5 Consequences; complete curves in modular varieties

We add some brief comments on the last theorem.

Firstly, it is worth noting that the immersed curves C in moduli space are likely in general to have singularities such as self intersections, although these can always be removed by passage to a finite sheeted covering of $\hat{\mathcal{M}}_g$. In fact it follows immediately from a topological property of the orbifold structure that *there is a finite Galois cover of the modular variety which is a compact complex (projective) manifold.* (This is proved in [42].) However, the genus of the resulting lifted Teichmüller–Veech curve is then much larger and may not be easy to estimate.

A second point concerns the general occurrence of surface groups as subgroups in Teichmüller modular groups.

The hyperbolic plane $\mathcal{U} = \tilde{X} \cong \mathbb{D}$ forms the fibre over the point $x = [X]$ of \mathcal{M}_g , with $X = \mathcal{U}/K$, say, for some Fuchsian group $K \cong \pi_1(X)$. This determines a different holomorphic embedding of a disc into the space $F_{g,k} \cong F(K) = \mathcal{T}(K') \cong \mathcal{T}_{g,k+1}$, and in the situation of Section 5.2 the Fuchsian group K which is acting on \mathbb{D} , and which is naturally contained in the corresponding extended modular group $\text{mod}(K)$, survives the projection homomorphism to the group $\text{Mod}(K) \cong \text{Mod}_{g,k}$ for some $k > 0$. However, it was proved by Kra [41] (and independently by Nag) that such fibre discs in Teichmüller spaces cannot be Teichmüller-geodesic. Consequently these two constructions determine completely distinct ways to realise Fuchsian groups acting on the upper half plane as geometric subobjects of the Teichmüller modular transformation group actions $\text{Mod}_{g,k}$.

As a final point, we observe that the degree of any covering which may occur in Theorem 11 for a T-curve in a given moduli space is bounded above by a linear function of g , at most $4(g + 1)$ since this is the maximum order of any abelian automorphism group for a complete curve of genus $g \geq 2$. This bound may not be sharp. It is also unclear whether other complete (not necessarily Teichmüller geodesic) curves exist in \mathcal{M}_g with larger automorphism group.

6 Towards a space of all \mathcal{Q} -curves: a universal modular family

Among the candidate theories which point towards the desired goal of a universal space of $\mathcal{Q} = \overline{\mathbb{Q}}$ -rational curves, so far achieved only in $g = 2$, there are links with Teichmüller theory stemming from both the holomorphic and combinatorial sides of the picture. The relationship we pursue here connects a class of finite index subgroups of the modular group $\mathrm{PSL}(2, \mathbb{Z})$ and a subset of the space of conjugacy classes of co-compact Fuchsian groups and there is evidence which suggests a deeper resonance with algebraic number theory. In particular there is a clear relationship with a modular action of $\mathrm{Gal}(\overline{\mathbb{Q}})$ given in work of J. G. Thompson [68], which we summarise below.

6.1 A moduli space of Fuchsian groups

The space \mathcal{D} of all lattice subgroups (defined to be discrete subgroups with finite volume quotient) of the Lie group $G = \mathrm{PSL}(2, \mathbb{R})$ has a natural structure of metric space as defined by Chabauty [8]. Within this there is a subset \mathcal{D}_0 consisting of all lattice subgroups with co-compact quotient. According to a fundamental theorem of A. M. Macbeath [43] this set is open in \mathcal{D} . Furthermore the geometric isomorphism type of the discrete groups in a connected component is fixed, and each component $\mathcal{D}(\Gamma)$ is, by standard results on Teichmüller-theory described earlier, a fibre space over the space of moduli for X_Γ the corresponding type of hyperbolic closed surface or 2-orbifold, with the fibre over the modular point (conformal class of surface) $x = [\Gamma]$ isomorphic to the unit tangent bundle over the quotient surface/orbifold \mathcal{U}/Γ ; the base is the moduli space $\mathcal{M}(\Gamma)$. This follows from results by Weil, by Macbeath [43], [45] and by Harvey (see [23], Chapter 9).

A structure of stratified space in the sense of Whitney can also be put on \mathcal{D}_0 , which takes note of the various topological types of orbifold, and there is a natural way to form the relative closure of \mathcal{D}_0 in the Chabauty space of G ; see again for instance [23], Chapter 9, where it is shown that the topological boundary structure of the Deligne–Mumford boundary can be replicated in this setting.

Later in this chapter we shall formulate a theorem that makes the subset of conjugacy classes of finite index subgroups of (p, q, ∞) -triangle groups into both a set of \mathbb{Q} -rational base points for the moduli spaces of compact surfaces representable as some finite sheeted covering of the sphere – this is just a paraphrase of Belyi’s theorem – and also a data-bank for T-disc models of all the corresponding algebraic curves defined by these coverings within the space of G -lattices.

Thus for the case of closed surface moduli spaces, each \mathcal{M}_g will contain (after passing to a suitable finite sheeted covering $\tilde{\mathcal{M}}_g$) complex affine models of all curves of genus g which carry a critically framed J–S form; these occur as subvarieties defined within a specific model of the moduli space, presumably with the standard \mathbb{Q} -structure that it has.

6.2 A representation of the absolute Galois group

Let $\Gamma(1) = \text{PSL}(2, \mathbb{Z})$, the classical modular group, act as usual on the upper half plane by Möbius transformations and denote by \mathcal{U}^* the completion by the *cusps* of $\Gamma(1)$, $\mathcal{U}^* = \mathcal{U} \cup \mathbb{Q} \cup \infty$ and $X^*(1) = \mathcal{U}^* / \Gamma(1)$ the corresponding compact quotient surface. This is the unique compact Riemann surface which contains the open surface $X(1) \cong \mathbb{C}$, so it is of course the Riemann sphere, but provided with a wealth of further information of an arithmetic nature, stemming from the isomorphism with the moduli space of marked complex tori, elliptic curves with their special representations coming from the classic theory of elliptic modular functions. There are many sources which the reader may consult on this fascinating topic, for instance [37] and [65].

Following Thompson ([68]), we construct an exact sequence of Galois groups associated with a certain space of meromorphic modular functions on this rational bordification, which represents from one aspect an action of the *absolute Galois group* $\text{Gal}(\overline{\mathbb{Q}})$ on Belyi coverings of the sphere; this is related to work of Ihara and others on braid and modular group representations of the absolute Galois group. The field of functions to be deployed forms a part of the regular representation of the modular group $\Gamma(1)$ on the space $\mathcal{H}(\mathcal{U})$ of functions holomorphic on the upper half plane with convergent Laurent–Puiseux expansions at each cusp which have the additional property that they are algebraic in the standard local parameter. For instance, near ∞ we require an expansion in the variable $q^{1/N} = \exp(2\pi i \tau / N)$ for some integer $N > 0$.

Let \mathbb{F}_0 denote the field of all functions meromorphic in $\hat{\mathcal{U}}$ which determine functions on some finite sheeted cover of the modular surface $X(1)^*$. Clearly \mathbb{F}_0 contains the subfield $\mathbb{C}(j)$, where $j(\tau)$ denotes the classical modular invariant, and it is well known (see for instance [63], [65]) that when the group $\Gamma(1)$ acts on \mathbb{F}_0 by $\sigma(\gamma, f) = f \circ \gamma$, the fixed field is precisely $\mathbb{C}(j)$.

In fact, it is known by general results on modular functions that for any subgroup Γ of finite index in $\Gamma(1)$ there is a (meromorphic) function invariant under precisely Γ ; standard methods of Galois theory then imply that $\text{Gal}(\mathbb{F}_0 / \mathbb{C}(j))$ is $\widehat{\Gamma(1)}$, the profinite completion of the modular group.

Now consider the following extension field of $\mathbb{C}(j)$ introduced in [68].

Definition. The *field of algebraic modular functions* \mathbb{K} is the relative algebraic closure of $\mathbb{Q}(j)$ in \mathbb{F}_0 .

We write \mathbf{G} for the Galois group of this field over $\mathbb{Q}(j)$. Then there is a short exact sequence connecting Galois groups for the field extension.

Theorem 12. *The subgroup fixing the intermediate field $\overline{\mathbb{Q}}(j)$ is isomorphic to $\widehat{\Gamma(1)}$ and there is a split exact sequence*

$$1 \rightarrow \widehat{\Gamma(1)} \longrightarrow \mathbf{G} \longrightarrow \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \rightarrow 1. \tag{1}$$

Here a constant field Galois automorphism $\sigma \in \text{Aut } \overline{\mathbb{Q}}$ acts on the field of algebraic modular functions via the action on coefficients of the polynomial ring $\overline{\mathbb{Q}}(j)[U]$: any $f \in \mathbb{K}$ is a root of some polynomial $P(j, U)$ and, by an observation of Thompson (see [68]), the image P^σ also has a root in \mathbb{K} .

The extension is split by the section $s : \text{Gal}(\overline{\mathbb{Q}}) \rightarrow \mathbf{G}$ defined in the following way: Let $f \in \mathbb{K}$ with $f = \sum_{n \geq M} a_n q^{n/N}$ near ∞ and let $\sigma \in \text{Gal}(\overline{\mathbb{Q}})$. Then

$$f \mapsto s^\sigma(f) = \sum_n a_n^\sigma q^{n/N}. \tag{2}$$

This establishes a base for a common action of the absolute Galois group and mapping class groups of all topological types on certain subspaces of the graded ring of modular forms. For the case of congruence subgroups, it is then possible to establish an arithmetic model for the function fields of all levels (this is done for instance in [65]) and generators for the ring of modular forms with (algebraic) integral q -coefficients at ∞ . In this sense, at least, the action of $\overline{\mathbb{Q}}$ on (Galois) dessins can be subsumed into this setting of a field of functions on \mathcal{U} .

For noncongruence subgroups, none of this arithmetic approach applies directly due to the lack of general methods such as a satisfactory theory of Hecke operators, but over a complex coefficient field one can define as usual the correct (integer weight k) action of $\gamma \in \text{SL}_2(\mathbb{R})$ on $\mathcal{H}(\mathcal{U})$ as

$$f \mapsto f[\gamma]_k(z) := f(\gamma(z)) \cdot (cz + d)^{-k} \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{R}). \tag{3}$$

The space $\mathbb{A}_4 = \bigcup_{\Gamma \leq \Gamma(1)} A_4(\Gamma)$ of all modular forms of weight 4 with poles of order at most 1 at cusps of some finite index subgroup Γ and having algebraic Fourier coefficients may then represent an appropriate arithmetic core within the cotangent bundles to moduli spaces of curves: the linear action of $\Gamma(1)$ on spaces $A_{2k}(\Gamma)$ of forms of weight $2k$, $k \geq 1$ for a subgroup Γ intertwines with that of the absolute Galois group on q -expansion coefficients. This territory has not been much explored for general (i.e. non-congruence) subgroups of $\Gamma(1)$, but the reader may consult several articles [4], [62] where this approach is applied to determining fields of definition and other arithmetic problems for non-congruence modular subgroups.

6.3 Modular forms with closed trajectories

We return to the bridge between the class of Jenkins–Strebel modular forms and the curves in moduli space discussed in Section 5.

Beginning with a finite index subgroup Γ of the modular group $\Gamma(1)$, let $f \in A_4(\Gamma)$ be a holomorphic automorphic form of weight 4, non-vanishing in \mathcal{U} , with zeros at some (non empty) finite subset of the cusps of Γ which includes ∞ .

In all the cases discussed in Section 4, where Γ was a finite index subgroup of some Hecke triangle group, it was possible to use the structure of the covering, or equivalently a fundamental domain for the group, triangulated in a canonical way, to

produce J–S forms. We restrict attention now to the classical modular group where a more direct approach is possible.

We make two observations which will be preparation for our final result. First, given any compact surface X furnished with a Belyi covering f of the sphere, there is a subgroup Γ of finite index in the modular group $\Gamma(1)$ such that the covering f is represented away from $f^{-1}(\infty)$ by the induced projection map of X_Γ onto $X(1)$, and with dessin the inverse image of the standard modular triangle decomposition of the sphere (with barycentric triangulation). From the many alternative combinatorial descriptions of the same ramified covering pattern which come from the various refinements of this polygonal decomposition of X , we choose a representation which involves subgroups of the cartographic subgroup $\Gamma_0(2)$ so as to coordinate with the joint action $\rho: \mathbf{G} \rightarrow \text{Aut}(\mathbb{F})$ on modular forms outlined in the exact sequence (1).

If the pattern of cells is a geometrical subdivision of a regular polygonal tessellation with an even number $2n$ of sides for each polygon, then there is a dessin with \bullet vertices at the corners, \star vertices at midpoints of edges and with a \circ vertex at the centre of each cell and a unique form on the surface with zeros only at the points marked \circ and with compact critical graph. This dessin is obtained from a multivalued holomorphic form on the complex line $X_0^* = \mathcal{U}/\Gamma_0(2) \cup \infty \cong \mathbb{P}^1$ which vanishes at ∞ ; it arises by use of Singerman's theorem on universal tessellations ([66]) which gives an isomorphism between this covering and the canonical covering of \mathbb{P}^1 that realises the tessellation by regular $2n$ -gons by a specific finite index subgroup of the $\{2, 2n, \infty\}$ -Hecke group. If the cell pattern chosen is irregular at the vertices, then we need to make a preliminary extra choice of covering surface which does have such a regular structure; for instance we could take the intersection $\Gamma \cap \Gamma_0(2n)$, a normal subgroup of that congruence group.

The kind of \mathcal{F} -structure that we saw in sections 2 and 3 arose from the prescription of a quadratic differential with *two* interacting sets of closed trajectories and in addition a finite order mapping class, represented as a rotation of the coordinates for the \mathcal{F} -structure. In the Thurston–Veech square-tiled examples, there is a $\pi/2$ rotational switch from vertical to horizontal which comes in when viewed in the modular group framework, for the reason that we need to use the paving by Euclidean squares, whereas in the case of the patterns of regular $2n$ -gons one obtains a quadratic differential for a Hecke group with closed trajectories in directions making angle π/n . In that case, the rotation acts as a translation at the cusp ∞ by a rational fraction, which induces a multiplication by some $2n$ -th root of unity in the algebraic action ρ on q -expansions.

In this way we see that the appropriate class of modular form for our purposes should have *closed trajectories in at least two independent directions related by a root of unity*.

Definition. An automorphic form φ of weight 4 for Γ is called *critically framed* if every non-critical trajectory is closed for the \mathcal{F} -structures obtained in two independent rationally related directions for φ on $X = \mathcal{U}/\Gamma$.

It seems quite likely that the study of these purely geometric objects should have intrinsic arithmetic interest, as they form natural ways to parametrise curves defined over number fields analogous to the intensively studied modular elliptic curves; the model curves in this setting are, however, usually noncongruence quotients.

6.4 Arithmetically defined curves in moduli space

We can now formulate a final result, valid for any complete Belyi curve X of genus $g \geq 2$.

Theorem 13. *Given any pair X, β with X a genus g curve and $\beta: X \rightarrow X(1)$ a Belyi map, there is a corresponding Veech curve in \mathcal{M}_g , an affine algebraic curve over $\overline{\mathbb{Q}}$ arising from a critically framed modular form φ in $A_4^1(X)$ which determines a T-disc modular parametrisation e_φ centred at X .*

The proof follows directly from the definitions and preliminary discussion of the preceding section.

The set of all such triples X, β, ω forms a subset of a fibre space over the modular tower $\hat{\mathcal{M}} = \bigcup_{g \geq 2} \mathcal{M}_g$ on which the absolute Galois group acts in the manner of equation (3) on Fourier q -coefficients of weight 4 modular forms. One expects that some curves will have many such immersed models in their moduli space, just as the hyperelliptic curves $y^2 + x^n = 1$ have at least two, possibly more, distinct T-disc parametrisations.

Having seen how the complex geometry of moduli spaces is linked to dessins d'enfant by way of the collection of modular Veech curves, we close by indicating another approach to the relationship between moduli, combinatorial patterns on closed surfaces and hyperbolic geometry. In a paper of Macbeath ([44]) originating from the same period as the ideas of Grothendieck and Thurston, it was shown that the set of points of \mathcal{M}_g which are represented by hyperbolic structures with Dirichlet fundamental region having an inscribed circle, that is simultaneously tangent to all sides, corresponds precisely to the multitude of genus g permutation representations of the modular group $\Gamma(1)$; further evidence, perhaps, of how important mathematical ideas have a life of their own. Macbeath's points in \mathcal{M}_g do represent curves defined over a number field, by Belyi's theorem.

References

- [1] W. Abikoff, *The real analytic theory of Teichmüller space*. Lecture Notes in Math. 820, Springer, Berlin 1980. 268
- [2] L. V. Ahlfors, *Lectures on quasiconformal mappings*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1987; second edition, with supplemental chapters by C. J. Earle, I. Kra, M. Shishikura and J. H. Hubbard, Univ. Lecture Ser. 38, Amer. Math. Soc., Providence, RI, 2006. 267

- [3] L. Ahlfors and L. Bers, Riemann's mapping theorem for variable metrics. *Ann. of Math.* (2) 72 (1960), 385–404. [259](#)
- [4] A. O. L. Atkin and H. P. F. Swinnerton-Dyer, Modular forms on non-congruence subgroups. In *Combinatorics* (Los Angeles, Calif., 1968), Proc. Sympos. Pure Math. 19, Amer. Math. Soc., Providence, R.I., 1971, 1–25. [286](#)
- [5] G. V. Belyĭ, On Galois extensions of a maximal cyclotomic field. *Izv. Akad. Nauk SSSR Ser. Mat.* 43 (1979), 267–276; English transl. *Math. USSR-Izv.* 14 (1980), 247–256. [252](#)
- [6] L. Bers, Quasiconformal mappings and Teichmüller's theorem. In *Analytic functions*, Princeton Math. Ser. 24, Princeton University Press, Princeton, N.J., 1960, 89–119. [268](#)
- [7] L. Bers, Fiber spaces over Teichmüller spaces. *Acta Math.* 130 (1973), 89–126. [264](#)
- [8] C. Chabauty, Limite d'ensembles et géométrie des nombres. *Bull. Soc. Math. France* 78 (1950), 143–151. [284](#)
- [9] P. B. Cohen and J. Wolfart, Dessins de Grothendieck et variétés de Shimura. *C. R. Acad. Sci. Paris Sér. I Math.* 315 (1992), 1025–1028. [281](#)
- [10] P. Cohen and J. Wolfart, Modular embeddings for some non-arithmetic Fuchsian groups. *Acta Arith.* 56 (1990), 93–110.
- [11] 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. [282](#)
- [12] A. Douady and J. Hubbard, On the density of Strebel differentials. *Invent. Math.* 30 (1975), 175–179. [278](#)
- [13] C. J. Earle and F. P. Gardiner, Teichmüller disks and Veech's \mathcal{F} -structures. In *Extremal Riemann surfaces* (San Francisco, CA, 1995), Contemp. Math. 201, Amer. Math. Soc., Providence, R.I., 1997, 165–189. [254](#), [266](#), [271](#), [272](#), [274](#), [276](#), [278](#), [280](#)
- [14] A. Eskin and A. Okounkov, Asymptotics of numbers of branched coverings of a torus and volumes of moduli spaces of holomorphic differentials. *Invent. Math.* 145 (2001), 59–103. [282](#)
- [15] D. Gabai, Convergence groups are Fuchsian groups. *Ann. of Math.* (2) 136 (1992), 447–510. [262](#)
- [16] F. P. Gardiner, *Teichmüller theory and quadratic differentials*. John Wiley & Sons, New York 1987. [258](#), [259](#), [261](#), [262](#), [264](#), [266](#), [278](#)
- [17] F. P. Gardiner and N. Lakic, *Quasiconformal Teichmüller theory*. Math. Surveys Monogr. 76, American Mathematical Society, Providence, R.I., 2000. [259](#)
- [18] G. González-Díez, Variations on Belyi's theorem. *Quart. J. Math. Oxford* (2) 57 (2006), 339–354. [252](#), [276](#)
- [19] G. González Díez and W. J. Harvey, On complete curves in moduli space I. *Math. Proc. Cambridge Philos. Soc.* 110 (1991), 461–466. [254](#)
- [20] G. González-Díez and W. J. Harvey, On families of algebraic curves with automorphisms. In *Complex manifolds and hyperbolic geometry* (Guanajuato, 2001), Contemp. Math. 311, Amer. Math. Soc., Providence, R.I., 2002, 261–276.
- [21] P. A. Griffiths, Complex-analytic properties of certain Zariski open sets on algebraic varieties. *Ann. of Math.* (2) 94 (1971), 21–51. [280](#)

- [22] A. Grothendieck, Esquisse d'un programme. In *Geometric Galois actions*. 1. Around Grothendieck's "Esquisse d'un programme" (Luminy, France, August 1995), Lond. Math. Soc. Lect. Note Ser. 242, Cambridge University Press, Cambridge 1997, 5–48; English transl. "Sketch of a programme", 241–283. [251](#), [252](#), [254](#), [255](#), [256](#)
- [23] W. J. Harvey (ed.), *Discrete groups and automorphic functions* (Cambridge, July 28–August 15, 1975). Academic Press, London 1977. [258](#), [261](#), [282](#), [284](#)
- [24] W. J. Harvey, Modular groups and representation spaces. In *Geometry of group representations* (Boulder, CO, 1987), Contemp. Math. 74, Amer. Math. Soc., Providence, R.I., 1988, 205–214. [282](#)
- [25] W. J. Harvey, On certain families of compact Riemann surfaces. In *Mapping class groups and moduli spaces of Riemann surfaces* (Göttingen, June 24–28, 1991 and Seattle, WA, August 6–10, 1991), Contemp. Math. 150, Amer. Math. Soc., Providence, R.I., 1993, 137–148. [273](#), [282](#)
- [26] W. J. Harvey, Drawings, triangle groups and algebraic curves. KCL Preprint MTH-97-12, Feb. 1997. [250](#)
- [27] W. J. Harvey and C. Maclachlan, On mapping-class groups and Teichmüller spaces. *Proc. London Math. Soc.* (3) 30 (1975), 496–512. [261](#)
- [28] E. Hecke, Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichung. *Math. Ann.* 112 (1936), 664–699; reprinted in *Mathematische Werke*, Vandenhoeck & Ruprecht, Göttingen 1959, 591–626. [270](#)
- [29] F. Herrlich and G. Schmithüsen, Dessins d'enfants and origamis. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume II, EMS Publishing House, Zurich 2008. [255](#)
- [30] P. Hubert and S. Lelièvre, Prime arithmetic Teichmüller discs in $\mathcal{H}(2)$. *Israel J. Math.* 151 (2006), 281–321. [281](#)
- [31] P. Hubert and T. A. Schmidt, Veech groups and polygonal coverings. *J. Geom. Phys.* 35 (2000), 75–91. [281](#)
- [32] J. A. Jenkins, *Univalent functions and conformal mapping*. Ergeb. Math. Grenzgeb. Neue Folge 18, Springer-Verlag, Berlin 1958. [266](#)
- [33] G. Jones and D. Singerman, Belyĭ functions, hypermaps and Galois groups. *Bull. London Math. Soc.* 28 (1996), 561–590. [255](#), [256](#), [276](#)
- [34] S. P. Kerckhoff, The asymptotic geometry of Teichmüller space. *Topology* 19 (1980), 23–41. [263](#), [267](#)
- [35] S. P. Kerckhoff, The Nielsen realization problem. *Ann. of Math.* (2) 117 (1983), 235–265. [262](#), [263](#)
- [36] S. Kerckhoff, H. Masur, and J. Smillie, Ergodicity of billiard flows and quadratic differentials. *Ann. of Math.* (2) 124 (1986), 293–311. [266](#)
- [37] A. W. Knap, *Elliptic curves*. Math. Notes 40, Princeton University Press, Princeton, N.J., 1992. [285](#)
- [38] K. Kodaira, A certain type of irregular algebraic surfaces. *J. Analyse Math.* 19 (1967), 207–215. [254](#)

- [39] M. Kontsevich and A. Zorich, Connected components of the moduli spaces of Abelian differentials with prescribed singularities. *Invent. Math.* 153 (2003), 631–678. [282](#)
- [40] I. Kra, The Carathéodory metric on abelian Teichmüller disks. *J. Analyse Math.* 40 (1981), 129–143. [282](#)
- [41] I. Kra, On the Nielsen-Thurston-Bers type of some self-maps of Riemann surfaces. *Acta Math.* 146 (1981), 231–270. [280](#), [283](#)
- [42] E. Looijenga, Smooth Deligne-Mumford compactifications by means of Prym level structures. *J. Algebraic Geom.* 3 (1994), 283–293. [282](#), [283](#)
- [43] A. M. Macbeath, Groups of homeomorphisms of a simply connected space. *Ann. of Math.* (2) 79 (1964), 473–488. [270](#), [284](#)
- [44] A. M. Macbeath, Generic Dirichlet polygons and the modular group. *Glasgow Math. J.* 27 (1985), 129–141. [288](#)
- [45] A. M. Macbeath and D. Singerman, Spaces of subgroups and Teichmüller space. *Proc. London Math. Soc.* (3) 31 (1975), 211–256. [264](#), [270](#), [284](#)
- [46] H. Masur, Dense geodesics in moduli space. In *Riemann surfaces and related topics* (State University New York, Stony Brook, N.Y., 1978), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 417–438. [274](#)
- [47] H. Masur, Closed trajectories for quadratic differentials with an application to billiards. *Duke Math. J.* 53 (1986), 307–314. [254](#)
- [48] C. T. McMullen, Billiards and Teichmüller curves on Hilbert modular surfaces. *J. Amer. Math. Soc.* 16 (2003), 857–885. [281](#), [282](#)
- [49] C. T. McMullen, Teichmüller geodesics of infinite complexity. *Acta Math.* 191 (2003), 191–223. [281](#)
- [50] C. T. McMullen, Dynamics of $SL_2(\mathbb{R})$ over moduli space in genus 2. *Ann. of Math.* 165 (2007), to appear. [281](#)
- [51] C. T. McMullen, Prym varieties and Teichmüller curves. *Duke Math. J.* 133 (2006), 569–590.
- [52] J. Milnor, On the 3-dimensional Brieskorn manifolds $M(p, q, r)$. In *Knots, groups, and 3-manifolds*, Ann. of Math. Stud. 84, Princeton University Press, Princeton, N.J., 1975, 175–225. [270](#), [280](#)
- [53] S. Morita, *Geometry of characteristic classes*. Transl. Math. Monographs 199, American Mathematical Society, Providence, R.I., 2001. [263](#)
- [54] S. Nag, *The complex analytic theory of Teichmüller spaces*. Canad. Math. Soc. Ser. Monographs Adv. Texts, John Wiley & Sons, New York 1988. [258](#), [262](#), [264](#)
- [55] S. Nag and D. Sullivan, Teichmüller theory and the universal period mapping via quantum calculus and the $H^{1/2}$ space on the circle. *Osaka J. Math.* 32 (1995), 1–34. [274](#)
- [56] R. Nevanlinna et al. (eds.), *Analytic Functions*. Princeton Math. Ser. 24, Princeton University Press, Princeton N.J., 1960. [251](#)
- [57] B. Riemann, *Gesammelte mathematische Werke und wissenschaftlicher Nachlaß*. Dover Publications, Inc., New York, N.Y., 1953. [251](#)
- [58] H. L. Royden, Automorphisms and isometries of Teichmüller space. 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, 369–383. [261](#)

- [59] T. A. Schmidt, Klein's cubic surface and a 'non-arithmetic' curve. *Math. Ann.* 309 (1997), 533–539.
- [60] L. Schneps and P. Lochak (eds.), *Geometric Galois actions*. Vol. 1, 2 (Luminy, France, August 1995). London Math. Soc. Lecture Note Ser. 242, 243, Cambridge University Press, Cambridge 1997.
- [61] L. Schneps (ed.), *The Grothendieck theory of dessins d'enfants* (Luminy, France, April 1993). London Math. Soc. Lecture Note Ser. 200, Cambridge University Press, Cambridge 1994. [255](#), [256](#)
- [62] A. J. Scholl, Modular forms on noncongruence subgroups. In *Séminaire de Théorie des Nombres, Paris 1985–86*, Progr. Math. 71, Birkhäuser, Boston 1987, 199–206. [286](#)
- [63] J.-P. Serre, *Cours d'arithmétique*. Deuxième édition revue et corrigée. Presses Universitaires de France, Paris 1977. [285](#)
- [64] G. B. Shabat and V. A. Voevodsky, Drawing curves over number fields. In *The Grothendieck Festschrift*, Vol. III, Progr. Math. 88, Birkhäuser, Boston 1990, 199–227. [252](#), [256](#)
- [65] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*. Publ. Math. Soc. Japan 11, Princeton University Press, Princeton, N.J., 1994. [285](#), [286](#)
- [66] D. Singerman, Universal tessellations. *Rev. Mat. Univ. Complut. Madrid* 1 (1988), 111–123. [275](#), [287](#)
- [67] K. Strebel, *Quadratic differentials*. *Ergeb. Math. Grenzgeb.* (3) 5, Springer-Verlag, Berlin 1984. [266](#), [278](#)
- [68] J. G. Thompson, Galois groups. In *Groups – St Andrews 1989*, Vol. 2, London Math. Soc. Lecture Note Ser. 160, Cambridge University Press, Cambridge 1991, 455–462. [284](#), [285](#), [286](#)
- [69] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)* 19 (1988), 417–431. [257](#), [267](#), [279](#)
- [70] W. A. Veech, Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. Math.* 97 (1989), 553–583; corrigendum, *ibid.* 103 (1991), 447. [266](#), [269](#), [271](#), [272](#), [276](#), [280](#)
- [71] W. A. Veech, Moduli spaces of quadratic differentials. *J. Analyse Math.* 55 (1990), 117–171. [272](#), [273](#), [274](#)
- [72] W. A. Veech, The billiard in a regular polygon. *Geom. Funct. Anal.* 2 (1992), 341–379. [269](#), [273](#)
- [73] A. Weil, The field of definition of a variety. *Amer. J. Math.* 78 (1956), 509–524. [276](#)
- [74] J. Wolfart, The 'obvious' part of Belyi's theorem and Riemann surfaces with many automorphisms. In *Geometric Galois actions*, Vol. 1, London Math. Soc. Lecture Note Ser. 242, Cambridge University Press, Cambridge 1997, 97–112. [252](#), [276](#)
- [75] S. Wolpert, On the homology of the moduli space of stable curves. *Ann. of Math.* (2) 118 (1983), 491–523. [282](#)
- [76] S. Wolpert, On obtaining a positive line bundle from the Weil-Petersson class. *Amer. J. Math.* 107 (1985), 1485–1507. [262](#), [282](#)

Chapter 6

On the boundary of Teichmüller disks in Teichmüller and in Schottky space

Frank Herrlich and Gabriela Schmithüsen

Contents

1	Introduction	294
2	Geodesic rays, Teichmüller disks and Teichmüller curves	300
	2.1 Teichmüller deformations	300
	2.2 Geodesic rays	303
	2.3 Teichmüller disks	303
	2.3.1 Teichmüller disks as a collection of geodesic rays	304
	2.3.2 Teichmüller disks by affine deformations	304
	2.3.3 Beltrami differentials	310
	2.4 Teichmüller curves	311
	2.4.1 Veech groups	311
	2.4.2 The action of the Veech group on the Teichmüller disk	312
	2.4.3 Veech groups and Teichmüller curves	314
3	Braungardt's construction of $\overline{\mathcal{T}}_{g,n}$	314
	3.1 Coverings with cusps	315
	3.2 The cusped universal covering of $\overline{M}_{g,n}$	317
	3.3 Teichmüller structures	319
4	Boundary points of Teichmüller curves	321
	4.1 Hitting the boundary via a Jenkins–Strebel ray	322
	4.1.1 X as patchwork of rectangles	323
	4.1.2 Stretching the cylinders	325
	4.1.3 S as patchwork of double annuli	326
	4.1.4 Contracting the central lines	327
	4.1.5 The end point of the Jenkins–Strebel ray	328
	4.1.6 Convergence	329
	4.2 Boundary points of Teichmüller disks	330
	4.2.1 Extending Teichmüller embeddings to the cusps of the mirror Veech group	331
	4.2.2 Boundary of Teichmüller disks that lead to Teichmüller curves	334
5	Schottky spaces	335
	5.1 Schottky coverings	336
	5.2 Relation to Teichmüller space	339

5.3	Schottky coverings of stable Riemann surfaces	341
5.4	$\bar{\mathcal{T}}_g$ as quotient of $\bar{\mathcal{T}}_g$	343
5.5	Teichmüller disks in Schottky space	346
	References	347

1 Introduction

One of the original motivations that led to the discovery of Teichmüller space was to better understand the classification of Riemann surfaces. Riemann himself already saw that the compact Riemann surfaces of genus g with n marked points on it depend on $3g - 3 + n$ complex parameters (if this number is positive). More precisely, there is a complex analytic space $M_{g,n}$ whose points correspond in a natural way to the isomorphism classes of such Riemann surfaces. $M_{g,n}$ is even an algebraic variety, but its geometry is not easy to understand. Most of the basic properties are known today, but many questions on the finer structure of $M_{g,n}$ are still open.¹

Many classification problems become more accessible if the objects are endowed with an additional structure or *marking*. The general strategy is to first classify the marked objects and then, in a second step, to try to understand the equivalence relation that forgets the marking. The markings that Teichmüller introduced for a compact Riemann surface X consist of orientation preserving diffeomorphisms $f: X_{\text{ref}} \rightarrow X$ from a reference Riemann surface X_{ref} to X . Markings (X, f) and (X', f') are considered the same if $f' \circ f^{-1}$ is homotopic to a biholomorphic map. Thus different markings of a fixed Riemann surface differ by a homotopy class of diffeomorphisms of X_{ref} . In other words the *mapping class group* (or *Teichmüller modular group*)

$$\Gamma_g = \text{Diffeo}^+(X_{\text{ref}}) / \text{Diffeo}^0(X_{\text{ref}}) \quad (1.1)$$

acts on the set \mathcal{T}_g of all marked Riemann surfaces of genus g , and the orbit space \mathcal{T}_g / Γ_g is equal to M_g (here $\text{Diffeo}^+(X_{\text{ref}})$ denotes the group of orientation preserving diffeomorphisms of X_{ref} and $\text{Diffeo}^0(X_{\text{ref}})$ the subgroup of those that are homotopic to the identity).

Teichmüller discovered that in each homotopy class of diffeomorphisms between compact Riemann surfaces X and X' there is a unique “extremal mapping”, i.e. a quasiconformal map with minimal dilatation. The logarithm of this dilatation puts a metric on \mathcal{T}_g , the *Teichmüller metric*. With it \mathcal{T}_g is a complete metric space, diffeomorphic to \mathbb{R}^{6g-6} , and Γ_g acts on \mathcal{T}_g by isometries. There is also a structure as complex manifold on \mathcal{T}_g , for which the elements of Γ_g act holomorphically and thus make the quotient map $\mathcal{T}_g \rightarrow M_g$ into an analytic map between complex spaces.

That the complex structure on \mathcal{T}_g is the “right one” for the classification problem can be seen from the fact that there is a family \mathcal{C}_g of Riemann surfaces over \mathcal{T}_g which in a very precise sense is universal. This family can be obtained as follows: By the uniformization theorem, the universal covering of a compact Riemann

¹Although we consider this general setting in a large part of this chapter, we shall restrict ourselves in this introduction to the case $n = 0$ and write, as usual, M_g instead of $M_{g,0}$ (and later \mathcal{T}_g instead of $\mathcal{T}_{g,0}$).

surface X of genus $g \geq 2$ is (isomorphic to) the upper half plane \mathbb{H} . Any marking $f: X_{\text{ref}} \rightarrow X$ induces an isomorphism f_* from $\pi_g = \pi_1(X_{\text{ref}})$, the fundamental group of the reference surface, to $\pi_1(X)$. We may obtain a holomorphic action of π_g on $\mathcal{T}_g \times \mathbb{H}$ as follows: for $\gamma \in \Gamma_g$, $x = (X, f) \in \mathcal{T}_g$ and $z \in \mathbb{H}$ put

$$\gamma(x, z) = (x, f_*(\gamma)(z)),$$

where we identify $\pi_1(X)$ with the group of deck transformations of the universal covering $\mathbb{H} \rightarrow X$. The quotient $\mathcal{C}_g = (\mathcal{T}_g \times \mathbb{H})/\pi_g$ is a complex manifold with a natural projection $p: \mathcal{C}_g \rightarrow \mathcal{T}_g$; the fibre $p^{-1}(X, f)$ is isomorphic to X . Moreover p is proper and therefore $p: \mathcal{C}_g \rightarrow \mathcal{T}_g$ is a *family of Riemann surfaces*. The representation of \mathcal{C}_g as a quotient of a manifold by an action of π_g is called a *Teichmüller structure* on this family. It follows from results of Bers on the uniformization of families (see e.g. [4, Theorem XVII]) that this family is *universal*, i.e. every other family of Riemann surfaces of genus g with a Teichmüller structure can be obtained as a pullback from $p: \mathcal{C}_g \rightarrow \mathcal{T}_g$. In a more fancy language: \mathcal{T}_g is a fine moduli space for Riemann surfaces of genus g with Teichmüller structure.

It follows by the same arguments that for any family $\pi: \mathcal{C} \rightarrow S$ of Riemann surfaces (over some complex space S) there is an analytic map $\mu = \mu_\pi: S \rightarrow M_g$, which maps $s \in S$ to the point in M_g that corresponds to the isomorphism class of the fibre $\pi^{-1}(s)$. Unfortunately, Γ_g does not act freely on \mathcal{T}_g ; therefore the quotient of \mathcal{C}_g by the action of Γ_g does not give a universal family over M_g : the fixed points of elements in Γ_g correspond to automorphisms of the Riemann surface, and the fibre over $[X] \in M_g$ in the family $\mathcal{C}_g/\Gamma_g \rightarrow M_g$ is the Riemann surface $X/\text{Aut}(X)$ (whose genus is strictly less than g if $\text{Aut}(X)$ is nontrivial). As a consequence, M_g is not a fine moduli space for Riemann surfaces, but only a “coarse” one (see e.g. [16, 1A] for a precise definition of fine and coarse moduli spaces).

There are several equivalent ways to define markings of Riemann surfaces and to describe Teichmüller space. Instead of classes of diffeomorphisms $f: X_{\text{ref}} \rightarrow X$ often conjugacy classes of group isomorphisms $\pi_g \rightarrow \pi_1(X)$ are used as markings. For the purpose of this chapter the approach to Teichmüller space via Teichmüller deformations is very well suited; it is developed in Section 2.1. The starting point is the observation that a holomorphic quadratic differential q on a Riemann surface X defines a flat structure μ on $X^* = X - \{\text{zeroes of } q\}$. Composing the chart maps of μ with a certain (real) affine map yields a new point in \mathcal{T}_g . Any point in \mathcal{T}_g is in a unique way such a Teichmüller deformation of a given base point $(X_{\text{ref}}, \text{id})$, cf. Section 2.2.

The main objects of interest in this chapter are *Teichmüller embeddings*, i.e. holomorphic isometric embeddings $\iota: \mathbb{H} \rightarrow \mathcal{T}_g$ (or $\iota: \mathbb{D} \rightarrow \mathcal{T}_g$), where \mathbb{H} (resp. \mathbb{D}) is given the hyperbolic metric and \mathcal{T}_g the Teichmüller metric, see Definition 2.4. The restriction of ι to a hyperbolic geodesic line in \mathbb{H} (or \mathbb{D}) is then a (real) geodesic line in \mathcal{T}_g in the usual sense. The image Δ_ι of such an embedding ι is called a *Teichmüller geodesic* or *Teichmüller disk* in \mathcal{T}_g . There are plenty of Teichmüller disks in \mathcal{T}_g . To see this note first that the tangent space to \mathcal{T}_g at a point $x = (X, f) \in \mathcal{T}_g$ is naturally isomorphic to the vector space $Q_X = H^0(X, \Omega_X^{\otimes 2})$ of holomorphic quadratic

differentials on X (this results from the Bers embedding of \mathcal{T}_g as a bounded open subdomain of Q_X). We shall explain in Section 2.3 in three different ways how one can, for a given holomorphic quadratic differential q on a Riemann surface X , construct a Teichmüller embedding $\iota: \mathbb{D} \rightarrow \mathcal{T}_g$ with $\iota(0) = x$ and $\iota'(0) = q$. This shows that for any $x \in \mathcal{T}_g$ and any (complex) tangent vector at x there is a Teichmüller disk passing through x in direction of the given tangent vector.

There are several natural and closely related objects attached to a Teichmüller disk Δ_ι (or a Teichmüller embedding $\iota: \mathbb{H} \rightarrow \mathcal{T}_g$): The first is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$ called the (*projective*) *Veech group* $\overline{\Gamma}_\iota$, cf. Section 2.4.1. If q is the quadratic differential on the Riemann surface X by which ι is induced, $\overline{\Gamma}_\iota$ consists of the derivatives of those diffeomorphisms of X that are affine with respect to the flat structure defined by q . Veech showed that this subgroup of $\mathrm{PSL}_2(\mathbb{R})$ is always discrete ([34, Proposition 2.7]).

A second group naturally attached to ι is the *stabilizer*

$$\mathrm{Stab}(\Delta_\iota) = \{\varphi \in \Gamma_g : \varphi(\Delta_\iota) = \Delta_\iota\}$$

of Δ_ι in the Teichmüller modular group. The pointwise stabilizer

$$\mathrm{Stab}^0(\Delta_\iota) = \{\varphi \in \Gamma_g : \varphi|_{\Delta_\iota} = \mathrm{id}_{\Delta_\iota}\}$$

is a finite subgroup of $\mathrm{Stab}(\Delta_\iota)$, and $\mathrm{Stab}(\Delta_\iota)/\mathrm{Stab}^0(\Delta_\iota)$ then is (via ι) a group of isometries of \mathbb{H} and thus a subgroup of $\mathrm{PSL}_2(\mathbb{R})$. This subgroup coincides with the projective Veech group $\overline{\Gamma}_\iota$, see Section 2.4.3.

Given a Teichmüller embedding ι we are also interested in the image C_ι of Δ_ι in the moduli space M_g . The map $\Delta_\iota \rightarrow C_\iota$ obviously factors through the Riemann surface $\mathbb{H}/\overline{\Gamma}_\iota$ or rather through its mirror image V_ι , see Section 2.4 and in particular 2.4.3. The typical case seems to be $\mathrm{Stab}(\Delta_\iota) = \{\mathrm{id}\}$ (although it is not trivial to give explicit examples). Much attention has been given in recent years to the other extreme case that $\overline{\Gamma}_\iota$ is a lattice in $\mathrm{PSL}_2(\mathbb{R})$. Then V_ι is of finite hyperbolic volume and hence a Riemann surface of finite type, or equivalently an algebraic curve. In this case the induced map $V_\iota \rightarrow C_\iota$ is birational (see [11]), i.e. V_ι is the desingularization (or normalization) of C_ι . It follows from a result of Veech ([34]) that V_ι (and hence also C_ι) cannot be projective. If $\overline{\Gamma}_\iota$ is a lattice, the affine curve C_ι is called a *Teichmüller curve*, cf. Section 2.4. First examples were given by Veech [34]; in them, $\overline{\Gamma}_\iota$ is a hyperbolic triangle group. Later more examples with triangle groups as Veech groups were found, see [20] for a comprehensive overview and [6] for recent results. Explicit examples for Teichmüller curves also with non triangle groups as Veech groups can be found e.g. in [27], [9] and [23]. Möller has shown ([28]) that every Teichmüller curve is, as a subvariety of M_g , defined over a number field. This implies that there are at most countably many Teichmüller curves.

A special class of Teichmüller curves is obtained by *origamis* (or *square-tiled surfaces*). They arise from finite coverings of an elliptic curve that are ramified over only one point. Given such a covering $p: X \rightarrow E$, the quadratic differential

$q = (p^*\omega_E)^2$ (where ω_E is the invariant holomorphic differential on E) induces a Teichmüller embedding whose Veech group is commensurable to $\mathrm{SL}_2(\mathbb{Z})$, see [15]. Lochak proposed in [23] a combinatorial construction for such coverings (which led to the name “origami”), and Schmithüsen [30] gave a group theoretic characterization of the Veech group. In [31], origamis and their Veech groups are systematically studied and numerous examples are presented. Origamis in genus 2 where q has one zero are classified in [19]. Using the description of origamis by gluing squares it is not difficult to see that there are, for any $g \geq 2$, infinitely many Teichmüller curves in M_g that come from origamis. In genus 3 there is even an explicit example of an origami curve that is intersected by infinitely many others, see [18].

We want to study boundary points of Teichmüller disks and Teichmüller curves; by this we mean, for a Teichmüller embedding $\iota: \mathbb{H} \rightarrow \mathcal{T}_g$, the closures of Δ_i and C_i in suitable (partial) compactifications of \mathcal{T}_g and M_g , respectively. For the moduli space we shall use the compactification \bar{M}_g by stable Riemann surfaces. Here a one-dimensional connected compact complex space X is called a *stable Riemann surface* if all singular points of X are ordinary double points, i.e. have a neighbourhood isomorphic to $\{(z, w) \in \mathbb{C}^2 : z \cdot w = 0, |z| < 1, |w| < 1\}$; moreover we require that every irreducible component L of X that is isomorphic to the projective line $\hat{\mathbb{C}} = \mathbb{P}^1(\mathbb{C})$ intersects $\overline{X - L}$ in at least three points. It was shown by Deligne and Mumford ([10]) that stable Riemann surfaces are classified by an irreducible compact variety \bar{M}_g that, like M_g , has the quality of a coarse moduli space. In fact, with the approach of Deligne–Mumford it is possible to classify stable algebraic curves over an arbitrary ground field: they construct a proper scheme over \mathbb{Z} of which \bar{M}_g is the set of complex-valued points. Some years later, Knudsen [22] showed that \bar{M}_g is a projective variety.

If $\iota: \mathbb{H} \rightarrow \mathcal{T}_g$ is a Teichmüller embedding such that C_i is a Teichmüller curve, the closure \bar{C}_i of C_i in \bar{M}_g is Zariski closed and therefore a projective curve. In particular, $\bar{C}_i - C_i$ consists of finitely many points, called the *cusps* of C_i . It is very interesting to know, for a given Teichmüller curve C_i , the number of cusps and the stable Riemann surfaces that correspond to the cusps. In the case that ι is induced by an origami there is an algorithm that determines (among other information) the precise number of cusps of C_i , see [30].

The boundary $\partial M_g = \bar{M}_g - M_g$ is a divisor, i.e. a projective subvariety of (complex) codimension 1. It has irreducible components $D_0, D_1, \dots, D_{[\frac{g}{2}]}$; the points in D_0 correspond to irreducible stable Riemann surfaces with a double point, while for $i = 1, \dots, [\frac{g}{2}]$, D_i classifies stable Riemann surfaces consisting of two nonsingular irreducible components that intersect transversally, one of genus i and the other of genus $g - i$. The combinatorial structure of the intersections of the D_i is best described in terms of the *intersection graph*. For a stable Riemann surface X , we define a graph $\Gamma(X)$ as follows: the vertices of $\Gamma(X)$ are the irreducible components of X , the edges are the double points (connecting two irreducible components of X which need not be distinct). For every graph Γ let $\bar{M}_g(\Gamma)$ be the set of points in \bar{M}_g corresponding to

stable Riemann surfaces with intersection graph isomorphic to Γ . It is not hard to see that for a given genus g , there are only finitely many graphs Γ with nonempty $\bar{M}_g(\Gamma)$, and that the $\bar{M}_g(\Gamma)$ are the strata of a stratification of \bar{M}_g . This means that each $\bar{M}_g(\Gamma)$ is a locally closed subset of \bar{M}_g (for the Zariski topology), that \bar{M}_g is the disjoint union of the $\bar{M}_g(\Gamma)$, and that the closure of each $\bar{M}_g(\Gamma)$ is a finite union of other $\bar{M}_g(\Gamma')$. A natural question in our context is: which $\bar{M}_g(\Gamma)$ contain cusps of Teichmüller curves? In [24] Maier showed that if Γ has no “bridge”, i.e. no edge e such that $\Gamma - e$ is disconnected, then the stratum $\bar{M}_g(\Gamma)$ contains points on a compactified Teichmüller curve \bar{C}_l with a Teichmüller embedding ι that corresponds to an origami. Möller and Schmithüsen observed that this condition on the graph is necessary if the Teichmüller curve comes from a quadratic differential which is the square of a holomorphic 1-form (or equivalently from a translation structure on X^*).

Most of our knowledge about cusps of Teichmüller curves comes from studying boundary points of Teichmüller disks in a suitable extension of Teichmüller space. Several different boundaries for Teichmüller space with very different properties have been studied, like the Thurston boundary or the one coming from the Bers embedding. In the framework of this chapter we look for a space $\bar{\mathcal{T}}_g$ in which \mathcal{T}_g is open and dense such that the action of the group Γ_g extends to an action on $\bar{\mathcal{T}}_g$, and the quotient space $\bar{\mathcal{T}}_g/\Gamma_g$ is equal to \bar{M}_g . Such a space is the “augmented” Teichmüller space $\hat{\mathcal{T}}_g$ introduced by Abikoff [1]. The points in $\hat{\mathcal{T}}_g$ are equivalence classes of pairs (X, f) , where X is a stable Riemann surface of genus g and $f: X_{\text{ref}} \rightarrow X$ is a *deformation*. This is a continuous surjective map such that there are finitely many loops c_1, \dots, c_k on X_{ref} with the property that f is a homeomorphism outside the c_i and maps each c_i to a single point P_i on X . Abikoff defined a topology on this space and showed that the quotient for the natural action of Γ_g on the pairs (X, f) is the moduli space \bar{M}_g as a topological space.

In his thesis [8], Braungardt introduced the concept of a covering of a complex manifold S with cusps over a divisor D . He showed that under mild assumptions on S there exists a universal covering \tilde{X} of this type which extends the usual holomorphic universal covering of $S - D$ by attaching “cusps” over D . \tilde{X} is no longer a complex manifold or a complex space, but Braungardt introduced a natural notion of holomorphic functions in a neighbourhood of a cusp and thus defined a sheaf $\mathcal{O}_{\tilde{X}}$ of rings (of holomorphic functions) on \tilde{X} . In this way \tilde{X} is a locally complex ringed space, and the quotient map $\tilde{X} \rightarrow \tilde{X}/\pi_1(S - D) = S$ is analytic for this structure. When applied to $S = \bar{M}_g$ and $D = \partial M_g$, Braungardt showed that the universal covering $\bar{\mathcal{T}}_g$ of \bar{M}_g with cusps over ∂M_g is, as a topological space with an action of Γ_g , homeomorphic to Abikoff’s augmented Teichmüller space. We shall reserve the symbol $\bar{\mathcal{T}}_g$ exclusively for this space (considered as a locally ringed space). In Section 3, we review Braungardt’s construction and results.

Our key technique to investigate boundary points of Teichmüller disks is the use of *Jenkins–Strebel rays*, see Definition 4.4. By this we mean a geodesic ray in \mathcal{T}_g

that corresponds by the construction in Section 2.2 to a Jenkins–Strebel quadratic differential on the Riemann surface X at the starting point of the ray. A Jenkins–Strebel differential decomposes X into cylinders swept out by horizontal trajectories. Mainly following [26] we give in Section 4.1 two explicit descriptions of the marked Riemann surfaces (X_K, f_K) (for $K > 1$) on a Jenkins–Strebel ray. This allows us to identify the boundary point (X_∞, f_∞) at the “end” of the ray as the stable Riemann surface that is obtained by contracting on X the core lines of the cylinders in a prescribed way, see Sections 4.1.5 and 4.1.6.

In the case that the Teichmüller embedding ι leads to a Teichmüller curve C_ι we show in Section 4.2 that all boundary points of Δ_ι are obtained in this way. This shows in particular that all cusps of Teichmüller curves are obtained by contracting, on a corresponding Riemann surface, the center lines of the cylinders of a Jenkins–Strebel differential. For the proof of this result we show that the Teichmüller embedding ι can be extended to a continuous embedding $\bar{\iota}: \mathbb{H} \cup \{\text{cusps of } \bar{\Gamma}_\iota^*\} \hookrightarrow \bar{\mathcal{T}}_g$, see Proposition 4.13. Moreover, if the Veech group $\bar{\Gamma}_\iota$ is a lattice in $\text{PSL}_2(\mathbb{R})$, the image of $\bar{\iota}$ is the closure of Δ_ι in $\bar{\mathcal{T}}_g$, see Proposition 4.14.

Since $\bar{\mathcal{T}}_g$ has these cusp singularities at the boundary that prevent it from being an ordinary complex space, whereas the boundary of \bar{M}_g is a nice divisor in a projective variety, it is interesting to look at spaces that lie properly between Teichmüller and moduli space and to ask for a boundary that fits somehow in between $\bar{\mathcal{T}}_g$ and \bar{M}_g . An example of such a space is provided by the *Schottky space* which goes back to the paper [32] of F. Schottky from 1887. He studied discontinuous groups that are freely generated by Möbius transformations $\gamma_1, \dots, \gamma_g$ (for some $g \geq 1$) chosen in such a way that there are disjoint closed Jordan domains $D_1, D'_1, \dots, D_g, D'_g$ such that γ_i maps D_i onto the complement of the interior of D'_i . The Riemann surface of such a *Schottky group* is compact of genus g . It can be shown that every compact Riemann surface X admits such a *Schottky uniformization* $X = \Omega / \Gamma$ (with $\Omega \subset \mathbb{P}^1(\mathbb{C})$ open and Γ a Schottky group), see Section 5.1. The covering $\Omega \rightarrow X$ is called a *Schottky covering*. It is minimal for the property that Ω is planar, i.e. biholomorphic to a subdomain of $\mathbb{P}^1(\mathbb{C})$; here minimality means that each unramified holomorphic covering $Y \rightarrow X$ with a planar manifold Y factors through Ω .

Schottky coverings are classified by a complex manifold S_g of dimension $3g - 3$, called the *Schottky space*. The natural map from \mathcal{T}_g to M_g factors through S_g , therefore there is a subgroup $\Gamma_g(\alpha)$ of the mapping class group Γ_g such that $\mathcal{T}_g / \Gamma_g(\alpha) = S_g$. Unfortunately the subgroup $\Gamma_g(\alpha)$ is not normal and depends on the choice of a certain group homomorphism α . As a consequence the induced map $S_g \rightarrow M_g$ is not the quotient for a group action. We review this classical but not so widely known material in Sections 5.1 and 5.2.

The concept of Schottky coverings can be extended to stable Riemann surfaces. If the analogous construction as for ordinary Riemann surfaces is applied to a surface X with nodes, we obtain a covering space Ω which is not planar, but on which nevertheless a free group Γ acts by holomorphic automorphisms with quotient space $\Omega / \Gamma = X$.

Although the groups Γ are no longer subgroups of $\mathrm{PSL}_2(\mathbb{C})$, it is possible to find parameters for them in almost the same way as for Schottky groups, namely by cross ratios of fixed points. It then turns out that these generalized Schottky coverings are classified by a complex manifold \bar{S}_g (which contains S_g as an open dense subset), see Section 5.3. This result was originally proved in [13]; here we show that it can easily be derived from Braungardt's characterization of $\tilde{\mathcal{T}}_g$ as the universal covering of \bar{M}_g with cusps over ∂M_g , see Section 5.4.

Finally we wonder what the image in S_g of a Teichmüller disk Δ_t in $\tilde{\mathcal{T}}_g$ might look like. In the general case we have no idea. Of course, the image may depend on the choice of the subgroup $\Gamma_g(\alpha)$ that gives the map $\tilde{\mathcal{T}}_g \rightarrow S_g$. In the special situation that C_t is a Teichmüller curve we prove that for suitable choice of α , the image of Δ_t in S_g is not a disk, see Proposition 5.21.

Acknowledgments. We would like to thank Volker Braungardt for allowing us to include his results on $\tilde{\mathcal{T}}_g$, and for his helpful comments on an earlier version of Section 3. We are also grateful to Pierre Lochak and Martin Möller for many valuable conversations on Teichmüller disks, Teichmüller curves, and their boundaries. Furthermore, we would like to thank Bill Abikoff for his useful suggestions that helped to improve the exposition considerably.

2 Geodesic rays, Teichmüller disks and Teichmüller curves

The aim of this section is to introduce Teichmüller disks and Teichmüller curves. We start by recalling in 2.1 the concept of Teichmüller deformations and using them we give a definition for the Teichmüller space $\tilde{\mathcal{T}}_g$ alternative to the one we gave in the introduction. This will help us to define geodesic rays in the Teichmüller space in 2.2. In 2.3 we introduce Teichmüller disks as complex version of geodesic rays giving different alternative definitions. Finally in 2.4 we introduce the Veech group and Teichmüller curves and summarize some facts about the interrelation between these objects.

2.1 Teichmüller deformations

As one of numerous possibilities, one can define the Teichmüller space as the space of Teichmüller deformations. We briefly recall this concept here. It is described e.g. in [2, Chapter I, §3].

At the end of this subsection we extend it to the corresponding concept for punctured Riemann surfaces and their Teichmüller space $\tilde{\mathcal{T}}_{g,n}$, cf. [2, Chapter II, §1].

Let $X = X_{\mathrm{ref}}$ be a fixed Riemann surface of genus $g \geq 2$ and q be a holomorphic quadratic differential on X . We refer to the zeros of q as *critical points*, all other points are *regular points*. Then on the surface

$$X^* = X - \{P \in X : P \text{ is a critical point of } q\}$$

the differential q naturally defines a *flat structure* μ , i.e. an atlas such that all transition maps are of the form $z \mapsto \pm z + c$, with some constant $c \in \mathbb{C}$. The charts of μ in regular points z_0 are given as

$$z \mapsto \int_{z_0}^z \sqrt{q(\xi)} d\xi. \tag{2.1}$$

One may deform this flat structure by composing each chart with the map

$$x + iy \mapsto Kx + iy = \frac{1}{2}(K + 1)z + \frac{1}{2}(K - 1)\bar{z} \quad (x, y \in \mathbb{R}) \tag{2.2}$$

with K an arbitrary real number > 1 . This defines a new flat structure on X^* which can uniquely be extended to a holomorphic structure on X .

We call X_K the Riemann surface that we obtain this way, $X_1 = X$ the surface with the original complex structure and $f_K: X_1 \rightarrow X_K$ the map that is topologically the identity. The map f_K is a Teichmüller map and has constant complex dilatation

$$k(z) = \frac{(f_K)_{\bar{z}}}{(f_K)_z} = \frac{K - 1}{K + 1}.$$

Its maximal dilatation $\sup_{z \in X} \frac{1+|k(z)|}{1-|k(z)|}$ (as a quasiconformal map) is equal to K .

Definition 2.1. Let q be a holomorphic quadratic differential on X and $K \in \mathbb{R}_{>1}$. The pair (X_K, f_K) as defined above is called the *Teichmüller deformation* of X of constant dilatation K with respect to q .

The pair (X_K, f_K) defines a point in the Teichmüller space \mathcal{T}_g which for simplicity we also denote as (X_K, f_K) . Since the constant dilatation of f_K is equal to K , the Teichmüller distance between the points (X_1, id) and (X_K, f_K) of \mathcal{T}_g is $\log(K)$.

If two holomorphic quadratic differentials on X are positive scalar multiples of each other, they define, for each K , the same point in \mathcal{T}_g . Thus one restricts to differentials with norm 1. By Teichmüller's existence and uniqueness theorems, see e.g. [2, Chapter I, (3.5), (4.2)], one can show that each point in \mathcal{T}_g is uniquely obtained as a Teichmüller deformation. If Q_X is the vector space of all holomorphic quadratic differentials on X and if Σ_X is the unit sphere in Q_X , one may thus write

$$\{(X, q, k) : q \in \Sigma_X, k \in (0, 1)\} \cup \{0\} = \mathcal{T}_g, \tag{2.3}$$

and the identification of the two sets is done by the map

$$(X, q, k) \mapsto (X_K, f_K) \quad \text{with } K = \frac{1+k}{1-k} \Leftrightarrow k = \frac{K-1}{K+1} \quad \text{and} \tag{2.4}$$

$$0 \mapsto \text{the base point } (X, \text{id}).$$

(X_K, f_K) depends of course by its definition on the differential q . In the following we shall denote the base point also by $(X, q, 0)$.

The map (2.4) is a homeomorphism. Here on the left-hand side of (2.3) one takes the topology obtained by identifying it with the open unit ball in Q_X . It follows in particular, that \mathcal{T}_g is contractible.

Teichmüller deformations can be understood as *affine deformations* in the following sense: Let us here and in what follows identify \mathbb{C} with \mathbb{R}^2 by the \mathbb{R} -linear map sending $(1, i)$ to the standard basis of \mathbb{R}^2 . Then the map in (2.2) is equal to the affine map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} K & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}.$$

Since composing charts with a biholomorphic map does not change the point in Teichmüller space, one obtains the same point (X_K, f_K) in \mathcal{T}_g if one composes each chart of the flat structure μ on X with the affine map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto D_K \cdot \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{with } D_K = \begin{pmatrix} \sqrt{K} & 0 \\ 0 & \frac{1}{\sqrt{K}} \end{pmatrix} \in \text{SL}_2(\mathbb{R}). \quad (2.5)$$

We shall use the following notations which are compatible with those in Section 2.3.2 where we introduce the general concept of affine deformations.

Definition 2.2. Let X be a compact Riemann surface of genus g , q a holomorphic quadratic differential, μ the flat structure defined by q . We call the flat structure defined by (2.5) μ_{D_K} and denote $(X, \mu) \circ D_K = (X, \mu_{D_K})$.

Note that (X, μ_{D_K}) is as Riemann surface isomorphic to X_K . Thus the point $[(X, \mu_{D_K}), \text{id}]$ in \mathcal{T}_g defined by the marking $\text{id}: X \rightarrow (X, \mu_{D_K})$ is equal to (X_K, f_K) .

Finally, let us turn to *Teichmüller deformations of punctured Riemann surfaces*: The definition is done almost in the same way as in the case without punctures, see [2, Chapter II, §1]. Suppose that g and n are natural numbers with $3g - 3 + n > 0$. Let X be a Riemann surface of genus g with n marked points P_1, \dots, P_n , and $X_{\text{ref}} = X_0 = X - \{P_1, \dots, P_n\}$.

In this case, one uses *admissible holomorphic quadratic differentials* on X_0 . They are by definition those meromorphic quadratic differentials on X that restrict to a holomorphic quadratic differential on X_0 and have at each puncture either a simple pole or extend holomorphically across the puncture, see [2, Chapter II, (1.4)]. The vector space of these differentials is called Q_{X_0} . For $q \in Q_{X_0}$ we define the *critical points* to be the marked points and all zeroes of q ; the remaining points are called *regular*. Now, the definition of Teichmüller deformation is done exactly as before, just always replacing Q_X by Q_{X_0} . One obtains in the same way:

$$\{(X, q, k) : q \in \Sigma_{X_0}, k \in (0, 1)\} \cup \{0\} = \mathcal{T}_{g,n}. \quad (2.6)$$

Here Σ_{X_0} is the unit ball in Q_{X_0} .

2.2 Geodesic rays

Let $X = X_{\text{ref}}$ be a Riemann surface of genus g . A holomorphic quadratic differential q on X naturally defines a geodesic embedding of $\mathbb{R}_{\geq 0}$ into \mathcal{T}_g with respect to the Teichmüller metric on \mathcal{T}_g as is described in the following.

Definition 2.3. Let q be a holomorphic quadratic differential on X and let γ be the map

$$\gamma = \gamma_q: \begin{cases} [0, \infty) \rightarrow \mathcal{T}_g, \\ t \mapsto (X_K, f_K) = (X, \mu_{D_K}) = (X, q, k) \\ \text{with } K = e^t, k = \frac{K-1}{K+1}. \end{cases} \quad (2.7)$$

The image of γ is called the *geodesic ray in \mathcal{T}_g in direction of q (or with respect to) q starting at (X, id) .*

Here we use the notation of the last section:

$$(X_K, f_K) \stackrel{\text{Def. 2.2}}{=} (X, \mu_{D_K}) \stackrel{(2.4)}{=} (X, q, k)$$

is the point in \mathcal{T}_g defined by the Teichmüller deformation of X of dilatation K with respect to q . Recall from the last section that the distance between the two points (X_K, f_K) and (X, id) in \mathcal{T}_g is $\log(K)$. Thus γ is an isometric embedding.

In fact, from the description of \mathcal{T}_g given in (2.3) one observes that all points in \mathcal{T}_g which have distance $\log(K)$ to the base point (X, id) are Teichmüller deformations of X of constant dilatation K with respect to a holomorphic quadratic differential. It follows that each isometric embedding of $[0, \infty)$ into \mathcal{T}_g is of the form (2.7).

2.3 Teichmüller disks

In this section we define Teichmüller disks. They can be found defined under this name e.g. in [29, p. 149/150] and [12, 8.1–8.2]. One may find comprehensive overviews e.g. in [34] and [11], or more recently [27] and [23], to pick only a few of numerous references where they occur. We introduce them here in detail comparing three different ways how to construct them. For completeness we have included most of the proofs.

Definition 2.4. Let $3g - 3 + n > 0$. A *Teichmüller disk* Δ_t is the image of a holomorphic isometric embedding

$$\iota: \mathbb{D} \hookrightarrow \mathcal{T}_{g,n}$$

of the complex unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ into the Teichmüller space. Here we take the Poincaré metric of constant curvature -1 on \mathbb{D} and the Teichmüller metric on $\mathcal{T}_{g,n}$. The embedding ι is also called *Teichmüller embedding*.

Instead of the unit disk \mathbb{D} one may take as well the upper half plane \mathbb{H} with the hyperbolic metric. We shall switch between these two models using the holomorphic isometry

$$f: \mathbb{H} \rightarrow \mathbb{D}, \quad t \mapsto \frac{i-t}{i+t}. \quad (2.8)$$

Thus Teichmüller disks are obtained equivalently as images of holomorphic isometric embeddings $\mathbb{H} \hookrightarrow \mathcal{T}_{g,n}$ of the upper half plane \mathbb{H} into the Teichmüller space $\mathcal{T}_{g,n}$.

How does one find such embeddings? Similarly as for geodesic rays, each holomorphic quadratic differential q on a Riemann surface X defines a Teichmüller disk. In the following we describe three alternative constructions starting from such a differential q that all lead to the same Teichmüller disk Δ_q . For simplicity we only consider the case $n = 0$ and $g \geq 2$. However the same constructions can be done in the general case of punctured surfaces.

2.3.1 Teichmüller disks as a collection of geodesic rays

Definition 2.5. Let q be a holomorphic quadratic differential on a Riemann surface X of genus g . Let ι_1 be the map

$$\iota_1: \begin{cases} \mathbb{D} \rightarrow \mathcal{T}_g, \\ z = r \cdot e^{i\varphi} \mapsto (X, e^{-i\varphi} \cdot q, r). \end{cases}$$

Here we use the definition of \mathcal{T}_g given by (2.3). Hence, $(X, e^{-i\varphi} \cdot q, r)$ is the point defined by the Teichmüller deformation of X of dilatation $K = \frac{1+r}{1-r}$ with respect to $q_{-\varphi} = e^{-i\varphi} \cdot q$.

We shall show in Proposition 2.8 that ι_1 is an isometric holomorphic embedding, thus the image Δ_{ι_1} of ι_1 is a Teichmüller disk.

The map ι_1 may be considered as a collection of geodesic rays in the following sense: Let τ_φ be the geodesic ray in \mathbb{D} starting from 0 in direction φ , i.e.:

$$\tau_\varphi: \begin{cases} [0, \infty) \rightarrow \mathbb{D}, \\ t \mapsto r(t) \cdot e^{i\varphi} \quad \text{with } r(t) = \frac{e^t - 1}{e^t + 1}. \end{cases}$$

Then $\iota_1 \circ \tau_\varphi: [0, \infty) \rightarrow \mathcal{T}_g$ is equal to the map given in (2.7) that defines the geodesic ray to the holomorphic quadratic differential $q_{-\varphi} = e^{-i\varphi} \cdot q$ on X .

Thus the Teichmüller disk Δ_{ι_1} is the union of all geodesic rays defined by the differentials $e^{i\varphi} \cdot q$ with $\varphi \in [0, 2\pi)$. Furthermore, $\iota_1 \circ \tau_\varphi$ is the parameterization by length of the restriction $\iota_1|_{R_\varphi}$ of ι_1 to the ray $R_\varphi = \{r \cdot e^{i\varphi} : r \in [0, 1)\}$.

2.3.2 Teichmüller disks by affine deformations. We now describe a second approach that starting from a holomorphic quadratic differential q leads to the same Teichmüller disk as in 2.3.1.

Recall from Section 2.1 that a holomorphic quadratic differential q defines on $X^* = X - \{\text{zeroes of } q\}$ a flat structure μ . The group $\text{SL}_2(\mathbb{R})$ acts on the flat structures

of X^* (as topological surface) in the following way: Let $B \in \mathrm{SL}_2(\mathbb{R})$ and μ be a flat structure on X^* . Composing each chart of μ with the affine map $z \mapsto B \cdot z$ gives a new flat structure on X^* which we denote $B \circ (X, \mu)$ or (X, μ_B) . In the special case $B = D_K$ we obtain the Teichmüller deformation of dilatation K , cf. Definition 2.2.

Definition 2.6. We call $(X, \mu_B) = B \circ (X, \mu)$ *affine deformation* of (X, μ) by the matrix B .

Note that for B_1, B_2 in $\mathrm{SL}_2(\mathbb{R})$ one may write

$$B_1 \circ (B_2 \circ (X, \mu)) = B_1 \circ (X, \mu_{B_2}) = (X, \mu_{B_1 B_2}) = B_1 \cdot B_2 \circ (X, \mu).$$

The flat structure μ_B defines in particular a complex structure on X . We identify here the complex plane \mathbb{C} with \mathbb{R}^2 as we already did in Section 2.1. In general the new complex structure will be different from the one defined by μ . Taking the identity $\mathrm{id}: (X, \mu) \rightarrow (X, \mu_B)$ on X as marking, we obtain a point $P_B = [(X, \mu_B), \mathrm{id}]$ in the Teichmüller space \mathcal{T}_g . By abuse of notation we shall sometimes denote this point also just as (X, μ_B) .

Thus one obtains the map

$$\hat{\iota}_2: \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathcal{T}_g, \quad B \mapsto P_B = [(X, \mu_B), \mathrm{id}] = (X, \mu_B).$$

If however the matrix $A = U$ is in $\mathrm{SO}_2(\mathbb{R})$ the map $\mathrm{id}: (X, \mu_B) \rightarrow (X, \mu_{U \cdot B})$ is holomorphic, thus the point in Teichmüller space is not changed, i.e.

$$U \in \mathrm{SO}_2(\mathbb{R}) \Rightarrow P_{UA} = P_A \text{ for all } A \in \mathrm{SL}_2(\mathbb{R}). \quad (2.9)$$

Hence $\hat{\iota}_2$ induces a map

$$\iota_2: \mathrm{SO}_2(\mathbb{R}) \backslash \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathcal{T}_g, \quad \mathrm{SO}_2(\mathbb{R}) \cdot B \mapsto P_B = [(X, \mu_B), \mathrm{id}] = (X, \mu_B).$$

Please note: The action of $\mathrm{SL}_2(\mathbb{R})$ on the flat structures $\{(X, \mu_A) : A \in \mathrm{SL}_2(\mathbb{R})\}$ does not descend to the image set $\{P_A : A \in \mathrm{SL}_2(\mathbb{R})\}$ in \mathcal{T}_g ; in particular $P_U = P_I \not\cong P_{AU} = P_A!$

The Teichmüller disk. One may identify $\mathrm{SO}_2(\mathbb{R}) \backslash \mathrm{SL}_2(\mathbb{R})$ with the upper half plane \mathbb{H} in the following way: Let $\mathrm{SL}_2(\mathbb{R})$ act by Möbius transformations on the upper half plane \mathbb{H} . This action is transitive and $\mathrm{SO}_2(\mathbb{R})$ is the stabilizing group of i . Thus the map

$$p: \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathbb{H}, \quad A \mapsto \overline{-A^{-1}(i)} \quad (2.10)$$

induces a bijection $\mathrm{SO}_2(\mathbb{R}) \backslash \mathrm{SL}_2(\mathbb{R}) \rightarrow \mathbb{H}$. Its inverse map is induced by

$$\mathbb{H} \rightarrow \mathrm{SL}_2(\mathbb{R}), \quad t \mapsto \frac{1}{\sqrt{\mathrm{Im}(t)}} \begin{pmatrix} 1 & \mathrm{Re}(t) \\ 0 & \mathrm{Im}(t) \end{pmatrix}.$$

Composing ι_2 from above with this bijection one obtains a map from \mathbb{H} to \mathcal{T}_g which we also denote by ι_2 .

Definition 2.7. Let q be a holomorphic quadratic differential on the Riemann surface X and μ the flat structure defined by q . Let ι_2 be the map

$$\iota_2: \mathbb{H} \rightarrow \mathcal{T}_g, \quad t \mapsto P_{A_t} = [(X, \mu_{A_t}), \text{id}]$$

with A_t chosen such that $-\overline{A_t^{-1}(i)} = t$.

Note that the identification of $\text{SO}_2(\mathbb{R}) \backslash \text{SL}_2(\mathbb{R})$ with \mathbb{H} given by p may seem a bit ponderous, but one has to compose $A \mapsto A^{-1}(i)$ with the reflection at the imaginary axis in order that ι_2 becomes holomorphic. We shall see this later in 2.3.3. In fact one has much more, as is stated in the following proposition.

Proposition 2.8. *The maps ι_1 and ι_2 are Teichmüller embeddings. They define the same Teichmüller disk*

$$\Delta_q = \Delta_{\iota_1} = \iota_1(\mathbb{D}) = \Delta_{\iota_2} = \iota_2(\mathbb{H}). \tag{2.11}$$

Proof. The proof is given in the rest of Subsection 2.3.2 and in 2.3.3:

In Proposition 2.12 we show that $\iota_2 = \iota_1 \circ f$ with f from (2.8) (see also Figure 1); thus it is sufficient to show only for one of them that it is isometric, and in the same manner for being holomorphic.

In Proposition 2.11 it is shown that ι_2 is isometric. In Subsection 2.3.3, it is shown that ι_1 is holomorphic (see Corollary 2.15). For this purpose we introduce an embedding $\iota_3: \mathbb{D} \rightarrow \mathcal{T}_g$, using Beltrami differentials, for which it is not difficult to see that it is holomorphic, and show that it is equal to ι_1 .

That ι_1 and ι_2 define the same Teichmüller disks then also follows from Proposition 2.12. □

In fact the described constructions do not only give some special examples but all Teichmüller disks are obtained like this: Each Teichmüller disk is equal to Δ_q as in (2.11) for some holomorphic quadratic differential q . And all Teichmüller embeddings are of the form $\iota_1: \mathbb{D} \hookrightarrow \mathcal{T}_g$ or equivalently $\iota_2: \mathbb{H} \hookrightarrow \mathcal{T}_g$, see [12, 7.4].

In order to see that ι_2 from Definition 2.7 is isometric we first calculate the Teichmüller distance between two affine deformations.

Teichmüller distance between two affine deformations. In what follows we constantly use the following fact about matrices in $\text{SL}_2(\mathbb{R})$:

Remark 2.9. Each matrix $A \in \text{SL}_2(\mathbb{R})$ with $A \notin \text{SO}_2(\mathbb{R})$ can be decomposed uniquely up to the minus signs as follows:

$$A = U_1 \cdot D_K \cdot U_2 \quad \text{with } U_1, U_2 \in \text{SO}_2(\mathbb{R}), \quad D_K = \begin{pmatrix} \sqrt{K} & 0 \\ 0 & \frac{1}{\sqrt{K}} \end{pmatrix}, \quad K > 1. \tag{2.12}$$

We may denote: $U_2 = U_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$.

This fact can e.g. be seen geometrically as follows: $SL_2(\mathbb{R})$ acts transitively on the upper half plane \mathbb{H} by Möbius transformations. The point $i \in \mathbb{H}$ can be mapped to $A(i) \neq i$ by first doing a stretching along the imaginary axis in direction ∞ and afterwards a rotation around i , i.e. $A(i) = U_1(D_K(i))$ with suitably chosen $U_1 \in SO_2(\mathbb{R})$ and D_K with $K > 1$ as in the remark. Since the stabilizer of i in $SL_2(\mathbb{R})$ is $SO_2(\mathbb{R})$, one has $A = U_1 \cdot D_K \cdot U_2$ with U_2 also in $SO_2(\mathbb{R})$. A short calculation gives the uniqueness claim.

In the following proposition let q again be a holomorphic quadratic differential on $X = X_{\text{ref}}$ and μ the flat structure that q defines.

Proposition 2.10. *Let A and B be in $SL_2(\mathbb{R})$ with $A \cdot B^{-1} \notin SO_2(\mathbb{R})$ and*

$$A \cdot B^{-1} = U_1 \cdot D_K \cdot U_2$$

with U_1, U_2 and D_K as in (2.12). Then the Teichmüller distance between the two points $P_A = [(X, \mu_A), \text{id}]$ and $P_B = [(X, \mu_B), \text{id}]$ in \mathcal{T}_g is $\log(K)$.

Proof. We proceed in three steps.

a) Suppose B is the identity matrix I and

$$A = D_K = \begin{pmatrix} \sqrt{K} & 0 \\ 0 & \frac{1}{\sqrt{K}} \end{pmatrix} \quad \text{for some } K \in \mathbb{R}_{>1}.$$

Thus we have in fact that $P_A = [(X, \mu_{D_K}), \text{id}]$ is the point in \mathcal{T}_g defined by the Teichmüller deformation of dilatation K with respect to q , see Definition 2.2. Hence the distance between P_A and the base point $(X_{\text{ref}}, \text{id}) = P_I$ is $\log(K)$.

b) Suppose again that $B = I$, but A is an arbitrary matrix in $SL_2(\mathbb{R})$.

Thus $A = U_1 \cdot D_K \cdot U_2$ and the map $\text{id}: (X, \mu) \rightarrow (X, \mu_A)$ is the composition of three maps:

$$(X, \mu) \xrightarrow{\text{id}} (X, \mu_{U_2}) \xrightarrow{\text{id}} (X, \mu_{D_K U_2}) \xrightarrow{\text{id}} (X, \mu_{U_1 D_K U_2}).$$

Since the first and the third map are biholomorphic the Teichmüller distance is again $\log(K)$.

More precisely, write $U_2 = U_\theta$ as in (2.12). Then μ_{U_2} is the flat structure obtained by composing each chart with $z \mapsto e^{i\theta} \cdot z$. This is equal to the flat structure defined by the quadratic differential $q_{2\theta} = (e^{i\theta})^2 \cdot q$ which is holomorphic on the Riemann surface X .

Now, $\text{id}: (X, \mu_{U_2}) \rightarrow (X, \mu_{D_K U_2})$ is (up to the stretching $z \mapsto \sqrt{K} \cdot z$) the Teichmüller deformation of dilatation K with respect to the holomorphic quadratic differential $q_{2\theta}$. Thus the distance between $P_A = P_{U_1 D_K U_2} \stackrel{(2.9)}{=} P_{D_K U_2}$ and the base point $P_B = P_I$ is $\log(K)$.

c) Let now A, B be arbitrary in $SL_2(\mathbb{R})$. The Teichmüller metric does not depend on the chosen base point. Thus we may consider P_B as base point and P_A as coming from the affine deformation defined by the matrix $A \cdot B^{-1}$. Then with the given decomposition $A \cdot B^{-1} = U_1 \cdot D_K \cdot U_2$ the distance is as in b) equal to $\log(K)$. \square

Proposition 2.11. ι_2 is an isometric embedding

Proof. We denote by ρ the Poincaré distance in \mathbb{H} and by $d_{\mathcal{T}}$ the Teichmüller distance in \mathcal{T}_g . Let t_1 and t_2 be arbitrary distinct points in \mathbb{H} . We may write $t_1 = p(A)$ and $t_2 = p(B)$ with A, B in $SL_2(\mathbb{R})$, p as in (2.10). Let $AB^{-1} = U_1 D_K U_2$ the decomposition of AB^{-1} from (2.12). ($AB^{-1} \notin SO_2(\mathbb{R})$ because $t_1 \neq t_2$)

$$\begin{aligned} \rho(t_1, t_2) &= \rho(-\overline{B^{-1}(i)}, -\overline{A^{-1}(i)}) = \rho(B^{-1}(i), A^{-1}(i)) = \rho(AB^{-1}(i), i) \\ &= \rho(U_1 D_K U_2(i), i) = \rho(U_1 D_K(i), i) \stackrel{\star}{=} \rho(D_K(i), i) \\ &= \rho(Ki, i) = \log(K) \stackrel{\text{Prop. 2.10}}{=} d_{\mathcal{T}}(P_B, P_A) = d_{\mathcal{T}}(\iota_2(t_1), \iota_2(t_2)). \end{aligned}$$

The equality \star is given since U_1 is a hyperbolic rotation with center i and thus does not change the distance to i . □

Now we show that ι_1 and ι_2 are “almost” the same map.

Proposition 2.12. ι_1 and ι_2 fit together. More precisely: $\iota_1 \circ f = \iota_2$, with the isomorphism $f: \mathbb{H} \rightarrow \mathbb{D}$ from (2.8).

The following diagram may be helpful while reading the proof. Some parts will be explained only after the proof; in particular the space $B(X)$ of Beltrami differentials will be introduced in 2.3.3.

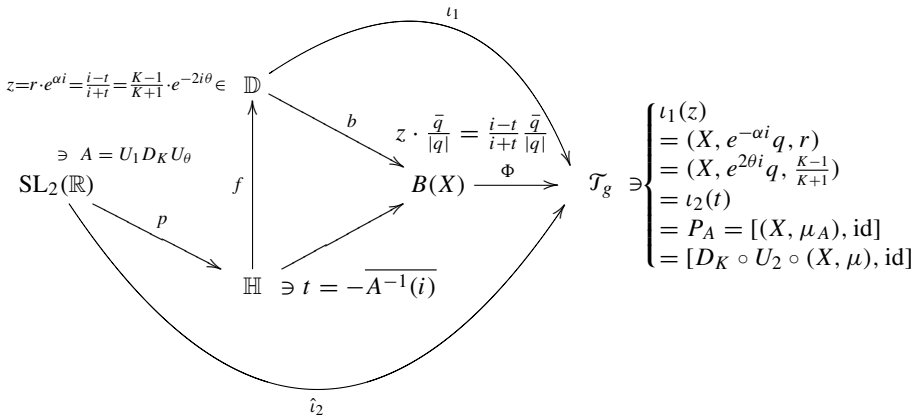


Figure 1. Diagram for alternative definitions of the Teichmüller disk Δ_g .

Proof. We proceed in two steps:

- (1) Let $A \in SL_2(\mathbb{R})$ be decomposed as in (2.12): $A = U_1 \cdot D_K \cdot U_2$, $U_2 = U_\theta$.

We show that $(f \circ p)(A) = r \cdot e^{-2i\theta}$ with $r = \frac{K-1}{K+1}$.

- (2) We show that $\iota_1(r \cdot e^{-2i\theta}) = \hat{\iota}_2(A)$.

Step 1. One may express $t := p(A)$ in terms of K and θ as follows:

$$\begin{aligned} t &= -\overline{A^{-1}(i)} = -U_2^{-1} D_K^{-1}(\bar{i}) = -U_2^{-1}\left(-\frac{i}{K}\right) = -\frac{\cos(\theta) \cdot \frac{-i}{K} + \sin(\theta)}{-\sin(\theta) \cdot \frac{-i}{K} + \cos(\theta)} \\ &= \frac{i \cos(\theta) - K \sin(\theta)}{i \sin(\theta) + K \cos(\theta)}. \end{aligned}$$

Now one has:

$$\begin{aligned} f(p(A)) = f(t) &= \frac{-t + i}{t + i} = \frac{-i \cos(\theta) + K \sin(\theta) + i(i \sin(\theta) + K \cos(\theta))}{i \cos(\theta) - K \sin(\theta) + i(i \sin(\theta) + K \cos(\theta))} \\ &= \frac{(K - 1)[\sin(\theta) + i \cos(\theta)]}{(K + 1)[- \sin(\theta) + i \cos(\theta)]} \\ &= \frac{K - 1}{K + 1} \cdot \frac{-(\sin(\theta) + i \cos(\theta))^2}{(\sin(\theta) - i \cos(\theta))(\sin(\theta) + i \cos(\theta))} \\ &= \frac{K - 1}{K + 1} (\cos(\theta) - i \sin(\theta))^2 = \frac{K - 1}{K + 1} \cdot e^{-2i\theta}. \end{aligned}$$

Step 2. $\iota_1(r \cdot e^{-2i\theta}) = (X, e^{2i\theta} \cdot q, r) \in \mathcal{T}_g$ is the point in the Teichmüller space that is obtained as Teichmüller deformation of dilatation $\frac{1+r}{1-r} = K$ with respect to the quadratic differential $e^{2i\theta} \cdot q$. Recall from the proof of Proposition 2.10 that this is precisely the point in \mathcal{T}_g defined by the affine deformation $D_K \circ U_\theta \circ (X, \mu) = (X, \mu_{D_K U_\theta}) = (X, \mu_{D_K U_\theta})$. Thus

$$(X, e^{2i\theta} \cdot q, r) = P_{D_K U_\theta} = P_{D_K U_2} \stackrel{(2.9)}{=} P_{U_1 D_K U_2} = P_A = \hat{\iota}_2(A). \quad (2.13)$$

□

Using (2.13) one may also describe the geodesic rays $\iota_1 \circ \tau_\varphi$ from 2.3.1 in the Teichmüller disk $\Delta_q = \Delta_{\iota_1} = \Delta_{\iota_2}$ as follows.

Corollary 2.13. *Define D_K, U_θ as in (2.12). The map*

$$[0, \infty) \rightarrow \mathcal{T}_g, \quad t \mapsto P_{D_K U_\theta} = [D_K \circ (X, \mu_{U_\theta}), \text{id}] \text{ with } K = e^t$$

is equal to $\iota_1 \circ \tau_{-2\theta}$.

It is thus by 2.3.1 the geodesic ray in direction of the quadratic differential $q_{2\theta} = e^{2\theta i} q$.

Proof. One has $t \xrightarrow{\tau_{-2\theta}} r(t) e^{-2\theta i} \xrightarrow{\iota_1} (X, e^{2\theta i} \cdot q, r(t)) \stackrel{(2.13)}{=} P_{D_K U_\theta}$. □

Hence, geometrically one obtains the geodesic ray to q_φ by rotating the flat structure by $U_{\frac{\varphi}{2}}$ and then stretching in vertical direction with dilatation K .

2.3.3 Beltrami differentials. In order to see that ι_1 and ι_2 are holomorphic we introduce an alternative way to define ι_1 using Beltrami differentials. We keep this aspect short and refer to e.g. [29] for more details.

Let

$$M(X) = \{(X_1, f) : X_1 \text{ Riemann surface, } f : X \rightarrow X_1 \text{ is a quasiconformal homeomorphism}\} / \approx$$

with $(X_1, f_1) \approx (X_2, f_2) \Leftrightarrow f_2 \circ f_1^{-1}$ is biholomorphic.

One has a natural projection $M(X) \rightarrow \mathcal{T}_g$. Furthermore $M(X)$ can be canonically identified with the open unit ball $B(X)$ in the Banach space $L^\infty_{(-1,1)}(X)$ of $(-1, 1)$ -forms by the bijection:

$$M(X) \rightarrow B(X), \quad (X_1, f) \mapsto \mu_f,$$

where μ_f is the Beltrami differential (or complex dilatation) of f , cf. [29, 2.1.4].

Thus one obtains a projection $\Phi : B(X) \rightarrow \mathcal{T}_g$. The map Φ is holomorphic ([29, 3.1]). Furthermore, for each quadratic differential q and for all $k \in (0, 1)$ the form $k \frac{\bar{q}}{|q|}$ is in $B(X)$ ([29, 2.6.3]) Thus one may define the map

$$\iota_3 : \begin{cases} \mathbb{D} \xrightarrow{b} B(X) \xrightarrow{\Phi} \mathcal{T}_g, \\ z \mapsto z \cdot \frac{\bar{q}}{|q|} \mapsto \Phi(z \cdot \frac{\bar{q}}{|q|}). \end{cases}$$

It is composition of two holomorphic maps and thus itself holomorphic.

We shall show in the following remark that $\iota_3 = \Phi \circ b = \iota_1$, cf. Figure 1.

Remark 2.14. For all $z_0 \in \mathbb{D} : \iota_3(z_0) = \iota_1(z_0)$.

Proof. Let $z_0 = r \cdot e^{i\alpha} \in \mathbb{D}$ and $A \in \text{SL}_2(\mathbb{R})$ with $f(p(A)) = z_0$.

Decompose $A = U_1 D_K U_2$ as in (2.12) with $U_2 = U_\theta$. Then by Step 1 of the proof of Proposition 2.12, $r = \frac{K-1}{K+1}$ and $\alpha = -2\theta$. Furthermore, by Proposition 2.12

$$\iota_1(z_0) = \hat{\iota}_2(A) = [(X, \mu_A), \text{id}] = [(X, \mu_{D_K U_2}), \text{id}].$$

Let us calculate the Beltrami differential of the Teichmüller deformation $f = \text{id} : X \rightarrow (X, \mu_{D_K U_2})$. We shall see that it is equal to $z_0 \cdot \frac{\bar{q}}{|q|}$. From this it follows that $\iota_1(z_0) = \iota_3(z_0)$.

One has $f = g \circ h$ with $h = \text{id} : X \rightarrow (X, \mu_{U_2})$ and $g = \text{id} : (X, \mu_{U_2}) \rightarrow (X, \mu_{D_K U_2})$. Locally in the charts of the flat structure defined by q , the maps g and h are given by

$$g : z \mapsto K \cdot \text{Re}(z) + i \cdot \text{Im}(z) \quad \text{and} \quad h : z \mapsto e^{i\theta} \cdot z.$$

Thus in terms of these charts one has:

$$\begin{aligned} f_z &= g_z \cdot h_z + g_{\bar{z}} \cdot \bar{h}_z = e^{i\theta} \cdot g_z, & f_{\bar{z}} &= g_z \cdot h_{\bar{z}} + g_{\bar{z}} \cdot \bar{h}_{\bar{z}} = e^{-i\theta} \cdot g_{\bar{z}} \\ \Rightarrow \frac{f_{\bar{z}}}{f_z} &= e^{-2i\theta} \cdot \frac{g_{\bar{z}}}{g_z} = e^{-2i\theta} \cdot \frac{K-1}{K+1} = e^{i\alpha} \cdot r = z_0. \end{aligned}$$

Hence the Beltrami differential of f is $z_0 \cdot \frac{\bar{q}}{|q|}$. □

One obtains immediately the following conclusion.

Corollary 2.15. $\iota_1 = \iota_3$ is holomorphic. By Proposition 2.12 ι_2 is also holomorphic.

2.4 Teichmüller curves

In this section we introduce Teichmüller curves and recall some properties of them, in particular their relation to Veech groups. This was explored by Veech in his article [34] and has been studied by many authors since then. Overviews and further properties can be found e.g. in [27], [11] or [20].

Let $\iota: \mathbb{D} \hookrightarrow \mathcal{T}_g$ be a Teichmüller embedding and $\Delta = \Delta_\iota = \iota(\mathbb{D})$ its image. We may consider the image of Δ_ι in the moduli space M_g under the natural projection $\mathcal{T}_g \rightarrow M_g$, cf. Section 1. In general it will be something with a large closure. But occasionally it is an algebraic curve. Such a curve is called Teichmüller curve.

Definition 2.16. If the image of the Teichmüller disk Δ in the moduli space M_g is an algebraic curve C , then C is called *Teichmüller curve*.

A surface (X, q) , with a Riemann surface X and a holomorphic quadratic differential q such that the Teichmüller disk $\Delta = \Delta_q$ defined by q projects to a Teichmüller curve is called *Veech surface*.

How can one decide whether a surface (X, q) induces a Teichmüller curve or not? An answer to this question is given by the Veech group, a subgroup of $\mathrm{SL}_2(\mathbb{R})$ associated to (X, q) . This is explained in the following two subsections.

2.4.1 Veech groups. Let X be a Riemann surface and q a holomorphic quadratic differential on X . Let μ be the flat structure on X defined by q . One obtains a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$ as follows: Let $\mathrm{Aff}^+(X, \mu)$ be the group of orientation preserving diffeomorphisms which are affine with respect to the flat structure μ , i.e. diffeomorphisms which are in terms of a local chart z of μ given by

$$z \mapsto A \cdot z + t, \quad \text{for some } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}), t \in \mathbb{C}.$$

As above we identify the complex plane \mathbb{C} with \mathbb{R}^2 . Furthermore, we denote for $z = x + iy$: $A \cdot z = ax + by + i(cx + dy)$.

Since μ is a flat structure, up to change of sign the matrix A does not depend on the charts. Thus one has a group homomorphism:

$$D: \mathrm{Aff}^+(X, \mu) \rightarrow \mathrm{PSL}_2(\mathbb{R}), \quad f \mapsto [A].$$

For simplicity we shall denote the image $[A]$ of the matrix A in $\mathrm{PSL}_2(\mathbb{R})$ often also just by A .

Definition 2.17. The image $\bar{\Gamma}(X, \mu) = D(\text{Aff}^+(X, \mu))$ of D is called the *projective Veech group* of (X, μ) .

We shall denote the projective Veech group also by $\bar{\Gamma}(X, q)$ and $\bar{\Gamma}_\iota$, where $\iota: \mathbb{D} \hookrightarrow \mathcal{T}_g$ or $\iota: \mathbb{H} \hookrightarrow \mathcal{T}_g$ is the Teichmüller embedding defined by q as described in 2.3. $\bar{\Gamma}(X, \mu)$ is a discrete subgroup of $\text{PSL}_2(\mathbb{R})$, see [34, Proposition 2.7].

2.4.2 The action of the Veech group on the Teichmüller disk. Recall that the projection $\mathcal{T}_g \rightarrow M_g$ from the Teichmüller space to the moduli space is given by the quotient for the action of the mapping class group

$$\Gamma_g = \text{Diffeo}^+(X) / \text{Diffeo}_0(X),$$

cf. (1.1) in the introduction. The action of $\text{Diffeo}^+(X)$ on \mathcal{T}_g is given by

$$\begin{aligned} \rho: \text{Diffeo}^+(X) &\rightarrow \text{Aut}(\mathcal{T}_g) \cong \Gamma_g, \quad \varphi \mapsto \rho_\varphi, \\ \text{with } \rho_\varphi: \mathcal{T}_g &\rightarrow \mathcal{T}_g, \quad (X_1, h) \mapsto (X_1, h \circ \varphi^{-1}). \end{aligned}$$

The affine group $\text{Aff}^+(X, \mu)$ acts as subgroup of $\text{Diffeo}^+(X)$ on \mathcal{T}_g . The following remark (cf. [11, Theorem 1]) determines this action restricted to the Teichmüller disk

$$\Delta = \Delta_q = \{P_B = [(X, \mu_B), \text{id}] \in \mathcal{T}_g : B \in \text{SL}_2(\mathbb{R})\}.$$

Remark 2.18. $\text{Aff}^+(X, \mu)$ stabilizes Δ . Its action on Δ is given as follows:

$$\begin{aligned} \varphi \in \text{Aff}^+(X, \mu), \quad B \in \text{SL}_2(\mathbb{R}) &\Rightarrow \rho_\varphi(P_B) = P_{BA^{-1}} \\ &\text{with } A \in \text{SL}_2(\mathbb{R}) \text{ a preimage of } D(\varphi) = [A]. \end{aligned}$$

Proof. Let $\varphi \in \text{Aff}^+(X)$, $B \in \text{SL}_2(\mathbb{R})$ and $A \in \text{SL}_2(\mathbb{R})$ be a preimage of $D(\varphi) = [A] \in \text{PSL}_2(\mathbb{R})$. In the following commutative diagram

$$\begin{array}{ccccc} (X, \mu) & \xrightarrow{\varphi^{-1}} & (X, \mu) & \xrightarrow{\text{id}} & (X, \mu_B) \\ & \searrow & & \searrow & \uparrow \\ & & & & (X, \mu_{BA^{-1}}) \end{array}$$

the map $(X, \mu_{BA^{-1}}) \rightarrow (X, \mu_B)$ is, as a composition of affine maps, itself affine. Its derivative is $D(\text{id} \circ \varphi^{-1} \circ \text{id}^{-1}) = BA^{-1}(BA^{-1})^{-1} = I$. Thus it is biholomorphic and $\rho_\varphi([(X, \mu_B), \text{id}]) = [(X, \mu_{BA^{-1}}), \text{id}]$. \square

It follows from Remark 2.18 that $\text{Aff}^+(X, \mu)$ is mapped by ρ to $\text{Stab}(\Delta)$, the global stabilizer of Δ in Γ_g . Furthermore $\rho: \text{Aff}^+(X, \mu) \rightarrow \text{Stab}(\Delta) \subseteq \Gamma_g$ is in fact an isomorphism: It is injective, see [11, Lemma 5.2] and surjective, see [11, Theorem 1]. Thus we have $\text{Aff}^+(X, \mu) \cong \text{Stab}(\Delta)$.

From Remark 2.18 it also becomes clear that the action of $\varphi \in \text{Aff}^+(X, \mu)$ depends only on $D(\varphi)$. Thus one obtains in fact an action of the projective Veech group $\overline{\Gamma}(X, \mu)$ on Δ .

Corollary 2.19. $\overline{\Gamma}(X, \mu) \subseteq \text{PSL}_2(\mathbb{R})$ acts on $\Delta = \{P_B \in \mathcal{T}_g : B \in \text{SL}_2(\mathbb{R})\}$ by

$$\rho_{[A]}(P_B) = P_{BA^{-1}} \tag{2.14}$$

where A is a preimage in $\text{SL}_2(\mathbb{R})$ of $[A]$.

Finally one may use the Teichmüller embedding $\iota_2: \mathbb{H} \rightarrow \mathcal{T}_g$ defined by q (cf. 2.7) in order to compare the action of $\overline{\Gamma}(X, \mu)$ on $\Delta = \Delta_\iota = \iota(\mathbb{H})$ with its action on \mathbb{H} via Möbius transformations. One obtains the diagram in the following remark (cf. [27, Proposition 3.2.]).

Remark 2.20. Let $A \in \text{PSL}_2(\mathbb{R})$. Denote by $A: \mathbb{H} \rightarrow \mathbb{H}$ its action as Möbius transformation on \mathbb{H} . The following diagram is commutative:

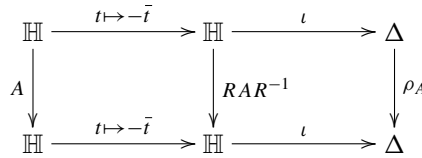
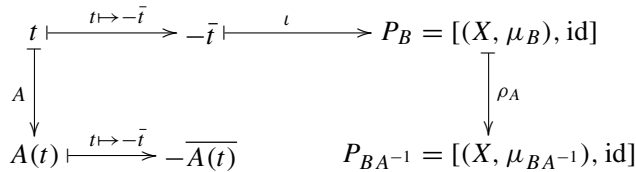


Figure 2

Here $R = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, thus R acts on $\mathbb{P}^1(\mathbb{C})$ by $z \mapsto -z$.

Proof. Let $t \in \mathbb{H}$. Choose some $B \in \text{SL}_2(\mathbb{R})$ with $-\overline{B^{-1}(i)} = -\bar{t}$, thus $\iota(-\bar{t}) = P_B = [(X, \mu_B), \text{id}]$ and using (2.14) we obtain the diagram:



The commutativity of the diagram in Figure 2 then follows from

$$RAR^{-1}(-\bar{t}) = -A(\bar{t}) = -\overline{A(t)}$$

and

$$-\overline{(BA^{-1})^{-1}(i)} = -A(\overline{B^{-1}(i)}) = -A(\bar{t}) = -\overline{A(t)}, \text{ thus } \iota(-\overline{A(t)}) = P_{BA^{-1}}. \quad \square$$

2.4.3 Veech groups and Teichmüller curves. In Remark 2.18 we saw that the affine group $\text{Aff}^+(X, \mu)$ maps isomorphically to the global stabilizer of the Teichmüller disk Δ in Γ_g . Denote by $\text{proj}: \mathcal{T}_g \rightarrow M_g$ the canonical projection. It then follows from Remark 2.20 that the map

$$\text{proj} \circ \iota: \mathbb{H} \rightarrow \text{proj}(\Delta) \subseteq M_g$$

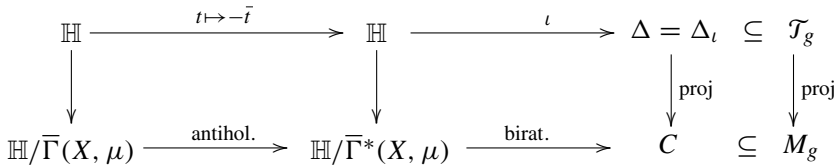
factors through $\mathbb{H}/R\bar{\Gamma}(X, \mu)R^{-1}$. We call

$$\bar{\Gamma}^*(X, \mu) = R\bar{\Gamma}(X, \mu)R^{-1}$$

the *mirror projective Veech group*, since $\mathbb{H}/\bar{\Gamma}^*(X, \mu)$ is a mirror image of $\mathbb{H}/\bar{\Gamma}(X, \mu)$, and refer to it also as $\bar{\Gamma}^*(X, q)$ or $\bar{\Gamma}_\iota^*$.

$\mathbb{H}/\bar{\Gamma}^*(X, \mu)$ is a surface of finite type and hence an algebraic curve if and only if $\bar{\Gamma}^*(X, \mu)$ is a lattice in $\text{PSL}_2(\mathbb{R})$. Altogether one obtains the following situation (cf. [27, Corollary 3.3]).

Corollary 2.21. *(X, q) induces a Teichmüller curve C if and only if $\bar{\Gamma}(X, \mu)$ is a lattice in $\text{PSL}_2(\mathbb{R})$. In this case the following diagram holds:*



In particular if $\bar{\Gamma}(X, \mu)$ is a lattice, then

- $\mathbb{H}/\bar{\Gamma}^*(X, \mu)$ is the normalization of the Teichmüller curve C ,
- $\mathbb{H}/\bar{\Gamma}(X, \mu)$ is antiholomorphic to $\mathbb{H}/\bar{\Gamma}^*(X, \mu)$.

3 Braungardt’s construction of $\bar{\mathcal{T}}_{g,n}$

Before we continue our study of Teichmüller disks and pass to the boundary, we want to explain the partial compactification $\bar{\mathcal{T}}_{g,n}$ of the Teichmüller space $\mathcal{T}_{g,n}$ that we shall use in the subsequent chapters. As mentioned in the introduction, $\bar{\mathcal{T}}_{g,n}$ will be a locally ringed space which, as a topological space, coincides with Abikoff’s augmented Teichmüller space $\hat{\mathcal{T}}_{g,n}$ (see the discussion following Proposition 3.9). The points of this space can be considered as marked stable Riemann surfaces (X, f) , where $f: X_{\text{ref}} \rightarrow X$ is a deformation map. The forgetful map $(X, f) \mapsto X$ defines a natural map from $\bar{\mathcal{T}}_{g,n}$ to the moduli space $\bar{M}_{g,n}$ of stable n -pointed Riemann surfaces of genus g . This map extends the projection $\mathcal{T}_{g,n} \rightarrow M_{g,n}$ and is in fact also the quotient map for the natural action of the mapping class group $\Gamma_{g,n}$. But the stabilizers of the

boundary points are infinite, and at the boundary the topology of $\overline{\mathcal{T}}_{g,n}$ is quite far from that of a manifold.

In his thesis [8], V. Braungardt gave a construction of $\overline{\mathcal{T}}_{g,n}$ which uses only the complex structure of $\overline{M}_{g,n}$ and the boundary divisor $\partial M_{g,n}$. Moreover his construction endows $\overline{\mathcal{T}}_{g,n}$ with the structure of a locally ringed space and he shows that it is a fine moduli space for “marked” stable Riemann surfaces. In this chapter we give a brief account of his approach.

3.1 Coverings with cusps

The basic idea of Braungardt’s construction is to study, for a complex manifold S , quotient maps $W \rightarrow W/G = S$ that have “cusps” over a divisor D in S . This concept, which will be explained in this section, generalizes the familiar ramified coverings. The key result is that, in the appropriate category of such quotient maps, there exists a universal object $p: \tilde{W} \rightarrow S$ with cusps over D .

In general \tilde{W} cannot be a complex manifold or even a complex space. Therefore we have to work in the larger category of *locally complex ringed spaces*, i.e. topological spaces W endowed with a sheaf \mathcal{O}_W of \mathbb{C} -algebras (called the *structure sheaf*) such that at each point $x \in W$ the stalk $\mathcal{O}_{W,x}$ is a local \mathbb{C} -algebra. The basic properties of such spaces can be found e.g. in [14, Chapter 1, § 1] (where they are called \mathbb{C} -ringed spaces).

In our situation Braungardt constructs a normal locally complex ringed space \tilde{W} such that the subspace $\tilde{W}_0 = \tilde{W} - p^{-1}(D)$ is a complex manifold and the restriction $p|_{\tilde{W}_0}: \tilde{W}_0 \rightarrow S_0 = S - D$ is the usual universal covering.

Example 3.1. The simplest example is well known and quite typical: Take S to be the unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ and $D = \{0\}$. The universal covering of $S - D$ is, of course, $\exp: \mathbb{H} \rightarrow \mathbb{D} - \{0\}$, $z \mapsto e^{2\pi iz}$. It turns out that the universal covering in Braungardt’s sense is $\hat{\mathbb{H}} = \mathbb{H} \cup i\infty$ with the *horocycle topology*, i.e. the sets $\mathbb{H}_R = \{z \in \mathbb{C} : \text{Im } z > R\} \cup i\infty$ for $R > 0$ form a basis of neighbourhoods of the point $i\infty$. Note that this topology is not the one induced from the Euclidean topology if $\mathbb{H} \cup \{\infty\}$ is considered as a subset of the Riemann sphere $\hat{\mathbb{C}}$.

$\hat{\mathbb{H}}$ is given the structure of a normal complex ringed space by taking $\mathcal{O}(U)$ to be the holomorphic functions on U for open subsets U of \mathbb{H} , and by defining $\mathcal{O}(\mathbb{H}_R)$ to be the set of holomorphic functions on $\{z \in \mathbb{C} : \text{Im } z > R\}$ that have a continuous extension to $i\infty$. Clearly $\mathcal{O}(\mathbb{H}_R)$ contains all functions of the form $z \mapsto e^{2\pi iz/n}$ for all $n \geq 1$.

We now give the precise definitions. We begin with the class of spaces that we need (cf. [8, 3.1.3]):

Definition 3.2. Let (W, \mathcal{O}_W) be a locally complex ringed space whose structure sheaf \mathcal{O}_W is a subsheaf of the sheaf $\mathcal{C}^\infty(W, \mathbb{C})$ of continuous complex valued functions on W .

a) A subset $B \subset W$ is called *analytic* if there is an open covering $(U_i)_{i \in I}$ of W and for each $i \in I$ there are finitely many elements $f_{i,1}, \dots, f_{i,n_i} \in \mathcal{O}_W(U_i)$ such that $B \cap U_i$ is the zero set of $\{f_{i,1}, \dots, f_{i,n_i}\}$.

b) We call (W, \mathcal{O}_W) an *R-space* if, for every open $U \subseteq W$ and every proper closed analytic subset $B \subset U$, a continuous function $f: U \rightarrow \mathbb{C}$ is in $\mathcal{O}_W(U)$ if and only if its restriction to $U - B$ is in $\mathcal{O}_W(U - B)$.

Note that all complex spaces are R-spaces: The required property is just Riemann's extension theorem, see [14, Chapter 7].

Definition 3.3. Let S be a complex manifold and $D \subset S$ a proper closed analytic subset. Then a surjective morphism $p: W \rightarrow S$ from an R-space (W, \mathcal{O}_W) to S is called a *covering with cusps over D* if there is a group G of automorphisms of W (as locally complex ringed space) such that

- (i) p is the quotient map $W \rightarrow W/G = S$,
- (ii) $W_0 = p^{-1}(S - D)$ is a complex manifold and $p|_{W_0}: W_0 \rightarrow S_0 = S - D$ is an unramified covering,
- (iii) for any $x \in W$ there is a basis of neighbourhoods U_x that are *precisely invariant* under the stabilizer G_x of x in G (i.e. $G_x(U_x) = U_x$ and $g(U_x) \cap U_x = \emptyset$ for each $g \in G - G_x$).

Note that, in particular, any ramified normal covering of complex manifolds is a covering in the sense of this definition (with cusps over the branch locus). As mentioned before, the basic result is (see [8, Satz 3.1.9])

Theorem 3.4. (i) *For any complex manifold S and any proper closed analytic subset $D \subset S$ there exists an initial object $p: (\tilde{W}, \mathcal{O}_{\tilde{W}}) \rightarrow S$ in the category of coverings of S with cusps over D ; it is called the universal covering with cusps over D . The restriction of p to $\tilde{W}_0 = p^{-1}(S_0)$ is the universal covering of S_0 , and the group $G = \text{Aut}(\tilde{W}/S)$ is the fundamental group $\pi_1(S_0)$.*

(ii) *If S' is an open submanifold of S and \tilde{W}' the universal covering of S' with cusps over $D' = D \cap S'$, then \tilde{W}'/H' embeds as an open subspace into \tilde{W} , where H' is the kernel of the homomorphism $\pi_1(S' - D') \rightarrow \pi_1(S - D) = G$.*

Proof. We only sketch the construction of the space $(\tilde{W}, \mathcal{O}_{\tilde{W}})$. The details that it satisfies all the required properties are worked out in [7]. For the proof of (ii) we refer to [8].

Let $S_0 = S - D$, $G = \pi_1(S_0)$ and $p_0: W_0 \rightarrow S_0$ the universal covering. \tilde{W} is obtained from W_0 by "filling in the holes above D " in such a way that the G -action extends from W_0 to \tilde{W} . More formally, the fibre \tilde{W}_s of \tilde{W} over any point $s \in S$ is

constructed as follows: let $\mathfrak{U}(s)$ be the set of open connected neighbourhoods of s in S ; for any $U \in \mathfrak{U}(s)$ denote by $X(U)$ the set of connected components of $p_0^{-1}(U)$. Then

$$\tilde{W}_s = \{(x_U)_{U \in \mathfrak{U}(s)} : x_U \in X(U), x_U \cap x_{U'} \neq \emptyset \text{ for all } U, U' \in \mathfrak{U}(s)\}.$$

Clearly $\tilde{W}_s = p_0^{-1}(s)$ for $s \in S_0$. Note that by definition, G acts transitively on each \tilde{W}_s , thus $\tilde{W}/G = S$. For any $x = (x_U) \in \tilde{W}$ define the sets $x_U \cup \{x\}$, $U \in \mathfrak{U}(s)$, to be open neighbourhoods of x . Finally define the structure sheaf by

$$\mathcal{O}_{\tilde{W}}(U) = \{f : U \rightarrow \mathbb{C} \text{ continuous} : f \text{ holomorphic on } U \cap \tilde{W}_0\} \quad (3.1)$$

for any open subset U of \tilde{W} . □

A key point in Braungardt’s proof of Theorem 3.4 is the existence of neighbourhoods U for any point $a \in D$ such that the natural homomorphism

$$\pi_1(U - D) \rightarrow \varinjlim_{U' \in \mathfrak{U}(a)} \pi_1(U' - D)$$

is an isomorphism. He calls such neighbourhoods *decent*. The importance of this notion is that if U is a decent neighbourhood of a point $a \in D$ and \bar{x}_U a connected component of $p^{-1}(U)$, then \bar{x}_U is precisely invariant under the stabilizer G_x in G of the unique point $x \in \bar{x}_U \cap p^{-1}(a)$.

Decent neighbourhoods in the above sense do not exist in general for singular complex spaces. For example, if S is a stable Riemann surface and $s \in S$ a node, $U - \{s\}$ is not even connected for small neighbourhoods U of s . Nevertheless the construction can be generalized to this case, and the proof of the theorem carries over to this case as Braungardt explains in [8, Anmerkung 3.1.4]; we therefore have:

Corollary 3.5. *Any stable Riemann surface has a universal covering with cusps over the nodes.*

Near the inverse image of a node, the universal covering of a stable Riemann surface looks like two copies of $\hat{\mathbb{H}}$ glued together in the cusps. If such a neighbourhood is embedded into the complex plane or $\mathbb{P}^1(\mathbb{C})$ it is called a *doubly cusped region*, cf. [25, VI.A.8].

3.2 The cusped universal covering of $\bar{M}_{g,n}$

Let us now fix nonnegative integers g, n such that $3g - 3 + n > 0$. We want to construct the space $\bar{\mathcal{T}}_{g,n}$ as the universal covering of $\bar{M}_{g,n}$ with cusps over the compactification (or boundary) divisor $\partial M_{g,n}$. But we cannot apply Theorem 3.4 directly to $\bar{M}_{g,n}$ since it is not a manifold, but only an orbifold (or smooth stack). Braungardt circumvents this difficulty by

Definition 3.6. A morphism $p : Y \rightarrow \bar{M}_{g,n}$ of locally complex ringed spaces is called a *covering with cusps over $D = \partial M_{g,n}$* if there is an open covering $(U_i)_{i \in I}$ of $\bar{M}_{g,n}$

and for each $i \in I$ a covering $q_i : U'_i \rightarrow U_i$ by a complex manifold U'_i such that $p|_{p^{-1}(U_i)}$ factors as $p^{-1}(U_i) \xrightarrow{p'_i} U'_i \xrightarrow{q_i} U_i$, where p'_i is a covering with cusps over $q_i^{-1}(D)$ (in the sense of Definition 3.3).

Then one can use Theorem 3.4 to prove

Proposition 3.7. *There is a universal covering $\overline{\mathcal{T}}_{g,n} \rightarrow \overline{M}_{g,n}$ with cusps over $\partial M_{g,n}$.*

Proof. We first construct local universal coverings and then glue them together. For any $s \in \overline{M}_{g,n}$ choose an open neighbourhood U and a covering $q' : U' \rightarrow U$ with a manifold U' . Let \tilde{W}' be the universal covering of U' with cusps over $D' = q'^{-1}(D)$. Let H' be the kernel of the homomorphism from $\pi_1(U' - D')$ to $\Gamma_{g,n}$. Theorem 3.4 (ii) suggests that the quotient \tilde{W}'/H' should be an open part of the universal covering of $\overline{M}_{g,n}$. All that remains to show is that the \tilde{W}'/H' glue together to a covering with cusps over D . This is done in [8, 3.2.1] \square

Locally $\overline{\mathcal{T}}_{g,n}$ looks like a product of a ball with some copies of the universal covering $\hat{\mathbb{H}}$ of \mathbb{D} with cusps over $\{0\}$ which was explained in Section 3.1:

Corollary 3.8. *Let $x \in \overline{\mathcal{T}}_{g,n}$ correspond to a stable Riemann surface X with k nodes. Then x has a neighbourhood that is isomorphic to*

$$\hat{\mathbb{H}}^k \times \mathbb{D}^{3g-3+n-k}.$$

Proof. Let $s \in \overline{M}_{g,n}$ be the image point of x . The deformation theory of stable Riemann surfaces gives us a map from \mathbb{D}^{3g-3+n} onto a neighbourhood of s such that the inverse image of $D = \partial M_{g,n}$ is the union of axes $D' = \{(z_1, \dots, z_{3g-3+n}) : z_1 \dots z_k = 0\}$, see [16, Section 3B]. The fundamental group of $\mathbb{D}^{3g-3+n} - D'$ is a free abelian group on k generators; they correspond to Dehn twists about the loops that are contracted in X . Thus the homomorphism $\pi_1(\mathbb{D}^{3g-3+n} - D') \rightarrow \Gamma_{g,n}$ is injective. By Proposition 3.7 and its proof the universal covering \tilde{W} of \mathbb{D}^{3g-3+n} with cusps over D' is therefore a neighbourhood of x . It is not hard to see that \tilde{W} is of the given form. \square

Our next goal is to compare $\overline{\mathcal{T}}_{g,n}$ to the *augmented* Teichmüller space $\hat{\mathcal{T}}_{g,n}$ introduced by Abikoff [1].

Proposition 3.9 (cf. [8], Satz 3.4.2). *$\overline{\mathcal{T}}_{g,n}$ is homeomorphic to the augmented Teichmüller space $\hat{\mathcal{T}}_{g,n}$.*

Before proving the proposition we summarize the definition and some properties of $\hat{\mathcal{T}}_{g,n}$: As a point set,

$$\hat{\mathcal{T}}_{g,n} = \{(X, f) : X \text{ a stable Riemann surface of type } (g, n), \quad (3.2)$$

$$f : X_{\text{ref}} \rightarrow X \text{ a deformation}\} / \sim .$$

As mentioned in the introduction, a deformation is a map that contracts some disjoint loops on X_{ref} to points (the nodes of X) and is a homeomorphism otherwise. The equivalence relation is the same as for $\mathcal{T}_{g,n}$: $(X, f) \sim (X', f')$ if and only if there is a biholomorphic map $h: X \rightarrow X'$ such that f' is homotopic to $h \circ f$.

Abikoff puts a topology on $\hat{\mathcal{T}}_{g,n}$ by defining neighbourhoods $U_{V,\varepsilon}$ of a point (X, f) for a compact neighbourhood V of the set of nodes in X and $\varepsilon > 0$:

$$U_{V,\varepsilon} = \{(X', f') : \text{there is a deformation } h: X' \rightarrow X, (1 + \varepsilon)\text{-quasiconformal} \\ \text{on } h^{-1}(X - V), \text{ such that } f \text{ is homotopic to } h \circ f'\} / \sim. \quad (3.3)$$

The action of the mapping class group $\Gamma_{g,n}$ extends continuously to $\hat{\mathcal{T}}_{g,n}$ ([1, Theorem 4]), and the orbit space $\hat{\mathcal{T}}_{g,n} / \Gamma_{g,n}$ is $\bar{M}_{g,n}$ (as a topological space).

Proof of Proposition 3.9. Braungardt shows (see [8, Hilfssatz 3.4.4]) that the stabilizer of a point $(X, f) \in \hat{\mathcal{T}}_{g,n}$ in $\Gamma_{g,n}$ is an extension of the free abelian group generated by the Dehn twists about the contracted loops by the holomorphic automorphism group $\text{Aut}(X)$ of X . For any V and ε , $\bigcap_{\sigma \in \text{Aut}(X)} \sigma(U_{V,\varepsilon})$ is invariant under the stabilizer of (X, f) , and for sufficiently small V and ε , it is precisely invariant. Therefore the quotient map $\hat{\mathcal{T}}_{g,n} \rightarrow \bar{M}_{g,n}$ is a covering with cusps over $\partial M_{g,n}$ in the sense of Definition 3.6, except that so far no structure sheaf has been defined on $\hat{\mathcal{T}}_{g,n}$. But this can be done in the same way as in (3.1). The universal property of $\bar{\mathcal{T}}_{g,n}$ then yields a map $p: \bar{\mathcal{T}}_{g,n} \rightarrow \hat{\mathcal{T}}_{g,n}$ compatible with the action of $\Gamma_{g,n}$ on both sides. To show that this map is an isomorphism we compare the stabilizers in $\Gamma_{g,n}$ for the points in both spaces. For a point in $\hat{\mathcal{T}}_{g,n}$ we just described this stabilizer, and the proof of Corollary 3.8 shows that for a corresponding point in $\bar{\mathcal{T}}_{g,n}$ it is also an extension of \mathbb{Z}^k by $\text{Aut}(X)$. \square

3.3 Teichmüller structures

In this section we explain how Braungardt extends the universal family of marked Riemann surfaces that is well known to exist over $\mathcal{T}_{g,n}$ to a family over $\bar{\mathcal{T}}_{g,n}$ which still is universal for the appropriate notion of marking or *Teichmüller structure*.

As above we fix a reference Riemann surface X_{ref} of type (g, n) ; let Q_1, \dots, Q_n be the marked points and $X_{\text{ref}}^0 = X_{\text{ref}} - \{Q_1, \dots, Q_n\}$. Let us also fix a universal covering $U_{\text{ref}} \rightarrow X_{\text{ref}}^0$ and identify $\pi_{g,n} = \pi_1(X_{\text{ref}}^0)$ with the group $\text{Aut}(U_{\text{ref}}/X_{\text{ref}}^0)$ of deck transformations.

A classical construction of the family $\mathcal{C}_{g,n}$ over $\mathcal{T}_{g,n}$ goes as follows (cf. [5]): For every point $x = (X, P_1, \dots, P_n, f) \in \mathcal{T}_{g,n}$ take a universal covering of $X^0 = X - \{P_1, \dots, P_n\}$ and arrange them so that they form an \mathbb{H} -bundle Ω^+ over $\mathcal{T}_{g,n}$. Then $\mathcal{C}_{g,n}$ is obtained as the quotient of Ω^+ by the natural action of $\pi_{g,n}$. More precisely, Ω^+ is defined as follows: to $x \in \mathcal{T}_{g,n}$ there corresponds the quasi-Fuchsian group $G_x = w^\mu G (w^\mu)^{-1}$, where $G = \text{Aut}(U_{\text{ref}}/X_{\text{ref}}^0) \cong \pi_{g,n}$ and w^μ is the quasicon-

formal automorphism of $\mathbb{P}^1(\mathbb{C})$ associated to x , see e.g. [21, 6.1.1]. The domain of discontinuity of G_x consists of two connected components $\Omega^-(x) = w^\mu(L)$ (where L is the lower half plane) and $\Omega^+(x) = w^\mu(\mathbb{H})$. Then $\Omega^+(x)/G_x = X^0$, whereas $\Omega^-(x)/G_x = X_{\text{ref}}^{0,*}$, the mirror image of X_{ref}^0 .

To extend this family we identify $\overline{\mathcal{T}}_{g,n}$ with $\widehat{\mathcal{T}}_{g,n}$ by Proposition 3.9. As explained in [1], any point $x = (X, P_1, \dots, P_n, f) \in \widehat{\mathcal{T}}_{g,n} - \overline{\mathcal{T}}_{g,n}$ corresponds to a *regular B-group* G_x . This means that G_x is a Kleinian group isomorphic to $\pi_{g,n}$ whose domain of discontinuity $\Omega(G_x)$ has a unique simply connected invariant component $\Omega^-(G_x)$ such that $\Omega^-(G_x)/G_x$ is isomorphic to $X_{\text{ref}}^{0,*}$. For the union $\Omega^+(G_x) = \Omega^+(x)$ of the other components of $\Omega(G_x)$ it holds that $\Omega^+(G_x)/G_x \cong X^0 - \{\text{nodes}\}$. To every node in X there corresponds a conjugacy class of parabolic elements in G_x , each of which is accidental (i.e. it becomes hyperbolic in the Fuchsian group $hG_x h^{-1}$, where $h: \Omega^-(G_x) \rightarrow \mathbb{H}$ is a conformal map). Near a fixed point of such a parabolic element, $\Omega^+(G_x)$ is a doubly cusped region, cf. the remark at the end of Section 3.1. If we denote by $\widehat{\Omega}^+(x)$ the union of $\Omega^+(G_x)$ with the fixed points of the parabolic elements in G_x (accidental or not), then $\widehat{\Omega}^+(x) \rightarrow X$ is the universal covering of X with cusps over the nodes (cf. Corollary 3.5).

Definition 3.10. Let

$$\widehat{\Omega}_{g,n}^+ = \{(x, z) \in \overline{\mathcal{T}}_{g,n} \times \mathbb{P}^1(\mathbb{C}) : z \in \widehat{\Omega}^+(x)\}.$$

On $\widehat{\Omega}_{g,n}^+$, $\pi_{g,n}$ acts in such a way that for fixed $x \in \overline{\mathcal{T}}_{g,n}$ the action on $\Omega^+(x)$ is that of G_x . $\overline{\mathcal{C}}_{g,n} = \widehat{\Omega}_{g,n}^+/\pi_{g,n}$ is called the *universal family* over $\overline{\mathcal{T}}_{g,n}$.

Braungardt shows ([8, Hilfssatz 4.2.1]) that $\Omega_{g,n}^+ = \{(x, z) \in \widehat{\Omega}_{g,n}^+ : x \in \overline{\mathcal{T}}_{g,n}, z \in \Omega^+(x)\}$ is an open subset of $\overline{\mathcal{T}}_{g,n} \times \mathbb{P}^1(\mathbb{C})$ and hence has a well defined structure of a complex ringed space. One can extend this structure sheaf to all of $\widehat{\Omega}_{g,n}^+$ in the same way as in (3.1). Then clearly $\overline{\mathcal{C}}_{g,n}$ is also a complex ringed space, and the fibre over x is isomorphic to the stable Riemann surface X represented by x .

To justify the name “universal” family for $\overline{\mathcal{C}}_{g,n}$ we introduce the notion of a Teichmüller structure: For a single smooth Riemann surface (X, P_1, \dots, P_n) of type (g, n) , a Teichmüller structure is just a marking: so far we used markings as classes of mappings $X_{\text{ref}} \rightarrow X$; equivalently a marking can be given as an isomorphism $\pi_{g,n} \rightarrow \pi_1(X - \{P_1, \dots, P_n\})$ inducing an isomorphism $\pi_g = \pi_1(X_{\text{ref}}) \rightarrow \pi_1(X)$ and respecting the orientation and the conjugacy classes of the loops around the Q_i resp. P_i . Yet another equivalent way to give a marking is as a universal covering $U \rightarrow X^0$ together with an isomorphism $\pi_{g,n} \rightarrow \text{Aut}(U/X^0)$. This last characterization also works for a stable Riemann surface if we take for U a universal covering with cusps over the nodes. Before we extend this definition to the relative situation we recall the notion of a family of stable Riemann surfaces.

Definition 3.11. Let S be a complex ringed space. A *family of stable Riemann surfaces* of type (g, n) over S is a complex ringed space \mathcal{C} together with a proper flat map $\pi: \mathcal{C} \rightarrow S$ such that the fibres $X_s = \pi^{-1}(s)$, $s \in S$, are stable Riemann surfaces of genus g . In addition we are given n disjoint sections $P_i: S \rightarrow \mathcal{C}$, $i = 1, \dots, n$, of π such that $P_i(s)$ is not a node on X_s . We denote by $\mathcal{C}^0 = \mathcal{C} - \bigcup_{i=1}^n P_i(S)$ the complement of the marked sections.

Definition 3.12. Let \mathcal{C}/S be a family of stable Riemann surfaces of type (g, n) over a complex ringed space S . A *Teichmüller structure* on \mathcal{C} is a complex ringed space \mathcal{U} together with a morphism $\mathcal{U} \rightarrow \mathcal{C}$ such that for every $s \in S$ the (restriction of the) fibre $U_s^0 \rightarrow X_s^0$ is a universal covering with cusps over the nodes, together with an isomorphism $\pi_{g,n} \rightarrow \text{Aut}(\mathcal{U}/\mathcal{C}^0)$.

Putting everything together we obtain

Theorem 3.13. $\overline{\mathcal{T}}_{g,n}$ is a fine moduli space for stable Riemann surfaces with Teichmüller structure. $\overline{\mathcal{C}}_{g,n} \rightarrow \overline{\mathcal{T}}_{g,n}$ is the universal family and $\hat{\Omega}_{g,n}^+ \rightarrow \overline{\mathcal{C}}_{g,n} = \hat{\Omega}_{g,n}^+/\pi_{g,n}$ is the universal Teichmüller structure.

Finally Braungardt gives a very elegant and conceptual description of $\overline{\mathcal{C}}_{g,n}$ which extends a classical result of Bers ([5, Theorem 9]) to the boundary:

Proposition 3.14. $\overline{\mathcal{T}}_{g,n+1}/\pi_{g,n}$ is in a natural way isomorphic to $\overline{\mathcal{C}}_{g,n}$.

Proof. The kernel of the obvious homomorphism $\Gamma_{g,n+1} \rightarrow \Gamma_{g,n}$ can be identified with $\pi_{g,n}$, which gives the action on $\overline{\mathcal{T}}_{g,n+1}$. The holomorphic map $\mathcal{T}_{g,n+1} \rightarrow \mathcal{T}_{g,n}$ which forgets the last marked point extends to a map $\overline{\mathcal{T}}_{g,n+1} \rightarrow \overline{\mathcal{T}}_{g,n}$ by a general property of universal coverings with cusps. The difficult step in Braungardt's proof is to show that the induced map $\overline{\mathcal{T}}_{g,n+1}/\pi_{g,n} \rightarrow \overline{\mathcal{T}}_{g,n}$ has the right fibres. For this purpose he constructs a map $\hat{\Omega}_{g,n}^+ \rightarrow \overline{\mathcal{T}}_{g,n+1}$ and shows that it is bijective and induces isomorphisms on the fibres over $\overline{\mathcal{T}}_{g,n}$. \square

4 Boundary points of Teichmüller curves

The aim of this chapter is to study the boundary points of the Teichmüller disks and Teichmüller curves introduced in Section 2 in $\overline{\mathcal{T}}_g$ and \overline{M}_g , respectively. Here and later, whenever we speak about $\overline{\mathcal{T}}_g$ and its boundary, we mean the bordification of the Teichmüller space described in Section 3.

In particular we shall derive, in Section 4.2, the following description of the boundary points of Teichmüller curves (see Proposition 4.14 and Corollary 4.15 for a more precise formulation):

Theorem 4.1. *One obtains the boundary points of a Teichmüller curve by contracting the centers of all cylinders in Jenkins–Strebel directions. They are determined by the parabolic elements in the associated mirror Veech group.*

This statement seems to be well known to the experts although we are not aware of a published proof.

In Section 4.1 we prepare for the proof of Theorem 4.1 by introducing Jenkins–Strebel rays. They are special geodesic rays in Teichmüller space which always converge to a point on the boundary. Following Masur [26], we describe this boundary point quite explicitly using the affine structure of the quadratic differential q that defines the Jenkins–Strebel ray.

In Section 4.2 we turn to the boundary points of Jenkins–Strebel rays that are contained in a Teichmüller disk. In particular if the Teichmüller disk descends to a Teichmüller curve in the moduli space, all its boundary points can be determined explicitly with the aid of the projective Veech group. One obtains Theorem 4.1 as a conclusion.

4.1 Hitting the boundary via a Jenkins–Strebel ray

In this section, we introduce Jenkins–Strebel rays and describe their end point on the boundary of $\widetilde{\mathcal{T}}_g$. As before, everything might be done as well for punctured surfaces and the moduli space $\widetilde{\mathcal{T}}_{g,n}$ with $3g - 3 + n > 0$, but for ease of notation, we restrict to the case $n = 0$.

Let X be a Riemann surface of genus $g \geq 2$, q a holomorphic quadratic differential on X . Recall from Section 2.1 that with q we have chosen a natural flat structure μ on the surface $X^* = X - \{\text{critical points of } q\}$ whose charts were given in (2.1). The maximal real curves in X^* which are locally mapped by these charts to horizontal (resp. vertical) line segments are called *horizontal* (resp. *vertical*) *trajectories*. A trajectory is *critical* if it ends in a critical point. Otherwise it is *regular*.

Definition 4.2. We say that a holomorphic quadratic differential q is *Jenkins–Strebel*, if all regular horizontal trajectories are closed.

Jenkins–Strebel differentials play an exceptional role in the following sense. Recall from Section 2.2 that each holomorphic quadratic differential defines a geodesic ray. If q is Jenkins–Strebel, then the geodesic ray defined by its negative $-q$ converges in $\widetilde{\mathcal{T}}_g$ to an end point on the boundary. This is described more precisely in the following proposition which was proven by Masur in [26]. We give a version of his proof with parts of the notation and arguments adapted to the context of our exposition.

Recall also from 2.2 that we obtain the geodesic ray to $-q$ as the image of the isometric embedding

$$\gamma = \gamma_{-q}: \begin{cases} [0, \infty) \rightarrow \widetilde{\mathcal{T}}_g, \\ t \mapsto (X_K, f_K) = [(X, \mu_{-q}) \circ \left(\begin{smallmatrix} K & 0 \\ 0 & 1 \end{smallmatrix}\right), \text{id}] \quad \text{with } K = e^t. \end{cases} \quad (4.1)$$

Note that here (X_K, f_K) is the Teichmüller deformation of X of dilatation K with respect to $-q$. Furthermore, μ_{-q} is the translation structure on X^* defined by $-q$.

Proposition 4.3. *Suppose $q \neq 0$ is a Jenkins–Strebel differential. For the geodesic ray defined by γ_{-q} in \mathcal{T}_g , one has:*

- a) *The ray converges towards a unique point (X_∞, f_∞) on the boundary of the Teichmüller space \mathcal{T}_g .*
- b) *One obtains this point by contracting the central lines of the horizontal cylinders defined by q as is described in 4.1.4.*

Definition 4.4. In the previous proposition, the geodesic ray defined by $-q$, i.e. the image of γ_{-q} in \mathcal{T}_g , is called a *Jenkins–Strebel ray*.

For the proof of Proposition 4.3 one may use two slightly different perspectives of the Jenkins–Strebel ray. They are described in 4.1.1, 4.1.2 and 4.1.3, 4.1.4. In 4.1.5 we describe the boundary point (X_∞, f_∞) . In 4.1.6 we show that the Jenkins–Strebel ray in fact converges towards this point.

Throughout Section 4.1, we assume that the differential q is Jenkins–Strebel.

4.1.1 X as patchwork of rectangles. One may regard X as a patchwork of rectangles in the complex plane, as is described in the following.

Since q is Jenkins–Strebel, the surface X , with the critical points and critical horizontal trajectories removed, is swept out by closed horizontal trajectories. More precisely, it follows from the work of Strebel (cf. [33], also see [26, Theorem B] which contains a list of the results we use here) that the surface X , except for the critical points and critical horizontal trajectories, is covered by a finite number of maximal *horizontal cylinders* Z_1, \dots, Z_p , i.e. annuli that are swept out by closed horizontal trajectories. For each Z_i one may choose a vertical trajectory β_i joining opposite boundary components of Z_i . If we remove β_i from Z_i , the remainder is mapped, by the natural chart w_i defined by μ (see (2.1)), to an open rectangle R_i in the complex plane. The horizontal and vertical edges have lengths

$$a_i = \int_{\alpha_i} |q(z)|^{\frac{1}{2}} dz \quad \text{and} \quad b_i = \int_{\beta_i} |q(z)|^{\frac{1}{2}} dz,$$

where α_i is any closed horizontal trajectory in the cylinder Z_i .

One may extend w_i^{-1} uniquely to a map from the closure \bar{R}_i of R_i to the closure of the annulus Z_i . Then the two horizontal edges of \bar{R}_i are mapped to the two horizontal boundary components of Z_i and the two vertical edges are both mapped to β_i . The critical points of q that lie on the boundary of Z_i define by their preimage marked points on the horizontal edges of \bar{R}_i and decompose them into segments.

For each such segment s on a horizontal edge of \bar{R}_i its image on X joins the annulus Z_i to an annulus Z_j possibly with $i = j$.

Thus the map $w_j \circ w_i^{-1}$ (w_i^{-1} is the extended map, w_j is locally the inverse map of the extended map w_j^{-1}) is an *identification map* between s and a segment on a horizontal edge of \bar{R}_j . (Images of critical points have to be excluded.)

These identification maps are of the form $z \mapsto \pm z + c$ with a constant $c \in \mathbb{C}$.

Conversely, given the closed rectangles $\bar{R}_1, \dots, \bar{R}_p$, the marked points on their horizontal edges and these identification maps, we may recover the surface X as follows: for each i glue the two vertical edges of \bar{R}_i by a translation and the horizontal edges (with the marked points removed) by the identification maps. In this way, one obtains a surface X^* with the flat structure on it inherited from the euclidean plane \mathbb{C} . By filling in the punctures at vertices, we obtain the original compact Riemann surface X .

In this sense one may consider X as a patchwork of the rectangles $\bar{R}_1, \dots, \bar{R}_p$. This description depends of course on the chosen holomorphic quadratic Jenkins–Strebel differential q .

Example 4.5. *Two Riemann surfaces X given as a patchwork of rectangles.* In the two examples in Figure 3 and Figure 4, the two vertical edges of each rectangle are glued by a translation, respectively. Horizontal segments with the same name are glued. The direction of the arrow indicates whether the identification is a translation or a rotation by 180° . In the example in Figure 3 one only has translations, in the example in Figure 4 only rotations.

In the first example the surface X is of genus 2 and all marked points are identified and thus give only one point on X . In the second example one obtains a surface of genus 0 with four marked points indicated by the four symbols \bullet, \star, \circ and \square .

Surface of genus 2 with 1 marked point

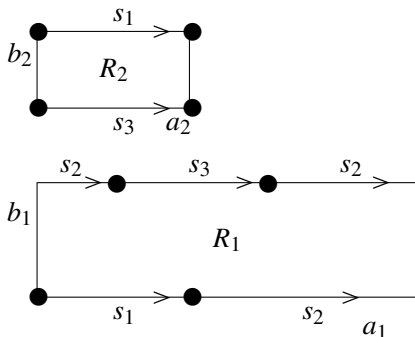


Figure 3

Surface of genus 0 with 4 marked points

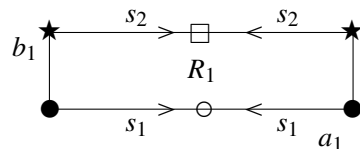


Figure 4

4.1.2 Stretching the cylinders. We shall now redescribe the Jenkins–Strebel ray defined by $-q$ by stretching the rectangles in the ‘patchwork’ from 4.1.1 in the vertical direction.

The flat structure defined by $-q = e^{\pi i} \cdot q$ is obtained from the flat structure μ defined by q by composing each chart with a rotation by $\frac{\pi}{2}$. Thus the deformation (X_K, f_K) of dilatation K with respect to $-q$ is equal to the affine deformation

$$\begin{pmatrix} K & 0 \\ 0 & 1 \end{pmatrix} \circ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \circ (X, \mu) = \begin{pmatrix} 0 & -K \\ 1 & 0 \end{pmatrix} \circ (X, \mu).$$

This defines by (2.9) the same point in \mathcal{T}_g as the affine deformation

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \circ \begin{pmatrix} 0 & -K \\ 1 & 0 \end{pmatrix} \circ (X, \mu) = \begin{pmatrix} 1 & 0 \\ 0 & K \end{pmatrix} \circ (X, \mu).$$

Thus the isometric embedding $\gamma = \gamma_{-q}$ in (4.1) is equivalently given by

$$\gamma_{-q}: \begin{cases} [0, \infty) \rightarrow \mathcal{T}_g, \\ t \mapsto (X_K, f_K) = \left[\begin{pmatrix} 1 & 0 \\ 0 & K \end{pmatrix} \circ (X, \mu), \text{id} \right], \quad K = e^t. \end{cases} \quad (4.2)$$

Recall again that here $(X_K, f_K) = (X_K^{-q}, f_K^{-q})$ is the Teichmüller deformation with respect to the differential $-q$.

Hence we obtain the point $\gamma_{-q}(t)$ as follows: Each chart of μ is composed with the map $x + iy \mapsto x + iKy$ ($x, y \in \mathbb{R}$) with $K = e^t$, and the marking is topologically the identity. Now, let X be given as a patchwork of the rectangles $\bar{R}_1, \dots, \bar{R}_p$ as in 4.1.1. Then we obtain the surface $X_K = X_K^{-q}$ in the following way: We stretch each rectangle \bar{R}_i , which has horizontal and vertical edges of lengths a_i and b_i , into a rectangle $\bar{R}_i(K)$ with horizontal and vertical edges of lengths a_i and $K \cdot b_i$. The identification maps of the horizontal segments are again translations or rotations identifying the same segments as before. The surface $X_K = X_K^{-q}$ then is the patchwork obtained from $\bar{R}_1(K), \dots, \bar{R}_p(K)$ as described in 4.1.1.

On \bar{R}_i , the diffeomorphism $f_K = f_K^{-q}$ has image $\bar{R}_i(K)$ and is given by

$$x + iy \mapsto x + iKy.$$

This glues to a well-defined diffeomorphism on X^* , which can be uniquely extended to X .

Example 4.6. *K-stretched surfaces.* One obtains the surface $X_K = X_K^{-q}$ from the surface X in Example 4.5 as the patchwork of the stretched rectangles in Figure 5 and Figure 6, respectively.

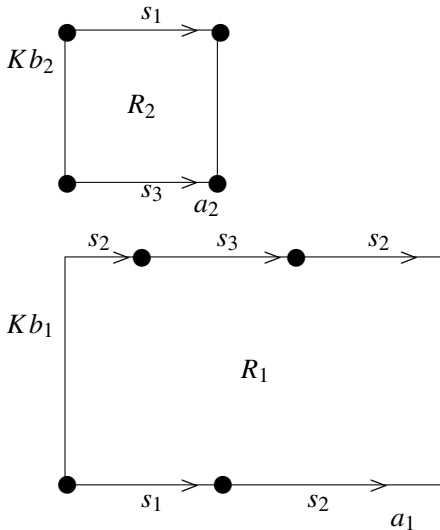


Figure 5

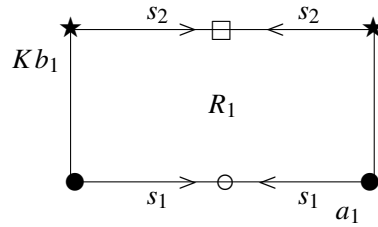


Figure 6

4.1.3 S as patchwork of double annuli. Recall that, in 4.1.1, we used μ to identify the horizontal cylinder Z_i on X with the euclidean cylinder defined by the rectangle R_i in \mathbb{C} ; we did so by adding the vertical boundary edges and identifying them by a translation. It turns out to be easier to describe the end point of the Jenkins–Strebel ray, if we identify the Z_i with so called double annuli A_i .

Definition 4.7. A cylinder Z of length a and height b defines a *double annulus* A as follows:

- Take two disjoint open annuli A^1 and A^2 given as

$$A^1 = A^2 = \{z \in \mathbb{C} : r \leq |z| < 1\} \quad \text{with } r = e^{-\pi \frac{b}{a}}.$$

- Glue their inner boundary lines $\{|z| = r\}$ by the map $z \mapsto \frac{1}{z} \cdot r^2$.
- We call the resulting surface A the *double annulus* of Z .

Remark 4.8. A is biholomorphic to Z .

The identification is given explicitly as follows:

- Z is biholomorphic to the Euclidean cylinder defined by the rectangle

$$\{z \in \mathbb{C} : 0 \leq \text{Re}(z) \leq a, 0 < \text{Im}(z) < b\}.$$

- Decompose the rectangle into two halves of height $\frac{b}{2}$, a lower half $R^1 = \{z \in \mathbb{C} : 0 \leq \text{Re}(z) \leq a, 0 < \text{Im}(z) \leq \frac{b}{2}\}$ and an upper half $R^2 = \{z \in \mathbb{C} : 0 \leq \text{Re}(z) \leq a, \frac{b}{2} \leq \text{Im}(z) < b\}$.

- The cylinder defined by R^1 is mapped to A^1 by $z \mapsto e^{2\pi i \frac{z}{a}}$.

The cylinder defined by R^2 is mapped to A^2 by $z \mapsto e^{2\pi i \frac{a+bi-z}{a}}$.

These maps respect the identifications and define a biholomorphic map from Z to A , as shown in Figure 7.

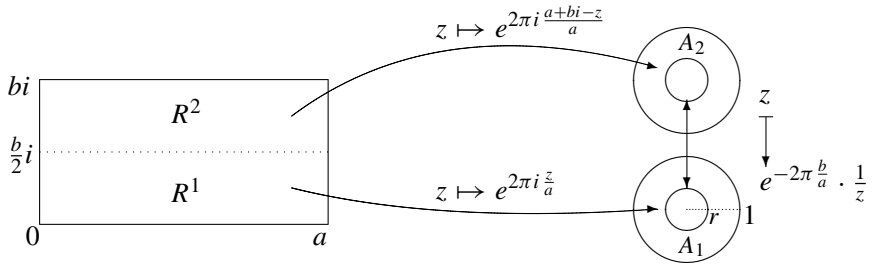


Figure 7

Consider the double annuli A_1, \dots, A_p defined by the cylinders Z_1, \dots, Z_p . The biholomorphic map $Z_i \rightarrow A_i$ extends to a continuous map from the closure of Z_i to the closure \bar{A}_i of A_i . The zeroes of q on the boundary of Z_i define marked points on the boundary of A_i and decompose it into segments. The surface X can now be described as a *patchwork of the closed double cylinders* $\bar{A}_1, \dots, \bar{A}_p$. The identification maps between the segments on the boundary of the A_i are essentially the same as in 4.1.1.

4.1.4 Contracting the central lines. Suppose that X is given as a patchwork of double annuli $\bar{A}_1, \dots, \bar{A}_p$ as in 4.1.3. We may describe the points $(X_K, f_K) = (X_K^{-q}, f_K^{-q})$ on the Jenkins–Strebel ray also as a patchwork of double annuli:

Let $A_i(K) = A_i^1(K) \cup A_i^2(K)$ ($i \in \{1, \dots, p\}$) be the double annulus from Definition 4.7 with $r = r_i(K) = r_i^K$ and define $X_K = X_K^{-q}$ to be the surface obtained by gluing the closures $\bar{A}_1(K), \dots, \bar{A}_p(K)$ with the same maps as $\bar{A}_1, \dots, \bar{A}_p$. Furthermore, define the diffeomorphism $f_K = f_K^{-q}$ on A_i by

$$f_K^{-q}: A_i^1 \rightarrow A_i^1(K) \quad \text{and} \quad A_i^2 \rightarrow A_i^2(K),$$

where

$$z = r \cdot e^{i\varphi} \mapsto r^K \cdot e^{i\varphi} \quad \text{on both parts.}$$

Then the following diagram is commutative:

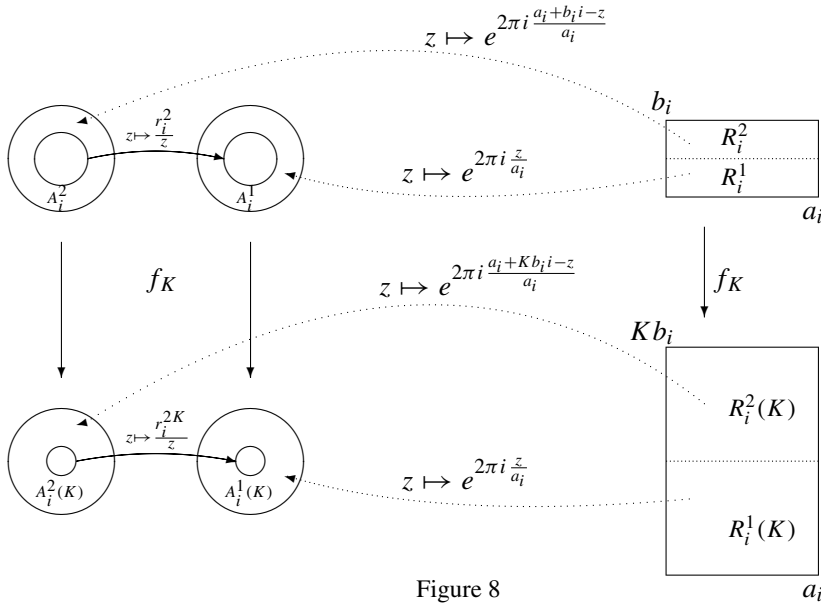


Figure 8

where $f_K = (r e^{i\varphi} \mapsto r^K e^{i\varphi})$ on the left side and $f_K = (x + iy \mapsto x + Kiy)$ on the right side of the diagram. Thus, in particular, we have defined here with $(X_K, f_K) = (X_K^{-q}, f_K^{-q})$ the same surface (up to isomorphism) and the same diffeomorphism as in 4.1.2.

4.1.5 The end point of the Jenkins–Strebel ray. We use the description of the Jenkins–Strebel ray in 4.1.4 to obtain its end point $(X_\infty, f_\infty) \in \overline{\mathcal{T}}_g$. Recall from 3.2 that a point in $\overline{\mathcal{T}}_g$ consists of a stable Riemann surface X_∞ and a deformation $f_\infty : X \rightarrow X_\infty$.

If $K \rightarrow \infty$ in 4.1.4, the interior radius $r_i(K) = r_i^K$ of the two annuli $A_i^1(K)$ and $A_i^2(K)$ that form the double annulus $A_i(K)$ tends to 0 ($i \in \{1, \dots, p\}$). $A_i(K)$ tends to a double cone $A_i(\infty)$ and the whole surface X_K to a stable Riemann surface X_∞ . More precisely, we define $A_i(\infty)$ and X_∞ as complex spaces in the following way.

Definition 4.9. Let $A_i^1(\infty)$ and $A_i^2(\infty)$ both be the punctured disk

$$\{z \in \mathbb{C} : 0 < |z| < 1\},$$

and let pt be an arbitrary point. The disjoint union

$$A_i(\infty) = A_i^1(\infty) \cup A_i^2(\infty) \cup \{pt\}$$

becomes a complex cone by the following chart:

$$\varphi : A_i(\infty) \rightarrow \{(z_1, z_2) \in \mathbb{C}^2 : z_1 \cdot z_2 = 0, |z_1|, |z_2| < 1\},$$

$$\varphi|_{A_i^1(\infty)} : z \mapsto (0, z), \quad \varphi|_{A_i^2(\infty)} : z \mapsto (z, 0), \quad \varphi(pt) = (0, 0).$$

The closures of the double cones $\bar{A}_1(\infty), \dots, \bar{A}_p(\infty)$ are glued to each other by the same identification maps as in the 'finite' case in 4.1.4. We call the resulting stable Riemann surface X_∞ . Topologically, X_∞ is obtained from the surface X by a contraction f_∞ of the middle curves of the cylinders.

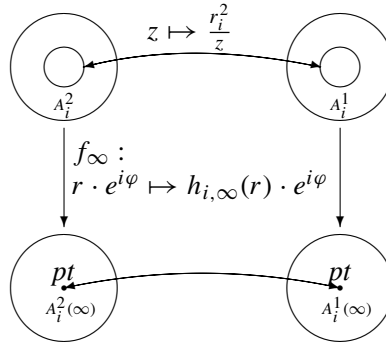


Figure 9

We now define the contraction f_∞ as the following map: Let A_i^1 and A_i^2 be the two annuli in Definition 4.7 that form the double annulus A_i ($i \in \{1, \dots, p\}$). Then f_∞ is given by

$$f_\infty : A_i^j \rightarrow A_i^j(\infty) \quad \text{for } j \in \{1, 2\},$$

$$z = r \cdot e^{i\varphi} \mapsto f_\infty(z) = h_{i,\infty}(r) \cdot e^{i\varphi},$$

with an arbitrary monotonously increasing diffeomorphism $h_{i,\infty} : [r_i, 1) \rightarrow [0, 1)$. The isotopy class of f_∞ is independent of the choices of $h_{i,\infty}$.

4.1.6 Convergence. We now show that, in the above notation, the Jenkins–Strebel ray γ_{-q} converges to the point (X_∞, f_∞) on the boundary of \mathcal{T}_g . Recall from (3.3) that a base of open neighbourhoods of (X_∞, f_∞) is given by the open sets

$$U_{V,\varepsilon}(X_\infty, f_\infty) = \{(X', f') : \text{there exists a deformation } \varphi : X' \rightarrow X_\infty$$

$$\text{such that } \varphi \circ f' \text{ is isotopic to } f_\infty$$

$$\text{and } \varphi|_{X' \setminus \varphi^{-1}(V)} \text{ has dilatation } < 1 + \varepsilon\},$$

for all compact neighbourhoods V of the singular points of X_∞ and for all $\varepsilon > 0$. We may restrict to open neighbourhoods V of the form

$$V = V(\kappa) = V_1 \cup \dots \cup V_p, \quad \kappa = (\kappa_1, \dots, \kappa_p), \quad 0 < \kappa_i < 1$$

where V_i is a double cone defined by

$$V_i = V_i^1 \cup V_i^2 \cup \{pt\} \quad \text{with } V_i^j = \{0 < |z| \leq \kappa_i\} \subseteq A_i^j(\infty), \quad j \in \{1, 2\}.$$

Lemma 4.10. *For each such $V = V(\kappa)$ and each $\varepsilon > 0$, there is some $K_0 \in \mathbb{R}_{>0}$ such that all points $(X_K, f_K) = (X_K^{-q}, f_K^{-q})$ with $K > K_0$ are in $U_{V,\varepsilon}(X_\infty, f_\infty)$.*

Proof. Choose K_0 such that $r_i^{K_0} < \kappa_i$ for all $i \in \{1, \dots, p\}$ and suppose that $K > K_0$. Define the diffeomorphism $\varphi: X_K \rightarrow X_\infty$ on $\bar{A}_i^j(K)$ by

$$\varphi: z = r \cdot e^{i\theta} \mapsto \begin{cases} z \in A_i^j(\infty), & \text{if } 1 > |z| \geq \kappa_i \\ h_K^i(r) \cdot e^{i\theta} \in A_i^j(\infty), & \text{if } \kappa_i \geq |z| > r_i^K \\ pt \in A_i^j(\infty), & \text{if } |z| = r_i^K \end{cases}$$

with an arbitrary monotonously increasing diffeomorphism $h_K^i: (r_i^K, \kappa_i) \rightarrow (0, \kappa_i)$. Then $\varphi \circ f_K$ is isotopic to f_∞ and $\varphi|_{X_K \setminus \varphi^{-1}(V)}$ is holomorphic, hence its dilatation is 1. Thus (X_K, f_K) is in $U_{V,\varepsilon}(X_\infty, f_\infty)$.

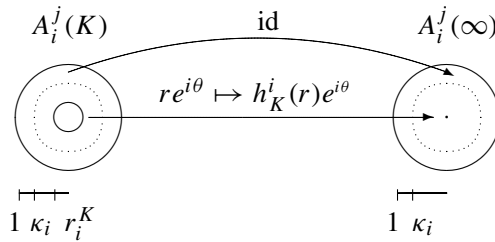


Figure 10

□

With Lemma 4.10 we have obtained the desired result and completed the proof of Proposition 4.3.

Corollary 4.11. *The Jenkins–Strebel ray defined by $-q$ converges to the point (X_∞, f_∞) on the boundary of \mathcal{T}_g .*

4.2 Boundary points of Teichmüller disks

In this section we study the boundary points of a Teichmüller disk $\Delta = \Delta_\iota$ in the bordification $\bar{\mathcal{T}}_g$ of the Teichmüller space; in particular, we consider the case that Δ_ι projects to an affine curve in the moduli space M_g . For convenience, we use the upper half plane model and consider Teichmüller embeddings as maps from \mathbb{H} to \mathcal{T}_g . We shall obtain Theorem 4.1 as our final result. We proceed in two steps:

- In 4.2.1, we show that a Teichmüller embedding $\iota: \mathbb{H} \hookrightarrow \mathcal{T}_g$ has a natural extension

$$\bar{\iota}: \mathbb{H} \cup \{\text{cusps of } \bar{\Gamma}_\iota^*\} \hookrightarrow \bar{\mathcal{T}}_g,$$

- In 4.2.2, we show that the image of $\bar{\iota}$ is the whole closure of Δ_i in $\bar{\mathcal{T}}_g$, if the Teichmüller disk Δ_i projects onto a Teichmüller curve in M_g .

It will follow from this that one obtains the boundary points of Δ_i precisely by contracting the central lines of the cylinders in “parabolic directions”. The parabolic directions correspond to the cusps of the projective mirror Veech group $\bar{\Gamma}_i^*$.

Throughout this section, we assume that $\iota: \mathbb{H} \hookrightarrow \mathcal{T}_g$ is a Teichmüller embedding to a fixed holomorphic quadratic differential q on $X = X_{\text{ref}}$ and that μ is the translation structure defined by q as in Section 2.1. Recall from Section 2.4 that the associated projective Veech group $\bar{\Gamma}_i = \bar{\Gamma}(X, \mu)$ and its mirror image $\bar{\Gamma}_i^* = R\bar{\Gamma}_i R^{-1}$ (with R as in Remark 2.20) are both Fuchsian groups in $\text{PSL}_2(\mathbb{R})$.

4.2.1 Extending Teichmüller embeddings to the cusps of $\bar{\Gamma}^*$. Let $\tilde{s} \in \mathbb{R}^\infty = \mathbb{R} \cup \{\infty\}$ be a cusp of the Fuchsian group $\bar{\Gamma}_i^*$, i.e. \tilde{s} is a fixed point of some parabolic element \tilde{A} of $\bar{\Gamma}_i^*$. We associate to \tilde{s} a point $\bar{\iota}(\tilde{s}) = (X_\infty(\tilde{s}), f_\infty(\tilde{s}))$ on the boundary of \mathcal{T}_g in the following way:

- In a natural way we associate to \tilde{s} a Jenkins–Strebel ray.
- We show that this Jenkins–Strebel ray is the image in \mathcal{T}_g of the hyperbolic ray in \mathbb{H} from i to \tilde{s} under ι .
- $\bar{\iota}(\tilde{s}) = (S_\infty(s), f_\infty(s))$ is defined to be the end point of the Jenkins–Strebel ray.

The Jenkins–Strebel ray associated to \tilde{s} . $A = R^{-1}\tilde{A}R$ is a parabolic element in the projective Veech group $\bar{\Gamma}_i$. Let v be its unit eigenvector.

By Proposition 2.4 in [34], the direction v is fixed by some affine diffeomorphism h of (X, μ) . The derivative of h is A and v is a Jenkins–Strebel direction. More precisely: The trajectories in the direction of v are preserved by h and each leaf is either closed or a saddle connection, i.e. connects two critical points.

As in 4.1.1, X decomposes into maximal cylinders of closed leaves parallel to v and the cylinders are bounded by saddle connections. The affine diffeomorphism h can be described nicely as follows: Passing to a power of h if necessary, one may assume that h fixes all critical points of q . Then h is the composition of Dehn twists along the core curves of the cylinders. Each trajectory is mapped by h to itself and the saddle connections are fixed pointwise.

Now, let us take the matrix

$$U = U_\theta \in \text{SO}_2(\mathbb{R}) \quad \text{such that } U \cdot v = \vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

with U_θ defined as in (2.12).

Consider the affine deformation $\text{id}: (X, \mu) \rightarrow (X, \mu_U) = (X, \mu) \circ U$ as in Definition 2.6. The vector v is mapped to \vec{e}_1 . Thus the same trajectories are now the horizontal ones.

Recall from 2.3.2 that the flat structure (X, μ_U) is defined by the quadratic differential $e^{2\theta i} \cdot q$. Thus $e^{2\theta i} \cdot q$ is Jenkins–Strebel. The ray is by (4.2) given as:

$$\begin{aligned} \gamma_{\tilde{s}} &= \gamma_{-e^{2\theta i} \cdot q}: [0, \infty) \rightarrow \mathcal{T}_g, \\ t &\mapsto \left[\begin{pmatrix} 1 & 0 \\ 0 & K \end{pmatrix} \circ (X, \mu_{U_\theta}), \text{id} \right] = [(X, \mu_{A_K}), \text{id}] \end{aligned}$$

with $K = e^t$ and $A_K = \begin{pmatrix} 1 & 0 \\ 0 & K \end{pmatrix} \cdot U_\theta$.

The Jenkins–Strebel ray $\gamma_{\tilde{s}}$ is the image of the geodesic ray in \mathbb{H} from i to the cusp \tilde{s} . From Remark 2.12 (see also Figure 1) one obtains that

$$\gamma_{\tilde{s}}(t) = [(X, \mu_{A_K}), \text{id}] = \hat{\iota}(A_K) = \iota(\overline{-A_K^{-1}(i)}).$$

Furthermore, we have

$$\overline{-A_K^{-1}(i)} = \overline{-U_\theta^{-1}(K \cdot i)} = -U_\theta^{-1}(-Ki) = RU_\theta^{-1}R^{-1}(Ki).$$

Thus the image of $\gamma_{\tilde{s}}$ is equal to the image of the ray $RU_\theta^{-1}R^{-1}(Ki)$ ($K \in [1, \infty)$) under ι . But the latter one is the geodesic ray in \mathbb{H} from i to $RU^{-1}R^{-1}(\infty) = -U^{-1}(\infty)$.

Observe finally that $-U^{-1}(\infty) = \tilde{s}$: Since $U \cdot v = \vec{e}_1$ for the eigenvector v of A , one has for the fixed point s of A that $U(s) = \infty$. Hence, one has for the fixed point \tilde{s} of $\tilde{A} = RAR^{-1}$ that $\tilde{s} = -s = -U^{-1}(\infty)$. Thus the Jenkins–Strebel ray defined by $\gamma_{\tilde{s}}$ is the image of the geodesic ray from i to \tilde{s} in \mathbb{H} under ι .

Finally we define $\bar{\iota}(\tilde{s}) = (X_\infty(\tilde{s}), f_\infty(\tilde{s})) \in \overline{\mathcal{T}}_g$ to be the end point of the Jenkins–Strebel ray $\gamma_{\tilde{s}}$. We then define the map $\bar{\iota}$ as follows.

Definition 4.12. $\bar{\iota}$ is the extension of ι defined by

$$\begin{aligned} \bar{\iota}: \mathbb{H} \cup \{\text{cusps of } \overline{\Gamma}_t^*\} &\rightarrow \overline{\mathcal{T}}_g, \\ t &\mapsto \begin{cases} \iota(t), & \text{if } t \in \mathbb{H} \\ \bar{\iota}(t) = (X_\infty(\tilde{s}), f_\infty(\tilde{s})), & \text{if } t = \tilde{s} \text{ is a cusp of } \overline{\Gamma}^*. \end{cases} \end{aligned}$$

We consider $\mathbb{H} \cup \{\text{cusps of } \overline{\Gamma}_t^*\}$ as topological space endowed with the horocycle topology as in Example 3.1.

Proposition 4.13. $\bar{\iota}$ is a continuous embedding.

Proof. $\bar{\iota}$ is continuous. Let s be a cusp of $\overline{\Gamma}_t^*$, i.e. s is a fixed point of some parabolic element $\tilde{A} \in \overline{\Gamma}_t^*$, and $c: [0, \infty) \rightarrow \mathbb{H}$ an arbitrary path in \mathbb{H} converging to s in the horocycle topology.

By Remark 2.20, the action of \tilde{A} on \mathbb{H} fits together with the action of $\rho(A) \in \Gamma_g$ on $\overline{\mathcal{T}}_g$. Both actions may be extended continuously to $\mathbb{H}_s = \mathbb{H} \cup \{s\}$ (endowed with the horocycle topology) and to $\overline{\mathcal{T}}_g$, respectively, and one obtains the following

commutative diagram:

$$\begin{array}{ccc}
 \mathbb{H}_s = \mathbb{H} \cup \{s\} & \xrightarrow{\bar{\iota}} & \overline{\mathcal{T}}_g \\
 p_{\tilde{A}} \downarrow & & \downarrow p \\
 \overline{\mathbb{H}_s / \langle \tilde{A} \rangle} & \xrightarrow{i_{\tilde{A}}} & \overline{M}_g
 \end{array}$$

Here the map $i_{\tilde{A}}: \overline{\mathbb{H}_s / \langle \tilde{A} \rangle} \rightarrow \overline{M}_g$ is the map induced by $\bar{\iota}$ and $\overline{\mathbb{H}_s / \langle \tilde{A} \rangle}$ is a disk with center $p_{\tilde{A}}(s)$.

Let W be a neighbourhood of

$$\bar{P}_\infty = i_{\tilde{A}}(p_{\tilde{A}}(s)) = p(\bar{\iota}(s)).$$

For i in an index set I , let P_∞^i be the preimages of \bar{P}_∞ in $\overline{\mathcal{T}}_g$ under p . One of them is $\bar{\iota}(s)$, again by the commutativity of the diagram.

Since $\{P_\infty^i : i \in I\}$ is discrete we may choose the neighbourhood W in such a manner that its preimage under p is of the form:

$$V = p^{-1}(W) = \bigcup_{i \in I} V_i \subseteq \overline{\mathcal{T}}_g$$

where the V_i are the connected components of V with $P_\infty^i \in V_i$ and V_i is invariant under the stabilizer of P_∞^i in the mapping class group Γ_g .

Furthermore, we may choose W such that the preimage of W under $i_{\tilde{A}}$ is a simply connected neighbourhood of $p_{\tilde{A}}(s)$. Then, again, the preimage

$$U = p_{\tilde{A}}^{-1}(i_{\tilde{A}}^{-1}(W))$$

is a neighbourhood of s in the horocycle topology.

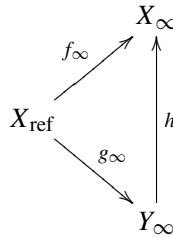
Thus an end piece of the path c is completely contained in U , i.e. there is some $l \in \mathbb{R}_{>0}$ such that $c([l, \infty))$ is contained in U .

Since the above diagram is commutative and the V_i are disjoint, the image of U is one of the V_i . This V_i then contains $\bar{\iota}(c([l, \infty))$. In addition, V_i has to contain the end piece of the Jenkins–Strebel ray that leads to s used to define $\bar{\iota}(s)$. Hence, V_i is the component that contains $\bar{\iota}(s)$.

Making W arbitrarily small, the neighbourhood U of s becomes arbitrarily small. Thus $\iota \circ c$ converges to $\bar{\iota}(s)$.

$\bar{\iota}$ is injective. Suppose there are two cusps s_1 and s_2 with $P_\infty = \bar{\iota}(s_1) = \bar{\iota}(s_2)$. Thus we have two Jenkins–Strebel rays defined by the negative of the Jenkins–Strebel differentials $q_1 = e^{i\theta_1} \cdot q$ and $q_2 = e^{i\theta_2} \cdot q$ with initial point $P_0 = \iota(i)$ and the same end point P_∞ in $\overline{\mathcal{T}}_g$. Let (X_∞, f_∞) and (Y_∞, g_∞) be the two marked stable Riemann surfaces defined by the two Jenkins–Strebel rays, respectively. Since they define the same point in $\overline{\mathcal{T}}_g$ the following diagram is commutative up to homotopy with some

biholomorphic h :



The core curves of the cylinders relative to the flat structure on X defined by q_1 are mapped by f_∞ to the singular points of X_∞ . Similarly the core curves coming from q_2 are mapped to the singular points of Y_∞ . Since the diagram is commutative up to isotopy, the two systems of core curves are homotopic. Thus the two Jenkins–Strebel rays are similar by definition, using the terminology in [26, Section 5]. From Theorem 2 in [26] it follows that there is some constant $M < \infty$ such that for two points $Q \neq R$ lying on the two Jenkins–Strebel rays which are equidistant from the initial point P_0 , one has $d(Q, R) \leq M$. But then, since ι is an isometric embedding, M would have to be an upper bound for the distance of equidistant points on two different geodesic rays in \mathbb{H} starting from i . This cannot be true. \square

4.2.2 Boundary of Teichmüller disks that lead to Teichmüller curves. Let now $\iota: \mathbb{H} \hookrightarrow \mathcal{T}_g$ be a Teichmüller embedding such that its image Δ_ι projects to a Teichmüller curve C in the moduli space M_g .

Proposition 4.14. *In this situation, the extended embedding from 4.2.1*

$$\bar{\iota}: \mathbb{H} \cup \{\text{cusps of } \bar{\Gamma}_\iota^*\} \hookrightarrow \bar{\Delta}_\iota \subseteq \bar{\mathcal{T}}_g$$

is surjective onto the closure $\bar{\Delta}_\iota$ of Δ_ι in $\bar{\mathcal{T}}_g$.

Proof. Recall from Corollary 2.21 that if ι leads to a Teichmüller curve then the projective Veech group $\bar{\Gamma} = \bar{\Gamma}_\iota$ is a lattice in $\text{PSL}_2(\mathbb{R})$, $\mathbb{H}/\bar{\Gamma}^*$ is a complex algebraic curve and $\mathbb{H}/\bar{\Gamma}^* \rightarrow C \subset M_g$ is the normalization of C . Thus it extends to a surjective morphism

$$\varphi: \overline{\mathbb{H}/\bar{\Gamma}^*} \rightarrow \bar{C} \subseteq \bar{M}_g,$$

where $\overline{\mathbb{H}/\bar{\Gamma}^*}$ and \bar{C} are the projective closure of $\mathbb{H}/\bar{\Gamma}^*$ and the closure of C in \bar{M}_g , respectively.

Furthermore, the map $\mathbb{H} \rightarrow \mathbb{H}/\bar{\Gamma}^*$ extends continuously to a surjective map $p_{\bar{\Gamma}}: \mathbb{H} \cup \{\text{cusps of } \bar{\Gamma}^*\} \rightarrow \overline{\mathbb{H}/\bar{\Gamma}^*}$, since $\bar{\Gamma}^*$ is a lattice in $\text{PSL}_2(\mathbb{R})$. Here we use the horocycle topology on $\mathbb{H} \cup \{\text{cusps of } \bar{\Gamma}^*\}$.

Thus one has the following commutative diagram of continuous maps:

$$\begin{array}{ccc}
 \mathbb{H} \cup \{\text{cusps of } \overline{\Gamma^*}\} & \xrightarrow{\bar{\iota}} & \overline{\Delta}_t \subseteq \overline{\mathcal{T}}_g \\
 p_{\overline{\Gamma}} \downarrow & & p|_{\overline{\Delta}_t} \downarrow \quad \downarrow p \\
 \overline{\mathbb{H}/\overline{\Gamma^*}} & \xrightarrow{\varphi} & \overline{C} \subseteq \overline{M}_g
 \end{array}$$

Let now P_∞ be a point on the boundary of Δ_t . Similarly as in the proof of the continuity of $\bar{\iota}$ we may choose a neighbourhood W of $p(P_\infty)$ in \overline{C} such that all connected components V_i of the preimage $p^{-1}(W)$ contain only one preimage of $p(P_\infty)$. One of them, let's say V_0 , contains of course P_∞ itself.

We choose an arbitrary path $c_t: [0, \infty) \rightarrow W \setminus \{p(P_\infty)\} \subseteq C$ that converges to $p(P_\infty)$. Let $\hat{c}_t: [0, \infty) \rightarrow V_0$ be an arbitrary lift of c_t via p in V_0 . Since we may choose W arbitrarily small, V_0 may become arbitrarily small and \hat{c}_t converges to P_∞ .

Now let $c: [0, \infty) \rightarrow \mathbb{H}$ be the preimage of \hat{c}_t under ι , i.e. the path such that $\iota \circ c = \hat{c}_t$. We project it by $p_{\overline{\Gamma}}$ to $\overline{\mathbb{H}/\overline{\Gamma^*}}$, i.e. we take the path $p_{\overline{\Gamma}} \circ c$. Its image under φ is $\varphi \circ p_{\overline{\Gamma}} \circ c = p \circ \hat{c}_t = c_t$ and converges to $p(P_\infty)$ in \overline{C} . Thus $p_{\overline{\Gamma}} \circ c$ converges in $\overline{\mathbb{H}/\overline{\Gamma^*}}$, since φ is an open map.

Since also $p_{\overline{\Gamma}}$ is open, c converges to some $t_\infty \in \mathbb{H} \cup \{\text{cusps of } \overline{\Gamma^*}\}$. By continuity of $\bar{\iota}$ one has $\bar{\iota}(t_\infty) = P_\infty$. Thus $\bar{\iota}$ is surjective onto $\overline{\Delta}_t$. □

One obtains immediately the following conclusions.

Corollary 4.15. *If $\iota: \mathbb{H} \hookrightarrow \mathcal{T}_g$ leads to a Teichmüller curve C , then*

- a) *the boundary points of the Teichmüller disk Δ_t are precisely the end points of the Jenkins–Strebel rays in Δ_t with initial point $\iota(i)$.*
- b) *These boundary points correspond to the fixed points of parabolic elements in the projective Veech group.*
- c) *Each boundary point of the Teichmüller curve C is obtained by contracting the core curves of the cylinders in the direction of v , where v is the eigenvector of a parabolic element in the Veech group.*

This finishes the proof of Theorem 4.1.

5 Schottky spaces

In this chapter we first recall the construction of Schottky coverings for smooth and stable Riemann surfaces. We use them to define markings called Schottky structures. In the smooth case they are classified by the well known Schottky space S_g , a complex

manifold of dimension $3g - 3$ (if $g \geq 2$). In [13] it was shown that also the Schottky structures on stable Riemann surfaces are parameterized by a complex manifold \bar{S}_g . Here we show how to obtain \bar{S}_g from Braungardt's extension $\bar{\mathcal{T}}_g$ of the Teichmüller space introduced in Section 3. In the last section of this chapter we study the image of a Teichmüller disk in the Schottky space.

5.1 Schottky coverings

We recall the basic definitions and properties of Schottky uniformization of Riemann surfaces. We introduce the Schottky space S_g and sketch, following [13], the construction of a universal family over it.

Definition 5.1. A group $\Gamma \subset \mathrm{PSL}_2(\mathbb{C})$ of Möbius transformations on $\mathbb{P}^1(\mathbb{C})$ is called a *Schottky group* if there are, for some $g \geq 1$, disjoint closed simply connected domains $D_1, D'_1, \dots, D_g, D'_g$ bounded by Jordan curves $C_i = \partial D_i, C'_i = \partial D'_i$, and generators $\gamma_1, \dots, \gamma_g$ of Γ such that $\gamma_i(C_i) = C'_i$ and $\gamma_i(D_i) = \mathbb{P}^1(\mathbb{C}) - \bar{D}'_i$ for $i = 1, \dots, g$. The generators $\gamma_1, \dots, \gamma_g$ are called a *Schottky basis* of Γ .

In Schottky's original paper [32], the D_i in the definition were disks. With the same notation let

$$F = F(\Gamma) = \mathbb{P}^1(\mathbb{C}) - \bigcup_{i=1}^g (\bar{D}_i \cup \bar{D}'_i) \quad \text{and} \quad \Omega = \Omega(\Gamma) = \bigcup_{\gamma \in \Gamma} \gamma(F).$$

It is well known, see e.g. [25, X.H.] that Γ is a Kleinian group, free of rank g with free generators $\gamma_1, \dots, \gamma_g$, that Ω is the region of discontinuity of Γ , and that $X = \Omega/\Gamma$ is a closed Riemann surface of genus g . The quotient map $\Omega \rightarrow X$ is called a *Schottky covering*.

An important fact is the following uniformization theorem:

Proposition 5.2. *Every compact Riemann surface X of genus $g \geq 1$ admits a Schottky covering by a Schottky group of rank g .*

Proof. The proof is based on the following construction that we shall extend to stable Riemann surfaces in Section 5.4: choose disjoint simple loops c_1, \dots, c_g on X which are independent in homology, i.e. $F = X - \bigcup_{i=1}^g c_i$ is connected. Then F is conformally equivalent to a plane domain that is bounded by $2g$ closed Jordan curves. For $i = 1, \dots, g$ denote by C_i and C'_i the two boundary components of F that result from cutting along c_i . Now let Φ_g be a free group on generators $\varphi_1, \dots, \varphi_g$, and take a copy F_w of F for every element $w \in \Phi_g$. The F_w are glued according to the following rule: if w and w' are reduced words in $\varphi_1, \dots, \varphi_g$ and if $w = w'\varphi_i$ then the boundary component C_i on $F_{w'}$ is glued to C'_i on F_w ; if w ends with φ_i^{-1} the roles of C_i and C'_i are interchanged. By this construction we obtain a plane domain Ω together with a holomorphic action of Φ_g on it: an element $\varphi \in \Phi_g$ maps the copy F_w to $F_{w\varphi}$. The

crucial step in the proof now is to show that this action extends to all of $\mathbb{P}^1(\mathbb{C})$, i.e. Φ_g acts by Möbius transformations. For this we refer to [3, Chapter IV, Theorem 19 F]. \square

Definition 5.3. Let \tilde{S}_g be the set of all $(\gamma_1, \dots, \gamma_g) \in \mathrm{PSL}_2(\mathbb{C})^g$ that generate a Schottky group Γ and form a Schottky basis for Γ . The set S_g of equivalence classes of g -tuples $(\gamma_1, \dots, \gamma_g) \in \tilde{S}_g$ under simultaneous conjugation is called the *Schottky space* of genus g .

For a point $s = (\gamma_1, \dots, \gamma_g) \in \tilde{S}_g$ let $\Gamma(s)$ be the Schottky group generated by $\gamma_1, \dots, \gamma_g$, $\Omega(s)$ the region of discontinuity of $\Gamma(s)$, and $X(s) = \Omega(s)/\Gamma(s)$ the associated Riemann surface. This leads to an alternative description of the Schottky space:

Remark 5.4. S_g is the set of equivalence classes of pairs (X, σ) , where X is a Riemann surface of genus g and $\sigma: \Phi_g \rightarrow \mathrm{PSL}_2(\mathbb{C})$ is an injective homomorphism such that $\Gamma := \sigma(\Phi_g)$ is a Schottky group and $\Omega(\Gamma)/\Gamma \cong X$.

(X, σ) and (X', σ') are equivalent if there is some $A \in \mathrm{PSL}_2(\mathbb{C})$ such that $\sigma'(\gamma) = A\sigma(\gamma)A^{-1}$ for all $\gamma \in \Phi_g$. Note that then X' is isomorphic to X .

To endow S_g with a complex structure we proceed as follows: Taking the fixed points and the multipliers of the γ_i we obtain an embedding of \tilde{S}_g as an open subdomain of $\mathbb{P}^1(\mathbb{C})^{3g}$. For $g = 1$ each equivalence class contains a unique Möbius transformation of the form $z \mapsto \lambda z$ for some $\lambda \in \mathbb{C}, 0 < |\lambda| < 1$. If $g \geq 2$ we find in each equivalence class in \tilde{S}_g a unique representative $(\gamma_1, \dots, \gamma_g)$ such that γ_1 and γ_2 have attracting fixed points 0 and 1, respectively, and γ_1 has repelling fixed point ∞ . This defines a section to the projection $\tilde{S}_g \rightarrow S_g$ and embeds S_g as a closed subspace of \tilde{S}_g which, moreover, lies in $\{0\} \times \{\infty\} \times \{1\} \times \mathbb{C}^{3g-3} \subseteq \mathbb{P}^1(\mathbb{C})^{3g}$. Thus we have shown

Proposition 5.5. a) S_1 is a punctured disk.

b) For $g \geq 2$, S_g carries a complex structure as an open subdomain of \mathbb{C}^{3g-3} .

Our next goal is to show that this complex structure on S_g is natural. The main step in this direction is

Proposition 5.6. The forgetful map $\mu: S_g \rightarrow M_g$, that sends $s = (X, \sigma)$ to the isomorphism class of X , is analytic and surjective.

Proof. The surjectivity of μ follows from Proposition 5.2. To show that μ is analytic we use the fact that M_g is a coarse moduli space for Riemann surfaces. Therefore it suffices to find a holomorphic family $\pi: \mathcal{C}_g \rightarrow S_g$ of Riemann surfaces over S_g which induces μ in the sense that for $s \in S_g$, $\mu(s)$ is the isomorphism class of the fibre $C_s = \pi^{-1}(s) \subset \mathcal{C}_g$.

The family \mathcal{C}_g is obtained as in Section 3.3: Let

$$\Omega_g = \{(s, z) \in S_g \times \mathbb{P}^1(\mathbb{C}) : z \in \Omega(s)\}.$$

Ω_g is a complex manifold on which the free group Φ_g acts holomorphically by $\varphi(s, z) = (s, \sigma(\varphi)(z))$ for $s = (X, \sigma) \in S_g$, $\varphi \in \Phi_g$ and $z \in \Omega(s)$.

The projection $\text{pr}_1 : \Omega_g \rightarrow S_g$ onto the first component factors through the orbit space $\mathcal{C}_g = \Omega_g / \Phi_g$, and the induced map $\pi : \mathcal{C}_g \rightarrow S_g$ is the family of Riemann surfaces we were looking for. \square

The family \mathcal{C}_g is in fact universal for Riemann surfaces with *Schottky structure*, a kind of marking that we now recall from [13, Section 1.3]:

Definition 5.7. a) Let $\mathcal{U} \rightarrow S$ be an analytic map of complex manifolds and $\Gamma \subset \text{Aut}(\mathcal{U}/S)$ a properly discontinuous subgroup. Then the analytic quotient map $\mathcal{U} \rightarrow \mathcal{U}/\Gamma = \mathcal{C}$ is called a *Schottky covering* if the induced map $\mathcal{C} \rightarrow S$ is a family of Riemann surfaces and if for every $x \in S$ the restriction $U_x \rightarrow C_x$ of the quotient map to the fibres is a Schottky covering.

b) A *Schottky structure* is a Schottky covering $\mathcal{U} \rightarrow \mathcal{U}/\Gamma = \mathcal{C}$ together with an equivalence class of isomorphisms $\sigma : \Phi_g \rightarrow \Gamma$, where σ and σ' are considered equivalent if they differ only by an inner automorphism of Φ_g .

Note that the construction in the proof of Proposition 5.6 endows the family \mathcal{C}_g/S_g with a Schottky structure.

A Schottky structure on a single Riemann surface X is given by a Schottky covering $\Omega \rightarrow \Omega/\Gamma = X$ and an isomorphism $\sigma : \Phi_g \rightarrow \Gamma$. Comparing the respective equivalence relations we find that the points (X, σ) in S_g correspond bijectively to the isomorphism classes of Riemann surfaces with Schottky structure. In fact a much stronger result holds:

Theorem 5.8. S_g is a fine moduli space for Riemann surfaces with Schottky structure.

Proof. Assume that \mathcal{C}/S is a family of Riemann surfaces and let $(\mathcal{U} \rightarrow \mathcal{U}/\Gamma = \mathcal{C}, \sigma : \Phi_g \xrightarrow{\sim} \Gamma)$ be a Schottky structure on \mathcal{C} . Then we have a map $f : S \rightarrow S_g$ which maps a point x to the isomorphism class of the Schottky covering $U_x \rightarrow C_x$. We have to show that f is analytic. Then the other properties of a fine moduli space follow easily from the definitions, namely that \mathcal{C} is the fibre product $\mathcal{C}_g \times_{S_g} S$ and that \mathcal{U} is isomorphic to $\Omega_g \times_{\mathcal{C}_g} \mathcal{C} = \Omega_g \times_{S_g} S$ such that the projection $\mathcal{U} \rightarrow \Omega_g$ onto the first factor is equivariant for the actions of Γ and Φ_g via the isomorphism σ .

The universal property of M_g as a coarse moduli space gives us, as above for μ , that the composition $\mu \circ f$ is analytic. Since μ has discrete fibres, it therefore suffices to show that f is continuous. This is quite subtle, see [13, § 3]. \square

5.2 Relation to Teichmüller space

In this section we explain that Schottky space can be obtained as a quotient space of the Teichmüller space which was introduced in Section 3.3. For this purpose we first endow the universal family $\mathcal{C}_{g,0}$ over the Teichmüller space $\mathcal{T}_g = \mathcal{T}_{g,0}$ with a Schottky structure as follows:

Let $a_1, b_1, \dots, a_g, b_g$ be a set of standard generators of π_g , the fundamental group of the reference surface X_{ref} ; this means that they satisfy the relation

$$\prod_{i=1}^g a_i b_i a_i^{-1} b_i^{-1} = 1.$$

Then b_1, \dots, b_g are homologically independent, hence the construction in the proof of Proposition 5.2 provides us with a corresponding Schottky covering $\Omega_{\text{ref}} \rightarrow X_{\text{ref}}$. The group $\text{Aut}(\Omega_{\text{ref}}/X_{\text{ref}})$ of deck transformations is isomorphic to the free group on b_1, \dots, b_g . Denoting $U_{\text{ref}} \rightarrow X_{\text{ref}}$ the universal covering, there is a covering map $U_{\text{ref}} \rightarrow \Omega_{\text{ref}}$ over X_{ref} . The group $\text{Aut}(U_{\text{ref}}/\Omega_{\text{ref}})$ is the kernel N_α of the homomorphism $\alpha: \pi_g \rightarrow \Phi_g$ which maps b_i to φ_i and a_i to 1; in other words, N_α is the normal closure in π_g of the subgroup generated by a_1, \dots, a_g .

In Section 3.3 we described the family $\Omega_{g,0}^+ \rightarrow \mathcal{C}_{g,0}$ of universal coverings of the surfaces in the family $\mathcal{C}_{g,0}$; the fundamental group π_g and hence also N_α acts on the fibres of this covering, and we obtain:

Remark 5.9. The induced map $\Omega_{g,0}^+/N_\alpha \rightarrow \mathcal{C}_{g,0}$ is a Schottky covering, and the universal Teichmüller structure $\tau: \pi_g \xrightarrow{\sim} \text{Aut}(\Omega_{g,0}^+/\mathcal{C}_{g,0})$ (cf. Theorem 3.13) descends via α to a Schottky structure $\sigma_\alpha: \Phi_g = \pi_g/N_\alpha \xrightarrow{\sim} \text{Aut}((\Omega_{g,0}^+/N_\alpha)/\mathcal{C}_{g,0})$ on $\mathcal{C}_{g,0}$.

By Theorem 5.8 this Schottky structure induces an analytic map $s_\alpha: \mathcal{T}_g \rightarrow S_g$. To describe s_α as the quotient map for a subgroup of the mapping class group Γ_g , we first identify Γ_g with the group $\text{Out}^+(\pi_g)$ of orientation preserving outer automorphisms of π_g ; then, to a diffeomorphism $f: X_{\text{ref}} \rightarrow X_{\text{ref}}$, we associate the induced automorphism $\varphi = f_*: \pi_g \rightarrow \pi_g$. It follows from the Dehn–Nielsen theorem that this gives an isomorphism $\Gamma_g \xrightarrow{\sim} \text{Out}^+(\pi_g)$. In this chapter, by $\varphi \in \Gamma_g$ we always mean an element of $\text{Out}^+(\pi_g)$.

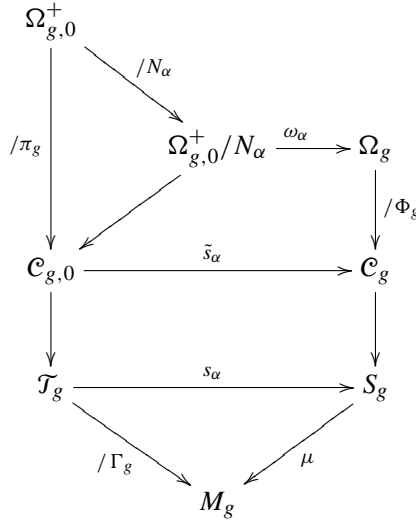
Proposition 5.10. a) s_α is the quotient map for the subgroup

$$\Gamma_g(\alpha) = \{\varphi \in \Gamma_g : \alpha \circ \varphi \equiv \alpha \text{ mod } \text{Inn}(\pi_g)\}$$

of the mapping class group Γ_g (where $\text{Inn}(\pi_g)$ denotes the group of inner automorphisms).

b) $s_\alpha: \mathcal{T}_g \rightarrow S_g$ is the universal covering of the Schottky space.

c) s_α lifts to maps \tilde{s}_α and ω_α that make the following diagram commutative:



Proof. a) Let $x = (X, f) \in \mathcal{T}_g$. Recall from Section 3.3 that the fibre over x in $\Omega_{g,0}^+$ is the component $\Omega^+(x)$ of the region of discontinuity of the quasi-Fuchsian group G_x associated to x . The universal Teichmüller structure on $\mathcal{C}_{g,0}$ induces an isomorphism $\tau_x: \pi_g \rightarrow G_x = \text{Aut}(\Omega^+(x)/X)$. From Remark 5.9 we see that the point $s_\alpha(x) = (X, \sigma) \in S_g$ is given by the restriction $\sigma_{\alpha,x}$ of σ_α to the fibre over x ; explicitly,

$$\sigma = \sigma_{\alpha,x}: \Phi_g = \pi_g/N_\alpha \xrightarrow{\sim} \text{Aut}((\Omega^+(x)/\tau_x(N_\alpha))/X) = G_x/\tau_x(N_\alpha).$$

For $\varphi \in \Gamma_g$ we have $s_\alpha(x) = s_\alpha(\varphi(x))$ if and only if $\sigma_{\alpha,x} = \sigma_{\alpha,\varphi(x)}$ up to an inner automorphism. Since $\tau_{\varphi(x)} = \tau_x \circ \varphi^{-1}$ this happens if and only if φ induces an inner automorphism on π_g/N_α , i.e. if and only if $\varphi \in \Gamma_g(\alpha)$.

b) This is clear from the fact that \mathcal{T}_g is simply connected and $\Gamma_g(\alpha)$ is torsion free, hence s_α is unramified. Using the construction in a) one can give a direct proof which in turn provides an independent proof that \mathcal{T}_g is simply connected, see [13, Proposition 6].

c) It follows from Remark 5.9 that $\Omega_{g,0}^+/N_\alpha \rightarrow \mathcal{C}_{g,0}$ is a Schottky covering. Therefore, by the universal property of S_g (Theorem 5.8), $\mathcal{C}_{g,0}$ is the fibre product $\mathcal{T}_g \times_{S_g} \mathcal{C}_g$, and \tilde{s}_α is the projection to \mathcal{C}_g . Moreover the Schottky covering $\Omega_{g,0}^+/N_\alpha \rightarrow \mathcal{C}_{g,0}$ is a pullback of the universal Schottky covering $\Omega_g \rightarrow \mathcal{C}_g$, i.e. $\Omega_{g,0}^+/N_\alpha = \mathcal{C}_{g,0} \times_{\mathcal{C}_g} \Omega_g$, and again ω_α is the projection to the second factor. \square

In fact, the action of $\Gamma_g(\alpha)$ on \mathcal{T}_g extends to $\Omega_{g,0}^+$, $\Omega_{g,0}^+/N_\alpha$ and $\mathcal{C}_{g,0}$; then \tilde{s}_α and ω_α are the quotient maps for these actions.

5.3 Schottky coverings of stable Riemann surfaces

In this and the following section we introduce a partial compactification \bar{S}_g of S_g that fits in between $\bar{\mathcal{T}}_g$ and \bar{M}_g . We have presented two different ways to define S_g , and we shall see that both are suited for extension to stable Riemann surfaces: The first way is to construct Schottky coverings for surfaces with nodes, define Schottky structures and find parameters for them. This approach was pursued in [13] and will be sketched in this section. The other possibility is to extend the action of $\Gamma_g(\alpha)$ to (part of) the boundary of $\bar{\mathcal{T}}_g$ and show that the quotient exists and has the desired properties; this will be done in Section 5.4.

Definition 5.11. Let X be a stable Riemann surface of genus g . A *cut system* on X is a collection of disjoint simple loops c_1, \dots, c_g on X , not passing through any of the nodes, such that $X - \bigcup_{i=1}^g c_i$ is connected.

Proposition 5.12. *On any stable Riemann surface there exist cut systems.*

Proof. Let $f: X_{\text{ref}} \rightarrow X$ be a deformation; we must find disjoint and homologically independent loops $\tilde{c}_1, \dots, \tilde{c}_g$ on X_{ref} that are disjoint from the loops a_1, \dots, a_k that are contracted by f . For this we complete a_1, \dots, a_k to a maximal system a_1, \dots, a_{3g-3} of homotopically independent disjoint loops (such a system decomposes X_{ref} into pairs of pants). Among the a_i we find a_{i_1}, \dots, a_{i_g} that are homologically independent. If $i_\nu > k$ we take $\tilde{c}_\nu = a_{i_\nu}$, and for $i_\nu \leq k$ we replace a_{i_ν} by a loop \tilde{c}_ν that is homotopic to a_{i_ν} and disjoint from it. \square

Once we have found c_1, \dots, c_g as above, we proceed as in the proof of Proposition 5.2 to construct a Schottky covering of X : Let $F = X - \bigcup_{i=1}^g c_i$, take a copy F_w of F for each $w \in \Phi_g$, and glue these copies exactly as before to obtain a space Ω . Of course, neither F nor Ω is planar whenever X has nodes. In all cases, the complex structure on X lifts to a structure of a one-dimensional complex space on Ω . The group Φ_g acts on this space by holomorphic automorphisms, i.e. there is an isomorphism $\Phi_g \rightarrow \Gamma = \text{Aut}(\Omega/X)$, and X is isomorphic to Ω/Γ as complex space.

Definition 5.13. The covering $\Omega \rightarrow X$ constructed above for a cut system $c = (c_1, \dots, c_g)$ on a stable Riemann surface X is called the *Schottky covering of X relative to c* . A covering of X is called a *Schottky covering* if it is the Schottky covering relative to some cut system.

The next goal is to define a space \bar{S}_g that classifies Schottky coverings in a way analogous to Definition 5.3. Since the covering space Ω is in general not a subspace of $\mathbb{P}^1(\mathbb{C})$ and thus the group of deck transformations not a subgroup of $\text{PSL}_2(\mathbb{C})$, we cannot directly extend 5.3.

A closer look at the construction of a Schottky covering $\Omega \rightarrow \Omega/\Gamma = X$ of a stable Riemann surface X shows the following:

Each irreducible component L of Ω is an open dense subset of a projective line; more precisely, the stabilizer of L in Γ is a Schottky group as in Definition 5.1, and L is its region of discontinuity. Moreover the intersection graph of the irreducible components of Ω is a tree (hence Ω is called a *tree of projective lines*).

Therefore, for each irreducible component L , there is a well defined projection $\pi_L: \Omega \rightarrow L$ which is the identity on L : For an arbitrary point $x \in \Omega$ there is a unique chain $L_0, L_1, \dots, L_n = L$ of mutually distinct components such that $x \in L_0$ and L_i intersects L_{i+1} for $i = 0, \dots, n-1$; then define $\pi_L(x)$ to be the intersection point of L_{n-1} and L .

An *end* of Ω is an equivalence class of infinite chains L_0, L_1, L_2, \dots of irreducible components as above (i.e. $L_i \neq L_j$ for $i \neq j$ and $L_i \cap L_{i+1} \neq \emptyset$), where two chains are equivalent if they differ only by finitely many components. Let $\Omega^* = \Omega \cup \{\text{ends of } \Omega\}$. Clearly the projection π_L to a component L can be extended to Ω^* .

For any three different points or ends y_1, y_2, y_3 in Ω^* there is a unique component $L = L(y_1, y_2, y_3)$ (called the *median* of the three points) such that the points $\pi_L(y_1), \pi_L(y_2), \pi_L(y_3)$ are distinct. Now given any four distinct points or ends y_1, \dots, y_4 in Ω^* we can define a *cross ratio* $\lambda(y_1, \dots, y_4)$ by taking the usual cross ratio of $\pi_L(y_1), \dots, \pi_L(y_4)$ on the median component $L = L(y_1, y_2, y_3)$ of the first three of them; note that $\lambda(y_1, \dots, y_4)$ will be 0, 1 or ∞ if $\pi_L(y_4)$ coincides with $\pi_L(y_1), \pi_L(y_2)$ or $\pi_L(y_3)$.

To obtain parameters for the group Γ observe that any $\gamma \in \Gamma$, $\gamma \neq 1$, has exactly two fixed points on the boundary of Ω , where boundary points of Ω are either points in the closure of a component, or ends of Ω ; one of the fixed points is attracting, the other repelling. For any four different (primitive) elements $\gamma_1, \dots, \gamma_4$ in Γ we define $\lambda(\gamma_1, \dots, \gamma_4)$ to be the cross ratio of their attracting fixed points. It is a remarkable fact that from these cross ratios both the space Ω and the group $\Gamma \subset \text{Aut}(\Omega)$ can be recovered. For any particular Schottky covering finitely many of them suffice, but for different Schottky coverings we must take, in general, the cross ratios of different elements of Φ_g . To parameterize all Schottky coverings we therefore have to use infinitely many of these cross ratios. We consider them as (projective) coordinates on an infinite product of projective lines $\mathbb{P}^1(\mathbb{C})$. The cross ratios satisfy a lot of algebraic relations, which define a closed subset B of this huge space. Every point of B represents a tree of projective lines Ω as above together with an action of Γ on it. \bar{S}_g is the open subset of B , where this action defines a Schottky covering. For details and in particular the technical complication caused by the presence of infinitely many variables and equations, see [13, §2] and [17]. In principle, one can proceed as in Section 5.1 to construct a family of stable Riemann surfaces over \bar{S}_g .

Given a family \mathcal{C}/S of stable Riemann surfaces over a complex manifold S , we can define the notion of a *Schottky covering* $\mathcal{U}/S \rightarrow \mathcal{C}/S$ and of a *Schottky structure* on \mathcal{U} exactly as in Definition 5.7, except that now \mathcal{U} is not assumed to be a manifold, but only a complex space. It is shown in [13, §3] that the family over \bar{S}_g carries a universal Schottky structure:

Theorem 5.14. \bar{S}_g is a fine moduli space for stable Riemann surfaces with Schottky structure.

5.4 \bar{S}_g as quotient of $\bar{\mathcal{T}}_g$

It is not possible to extend the quotient map $s_\alpha: \mathcal{T}_g \rightarrow S_g$ constructed in Section 5.2 to the whole boundary of \mathcal{T}_g in $\bar{\mathcal{T}}_g$. Instead we shall, for each α , identify a part $\bar{\mathcal{T}}_g(\alpha)$ of $\bar{\mathcal{T}}_g$ to which the action of $\Gamma_g(\alpha)$ and hence the morphism s_α can be extended. It will turn out that the quotient space is the extended Schottky space \bar{S}_g described in the previous section.

We begin with the definition of the admissible group homomorphisms α and the associated parts $\bar{\mathcal{T}}_g(\alpha)$ of $\bar{\mathcal{T}}_g$.

Definition 5.15. a) A surjective homomorphism $\alpha: \pi_g \rightarrow \Phi_g$ is called *symplectic* if there are standard generators $a_1, b_1, \dots, a_g, b_g$ of π_g (in the sense of Section 5.2) such that $\alpha(a_i) = 1$ for $i = 1, \dots, g$.

b) Recall from Section 3 that a point in $\bar{\mathcal{T}}_g$ can be described as an equivalence class of pairs (X, f) , where X is a stable Riemann surface and $f: X_{\text{ref}} \rightarrow X$ is a deformation (see Corollary 3.9).

For a symplectic homomorphism $\alpha: \pi_g \rightarrow \Phi_g$ let

$$\bar{\mathcal{T}}_g(\alpha) = \{(X, f) \in \bar{\mathcal{T}}_g : \ker(\pi_1(f)) \subseteq \ker(\alpha)\}.$$

Proposition 5.16. a) $\bar{\mathcal{T}}_g(\alpha)$ is an open subset of $\bar{\mathcal{T}}_g$; it contains \mathcal{T}_g and is invariant under the group $\Gamma_g(\alpha)$ introduced in Proposition 5.10.

b) $\bar{\mathcal{T}}_g$ is the union of the $\bar{\mathcal{T}}_g(\alpha)$, where α runs through the symplectic homomorphisms.

c) The restriction to $\bar{\mathcal{T}}_g(\alpha)$ of the universal covering $p: \bar{\mathcal{T}}_g \rightarrow \bar{M}_g$ is surjective for every symplectic α .

Proof. a) Let (X, f) be a point in $\bar{\mathcal{T}}_g$ and c_1, \dots, c_k the loops on X_{ref} that are contracted under f . Then the kernel of $\pi_1(f): \pi_g \rightarrow \pi_1(X)$ is the normal subgroup generated by c_1, \dots, c_k . The local description of $\bar{\mathcal{T}}_g$ in Corollary 3.8 shows that there is a neighbourhood U of (X, f) in $\bar{\mathcal{T}}_g$ such that for every $(X', f') \in U$ the map $f': X_{\text{ref}} \rightarrow X'$ contracts a subset of $\{c_1, \dots, c_k\}$. Hence the kernel of $\pi_1(f')$ is contained in $\ker(\pi_1(f))$. Thus if $(X, f) \in \bar{\mathcal{T}}_g(\alpha)$, also $U \subseteq \bar{\mathcal{T}}_g(\alpha)$. The remaining assertions are clear.

b) Again let (X, f) be a point in $\bar{\mathcal{T}}_g$ and c_1, \dots, c_k the loops on X_{ref} contracted by f . By Proposition 5.12 we can find a cut system a_1, \dots, a_g on X and a corresponding Schottky covering. This covering induces a surjective homomorphism $\pi_1(X) \rightarrow \Phi_g$. Composing this homomorphism with $\pi_1(f)$ yields a homomorphism $\alpha: \pi_g \rightarrow \Phi_g$ which corresponds to a Schottky covering of X_{ref} (relative to the cut

system $f^{-1}(a_1), \dots, f^{-1}(a_g)$ and hence is symplectic. By construction, c_1, \dots, c_k are in the kernel of α .

c) Let $\alpha: \pi_g \rightarrow \Phi_g$ be symplectic and $a_1, b_1, \dots, a_g, b_g$ standard generators of π_g such that $\alpha(a_i) = 1$ for all i . For an arbitrary stable Riemann surface X choose a deformation $f: X_{\text{ref}} \rightarrow X$ and let c_1, \dots, c_k be the loops that are contracted by f . As in the proof of b) we find standard generators $a'_1, b'_1, \dots, a'_g, b'_g$ such that the c_j are contained in the normal subgroup generated by the a'_i . The map $a_i \mapsto a'_i, b_i \mapsto b'_i$ defines an automorphism φ of π_g and thus an element of Γ_g . Then by construction $(X, f \circ \varphi)$ lies in $\overline{\mathcal{T}}_g(\alpha)$ and $p(X, f \circ \varphi) = X$. \square

As a side remark we note that $\overline{\mathcal{T}}_g(\alpha)$ is not only invariant under $\Gamma_g(\alpha)$, but also under the larger “handlebody” group

$$H_g(\alpha) = \{\varphi \in \Gamma_g : \varphi(N_\alpha) = N_\alpha\}$$

(where $N_\alpha = \ker(\alpha)$ as in Section 5.2). Note that $H_g(\alpha)$ is the normalizer of $\Gamma_g(\alpha)$ in Γ_g , and that we have an exact sequence

$$1 \rightarrow \Gamma_g(\alpha) \rightarrow H_g(\alpha) \rightarrow \text{Out}(\Phi_g) \rightarrow 1.$$

The quotient space $\hat{S}_g = \overline{\mathcal{T}}_g/H_g(\alpha) = S_g/\text{Out}(\Phi_g)$ is a parameter space for Schottky groups of rank g (without any marking).

Proposition 5.17. *For any symplectic homomorphism $\alpha: \pi_g \rightarrow \Phi_g$, the quotient space $\overline{\mathcal{T}}_g(\alpha)/\Gamma_g(\alpha)$ is a complex manifold $\overline{S}_g(\alpha)$.*

Proof. This is a local statement which is clear for points $(X, f) \in \mathcal{T}_g$ since $\Gamma_g(\alpha)$ is torsion free. For an arbitrary $x = (X, f) \in \overline{\mathcal{T}}_g$ we saw in Section 3.2 that the Dehn twists τ_1, \dots, τ_k around the loops c_1, \dots, c_k that are contracted by f generate a finite index subgroup Γ_x^0 of the stabilizer Γ_x of x in Γ_g (the quotient being the finite group $\text{Aut}(X)$). Let α be a symplectic homomorphism with respect to standard generators $a_1, b_1, \dots, a_g, b_g$, and assume $(X, f) \in \overline{\mathcal{T}}_g(\alpha)$. Since the c_i are in the normal subgroup generated by a_1, \dots, a_g , they do not intersect any of the a_j and thus $\tau_i(a_j) = a_j$ for all i and j . This shows $\Gamma_x \subseteq \Gamma_g(\alpha)$.

Now choose a neighbourhood U of $x = (X, f)$ in $\overline{\mathcal{T}}_g(\alpha)$ which is precisely invariant under Γ_x . Then it follows, from Proposition 3.7 (and Definition 3.6), that U/Γ_x is a complex manifold. \square

For any two sets $a_1, b_1, \dots, a_g, b_g$ and $a'_1, b'_1, \dots, a'_g, b'_g$ of standard generators, $a_i \mapsto a'_i, b_i \mapsto b'_i$ defines an automorphism of π_g . Therefore for any two symplectic homomorphisms α and α' there is an automorphism $\psi \in \Gamma_g$ such that $\alpha = \alpha' \circ \psi$. Then clearly $N_\alpha = \psi(N_{\alpha'})$ and $\Gamma_g(\alpha') = \psi\Gamma_g(\alpha)\psi^{-1}$. This shows that, as an automorphism of $\overline{\mathcal{T}}_g$, ψ maps $\overline{\mathcal{T}}_g(\alpha)$ to $\overline{\mathcal{T}}_g(\alpha')$ and descends to an isomorphism $\overline{\psi}: \overline{S}_g(\alpha) \rightarrow \overline{S}_g(\alpha')$. We have shown:

Remark 5.18. The complex manifolds $\overline{S}_g(\alpha)$ are isomorphic for all symplectic homomorphisms α .

It remains to show that the $\overline{S}_g(\alpha)$ coincide with the fine moduli space \overline{S}_g of Section 5.3. This is achieved by showing that $\overline{S}_g(\alpha)$ satisfies the same universal property as \overline{S}_g :

Proposition 5.19. *For any symplectic α , $\overline{S}_g(\alpha)$ is a fine moduli space for stable Riemann surfaces with Schottky structure and hence isomorphic to \overline{S}_g .*

Proof. The idea of the proof is to endow the universal family over $\overline{\mathcal{T}}_g(\alpha)$ with a Schottky structure and to transfer this to a Schottky structure on the image family over $\overline{S}_g(\alpha)$.

Before explaining this for the whole family we consider a single stable Riemann surface X . Let d_1, \dots, d_k be the nodes on X , $f: X_{\text{ref}} \rightarrow X$ a deformation and $\alpha: \pi_g \rightarrow \Phi_g$ a symplectic homomorphism such that $x = (X, f) \in \overline{\mathcal{T}}_g(\alpha)$. In Section 3.3 we described the universal covering $\hat{\Omega}^+(x) \rightarrow \hat{\Omega}^+(x)/G_x = X$ of X with cusps over the nodes. Recall that $\hat{\Omega}^+(x)$ is the union of the plane region $\Omega^+(x)$ with the common boundary points of the doubly cusped regions lying over the nodes d_i , and that G_x is isomorphic to π_g .

Remark 5.20. Using the above notation, let $\rho: \pi_g \rightarrow G_x$ be an isomorphism and $N_\alpha^{G_x} = \ker(\alpha \circ \rho^{-1}) \subseteq G_x$. Then $\Omega = \hat{\Omega}^+(x)/N_\alpha^{G_x}$ is a complex space, $G_x/N_\alpha^{G_x} \cong \Phi_g$ acts holomorphically on Ω , and $\Omega \rightarrow \Omega/\Phi_g = X$ is a Schottky covering.

Proof. The key observation is that the stabilizer in G_x of a point $\tilde{d}_i \in \hat{\Omega}^+(x)$ lying over d_i is generated by an element γ_i corresponding under ρ to a conjugate of the loop $f^{-1}(d_i)$. Since we assumed $(X, f) \in \overline{\mathcal{T}}_g(\alpha)$, we have $\gamma_i \in N_\alpha^{G_x}$. This shows that Ω is a complex space, more precisely: a Riemann surface with nodes. The other assertions then follow directly from the definitions. \square

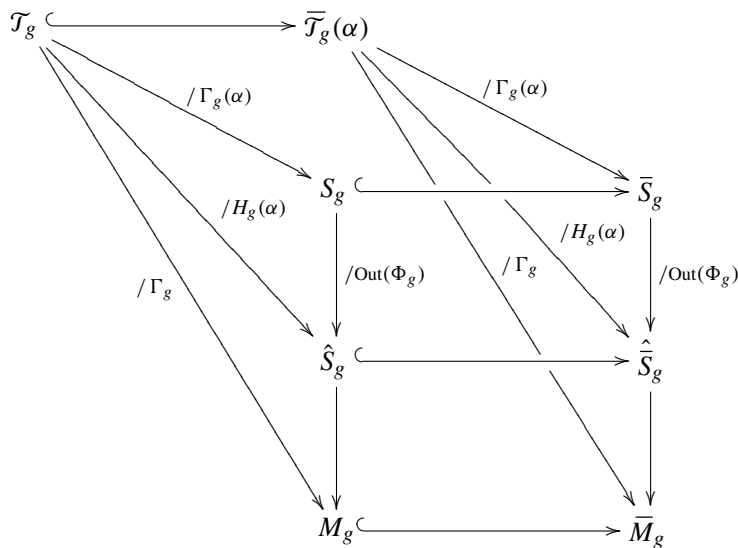
The above construction can be carried over to families in the following way: First consider the universal family $\overline{\mathcal{C}}_g$ over $\overline{\mathcal{T}}_g$ and the universal Teichmüller structure $\hat{\Omega}_g^+ \rightarrow \overline{\mathcal{C}}_g$ on it. Denote by $\overline{\mathcal{C}}_g(\alpha)$ resp. $\hat{\Omega}_g^+(\alpha)$ the restriction to $\overline{\mathcal{T}}_g(\alpha)$. Then the quotient space $\hat{\Omega}_g^+(\alpha)/N_\alpha$ is a complex space on which $\Phi_g = \pi_g/N_\alpha$ acts. The quotient map $\hat{\Omega}_g^+(\alpha)/N_\alpha \rightarrow \overline{\mathcal{C}}_g(\alpha)$ is a Schottky covering and the identification of Φ_g with the group of deck transformations defines a Schottky structure.

The group $\Gamma_g(\alpha)$ acts not only on $\overline{\mathcal{T}}_g(\alpha)$, but also on $\hat{\Omega}_g^+(\alpha)$ as follows: for $\varphi \in \Gamma_g(\alpha)$ and $(x, z) \in \hat{\Omega}_g^+(\alpha)$ with $x \in \overline{\mathcal{T}}_g(\alpha)$ and $z \in \hat{\Omega}^+(x)$ we set

$$\varphi(x, z) = (\varphi(x), z).$$

Note that the groups G_x and $G_{\varphi(x)}$ are the same (only the isomorphism with π_g has changed); therefore $\hat{\Omega}^+(x) = \hat{\Omega}^+(\varphi(x))$. This action, which is trivial on the fibres, descends to actions of $\Gamma_g(\alpha)$ on $\hat{\Omega}_g^+(\alpha)/N_\alpha$ and on $\bar{\mathcal{C}}_g(\alpha)$. The respective orbit spaces give a family $\bar{\mathcal{C}}_g = \bar{\mathcal{C}}_g(\alpha)/\Gamma_g(\alpha)$ over $\bar{S}_g(\alpha)$ and a Schottky structure on it. Using the universal property of the family over $\bar{\mathcal{T}}_g(\alpha)$ (see Theorem 3.13) and the fact that Schottky structures are locally induced by Teichmüller structures, we find that the Schottky structure on $\bar{\mathcal{C}}_g$ is in fact universal. \square

The following diagram collects the relations between the spaces introduced and used in this section. The horizontal maps are open embeddings, the last two vertical maps are analytic with discrete fibres; all other maps in the diagram are quotient maps for the groups indicated (to be precise, the map from $\bar{\mathcal{T}}_g(\alpha)$ to \bar{M}_g is the restriction of the orbit map for the action of Γ_g on $\bar{\mathcal{T}}_g$).



5.5 Teichmüller disks in Schottky space

Let $\iota: \mathbb{H} \rightarrow \mathcal{T}_g$ be a Teichmüller embedding as in Definition 2.4 and $\Delta = \iota(\mathbb{H})$ its image in \mathcal{T}_g . Let $\text{Stab}(\Delta)$ be the stabilizer of Δ in Γ_g . We have seen in Section 2.4.3 that $\text{Stab}(\Delta)$ maps surjectively to the projective Veech group $\bar{\Gamma}_\iota$ of ι (see Definition 2.17); the kernel of this map is the pointwise stabilizer of Δ .

In this section we assume that $\bar{\Gamma}_\iota$ is a lattice in $\text{PSL}_2(\mathbb{R})$, or equivalently that the image C_ι of Δ in M_g is a Teichmüller curve (cf. Corollary 2.21). As mentioned in the introduction, Veech showed that C_ι is not a projective curve and thus cannot be closed in \bar{M}_g .

Proposition 5.21. *Let $\iota: \mathbb{H} \rightarrow \mathcal{T}_g$ be a Teichmüller embedding such that $\bar{\Gamma}_l$ is a lattice in $\mathrm{PSL}_2(\mathbb{R})$. Then there exists a symplectic homomorphism $\alpha: \pi_g \rightarrow \Phi_g$ such that*

$$\mathrm{Stab}(\Delta) \cap \Gamma_g(\alpha) \neq \{1\}.$$

Since $\Gamma_g(\alpha)$ is torsion free, this implies that the intersection is infinite. As a consequence, the image of the Teichmüller disk Δ in the Schottky space S_g is the quotient by an infinite group and in particular not isomorphic to a disk.

Proof. Denote by $\bar{\Delta}$ and \bar{C}_l the closures of Δ and C_l in $\bar{\mathcal{T}}_g$ and \bar{M}_g , respectively. Since C_l is not closed, we can find a point $z \in \bar{C}_l - C_l$; let $x \in \bar{\Delta}$ be a point above z . By Proposition 5.16 b) there is a symplectic homomorphism α such that $x \in \bar{\mathcal{T}}_g(\alpha)$.

Let $\bar{s}_\alpha: \bar{\mathcal{T}}_g(\alpha) \rightarrow \bar{S}_g$ be the quotient map for $\Gamma_g(\alpha)$ (see Proposition 5.17 and Proposition 5.19) and let $D(\iota) = s_\alpha(\Delta)$ be the image of Δ in S_g . Then the closure $\bar{D}(\iota)$ of $D(\iota)$ in \bar{S}_g contains $\bar{s}_\alpha(x)$, and we have $\bar{C}_l = \bar{\mu}^{-1}(\bar{D}(\iota))$, cf. the diagram at the end of Section 5.4.

By our assumption, \bar{C}_l is Zariski closed in \bar{M}_g . Therefore $\bar{\mu}^{-1}(\bar{C}_l)$ is an analytic subset of \bar{S}_g . $\bar{D}(\iota)$ is an irreducible component of $\bar{\mu}^{-1}(\bar{C}_l)$ and hence also an analytic subset.

Recall, from Corollary 2.21, that $\Delta/\mathrm{Stab}(\Delta)$ is the normalization of C_l . Furthermore, by Proposition 4.14, $\bar{\Delta}$ is isomorphic to $\mathbb{H} \cup \{\text{cusps of } \bar{\Gamma}_l^*\}$. Therefore $\bar{\Delta}/\mathrm{Stab}(\Delta)$ is the normalization of \bar{C}_l . The restriction of the quotient map $\bar{\Delta} \rightarrow \bar{\Delta}/\mathrm{Stab}(\Delta)$ to the intersection $\bar{\Delta}_\alpha = \bar{\Delta} \cap \bar{\mathcal{T}}_g(\alpha)$ factors through \bar{s}_α . If the intersection $\mathrm{Stab}(\Delta) \cap \Gamma_g(\alpha)$ was trivial, this restriction would be an isomorphism. But then $\bar{\Delta}_\alpha$ would be isomorphic to an analytic subset of a complex manifold. This is impossible since $\bar{\Delta}_\alpha$ contains $x \in \bar{\mathcal{T}}_g - \mathcal{T}_g$ and hence is not a complex space. \square

References

- [1] W. Abikoff, Degenerating families of Riemann surfaces. *Ann. of Math.* (2) 105 (1977), 29–44. [298](#), [318](#), [319](#), [320](#)
- [2] W. Abikoff, *The real analytic theory of Teichmüller space*. Lecture Notes in Math. 820, Springer-Verlag, Berlin 1980. [300](#), [301](#), [302](#)
- [3] L. V. Ahlfors and L. Sario, *Riemann surfaces*. Princeton Math. Ser. 26, Princeton University Press, Princeton, N.J., 1960. [337](#)
- [4] L. Bers, Uniformization, moduli, and Kleinian groups. *Bull. London Math. Soc.* 4 (1972), 257–300. [295](#)
- [5] L. Bers, Fiber spaces over Teichmüller spaces. *Acta. Math.* 130 (1973), 89–126. [319](#), [321](#)
- [6] I. Bouw and M. Möller, Teichmüller curves, triangle groups, and Lyapunov exponents. Preprint, 2005. arXiv:math.AG/0511738. [296](#)
- [7] V. Braungardt, Abschluß von Überlagerungen. Diploma Thesis, Karlsruhe 1996. [316](#)

- [8] V. Braungardt, Überlagerungen von Modulräumen für Kurven. PhD Thesis, Karlsruhe 2001. [298](#), [315](#), [316](#), [317](#), [318](#), [319](#), [320](#)
- [9] K. Calta, Veech surfaces and complete periodicity in genus two. *J. Amer. Math. Soc.* 17 (2004), 871–908. [296](#)
- [10] 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. [297](#)
- [11] C. J. Earle and F. P. Gardiner, Teichmüller disks and Veech’s \mathcal{F} -structures. In *Extremal Riemann surfaces* (San Francisco, CA, 1995), Contemp. Math. 201, Amer. Math. Soc., Providence, R.I., 1997, 165–189. [296](#), [303](#), [311](#), [312](#)
- [12] F. P. Gardiner and N. Lakic, *Quasiconformal Teichmüller theory*. Math. Surveys Monogr. 76, Amer. Math. Soc., Providence, R.I., 2000. [303](#), [306](#)
- [13] L. Gerritzen and F. Herrlich, The extended Schottky space. *J. Reine Angew. Math.* 389 (1988), 190–208. [300](#), [336](#), [338](#), [340](#), [341](#), [342](#)
- [14] H. Grauert and R. Remmert, *Coherent analytic sheaves*. Grundlehren Math. Wiss. 265, Springer-Verlag, Berlin 1984. [315](#), [316](#)
- [15] E. Gutkin and C. Judge, Affine mappings of translation surfaces: geometry and arithmetic. *Duke Math. J.* 103 (2000), 191–213. [297](#)
- [16] J. Harris and I. Morrison, *Moduli of curves*. Grad. Texts in Math. 187, Springer-Verlag, New York 1998. [295](#), [318](#)
- [17] F. Herrlich, Moduli for stable marked trees of projective lines. *Math. Ann.* 291 (1991), 643–661. [342](#)
- [18] F. Herrlich and G. Schmithüsen, An extraordinary origami curve. Preprint, 2005. arXiv:math.AG/0509195. [297](#)
- [19] P. Hubert and S. Lelièvre, Prime arithmetic Teichmüller discs in $\mathcal{H}(2)$. *Israel J. Math.* 151 (2006), 281–321. [297](#)
- [20] P. Hubert and T. A. Schmidt, Invariants of translation surfaces. *Ann. Inst. Fourier (Grenoble)* 51 (2001), 461–495. [296](#), [311](#)
- [21] Y. Imayoshi and M. Taniguchi, *An introduction to Teichmüller spaces*. Springer-Verlag, Tokyo 1992. [320](#)
- [22] F. Knudsen, The projectivity of the moduli space of stable curves. I, II, III. *Math. Scand.* 39 (1976), 19–55; 52 (1983), 161–199, 200–212. [297](#)
- [23] P. Lochak, On arithmetic curves in the moduli spaces of curves. *J. Inst. Math. Jussieu* 4 (2005), 443–508. [296](#), [297](#), [303](#)
- [24] M. Maier, Schnittgraphen von stabilen Kurven, die als Randpunkte von Origamikurven vorkommen. Diploma Thesis, Karlsruhe 2005. [298](#)
- [25] B. Maskit, *Kleinian groups*. Grundlehren Math. Wiss. 287, Springer-Verlag, Berlin 1988. [317](#), [336](#)
- [26] H. Masur, On a class of geodesics in Teichmüller space. *Ann. of Math.* (2) 102 (1975), 205–221. [299](#), [322](#), [323](#), [334](#)
- [27] C. T. McMullen, Billiards and Teichmüller curves on Hilbert modular surfaces. *J. Amer. Math. Soc.* 16 (2003), 857–885. [296](#), [303](#), [311](#), [313](#), [314](#)

- [28] M. Möller, Variations of Hodge structure of Teichmüller curves. *J. Amer. Math. Soc.* 19 (2006), 327–344. [296](#)
- [29] S. Nag, *The complex analytic theory of Teichmüller spaces*. Canad. Math. Soc. Ser. Monographs Adv. Texts, John Wiley & Sons, New York 1988. [303](#), [310](#)
- [30] G. Schmithüsen, An algorithm for finding the Veech group of an origami. *Experiment. Math.* 13 (2004), 459–472. [297](#)
- [31] G. Schmithüsen, Veech groups of origamis. PhD Thesis, Karlsruhe 2005. [297](#)
- [32] F. Schottky, Ueber eine specielle Function, welche bei einer bestimmten linearen Transformation ihres Arguments unverändert bleibt. *J. Reine Angew. Math.* 101 (1887), 227–272. [299](#), [336](#)
- [33] K. Strebel, *Quadratic differentials*. *Ergeb. Math. Grenzgeb.* (3) 5, Springer-Verlag, Berlin 1984. [323](#)
- [34] W. A. Veech, Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. Math.* 97 (1989), 553–583. [296](#), [303](#), [311](#), [312](#), [331](#)

Part B

The group theory, 1

Chapter 7

Introduction to mapping class groups of surfaces and related groups

Shigeyuki Morita

Contents

1	Introduction	353
2	Diffeomorphism groups and diffeotopy groups of differentiable manifolds	354
3	Mapping class groups of surfaces	357
4	Automorphism groups of free groups and IA automorphism groups	361
5	Dehn twists	363
6	Mapping class groups acting on the homology of surfaces and the Torelli groups	365
7	Johnson homomorphisms	371
8	Teichmüller space and Outer Space	376
9	Symplectomorphism groups of surfaces	378
10	Extensions of the Johnson homomorphism and the flux homomorphism	380
	References	382

1 Introduction

Let Σ_g denote a closed oriented surface of genus g . The mapping class group of Σ_g , denoted by \mathcal{M}_g , is the group of isotopy classes of orientation preserving diffeomorphisms of Σ_g . This group has been investigated from various points of view for many years.

First of all, this group has been one of the main objects in the combinatorial group theory, the other one being the automorphism group of a finitely generated free group. Secondly, \mathcal{M}_g acts on the Teichmüller space and the quotient space is the moduli space of genus g Riemann surfaces which is a very important space in both algebraic geometry and complex analysis. Thirdly, this group has been playing crucial roles in the theory of 3-manifolds in relation to Heegaard decompositions as well as the geometry of surface bundles over the circle.

Reflecting this situation, there exist already many survey papers concerning various aspects of the mapping class group. We have the famous book by Birman [7] and several survey papers such as [8], [9], [27], [10]. Ivanov's paper [34] gives a very nice

introduction to the present state of the study of the mapping class group (see also [32], [33]). Also we have the survey paper [37] by Johnson on the structure of the Torelli group which is a very important subgroup of the mapping class group.

There are also survey papers on the cohomological structure of the mapping class group or the moduli space of Riemann surfaces such as Harer [26], Hain–Looijenga [24] and the author [69], [71].

The purpose of this chapter is to describe somewhat different points of view in the study of the mapping class group and to suggest possible new directions for future research. More precisely, we would like to consider this as a special case of the study of the structure of the diffeomorphism group as well as the diffeotopy group of general C^∞ manifolds. We also would like to seek for similarity and difference between the structures of the mapping class group and some of its closely related groups such as the automorphism groups of free groups and the symplectomorphism groups of surfaces. We refer to Vogtmann [84] for a survey of the study of the automorphism groups of free groups.

This work was partially supported by JSPS Grant No. 16204005.

2 Diffeomorphism groups and diffeotopy groups of differentiable manifolds

Let M be a closed C^∞ manifold. Then it is a very important problem to determine the set of all the isomorphism classes of differentiable M -bundles

$$\pi : E \longrightarrow X$$

over a given C^∞ manifold X . By a standard technique in topology, this problem can be translated into the following one in homotopy theory. We denote by $\text{Diff } M$ the diffeomorphism group of M equipped with the C^∞ topology and let $\text{BDiff } M$ be its classifying space. Then we have a natural identification

$$\{\text{isomorphism classes of smooth } M\text{-bundles over } X\} \cong [X, \text{BDiff } M]$$

where the right hand side denotes the set of all the homotopy classes of continuous mappings from X to $\text{BDiff } M$. Hence we meet with the problem of determining the homotopy type of $\text{Diff } M$, in particular the computation of the homotopy groups $\pi_i(\text{Diff } M)$. However this is an extremely difficult problem for a general differentiable manifold M .

Let us consider the simplest case where $X = S^1$. Any smooth M -bundle

$$\pi : E \longrightarrow S^1$$

over S^1 can be described as follows. Choose a point $x \in S^1$ and cut the total space E along the fiber $\pi^{-1}(x)$ over x . Then we obtain an M -bundle over the interval I so that it is diffeomorphic to the product $I \times M$. Observe here that we have chosen a

diffeomorphism $\pi^{-1}(x) \cong M$. Now the given bundle E can be recovered as

$$E = I \times M / (1, p) \sim (0, \varphi(p)) \quad (p \in M)$$

for a certain element $\varphi \in \text{Diff } M$. Namely E is obtained from $I \times M$ by pasting $\{1\} \times M$ to $\{0\} \times M$ by the diffeomorphism φ . Note that the element φ is well-defined up to isotopy and also note that, if we change the identification $\pi^{-1}(x) \cong M$, then the element φ changes into its conjugate element $\psi\varphi\psi^{-1}$. Thus we obtain

$$\begin{aligned} & \{\text{isomorphism classes of smooth } M\text{-bundles over } S^1\} \\ & \cong [S^1, \text{BDiff } M] \\ & \cong \{\text{isotopy classes of elements of } \text{Diff } M\} / \text{conjugacy}. \end{aligned}$$

From the above consideration, it is natural to introduce the group consisting of all the isotopy classes of elements of $\text{Diff } M$ which we denote by $\mathcal{D}(M)$ and call it the *diffeotopy group* of M . It can also be described as the group of path components of the topological group $\text{Diff } M$, namely

$$\mathcal{D}(M) = \pi_0(\text{Diff } M).$$

Alternatively, we can also write

$$\mathcal{D}(M) = \text{Diff } M / \text{Diff}_0 M$$

where $\text{Diff}_0 M$ denotes the identity component of $\text{Diff } M$. Thus we have an extension

$$1 \longrightarrow \text{Diff}_0 M \longrightarrow \text{Diff } M \longrightarrow \mathcal{D}(M) \longrightarrow 1.$$

This simplest case, namely the determination of $\mathcal{D}(M)$ is already a very difficult problem in general. In fact, the case where M is an n -dimensional sphere S^n was one of the most important subjects during the early years of differential topology. By virtue of the foundational work of Cerf [13], [14] as well as the solution of the generalized Poincaré conjecture due to Smale [81], there is a natural isomorphism

$$\mathcal{D}_+(S^n) \cong \theta_{n+1} \quad (n \geq 5)$$

where $\mathcal{D}_+(S^n) = \pi_0(\text{Diff}_+ S^n)$ denotes the *orientation preserving* diffeotopy group of S^n and θ_n denotes the group of homotopy n -spheres introduced and studied by Kervaire and Milnor [44].

Since $\mathcal{D}(M)$ is the quotient group of the diffeomorphism group $\text{Diff } M$ divided by the equivalence relation of isotopy which is stronger than (or sometimes equal to) that of homotopy, $\mathcal{D}(M)$ acts on any homotopy invariants of M such as the fundamental group $\pi_1 M$ and the homology group $H_*(M; \mathbb{Z})$. We first consider the case of $\pi_1 M$. For any abstract group Γ , let $\text{Aut } \Gamma$ denote the automorphism group of Γ . Any element $\gamma \in \Gamma$ defines that of $\text{Aut } \Gamma$ which represents the inner automorphism of Γ by the element γ . This induces a homomorphism $\Gamma \rightarrow \text{Aut } \Gamma$. Let $\text{Inn } \Gamma$ denote the image of this homomorphism. Clearly the group Γ is abelian if and only if $\text{Inn } \Gamma$ is the trivial subgroup of $\text{Aut } \Gamma$. It is easy to see that $\text{Inn } \Gamma$ is a normal subgroup of $\text{Aut } \Gamma$. The quotient group $\text{Aut } \Gamma / \text{Inn } \Gamma$ is denoted by $\text{Out } \Gamma$ and it is called the

outer automorphism group of Γ . In these terminologies, we can say that the action of $\mathcal{D}(M)$ on $\pi_1 M$ induces a homomorphism

$$\rho_\pi : \mathcal{D}(M) \longrightarrow \text{Out } \pi_1 M.$$

Also the action of $\mathcal{D}(M)$ on the homology group $H^*(M; \mathbb{Z})$ induces another homomorphism

$$\rho_H : \mathcal{D}(M) \longrightarrow \text{Aut } H_*(M; \mathbb{Z}).$$

Then there arise natural questions about these homomorphisms. For example we could ask whether the homomorphisms ρ_π, ρ_H are surjective or not. We could also ask whether they are injective or not. In the cases where these questions are answered negatively, the problem of describing the images as well as the kernels of these homomorphisms arise. It turns out that these questions depend on the global topology of the manifold M and the above problems are often very difficult to be settled.

If there is given a geometric structure on M , then the automorphism group of this structure, which is considered to be a subgroup of $\text{Diff } M$, is one of the basic objects to be studied. Here we would like to mention a few examples.

The first obvious example is the case where there is given a Riemannian metric on M . Then the corresponding automorphism group is nothing but the isometry group $\text{Isom } M$. If M is compact, then this group is known to be a compact Lie group sitting inside $\text{Diff } M$. For example, for the n -sphere S^n with the standard metric, we have $\text{Isom } S^n = \text{O}(n+1) \subset \text{Diff } S^n$. It has been one of the main problems in the theory of differentiable transformation groups to study possible subgroups of $\text{Diff } M$ which are Lie transformation groups for a given manifold M .

The second example is the case where there is given a volume form ν on M . Then we can consider the subgroup of $\text{Diff } M$, denoted by $\text{Diff}^\nu M$, which consists of those diffeomorphisms which preserve the form ν . It is usually called the *volume preserving diffeomorphism group* of M . Moser's theorem in [72] implies that the inclusion $\text{Diff}^\nu M \subset \text{Diff}_+ M$ is a homotopy equivalence. Hence the classifying spaces $\text{BDiff}^\nu M$ and $\text{BDiff}_+ M$ have the same homotopy type. In particular, we have a *bijection*

$$\pi_0(\text{Diff}^\nu M) \cong \pi_0(\text{Diff}_+ M) = \mathcal{D}_+ M.$$

However the two groups $\text{Diff}^\nu M$ and $\text{Diff}_+ M$ seem to have considerably different properties as abstract groups and there should be many interesting problems here.

The third example is the case where there is given a *symplectic form* ω on a $2n$ dimensional manifold M . A symplectic form is, by definition, a closed 2-form ω such that ω^n is a volume form on M . The pair (M, ω) is called a *symplectic manifold* and the subgroup $\text{Symp}(M, \omega) \subset \text{Diff } M$ consisting of those diffeomorphisms which preserve the form ω is called the *symplectomorphism group* of (M, ω) . Recently there have been obtained many interesting deep results in geometry and topology of symplectic manifolds as well as those of symplectomorphism groups (see the book [61] by McDuff and Salamon which gives an excellent introduction to this field). Let $\text{Symp}_0(M, \omega)$ be the identity component of $\text{Symp}(M, \omega)$ which is a normal subgroup. The quotient

group $\mathcal{M}(M, \omega) = \text{Symp}(M, \omega)/\text{Symp}_0(M, \omega)$ is called the symplectic mapping class group of (M, ω) . Thus we have the following extension.

$$1 \longrightarrow \text{Symp}_0(M, \omega) \longrightarrow \text{Symp}(M, \omega) \longrightarrow \mathcal{M}(M, \omega) \longrightarrow 1.$$

We have an obvious natural homomorphism $\mathcal{M}(M, \omega) \rightarrow \mathcal{D}_+(M)$ and there are many interesting problems concerning it. For example, we can ask about the structure of the image as well as the kernel of this homomorphism. See Banyaga [3] and McDuff [60] for basic results concerning the structures of the groups $\text{Diff}^v M$ and $\text{Symp}(M, \omega)$.

Recall that any C^∞ manifold M admits a real analytic structure (see Whitney [86]). Our final example here is the subgroup $\text{Diff}^\omega M \subset \text{Diff} M$ consisting of *real analytic* diffeomorphisms of M (with respect to a fixed real analytic structure). It seems to be an interesting problem to investigate whether there exist differences in algebraic structures between the two groups $\text{Diff}^\omega M$ and $\text{Diff} M$.

3 Mapping class groups of surfaces

The *mapping class group* of a closed oriented surface Σ_g of genus g , which we denote by \mathcal{M}_g , is by definition the *oriented* diffeotopy group of Σ_g . Namely it is the group consisting of all the isotopy classes of *orientation preserving* diffeomorphisms of Σ_g . If we denote by $\text{Diff}_+ \Sigma_g$ the group of orientation preserving diffeomorphisms of Σ_g equipped with the C^∞ topology, then we have

$$\mathcal{M}_g = \pi_0(\text{Diff}_+ \Sigma_g).$$

It is easy to see that we have an extension

$$1 \longrightarrow \mathcal{M}_g \longrightarrow \mathcal{D}(\Sigma_g) \longrightarrow \mathbb{Z}/2 \longrightarrow 1$$

where the homomorphism $\mathcal{D}(\Sigma_g) \rightarrow \mathbb{Z}/2$ is induced by the action of $\mathcal{D}(\Sigma_g)$ on the set of orientations on Σ_g or equivalently on the group $H_2(\Sigma_g; \mathbb{Z}) \cong \mathbb{Z}$.

As is well known, the topology and also the geometry of surfaces Σ_g can be roughly divided into three classes, namely the cases where $g = 0$, $g = 1$ and $g \geq 2$. From the topological point of view, the fundamental group $\pi_1 \Sigma_g$ is trivial for $g = 0$, rank 2 abelian for $g = 1$ and non-abelian for $g \geq 2$. On the other hand, from the geometrical point of view, each surface Σ_g admits a Riemannian metric of constant Gaussian curvature K where $K \equiv 1, 0, -1$ for $g = 0, g = 1, g \geq 2$ respectively. Furthermore for the latter two cases, there exist plenty of such metrics up to isotopy and they fit together to make a nice topological space called the *Teichmüller space*. It turns out that the structure of the mapping class group \mathcal{M}_g reflects this rough classification of surfaces rather closely as follows.

First of all, we consider the case where $g = 0$, namely the case of the sphere S^2 . Then a theorem of Smale [80] implies that the inclusion

$$\text{O}(3) \subset \text{Diff} S^2$$

is a homotopy equivalence. It follows, in particular, that the subgroup $\text{Diff}_+ S^2$ is connected. Therefore any orientation preserving diffeomorphism of S^2 is isotopic to the identity. Hence the genus 0 mapping class group \mathcal{M}_0 is the trivial group.

Next we assume that $g \geq 1$ and consider the action of \mathcal{M}_g on $\pi_1 \Sigma_g$ as in §2 where we considered the case of a general manifold M . It is known that the surface Σ_g for $g \geq 1$ is an Eilenberg–MacLane space $K(\pi_1 \Sigma_g, 1)$ meaning that the higher homotopy groups $\pi_i \Sigma_g$ vanish for all $i \geq 2$. Here for a given (abstract) group π and a positive integer n , any topological space X with the property that

$$\pi_i X \cong \begin{cases} \pi & (i = n) \\ 0 & (i \neq n) \end{cases}$$

is called an *Eilenberg–MacLane space* $K(\pi, n)$ (we assume that π is an abelian group in the cases where $n \geq 2$).

In fact, in the genus 1 case, the torus T^2 is expressed as $\mathbb{R}^2/\mathbb{Z}^2$ so that its universal covering manifold is the plane \mathbb{R}^2 . Hence all the higher homotopy groups of T^2 vanish and T^2 is a $K(\mathbb{Z}^2, 1)$. In the cases where $g \geq 2$, there exist Riemannian metrics on Σ_g which have constant negative curvature -1 . It follows that its universal covering manifold is isometric to the upper half plane $\mathbb{H} = \{(x, y) \in \mathbb{R}^2; y > 0\}$ equipped with the Poincaré metric. Hence Σ_g is a $K(\pi_1 \Sigma_g, 1)$. Here recall the standard presentation of the fundamental group $\pi_1 \Sigma_g$ which is expressed as

$$\begin{aligned} \pi_1 \Sigma_g &= \langle \alpha_1, \dots, \alpha_g, \beta_1, \dots, \beta_g; \zeta \rangle && (2g \text{ generators}), \\ \zeta &= [\alpha_1, \beta_1] \dots [\alpha_g, \beta_g] && (\text{defining relation}). \end{aligned}$$

Now we state a classical theorem which is usually called the *Dehn–Nielsen theorem*.

Theorem 3.1 (Dehn–Nielsen, Baer). *The natural action of \mathcal{M}_g on $\pi_1 \Sigma_g$ induces an isomorphism*

$$\mathcal{M}_g \cong \text{Out}_+ \pi_1 \Sigma_g.$$

Since $\pi_1 T^2 \cong \mathbb{Z}^2$ is an abelian group, $\text{Aut } \mathbb{Z}^2 = \text{Out } \mathbb{Z}^2 \cong \text{GL}(2, \mathbb{Z})$. The subscript $+$ in $\text{Out}_+ \mathbb{Z}^2$ means, in this case, that we consider only matrices with determinant 1. Thus we can write

$$\mathcal{M}_1 \cong \text{SL}(2, \mathbb{Z}).$$

In order to interpret the subscript $+$ for the general case, we briefly recall the definition of the homology group as well as the cohomology group of an abstract group π (see Brown’s book [12] for details). It is known that, there exists a $K(\pi, 1)$ which is a CW complex. Furthermore it is uniquely defined up to homotopy equivalences. Hence for any π -module M (namely M is a module and there is given a homomorphism $\pi \rightarrow \text{Aut } M$), we can define the (co)homology group of π with coefficients in M by

setting

$$\begin{aligned} H_*(\pi; M) &= H_*(K(\pi, 1); M), \\ H^*(\pi; M) &= H^*(K(\pi, 1); M) \end{aligned}$$

where M on the right hand sides denotes the local system over $K(\pi, 1)$ induced by the given π action on the module M . Any group homomorphism $\rho: \pi \rightarrow \pi'$ induces a homomorphism $\rho_*: H_*(\pi) \rightarrow H_*(\pi')$ and similarly for the homology with twisted coefficients as well as the cohomology group. In particular, we have a homomorphism

$$\text{Aut } \pi \longrightarrow \text{Aut } H_*(\pi; M).$$

It is well known (and not so difficult to see) that the inner automorphisms induce the trivial action on the homology group so that we obtain a homomorphism

$$\text{Out } \pi \longrightarrow \text{Aut } H_*(\pi; M).$$

Now for any $g \geq 1$, Σ_g is a $K(\pi_1 \Sigma_g, 1)$ as mentioned above. Hence we have $H_*(\pi_1 \Sigma_g; \mathbb{Z}) = H_*(\Sigma_g; \mathbb{Z})$. In particular $H_2(\pi_1 \Sigma_g; \mathbb{Z}) \cong \mathbb{Z}$ and we have a homomorphism

$$\text{Out } \pi_1 \Sigma_g \longrightarrow \text{Aut } H_2(\pi_1 \Sigma_g; \mathbb{Z}) \cong \text{Aut } \mathbb{Z} \cong \mathbb{Z}/2.$$

The group $\text{Out}_{+\pi_1 \Sigma_g}$ in Theorem 3.1 is defined to be the kernel of the above homomorphism. Sometimes this group is called the *orientation preserving* outer automorphism group of $\pi_1 \Sigma_g$ because it is the subgroup of the whole group consisting of those outer automorphisms which are induced from orientation preserving diffeomorphisms of Σ_g .

Now we would like to mention the methods of proving Theorem 3.1 somewhat historically.

First of all, for any topological space X , let $\mathcal{E}(X)$ denote the set of all the homotopy classes of self homotopy equivalences of X . The composition of mappings induces a natural group structure on $\mathcal{E}(X)$. Next for any *topological* manifold M , let $\text{Homeo } M$ denote the group of all the homeomorphisms of M equipped with the compact open topology and let $\mathcal{H}(M) = \text{Homeo } M / \text{Homeo}_0 M$ denote the quotient group divided by the identity component $\text{Homeo}_0 M$ of $\text{Homeo } M$. It is called the *homeotopy group* of M . Now if M is a C^∞ manifold, then there is a natural sequence of forgetful homomorphisms

$$\mathcal{D}(M) \longrightarrow \mathcal{H}(M) \longrightarrow \mathcal{E}(M).$$

One can also introduce the equivalence relation on $\text{Diff } M$, $\text{Homeo } M$ induced by the homotopy of mappings to obtain variants of diffeotopy or homeotopy groups.

In general, these groups are all different from each other and they have their own meanings and properties. However, a very important and characteristic phenomenon occurs in dimension 2 and that is the fact that they are *all* equal for surfaces. Namely

we have isomorphisms

$$\begin{aligned}\mathcal{D}(\Sigma_g) &\cong \mathcal{H}(\Sigma_g) \cong \mathcal{E}(\Sigma_g), \\ \mathcal{M}_g = \mathcal{D}_+(\Sigma_g) &\cong \mathcal{H}_+(\Sigma_g) \cong \mathcal{E}_+(\Sigma_g)\end{aligned}$$

where the subscripts $+$ means appropriate subgroups of index 2 consisting of orientation preserving elements. Now elementary homotopy theory implies that there are canonical isomorphisms

$$\mathcal{E}(\Sigma_g) \cong \text{Out } \pi_1 \Sigma_g, \quad \mathcal{E}_+(\Sigma_g) \cong \text{Out}_+ \pi_1 \Sigma_g.$$

In fact the isomorphism $\mathcal{E}(X) \cong \text{Out } \pi_1 X$ holds for any $K(\pi, 1)$ space X .

Dehn and then Nielsen [77] proved that the natural map

$$\mathcal{H}_+(\Sigma_g) \longrightarrow \text{Out}_+ \pi_1 \Sigma_g$$

is *surjective*. The *injectivity* of the same map was proved by Baer [2] and, much later, reproved by Epstein [19]. It may be said that Dehn and Nielsen essentially proved that the natural map

$$\mathcal{D}_+(\Sigma_g) \longrightarrow \text{Out}_+ \pi_1 \Sigma_g$$

is *surjective*, although it is unclear how they recognized the concept of diffeomorphisms as well as homeomorphisms which are now strictly distinguished. The *injectivity* of the above map can be obtained by adapting the proofs of Baer and Epstein from the context of homeomorphisms to that of diffeomorphisms which are known to be possible in this low dimensional case.

There are variants of the mapping class group and analogues of Theorem 3.1 for them as follows. First, if we choose a base point $*$ \in Σ_g , then we can consider the subgroup $\text{Diff}_+(\Sigma_g, *) \subset \text{Diff}_+ \Sigma_g$ consisting of all the orientation preserving diffeomorphisms of Σ_g which fix the base point $*$. Then we set

$$\mathcal{M}_{g,*} = \pi_0(\text{Diff}_+(\Sigma_g, *))$$

and call it the mapping class group of Σ_g *relative* to the base point. There is the forgetful homomorphism $\mathcal{M}_{g,*} \rightarrow \mathcal{M}_g$ which is an isomorphism for $g = 0, 1$ and in the cases $g \geq 2$, the kernel of this homomorphism is known to be canonically isomorphic to $\pi_1 \Sigma_g$. Thus we have an extension

$$1 \longrightarrow \pi_1 \Sigma_g \longrightarrow \mathcal{M}_{g,*} \longrightarrow \mathcal{M}_g \longrightarrow 1 \quad (g \geq 2).$$

Next if we choose an embedded disk $D^2 \subset \Sigma_g$, then we can consider the subgroup $\text{Diff}(\Sigma_g, D^2) \subset \text{Diff}_+ \Sigma_g$ consisting of all the diffeomorphisms of Σ_g which restrict to the identity of D^2 . Then we set

$$\mathcal{M}_{g,1} = \pi_0(\text{Diff}(\Sigma_g, D^2))$$

and call it the mapping class group of Σ_g *relative* to D^2 . Alternatively, we can consider the compact surface $\Sigma_g^0 = \Sigma_g \setminus \text{Int } D^2$ and the diffeomorphism group $\text{Diff}(\Sigma_g^0, \partial \Sigma_g^0)$ consisting of all diffeomorphisms of Σ_g^0 which restrict to the identity on the boundary.

Then we can also define

$$\mathcal{M}_{g,1} = \pi_0(\text{Diff}(\Sigma_g^0, \partial \Sigma_g^0)).$$

If we choose a base point $*$ in $D^2 \subset \Sigma_g$, then we have the forgetful homomorphism $\mathcal{M}_{g,1} \rightarrow \mathcal{M}_{g,*}$, and in the case $g \geq 1$ the kernel of this homomorphism is known to be isomorphic to \mathbb{Z} so that we have an extension

$$1 \longrightarrow \mathbb{Z} \longrightarrow \mathcal{M}_{g,1} \longrightarrow \mathcal{M}_{g,*} \longrightarrow 1 \quad (g \geq 1). \quad (3.1)$$

Now we can state the analogue of Theorem 3.1 for the mapping class groups $\mathcal{M}_{g,*}$ and $\mathcal{M}_{g,1}$ as follows. This should also be considered as a classical theorem going back to Magnus [54] and Zieschang [87].

Theorem 3.2. *There are natural isomorphisms*

$$\begin{aligned} \mathcal{M}_{g,*} &\cong \text{Aut}_+ \pi_1 \Sigma_g, \\ \mathcal{M}_{g,1} &\cong \{ \varphi \in \text{Aut } \pi_1 \Sigma_g^0; \varphi(\zeta) = \zeta \} \end{aligned}$$

where $\zeta = [\alpha_1, \beta_1] \dots [\alpha_g, \beta_g]$ denotes the single defining relation of $\pi_1 \Sigma_g$ with respect to a standard generating system $\alpha_1, \dots, \alpha_g, \beta_1, \dots, \beta_g$.

We can also consider the mapping class groups of Σ_g relative to finitely many distinguished points as well as finitely many embedded disks on Σ_g . However here we omit them.

4 Automorphism groups of free groups and IA automorphism groups

Let F_n be a free group of rank n . We denote by $\text{Aut } F_n$ the *automorphism group of the free group* F_n . If $n = 1$, then clearly $\text{Aut } \mathbb{Z} \cong \mathbb{Z}/2$. Henceforth we assume that $n \geq 2$. Then the homomorphism $\iota: F_n \rightarrow \text{Aut } F_n$ defined by

$$\iota(\alpha)\gamma = \alpha\gamma\alpha^{-1} \quad (\alpha, \gamma \in F_n)$$

is easily seen to be injective. As was already mentioned in §2 in a general setting, the image $\text{Im } \iota$ is denoted by $\text{Inn } F_n$ and called the *inner* automorphism group of F_n . It can be checked that $\text{Inn } F_n$ is a normal subgroup of $\text{Aut } F_n$ so that we can consider the quotient group

$$\text{Out } F_n = \text{Aut } F_n / \text{Inn } F_n$$

which is called the *outer automorphism group of the free group* F_n .

These groups $\text{Aut } F_n$ and $\text{Out } F_n$ have been one of the main objects of combinatorial group theory going back to the works of Nielsen and then Magnus from the late 1910s to the 1930s.

$\text{Aut } F_n$ acts naturally on the abelianization of F_n , which is a free abelian group of rank n . If we choose a system $\gamma_1, \dots, \gamma_n$ of generators for F_n , then we obtain a homomorphism

$$\rho_0: \text{Aut } F_n \longrightarrow \text{GL}(n, \mathbb{Z}). \quad (4.1)$$

It is easy to see that ρ_0 is trivial on the subgroup $\text{Inn } F_n$. Hence we obtain a homomorphism

$$\rho_0: \text{Out } F_n \longrightarrow \text{GL}(n, \mathbb{Z}). \quad (4.2)$$

Nielsen [74] proved that the above homomorphism is an isomorphism for $n = 2$ so that

$$\text{Out } F_2 \cong \text{GL}(2, \mathbb{Z}).$$

However for $n \geq 3$, Nielsen [75] also observed that ρ_0 is not injective and in [76] he proved that the following four elements

- (i) $\gamma_1 \rightarrow \gamma_2, \gamma_2 \rightarrow \gamma_1, \gamma_i \rightarrow \gamma_i \quad (i = 3, \dots, n),$
- (ii) $\gamma_1 \rightarrow \gamma_1^{-1}, \gamma_i \rightarrow \gamma_i \quad (i = 2, \dots, n),$
- (iii) $\gamma_1 \rightarrow \gamma_1\gamma_2, \gamma_i \rightarrow \gamma_i \quad (i = 2, \dots, n),$
- (iv) $\gamma_1 \rightarrow \gamma_2, \gamma_2 \rightarrow \gamma_3, \dots, \gamma_n \rightarrow \gamma_1$

generate $\text{Aut } F_n$ and hence $\text{Out } F_n$. By this he was able to prove that the homomorphisms ρ_0 (4.1), (4.2) above are *surjective*. He also gave a finite complete set of defining relations in terms of the above generators, for both of $\text{Aut } F_n$ and $\text{Out } F_n$. Later McCool [58] gave a simpler finite presentation for $\text{Aut } F_n$. Also Gersten [21] gave a finite presentation for the subgroup $\text{Aut}_+ F_n$ which is the full inverse image under ρ_0 of the subgroup $\text{GL}^+(n, \mathbb{Z}) \subset \text{GL}(n, \mathbb{Z})$ consisting of matrices with determinant 1.

The kernels of the homomorphisms ρ_0 are called *IA (outer) automorphism groups* of F_n which we denote by IAut_n and IOut_n respectively. Thus we have group extensions

$$\begin{aligned} 1 &\longrightarrow \text{IAut}_n \longrightarrow \text{Aut } F_n \xrightarrow{\rho_0} \text{GL}(n, \mathbb{Z}) \longrightarrow 1, \\ 1 &\longrightarrow \text{IOut}_n \longrightarrow \text{Out } F_n \xrightarrow{\rho_0} \text{GL}(n, \mathbb{Z}) \longrightarrow 1. \end{aligned}$$

Magnus [53] proved that the group IAut_n is finitely generated. On the other hand, Baumslag–Taylor [4] proved that IAut_n is torsion free.

There is a close connection between the mapping class group and the automorphism groups of free groups. More precisely, we have the following two explicit relations. One is the realization of the mapping class group $\mathcal{M}_{g,1}$ of Σ_g relative to an embedded disk $D \subset \Sigma_g$ as a subgroup of $\text{Aut } F_{2g}$ described as

$$\mathcal{M}_{g,1} = \{\varphi \in \text{Aut } F_{2g}; \varphi(\zeta) = \zeta\} \subset \text{Aut } F_{2g} \quad (4.3)$$

where $\zeta = [\gamma_1, \gamma_2] \dots [\gamma_{2g-1}, \gamma_{2g}] \in F_{2g}$. This follows from Theorem 3.2 because $\pi_1 \Sigma_g^0$ is isomorphic to F_{2g} . The other is given as follows. The subgroup $\mathbb{Z} = \text{Ker}(\mathcal{M}_{g,1} \rightarrow \mathcal{M}_{g,*})$ described in (3.1) is generated by the Dehn twist (see the next

section §5) along a simple closed curve parallel to the boundary of the embedded disk $D^2 \subset \Sigma_g$ and its action on $\pi_1 \Sigma_g^0$ is the conjugation by the element ζ above. Since $\mathcal{M}_{g,1}/\mathbb{Z}$ is canonically isomorphic to $\mathcal{M}_{g,*}$, we obtain a representation

$$\mathcal{M}_{g,*} \longrightarrow \text{Aut } F_{2g}/\text{Inn } F_{2g} = \text{Out } F_{2g}.$$

It is known that this representation is injective so that we can consider $\mathcal{M}_{g,*}$ as a subgroup of $\text{Out } F_{2g}$

$$\mathcal{M}_{g,*} \subset \text{Out } F_{2g}.$$

The comparison of various group theoretical properties between the mapping class groups $\mathcal{M}_{g,*}$, \mathcal{M}_g and automorphism groups of free groups $\text{Aut } F_n$, $\text{Out } F_n$ have been an important subject since the very beginning of combinatorial group theory. Recently, this tendency is strengthened in a wider framework including geometric viewpoints.

Finally we mention the following result of Laudenbach [50] which shows that, up to a certain finite group, the (outer) automorphism groups of free groups are naturally isomorphic to the diffeotopy groups of certain 3-manifolds. Let $n S^1 \times S^2$ denote the connected sum of n -copies of $S^1 \times S^2$. Then there are the following exact sequences

$$\begin{aligned} 1 &\longrightarrow (\mathbb{Z}/2)^n \longrightarrow \text{Out } F_n \longrightarrow \mathcal{D}(n S^1 \times S^2) \longrightarrow 1, \\ 1 &\longrightarrow (\mathbb{Z}/2)^n \longrightarrow \text{Aut } F_n \longrightarrow \mathcal{D}(n S^1 \times S^2, \text{rel } D^3) \longrightarrow 1 \end{aligned}$$

where $\mathcal{D}(n S^1 \times S^2, \text{rel } D^3)$ denotes the group of path components of those diffeomorphisms of $n S^1 \times S^2$ which are the identity on an embedded disk $D^3 \subset n S^1 \times S^2$.

5 Dehn twists

So far we have not mentioned explicit examples of elements of the mapping class group \mathcal{M}_g . Here we describe the most important construction of such elements which is called the *Dehn twist* because it was introduced by Dehn.

Suppose that there is given a simple closed curve C on Σ_g and also recall that there is specified an orientation on Σ_g . Then we can define an element $\tau_C \in \mathcal{M}_g$, which is called the (right handed) Dehn twist along C , as follows. Let us choose an embedding

$$i: S^1 \times [-1, 1] \longrightarrow \Sigma_g$$

of an annulus into Σ_g such that

- (i) $i(S^1 \times \{0\}) = C$ and
- (ii) i preserves the orientations

where we give S^1 and $[-1, 1]$ the standard orientations and the annulus $S^1 \times [-1, 1]$ the product orientation of them. Then we define a diffeomorphism φ_0 of the annulus by

$$\varphi_0(\theta, t) = (\theta + f(t), t)$$

where $0 \leq \theta \leq 2\pi$ and $-1 \leq t \leq 1$ are the coordinates of S^1 and $[-1, 1]$ respectively, and $f: [-1, 1] \rightarrow \mathbb{R}$ is a C^∞ function such that

$$f(t) = \begin{cases} 0 & (-1 \leq t \leq -\frac{1}{2}) \\ 2\pi & (\frac{1}{2} \leq t \leq 1) \end{cases}$$

and f is strictly increasing on the interval $[-\frac{1}{2}, \frac{1}{2}]$. Now recall that for any diffeomorphism ψ of a C^∞ manifold M , the closed set

$$\text{supp}(\psi) = \overline{\{p \in M; \psi(p) \neq p\}} \subset M$$

is called the *support* of φ . By the definition of φ_0 , it is clear that $\text{supp}(\varphi_0)$ is contained in $\text{Int}(S^1 \times [-1, 1])$. Hence we can define a diffeomorphism $\varphi \in \text{Diff}_+ \Sigma_g$ by setting $\varphi = \varphi_0$ on $i(S^1 \times [-1, 1])$ and extend it by the identity over the whole of Σ_g . It can be checked that the isotopy class of φ does not depend on the choice of the function f nor the orientation preserving embedding i (observe here that the *opposite* embedding $\bar{i}: S^1 \times [-1, 1] \rightarrow \Sigma_g$ defined by $\bar{i}(\theta, t) = i(-\theta, -t)$ does not change the isotopy class of φ). Furthermore it depends only on the isotopy class of the simple closed curve C . We denote the resulting mapping class by $\tau_C \in \mathcal{M}_g$ and call it the *right handed* Dehn twist along C . This is because a path, which crosses the simple closed curve C transversely at a point, will be transferred by φ to a path which, after getting near to C , goes around C once to the right direction (with respect to the orientation of Σ_g) and then goes on as before. The inverse τ_C^{-1} is called the *left handed* Dehn twist along C . Note here that if we reverse the orientation of Σ_g , then the right handed Dehn twist is changed into the left handed one and vice versa. Also note that the orientation of the simple closed curve C itself has nothing to do with the definition of the Dehn twist.

If a simple closed curve C on Σ_g bounds a disk, then it is easy to see that the corresponding Dehn twist is the identity in \mathcal{M}_g . A simple closed curve on Σ_g is called *essential* if it does not bound a disk. We define

$$\mathcal{S}(\Sigma_g) = \{\text{isotopy classes of essential simple closed curves on } \Sigma_g\}.$$

In summary, we obtain a mapping

$$\mathcal{S}(\Sigma_g) \ni [C] \longmapsto \tau_C \in \mathcal{M}_g.$$

One simple but important property of the Dehn twists is that the equality

$$\tau_{\varphi(C)} = \varphi \circ \tau_C^{\varepsilon(\varphi)} \circ \varphi^{-1} \tag{5.1}$$

holds for any simple closed curve C and any element $\varphi \in \text{Diff } \Sigma_g$, where $\varepsilon(\varphi) = 1$ or -1 if φ preserves (or reverses) the orientation of Σ_g .

It may appear first that the definition of the Dehn twists is so simple that they will cover a relatively small part of the mapping class group. However, if one observes that there are enormously many simple closed curves on Σ_g and two (or more) simple closed curves can meet each other in a very complicated way, one can easily understand

that products of Dehn twists along various simple closed curves can express very complicated elements in \mathcal{M}_g . In fact, Dehn [17] proved that finitely many Dehn twists generate \mathcal{M}_g . Later Lickorish [51] proved that a certain system of $3g - 1$ Dehn twists, which are now called the Lickorish generators, generates \mathcal{M}_g . Then Humphries [30] proved that $2g + 1$ members among the Lickorish generators already generate \mathcal{M}_g . He also proved that this number $2g + 1$ is the minimum of the number of Dehn twists which can generate \mathcal{M}_g .

As for the presentation of the mapping class group, McCool [59] proved that \mathcal{M}_g is finitely presentable without giving an explicit presentation. Hatcher and Thurston [28] gave a method of obtaining a finite presentation and it was finally completed by the work of Wajnryb [85].

6 Mapping class groups acting on the homology of surfaces and the Torelli groups

The mapping class group \mathcal{M}_g acts on the first homology group of Σ_g naturally. Assume here that $g \geq 1$ and we denote simply by H the first integral homology group $H_1(\Sigma_g; \mathbb{Z})$ of Σ_g . As an abstract group, H is a free abelian group of rank $2g$. However, there exists an important additional structure on H coming from the geometry of Σ_g . More precisely, the intersection numbers of elements of H give rise to a bilinear mapping

$$\mu: H \times H \longrightarrow \mathbb{Z}.$$

We denote by $u \cdot v$ ($u, v \in H$) the intersection number $\mu(u, v)$. Then $v \cdot u = -u \cdot v$ so that μ is *skew symmetric*. The natural action of \mathcal{M}_g on H comes from orientation preserving diffeomorphisms of the surface Σ_g . Hence it clearly preserves the intersection pairing μ so that we obtain a homomorphism

$$\rho_0: \mathcal{M}_g \longrightarrow \text{Aut}(H, \mu) \tag{6.1}$$

where $\text{Aut}(H, \mu)$ denotes the automorphism group of H preserving μ . Namely

$$\text{Aut}(H, \mu) = \{f \in \text{Aut } H; f(u) \cdot f(v) = u \cdot v \text{ for any } u, v \in H\}.$$

Let us study how this condition can be expressed in terms of that of matrices representing elements of $\text{Aut}(H, \mu)$. For this, choose a basis $x_1, \dots, x_g, y_1, \dots, y_g$ of H such that

$$\begin{aligned} x_i \cdot y_j &= \delta_{ij}, \\ x_i \cdot x_j &= y_i \cdot y_j = 0 \quad (i, j = 1, \dots, g). \end{aligned}$$

A basis with this property is called a *symplectic basis*. It is easy to see that there exist infinitely many such bases because

$$\begin{aligned} \langle x_1, \dots, x_g, y_1, \dots, y_g \rangle \text{ is a symplectic basis} &\implies \\ \langle x_1 + ny_1, x_2, \dots, x_g, y_1, \dots, y_g \rangle \text{ is also a symplectic basis} \end{aligned}$$

for any $n \in \mathbb{Z}$, for example.

Now if we fix a symplectic basis of H , then the automorphism group $\text{Aut } H$ can be identified as

$$\text{Aut } H = \text{GL}(n, \mathbb{Z})$$

by associating to each element in $\text{Aut } H$ the corresponding matrix with respect to the given symplectic basis. More precisely, first we express any two elements $u, v \in H$ as linear combinations

$$\begin{aligned} u &= u_1x_1 + \cdots + u_gx_g + u_{g+1}y_1 + \cdots + u_{2g}y_g, \\ v &= v_1x_1 + \cdots + v_gx_g + v_{g+1}y_1 + \cdots + v_{2g}y_g \end{aligned}$$

with respect to the above symplectic basis and then we identify the two elements u, v with the following $2g$ -dimensional column vectors

$$u = {}^t(u_1, \dots, u_g, u_{g+1}, \dots, u_{2g}), \quad v = {}^t(v_1, \dots, v_g, v_{g+1}, \dots, v_{2g})$$

in \mathbb{R}^{2g} . Now set

$$J = \begin{pmatrix} O & E \\ -E & O \end{pmatrix} \in \text{GL}(2g, \mathbb{Z}).$$

Then we can write

$$u \cdot v = u_1v_{g+1} + \cdots + u_gv_{2g} - u_{g+1}v_{2g} - \cdots - u_{2g}v_g = (u, Jv)$$

where (u, Jv) denotes the standard Euclidean inner product of two vectors $u, Jv \in \mathbb{R}^{2g}$. Now a matrix $A \in \text{GL}(2g, \mathbb{Z})$ preserves the intersection pairing μ if and only if the condition

$$Au \cdot Av = u \cdot v \quad \text{for any } u, v \in H \tag{6.2}$$

holds. On the other hand we have

$$\begin{aligned} Au \cdot Av &= (Au, JAv) = (u, {}^tAJAv), \\ u \cdot v &= (u, Jv). \end{aligned}$$

It follows that A satisfies the condition (6.2) if and only if

$${}^tAJA = J.$$

Based on the above consideration, we define a subgroup

$$\text{Sp}(2g, \mathbb{Z}) = \{A \in \text{GL}(2g, \mathbb{Z}); {}^tAJA = J\}$$

of $\text{GL}(2g, \mathbb{Z})$. This group is a discrete subgroup of the symplectic group $\text{Sp}(2g, \mathbb{R})$ consisting of *unimodular* symplectic matrices. Sometimes the group $\text{Sp}(2g, \mathbb{Z})$ is

called the *Siegel modular group* because it plays a fundamental role in the theory of Siegel modular forms.

In conclusion, if we fix a symplectic basis of H , then we have an isomorphism

$$\text{Aut}(H, \mu) \cong \text{Sp}(2g, \mathbb{Z})$$

and (6.1) induces a homomorphism

$$\rho_0: \mathcal{M}_g \longrightarrow \text{Sp}(2g, \mathbb{Z}).$$

It is easy to see that $\text{Sp}(2, \mathbb{Z}) = \text{SL}(2, \mathbb{Z})$ so that ρ_0 is an isomorphism for the case $g = 1$. In the cases where $g \geq 2$, it was classically known, going back to a work of Burkhardt at the end of the 19th century and later works of Dehn and Nielsen, that ρ_0 is *surjective* (cf. [55]). It was recognized that ρ_0 has a non-trivial kernel which is a normal subgroup of \mathcal{M}_g . This group was named the *Torelli group* after an Italian mathematician and was known for some time among complex analysts and algebraic geometers. However it was relatively recently that the Torelli group called the attention of topologists. Probably Birman's paper [6] published in 1971 is the earliest work on this group by topologists. Then Johnson began a systematic study of this group in the late 1970s and obtained foundational results concerning the structure of the Torelli group within several years. We refer the readers to his survey paper [37] as well as [36], [38] [39], [40]. Following his notation, the Torelli group is usually denoted by \mathcal{I}_g . Thus we have a group extension

$$1 \longrightarrow \mathcal{I}_g \longrightarrow \mathcal{M}_g \xrightarrow{\rho_0} \text{Sp}(2g, \mathbb{Z}) \longrightarrow 1.$$

It is a classical result, going back to Grothendieck and Serre, that

the Torelli group \mathcal{I}_g is torsion free.

This can be shown as follows. Suppose that there exists a non-trivial element $\varphi \in \mathcal{I}_g$ which has a finite order, say $d > 0$. Then by Nielsen [78], there exists a diffeomorphism $\tilde{\varphi}: \Sigma_g \rightarrow \Sigma_g$ such that $\tilde{\varphi}^d = \text{id}$ and the mapping class of $\tilde{\varphi}$ is the given one φ . Then consider the quotient $\Sigma_g/G_{\tilde{\varphi}}$ of Σ_g divided by the action of the cyclic group $G_{\tilde{\varphi}} \cong \mathbb{Z}/d$ generated by the element $\tilde{\varphi}$. It is easy to see that this quotient space is homeomorphic to a closed surface of some genus h , because the projection $\Sigma_g \rightarrow \Sigma_g/G_{\tilde{\varphi}}$ must be a branched covering along a finite set consisting of fixed points of $\tilde{\varphi}$ on Σ_g . Now there is a general fact on the rational cohomology of the quotient space X/G divided by a properly discontinuous action of a discrete group G on X , due originally to Grothendieck, that $H^*(X/G; \mathbb{Q}) \cong H^*(X; \mathbb{Q})^G$. If we apply this to the above simplest case of a finite group action, we obtain isomorphisms

$$H^*(\Sigma_h; \mathbb{Q}) \cong H^*(\Sigma_g/G_{\tilde{\varphi}}; \mathbb{Q}) \cong H^*(\Sigma_g; \mathbb{Q})^{G_{\tilde{\varphi}}} \cong H^*(\Sigma_g; \mathbb{Q})$$

where the last isomorphism comes from the assumption that $\tilde{\varphi}$ acts trivially on the homology (and hence cohomology) of Σ_g . We can now conclude that $g = h$ which is a contradiction because the condition $g \geq 2$ should imply that $h < g$. Observe

here that the genus 1 surface, namely the torus T^2 admits a free \mathbb{Z}/d action and the quotient space is again diffeomorphic to T^2 .

Now one of the foundational results of Johnson mentioned above is that \mathcal{I}_g is finitely generated for any $g \geq 3$. To prove this, he introduced the following two types of elements of \mathcal{M}_g . One is the *BP-map* (BP for bounding pair) defined as follows. Suppose that there are given two disjoint simple closed curves C and D on Σ_g which satisfy the condition that if we cut Σ_g along C and D , then the resulting surface is disconnected. In other words, the disjoint union $C \cup D$ bounds a subsurface of Σ_g . We say that C and D are a bounding pair. In this case, we call the element

$$\tau_C \tau_D^{-1} \in \mathcal{M}_g$$

the BP-map corresponding to the above bounding pair. The other type is the *BSCC-map* (BSCC for bounding simple closed curve) defined as follows. Suppose that there is given a simple closed curve C on Σ_g such that if we cut Σ_g along C , then the resulting surface is disconnected. In other words, the simple closed curve C bounds a subsurface of Σ_g . We say that C is a bounding simple closed curve. In this case, we call the element

$$\tau_C \in \mathcal{M}_g$$

the BSCC-map corresponding to the bounding simple closed curve C .

In fact, the following important fact holds:

$$\text{BP-map, BSCC-map} \in \mathcal{I}_g. \tag{6.3}$$

To see this, let us study how a Dehn twist τ_C along a simple closed curve C acts on $H_1(\Sigma_g; \mathbb{Z})$. Let $u \in H_1(\Sigma_g; \mathbb{Z})$ be a homology class and choose an oriented curve E on Σ_g which represents u . We can assume that E intersects C transversely at finitely many points. Let us choose an orientation on C and let $v \in H_1(\Sigma_g; \mathbb{Z})$ be the homology class represented by C with this orientation. Locally C divides the regular neighborhood of C into two parts. If we identify a closed regular neighborhood of the oriented C with $S^1 \times [-1, 1]$ according to the given orientation on the surface in such a way that the oriented C is identified with the oriented $S^1 \times \{0\}$, then we can distinguish the above two pieces by calling them the negative and positive sides respectively. Now we count the number of the intersection points $C \cap E$ algebraically by giving $+1$ if the oriented curve E intersects C from negative to positive direction and -1 if E intersects C from positive to negative direction. Let m be the totality of these ± 1 numbers. Then we have

$$\tau_C(u) = u + mv.$$

Observe here that, if we reverse the orientation of C , then both m and v change signs so that the above formula remains unchanged. Now we can check the above fact (6.3) as follows. First let C, D be a BP-pair. Then we can give orientations on them so that the resulting homology classes are the same. On the other hand, in the above computation, we have $\tau_C(u) = \tau_D(u)$ for any u so that $\tau_C \tau_D^{-1}$ acts on the homology

trivially. Next, if C is a BSCC, then clearly the corresponding homology class is trivial, whence the claim follows.

The *genus* of a BP-map (or BSCC-map) is defined as follows. If we cut Σ_g along a BP-pair C, D (or a BSCC C), then we obtain two surfaces. The genus of the relevant map is defined to be the smaller genus of these two surfaces.

Theorem 6.1 (Johnson [38]). *The Torelli group \mathcal{I}_g is finitely generated for any $g \geq 3$.*

The method of proving this theorem was roughly as follows. Johnson constructed a certain finite set of BP-maps of all genera between 1 and $g - 2$ and showed that the subgroup of \mathcal{I}_g generated by them is a *normal* subgroup. Since he had already proved in [35] that \mathcal{I}_g is normally generated by just one BP-map of genus 1, the proof was completed.

Johnson considered also the subgroup

$$\mathcal{K}_g = \text{the subgroup of } \mathcal{M}_g \text{ generated by all the BSCC-maps}$$

of the mapping class group. Since any BSCC-map is contained in \mathcal{I}_g , \mathcal{K}_g is a subgroup of \mathcal{I}_g . Also it is easy to see that property (5.1) implies any conjugate element of a BSCC-map is again a BSCC-map. It follows that \mathcal{K}_g is a normal subgroup of \mathcal{M}_g (and \mathcal{I}_g). More strongly, it is known that \mathcal{K}_g is a characteristic subgroup of \mathcal{M}_g (and \mathcal{I}_g). This follows from a result of Ivanov [31] (see also [57]) that any automorphism of \mathcal{M}_g is induced by an inner automorphism of $\mathcal{D}(\Sigma_g)$ for any $g \geq 3$ and a similar result for the case of the Torelli group \mathcal{I}_g due to Farb and Ivanov [20].

It can be shown that \mathcal{K}_g coincides with \mathcal{I}_g for $g = 2$. However Johnson [39] proved that the quotient $\mathcal{I}_g/\mathcal{K}_g$ is an infinite group for any $g \geq 3$. More precisely, choose a symplectic basis $x_1, \dots, x_g, y_1, \dots, y_g$ of $H = H_1(\Sigma_g; \mathbb{Z})$ as before. The element

$$\omega_0 = x_1 \wedge y_1 + \dots + x_g \wedge y_g \in \Lambda^2 H$$

is called the *symplectic class*. It is known that this element is well defined independent of the choice of symplectic bases. It is easy to see that the mapping

$$H \ni u \longmapsto u \wedge \omega_0 \in \Lambda^3 H$$

is injective. Hence H can be considered as a submodule of $\Lambda^3 H$ so that we can consider the quotient module $\Lambda^3 H/H$. Now Johnson [36] constructed a homomorphism

$$\tau: \mathcal{I}_g \longrightarrow \Lambda^3 H/H$$

and showed that it is surjective and vanishes on the subgroup \mathcal{K}_g . This homomorphism is called now the Johnson homomorphism (see §7 for more details). Later he proved in [39] that $\text{Ker } \tau$ is precisely the subgroup \mathcal{K}_g . Thus we have an extension

$$1 \longrightarrow \mathcal{K}_g \longrightarrow \mathcal{I}_g \xrightarrow{\tau} \Lambda^3 H/H \longrightarrow 1.$$

Because of these basic works, sometimes the group \mathcal{K}_g is called the *Johnson subgroup* or Johnson kernel.

To conclude this section, we would like to make a list which indicates the known results concerning finite generation as well as finite presentability of groups such as the automorphism group of F_n , the IA automorphism group of F_n , the mapping class group, the Torelli group and the Johnson subgroup \mathcal{K}_g (the groups $\text{Out } F_n$ and IOut_n are not included in the list because the results for these groups are the same as $\text{Aut } F_n$ and IAut_n respectively). We refer the readers to the cited original papers and also to the survey papers [84], [10], [34], [37] for details.

Groups	Generators	Relations
$\text{Aut } F_n$	finitely generated Nielsen [76] McCool [58]	finitely presented Nielsen [76] McCool [58]
$\text{Aut}_+ F_n$	Gersten [21]	Gersten [21]
IAut_n	finitely generated Magnus [53]	unknown ($n \geq 4$) not finitely presentable ($n = 3$) Krstić–McCool [48]
\mathcal{M}_g	finitely generated Dehn [17] Lickorish [51] Humphries [30]	finitely presentable McCool [59] Hatcher–Thurston [28] finitely presented Wajnryb [85] Gervais [22]
\mathcal{I}_g	finitely generated ($g \geq 3$) Johnson [38] infinitely generated ($g = 2$) Mess [62]	unknown ($g \geq 3$) infinitesimal finite presentation Hain [23] free group ($g = 2$) Mess [62]
\mathcal{K}_g	infinitely generated Biss–Farb [11] ($g \geq 3$)	unknown ($g \geq 3$)

Problem 1. Complete the above list by filling in the “unknown” blanks.

7 Johnson homomorphisms

In this section, we define so called *Johnson homomorphisms* which give homomorphisms defined on a certain series of subgroups of the mapping class group into certain abelian groups.

In order to do so, we first describe a method of investigating the structure of a given abstract group Γ by approximating it by a series of nilpotent groups. This method is due originally to Malcev [56]. The first approximation is the abelianization Γ^{ab} of Γ . This can be algebraically expressed as follows. Let Γ_1 denote the commutator subgroup of Γ . Namely, it is the subgroup of Γ generated by the commutators $[\gamma_1, \gamma_2] = \gamma_1\gamma_2\gamma_1^{-1}\gamma_2^{-1}$ ($\gamma_1, \gamma_2 \in \Gamma$). It is easy to see that Γ_1 is a normal subgroup of Γ and, as is well known, we have a canonical isomorphism

$$\Gamma^{\text{ab}} \cong \Gamma/\Gamma_1.$$

Next we consider the *second* commutator subgroup Γ_2 which is defined to be the subgroup of Γ generated by the two-fold commutators

$$[\gamma_1, [\gamma_2, \gamma_3]] = \gamma_1[\gamma_2, \gamma_3]\gamma_1^{-1}[\gamma_2, \gamma_3]^{-1} \quad (\gamma_1, \gamma_2, \gamma_3 \in \Gamma).$$

Then it can be checked that Γ_2 is a normal subgroup of Γ and Γ/Γ_2 is a two-step nilpotent group. More precisely, the quotient group Γ_1/Γ_2 is an abelian group and Γ/Γ_2 can be described by the following group extension

$$0 \longrightarrow \Gamma_1/\Gamma_2 \longrightarrow \Gamma/\Gamma_2 \longrightarrow \Gamma/\Gamma_1 = \Gamma^{\text{ab}} \longrightarrow 1 \quad (7.1)$$

which is a *central* extension of $\Gamma/\Gamma_1 = \Gamma^{\text{ab}}$ by Γ_1/Γ_2 . We can continue this procedure to obtain a series of nilpotent groups N_k ($k = 1, 2, \dots$) which approximate Γ as follows, where $N_1 = \Gamma^{\text{ab}}$ and $N_2 = \Gamma/\Gamma_2$. We set $\Gamma_0 = \Gamma$ and for each $k = 1, 2, \dots$, we inductively define

$$\begin{aligned} \Gamma_k &= [\Gamma, \Gamma_{k-1}] \\ &= \text{the subgroup of } \Gamma \text{ generated by } k\text{-fold commutators.} \end{aligned}$$

It can be checked that Γ_k is a normal subgroup of Γ and the series $\{\Gamma_k\}_k$ of normal subgroups of Γ is called the *lower central series* of Γ . It can be easily checked that these subgroups of Γ are all *characteristic subgroups* meaning that any automorphism $f \in \text{Aut } \Gamma$ of Γ preserves them. Now we set $N_k = \Gamma/\Gamma_k$ and call this group the *k-th nilpotent quotient* of Γ . In fact, the quotient $C_k = \Gamma_{k-1}/\Gamma_k$ is easily seen to be an abelian group and furthermore we have a central extension

$$0 \longrightarrow C_k \longrightarrow N_k \longrightarrow N_{k-1} \longrightarrow 1 \quad (k = 2, 3, \dots). \quad (7.2)$$

Hence N_k is a k -step nilpotent group and we obtain an inverse system

$$\dots \longrightarrow N_k \longrightarrow N_{k-1} \longrightarrow \dots \longrightarrow N_3 \longrightarrow N_2 \longrightarrow N_1 = \Gamma^{\text{ab}}$$

of nilpotent groups to which there is a homomorphism from the given group Γ .

Next we ignore any torsion in the above argument and consider everything over \mathbb{Q} . In some sense, this procedure can be understood as taking tensor products with \mathbb{Q} . The first step is straightforward. Namely we just take the usual tensor product of Γ^{ab} with \mathbb{Q}

$$N_1 \otimes \mathbb{Q} = \Gamma^{\text{ab}} \otimes \mathbb{Q}$$

which is a vector space over \mathbb{Q} . For the second step, recall that any central extension of a group Γ by an abelian group C is classified by its extension class which is a second cohomology class defined in $H^2(\Gamma; C)$ (see [12]). If we apply this to the central extension (7.1), we obtain a certain element

$$\chi_2(\Gamma) \in H^2(\Gamma^{\text{ab}}; \Gamma_1/\Gamma_2).$$

Application of the natural homomorphisms $\Gamma^{\text{ab}} \rightarrow \Gamma^{\text{ab}} \otimes \mathbb{Q}$ and $\Gamma_1/\Gamma_2 \rightarrow \Gamma_1/\Gamma_2 \otimes \mathbb{Q}$ to the above element gives

$$\chi_2^{\mathbb{Q}}(\Gamma) \in H^2(\Gamma^{\text{ab}} \otimes \mathbb{Q}; \Gamma_1/\Gamma_2 \otimes \mathbb{Q}).$$

This yields a central extension

$$0 \longrightarrow \Gamma_1/\Gamma_2 \otimes \mathbb{Q} \longrightarrow N_2 \otimes \mathbb{Q} \longrightarrow \Gamma^{\text{ab}} \otimes \mathbb{Q} \longrightarrow 1 \tag{7.3}$$

by which the group $N_2 \otimes \mathbb{Q}$ is defined. We can inductively continue this procedure and we eventually obtain a series of central extensions

$$0 \longrightarrow C_k \otimes \mathbb{Q} \longrightarrow N_k \otimes \mathbb{Q} \longrightarrow N_{k-1} \otimes \mathbb{Q} \longrightarrow 1 \quad (k = 2, 3, \dots). \tag{7.4}$$

In this way, a series of nilpotent groups $\{N_k \otimes \mathbb{Q}\}_k$ is defined. The inverse system

$$\dots \longrightarrow N_k \otimes \mathbb{Q} \longrightarrow N_{k-1} \otimes \mathbb{Q} \longrightarrow \dots \longrightarrow N_3 \otimes \mathbb{Q} \longrightarrow N_2 \otimes \mathbb{Q} \longrightarrow N_1 \otimes \mathbb{Q}$$

of nilpotent groups is called the *Malcev completion* of the given group Γ .

Example 7.1. One of the most important examples of the Malcev completions which appears in the theory of free groups as well as the mapping class group is that of free groups. Here we briefly describe it. Let F_n denote a free group of rank n and we denote the abelianization $H_1(F_n; \mathbb{Z})$ of F_n simply by H which is a free abelian group of rank n . We consider the free graded Lie algebra generated by the elements of H which we denote by

$$\mathcal{L} = \bigoplus_{k=1}^{\infty} \mathcal{L}_k$$

as follows. The degree 1 part \mathcal{L}_1 is defined to be H itself. Then we consider the bracket $[u, v] \in \mathcal{L}_2$ of two elements $u, v \in H$. The skew commutativity $[v, u] = -[u, v]$ of the Lie algebra implies that $\mathcal{L}_2 = \Lambda^2 H$. Next we consider the bracket

$$[,] : \mathcal{L}_1 \otimes \mathcal{L}_2 = H \otimes \Lambda^2 H \longrightarrow \mathcal{L}_3.$$

The Jacobi identity of the Lie algebra forces that the submodule $\Lambda^3 H \subset H \otimes \Lambda^2 H$ (defined by the correspondence $u \wedge v \wedge w \rightarrow u \otimes [v, w] + v \otimes [w, u] + w \otimes [u, v]$)

must vanish under the above map. There are no other constraints so that $\mathcal{L}_3 = H \otimes \Lambda^2 H / \Lambda^3 H$. Going further, the complexity of enumerating all the relations imposed by the structure of the Lie algebra increases. However we can avoid this difficulty by embedding \mathcal{L} into the tensor algebra

$$T^*(H) = \bigoplus_{k=1}^{\infty} H^{\otimes k}.$$

The degree 1 parts of \mathcal{L} and $T^*(H)$ are the same, namely both are H . The second term can be embedded as

$$\mathcal{L}_2 = \Lambda^2 H \ni [u, v] = u \wedge v \longmapsto u \otimes v - v \otimes u \in H^{\otimes 2}.$$

As for the third term, consider the linear mapping

$$\begin{aligned} H \otimes \Lambda^2 H \ni u \otimes [v, w] \\ \longmapsto u \otimes v \otimes w - u \otimes w \otimes v - v \otimes w \otimes u + w \otimes v \otimes u \in H^{\otimes 3}. \end{aligned}$$

It is easy to check that the kernel of this map is precisely $\Lambda^3 H$ so that we obtain an embedding

$$\begin{aligned} \mathcal{L}_3 = H \otimes \Lambda^2 H / \Lambda^3 H \ni u \otimes [v, w] \\ \longmapsto u \otimes v \otimes w - u \otimes w \otimes v - v \otimes w \otimes u + w \otimes v \otimes u \in H^{\otimes 3}. \end{aligned}$$

Then we can inductively define \mathcal{L}_k as the image of the linear mapping

$$H^{\otimes k} \supset H \otimes \mathcal{L}_{k-1} \ni u \otimes \xi \longmapsto u \otimes \xi - \xi \otimes u \in H^{\otimes k}.$$

Thus $\mathcal{L} = \bigoplus_k \mathcal{L}_k$ is realized as a submodule of $T^*(H)$. The elements in \mathcal{L} are called *Lie elements* of $T^*(H)$. Now it is a classical result that \mathcal{L} is isomorphic to the graded module associated to the lower central series of the free group F_n . Namely there exists a canonical isomorphism

$$(F_n)_{k-1} / (F_n)_k \cong \mathcal{L}_k$$

where $(F_n)_k$ denotes the k -th term in the lower central series of F_n (the first one $F_n / (F_n)_1 \cong H$ gives the abelianization). Thus we have a series of central extensions

$$0 \longrightarrow \mathcal{L}_k \longrightarrow N_k(F_n) \longrightarrow N_{k-1}(F_n) \longrightarrow 1 \quad (k = 2, 3, \dots)$$

where $N_k(F_n)$ denotes the k -th nilpotent quotient of F_n . See [55] for details.

Now we define the Johnson homomorphisms. First we begin with the case of automorphism groups of free groups. This case was considered first by Andreadakis [1] before the works of Johnson.

We can define a series $\text{Aut } F_n(k)$ ($k = 1, 2, \dots$) of subgroups of $\text{Aut } F_n$ as follows. Any member $(F_n)_k$ in the lower central series of F_n is a characteristic subgroup in the sense that it is preserved by any automorphism $\varphi \in \text{Aut } F_n$. Hence we obtain a series

of representations

$$p_k : \text{Aut } F_n \longrightarrow \text{Aut } N_k(F_n) \quad (k = 1, 2, \dots).$$

The first one p_1 is nothing but the natural homomorphism

$$\text{Aut } F_n \longrightarrow \text{Aut } N_1(F_n) = \text{GL}(n, \mathbb{Z}).$$

Now we set

$$\text{Aut } F_n(k) = \text{Ker } p_k = \{\varphi \in \text{Aut } F_n; \varphi \text{ acts on } N_k(F_n) \text{ trivially}\}.$$

The first one $\text{Aut } F_n(1)$ is nothing but the subgroup IAut_n .

Now let $\varphi \in \text{Aut } F_n(k)$ be any element. Then for each element $\gamma \in F_n$,

$$\varphi(\gamma)\gamma^{-1} \in (F_n)_k$$

because, by the assumption, φ acts on $N_k(F_n) = F_n/(F_n)_k$ trivially. Consider the image of $\varphi(\gamma)\gamma^{-1}$ in $\mathcal{L}_{k+1} = (F_n)_k/(F_n)_{k+1}$ which we denote by $[\varphi(\gamma)\gamma^{-1}]$. This procedure defines a mapping

$$F_n \ni \gamma \longmapsto [\varphi(\gamma)\gamma^{-1}] \in \mathcal{L}_{k+1}.$$

It can be shown that the above mapping factors through the abelianization of F_n so that we obtain a mapping

$$H = (F_n)^{\text{ab}} \longrightarrow \mathcal{L}_{k+1}.$$

Hence we can define a mapping

$$\tau_k : \text{Aut } F_n(k) \longrightarrow \text{Hom}(H, \mathcal{L}_{k+1}) \quad (7.5)$$

by setting

$$\tau_k(\varphi)([\gamma]) = [\varphi(\gamma)\gamma^{-1}] \quad (\gamma \in F_n).$$

Finally it can be checked that the above mapping (7.5) is in fact a *homomorphism* and this is called the k -th Johnson homomorphism for the automorphism groups of free groups.

By the definition of the homomorphism τ_k , it is easy to see that

$$\text{Ker } \tau_k = \text{Aut } F_n(k+1).$$

Therefore we have an injection

$$\text{Aut } F_n(k)/\text{Aut } F_n(k+1) \subset \text{Hom}(H, \mathcal{L}_{k+1}).$$

If we make the direct sum over k , we obtain an injection

$$\bigoplus_{k=1}^{\infty} \text{Aut } F_n(k)/\text{Aut } F_n(k+1) \subset \bigoplus_{k=1}^{\infty} \text{Hom}(H, \mathcal{L}_{k+1}).$$

It is known that both of the above graded modules have the structure of graded Lie algebras over \mathbb{Z} and it is a very important problem to identify the left hand side as an explicit Lie subalgebra of the right hand side.

Problem 2. Determine the graded module

$$\bigoplus_{k=1}^{\infty} \text{Aut } F_n(k) / \text{Aut } F_n(k+1)$$

associated to the filtration $\{\text{Aut } F_n(k)\}_k$ of the group $\text{Aut } F_n$ as a Lie subalgebra of the graded Lie algebra

$$\bigoplus_{k=1}^{\infty} \text{Hom}(H, \mathcal{L}_{k+1}).$$

Next we consider the case of the mapping class group. Here, for simplicity, we consider only the case of a compact surface $\Sigma_g^0 = \Sigma_g \setminus \text{Int}D^2$ with one boundary component. Then $\Gamma = \pi_1 \Sigma_g^0$ is a free group of rank $2g$. As before, we denote $H_1(\Sigma_g^0; \mathbb{Z})$ simply by H which is a free abelian group of rank $2g$. By Theorem (3.2), the mapping class group $\mathcal{M}_{g,1}$ of Σ_g^0 is a subgroup of $\text{Aut } F_{2g}$. Hence we can define a filtration $\{\mathcal{M}_{g,1}(k)\}_k$ of $\mathcal{M}_{g,1}$ by simply restricting that of $\text{Aut } F_{2g}$ to the subgroup $\mathcal{M}_{g,1}$. The first term $\mathcal{M}_{g,1}(1)$ in this filtration is nothing but the Torelli group $\mathcal{I}_{g,1}$. It turns out that, in the case of the mapping class group, there are important additional structures which do not exist in the case of $\text{Aut } F_{2g}$. First notice that the Poincaré duality theorem for the (co)homology of the surface implies that there is a natural isomorphism

$$H^* = \text{Hom}(H, \mathbb{Z}) = H^1(\Sigma_g^0; \mathbb{Z}) \cong H.$$

It follows that we can replace the target $\text{Hom}(H, \mathcal{L}_{k+1})$ of the k -th Johnson homomorphism (7.5) by $H \otimes \mathcal{L}_{k+1}$. Johnson [36] proved that the image of the first Johnson homomorphism

$$\tau_1: \mathcal{I}_{g,1} \longrightarrow H \otimes \Lambda^2 H$$

is precisely the submodule

$$\Lambda^3 H \subset H \otimes \Lambda^2 H.$$

Generalizing this fact, it was proved in [66] that the target of τ_k can be narrowed, for any k , as follows. We define a submodule \mathcal{H}_k of $H \otimes \mathcal{L}_{k+1}$ by setting

$$\mathcal{H}_k = \text{Ker}(H \otimes \mathcal{L}_{k+1} \xrightarrow{[\cdot, \cdot]} \mathcal{L}_{k+2})$$

where $[\cdot, \cdot]$ denotes the bracket operation in the graded Lie algebra \mathcal{L} . It can be checked that the graded submodule

$$\bigoplus_{k=1}^{\infty} \mathcal{H}_k \subset \bigoplus_{k=1}^{\infty} H \otimes \mathcal{L}_{k+1}$$

is a graded Lie subalgebra.

Problem 3. Determine the graded module

$$\bigoplus_{k=1}^{\infty} \mathcal{M}_{g,1}(k) / \mathcal{M}_{g,1}(k+1)$$

associated to the filtration $\{\mathcal{M}_{g,1}(k)\}_k$ of the mapping class group $\mathcal{M}_{g,1}$ as a Lie subalgebra of the graded Lie algebra

$$\bigoplus_{k=1}^{\infty} \mathcal{H}_k.$$

A similar problem for the usual mapping class group \mathcal{M}_g can be formulated by making use of the result of Labute [49].

8 Teichmüller space and Outer Space

There are two important spaces on which the mapping class group \mathcal{M}_g and the outer automorphism group $\text{Out } F_n$ of a free group act canonically. One is the classical *Teichmüller space*, introduced by Teichmüller in the 1930s, and the other is the *Outer Space* defined in the 1980s by Culler and Vogtmann [16]. Here we briefly describe the definitions of them which can be given in parallel with each other.

The Teichmüller space of Σ_g , denoted by \mathcal{T}_g , is defined to be the space of all the orientation preserving diffeomorphisms

$$f: \Sigma_g \longrightarrow M$$

from Σ_g to compact Riemann surfaces M of genus g divided by a certain equivalence relation. More precisely

$$\mathcal{T}_g = \{f: \Sigma_g \rightarrow M; M \text{ is a Riemann surface of genus } g\} / \sim$$

where two orientation preserving diffeomorphisms

$$f: \Sigma_g \longrightarrow M, \quad f': \Sigma_g \longrightarrow M'$$

are equivalent if there exists an isomorphism $h: M \rightarrow M'$ of Riemann surfaces (namely a biholomorphism) such that the following diagram

$$\begin{array}{ccc} \Sigma_g & \xrightarrow{f} & M \\ \parallel & & \downarrow h \\ \Sigma_g & \xrightarrow{f'} & M' \end{array}$$

is homotopy commutative. Let $[f: \Sigma_g \rightarrow M] \in \mathcal{T}_g$ denote the equivalence class represented by $f: \Sigma_g \rightarrow M$. If we pull back the complex structure on M by the

diffeomorphism f , we obtain a complex structure on Σ_g whose induced orientation coincides with the given orientation on Σ_g . Since homotopic diffeomorphisms of Σ_g are mutually isotopic, the isotopy class of the pull back complex structure on Σ_g depends only on the element $[f: \Sigma_g \rightarrow M]$. Thus we can also write

$$\mathcal{T}_g = \{\text{isotopy classes of complex structure on } \Sigma_g\}.$$

In the cases where $g \geq 2$, a complex structure on Σ_g is the same as a hyperbolic structure (that is a Riemannian structure with constant negative curvature -1) by the classical uniformization theorem. Therefore we have yet one more description:

$$\mathcal{T}_g = \{\text{isotopy classes of hyperbolic structure on } \Sigma_g\} \quad (g \geq 2).$$

The space of all the complex (or hyperbolic) structures on Σ_g has a natural C^∞ topology and it induces a topology on \mathcal{T}_g . It is a classical result of Teichmüller that \mathcal{T}_g is homeomorphic to \mathbb{R}^{6g-6} for any $g \geq 2$. The mapping class group \mathcal{M}_g acts on \mathcal{T}_g from the right by

$$\mathcal{T}_g \times \mathcal{M}_g \ni ([f: \Sigma_g \rightarrow M], \varphi) \longmapsto [f \circ \tilde{\varphi}: \Sigma_g \rightarrow M] \in \mathcal{T}_g$$

where $\tilde{\varphi} \in \text{Diff}_+ \Sigma_g$ is a lift of $\varphi \in \mathcal{M}_g$. This action is known to be properly discontinuous. The quotient space

$$\mathbf{M}_g = \mathcal{T}_g / \mathcal{M}_g$$

is called the *moduli space of Riemann surfaces* of genus g which consists of all the isomorphism classes of genus g Riemann surfaces.

The Outer Space X_n , which is an analogue of the Teichmüller space where we replace \mathcal{M}_g with $\text{Out } F_n$, was defined by Culler and Vogtmann [16] as follows. A *metric graph* Γ is a graph (one dimensional finite complex) such that (i) the valencies at vertices are all ≥ 3 and (ii) there is given a length on every edge such that the sum is equal to 1. Let R_n denote the wedge of labeled n circles S_i^1 ($i = 1, \dots, n$) so that there is given an isomorphism $\pi_1 R_n \cong F_n$. Then the Outer Space (of rank n) is defined as the set of homotopy equivalences $f: R_n \rightarrow \Gamma$ from R_n to metric graphs Γ divided by a certain equivalence relation. More precisely

$$X_n = \{f: R_n \rightarrow \Gamma; \Gamma \text{ is a metric graph with } \pi_1 \Gamma \cong F_n\} / \sim$$

where two homotopy equivalences

$$f: R_n \longrightarrow \Gamma, \quad f': R_n \longrightarrow \Gamma'$$

are equivalent if there exists an isometry $h: \Gamma \rightarrow \Gamma'$ of metric graphs such that the following diagram

$$\begin{array}{ccc} R_n & \xrightarrow{f} & \Gamma \\ \parallel & & \downarrow h \\ R_n & \xrightarrow{f'} & \Gamma' \end{array}$$

is homotopy commutative. Let $[f: R_n \rightarrow \Gamma] \in X_n$ denote the equivalence class represented by $f: R_n \rightarrow \Gamma$. The group $\text{Out } F_n$ acts on X_n from the right by

$$X_n \times \text{Out } F_n \ni ([f: R_n \rightarrow \Gamma], \varphi) \mapsto [f \circ \tilde{\varphi}: R_n \rightarrow \Gamma] \in X_n$$

where $\tilde{\varphi}: R_n \rightarrow R_n$ is a homotopy equivalence which represents $\varphi \in \text{Out } F_n$. There is a natural topology on X_n and Culler and Vogtmann proved that X_n is contractible and the above action is properly discontinuous. The quotient space

$$\mathbf{G}_n = X_n / \text{Out } F_n$$

is called the *moduli space of metric graphs* of rank n which consists of all the isometry classes of metric graphs of rank n . We refer to the survey paper [84] by Vogtmann and also Bestvina [5] for recent results concerning the Outer Space as well as $\text{Out } F_n$.

In general, it is a very important problem to determine the (co)homology groups of the moduli spaces associated to various geometrical objects. In the above, we have the moduli spaces of Riemann surfaces and the moduli space of metric graphs. There have been obtained many results concerning the cohomology of these moduli spaces (we refer to the survey papers [26], [45], [83], [84], [71] as well as original papers [25], [73], [64], [65], [63], [52], [29], [15]). However the cohomological structures of them are far from being very well understood.

Problem 4. Study the (co)homology groups of the moduli space \mathbf{M}_g of Riemann surfaces and the moduli space \mathbf{G}_n of metric graphs.

9 Symplectomorphism groups of surfaces

As in §1, let $\text{Diff}_+ \Sigma_g$ denote the orientation preserving diffeomorphism group of Σ_g . Let us choose an area form ω on Σ_g . Then, by the dimension reason, it can be considered also as a *symplectic form* on Σ_g . We denote by

$$\text{Symp } \Sigma_g = \{\varphi \in \text{Diff}_+ \Sigma_g; \varphi^* \omega = \omega\}$$

the subgroup of $\text{Diff}_+ \Sigma_g$ consisting of those diffeomorphisms which preserve the form ω . We call it the symplectomorphism group of the symplectic manifold (Σ_g, ω) or the orientation and area preserving diffeomorphism group of Σ_g with respect to the area form ω .

As was already mentioned in §1, in general, the volume preserving diffeomorphism group $\text{Diff}^v M$ of a C^∞ manifold M with respect to a given volume form ν and also the symplectomorphism group $\text{Symp}(M, \omega)$ of a symplectic manifold (M, ω) are both very important objects of geometry and topology. Recently there has been rapid progress in a topological study of symplectic manifolds, under the name of *symplectic topology* (see [61] for foundations and generalities of this theory).

The case of surfaces is the simplest one. However it is at the same time very important because the symplectic and the volume preserving contexts are the same in

this case so that we can expect very rich structures here. In the following, we would like to describe one feature of these structures, namely the one which is induced by a basic concept in symplectic topology called the flux homomorphism.

If we apply Moser’s theorem in [72], mentioned in §2, to $\text{Symp } \Sigma_g$, we can conclude that the inclusion

$$\text{Symp } \Sigma_g \subset \text{Diff}_+ \Sigma_g$$

is a homotopy equivalence. It follows that the symplectic mapping class group $\mathcal{SD}(\Sigma_g, \omega)$ of the symplectic manifold (Σ_g, ω) can be canonically identified with the usual mapping class group \mathcal{M}_g and we obtain the following exact sequence

$$1 \longrightarrow \text{Symp}_0 \Sigma_g \longrightarrow \text{Symp } \Sigma_g \longrightarrow \mathcal{M}_g \longrightarrow 1.$$

In particular, the natural homomorphism $\text{Symp } \Sigma_g \rightarrow \mathcal{M}_g$ is surjective. Let us see this fact somewhat more explicitly. One form of Moser’s theorem cited above can be stated as follows. Let M be a closed oriented C^∞ manifold and let ν, ν' be any two volume forms on M . Then there exists a diffeomorphism φ of M , which can be chosen to be isotopic to the identity, such that $\nu' = c\varphi^*\nu$ where c is a constant defined by

$$\int_M \nu' = c \int_M \nu.$$

Now let $\varphi \in \mathcal{M}_g$ be any element and let $\tilde{\varphi} \in \text{Diff}_+ \Sigma_g$ be its lift. Consider the form $\tilde{\varphi}^*\omega$ which is another area form on Σ_g . Hence by the above theorem of Moser, there exists an element $\psi \in \text{Diff}_+ \Sigma_g$, which is isotopic to the identity, such that $\tilde{\varphi}^*\omega = \psi^*\omega$. If we set $\tilde{\varphi}' = \tilde{\varphi}\psi^{-1}$, then $(\tilde{\varphi}')^*\omega = \omega$ so that $\tilde{\varphi}'$ belongs to $\text{Symp } \Sigma_g$. On the other hand, since ψ is isotopic to the identity, the projection of $\tilde{\varphi}'$ to \mathcal{M}_g is the same as that of $\tilde{\varphi}$ which is the given element φ . We can now conclude that the mapping $\text{Symp } \Sigma_g \rightarrow \mathcal{M}_g$ is surjective as required. It might be amusing to observe here that the Dehn twist along a simple closed curve C on Σ_g , defined in §5, preserves any area form on Σ_g whose restriction to a cylindrical neighborhood of C is equal to the 2-form $d\theta \wedge dt$.

Now we describe the *flux homomorphism* briefly (see [61] for details). It is defined for a general symplectic manifold (M, ω) . Let $\text{Symp}_0(M, \omega)$ denote the identity component of $\text{Symp}(M, \omega)$ as before. Then the flux homomorphism is a homomorphism

$$\text{Flux} : \widetilde{\text{Symp}}_0(M, \omega) \longrightarrow H^1(M; \mathbb{R}) \tag{9.1}$$

from the universal covering group of $\text{Symp}_0(M, \omega)$ to the first real cohomology group of M defined as follows. For each element $\varphi \in \text{Symp}_0(M, \omega)$, let $\varphi_t \in \text{Symp}_0(M, \omega)$ be an isotopy such that $\varphi_0 = \text{id}$ and $\varphi_1 = \varphi$. Then

$$\text{Flux}(\{\varphi_t\}) = \int_0^1 i_{\dot{\varphi}_t} \omega dt$$

where $\dot{\varphi}_t$ denotes the vector field associated to φ_t , which is considered as a one-parameter family of transformations of M , and i denotes the interior product. It can be checked that the above value depends only on the homotopy class of the curve $\{\varphi_t\}$

in $\text{Symp}_0(M, \omega)$ with fixed endpoints. Hence we have the induced map described in (9.1). Furthermore it can be checked that Flux is a homomorphism and also that it is surjective. We have an exact sequence

$$1 \longrightarrow \pi_1 \text{Symp}_0(M, \omega) \longrightarrow \widetilde{\text{Symp}}_0(M, \omega) \longrightarrow \text{Symp}_0(M, \omega) \longrightarrow 1$$

and the subgroup

$$\Gamma_\omega = \text{Flux}(\pi_1 \text{Symp}_0(M, \omega)) \subset H^1(M; \mathbb{R})$$

is called the *flux group*. Then (9.1) induces the following homomorphism which is also called the flux homomorphism

$$\text{Flux} : \text{Symp}_0(M, \omega) \longrightarrow H^1(M; \mathbb{R}) / \Gamma_\omega. \tag{9.2}$$

Very recently, Ono [79] proved a long standing conjecture that the flux group is a *discrete* subgroup of $H^1(M; \mathbb{R})$ for any compact symplectic manifold M .

Now in our case of surfaces, by Moser’s theorem $\text{Symp}_0 \Sigma_g$ is homotopy equivalent to $\text{Diff}_+ \Sigma_g$ which in turn is known by Earle and Eells [18] to be homotopy equivalent to T^2 for the case $g = 1$ and contractible for any $g \geq 2$. Hence we obtain homomorphisms

$$\begin{aligned} \text{Flux} : \text{Symp}_0 T^2 &\longrightarrow H^1(T^2; \mathbb{R}) / H^1(T^2; c\mathbb{Z}), \\ \text{Flux} : \text{Symp}_0 \Sigma_g &\longrightarrow H^1(\Sigma_g; \mathbb{R}) \quad (g \geq 2) \end{aligned}$$

where c denotes the total area of T^2 with respect to ω .

10 Extensions of the Johnson homomorphism and the flux homomorphism

Assume that a group G acts on a module M by automorphisms. In other words, suppose that a homomorphism

$$G \longrightarrow \text{Aut } M$$

is given. Then we can give the direct product $M \times G$ a natural structure of a group by setting

$$(m, g)(n, h) = (m + g(n), gh) \quad (m, n \in M, g, h \in G).$$

The resulting group is denoted by $M \rtimes G$ and called the *semi-direct product* of M and G or *split extension* of G by M .

We mention the following three results which have a similar formal nature to each other. The first one is given in [67], where it was proved that the first Johnson homomorphism $\tau_1 : \mathcal{I}_{g,1} \longrightarrow \Lambda^3 H$ can be extended to a homomorphism

$$\rho_1 : \mathcal{M}_{g,1} \longrightarrow \Lambda^3 H_{\mathbb{Q}} \rtimes \text{Sp}(2g, \mathbb{Z})$$

where $H_{\mathbb{Q}} = H \otimes \mathbb{Q}$. Similar results hold for the other types of the mapping class groups \mathcal{M}_g and $\mathcal{M}_{g,*}$. The second one is due to Kawazumi [41] who proved, among other things, that the first Johnson homomorphism $\tau_1 : \text{IAut}_n \rightarrow H^* \otimes \Lambda^2 H$ can be extended to a homomorphism

$$\rho_1 : \text{Aut } F_n \rightarrow V_{\mathbb{Q}} \rtimes \text{GL}(n, \mathbb{Z})$$

where $V = H^* \otimes \Lambda^2 H$ and $V_{\mathbb{Q}} = V \otimes \mathbb{Q}$. A similar result holds for $\text{Out } F_n$. The third one is given in [46] where it was proved that the flux homomorphism $\text{Flux} : \text{Symp}_0 \Sigma_g \rightarrow H^1(\Sigma_g; \mathbb{R})$ can be extended to a homomorphism

$$\widetilde{\text{Flux}} : \text{Symp } \Sigma_g \rightarrow H^1(\Sigma_g; \mathbb{R}) \rtimes \text{Sp}(2g, \mathbb{Z}).$$

Thus we have the following three commutative diagrams.

$$\begin{array}{ccccccc} 1 & \longrightarrow & \mathcal{I}_{g,1} & \longrightarrow & \mathcal{M}_{g,1} & \longrightarrow & \text{Sp}(2g, \mathbb{Z}) \longrightarrow 1 \\ & & \downarrow \tau_1 & & \downarrow \rho_1 & & \parallel \\ 1 & \longrightarrow & \Lambda^3 H & \longrightarrow & \Lambda^3 H_{\mathbb{Q}} \rtimes \text{Sp}(2g, \mathbb{Z}) & \longrightarrow & \text{Sp}(2g, \mathbb{Z}) \longrightarrow 1 \end{array}$$

$$\begin{array}{ccccccc} 1 & \longrightarrow & \text{IAut}_n & \longrightarrow & \text{Aut } F_n & \longrightarrow & \text{GL}(n, \mathbb{Z}) \longrightarrow 1 \\ & & \downarrow \tau_1 & & \downarrow \rho_1 & & \parallel \\ 1 & \longrightarrow & V & \longrightarrow & V_{\mathbb{Q}} \rtimes \text{GL}(n, \mathbb{Z}) & \longrightarrow & \text{GL}(n, \mathbb{Z}) \longrightarrow 1 \end{array}$$

$$\begin{array}{ccccccc} 1 & \longrightarrow & \text{Symp}_0 \Sigma_g & \longrightarrow & \text{Symp } \Sigma_g & \longrightarrow & \mathcal{M}_g \longrightarrow 1 \\ & & \downarrow \text{Flux} & & \downarrow \widetilde{\text{Flux}} & & \parallel \\ 1 & \longrightarrow & H_{\mathbb{R}} & \longrightarrow & H_{\mathbb{R}} \rtimes \mathcal{M}_g & \longrightarrow & \mathcal{M}_g \longrightarrow 1 \end{array}$$

where $H_{\mathbb{R}} = H^1(\Sigma_g; \mathbb{R})$.

The first two diagrams can be extended further by considering higher nilpotent quotients of the Torelli group $\mathcal{I}_{g,1}$ and the group IAut_n . However, as for the last one, there is no such extension because the kernel of the flux homomorphism, which is denoted by $\text{Ham } \Sigma_g$ and called the Hamiltonian symplectomorphism group, is known to be perfect by Thurston [82] (see also Banyaga [3] for the generalization of this fact to general symplectic manifolds). There are several results which make use of the above three commutative diagrams, see [68], [42], [43], [70], [41], [46], [47] and references in them. However it seems likely that there should exist further interesting facts to be uncovered along these lines.

Problem 5. Give further applications as well as generalizations of the above three commutative diagrams.

References

- [1] S. Andreadakis, On the automorphisms of free groups and free nilpotent groups. *Proc. London Math. Soc.* 15 (1965), 239–268. [373](#)
- [2] R. Baer, Isotopien auf Kurven von orientierbaren, geschlossenen Flächen. *J. Reine Angew. Math.* 159 (1928), 101–116. [360](#)
- [3] A. Banyaga, Sur la structure du groupe des difféomorphismes qui préservent une forme symplectique. *Comment. Math. Helv.* 53 (1978), 174–227. [357](#), [381](#)
- [4] G. Baumslag and T. Taylor, The centre of groups with one defining relator. *Math. Ann.* 175 (1968), 315–319. [362](#)
- [5] M. Bestvina, The topology of $\text{Out}(F_n)$. In *Proceedings of the International Congress of Mathematicians* (Beijing, 2002), Vol. II, Higher Ed. Press, Beijing 2002, 373–384. [378](#)
- [6] J. Birman, On Siegel’s modular group. *Math. Ann.* 191 (1971), 59–68. [367](#)
- [7] J. Birman, *Braids, Links and Mapping Class Groups*. Ann. of Math. Stud. 82, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo 1974. [353](#)
- [8] J. Birman, Mapping class groups of surfaces: a survey. In *Discontinuous Groups and Riemann Surfaces* (L. Greenberg, ed.), Ann. of Math. Stud. 79, Princeton University Press, Princeton, N.J., 1974, 57–71. [353](#)
- [9] J. Birman, The algebraic structure of surface mapping class groups. In *Discrete Groups and Automorphic Functions* (W. J. Harvey, ed.), Academic Press, London 1977, 163–198. [353](#)
- [10] J. Birman, Mapping class groups of surfaces. In *Braids* (J. Birman and A. Libgober, eds.), Contemp. Math. 78, Amer. Math. Soc., Providence, RI, 1988, 13–43. [353](#), [370](#)
- [11] D. Biss and B. Farb, \mathcal{K}_g is not finitely generated. *Invent. Math.* 163 (2006), 213–226. [370](#)
- [12] K. S. Brown, *Cohomology of Groups*. Grad. Texts in Math. 87, Springer-Verlag, Berlin 1982. [358](#), [372](#)
- [13] J. Cerf, *Sur les difféomorphismes de la sphère de dimension trois* ($\Gamma_4 = 0$). Lecture Notes in Math. 53, Springer-Verlag, Berlin 1968. [355](#)
- [14] J. Cerf, La stratification naturelle des espaces de fonctions différentiables réelles et le théorème de la pseudo-isotopie. *Inst. Hautes Études Sci. Publ. Math.* 39 (1970), 5–173. [355](#)
- [15] J. Conant and K. Vogtmann, Morita classes in the homology of automorphism groups of free groups. *Geom. Topol.* 8 (2004), 1471–1499. [378](#)
- [16] M. Culler and K. Vogtmann, Moduli of graphs and automorphisms of free groups. *Invent. Math.* 84 (1986), 91–119. [376](#), [377](#)
- [17] M. Dehn, Die Gruppe der Abbildungsklassen. *Acta Math.* 69 (1938), 135–206. [365](#), [370](#)
- [18] C. J. Earle and J. Eells, The diffeomorphism group of a compact Riemann surface. *Bull. Amer. Math. Soc.* 73 (1967), 557–559. [380](#)
- [19] D. Epstein, Curves on 2-manifolds and isotopies. *Acta Math.* 115 (1966), 83–107. [360](#)
- [20] B. Farb and N. Ivanov, The Torelli geometry and its applications. *Math. Res. Lett.* 12 (2005), 293–301. [369](#)

- [21] S. M. Gersten, A presentation for the special automorphism group of a free group. *J. Pure Appl. Algebra* 33 (1984), 269–279. [362](#), [370](#)
- [22] S. Gervais, A finite presentation for the mapping class group of a punctured surface. *Topology* 40 (2001), 703–725. [370](#)
- [23] R. Hain, Infinitesimal presentations of the Torelli groups. *J. Amer. Math. Soc.* 10 (1997), 597–651. [370](#)
- [24] R. Hain and E. Looijenga, Mapping class groups and moduli spaces of curves. *Proc. Sympos. Pure Math.* 62 (2) (1997), 97–142. [354](#)
- [25] J. Harer, Stability of the homology of the mapping class group of an orientable surface. *Ann. of Math.* 121 (1985), 215–249. [378](#)
- [26] J. Harer, *The cohomology of the moduli space of curves*. Lecture Notes in Math. 1337, Springer-Verlag, Berlin 1988, 138–221. [354](#), [378](#)
- [27] W. Harvey, Geometric structure of surface mapping class groups. In *Homological Group Theory* (C. T. C. Wall, ed.), London Math. Soc. Lecture Note Ser. 36, Cambridge University Press, Cambridge 1979, 255–269. [353](#)
- [28] A. Hatcher and W. Thurston, A presentation for the mapping class group of a closed orientable surface. *Topology* 19 (1980), 221–237. [365](#), [370](#)
- [29] A. Hatcher and K. Vogtmann, Rational homology of $\text{Aut}(F_n)$. *Math. Res. Lett.* 5 (1998), 759–780. [378](#)
- [30] S. P. Humphries, Generators for the mapping class group. In *Topology of low-dimensional manifolds*, Lecture Notes in Math. 722, Springer-Verlag, Berlin 1979, 44–47. [365](#), [370](#)
- [31] N. Ivanov, Algebraic properties of the Teichmüller modular group. *Soviet Math. Dokl.* 29 (1984), 288–291. [369](#)
- [32] N. Ivanov, Algebraic properties of mapping class groups of surfaces. In *Geometric and algebraic topology*, Banach Center Publ. 18, PWN-Polish Scientific Publishers, Warsaw 1986, 15–35. [354](#)
- [33] N. Ivanov, *Subgroups of Teichmüller Modular Groups*. Transl. Math. Monogr. 115, Amer. Math. Soc., Providence, RI, 1992. [354](#)
- [34] N. Ivanov, Mapping class groups. In *Handbook of Geometric Topology*, North-Holland, Amsterdam 2002, 523–634. [353](#), [370](#)
- [35] D. Johnson, Homeomorphisms of a surface which act trivially on homology. *Proc. Amer. Math. Soc.* 75 (1979), 119–125. [369](#)
- [36] D. Johnson, An abelian quotient of the mapping class group \mathcal{I}_g . *Math. Ann.* 249 (1980), 225–242. [367](#), [369](#), [375](#)
- [37] D. Johnson, A survey of the Torelli group. *Contemp. Math.* 20 (1983), 165–179. [354](#), [367](#), [370](#)
- [38] D. Johnson, The structure of the Torelli group I: A finite set of generators for \mathcal{I}_g . *Ann. of Math.* 118 (1983), 423–442. [367](#), [369](#), [370](#)
- [39] D. Johnson, The structure of the Torelli group II: A characterization of the group generated by twists on bounding simple closed curves. *Topology* 24 (1985), 113–126. [367](#), [369](#)
- [40] D. Johnson, The structure of the Torelli group III: The abelianization of \mathcal{I}_g . *Topology* 24 (1985), 127–144. [367](#)

- [41] N. Kawazumi, Cohomological aspects of the Magnus expansions. Preprint, 2006; arXiv:math.GT/0505497. [381](#)
- [42] 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. Res. Lett.* 3 (1996), 629–641. [381](#)
- [43] N. Kawazumi and S. Morita, The primary approximation to the cohomology of the moduli space of curves and cocycles for the Mumford-Morita-Miller classes. Preprint. [381](#)
- [44] M. Kervaire and J. Milnor, Groups of homotopy spheres, I. *Ann. of Math.* 77 (1963), 504–537. [355](#)
- [45] F. Kirwan, Cohomology of moduli spaces. In *Proceedings of the International Congress of Mathematicians* (Beijing, 2002), Vol. I, Higher Ed. Press, Beijing 2002, 363–382. [378](#)
- [46] D. Kotschick and S. Morita, Signature of foliated surface bundles and the group of symplectomorphisms of surfaces. *Topology* 44 (2005), 131–149. [381](#)
- [47] D. Kotschick and S. Morita, Characteristic classes of foliated surface bundles with area-preserving total holonomy. *J. Differential Geom.* 75 (2007), 273–302. [381](#)
- [48] S. Krstić and J. McCool, The non-finite presentability of $IA(F_3)$ and $GL_2(\mathbb{Z}[t, t^{-1}])$. *Invent. Math.* 129 (1997), 595–606. [370](#)
- [49] J.P. Labute, On the descending central series of groups with a single defining relation. *J. Algebra* 14 (1970), 16–23. [376](#)
- [50] F. Laudenbach, Sur les 2-sphère d’une variété de dimension 3. *Ann. of Math.* 97 (1973), 57–81. [363](#)
- [51] W. Lickorish, A finite set of generators for the homeotopy group of a 2-manifold. *Proc. Cambridge Philos. Soc.* 61 (1964), 769–778. [365](#), [370](#)
- [52] I. Madsen and M. Weiss, The stable moduli space of Riemann surfaces: Mumford’s conjecture. *Ann. of Math.* 165 (3) (2007), to appear; arXiv:math.AT/0212321. [378](#)
- [53] W. Magnus, Über n -dimensionale Gittertransformationen. *Acta Math.* 64 (1934), 353–367. [362](#), [370](#)
- [54] W. Magnus, Über Automorphismen von Fundamentalgruppen berandeter Flächen. *Math. Ann.* 109 (1934), 617–648. [361](#)
- [55] W. Magnus, A. Karrass and D. Solitar, *Combinatorial Group Theory*. Wiley, New York 1966. [367](#), [373](#)
- [56] A. Malcev, On a class of homogeneous spaces. *Izv. Akad. Nauk. SSSR Ser. Mat.* 13 (1949), 9–32; English transl. *Amer. Math. Soc. Transl.* 39 (1962), 775–790. [371](#)
- [57] J. McCarthy, Automorphisms of surface mapping class groups. A recent theorem of N. Ivanov. *Invent. Math.* 84 (1986), 49–71. [369](#)
- [58] J. McCool, A presentation for the automorphism group of a free group of finite rank. *J. London Math. Soc.* (2) 8 (1974), 259–266. [362](#), [370](#)
- [59] J. McCool, Some finitely presented subgroups of the automorphism group of a free group. *J. Algebra* 35 (1975), 205–213. [365](#), [370](#)
- [60] D. McDuff, A survey of topological properties of groups of symplectomorphisms. In *Topology, Geometry and Quantum Field Theory*, Cambridge University Press, Cambridge 2004, 173–193. [357](#)

- [61] D. McDuff and D. A. Salamon, *Introduction to Symplectic Topology*. Second edition, Oxford Math. Monogr., Oxford University Press, New York 1998. [356](#), [378](#), [379](#)
- [62] G. Mess, The Torelli groups for genus 2 and 3 surfaces. *Topology* 31 (1992), 775–790. [370](#)
- [63] E. Y. Miller, The homology of the mapping class group. *J. Differential Geom.* 24 (1986), 1–14. [378](#)
- [64] S. Morita, Characteristic classes of surface bundles. *Bull. Amer. Math. Soc.* 11 (1984), 386–388. [378](#)
- [65] S. Morita, Characteristic classes of surface bundles. *Invent. Math.* 90 (1987), 551–577. [378](#)
- [66] S. Morita, Abelian quotients of subgroups of the mapping class group of surfaces. *Duke Math. J.* 70 (1993), 699–726. [375](#)
- [67] S. Morita, The extension of Johnson’s homomorphism from the Torelli group to the mapping class group. *Invent. Math.* 111 (1993), 197–224. [380](#)
- [68] S. Morita, A linear representation of the mapping class group of orientable surfaces and characteristic classes of surface bundles. In *Topology and Teichmüller spaces* (Katinkulta, 1995), World Scientific, River Edge, N.J., 1996, 159–186. [381](#)
- [69] S. Morita, Structure of the mapping class groups of surfaces: a survey and a prospect. *Geometry and Topology Monographs* 2 (1999), *Proceedings of the Kirbyfest*, 349–406. [354](#)
- [70] S. Morita, Structure of the mapping class group and symplectic representation theory. In *Essays on geometry and related topics*, Vol. 2, Monogr. Enseign. Math. 38, L’Enseignement Math., Geneva 2001, 577–596. [381](#)
- [71] S. Morita, Cohomological structure of the mapping class group and beyond. In *Problems on Mapping Class Groups and Related Topics* (Benson Farb, ed.), Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI, 2006, 329–354. [354](#), [378](#)
- [72] J. Moser, On the volume forms. *Trans. Amer. Math. Soc.* 120 (1965), 286–294. [356](#), [379](#)
- [73] D. Mumford, Towards an enumerative geometry of the moduli space of curves. In *Arithmetic and Geometry*, Progr. Math. 36, (1983), 271–328. [378](#)
- [74] J. Nielsen, Die Isomorphismen der allgemeinen unendlichen Gruppe mit zwei Erzeugenden. *Math. Ann.* 78 (1917), 385–397. [362](#)
- [75] J. Nielsen, Die Gruppe der dreidimensionalen Gittertransformationen. *Danske Vid. Selsk. Math. Fys. Medd.* 12 (1924), 1–29. [362](#)
- [76] J. Nielsen, Die Isomorphismengruppe der freien Gruppen. *Math. Ann.* 91 (1924), 169–209. [362](#), [370](#)
- [77] J. Nielsen, Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen I. *Acta. Math.* 50 (1927), 189–358. [360](#)
- [78] J. Nielsen, Abbildungsklassen endlicher Ordnung. *Acta. Math.* 75 (1942), 23–115. [367](#)
- [79] K. Ono, Floer-Novikov cohomology and the flux conjecture. Preprint 2004. [380](#)
- [80] S. Smale, Diffeomorphisms of the 2-sphere. *Proc. Amer. Math. Soc.* 10 (1959), 621–626. [357](#)

- [81] S. Smale, Generalized Poincaré's conjecture in dimensions greater than four. *Ann. of Math.* 74 (1961), 391–406. [355](#)
- [82] W. Thurston, On the structure of volume-preserving diffeomorphisms. Unpublished manuscript. [381](#)
- [83] R. Vakil, The moduli space of curves and its tautological ring. *Notices Amer. Math. Soc.* 50 (2003), 647–658. [378](#)
- [84] K. Vogtmann, Automorphisms of free groups and Outer Space. *Geom. Dedicata* 94 (2002), 1–31. [354](#), [370](#), [378](#)
- [85] B. Wajnryb, A simple presentation for the mapping class group of an orientable surface. *Israel J. Math.* 45 (1983), 157–174. [365](#), [370](#)
- [86] H. Whitney, Differentiable manifolds. *Ann. of Math.* 37 (1936), 645–680. [357](#)
- [87] H. Zieschang, Discrete groups of motions of the plane and plane group diagrams. *Uspehi Mat. Nauk* 29 (1966), 195–212 (in Russian). [361](#)

Chapter 8

Geometric survey of subgroups of mapping class groups

Lee Mosher

Contents

1	Introduction	387
2	The action of \mathcal{MCG} on compactified Teichmüller space	389
3	The classification of mapping classes	392
4	The Tits alternative and abelian subgroups	397
5	Limit sets	398
6	Convex cocompact subgroups	400
7	Teichmüller discs and their stabilizers	404
8	Surface subgroups and the Leininger–Reid combination theorem	407
	References	408

1 Introduction

Let S be a finite type surface, obtained from a closed oriented surface of genus $g \geq 0$ by removing $p \geq 0$ punctures. The (orientation preserving) *mapping class group* of S is denoted $\mathcal{MCG} = \mathcal{MCG}(S) = \text{Homeo}_+(S)/\text{Homeo}_0(S)$ where $\text{Homeo}_+(S)$ is the group of orientation preserving homeomorphisms and $\text{Homeo}_0(S)$ is the normal subgroup of homeomorphisms isotopic to the identity. The *Teichmüller space* of S , denoted $\mathcal{T} = \mathcal{T}(S)$, is the space of conformal structures on S modulo isotopy, or equivalently the space of complete, finite area hyperbolic structures on S modulo isotopy; the phrase “modulo isotopy” means “modulo the action of $\text{Homeo}_0(S)$ ”. Thurston’s boundary of Teichmüller space, denoted $\mathcal{PMF} = \mathcal{PMF}(S)$, is the space of projective Whitehead classes of measured foliations on S . There is a natural topology on the set $\overline{\mathcal{T}} = \mathcal{T} \cup \mathcal{PMF}$ which makes it homeomorphic to a ball of dimension $6g - g + 2p$ with interior \mathcal{T} and boundary \mathcal{PMF} . The group \mathcal{MCG} acts naturally on \mathcal{T} and on \mathcal{PMF} , extending to an action by homeomorphisms on $\overline{\mathcal{T}}$.

The theme of this survey is that subgroups of \mathcal{MCG} can be studied via the geometric/dynamical properties of their action on $\overline{\mathcal{T}} = \mathcal{T} \cup \mathcal{PMF}$, just as discrete subgroups of the isometries of hyperbolic space \mathbb{H}^n can be studied via their action on

$\overline{\mathbb{H}}^n = \mathbb{H}^n \cup S_{\infty}^{n-1}$. Recent problem lists with this theme in mind include [38] and [40]. We will survey several topics exemplifying this theme, in the following sections:

§3 The Thurston–Bers classification of mapping classes of \mathcal{MCG} .

§4 The Tits alternative.

§5 Limit sets and domains of discontinuity.

§6 Convex cocompact subgroups.

§7 Stabilizers of Teichmüller discs.

§8 Combination theorems and surface subgroups.

Here is a brief description of each of these topics, to be filled out in the rest of this chapter.

Thurston [43] and Bers [1] classified individual elements of \mathcal{MCG} , or equivalently cyclic subgroups of \mathcal{MCG} , via the geometry and dynamics of their actions on $\overline{\mathcal{T}}$. Finite order mapping classes have fixed points in \mathcal{T} . Infinite order irreducible mapping classes have positive translation distance in \mathcal{T} , which is minimized along a unique Teichmüller geodesic, and they act on $\overline{\mathcal{T}}$ with source-sink dynamics. Finite order mapping classes have fixed points in \mathcal{T} , at which the translation distance of zero is minimized. And infinite order reducible mapping classes have translation distance in \mathcal{T} whose infimum is not realized.

The *Tits alternative*, originally formulated and proved for lattices in semisimple Lie groups, was also formulated and proved for subgroups of \mathcal{MCG} independently by McCarthy [33] and Ivanov [25]. The Tits alternative says that each subgroup of \mathcal{MCG} has either an abelian subgroup of finite index, or a free subgroup of rank ≥ 2 .

Limit sets and domains of discontinuity of subgroups of \mathcal{MCG} were studied by McCarthy–Papadopoulos [34], in analogy to limit sets and domains of discontinuity for discrete subgroups of $\text{Isom}(\mathbb{H}^n)$.

Convex cocompact subgroups of \mathcal{MCG} , including Schottky subgroups of \mathcal{MCG} , were originally studied by Farb and Mosher [9] and further developed by Kent–Leininger [27] and Hamenstädt [17], in analogy to convex cocompact subgroups and Schottky subgroups of $\text{Isom}(\mathbb{H}^n)$.

Teichmüller discs in \mathcal{T} are isometrically embedded copies of the hyperbolic plane, and their stabilizers are analogues in \mathcal{MCG} of Fuchsian subgroups of $\text{Isom}(\mathbb{H}^n)$, which are stabilizers of isometrically embedded copies of \mathbb{H}^2 in \mathbb{H}^n .

The *Leininger–Reid combination theorem* [29] gives a method for building closed surface subgroups of \mathcal{MCG} , by combining certain Teichmüller disc stabilizer subgroups, in analogy to the Maskit combination theorem and related results for building discrete subgroups of $\text{Isom}(\mathbb{H}^n)$ by combining simpler discrete subgroups.

Because of our focus on the geometric/dynamic properties of subgroups of \mathcal{MCG} , we have nothing to say about the very rich and interesting algebraic tools that are used to study subgroups, particularly the Torelli subgroup and the lower central series of \mathcal{MCG} .

This work was partially supported by NSF Grant No. 0405979.

2 The action of \mathcal{MCG} on compactified Teichmüller space

We start by establishing what will be the basic notions and notations for this survey. References for this material include [24] and [10].

Teichmüller space and the Thurston boundary. An *essential curve* on S is a simple closed curve which bounds neither a disc nor a once-punctured disc. The isotopy class of an essential curve c is denoted $[c]$. Let \mathcal{C} be the set of isotopy classes of essential curves on S . An *essential curve family* is a finite set of pairwise disjoint essential curves, no two of which are isotopic. The *Teichmüller space* of S , denoted $\mathcal{T}(S)$, is the set of complete, finite area hyperbolic structures on S , or equivalently the set of conformal structures with removable singularities at the punctures, modulo isotopy. Isotopy classes, in \mathcal{C} or in \mathcal{T} or in other contexts, will be denoted formally with square brackets $[\cdot]$, although informally these brackets will often be dropped, as in the next sentence. There is an embedding $\mathcal{T}(S) \rightarrow \mathbb{R}^{\mathcal{C}}$ defined by assigning to each $\sigma \in \mathcal{T}$ and $c \in \mathcal{C}$ the minimal length of curves isotopic to c in the hyperbolic structure σ . The topology on $\mathcal{T}(S)$ is defined so that the map $\mathcal{T}(S) \rightarrow \mathbb{R}^{\mathcal{C}}$ is a homeomorphism onto its image. The group $\mathcal{MCG}(S)$ acts naturally on \mathcal{C} and on $\mathcal{T}(S)$, and so the embedding $\mathcal{T}(S) \rightarrow \mathbb{R}^{\mathcal{C}}$ is \mathcal{MCG} -equivariant.

A *quadratic differential* on $\sigma \in \mathcal{T}$ is, formally, a holomorphic section of the symmetric square of the cotangent bundle of σ , whose integral is finite. Such sections can be added and multiplied by complex scalars, and so the quadratic differentials on σ form a vector space QD_{σ} over the complex numbers. Informally, to each conformal coordinate z for σ , a quadratic differential θ assigns an expression $f(z)dz^2$, where $f(z)$ is holomorphic and such that the expression behaves correctly under coordinate change. For $p \in S$ outside of a finite singular set, θ has a local coordinate z in which its expression is dz^2 and $z(p) = 0$, and z is unique up to multiplication by ± 1 . On each point of the singular set, including each puncture, θ has a local coordinate z in which its expression is $z^{k-2}dz^2$ for some integer k , where $k \geq 3$ if p is not a puncture and $k \geq 1$ if p is a puncture, and this coordinate z is well-defined up to multiplication by a k th root of unity. The *horizontal and vertical measured foliations* of θ are defined in each regular canonical coordinate $z = x + iy$ as follows: the horizontal foliation has leaves on which y is constant, with transverse measure $|dx|$; the vertical foliation has leaves on which x is constant, with transverse measure $|dy|$. At a singularity p near which θ has the canonical expression $z^{k-2}dz^2$ we say that the horizontal and vertical measured foliations have *k -pronged singularities* at p .

Teichmüller's theorem says that for any $\sigma, \sigma' \in \mathcal{T}$ there exist unique choice of quadratic differentials θ on σ called the *initial quadratic differential*, θ' on σ' called the *terminal quadratic differential*, and a number $d \geq 0$, such that if $z = x + iy$ is a regular local coordinate for θ then (up to isotopy of σ') the expression $z' = e^d x + i e^{-d} y$ defines a regular local coordinate for θ' . This number d is called the *Teichmüller distance* between σ and σ' , and it defines a metric on $\mathcal{T}(S)$ which gives the same topology as that defined above.

Composing the embedding $\mathcal{T}(S) \rightarrow \mathbb{R}^{\mathcal{C}}$ with the projectivization $\mathbb{R}^{\mathcal{C}} \rightarrow P\mathbb{R}^{\mathcal{C}}$, the composition $\mathcal{T}(S) \rightarrow P\mathbb{R}^{\mathcal{C}}$ is still an embedding, and its image is precompact. The closure of the image minus the image defines *Thurston's boundary* for $\mathcal{T}(S)$, and this boundary is characterized as follows. A *measured foliation* on S is a foliation with a finite singular set, equipped with a positive transverse Borel measure, so that at each singularity p including punctures there exists $k \geq 1$ such that the singularity is locally modelled on a k -pronged singularity of $z^{k-2}dz^2$, where $k \geq 3$ if p is not a puncture. There is an equivalence relation on measured foliations generated by isotopy and by the *Whitehead move*, a move which collapses to a point each leaf segment whose endpoints are a pair of singularities at least one of which is not a puncture. The equivalence classes are called *Whitehead classes*, and the set of Whitehead classes of measured foliations is denoted \mathcal{MF} (again we use $[\cdot]$ to denote Whitehead class, except that the notation is often dropped). There is an embedding $\mathcal{MF} \rightarrow \mathbb{R}^{\mathcal{C}}$ defined by assigning to each $\mathcal{F} \in \mathcal{MF}$ and $c \in \mathcal{C}$ the number $\langle \mathcal{F}, c \rangle \geq 0$ which is the minimal integral of curves homotopic to c with respect to the transverse measure on \mathcal{F} . The topology on \mathcal{MF} is chosen so that the map $\mathcal{MF} \rightarrow \mathbb{R}^{\mathcal{C}}$ is a homeomorphism onto its image. When the transverse measure on \mathcal{F} is multiplied by a number $r > 0$ we denote the result $r\mathcal{F}$, and this generates the relation of *projective equivalence* on \mathcal{MF} . The space of projective equivalence classes is denoted \mathcal{PMF} , and the resulting map $\mathcal{PMF} \rightarrow P\mathbb{R}^{\mathcal{C}}$ is an embedding.

Thurston's compactification theorem says that the disjoint union of \mathcal{T} and \mathcal{PMF} embeds into $P\mathbb{R}^{\mathcal{C}}$, the image is a closed ball of dimension $6g - 6 + 2p$, the interior of this ball is the image of \mathcal{T} , and its boundary sphere is the image of \mathcal{PMF} . There is a natural action of \mathcal{MCG} on \mathcal{PMF} with respect to which the embedding $\mathcal{PMF} \rightarrow P\mathbb{R}^{\mathcal{C}}$ is equivariant, and hence \mathcal{MCG} acts homeomorphically on the ball $\overline{\mathcal{F}} = \mathcal{T} \cup \mathcal{PMF}$.

Consider an essential closed curve $c \in \mathcal{C}$. The *enlargement* of c is a measured foliation \mathcal{F}_c well-defined up to Whitehead equivalence as follows. First one chooses a spine for $S - c$ containing all punctures. The complement of this spine is an open annulus with core c . Then one foliates this annulus by curves isotopic to c , which together with the spine defines \mathcal{F}_c as a singular foliation. The transverse measure on \mathcal{F}_c is defined so that a curve in the annulus from boundary to boundary transverse to \mathcal{F}_c has total transverse measure 1. This construction generalizes to the *enlargement of a weighted multicurve*, a formal sum $a_1c_1 + \cdots + a_kc_k$ where $\{c_1, \dots, c_k\}$ is an essential curve system and $a_1, \dots, a_k > 0$; in this case one chooses a spine for $S - (c_1 \cup \cdots \cup c_k)$ containing all punctures, whose complement is a union of open annuli with cores c_1, \dots, c_k respectively; then one foliates these annuli by closed curves parallel to the cores; then one assigns total transverse measures a_1, \dots, a_k , respectively, to these foliated annuli. The set of points in \mathcal{PMF} represented by enlargements of weighted multicurves is dense. In fact, the set of points in \mathcal{MF} represented by enlargements of essential closed curves is dense.

The intersection pairing $\mathcal{MF} \times \mathcal{C} \rightarrow [0, +\infty)$ taking (\mathcal{F}, c) to $\langle \mathcal{F}, c \rangle$ extends to a continuous intersection pairing $\mathcal{MF} \times \mathcal{MF} \rightarrow [0, +\infty)$, called the *inter-*

section number of $\mathcal{F}, \mathcal{F}'$; uniqueness of the extension follows from denseness of \mathcal{C} in \mathcal{MF} .

A pair of measured foliations $\mathcal{F}, \mathcal{F}'$ is said to *jointly fill* the surface S if for every $c \in \mathcal{C}$ either $\langle \mathcal{F}, c \rangle$ or $\langle \mathcal{F}', c \rangle$ is nonzero, and $\langle \mathcal{F}, \mathcal{F}' \rangle$ is also nonzero. This happens if and only if there exists a conformal structure $\sigma \in \mathcal{T}$ and a quadratic differential θ on σ such that $\mathcal{F}, \mathcal{F}'$ are Whitehead equivalent to the horizontal and vertical measured foliations of θ ; moreover, σ and θ are uniquely determined by $\mathcal{F}, \mathcal{F}'$.

The set of geodesic lines in \mathcal{T} for the Teichmüller metric can be parameterized as follows. Choose measured foliations $\mathcal{F}, \mathcal{F}'$ which jointly fill. For each $r \in \mathbb{R}$, the measured foliations $e^r \mathcal{F}, e^{-r} \mathcal{F}'$ jointly fill and determine a conformal structure σ_r . The map $r \mapsto \sigma_r$ is a geodesic embedding $\mathbb{R} \rightarrow \mathcal{T}$. The image of this geodesic depends only on the projective classes $P[\mathcal{F}], P[\mathcal{F}']$, and we say that this is the *Teichmüller geodesic* with directions $P[\mathcal{F}], P[\mathcal{F}']$.

The set of geodesic rays in the Teichmüller metric can be parameterized as follows. First we use a theorem of Hubbard and Masur [21] which says that for each $\sigma \in \mathcal{T}$ and $P[\mathcal{F}] \in \mathcal{PMF}$ there is up to real scalar multiple a unique $\theta \in \text{QD}_\sigma$ whose horizontal measured foliation is in the given projective class $P[\mathcal{F}]$. For each $d > 0$ we can then transform each regular canonical coordinate $z = x + iy$ of θ to get a coordinate $z' = e^d x + i e^{-d} y$ defining an isometric embedding $[0, \infty) \rightarrow \mathcal{T}$. The image of this map is called the *Teichmüller ray with basepoint σ and direction $P[\mathcal{F}]$* .

A measured foliation \mathcal{F} is *arational* if $\langle \mathcal{F}, c \rangle > 0$ for every $c \in \mathcal{C}$. Equivalently, there is no closed loop in S which is everywhere parallel to \mathcal{F} , nor is there any closed path from puncture to puncture which is everywhere parallel to \mathcal{F} . If $\mathcal{F}, \mathcal{F}'$ are arational, and if $\langle \mathcal{F}, \mathcal{F}' \rangle > 0$, then $\mathcal{F}, \mathcal{F}'$ jointly fill S .

Virtual torsion free. Serre [42] originally proved that the group \mathcal{MCG} has a finite index subgroup which is torsion free. For another proof see [25] Corollary 1.5.

Theorem 2.1. *The kernel of the homomorphism $\mathcal{MCG}(S) \rightarrow \text{GL}(H_1(S; \mathbb{Z}/3))$ is torsion free. This kernel has index bounded above by $|\text{GL}(H_1(S; \mathbb{Z}/3))|$ which is finite because $H_1(S; \mathbb{Z}/3)$ has finite rank $\leq 2g + p$ over $\mathbb{Z}/3$.*

Note that the bound in the index of the kernel is exponential in g and p .

Gromov hyperbolic metric spaces. Much of this survey is motivated by analogues in the theory of discrete groups of isometries of Gromov hyperbolic metric spaces, particularly hyperbolic spaces \mathbb{H}^n . We recall some of the notions of this theory, whose details can be found in, for example, [15], [13], [8], [5].

Let X be a metric space. X is *proper* if closed balls are compact. X is a *geodesic metric space* if for any $x, y \in X$ there exists an isometrically embedded path $\gamma: [0, d(x, y)] \rightarrow X$, a *geodesic*, from x to y . We shall use the notation $[x, y]$ to refer to the image of some geodesic. The *Hausdorff distance* between two sets $A, B \subset X$ is the infimum of $\delta > 0$ such that $A \subset N_\delta(B)$ and $B \subset N_\delta(A)$.

A geodesic metric space X is *hyperbolic* in the sense of Gromov if there exists $\delta > 0$ such that for any x, y, z and any geodesics $[x, y]$, $[y, z]$, $[z, x]$ in X , $[x, y]$ is contained in the δ -neighborhood of $[y, z] \cup [z, x]$. The *boundary* ∂X is defined to be the set of geodesic rays modulo the relation of finite Hausdorff distance; given a geodesic ray ρ , let $[\rho]$ denote the corresponding point of ∂X . There is a natural *Gromov topology* on $X \cup \partial X$, whose definition we shall not recount, but we shall mention several important properties:

- If X is a proper geodesic metric space then ∂X and $X \cup \partial X$ are compact.
- Every geodesic ray ρ converges in $X \cup \partial X$ to $[\rho]$.
- If X is proper then for any $x \in X$ and any $\xi \in \partial X$ there exists a geodesic ray ρ based at x and converging to ξ , and for any $\xi, \eta \in \partial X$ there exists a bi-infinite geodesic whose two ends converge to ξ, η .

Of course, the primary example of the Gromov topology is the union of \mathbb{H}^n with its sphere at infinity S_∞^{n-1} . When X is not proper then the compactness properties above can fail.

Another important example for us of a Gromov hyperbolic metric space is the curve complex of S , discussed below.

3 The classification of mapping classes

The discrete cyclic groups of isometries of \mathbb{H}^n or of any proper, geodesic, Gromov hyperbolic metric space X fall into three classes: *elliptic* groups which are finite and which fix a point of X ; *parabolic* groups which are infinite cyclic and which fix a point of ∂X but no point of X ; and *loxodromic* groups which are infinite cyclic and which translate along a bi-infinite geodesic of X .

Thurston discovered an analogous classification of elements of \mathcal{MCG} [43]; see also [10] and [6]. Work of Bers strengthens the analogy in terms of the action on \mathcal{T} [1].

There are two major classifications of a mapping class $\phi \in \mathcal{MCG}$: whether ϕ is finite order; and whether ϕ is *reducible* meaning that there exists an essential curve family \mathcal{C} , called a *reduction family* for ϕ , such that $\phi[\mathcal{C}] = [\mathcal{C}]$. If ϕ is not reducible then it is *irreducible*.

A homeomorphism $\Phi: S \rightarrow S$ is said to be *pseudo-Anosov* if there exists a transverse pair of measured foliations $\mathcal{F}^s, \mathcal{F}^u$ which jointly fill the surface, called the *stable and unstable measured foliations* of Φ , and there exists a number $\lambda > 1$ called the *expansion factor* of Φ , such that Φ takes each leaf of \mathcal{F}^s to a leaf of \mathcal{F}^s , each leaf of \mathcal{F}^u to a leaf of \mathcal{F}^u , Φ contracts the leaves of \mathcal{F}^s by a factor of λ with respect to the \mathcal{F}^u measure, and Φ expands the leaves of \mathcal{F}^u by a factor of λ with respect to the \mathcal{F}^s measure. To say this more briefly, $\Phi(\mathcal{F}^s) = \lambda \mathcal{F}^s$ and $\Phi(\mathcal{F}^u) = \lambda^{-1} \mathcal{F}^u$.

A mapping class $\phi \in \mathcal{MCG}(S)$ is said to be pseudo-Anosov if it is represented by a pseudo-Anosov homeomorphism.

Thurston proved that every mapping class $\phi \in \mathcal{MCG}(S)$ falls into one of three types: finite order, reducible, or pseudo-Anosov; in other words, every infinite order, irreducible mapping class is *pseudo-Anosov*. The first two types are not mutually exclusive, but a mapping class is pseudo-Anosov if and only if it is infinite order and irreducible.

There are several necessary and/or sufficient conditions for the three types of mapping classes, expressed in terms of the geometry and dynamics of the action on compactified Teichmüller space.

Finite order mapping classes. A mapping class $\phi \in \mathcal{MCG}(S)$ is of finite order if and only if it has a fixed point in \mathcal{T} , in which case there exists a conformal structure σ on S and a homeomorphism Φ representing ϕ such that $\Phi(\sigma) = \sigma$. It follows that Φ is a finite order homeomorphism.

A sufficient condition for $\phi \in \mathcal{MCG}(S)$ to be of finite order, proved in [10], is the existence of an arational measured foliation \mathcal{F} such that $\phi[\mathcal{F}] = [\mathcal{F}]$.

It is possible for a finite order mapping class ϕ to be irreducible. This happens if and only if, for Φ a finite order homeomorphism representing ϕ , the quotient orbifold S/Φ is a sphere with three cone points.

There is a bound on the order of a finite order element of $\mathcal{MCG}(S)$, in fact a bound on the order of a finite subgroup $G < \mathcal{MCG}(S)$, depending only on the topology of S . One way to see this is by applying Theorem 2.1, which gives a bound which is exponential in g and p . A linear bound is obtained using the much deeper theorem of Gabai [11] and Casson–Jungreis [7] which says that G is realized as a finite group of homeomorphisms of S , also denoted G . We therefore have $|G| = \chi(S)/\chi(S/G) = (2 - 2g - p)/\chi(S/G)$, where the denominator is the orbifold Euler characteristic of the quotient orbifold S/G . The numerator and denominator of this fraction being negative, $|G|$ is maximized when $\chi(S/G)$ is maximized.

The maximum value of the Euler characteristic of a compact, oriented hyperbolic 2-orbifold is realized by the spherical orbifold with three cone points of angles $\frac{2\pi}{2}$, $\frac{2\pi}{3}$, $\frac{2\pi}{7}$, whose Euler characteristic equals $2 - (1 - \frac{1}{2}) - (1 - \frac{1}{3}) - (1 - \frac{1}{7}) = \frac{1}{42}$. It follows that $|G| \leq 42(2g - 2)$ when S is compact.

When S has one or more punctures there is a slightly better bound, because the maximum value of the Euler characteristic of a noncompact, finite area, oriented hyperbolic 2-orbifold is achieved by the modular space of $\mathrm{SL}(2, \mathbb{Z})$, a once-punctured sphere with two cone points of angles $\frac{2\pi}{2}$, $\frac{2\pi}{3}$ and with Euler characteristic $2 - (1 - \frac{1}{2}) - (1 - \frac{1}{3}) = \frac{1}{6}$, and so $|G| \leq 6(2g - 2 + p)$.

To summarize:

Theorem 3.1. *If G is a finite subgroup of \mathcal{MCG} then*

$$|G| \leq \begin{cases} 84g - 84 & \text{if } p = 0, \\ 12g - 12 + 6p & \text{if } p > 0. \end{cases}$$

□

Pseudo-Anosov mapping classes. Bers proved [1] that a mapping class $\phi \in \mathcal{MCG}(S)$ is pseudo-Anosov if and only if the infimum of $d([\sigma], \phi[\sigma])$ is positive and is achieved by some $[\sigma] \in \mathcal{T}$. In this case, by combining with the fact that the stable and unstable foliations $\mathcal{F}^s, \mathcal{F}^u$ are uniquely ergodic, it follows that the set of points $[\sigma]$ where the infimum is achieved is a bi-infinite geodesic in \mathcal{T} along which ϕ translates, with translation distance equal to $\log(\lambda)$. This geodesic is called the *axis* for ϕ in \mathcal{T} . This axis has two ideal endpoints in \mathcal{PMF} , namely $P[\mathcal{F}^s]$ and $P[\mathcal{F}^u]$. The action of ϕ on $\overline{\mathcal{T}}$ is a *source sink action*, with *source* $P[\mathcal{F}^s]$ and *sink* $P[\mathcal{F}^u]$, meaning that for every point $x \in \overline{\mathcal{T}} - \{P[\mathcal{F}^s], P[\mathcal{F}^u]\}$,

$$\lim_{i \rightarrow -\infty} \phi^i(x) = P[\mathcal{F}^s] \quad \text{and} \quad \lim_{i \rightarrow +\infty} \phi^i(x) = P[\mathcal{F}^u].$$

It follows that the fixed point set of ϕ in $\overline{\mathcal{T}}$ is

$$\text{Fix}(\phi) = \{P[\mathcal{F}^s], P[\mathcal{F}^u]\}.$$

A sufficient (and necessary) condition for $\phi \in \mathcal{MCG}(S)$ to be pseudo-Anosov is the existence of an arational measured foliation \mathcal{F} and $\lambda > 0, \lambda \neq 1$ such that $\phi[\mathcal{F}] = \lambda[\mathcal{F}]$; if this happens then $P[\mathcal{F}] = P[\mathcal{F}^s]$ or $P[\mathcal{F}^u]$. See [10] for the proof.

Reducible mapping classes. If a mapping class ϕ is reducible, and if \mathcal{C} is a reducing family for ϕ , then there exists a representative Φ of ϕ such that $\Phi(\mathcal{C}) = \mathcal{C}$. Given a component S' of $S - \mathcal{C}$, let $n \geq 1$ be the first return time of Φ to S' , meaning that $\Phi^n(S') = S'$ and $\Phi^i(S') \neq S'$ for $0 < i < n$. Then the first return time n and the mapping class of the first return map Φ^n are determined by ϕ , independent of the representative Φ . These mapping classes are called the *component mapping classes* of ϕ relative to \mathcal{C} .

A reducing family \mathcal{C} for a mapping class ϕ is *complete* if all of the component mapping classes of ϕ relative to \mathcal{C} are finite order or pseudo-Anosov. A minimal, complete reducing family \mathcal{C} for ϕ is unique up to isotopy; see for example [4], [18]. If $\mathcal{C}, \mathcal{C}'$ are any two minimal, complete reducing families for ϕ , if Φ, Φ' are representatives of ϕ such that $\Phi(\mathcal{C}) = \mathcal{C}, \Phi'(\mathcal{C}') = \mathcal{C}'$, and if $\Psi \in \text{Homeo}_0(S)$ satisfies $\Psi(\mathcal{C}) = \mathcal{C}'$, then Ψ conjugates the component mapping classes of ϕ relative to \mathcal{C} to the component mapping classes of ϕ relative to \mathcal{C}' . We may therefore speak, in a well-defined manner, of the *component mapping classes of ϕ* , meaning the component mapping classes of ϕ relative to any minimal, complete reducing family.

Bers proved [1] that a mapping class ϕ is infinite order and reducible if and only if the infimum of $d([\sigma], \phi[\sigma])$ is not achieved by any $[\sigma] \in \mathcal{T}$. Moreover, two subcases are distinguished: if the infimum equals zero then each component mapping class of ϕ has finite order; whereas if the infimum is positive then some component mapping class of ϕ is pseudo-Anosov.

A sufficient (and necessary) condition for $\phi \in \mathcal{MCG}(S)$ to be reducible is the existence of a non-arational measured foliation \mathcal{F} such that $\phi[\mathcal{F}] = \mathcal{F}$.

A complete description of the dynamics of a reducible mapping class acting on $\mathcal{P}\mathcal{M}\mathcal{F} = \partial\mathcal{T}$ is obtained in [25].

Independence of pseudo-Anosov homeomorphisms. In a discrete group acting on \mathbb{H}^n , given two loxodromic elements ϕ, ϕ' , either $\text{Fix}(\phi) = \text{Fix}(\phi')$ or $\text{Fix}(\phi) \cap \text{Fix}(\phi') = \emptyset$. The same is true for pseudo-Anosov homeomorphisms. Proofs are given in [25], [34]. We shall give a proof here.

Lemma 3.2. *For any pseudo-Anosov $\phi, \phi' \in \mathcal{MCG}(S)$, either $\text{Fix}(\phi) = \text{Fix}(\phi')$ or $\text{Fix}(\phi) \cap \text{Fix}(\phi') = \emptyset$. In the latter case we say that ϕ, ϕ' are independent.*

Proof. Supposing that $\text{Fix}(\phi) \cap \text{Fix}(\phi') \neq \emptyset$, up to inverses of ϕ, ϕ' and up to isotopy we may assume that $\mathcal{F} := \mathcal{F}^u(\phi) = \mathcal{F}^u(\phi')$, and that $\phi[\mathcal{F}] = \lambda[\mathcal{F}]$ and $\phi'[\mathcal{F}] = \lambda'[\mathcal{F}]$ for some $\lambda, \lambda' > 1$. Let Γ be the intersection of $\text{Stab}(P[\mathcal{F}])$ and the kernel of the homomorphism $\mathcal{MCG}(S) \rightarrow \text{GL}(H_1(S; \mathbb{Z}/3))$, and so by Theorem 2.1 it follows that Γ has finite index in $\text{Stab}(P[\mathcal{F}])$. It suffices to prove that Γ is infinite cyclic, for in that case ϕ, ϕ' have positive powers in Γ , implying that $\phi^m = \phi'^n$ for some $m, n \geq 0$, implying that $[\mathcal{F}^s(\phi)] = [\mathcal{F}^s(\phi^m)] = [\mathcal{F}^s(\phi'^n)] = [\mathcal{F}^s(\phi')]$.

Consider the homomorphism $\ell: \Gamma \rightarrow \mathbb{R}_+$ which is characterized by the equation $\psi[\mathcal{F}] = \ell(\psi)[\mathcal{F}]$. Suppose that $\psi \in \text{kernel}(\ell)$. As seen above, ψ has finite order. But $\psi \in \text{kernel}(\mathcal{MCG}(S) \rightarrow \text{GL}(H_1(S; \mathbb{Z}/3)))$ and so by Theorem 2.1, ψ is the identity. The homomorphism ℓ is thus injective. But the image of ℓ is discrete – indeed, the set of expansion factors of pseudo-Anosov elements of $\mathcal{MCG}(S)$ is bounded away from 1, because each is an eigenvalue of a square integer matrix of bounded size [10]. It follows that Γ is infinite cyclic. \square

The ping pong argument and free subgroups. Here is the central construction of the Tits alternative:

Lemma 3.3. *For any two independent pseudo-Anosov mapping classes ϕ, ψ , for sufficiently large integers $m, n > 0$ the elements ϕ^m, ψ^n freely generate a free subgroup of $\mathcal{MCG}(S)$.*

The proof of this lemma is a standard ping pong argument, using that ϕ, ψ each act with source-sink dynamics on \mathcal{T} , and that their fixed points sets are disjoint. A general ping pong lemma stated in a topological context is as follows:

Lemma 3.4 (Topological ping pong). *Let G act on a topological space X and let $\phi_1, \dots, \phi_k \in G$. Suppose that there exists a pairwise disjoint collection of open sets $U_i^-, U_i^+, i = 1, \dots, k$ such that for each $i \in \{1, \dots, k\}$ and each $\varepsilon \in \{-, +\}$ we have*

$$\phi_i^\varepsilon(X - U_i^{-\varepsilon}) \subset U_i^\varepsilon.$$

Then G is freely generated by ϕ_1, \dots, ϕ_k .

The way that Lemma 3.3 is proved is to note that when ϕ, ψ are independent pseudo-Anosov mapping classes, and if $U_\phi^s, U_\phi^u, U_\psi^s, U_\psi^u$ are pairwise disjoint neighborhoods of the fixed points $P[\mathcal{F}_\phi^s], P[\mathcal{F}_\phi^u], P[\mathcal{F}_\psi^s], P[\mathcal{F}_\psi^u]$, then by applying the source–sink dynamics of ϕ and of ψ , the hypotheses of the Ping Pong Lemma are satisfied for ϕ^m, ψ^n as long as m, n are sufficiently large.

One of the main themes of the rest of this survey is that by taking still more care with the proof of Lemma 3.3, one can show that the free subgroup $\langle \phi^m, \psi^n \rangle$ satisfies stronger geometric properties reminiscent of convex cocompact discrete subgroups of hyperbolic isometries.

The virtual centralizer and virtual normalizer of a pseudo-Anosov cyclic subgroup. Consider a group G and a subgroup $H < G$. The *virtual centralizer* $VC(H)$ of H in G is the subgroup of all $g \in G$ which commute with a finite index subgroup of H . The *virtual normalizer* $VN(H)$ is the subgroup of all $g \in G$ such that $gHg^{-1} \cap H$ has finite index in gHg^{-1} and in H .

The virtual centralizer and virtual normalizer of an infinite cyclic pseudo-Anosov subgroup each have a particularly nice geometric/dynamic description, and the proof of this description uses the ping pong argument in a nice way. Given an action of a group G on a space X , for a subspace $S \in X$ denote $\text{Stab}(S) = \{g \in G \mid g(S) = S\}$. If S is homeomorphic to a line denote $\text{Stab}_+(S)$ to be the subgroup of $\text{Stab}(S)$ that preserves the orientation on S .

Theorem 3.5. *Given a pseudo-Anosov $\phi \in \mathcal{MCG}(S)$ with stable and unstable foliations $[\mathcal{F}^s], [\mathcal{F}^u]$ and Teichmüller axis A , we have*

$$\begin{aligned} VC(\phi) &= \text{Stab}(P[\mathcal{F}^s]) = \text{Stab}(P[\mathcal{F}^u]) = \text{Stab}_+(A), \\ VN(\phi) &= \text{Stab}\{P[\mathcal{F}^s], P[\mathcal{F}^u]\} = \text{Stab}(A). \end{aligned}$$

The group $\text{Stab}_+(A)$ has index at most 2 in $\text{Stab}(A)$, it has an infinite cyclic subgroup of finite index, and the index in $\text{Stab}_+(A)$ of the maximal such subgroup is bounded by the maximum order of a finite subgroup of $\mathcal{MCG}(S)$.

Proof. The group $\text{Stab}(A)$ acts properly discontinuously on the Teichmüller geodesic A and so is virtually cyclic. The inclusions $\text{Stab}_+(A) \subset VC(\phi)$ and $\text{Stab}(A) \subset VN(\phi)$ follow immediately. The inclusions $\text{Stab}(P[\mathcal{F}^s]), \text{Stab}(P[\mathcal{F}^u]) \subset \text{Stab}_+(A)$ and the equation $\text{Stab}\{P[\mathcal{F}^s], P[\mathcal{F}^u]\} = \text{Stab}(A)$ are obvious.

Suppose that $\psi \in \mathcal{MCG}(S) - \text{Stab}\{P[\mathcal{F}^s], P[\mathcal{F}^u]\}$. Then $\phi' = \psi\phi\psi^{-1}$ is pseudo-Anosov, and it is independent of ϕ by Lemma 3.2. Some powers of ϕ and ϕ' therefore generate a free group of rank 2, and hence $\psi \notin VN(\phi)$.

Suppose next that $\psi \in \mathcal{MCG}(S) - \text{Stab}(P[\mathcal{F}^u])$, and again consider the pseudo-Anosov mapping class $\phi' = \psi\phi\psi^{-1}$. If $\psi(P[\mathcal{F}^u]) \neq P[\mathcal{F}^s]$ then ϕ' is independent of ϕ and $\psi \notin VN(\phi)$, so $\psi \notin VC(\phi)$. If $\psi(P[\mathcal{F}^u]) = P[\mathcal{F}^s]$ but $\psi(P[\mathcal{F}^s]) \neq P[\mathcal{F}^u]$ then $\text{Fix}(\phi), \text{Fix}(\phi')$ are not disjoint and not equal, violating Lemma 3.2.

And if ψ interchanges $P[\mathcal{F}^u]$ and $P[\mathcal{F}^s]$ then ϕ' translates along A in the direction opposite to ϕ and so $\psi \notin \text{VC}(\phi)$. \square

4 The Tits alternative and abelian subgroups

The following theorem was proved independently by Ivanov [25] and McCarthy [33]:

Theorem 4.1 (The Tits alternative). *For each finite type surface S there exist constants I, R such that for any subgroup G of $\mathcal{MCG}(S)$, either G contains a free subgroup of rank 2, or G contains an abelian subgroup of index $\leq I$ and rank $\leq R$.*

Ivanov gave an underlying geometric description of each subgroup of $\mathcal{MCG}(S)$, from which the above Tits alternative follows quickly. A subgroup $G < \mathcal{MCG}(S)$ is *reducible* if there exists an essential curve family which is invariant under the action of each element of G . If no such curve family exists then G is *irreducible*.

Theorem 4.2. *If G is an infinite, irreducible subgroup of $\mathcal{MCG}(S)$ then G contains a pseudo-Anosov element.*

For the proof see Ivanov's book [25].

Proof of Tits alternative, from Theorem 4.2. Let G be an arbitrary subgroup of \mathcal{MCG} .

Suppose first that G is irreducible. Applying Theorem 4.2, G contains a pseudo-Anosov mapping class ϕ . If $G \subset \text{Stab}(\text{Fix}(\phi)) = \text{VN}(\phi)$ then by Theorem 3.5 the group G has an infinite cyclic subgroup of index at most twice the maximum size of a finite subgroup of \mathcal{MCG} . If $G \not\subset \text{Stab}(\text{Fix}(\phi))$, say $\psi \in G - \text{Stab}(\text{Fix}(\phi))$, then by Lemma 3.2 $\psi\phi\psi^{-1}$ is a pseudo-Anosov element independent of ϕ and so by the ping pong argument G contains a free group of rank 2.

Suppose now that G is reducible. Let \mathcal{C} be a maximal essential curve family invariant under G . The group G acts on the set of oriented elements of \mathcal{C} ; let $G_{\mathcal{C}}$ denote the kernel of this action, a finite index subgroup of G whose index is bounded above by $2^n |\mathcal{C}|!$. Letting $\{S_i\}$ be the components of $S - \mathcal{C}$, it follows that $G_{\mathcal{C}}$ preserves each S_i up to isotopy, and so there is a well-defined restriction homomorphism $r_i: G_{\mathcal{C}} \rightarrow \mathcal{MCG}(S_i)$ whose image we denote $G | S_i$. By maximality of \mathcal{C} it follows that $G | S_i$ is irreducible. If $G | S_i$ has two independent pseudo-Anosov elements then $G | S_i$ has a free subgroup of rank ≥ 2 , and so therefore does G .

We have reduced to the case that $G | S_i$ is either finite or is contained in $\text{VN}(\phi_i)$ for some pseudo-Anosov element $\phi_i \in \mathcal{MCG}(S_i)$. In this case we find a free abelian subgroup of finite index in G . Let $A_i \subset G | S_i$ denote either the trivial subgroup or an infinite cyclic group, with index bounded as in Theorem 3.1 or 3.5. Consider the subgroup $H = \bigcap_i r_i^{-1}(A_i) < G$. The number of subsurfaces S_i is bounded in terms of g and p , and so we obtain a finite bound on the index of H in G , in terms

of g and p . Restricted to each S_i , the subgroup H is either trivial or infinite cyclic, in the latter case generated by a pseudo-Anosov element of S_i . H also contains Dehn twists about the curves in \mathcal{C} . These pseudo-Anosov subsurface mapping classes and the Dehn twists all commute with each other. Thus H is a free abelian group, whose rank is bounded by $|\mathcal{C}| + |\{S_i\}|$, which is bounded in terms of g and p . \square

Note that this proof gives a complete description of all free abelian subgroups of $\mathcal{MCG}(S)$: each irreducible one is infinite cyclic generated by a pseudo-Anosov mapping class; and each reducible one is freely generated by Dehn twists along curves \mathcal{C} and pseudo-Anosov mapping classes on components $\{S_i\}$ of $S - \mathcal{C}$. To compute the maximal rank exactly, note that on any subsurface S_i that supports a pseudo-Anosov homeomorphism, one can replace that homeomorphism by a Dehn twist, enlarging \mathcal{C} without decreasing the rank. The maximum rank is thus achieved when each S_i is a pair of pants, in which case the rank equals the maximum value of $|\mathcal{C}|$, which equals $3g - 3 + p$.

5 Limit sets

Given a Gromov hyperbolic, geodesic metric space X and a finitely generated, discrete subgroup $G < \text{Isom}(X)$, its *limit set* $\Lambda_G \subset \partial X$ is, by definition, the accumulation set in ∂X of any orbit of G in X . As long as G is nonelementary (that is, not virtually abelian), the limit set may be characterized as the closure of the set of fixed points of loxodromic elements of G , or as the unique minimal closed nonempty G -invariant subset of ∂X . Assuming in addition that X is proper, it follows that Λ_G is compact, and the *domain of discontinuity* $\Delta_G = \partial X - \Lambda_G$ is characterized as the unique maximal open G -invariant set on which the action of G is properly discontinuous.

In [31], Masur investigated notions of the limit set and domain of discontinuity for the *handlebody subgroup* of $\mathcal{MCG}(S)$ when S is the boundary of a compact genus g handlebody H ; this subgroup consists of the mapping classes of homeomorphisms of S that extend to H .

In [34], McCarthy and Papadopoulos generalized Masur's results to arbitrary subgroups of $\mathcal{MCG}(S)$. The results are not quite as pretty as they are for discrete groups of hyperbolic isometries. But as we shall see, the results of Kent and Leininger [27] show that limit sets of convex cocompact subgroups of \mathcal{MCG} behavior very much like their counterparts in hyperbolic geometry.

We start with some elementary cases. When G is a finite subgroup of \mathcal{MCG} , the natural choice of the limit set is $\Lambda_G = \emptyset$, and the action of G on $\Delta_G = \mathcal{PMF}$ is properly discontinuous. When G is virtually cyclic, containing a pseudo-Anosov cyclic subgroup $\langle \phi \rangle$ with finite index, let $\Lambda_G = \{P[\mathcal{F}_\phi^s], P[\mathcal{F}_\phi^u]\}$. In the case where the action of G does not fix each element of Λ_G , that is when $\text{VC}\langle \phi \rangle$ is an index 2 subgroup of $\text{VN}\langle \phi \rangle$, then Λ_G is the unique nonempty minimal closed G -invariant subset of \mathcal{PMF} . In the other case, when $\text{VC}\langle \phi \rangle = \text{VN}\langle \phi \rangle$, then the two points of Λ_G

are the only nonempty minimal closed G -invariant subsets of \mathcal{PMF} . In either case, the action of G on $\Delta_G = \mathcal{PMF} - \Lambda_G$ is properly discontinuous.

Consider now the case where G is an infinite, irreducible subgroup of \mathcal{MCG} and G has no finite index cyclic subgroup. Define $\Lambda_0 \subset \mathcal{PMF}$ to be the set of fixed points of pseudo-Anosov elements of G . Define $\Lambda_G \subset \mathcal{PMF}$ to be the closure of Λ_0 . Define $Z\Lambda_G$ to be the “zero set” of Λ_G , the set of all $P[\mathcal{F}] \in \mathcal{PMF}$ such that for some $P[\mathcal{F}'] \in \Lambda_G$ we have $\langle [\mathcal{F}], [\mathcal{F}'] \rangle = 0$; this is well-defined independent of the choice of $[\mathcal{F}], [\mathcal{F}']$ in their projective Whitehead classes. Define $\Delta_G = \mathcal{PMF} - Z\Lambda_G$.

Theorem 5.1 ([34]). *For G an infinite, irreducible subgroup of \mathcal{MCG} , the set Λ_G is the unique minimal nonempty closed G -invariant subset of \mathcal{PMF} . If $\Lambda_G \neq \mathcal{PMF}$ then $\Delta_G \neq \emptyset$, and the action of G on Δ_G is properly discontinuous.*

We turn now to the case that G is an infinite, reducible subgroup of $\mathcal{MCG}(S)$. In general there will be no unique closed minimal G -invariant subset of \mathcal{PMF} . For example, if G is a free abelian group generated by Dehn twists about the curves in an essential curve family \mathcal{C} , then G has many, many fixed points, namely any point in \mathcal{PMF} having zero intersection number with each curve in \mathcal{C} . Nevertheless, as McCarthy and Papadopoulos show, there is still a reasonable candidate for a limit set and a domain of discontinuity.

Let \mathcal{C} be an essential curve family invariant under G such that if $\{S_i\}$ is the set of components of $S - \mathcal{C}$, then the restriction $G|_{S_i} < \mathcal{MCG}(S_i)$ is an irreducible subgroup of $\mathcal{MCG}(S_i)$. As is shown in, say, [34], there exists a unique such family \mathcal{C} which is minimal with respect to inclusion. Reindex the S_i so that S_1, \dots, S_n are the surfaces on which the restriction groups $G|_{S_i}$ are infinite. Each of the groups $G|_{S_i}$ for $i = 1, \dots, n$ therefore contains a pseudo-Anosov element of $\mathcal{MCG}(S_i)$. The stable and unstable foliations of pseudo-Anosov elements of $G|_{S_i}$, being measured foliations on S_i , may by enlargement be regarded as measured foliations on S ; let $\Lambda_0^i \subset \mathcal{PMF}$ be the set of all points in \mathcal{PMF} obtained in this manner. Let $\Lambda^i \subset \mathcal{PMF}$ be the closure of Λ_0^i . Define

$$\Lambda_G = \left(\bigcup_{i=1}^n \Lambda^i \right) \cup \left(\bigcup_{c \in \mathcal{C}} P[c] \right)$$

Define the zero set $Z\Lambda_G$ and the domain $\Delta_G = \mathcal{PMF} - Z\Lambda_G$ as above.

Theorem 5.2 ([34]). *For each infinite, reducible subgroup of \mathcal{MCG} , the set Λ_G is a nonempty closed G -invariant subset of \mathcal{PMF} . If $\Lambda_G \neq \mathcal{PMF}$ then $\Delta_G \neq \emptyset$, and the action of G on Δ_G is properly discontinuous.*

The set Λ_G shall be called the *limit set* and Δ_G the *domain of discontinuity* of G .

6 Convex cocompact subgroups

Given a finitely generated, discrete subgroup $G < \text{Isom}(\mathbb{H}^n)$, the *convex hull* of its limit set Λ_G is the smallest closed, convex subset $\mathcal{H}_G \subset \mathbb{H}^n$ whose set of accumulation points in S_∞^{n-1} is Λ_G . The action of G on \mathbb{H}^n preserves \mathcal{H}_G , and G is said to be *convex cocompact* if G acts cocompactly on \mathcal{H}_G . Several conditions are equivalent to convex cocompactness: some (every) orbit of G in \mathcal{H}_G is quasiconvex; the action of G on Δ_G is cocompact; the action of G on $\mathbb{H}^n \cup \Delta_G$ is cocompact. Moreover, if G is convex cocompact then G is word hyperbolic, there is a G -equivariant homeomorphism from the Gromov boundary ∂G to Λ_G , and this homeomorphism extends to a G -equivariant continuous map from the Gromov compactification $G \cup \partial G$ to $\mathcal{H}_G \cup \Lambda_G$.

These notions can be generalized with care to a finitely generated, discrete group G of isometries of any Gromov hyperbolic, geodesic metric space X .

The theory of convex cocompact subgroups of \mathcal{MCG} , originated in [9] and further developed in [27], and [17], gives analogues to these results about convex cocompact subgroups of \mathcal{MCG} . In addition, the theory develops further equivalent characterizations of convex compactness that are special to the setting of \mathcal{MCG} and are important for applications. Thus far, the only examples of convex cocompact subgroups of \mathcal{MCG} are free [9] or virtually free examples as constructed by Honglin Min in her dissertation [36].

We define a subset $X \subset \mathcal{T}$ to be *quasiconvex* if there exist a constant C such that for every $x, y \in X$, each point of the Teichmüller geodesic $[x, y]$ is within distance C of some point of X .

Given a closed subset $\Lambda \subset \mathcal{PMF}$, if every pair $\xi \neq \eta \in \Lambda$ jointly fills S then we define the *weak hull* of Λ to be the subset of \mathcal{T} defined by

$$\mathcal{H}(\Lambda) = \bigcup \{ \overleftrightarrow{(\xi, \eta)} \mid \xi \neq \eta \in \Lambda \}$$

and we say that the *weak hull of Λ is defined*. Otherwise, if there exist $\xi \neq \eta \in \Lambda$ which do not jointly fill S then we say that the weak hull of Λ is undefined.

Theorem 6.1 ([9]). *For any finitely generated subgroup $G < \mathcal{MCG}$, the following are equivalent.*

- *Every (some) orbit of G on \mathcal{MCG} is quasiconvex in \mathcal{T} .*
- *G is word hyperbolic, and there exists a G -equivariant embedding $\partial G \hookrightarrow \mathcal{PMF}$ with image denoted Λ , such that the weak hull of Λ is defined, the action of G on $\text{Hull}(\Lambda)$ is cocompact, and the extension of the map $\partial G \hookrightarrow \mathcal{PMF}$ by any G -equivariant map $G \rightarrow \text{Hull}(\Lambda)$ is a continuous map $G \cup \partial G \rightarrow \text{Hull}(\Lambda) \cup \Lambda$.*

A group satisfying these conditions is called *convex cocompact*. When G is convex cocompact then every infinite order element of G is pseudo-Anosov.

The connection of convex cocompactness with the limit set Λ_G and domain of discontinuity Δ_G as defined by McCarthy–Papadopoulos is given in the following results of Kent and Leininger:

Theorem 6.2 ([27]). *For any finitely generated subgroup $G < \mathcal{MCG}$, the following are equivalent:*

- G is convex cocompact.
- The weak hull Hull_G of the McCarthy–Papadopoulos limit set Λ_G is defined, and the action of G on Hull_G is cocompact.
- The action of G on $\mathcal{T} \cup \Delta_G$ is cocompact.

Furthermore, if G is convex cocompact, then the image of the embedding $\partial G \rightarrow \mathcal{PMF}$ is the McCarthy–Papadopoulos limit set Λ_G , and $Z\Lambda_G = \Lambda_G$, and so $\Delta_G = \mathcal{PMF} - \Lambda_G$.

Convex cocompactness on the curve complex. The curve complex $\mathcal{CC} = \mathcal{CC}(S)$ is the simplicial complex whose vertex set is \mathcal{C} , and whose n -simplices are the isotopy classes of essential curve families of cardinality $n + 1$. To be precise, a set of distinct vertices $[c_0], \dots, [c_n] \in \mathcal{C}$ spans an n -simplex of \mathcal{CC} if and only if there are representatives c_0, \dots, c_n which are pairwise disjoint. The complex \mathcal{CC} is locally infinite, in fact the link of every simplex of positive codimension is infinite.

The natural action of \mathcal{MCG} on \mathcal{CC} has finitely many orbits of simplices. The number of orbits of vertices is

$$m = \begin{cases} 1 + \lceil \frac{g-1}{2} \rceil & \text{if } g \geq 1, p = 0, \\ \lceil \frac{p-3}{2} \rceil & \text{if } g = 0, p \geq 4, \\ 1 + \lceil \frac{(g+1)(p+1)-4}{2} \rceil & \text{if } g, p \geq 1. \end{cases}$$

When $g \geq 1$ there is one orbit of nonseparating curves, but in genus zero every curve is separating. The orbit of a separating curve is determined by how it partitions the genus and the number of punctures. When $p = 0$, the relevant partition is $g = g_1 + g_2$ with $g_1, g_2 \in \{1, \dots, g\}$, and when $g = 0$ the relevant partition is $p = p_1 + p_2$ with $p_1, p_2 \in \{2, \dots, p\}$. When $g \geq 1, p \geq 1$ then the relevant partition is an unordered partition of the ordered pair $(g, p) = (g_1, p_1) + (g_2, p_2)$, where the four values $(g_i, p_i) = (0, 0), (g, p), (0, 1), (g, p - 1)$ are forbidden but all other values with $g_i \in \{0, \dots, g\}$ and $p_i \in \{0, \dots, p\}$ are allowed.

The large scale geometric properties of \mathcal{CC} and of \mathcal{MCG} are related in the following manner.

Given a finitely generated group G and a finite collection of finitely generated subgroups H_1, \dots, H_m , assume that the generators of G include generators of each of the subgroups H_1, \dots, H_m , so that the Cayley graph of H_i embeds, H_i -equivariantly, in the Cayley graph Γ of G . Define the *coned off Cayley graph* of G relative to H_1, \dots, H_m by adding a new vertex to Γ for each left coset gH_i of each H_i , and adding new edges connecting the new vertex to all the vertices of gH_i . We say that Γ is *weakly hyperbolic* relative to H_1, \dots, H_m if the coned off Cayley graph is a Gromov hyperbolic metric space.

Choose unique representatives $[c_1], \dots, [c_m]$ for the orbits of the action of \mathcal{MCG} on \mathcal{C} , and let $H_i = \text{Stab}[c_i] < \mathcal{MCG}$. It is easy to see [32] that \mathcal{CC} is equivariantly quasi-isometric to the coned off Cayley graph of \mathcal{MCG} relative to H_1, \dots, H_m . Much deeper is the question of what, exactly, is the large scale geometric behavior of these two spaces:

Theorem 6.3 ([32]). *The curve complex \mathcal{CC} is a Gromov hyperbolic metric space, and \mathcal{MCG} is weakly hyperbolic relative to stabilizer subgroups of representatives of orbits in \mathcal{C} .*

The boundary of the curve complex was identified in the following theorem of Klarreich [28]; see also [16]. Two points in \mathcal{PMF} are *topologically equivalent* if they are represented by singular foliations which, after forgetting measure, are identical.

Theorem 6.4. *There is an \mathcal{MCG} equivariant quotient map from the subspace of \mathcal{PMF} consisting of classes of arational measured foliations to the Gromov boundary of \mathcal{CC} , such that the decomposition elements of this quotient map are the topological equivalence classes of arational measured foliations.*

Using the fact that \mathcal{CC} is convex cocompact, it makes sense to compare convex cocompactness of a subgroup of \mathcal{MCG} with respect to its actions on \mathcal{T} and on \mathcal{CC} . This is accomplished in the following theorem proved independently, and with different techniques, by Kent–Leininger and Hamenstädt:

Theorem 6.5 ([27], [17]). *A finitely generated subgroup $G < \mathcal{MCG}$ is convex cocompact if and only if its action on \mathcal{CC} is convex cocompact.*

Schottky and virtual Schottky subgroups of mapping class groups. A convex cocompact, free subgroup of \mathcal{MCG} is called a *Schottky subgroup*. The ubiquity of Schottky subgroups was first demonstrated by Farb and myself in [9], using a result of [37] which is to be explained below.

Theorem 6.6 ([9]). *For any independent set of pseudo-Anosov mapping classes $\phi_1, \dots, \phi_K \in \mathcal{MCG}(S)$, there exists an integer $B \geq 1$ such that for all integers $\beta_1, \dots, \beta_K \geq B$, the mapping classes $\phi_1^{\beta_1}, \dots, \phi_K^{\beta_K}$ freely generates a Schottky subgroup of $\mathcal{MCG}(S)$.*

Kent and Leininger [27] and independently Hamenstädt [17] discovered a more direct and natural setting for this theorem, using a ping pong argument in the curve complex. Here is their proof.

The action of ϕ_k on \mathcal{PMF} has source sink dynamics, with source and sink each being arational, and for $k = 1, \dots, K$ the $2K$ fixed points are pairwise topologically inequivalent, because they are pairwise distinct and each is uniquely ergodic. By Theorem 6.4, it follows that the action of each ϕ_k on $\partial\mathcal{CC}$ has source sink dynamics.

and so ϕ_k is a hyperbolic isometry on the Gromov hyperbolic metric space $\mathcal{C}\mathcal{C}$, and that their fixed point sets in $\partial\mathcal{C}\mathcal{C}$ are pairwise disjoint. To conclude that $\phi_1^{\beta_1}, \dots, \phi_K^{\beta_K}$ freely generate a convex cocompact subgroup of \mathcal{MCG} , we apply the following folk theorem:

Theorem 6.7 (Hyperbolic ping pong). *If X is a Gromov hyperbolic geodesic metric space, and if ϕ_1, \dots, ϕ_K are hyperbolic isometries of X with pairwise disjoint fixed points in ∂X , then there exists an integer $B \geq 1$ such that for all integers $\beta_1, \dots, \beta_K \geq B$, the isometries $\phi_1^{\beta_1}, \dots, \phi_K^{\beta_K}$ freely generate a convex cocompact group of isometries.*

It seems hard to track down an early proof of hyperbolic ping pong, although it was certainly known to the very earliest practitioners of Gromov hyperbolicity. Here is a sketch of a proof, pretty much the same as the proof given in [27].

One may choose λ, ε quasigeodesic lines $\gamma_1, \dots, \gamma_K$ in X that are axes for ϕ_1, \dots, ϕ_K , respectively (if X were proper then one could choose $\lambda = 1, \varepsilon = 0$, in other words, geodesics). Fix a base point $p \in X$, and let D be an upper bound for the distance from p to $\gamma_1, \dots, \gamma_K$. Also, let L be a lower bound for the length of a fundamental domain of each ϕ_1, \dots, ϕ_K along its axis $\gamma_1, \dots, \gamma_K$. Fix B and $\beta_1, \dots, \beta_K \geq B$, let G be the group generated by $\phi_1^{\beta_1}, \dots, \phi_K^{\beta_K}$.

We know by topological ping pong that if B is sufficiently large then G is freely generated by $\phi_1^{\beta_1}, \dots, \phi_K^{\beta_K}$. It suffices to show that for each word $w = w_1 \dots w_m$ in the generators, a geodesic from p to $w(p)$ stays uniformly close to the orbit $G \cdot p$. We describe a path γ_w from p to $w(p)$ as follows. Denote $w_i = \phi_{k_i}^{\beta_{k_i}}$. Starting from p , jump a distance at most D onto the axis γ_{k_1} , travel along γ_1 for β_1 fundamental domains, thereby travelling a distance of at least BL , then jump a distance at most D to the point $w_1(p)$. Next, jump a distance at most D onto the axis $w_1(\gamma_{k_2})$, travel along this axis for β_2 fundamental domains, thereby travelling a distance of at least BL , then jump a distance at most D to the point $w_1 w_2(p)$. Continuing in this manner we get a path γ_w from p to $w(p)$, which stays in a $D + BL$ neighborhood of $G \cdot p$. The path γ_w is a $(BL + 2D)$ local $\lambda, (\varepsilon + 2D)$ quasigeodesic. If B is sufficiently large, it follows that γ_w is a λ, ε' quasigeodesic for ε' depending only on the previous constants and the hyperbolicity constant of X [5]. Now we use the fact that γ_w has Hausdorff distance at most E from any geodesic ℓ with the same endpoints $p, w(p)$, where E depends only on λ, ε' , and the hyperbolicity constant for X . It follows that ℓ stays within distance $E + D + BL$ of $G \cdot p$.

Virtual Schottky subgroups. A *virtual Schottky subgroup* of \mathcal{MCG} is a subgroup that contains a Schottky subgroup with finite index. Clearly convex cocompactness is preserved under passage to finite index supergroups, so every virtual Schottky subgroup is convex cocompact. The following recent result, to appear in the dissertation of Honglin Min, is another example of a ping pong argument in the mapping class group:

Theorem 6.8 ([36]). *If A, B are finite subgroups of \mathcal{MCG} , if $\phi \in \mathcal{MCG}(S)$ is pseudo-Anosov, and if the virtual normalizer of ϕ has trivial intersection with A and B , then for all sufficiently large n the subgroups A and $\phi^n B \phi^{-n}$ freely generate their free product in \mathcal{MCG} , and this subgroup is virtually Schottky.*

Hyperbolic extensions of surface groups When the surface S is closed and oriented, there is a canonical short exact sequence

$$1 \rightarrow \pi_1(S, p) \rightarrow \mathcal{MCG}(S - p) \rightarrow \mathcal{MCG}(S) \rightarrow 1.$$

See [3].

Given a subgroup $H \subset \mathcal{MCG}(S)$, let Γ_H be its preimage in $\mathcal{MCG}(S - p)$, and so we have an extension

$$1 \rightarrow \pi_1(S, p) \rightarrow \Gamma_H \rightarrow H \rightarrow 1.$$

Thurston's geometrization theorem (see [39]) shows that if H is an infinite cyclic, pseudo-Anosov subgroup of $\mathcal{MCG}(S)$ then Γ_H is a word hyperbolic group.

The following theorem was the sequel to the theory of convex cocompact groups. It is proved by an application of the Bestvina–Feighn combination theorem.

Theorem 6.9 ([37]). *If ϕ_1, \dots, ϕ_K are independent pseudo-Anosov elements of \mathcal{MCG} then sufficiently high powers of ϕ_1, \dots, ϕ_K freely generate a subgroup H such that Γ_H is word hyperbolic.*

With this result as motivation, in [9] Farb and I set out to develop a geometric understanding of when Γ_H is word hyperbolic, attempting to prove that for this to be true it was necessary and sufficient that H be convex cocompact. We proved necessity, and when H is free we proved sufficiency by using the Bestvina–Feighn combination theorem [2]. A completely general proof of sufficiency was recently given by Hamenstädt [17], so that now one can state:

Theorem 6.10 ([9], [17]). *For any subgroup $H \subset \mathcal{MCG}$, the group Γ_H is word hyperbolic if and only if H is convex cocompact.*

7 Teichmüller discs and their stabilizers

One very well studied class of subgroups of \mathcal{MCG} is stabilizers of Teichmüller discs. A *Teichmüller disc* in \mathcal{T} can be defined in one of the following equivalent ways.

First, consider $\sigma \in \mathcal{T}$ and $\theta \in \text{QD}_\sigma$. As ω varies over the unit circle in the complex plane, consider the family of quadratic differentials $\omega\theta$. This defines a family of rays based at σ parameterized by the circle, and the union of these rays is defined to be the *Teichmüller disc with basepoint σ and direction θ* , denoted $D(\sigma, \theta)$. A somewhat more invariant way to say almost the same thing is that $D(\sigma, \theta)$ is the set of points

in \mathcal{T} obtained by choosing an element of $\mathrm{PSL}(2, \mathbb{R})$ acting affinely on the Euclidean plane, and applying that element to the regular canonical coordinates for θ to get a new conformal structure.

Second, a Teichmüller disc D is a maximal subset which is the image of a holomorphic embedding of the Poincaré disc into \mathcal{T} . A theorem of Royden [41] in the compact case and Gardiner [12] in the noncompact case says that any two points $\sigma \neq \sigma' \in \mathcal{T}$ lie on a unique such disc D , and moreover that the Teichmüller distance between σ and σ' equals their distance in the Poincaré metric on D . The identification of D with $D(\sigma, \theta)$ for some θ is obtained by choosing θ to be the initial quadratic differential from σ to σ' .

To each Teichmüller disc D there is associated an embedded circle $S^1(D) \subset \mathcal{PMF}$, defined as follows. Choose $\sigma \in \mathcal{T}$ and $\theta \in \mathrm{QD}_\sigma$ so that $D = D(\sigma, \theta)$. As ω varies over the unique circle in the complex plane, the horizontal measured foliations of $\omega\theta$ trace out the desired circle $S^1(D)$. Alternatively, a point $P[\mathcal{F}] \in \mathcal{PMF}$ is in $S^1(D)$ if and only if it is represented as a linear measured foliation in any regular canonical coordinate z for θ ; in other words, there exists an \mathbb{R} -linear differential form $a dx + b dy$ such that $P[\mathcal{F}]$ is represented in any z by $|a dx + b dy|$. This makes it clear that $S^1(D)$ is well-defined independent of $\sigma \in D$, because any other $\sigma' \in D$ is obtained by a Teichmüller deformation with initial quadratic differential $\omega\theta$ for some ω in the unit circle, resulting in a terminal quadratic differential θ' on σ' , and linearity with respect to θ and to θ' are clearly equivalent.

Let $D \subset \mathcal{T}$ be a Teichmüller disc, and let $\mathrm{Stab}(D) = \{\phi \in \mathcal{MCG} \mid \phi(D) = D\}$ be its stabilizer. The group $\mathrm{Stab}(D)$, or its image in $\mathrm{PSL}(2, \mathbb{R})$ when regarding D as the upper half plane, is sometimes called a *Veech group*, and the quotient orbifold $D/\mathrm{Stab}(D)$, a proper holomorphic curve in the moduli space of S , is sometimes called a *Veech surface*. These terms are sometimes restricted to the case where $\mathrm{Stab}(D)$ is a lattice, meaning that $D/\mathrm{Stab}(D)$ has finite hyperbolic area.

Regarding D as the hyperbolic plane, the elliptic–parabolic–loxodromic trichotomy for $\mathrm{Isom}(\mathbb{H}^2)$ becomes not just an analogy but a strict correspondence with the finite order–reducible–pseudo-Anosov dichotomy for elements of $\mathrm{Stab}(D)$; for details see for example [44]. To be precise, fix $\phi \in \mathrm{Stab}(D)$. Then ϕ is loxodromic on D if and only if it is pseudo-Anosov in \mathcal{MCG} , ϕ is parabolic on D if and only if it is reducible in \mathcal{MCG} , and ϕ is elliptic on D if and only if it is finite order in \mathcal{MCG} . The case where ϕ is parabolic can be analyzed more closely. In this case, letting $D = D(\sigma, \theta)$, the action of ϕ is a shear transformation on any regular canonical coordinate z , parallel to the leaves of some constant slope measured foliation \mathcal{F} called the *shear foliation* for ϕ ; this foliation corresponds to the point of $S^1(D)$ fixed by ϕ . Each leaf of \mathcal{F} is compact, and so \mathcal{F} is obtained by enlarging some weighted family of simple closed curves c_1, \dots, c_n . In particular, the essential curve family $\{c_1, \dots, c_n\}$ is a canonical reducing system for ϕ , and the first return of ϕ to each complementary component of $S - (c_1 \cup \dots \cup c_n)$ is of finite order, so ϕ is the composition of a finite order mapping class that is reduced by $\{c_1, \dots, c_n\}$ and a product of Dehn twists along the components of c_1, \dots, c_n .

Using this description, we can get a fairly precise understanding of the case when $\text{Stab}(D)$ is a reducible subgroup of \mathcal{MCG} . Assuming this is so, let $\phi \in \text{Stab}(D)$ be of infinite order, with shear foliation \mathcal{F} . For any other infinite order $\phi' \in \text{Stab}(D)$, the shear foliation of ϕ' must have the same slope as \mathcal{F} and so represent the same point in $S^1(D)$. Otherwise, if the shear foliation \mathcal{F}' of ϕ' is distinct from \mathcal{F} in $S^1(D)$, then high powers of ϕ and ϕ' would generate a loxodromic element of $\text{Stab}(D)$, by the proof of the Tits alternative for \mathbb{H}^2 (this is not hyperbolic ping pong, because ϕ and ϕ' do not have source–sink dynamics, but a variation of hyperbolic pong-pong works). Thus, every infinite order element of $\text{Stab}(D)$ stabilizes the same point in $S^1(D)$, in fact every such element acts as a shear transformation on θ with respect to the same linear measured foliation \mathcal{F}_ω . But this implies that $\text{Stab}(D)$ is virtually cyclic, generated by a multiple Dehn twist about components of the curves c_1, \dots, c_n whose enlargement is \mathcal{F}_ω . In this case we can identify the McCarthy–Papadopoulos limit set of $\text{Stab}(D)$: it is the simplex in \mathcal{PMF} spanned by the projective classes of c_1, \dots, c_n .

Henceforth we shall only be interested in the case when $\text{Stab}(D)$ is infinite and irreducible. In this case, since $S^1(D)$ is invariant under $\text{Stab}(D)$ by construction, and since $S^1(D)$ is obviously closed in \mathcal{PMF} , by applying Theorem 5.1 we obtain:

Theorem 7.1. *If $\text{Stab}(D)$ is infinite and irreducible then the McCarthy–Papadopoulos limit set of $\text{Stab}(D)$ is contained in $S^1(D)$.*

Thurston [43] gave the first examples of Teichmüller discs D such that $\text{Stab}(D)$ is a lattice, meaning that it acts on D with cofinite area. Start with the linear action of $\text{SL}(2, \mathbb{Z})$ on $T^2 = \mathbb{R}^2/\mathbb{R}^2$, then choose a finite subset $A \subset T^2$ invariant under this action, and let $S \rightarrow T^2$ be a branched cover of T^2 branched over A . Lifting the conformal structure from T^2 to S defines $\sigma \in \mathcal{T}(S)$, and lifting the quadratic differential $(dx + i dy)^2$ from T^2 to S defines $\theta \in \text{QD}_\sigma$. There exists a finite index subgroup $G < \text{SL}(2, \mathbb{Z})$ that lifts to S , and G stabilizes $D(\sigma, \theta)$ and acts with cofinite area. Combining with Theorem 7.1 it follows that the McCarthy–Papadopoulos limit set of G is the circle $S^1(D)$.

Veech [44] gave the first examples where $\text{Stab}(D)$ is a lattice that does not arise from Thurston’s construction of branching over the torus. Veech also developed some of the general theory of $\text{Stab}(D)$, for instance establishing that $\text{Stab}(D)$ never acts cocompactly on D . When $\text{Stab}(D)$ is a lattice, it follows that there are finitely many maximal parabolic subgroups of $\text{Stab}(D)$ up to conjugacy, and the Veech surface $D/\text{Stab}(D)$ is a finite area holomorphic curve in the moduli space of S .

McMullen [35] proved that if S is the closed surface of genus 2, if $D \subset \mathcal{T}(S)$ is a Teichmüller disc, and if $\text{Stab}(D)$ contains a hyperbolic element, then the limit set of $\text{Stab}(D)$ is all of $S^1(D)$. In the same paper McMullen also gave examples where $\text{Stab}(D)$ is infinitely generated, a phenomenon independently discovered by Hubert and Schmidt [23].

For a further survey of stabilizers of Teichmüller discs, and references to the literature, see the section “Veech surfaces” in the problem list [22], and the chapters by Harvey [19] and by Herrlich and Schmithüsen [20] in this Handbook.

8 Surface subgroups and the Leininger–Reid combination theorem

The only known examples of convex cocompact subgroups are free groups and virtually free groups. Even in $\text{Isom}(\mathbb{H}^n)$, constructing nonfree discrete subgroups can be subtle. One of the pioneering tools in this regard is the Maskit combination theorem [30].

One of the most interesting questions about subgroups of mapping class groups is whether there exists closed surfaces S, F of genus ≥ 2 and a convex cocompact subgroup of $\mathcal{MCG}(S)$ isomorphic to $\pi_1(F)$. If such a subgroup existed then, by Hamenstädt’s theorem, one would obtain the first example of a compact 4-manifold M with word hyperbolic fundamental group that fibers over a surface. The preprint [26] of M. Kapovich contains a theorem that says if such an M existed then M could not have a $\mathbb{C}\mathbb{H}^2$ structure.

It is even hard to construct any surface subgroup $\pi_1(F)$ of $\mathcal{MCG}(S)$ with the genus of F at least two. The first examples were given by Harvey and González–Díez [14], but they have many non-pseudo-Anosov elements and so cannot be convex cocompact.

The closest approach yet to a convex cocompact surface subgroup of \mathcal{MCG} is the following theorem of Leininger and Reid, whose statement is reminiscent of the Maskit combination theorem.

The starting point of the theorem is the construction of a Teichmüller disc D and finite index subgroup $G < \text{Stab}(D)$ such that $\text{Stab}(D)$ (and hence also G) is a lattice, G is torsion free, and G has exactly one parabolic subgroup, generated by a multitwist τ about a multicurve c such that no component of $S - c$ is a three holed sphere. It follows that there exists a homeomorphism ψ that preserves c and is pseudo-Anosov on each component of $S - c$.

Theorem 8.1 ([29]). *With the objects as denoted above, for each sufficiently large k the group generated by G and $\phi^k G \phi^{-k}$ is a free product with amalgamation along $\langle \tau \rangle$, and is therefore isomorphic to $\pi_1(F)$ for an oriented surface F of genus ≥ 2 . Moreover, every element of $\pi_1(F) < \mathcal{MCG}(S)$ is pseudo-Anosov except for elements conjugate to powers of τ .*

A Leininger–Reid subgroup cannot be convex cocompact, because it contains a multitwist. In [38], a notion of geometric finiteness for subgroups of \mathcal{MCG} is proposed, and the question is posed whether Leininger–Reid subgroups are geometrically finite.

References

- [1] L. Bers, An extremal problem for quasiconformal mappings and a theorem by Thurston. *Acta Math.* 141 (1978), 73–98. [388](#), [392](#), [394](#)
- [2] M. Bestvina and M. Feighn, A combination theorem for negatively curved groups. *J. Differential Geom.* 35 (1) (1992), 85–101. [404](#)
- [3] J. Birman, *Braids, links, and mapping class groups*. Ann. of Math. Stud., 82, Princeton University Press, Princeton, N.J., 1974. [404](#)
- [4] J. Birman, A. Lubotzky, and J. McCarthy, Abelian and solvable subgroups of the mapping class groups. *Duke Math. J.* 50 (4) (1983), 1107–1120. [394](#)
- [5] J. Cannon, The theory of negatively curved spaces and groups. In *Ergodic theory, symbolic dynamics, and hyperbolic spaces* (C. Series, T. Bedford, M. Keane, eds.), Oxford Sci. Publ., Oxford University Press, New York 1991, 315–369. [391](#), [403](#)
- [6] A. Casson and S. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*. London Math. Soc. Student Texts 9, Cambridge University Press, Cambridge 1988. [392](#)
- [7] A. Casson and D. Jungreis, Convergence groups and Seifert fibered 3-manifolds. *Invent. Math.* 118 (1994), 441–456. [393](#)
- [8] M. Coornaert, T. Delzant, and A. Papadopoulos, *Géométrie et théorie des groupes*. Lecture Notes in Math. 1441, Springer-Verlag, Berlin 1990. [391](#)
- [9] B. Farb and L. Mosher, Convex cocompact subgroups of mapping class groups. *Geom. Topol.* 6 (2002), 91–152. [388](#), [400](#), [402](#), [404](#)
- [10] A. Fathi, F. Laudenbach, and V. Poénaru (eds.), *Travaux de Thurston sur les surfaces*. *Astérisque* 66–67 (1979). [389](#), [392](#), [393](#), [394](#), [395](#)
- [11] D. Gabai, Convergence groups are Fuchsian groups. *Ann. of Math.* 136 (1992), 447–510. [393](#)
- [12] F. P. Gardiner, Approximation of infinite-dimensional Teichmüller spaces. *Trans. Amer. Math. Soc.* 282 (1) (1984), 367–383. [405](#)
- [13] E. Ghys and P. de la Harpe (eds.), *Sur les groupes hyperboliques d’après Mikhael Gromov*. Progr. Math. 83, Birkhäuser, Boston, MA, 1990. [391](#)
- [14] G. González-Díez and W. J. Harvey, Surface groups inside mapping class groups. *Topology* 38 (1) (1999), 57–69. [407](#)
- [15] M. Gromov, Hyperbolic groups. In *Essays in group theory* (S. Gersten, ed.), Math. Sci. Res. Inst. Publ. 8, Springer-Verlag, New York 1987, 75–263. [391](#)
- [16] 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, 2005, 187–207. [402](#)
- [17] U. Hamenstädt, Word hyperbolic extensions of surface groups. Preprint, 2005; arXiv:math.GT/0505244. [388](#), [400](#), [402](#), [404](#)
- [18] M. Handel and W. Thurston, New proofs of some results of Nielsen. *Adv. in Math.* 56 (2) (1985), 173–191. [394](#)
- [19] W. Harvey, Teichmüller disks, triangle groups and Grothendieck dessins. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 249–292. [407](#)
- [20] F. Herrlich and G. Schmithüsen, On the boundary of Teichmüller disks in Teichmüller space and in Schottky space. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 293–349. [407](#)

- [21] J. Hubbard and H. Masur, Quadratic differentials and foliations. *Acta Math.* 142 (1979), 221–274. [391](#)
- [22] P. Hubert, H. Masur, T. Schmidt, and A. Zorich, Problems on billiards, flat surfaces, and translation surfaces. Preprint, 2005; <http://www.math.uic.edu/~masur/papers.html>. [407](#)
- [23] P. Hubert and T. Schmidt, Infinitely generated Veech groups. *Duke Math. J.* 123 (1) (2004), 49–69. [406](#)
- [24] Y. Imayoshi and M. Taniguchi, An introduction to Teichmüller spaces. Springer-Verlag, Tokyo 1992. [389](#)
- [25] N. V. Ivanov, Subgroups of Teichmüller modular groups. Transl. Math. Monogr. 115, Amer. Math. Soc., Providence, RI, 1992. [388](#), [391](#), [395](#), [397](#)
- [26] M. Kapovich, On normal subgroups in the fundamental groups of complex surfaces. Preprint, 1998; arXiv:math.GT/9808085. [407](#)
- [27] R. P. Kent, IV and C. J. Leininger, Shadows of mapping class groups: capturing convex co-cocompactness. *Geom. Funct. Anal.* (2007), to appear; arXiv:math.GT/0505114. [388](#), [398](#), [400](#), [401](#), [402](#), [403](#)
- [28] E. Klarreich, The boundary at infinity of the curve complex and the relative Teichmüller space. Preprint, 1999. [402](#)
- [29] C. Leininger and A. Reid, A combination theorem for Veech subgroups of the mapping class group. *Geom. Funct. Anal.* 16 (2) (2006), 403–436. [388](#), [407](#)
- [30] B. Maskit, *Kleinian groups*. Grundlehren Math. Wiss. 287, Springer-Verlag, Berlin 1988. [407](#)
- [31] H. Masur, Measured foliations and handlebodies. *Ergodic Theory Dynam. Systems* 6 (1) (1986), 99–116. [398](#)
- [32] H. Masur and Y. Minsky, Geometry of the complex of curves, I. Hyperbolicity. *Invent. Math.* 138 (1) (1999), 103–149. [402](#)
- [33] J. McCarthy, A “Tits-alternative” for subgroups of surface mapping class groups. *Trans. Amer. Math. Soc.* 291 (2) (1985), 582–612. [388](#), [397](#)
- [34] J. McCarthy and A. Papadopoulos, Dynamics on Thurston’s sphere of projective measured foliations. *Comment. Math. Helv.* 64 (1) (1989), 133–166. [388](#), [395](#), [398](#), [399](#)
- [35] C. McMullen, Teichmüller geodesics of infinite complexity. *Acta Math.* 191 (2) (2003), 191–223. [406](#)
- [36] H. Min, Hyperbolic graphs of surface groups. Ph.D. thesis, Rutgers University, Newark, 2007, in preparation. [400](#), [404](#)
- [37] L. Mosher, A hyperbolic-by-hyperbolic hyperbolic group. *Proc. Amer. Math. Soc.* 125 (12) (1997), 3447–3455. [402](#), [404](#)
- [38] L. Mosher, Problems in the geometry of surface group extensions. In *Problems on mapping class groups and related topics*, Proc. Sympos. Pure Math., Amer. Math. Soc. 74, 2006, 245–256. [388](#), [407](#)
- [39] J.-P. Otal, Le théorème d’hyperbolisation pour les variétés fibrées de dimension 3. *Astérisque* 235 (1996). [404](#)

- [40] A. Reid, Surface subgroups of mapping class groups. Problems on mapping class groups and related topics, In *Problems on mapping class groups and related topics*, Proc. Sympos. Pure Math., Amer. Math. Soc. 74, 2006, 257–268. [388](#)
- [41] H. Royden, Automorphisms and isometries of Teichmüller space. In *Advances in the Theory of Riemann Surfaces*, Ann. of Math. Stud. 66, Princeton University Press, Princeton, N.J., 1971, 369–383. [405](#)
- [42] J.-P. Serre, Rigidité du foncteur de Jacobi d'échelon $n \geq 3$. Appendix to A. Grothendieck, *Techniques de construction en géométrie analytique, X. Construction de l'espace de Teichmüller*, Séminaire Henri Cartan, no. 17, Secrétariat mathématique de l'ENS, Paris, 1960/61. [391](#)
- [43] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc.* 19 (1988), 417–431. [388](#), [392](#), [406](#)
- [44] W. A. Veech, Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. Math.* 97 (3) (1989), 553–583; erratum *ibid* 103 (2) (1991), 447. [405](#), [406](#)

Chapter 9

Deformations of Kleinian groups

Albert Marden

Contents

1	Introduction	411
2	Kleinian groups	412
	2.1 The 3-dimensional point of view	413
	2.2 The classical examples	418
	2.3 The convex core	419
	2.4 The Hyperbolization Theorem	420
	2.5 Uniqueness	421
3	Boundary deformations of geometrically finite groups	422
	3.1 The quasiconformal deformation space	423
4	Geometric deformations of the groups; the representation variety	425
	4.1 Deforming the convex core	427
5	The deformation space boundary	427
6	Deformation of surface groups	429
	6.1 Bers slices	430
	6.2 Earthquakes	432
	6.3 Pleated surfaces	434
	6.4 Complex scaling	435
	6.5 Complex earthquakes	438
7	Complex scaling for general laminations	438
	7.1 Sullivan's theorem	440
	7.2 Earthquake disks	441
	7.3 Appendix: holomorphic motion	442
	References	443

1 Introduction

The goal of this chapter is

- to discuss how and why the deformations of hyperbolic 3-manifolds are controlled by the deformation of the complex structures of their boundary components;
- to discuss the approach to deforming Riemann surfaces by directly deforming their hyperbolic structures.

In this exposition, we assume readers are familiar with the classical theory of Teichmüller spaces of Riemann surfaces, and with the hyperbolic plane \mathbb{H}^2 and space \mathbb{H}^3 . We will include short introductions to Kleinian groups and 3-manifolds, earthquakes, complex scaling, pleated surfaces, and in an appendix, holomorphic motions.

We thank Lee Mosher for helpful comments.

2 Kleinian groups

We will represent \mathbb{H}^3 by one of the two common models that best fits the situation at hand: (i) Upper halfspace (UHS) $\{(z, t) : z \in \mathbb{C}, t > 0\}$ with the metric $ds = \frac{|dx|}{t}$, or (ii) the ball $\{x \in \mathbb{R}^3 : |x| < 1\}$ with the metric $ds = \frac{2|dx|}{1-|x|^2}$. Here $|dx|$ is the euclidean distance element. These are referred to as the “conformal models” because they represent the angles correctly. Stereographic projection extends from $\mathbb{S}^2 \rightarrow \mathbb{C} \cup \infty$ to an isometry of the ball model onto the UHS model. Möbius transformations extend naturally from $\mathbb{C} \cup \infty \cong \mathbb{S}^2$ to UHS and form there the full group of orientation preserving isometries. Stereographic projection conjugates this group to the orientation preserving isometries of the ball model.

A *Kleinian group* is a discrete group G of (orientation preserving) Möbius transformations acting on hyperbolic 3-space \mathbb{H}^3 and its boundary at infinity which we will denote as $\partial\mathbb{H}^3 \cong \mathbb{S}^2$. For current expositions of the theory, see [62], [34], [47], [45].

For the Kleinian groups G employed in this chapter, we will assume that (i) G is nonelementary in that its limit set (see below) contains at least three points, (ii) G has no elliptic elements (no torsion) – except as explicitly mentioned otherwise, and (iii) G is finitely generated.

It formally follows from Theorem 2.3 that (iii) implies G is finitely presented [55].

The ordinary set of G is denoted by $\Omega(G) \subset \mathbb{S}^2$ and the limit set by $\Lambda(G) \subset \mathbb{S}^2$. The limit set $\Lambda(G)$ is the closed, perfect set which is the closure of the set of fixed points of G . It is also the set of accumulation points of the orbit $\{G(x)\}$ of any $x \in \mathbb{H}^3 \cup \partial\mathbb{H}^3$. Its complementary open set $\Omega(G)$ on $\partial\mathbb{H}^3$, which may be empty, is the largest open set in $\partial\mathbb{H}^3$ in which G is properly discontinuous. If $\Omega(G) \neq \mathbb{S}^2$, it has either one, two or infinitely many components each of which is either simply or infinitely connected.

The Ahlfors conjecture was that, if G is finitely generated, either $\Lambda(G) = \mathbb{S}^2$, or it has zero Lebesgue area. This is now a theorem; the *coup de grâce* was provided by confirmation of the Tameness Conjecture (see below).

Under our assumptions, $\mathcal{M}(G) = (\Omega(G) \cup \mathbb{H}^3)/G$ is a 3-manifold with boundary $\partial\mathcal{M}(G) = \Omega(G)/G$, if $\Omega(G) \neq \emptyset$; $\pi_1(\mathcal{M}(G)) \cong G$. Each component of $\partial\mathcal{M}(G)$ is a Riemann surface; its complex structure is inherited from that of \mathbb{S}^2 . On the other hand, the interior of $\mathcal{M}(G)$ has a complete hyperbolic structure. While $\partial\mathcal{M}(G)$ is “at ∞ ” in the metric, it is closely related to $\mathcal{M}(G)$ by virtue of the hyperbolic geometry of lines and planes in \mathbb{H}^3 .

The modern theory was ushered in by

Theorem 2.1 (Ahlfors’ Finiteness Theorem [2]). *If G is finitely generated, $\partial\mathcal{M}(G)$ has at most a finite number of components. Each component is a closed Riemann surface with at most a finite number of punctures (and, if elliptics are allowed, a finite number of cone points).*

The group G and manifold $\mathcal{M}(G)$ are called *geometrically finite* if G has a finite sided fundamental polyhedron in \mathbb{H}^3 ; in terms of $\mathcal{M}(G)$ this means that the manifold is “essentially” compact, that is, compact except for cusps (see below). Automatically, the interior of a geometrically finite manifold is homeomorphic to the interior of a compact 3-manifold, possibly with boundary. If G and $\mathcal{M}(G)$ are not geometrically finite, they are called *geometrically infinite*.

2.1 The 3-dimensional point of view

A simple loop $\gamma \subset \partial\mathcal{M}(G)$ is called *compressible* if it bounds an essential disk $D \subset \mathcal{M}(G)$. This is an embedded disk with the properties (i) it is not homotopic, relative to γ , into $\partial\mathcal{M}(G)$, and (ii) $D \cap \partial\mathcal{M}(G) = \partial D$. It is the famous Dehn’s Lemma and Loop Theorem that tells us that γ is compressible if and only if it is homotopic to a point in $\mathcal{M}(G)$ but not in $\partial\mathcal{M}(G)$.

The 3-manifold $\mathcal{M}(G)$ is said to be *boundary incompressible* if the inclusion $\pi_1(\partial\mathcal{M}(G)) \hookrightarrow \pi_1(\mathcal{M}(G))$ is injective on each boundary component. Equivalently, $\mathcal{M}(G)$ is boundary incompressible if and only if each component of $\Omega(G)$ is simply connected: There are no compressible loops.

We will also have cause to refer to *essential cylinders* C embedded in $\mathcal{M}(G)$. This means that (i) C is not homotopic, relative to ∂C , into $\partial\mathcal{M}(G)$, and (ii) $C \cap \partial\mathcal{M}(G) = \partial C$. Thus an annular region in the boundary is not an essential cylinder. Essential cylinders arise from two disjoint loops on $\partial\mathcal{M}(G)$ which are freely homotopic in $\mathcal{M}(G)$ but not in $\partial\mathcal{M}(G)$.

A manifold without essential cylinders is called *acylindrical*.

The role of parabolics. Before continuing with our story we have to come to terms with the special role played by parabolic transformations. Suppose ζ is the fixed point of a parabolic element of G . The group of all elements of G that fix ζ is either conjugate to $\langle z \mapsto z + 1 \rangle$, or is a free abelian group of rank two conjugate to $\langle z \mapsto z + 1, z \mapsto z + \tau \rangle$, for some τ with $\text{Im } \tau > 0$.

Within \mathbb{H}^3 there is a certain “universal horoball” at ζ which is preserved by the parabolic group but mapped disjoint to itself by every other member of G . The same size horoball works for all groups. In the rank two case, the boundary horosphere projects to a flat torus within $\mathcal{M}(G)$ called a *cusped torus*. It bounds the projection of the horoball which is called a *solid cusp torus*; topologically it is $\{w : 0 < |w| \leq 1\} \times \mathbb{S}^1$.

In the rank one case, the projection of the horosphere called a *cusped cylinder*. It bounds in $\mathcal{M}(G)$ the projection of the horoball which is called a *solid cusp cylinder*:

topologically it is just $\{w : 0 < |w| \leq 1\} \times (-\infty, +\infty)$. In the good cases, in particular in geometrically finite groups, the solid cusp cylinder can be replaced by a *solid pairing tube*. This arises from a pair of punctures on $\partial\mathcal{M}(G)$, paired by the following phenomenon. Take small simple loops surrounding each of the punctures. There is a cylinder $C \subset \mathcal{M}(G)$, bounded by these two loops, such that C is relative boundary of solid pairing tube of the form $\{w : 0 < |w| \leq 1\} \times [-1, +1]$. Up in $\Omega(G)$ the parabolic fixed point ζ supports a *double horodisk*: a pair of invariant, mutually disjoint disks, tangent at ζ ; the pair is bounded by a *double horocycle*.

Within the corresponding manifold, these two structures are referred to as *rank one cusps* and *rank two cusps*.

For a simple, explicit example consider the group $G = \langle z \mapsto z + 1 \rangle$; \mathbb{C}/G is a twice punctured sphere. The horocycles $\{z : \text{Im } z = \pm 1\}$ project to simple loops about the two punctures. They bound a pairing cylinder C formed by the projection of a piece of each vertical halfplane based on the horocycles, and a section of the horosphere $t = 1$ in UHS.

The Ahlfors Finiteness Theorem can be extended to state that a finitely generated group also has at most a finite number of conjugacy classes of rank one and rank two parabolic subgroups, see [27], [29].

Geometrically finite manifolds. We have already introduced the class of geometrically finite groups/manifolds. The central role this class plays in the theory is confirmed by the following theorem with details provided in [11], [12].

Theorem 2.2 (The Density Theorem). *Geometrically finite groups are dense in all finitely generated groups.*

The topology used here is the topology of algebraic convergence to be introduced in §4 below.

In a geometrically finite manifold all the punctures on $\partial\mathcal{M}(G)$ are arranged in pairs, paired by solid pairing tubes. The puncture pairs are in one-one correspondence with the conjugacy classes of rank one parabolic subgroups of G . In fact a geometrically finite manifold is characterized by the following fact: It has a finite number of solid cusp tori and solid pairing tubes such that when their interiors are removed, there results a compact manifold, which we have denoted by $\mathcal{M}_0(G)$. In the presence of cusps, the terms boundary incompressibility and acylindricity are often applied not to $\mathcal{M}(G)$ itself but to the compactified $\mathcal{M}_0(G)$.

Finite volume manifolds are geometrically finite. In particular, they cannot have any rank one cusps. Typical examples are most knot complements, see §2.4.

Geometrically infinite manifolds. The basic tool in exploring these is the following:

Theorem 2.3 (Scott Compact Core [56], [55]). *Assume G is finitely generated. There exists a compact, connected submanifold $C \subset \text{Int}(\mathcal{M}(G))$ with the properties (i) the inclusion $\pi_1(C) \hookrightarrow \pi_1(\mathcal{M}(G))$ is an isomorphism, and (ii) each complementary component of $\text{Int}(C)$ in $\text{Int}(\mathcal{M}(G))$ is bounded by a single component of ∂C .*

The submanifold $C = C(G)$ is called a *compact core*. The existence of the core, which was discovered independently by Shalen, was initially used for the purpose of proving that G is finitely presented.

Each boundary component S of C corresponds to the *end* E_S of $\mathcal{M}(G)$ determined by the complementary component of C bounded by S . The end E_S is called geometrically finite if there is a boundary component R of $\mathcal{M}(G)$ such that $E_S \cong S \times [0, 1)$ is bounded by S and R : S is parallel to R . But here we are interested in the geometrically infinite case. The boundary component R that “should be” parallel to S has gone missing. What is then the topology of E_S ? Has the missing boundary component R left some trace?

When there are parabolics, one often uses instead of the compact core as above, a different compact core that takes account of the effect of parabolics.

We will now use the notation $\mathcal{M}_0(G)$ for the result of removing the (open) solid cusp tori and solid cusp cylinders from $\mathcal{M}(G)$, which may not be geometrically finite. The *relative compact core* $C_{\text{rel}} = C_{\text{rel}}(G)$ [36] is constructed with respect to $\mathcal{M}_0(G)$ to have the following properties: (i) each torus component of $\partial\mathcal{M}_0(G)$ is a component of ∂C_{rel} , (ii) each cylinder component of $\partial\mathcal{M}_0(G)$ intersects ∂C_{rel} in a closed annular region, (iii) the inclusion $\pi_1(C_{\text{rel}}) \hookrightarrow \pi_1(\mathcal{M}(G))$ is an isomorphism, and (iv) each complementary component in $\text{Int}(\mathcal{M}_0(G))$ is bounded by a single component of ∂C_{rel} .

The *fundamental theorem of geometrically infinite manifolds* is this:

Theorem 2.4 (Tameness Theorem [1], [13]). *For any finitely generated group G , the interior of $\mathcal{M}(G)$ is homeomorphic to the interior of a compact manifold.*

In particular each end E_S is topologically $S \times [0, 1)$. The Tameness Theorem is far from obvious (think of wild spheres, etc.) and resolves the long standing conjecture called the Marden conjecture. It has many ramifications in the theory, not least in providing the final step needed to confirm the Ahlfors conjecture.

The Density Theorem 2.2 was first formulated as a conjecture by Bers for surface groups but quickly gained notoriety in the trade as applicable to all finitely generated groups. It was finally proved at the end of a long line of important preliminary results by many authors. It is a deep result as in its full implication it needs the Tameness Theorem and the Ending Lamination Theorem. We will not discuss the latter here but remain content to note the following. Thurston originally discovered, in certain cases, that each “missing” boundary component of a geometrically infinite manifold is characterized by an “ending lamination”, that can be realized as a lamination on the relevant boundary component S of the core $C(G)$. He conjectured that this situation was true in general. His vision has now been confirmed in full generality, using Minsky’s bilipschitz model of hyperbolic manifolds [48], by Brock, Canary and Minsky [11], [12]. The work also requires a deep study of the complex of simple curves on a surface carried out by Masur and Minsky. Rather than itemize all the contributions here we refer to the cited references for a full account and bibliography. We will say a bit more on this subject at the end of §5.

Extending boundary mappings. Suppose $F: \Omega(G) \rightarrow \Omega(H)$ is a quasiconformal mapping that induces an isomorphism $\varphi: G \rightarrow H$ and $f: \partial\mathcal{M}(G) \rightarrow \partial\mathcal{M}(H)$ is its projection. Does f , or a mapping homotopic to f , have an extension to a quasiconformal map $\mathcal{M}(G) \rightarrow \mathcal{M}(H)$? This is not true in general but it is true for geometrically finite manifolds $\mathcal{M}(G)$. However, there does not seem to be a “canonical” way to construct the extension. Instead, the extension can be constructed by topological means. Assume first that $\mathcal{M}(G)$ is compact.

Denote the components of $\partial\mathcal{M}(G)$ by $\{R_i\}$, and the components over them by $\{\Omega_{i,j}\} \subset \Omega(G)$. Denote the stabilizer of $\Omega_{i,j}$ by $G_{i,j}$. Then $G_{i,j} \cong \pi_1(R_i)/\pi_c(R_i)$ where $\pi_c(R_i) \subset \pi_1(R_i)$ denotes the normal subgroup generated by free homotopy classes of compressing loops on R_i . The subgroup $\pi_c(R_i)$ is also the kernel of the inclusion map $\iota: \pi_1(R_i) \hookrightarrow \pi_1(\mathcal{M}(G))$.

Citing [31, Theorem 13.9, Corollary 13.7], f is homotopic on $\partial\mathcal{M}(G)$ to a homeomorphism f_1 which has an extension to a homeomorphism between the manifolds $\mathcal{M}(G) \rightarrow \mathcal{M}(H)$. According to [50], f_1 in turn is homotopic to a (quasiconformal) diffeomorphism f_2 . Now f_2 is homotopic to f on $\partial\mathcal{M}(G)$. Choose the lift F_2 of f_2 to $\mathbb{H}^3 \cup \Omega(G)$ so that F_2 also induces φ and is homotopic to F on $\Omega(G)$. Now F_2 has a quasiconformal extension to all \mathbb{S}^2 .

If there are parabolics we have to replace the manifold by the compact manifold $\mathcal{M}_0(G)$ resulting from the removal of the solid pairing tubes and the solid cusp tori. The resulting mapping f_2 sends the pairing cylinders and cusp tori on $\mathcal{M}(G)$ to those on $\mathcal{M}(H)$. It then needs to be extended inside the solid pairing tubes and cusp tori.

Although it would suffice in the applications to replace (F, f) by (F_2, f_2) , we will present a stronger result.

We started above with a φ -equivariant quasiconformal map $F: \Omega(G) \rightarrow \Omega(H)$. We found a φ -equivariant map F_2 of \mathbb{S}^2 whose restriction to $\Omega(G)$ is homotopic to F . Using the density of the loxodromic fixed points in $\Lambda(G)$ one can verify:

Lemma 2.5. *F has an extension to a homeomorphism of \mathbb{S}^2 that satisfies $F(\zeta) = F_2(\zeta)$ for all $\zeta \in \Lambda(G)$.*

Set $H = F_2^{-1} \circ F: \Omega(G) \rightarrow \Omega(G)$. The map H is homotopic to the identity on each component of $\Omega(G)$, induces the identity automorphism of G , and is equal to the identity on $\Lambda(G)$. Let γ_z be a shortest geodesic from z to $H(z)$. There is a constant $C_1 < \infty$ such that $L_h(z) = d_h(z, H(z)) < C_1$ for all $z \in \Omega(G)$ (lift from the quotient). Here $d_h(\cdot, \cdot)$ denotes the shortest hyperbolic distance on $\Omega(G)$. From this it follows if $z \rightarrow \zeta \in \Lambda(G)$ in the spherical metric, then $\lim \gamma_z = \zeta$ uniformly on γ_z . That is, there exists a constant C_2 such that $d(\gamma_z, \Lambda(G)) < C_2$ for all $z \in \Omega(G)$. Here $d(\cdot, \cdot)$ denotes spherical distance. It follows that $d(w, \Lambda(G)) < C_2 d(z, \Lambda(G))$, for all $w \in \gamma_z$ and some $C_2 < \infty$ (actually we only need these estimates for $z \in \Omega(G)$ near a point $\zeta \in \Lambda(G)$).

I am grateful to Vlad Markovic for allowing inclusion of his unpublished result as follows.

Proposition 2.6 (Markovic). *H is quasiconformal on \mathbb{S}^2 ; hence F itself is the restriction to $\Omega(G)$ of an equivariant quasiconformal map of \mathbb{S}^2 .*

Proof. Markovic's proof is as follows. Set $X = \Lambda(G)$. From [51] we know that X has the property of uniform perfectness. That is, for the hyperbolic metric $\rho(w)|dw|$ in each component of $\Omega(G)$ and some constant $C_3 > 0$,

$$\frac{C_3|dw|}{d(w, X)} < \rho(w)|dw| < \frac{2|dw|}{d(w, X)}.$$

Upon integrating over a shortest geodesic γ_z of hyperbolic length $L_h(z)$ from z to $H(z)$, we find that $C_3d(z, H(z)) \leq L_h(z) \sup_{w \in \gamma_z} d(w, X) < L_h C_2 d(z, X)$. In other terms, $d(z, H(z)) < C_4 d(z, \zeta)$ for any $\zeta \in X$. Now $d(H(z), \zeta) \leq d(H(z), z) + d(z, \zeta)$. Consequently for some constant C_5 , $d(H(z), \zeta) < C_5 d(z, \zeta)$. The same holds if we replace z by $H(z)$. We conclude that

$$\frac{d(H(z), \zeta)}{C_5} \leq d(z, \zeta) \leq C_5 d(H(z), \zeta).$$

So the ratio of distances to ζ is uniformly bounded between 0 and ∞ as $z \rightarrow \zeta$. We are now in position to apply the geometric definition of quasiconformality to show that $H(\zeta)$ is quasiconformal at ζ . Since ζ was arbitrarily chosen this proves H is quasiconformal on $\Lambda(G)$. \square

After applying Theorem 2.9 (iii) we end up with the following useful results.

Theorem 2.7. *Suppose G is geometrically finite, and $F: \Omega(G) \rightarrow \Omega(H)$ is a quasiconformal map inducing an isomorphism $\varphi: G \rightarrow H$. Then F is the restriction of a φ -equivariant quasiconformal map $F: \mathbb{H}^3 \cup \mathbb{S}^2 \rightarrow \mathbb{H}^3 \cup \mathbb{S}^2$. It projects to a quasiconformal map $f: \mathcal{M}(G) \rightarrow \mathcal{M}(H)$.*

Corollary 2.8. *With the notation of Theorem 2.7, assume $H = G$ and $\varphi = \text{id}$ ($\varphi = \text{id}$ if and only if F pointwise fixes $\Lambda(G)$). Then $f: \text{Int}(\mathcal{M}(G)) \rightarrow \text{Int}(\mathcal{M}(H))$ is homotopic to the identity map.*

Corollary 2.8 is significant in the case that $\partial\mathcal{M}(G)$ is compressible and f is not homotopic to the identity as a self-mapping of each component of $\partial\mathcal{M}(G)$.

Instead of a mapping between boundary components, we will now start afresh with a quasiconformal mapping $F: \mathbb{S}^2 \rightarrow \mathbb{S}^2$ that induces an isomorphism $\varphi: G \rightarrow H$ satisfying $F(g(z)) = \varphi(g)F(z)$ for all $g \in G, z \in \mathbb{S}^2$; here G and H are not necessarily geometrically finite. If the restriction of F to $\Omega(G)$ is conformal $\Omega(G) \rightarrow \Omega(H)$ then F is Möbius and the two groups are conjugate. Here we are again using the fact that $\Lambda(G)$ has zero area. We now have a choice of extension theorems to \mathbb{H}^3 and $\mathcal{M}(G)$.

Theorem 2.9 (Basic Extension Theorems). *Suppose G, H are arbitrary Kleinian groups and $F: \mathbb{S}^2 \rightarrow \mathbb{S}^2$ is an equivariant K -quasiconformal mapping that induces an isomorphism $\varphi: G \rightarrow H$. Then:*

- (i) [18], [67] *The map F has an equivariant extension to \mathbb{H}^3 that is a homeomorphism which also induces φ ; its projection $f: \mathcal{M}(G) \rightarrow \mathcal{M}(H)$ is an orientation preserving homeomorphism.*
- (ii) [66] *The map F has an equivariant (L, a) -quasi-isometric extension for some $L = L(K), a = a(K)$; its projection $f: \mathcal{M}(G) \rightarrow \mathcal{M}(H)$ is a (L, a) -quasi-isometric mapping.*
- (iii) [39, Corollary B.23] *The map F has an equivariant extension to a $K^{3/2}$ -bilipschitz C^1 -diffeomorphism of \mathbb{H}^3 ; its projection $f: \mathcal{M}(G) \rightarrow \mathcal{M}(H)$ is a $K^{3/2}$ -bilipschitz C^1 -diffeomorphism (and hence a quasiconformal map).*

A mapping f of \mathbb{H}^3 is called (L, a) -quasi-isometric if there exist finite constants $1 \leq L$ and $a \geq 0$ such that in the hyperbolic metric

$$\frac{1}{L}d(x, y) - a \leq d(f(x), f(y)) \leq Ld(x, y) + a.$$

Thus a quasi-isometric map need not be continuous but at long range it is essentially bilipschitz. It is L -bilipschitz if $a = 0$. Quasi-isometric maps can be extended to $\partial\mathbb{H}^3$ and the extension is a quasiconformal mapping of \mathbb{S}^2 .

2.2 The classical examples

Fuchsian and quasi-Fuchsian groups. A Fuchsian group is a discrete group G that preserves a round disk in \mathbb{S}^2 , which we may regard as \mathbb{H}^2 . If $R = \mathbb{H}^2/G$ then $\mathcal{M}(G) \cong R \times [0, 1]$. One can see the fibering explicitly in the upper halfspace model by taking the family of euclidean halfplanes in upper halfspace, bounded by the real axis. This is also a good place to see the pairing of punctures between R and R' , the G -quotient of the exterior of the disk or upper halfplane used to form R .

A quasi-Fuchsian group is a quasiconformal deformation of a Fuchsian group G : there exists a quasiconformal map $f: \mathbb{S}^2 \rightarrow \mathbb{S}^2$ that induces an isomorphism $\varphi: G \rightarrow H$ by $f \circ g(z) = \varphi(g) \circ f(z)$ for all $z \in \mathbb{S}^2$ and all $g \in G$. Such f are characterized by their *Beltrami differentials* $\mu = f_{\bar{z}}/f_z$ with $\mu(g)\bar{g}'/g' = \mu$ for all $g \in G$ and a.e. $z \in \mathbb{S}^2$. It is a classical result of Bowen and Sullivan that the limit set $\Lambda(G)$ is the quasiconformal image of a circle and has Hausdorff dimension strictly between one and two, unless it is again a circle. The 3-manifold $\mathcal{M}(G)$ retains the product structure $R \times [0, 1]$. Conversely, if a manifold $\mathcal{M}(G)$ has such a product structure, G is quasi-Fuchsian.

Schottky groups. These are groups generated as follows. Take an even number of round disks in \mathbb{C} with disjoint closures, Arrange them in pairs. For each pair choose a Möbius transformation T_i that sends the exterior of one disk to the interior of its

partner. The group $G = \langle T_1, \dots, T_g \rangle$ is called a Schottky group. More generally, the disks need only be regions bounded by Jordan curves but one must postulate the existence of pairing transformations that take the exterior of one Jordan curve to the interior of its partner. Schottky groups are free groups and, if $g \geq 2$, $\Lambda(G)$ is totally disconnected of positive Hausdorff dimension.

If G is Schottky, $\mathcal{M}(G)$ is a handlebody bounded by the closed surface $R = \Omega(G)/G$ of genus $g \geq 1$ ($g = 1$ arises from a cyclic group for which $\mathcal{M}(G)$ is directly seen to be a solid torus – a bagel). A handlebody of genus g is a compact 3-manifold such that there are g mutually disjoint compressing curves on R bounding mutually disjoint disks in $\mathcal{M}(G)$. If $\mathcal{M}(G)$ is cut along the disks, there results a topological ball. conversely if an $\mathcal{M}(G)$ has this structure, G is Schottky – but is not necessarily generated from pairs of circles. For a full analysis of circle generated Schottky groups see the recent paper [28].

2.3 The convex core

The *convex core* $\mathcal{C}(G)$ of $\mathcal{M}(G)$ is the smallest (hyperbolically) convex submanifold of $\mathcal{M}(G)$ that contains all the geodesics. It may well coincide with $\mathcal{M}(G)$.

The convex core is constructed in \mathbb{H}^3 as the convex hull $\mathcal{C}^*(G)$ of $\Lambda(G)$ and then projected to $\mathcal{M}(G)$. The construction is as follows.

Before introducing a group, consider a region $\Omega \subset \mathbb{C}$ without isolated boundary points, $\Omega \neq \mathbb{C}$. A maximal disk $D \subset \Omega$ is a round disk that is not contained in any larger disk. Its closure intersects $\partial\Omega$ in at least two points. We will call the circle $C = \partial D$ a maximal circle. Erect the hyperbolic plane $P_C \subset \mathbb{H}^3$ (in the upper halfspace model) rising from C . It divides \mathbb{H}^3 into two halfspaces. Denote by $H(P_C)$ the halfspace bordering $\mathbb{S}^2 \setminus D$. The convex hull of the *complement* in \mathbb{S}^2 of Ω is defined to be

$$\mathcal{C}^* = \bigcap_C H(P_C),$$

the intersection being taken over all maximal circles C in Ω .

The relative boundary $\text{Dome}(\Omega) = \partial\mathcal{C}^* \subset \mathbb{H}^3$ lies over Ω as the dome lies over the floor in a domed stadium. It is the union of “flat pieces” and “bending lines”. Each flat piece is an open hyperbolic polygon, possibly with infinitely many edges, contained in some plane P_C . The components of the complementary closed set are geodesics with endpoints in $\partial\Omega(G)$. They are called bending lines.

An isolated bending line (isolated on both sides) is the common edge of two flat pieces. There is a well determined (exterior) *bending angle* $0 < \theta < \pi$ measured so that the limiting case of zero bending means there is no bending at all. There is a *bending measure* μ defined on transverse segments τ where in our applications we can take τ as a geodesic segment with endpoints in flat pieces. The measure is constructed by taking the limit of the bending of finite approximates to the dome [20].

Suppose now that Ω is simply connected. The ambient hyperbolic metric in \mathbb{H}^3 restricts to a path metric on $\text{Dome}(\Omega)$, referred to as its hyperbolic metric. A result of Thurston is that there is an isometry of $\text{Dome}(\Omega)$ onto \mathbb{H}^2 , uniquely determined up to postcomposition with a Möbius transformation; it may be regarded as the Riemann map.

There is a canonical map $r(z) : \Omega \rightarrow \text{Dome}(\Omega)$, called the *nearest point retraction* that relates Ω to its dome: Given $z \in \Omega$, consider the family of horospheres in \mathbb{H}^3 based at z . There is one that meets the dome at exactly one point, which is defined to be $r(z)$. If there are no isolated bending lines, r is a homeomorphism. If there is an isolated bending line ℓ with bending angle α then there are two planes P_1, P_2 that intersect with angle α along ℓ . Denote by P'_1, P'_2 the planes orthogonal to P_1, P_2 along ℓ . The exterior bending angle α is also the vertex angle of the crescent in Ω determined by P'_1, P'_2 . This is the crescent which is the preimage of ℓ under r . The retraction r extends to $\partial\Omega = \partial\text{Dome}(\Omega)$ and fixes every point.

The nearest point retraction r is known to be 2-lipschitz in the respective hyperbolic metrics [23]. If in addition Ω is euclidean convex, r is also 2-quasiconformal [24].

Finally we will bring in a Kleinian group G . Construct the convex hull $\mathcal{C}^*(G)$ of $\Lambda(G)$ by repeating our construction on each component of $\Omega(G)$, if $\Omega(G) \neq \emptyset$. Then $\mathcal{C}^*(G)$, like $\Omega(G)$, is invariant under G . Its projection $\mathcal{C}(G)$ to $\mathcal{M}(G)$ is called the *convex core* of $\mathcal{M}(G)$. The projection of r maps $\partial\mathcal{M}(G)$ onto the relative boundary $\partial\mathcal{C}(G) \subset \text{Int}(\mathcal{M}(G))$. If $\partial\mathcal{M}(G) = \emptyset$ then we set $\mathcal{C}(G) = \mathcal{M}(G)$.

Chris Bishop used $\text{Dome}(\Omega)$ to develop a new way to numerically estimate the Riemann map of a simply connected $\Omega \rightarrow \mathbb{D}$. He makes use of the crescent system under a finite approximation to the dome to deform Ω back to \mathbb{D} [7].

A manifold is geometrically finite if and only if its convex core has finite volume.

2.4 The Hyperbolization Theorem

In the 60s and early 70s as the theory was being developed, the unsettling question that increasingly came to the fore was this. Is the class of hyperbolic manifolds a rather sporadic class within all 3-manifolds, or does it represent a large class? The answer was provided by Thurston in the mid 70s by bringing in startlingly original and powerful new tools.

His Hyperbolization Theorem concerns orientable compact 3-manifolds M^3 with possible boundary with the following two properties:

- Every embedded 2-sphere bounds a ball (irreducibility).
- Every map f of a torus T into M^3 is homotopic to a map of T into ∂M^3 , provided $f_* : \pi_1(T) \hookrightarrow \pi_1(M^3)$ is injective (atoroidality).

It is known every M^3 may be cut along a finite number of spheres and incompressible tori so that the resulting pieces have these two properties, after capping off the exposed spheres with balls.

Theorem 2.10 (The Hyperbolization Theorem). *There are two cases.*

- (i) *If $\partial M^3 \neq \emptyset$ then the interior of M^3 has a complete, geometrically finite, hyperbolic structure.*
- (ii) *If M^3 is a closed manifold that contains an incompressible surface then M^3 has a hyperbolic structure.*

An incompressible surface is one containing no compressing loop bounding a disk in its complement. There exist such surfaces whenever $\pi_1(M^3)$ is “sufficiently large”, which is the case if the first homology group contains infinitely many elements.

For Thurston’s exposition see [63] and for complete accounts of the proof see [34], [52], [53].

It is conjectured that a closed, irreducible manifold M^3 is hyperbolic whenever $\pi_1(M^3)$ contains no rank two abelian subgroups. Confirmation of this would completely decide the hyperbolization question.¹

There is a sharper form of the theorem that allows specification of rank one parabolic subgroups – pairs of punctures.

A striking corollary drawn by Thurston is that a knot complement in \mathbb{S}^3 has a hyperbolic structure if and only if it is not a torus knot or a satellite knot.

2.5 Uniqueness

The following theorem appeared not long after Ahlfors’ Finiteness Theorem. It was quite surprising at the time as hyperbolic surfaces of finite area are most definitely not rigid.

Theorem 2.11 (Mostow’s Rigidity Theorem). *If $\mathcal{M}(G)$ has finite volume and $\varphi: G \rightarrow H$ is an isomorphism to a Kleinian group H , then φ is induced by an isometry $\mathcal{M}(G) \rightarrow \mathcal{M}(H)$.*

Mostow originally proved his theorem [49] for closed manifolds. It was extended to finite volume manifolds independently by [42] (in dimension 3) and [54] (in $n \geq 3$ dimensions).

Taking account of boundary components, Mostow’s theorem was generalized to manifolds with boundary in [42] as follows:

Theorem 2.12. *Suppose G is a geometrically finite Kleinian group and $F: \Omega(G) \rightarrow \Omega(H)$ is a conformal map that induces an isomorphism $\varphi: G \rightarrow H$ by $F \circ g(z) = \varphi(g) \circ F(z)$ for all $z \in \Omega(G)$ and $g \in G$. Then F is the restriction of a Möbius transformation and $\mathcal{M}(H)$ is isometric to $\mathcal{M}(G)$.*

The proof uses the fact that F has an extension to \mathbb{H}^3 that in turn extends to a quasiconformal map of \mathbb{S}^2 . This was discussed in §2.1.

¹*Added in proof:* As a consequence of Grigori Perelman’s recent work, this is now known to be true.

3 Boundary deformations of geometrically finite groups

In this section we will work with a geometrically finite $\mathcal{M}(G)$ with nonempty boundary. Its boundary is the union of a finite number of Riemann surfaces $\{R_i\}$, each of which is a closed surface of genus $g_i \geq 0$, with $n_i \geq 0$ punctures. Since each component of $\Omega(G)$ has a hyperbolic metric the same is true of each R_i .

As we have seen elsewhere in this volume, the Teichmüller space $\mathcal{T}(R_i)$ is a complex manifold of dimension $3g_i + n_i - 3$. This is positive except in the case that R_i is the triply punctured sphere.

We would like to apply basic Teichmüller theory in succession to each component of $\partial\mathcal{M}(G)$. Namely suppose f is a quasiconformal map of each component R_i to another Riemann surface. Let μ_i denote its Beltrami differential on R_i . Provided each component of $\Omega(G)$ is simply connected and hence serves as the universal covers of the $\{R_i\}$, we can lift the $\{\mu_i\}$ to a Beltrami differential μ on $\Omega(G)$ by setting $\mu(\pi^{-1}(z)) = \mu_i(z)$, if $z \in R_i$. The collection μ of the lifted $\{\mu_i\}$ on $\Omega(G)$ satisfies $\mu(z)\bar{g}'/g' = \mu(z)$ for all $g \in G$ and a.e. $z \in \mathbb{S}^2$,

Solve the Beltrami equation on \mathbb{S}^2 , using the fact the limit set has zero area. The solution F is quasiconformal on \mathbb{S}^2 and is uniquely determined up to postcomposition by a Möbius transformation, which can be used to normalize: we may assume $(0, 1, \infty) \in \Lambda(G)$ and require that F fix each of these points. The map F induces an isomorphism $\theta: G \rightarrow H$ to another group H , which necessarily geometrically finite as well. In view of Proposition 2.7 we may assume the restriction of F to $\Omega(G)$ projects to a quasiconformal map $f: \partial\mathcal{M}(G) \rightarrow \partial\mathcal{M}(H)$ that extends to a quasiconformal map $\mathcal{M}(G) \rightarrow \mathcal{M}(H)$.

On the one hand if we regard $\partial\mathcal{M}(G)$ simply as a union of Riemann surfaces with no particular relation to the interior of the manifold or to deformations of G itself then we would find that the Teichmüller space of $\partial\mathcal{M}(G)$ is simply the abstract product of the Teichmüller spaces of the individual boundary surfaces. When all components of $\Omega(G)$ are simply connected, they are representations of the universal covers of the components of $\partial\mathcal{M}(G)$. The classical theory is then operative in that nontrivial deformations of components of $\partial\mathcal{M}(G)$ correspond to nontrivial deformations of G .

So all is fine if $\mathcal{M}(G)$ is boundary incompressible. Otherwise we have a little problem.

Consider an infinitely connected component $\Omega_{i,j}$ over R_i that is stabilized by $G_{i,j}$. For example $\Omega_{i,j}$ could be the ordinary set of a Schottky group. Consider a compressing loop $\gamma^* \subset \Omega_{i,j}$, the lift of a compressing loop $\gamma \in R_i \subset \partial\mathcal{M}(G)$. For example γ may be part of a generating set of $\pi_1(R_i)$, say a loop about a handle. It certainly figures in the deformation of the Riemann surface R_i . But since γ^* determines the identity element of $G_{i,j} \subset G$ it can play no role at all in the deformation of G .

Fix a basepoint $O \in R_i$ for each i . Consider the normal subgroup $\pi_c(R_i) \subset \pi_1(R_i)$ consisting of loops, basepoint O , freely homotopic to compressing loops of R_i . The quotient group $\pi_1(R_i)/\pi_c(R_i)$ is isomorphic to the stabilizer $G_{i,j}$ of any component

$\Omega_{i,j}$ of $\Omega(G)$ over R_i . A automorphism h of R_i lifts to a component $\Omega_{i,j}$ if and only if $h_*(\pi_c(R_i)) = \tau\pi_c(R_i)\tau^{-1}$, for some arc τ .

The way out of the conudrum has already been suggested in our discussion in §2.1 of extending boundary maps, especially Corollary 2.8. We will consider quasiconformal mappings $F: \Omega(G) \rightarrow \Omega(H)$ that induce isomorphisms $\theta: G \rightarrow H$ in the sense that $F \circ g = \theta(g) \circ F$ for all $g \in G$. We may assume by Proposition 2.7 that F extends to a quasiconformal mapping of \mathbb{S}^2 . Denote the projection by $f: \partial\mathcal{M}(G) \rightarrow \partial\mathcal{M}(H)$.

3.1 The quasiconformal deformation space

Define the quasiconformal deformation space of G to be

$$\mathfrak{T}(G) = \{\theta: G \rightarrow H \text{ is an isomorphism induced by a qc map } F: \mathbb{S}^2 \rightarrow \mathbb{S}^2\} / \equiv .$$

The quasiconformal map F satisfies $F \circ g(z) = \theta(g) \circ F(z)$ for all $g \in G$ and $z \in \mathbb{S}^2$. Here “ \equiv ” means that mappings F that give rise to conjugate isomorphisms are to be identified.

We had *normalized* quasiconformal mappings F of \mathbb{S}^2 by replacing G by a conjugate so that $(0, 1, \infty) \in \Lambda(G)$ and then requiring that F fix these points. Equivalently then, $\mathfrak{T}(G)$ can be defined in terms of equivalence classes of normalized quasiconformal mappings F of \mathbb{S}^2 which induce isomorphisms $\theta: G \rightarrow H$ where the equivalence relation is: $F_1 \sim F_2$ if and only if the projection $f_2^{-1} \circ f_1: \partial\mathcal{M}(G) \rightarrow \partial\mathcal{M}(G)$ extends to a mapping $\text{Int}(\mathcal{M}(G)) \rightarrow \text{Int}(\mathcal{M}(G))$ homotopic to the identity.

Alternatively, $F_1 \sim F_2$ (i) if and only if F_1 and F_2 induce the same isomorphism $\theta: G \rightarrow H$, or (ii) if and only if $F_1(\zeta) = F_2(\zeta)$ for all $\zeta \in \Lambda(G)$. In the latter case, if $F_2^{-1} \circ F_1$ is conformal on a component Ω of $\Omega(G)$, it is the identity there, since it extends to the identity on $\partial\Omega$. For further details consult [46].

We have yet to explain the relation of $\mathfrak{T}(G)$ to the product of the classical Teichmüller spaces of the components: $\mathcal{T}(R_1) \times \cdots \times \mathcal{T}(R_k)$.

To do this we have to recognize that a stronger equivalence can be imposed:

$$F_1 \simeq F_2 \text{ if and only if } f_2^{-1} \circ f_1: \partial\mathcal{M}(G) \rightarrow \partial\mathcal{M}(G) \text{ is homotopic to id.}$$

It follows that $F_1 \sim F_2$ in the earlier definition. These two equivalences differ only when $\partial\mathcal{M}(G)$ is compressible.

To mirror the difference in the two equivalence relations we introduce the group $X(G)$ consisting of equivalence classes of quasiconformal automorphisms $h: \partial\mathcal{M}(G) \rightarrow \partial\mathcal{M}(G)$ which extend to maps $\mathcal{M}(G) \rightarrow \mathcal{M}(G)$ that are homotopic to the identity on the interior – Corollary 2.8. Two such maps h_1, h_2 are equivalent $h_1 \simeq h_2$ if and only if $h_2^{-1} \circ h_1$ is homotopic to the identity on $\partial\mathcal{M}(G)$; specifically, $h_2^{-1} \circ h_1$ maps each component R_i onto itself and is homotopic on R_i to the identity. If $\mathcal{M}(G)$ is boundary incompressible, $X(G) = \text{id}$.

Denote by $\text{Mod}_0(R_i)$ the group of homotopy classes of quasiconformal mappings $h: R_i \rightarrow R_i$ which extend to $\mathcal{M}(G)$ to be homotopic in $\text{Int}(\mathcal{M}(G))$ to the identity map of the interior. If the $\Omega_{i,j}$ are simply connected, $\text{Mod}_0(R_i) = \text{id}$. Such a map h fixes the punctures on R_i , and the set of compressing loops. Extend h to all $\partial\mathcal{M}(G)$ by setting it equal to the identity on $R_m, m \neq i$. Then h extends to $\mathcal{M}(G)$ and (its restriction to the interior) is homotopic in $\text{Int}(\mathcal{M}(G))$ to the identity. In other terms, h is the projection of a quasiconformal automorphism H of each component $\Omega_{i,j}$ over R_i with the property that H induces the identity automorphism of $\text{Stab}(\Omega_{i,j})$ and extends continuously to the identity map of $\partial\Omega_{i,j}$. A typical element of $\text{Mod}_0(R_i)$ arises from a Dehn twist h about a compressing curve γ of some R_i ; γ bounds a disk D within $\mathcal{M}(G)$. Thicken D to D' and extend h , first to D' so that it is the identity on each face of D' , and then to all $\mathcal{M}(G)$.

Consequently the group $X(G)$ splits into a direct product

$$X(G) = \text{Mod}_0(R_1) \times \text{Mod}_0(R_2) \times \cdots \times \text{Mod}_0(R_k).$$

The group $\text{Mod}_0(R_i)$ acts without fixed points on $\mathcal{T}(R_i)$. For suppose, for example, that $h \in \text{Mod}_0(R_i)$ fixes the origin (R_i, id) in $\mathcal{T}(R_i)$. Then h is homotopic to a conformal map $h_0: R_i \rightarrow R_i$. Now h and then h_0 lift to automorphisms h^* and h_0^* of $\Omega_{i,j}$ over R_i ; we can choose h^* to be homotopic in $\Omega_{i,j}$ to h_0^* . We know that h^* extends continuously to $\partial\Omega_{i,j}$ and fixes every point. The same is therefore true of h_0^* . But h_0^* is the identity since it is a conformal automorphism. Therefore h^* is homotopic in $\Omega_{i,j}$ to the identity and h is homotopic in R_i to the identity.

Finally classical results using a projection from the space of Beltrami differentials with respect to G on $\Omega(G)$ imply that $\mathfrak{T}(G)$ is a complex analytic manifold.

As a consequence of our discussion, we arrive at a 3D interpretation of the following result proved earlier in terms of Fuchsian equivalents of the surfaces $\{R_i\}$ by Bers [3], Maskit [46], and Kra [35]:

Theorem 3.1. *Suppose G is geometrically finite. Denote the components of $\partial\mathcal{M}(G)$ by $\{R_i\}$. Then*

$$\mathfrak{T}(G) = \mathcal{T}(R_1)/\text{Mod}_0(R_1) \times \cdots \times \mathcal{T}(R_k)/\text{Mod}_0(R_k).$$

Here $\text{Mod}_0(R_i)$ is the fixed point free subgroup of biholomorphic automorphisms of $\mathcal{T}(R_i)$ generated by automorphisms of R_i that extend to be homotopic to the identity in the interior of $\mathcal{M}(G)$.

The product $\mathcal{T}(R_1) \times \cdots \times \mathcal{T}(R_k)$ is the universal cover of $\mathfrak{T}(G)$; they are identical if and only if $\partial\mathcal{M}(G)$ is incompressible. The spaces $\mathfrak{T}(G)$ and $\mathcal{T}(R_i)/\text{Mod}_0(R_i)$ are complex analytic manifolds; $\mathfrak{T}(G)$ has dimension

$$\sum_{i=1}^m (3g_i + n_i - 3),$$

where g_i is the genus of the i 'th component of $\partial\mathcal{M}(G)$ and n_i is the number of its punctures.

If $\mathcal{M}(G)$ is geometrically infinite and boundary incompressible Theorem 3.1 still holds; in this case $\text{Mod}_0(R_i) = \text{id}$. Thus if $\Omega(G)$ is connected and simply connected (a singly degenerate group in $\partial\mathcal{B}(G)$ – see §6.1) then $\mathfrak{T}(G)$ has dimension $(3g + n - 3)$. In contrast if G_1 is Fuchsian with \mathbb{H}^2/G_1 conformal to $\Omega(G)/G$ then $\mathfrak{T}(G_1)$ has dimension $(6g + 2n - 6)$. If all components R_i are triply punctured spheres, G is quasiconformally rigid, geometrically finite or not.

Remark 3.2. Suppose all components Ω_i are simply connected but that there may be torsion in G_i . Then in addition to the punctures on each component $R_i \subset \partial\mathcal{M}(G)$ there will be $b_i \geq 0$ “branch” or “cone” points. In this case the dimension count will be

$$\sum (3g_i + b_i + n_i - 3).$$

For it is an interesting fact that $\mathcal{T}(R_i)$ is biholomorphically equivalent to $\mathcal{T}(R'_i)$ where R'_i is the result of removing the cone points. That is, the dimension is the same whether you have $b_i + n_i$ punctures, or n_i punctures and b_i cone points.

The basis for the equivalence is the following fact [41], [5]: A homeomorphism $f: R_i \rightarrow R_i$ lifts to a homeomorphism f^* of Ω_i which induces the identity automorphism of G_i if and only if f is homotopic in R'_i to the identity map. This follows from the fact that γ is freely homotopic in R'_i to $f(\gamma)$ for all simple loops $\gamma \subset R'_i$.

There is an analogue of Theorem 3.1 that covers the case that G has elliptics; see [46], [35] for details.

4 Geometric deformations of the groups; the representation variety

Assume G is geometrically finite. What we will call its *representation variety* (actually it is the quotient of an open subset of an affine algebraic variety) consists of all its type preserving homomorphisms onto nonelementary subgroups of $\text{PSL}(2, \mathbb{C})$:

$$\mathfrak{R}(G) = \{\text{type preserving homomorphisms of } G \text{ to nonelementary groups}\} / \equiv$$

A homomorphism φ is called type preserving if $\varphi(g)$ is parabolic whenever $g \in G$ is so. Here “ \equiv ” means that two homomorphisms φ, φ_1 are identified if $\varphi_1(G) = T\varphi(G)T^{-1}$, for some Möbius transformation T ; we do not wish to distinguish between conjugate groups. In this general context *elementary groups* are those with the property that any two elements of infinite order have a common fixed point.

We are assuming the targets are nonelementary because it is natural to do so in the context of the theory of Kleinian groups, and also because, with this assumption, $\mathfrak{R}(G)$ is a Hausdorff space. The largest class of groups left out are purely elliptic groups, that is, groups conjugate to a group of rotations of \mathbb{S}^2 .

The group G has a finite presentation. We can imagine changing the entries in the generating matrices in such a way that the new matrices continue to satisfy the

relation equations. Doing so will give a homomorphic image. The image group H is not likely to be discrete, or even finitely presented, unless we remain close enough to the initial G .

The natural topology in $\mathfrak{R}(G)$ is given by the topology of *algebraic convergence*. A sequence $\{\langle T_{1,k}, \dots, T_{N,k} \rangle\}$ of N -generator groups converges, necessarily to $\langle T_1, \dots, T_N \rangle$, if and only if $\lim_{k \rightarrow \infty} T_{i,k} = T_i$ is a Möbius transformation for all i .

For our purposes the most important part of $\mathfrak{R}(G)$ is its *discreteness locus*. This is the following *closed* subset:

$$\mathfrak{R}_{\text{disc}}(G) = \{\theta \in \mathfrak{R}(G) : \theta : G \rightarrow H \text{ is an isomorphism to a discrete group } H\}.$$

It is closed because of Jørgensen's theorem [32] that the algebraic limit H of a sequence of isomorphisms $\{\theta_n : G \rightarrow G_n\}$ to nonelementary, discrete groups is likewise discrete and nonelementary; moreover $\theta = \lim \theta_n : G \rightarrow H$ is an isomorphism.

Theorem 4.1. $\mathfrak{T}(G)$ is the connected component of $\text{Int}(\mathfrak{R}_{\text{disc}}(G))$ that contains the identity.

Theorem 4.1 is a direct consequence of the fact [42] that $\mathfrak{T}(G)$ is open in $\mathfrak{R}(G)$ for a geometrically finite group G . This is because the quasiconformal deformations of G are “strongly stable” in Bers’ terminology: Given $\varepsilon > 0$ and a quasiconformal image $H = \varphi(G)$, there is a neighborhood N of H in $\mathfrak{R}(G)$ such that for each group $H' = \rho(G) \in N$, there is a $(1 + \varepsilon)$ -quasiconformal map of \mathbb{S}^2 that induces the isomorphism $\rho\varphi^{-1} : H \rightarrow G \rightarrow H'$. Geometrically infinite groups in $\mathfrak{R}_{\text{disc}}(G)$ are not similarly stable.

In particular, the dimension of $\mathfrak{R}(G)$ is the same as that of $\mathfrak{T}(G)$.

Now the image H of an isomorphism $\theta \in \text{Int}(\mathfrak{R}_{\text{disc}}(G))$ is geometrically finite and is a quasiconformal deformation of G . But the isomorphism θ *itself* may not be induced by a quasiconformal map of \mathbb{S}^2 . Thus the interior of the discreteness locus may have many components only one of which is $\mathfrak{T}(G)$.

In fact each component \mathcal{D} of $\text{Int}(\mathfrak{R}_{\text{disc}}(G))$ is itself the quasiconformal deformation space of any fixed $\theta : G \rightarrow G' \in \mathcal{D}$.

The definitive study of the components of $\mathfrak{R}_{\text{disc}}(G)$ is carried out in [15] (there, the notation $\mathcal{A}\mathcal{H}(G)$ is used for what here is denoted $\mathfrak{R}_{\text{disc}}(G)$). In brief, the number of components of the interior is determined by the number of different homotopy equivalences between $\mathcal{M}(G)$ and the target manifolds, or equivalently, the number of essentially different isomorphisms between G and the target groups. In the case G is without parabolics, and $\partial\mathcal{M}(G) \neq \emptyset$, there are only a finite number of manifolds which are homotopy equivalent but not homeomorphic to $\mathcal{M}(G)$ [61]. If G is a Fuchsian closed surface group, then $\text{Int}(\mathfrak{R}_{\text{disc}}(G)) = \mathfrak{T}(G)$.

4.1 Deforming the convex core

Assume for simplicity that $\mathcal{M}(G)$ is compact, boundary incompressible, and acylindrical. Recall that corresponding to each component R_i of $\partial\mathcal{M}(G)$ is a relative boundary component C_i of the convex core $\mathcal{C}(G)$ parallel to it. Each C_i is a simply connected, embedded, pleated surface (see §§6.2–6.3) with bending lamination (Λ_i, μ_i) ; (Λ, μ) denotes the vector of individual laminations. Conversely we have the following special case of a more general result of Bonahon and Otal.

Theorem 4.2 ([10]). *Suppose $\mathcal{M}(G)$ compact, boundary incompressible and acylindrical. Assume we are given a measured lamination (Λ, μ) on $\partial\mathcal{M}(G)$. There exists a new geometrically finite hyperbolic structure $\mathcal{M}(G_\mu)$ on $\mathcal{M}(G)$ whose convex core boundary $\partial\mathcal{C}(G_\mu)$ has the bending lamination (Λ, μ) if and only if on each closed leaf γ of Λ , $0 < \mu(\gamma) < \pi$.*

If Λ is the union of a finite number of closed leaves, then G_μ is uniquely determined by (Λ, μ) up to Möbius equivalence.

Consider the case that the bending lamination consists of a maximal system of $\sum(3g_i - 3)$ simple closed geodesics on the convex hull boundary $\mathcal{C}(G_\mu)$; here g_i is the genus of the i 'th component of $\partial\mathcal{M}(G)$. Choi and Series [17] show that the complex lengths of these $\sum(3g_i - 3)$ -geodesics serve as parameters for the local deformations of G_μ in its representation variety $\mathfrak{R}(G_\mu)$: The lift to \mathbb{H}^3 of a leaf ℓ determines an element $T \in G_\mu$. The complex length of ℓ is $\log k + i\theta \pmod{2\pi i}$ if $T: z \mapsto ke^{i\theta}z, k > 1$, in normalized form.

As yet, the theory of bending laminations is not complete.

5 The deformation space boundary

The closure $\overline{\mathfrak{T}(G)}$ of $\mathfrak{T}(G)$ in $\mathfrak{R}(G)$ lies in $\mathfrak{R}_{\text{disc}}(G)$. It is an interesting fact [34, Theorem 8.44], [60] (see also [40, appendix]) that $\mathfrak{T}(G)$ is the interior of $\overline{\mathfrak{T}(G)}$. That is, $\mathfrak{T}(G)$ is not like an open ball with a slit to the boundary removed.

Denote the boundary of $\mathfrak{T}(G)$ in $\mathfrak{R}_{\text{disc}}(G)$ by $\partial\mathfrak{T}(G)$. The study of the boundary is perhaps the most fascinating aspect of the theory; as we approach a boundary point from the interior, some sort of degeneration of the manifold must occur. The topology of $\overline{\mathfrak{T}(G)}$ is interesting as well.

In general, $\overline{\mathfrak{T}(G)}$ is not compact and not a manifold [40]. However it is compact if $\mathcal{M}(G)$ is acylindrical (Thurston's Compactness Theorem [64]). For when there are essential cylinders, the cylinder can be squeezed to a point – the cyclic fundamental group of the cylinder becomes a cyclic parabolic group. This process destroys G . For a simple example just squeeze to a point a simple geodesic not parallel to a boundary component on \mathbb{H}^2/F where F is Fuchsian and watch what happens to F .

The topology we have been using in $\mathfrak{R}(G)$ is the topology of algebraic convergence: a sequence of N -generator groups converges to a group, if their generators converge and generate the group. The algebraic limit of a sequence $\{G_n\}$ of nonelementary N -generator discrete groups is also discrete and nonelementary [32].

But there is a fundamentally different notion of convergence, first discovered by Troels Jørgensen. It is akin to “tangential convergence”, thinking of a sequence in the unit disk converging tangentially to the unit circle, whereas algebraic convergence is akin to radial convergence. It relates to the convergence of the quotient manifolds rather than convergence of the generators of the fundamental group.

A sequence of N -generator groups $\{G_n\}$ is said to *converge geometrically* to a group H if and only if the following hold:

- (i) If $\lim g_n = h$ exists, $g_n \in G_n$, then $h \in H$.
- (ii) If $h \in H$ then $h = \lim g_n$, $g_n \in G_n$.

The group H is called the geometric limit of $\{G_n\}$. As was first pointed out by Jørgensen, H is not necessarily finitely generated!

The basic fact needed to explore the boundary is this:

Theorem 5.1 ([33]). *Suppose the sequence $\{G_n\}$ of N -generator, nonelementary Kleinian groups converges algebraically to G . Then there is a geometrically converging subsequence $\{G_k\}$. The limit H of any geometrically converging subsequence contains G ; consequently $\mathcal{M}(H)$ is a covering manifold of $\mathcal{M}(G)$.*

Furthermore, there is a sequence of surjective homomorphisms $\{\psi_k : H \rightarrow G_k\}$ such that $\lim \psi_k(h) = h$ for all $h \in H$.

The following result, part (i) of which builds on prior work of a number of researchers, gives considerable insight into when a geometric limit of N -generator groups differs from the algebraic. We have formulated it to take account of the Tameness Theorem.

Theorem 5.2. *Suppose $\theta_n : G \rightarrow G_n$ is a sequence of isomorphisms converging algebraically to H . The sequence also converges geometrically to H under one of the following situations:*

- (i) [26] *If $\Omega(H) \neq \emptyset$, and G has no “new parabolics”.*
- (ii) [14] *If $\Omega(G) = \emptyset$.*

In each case $\lim \Lambda(G_n) = \Lambda(H)$ in Hausdorff convergence.

No new parabolics means that $h \in H$ is parabolic only when $\theta_n \theta^{-1}(h)$ is parabolic for all large indices. It is not required that the algebraic limit be geometrically finite, which it is not, in general.

There is a countable number of special points on $\partial\mathfrak{T}(G)$ called *cusps*. A cusp is a geometrically finite boundary group. It comes from pinching simple loops on $\partial\mathcal{M}(G)$.

In §6.1 we will discuss this in more detail. A *maximal cusp* is a cusp H with $\partial\mathcal{M}(H)$ a union of triply punctured spheres. Maximal cusps are rigid.

One of the deepest and most fascinating topics in the theory is to describe the nature of (i) the geometrically infinite boundary groups, and (ii) the geometric limits at the boundary. Fundamental in this investigation is the Tameness Theorem 2.4 which says that each end of $\text{Int}(\mathcal{M}(G))$ is homeomorphic to a product $S \times [0, 1)$ where S is a finitely punctured compact hyperbolic surface. An end of a geometrically finite manifold is a neighborhood of a Riemann surface boundary component; for this case the conclusion of the tameness theorem has long been known. It is for geometrically infinite ends that the theorem is not at all obvious.

As already mentioned in §2.1, Thurston discovered the trace of the “missing boundary component” of each infinite end: The end E_S is characterized by a sequence of geodesics $\{\gamma_n\}$ that exit it. As a consequence of the Tameness theorem, the geodesics $\{\gamma_n\}$ are homotopic to simple geodesics $\{\gamma_n^*\}$ on S such that the sequence $\{\gamma_n^*\}$ converges to a geodesic lamination called the *ending lamination*. When S , which represents the end, is incompressible the ending lamination is uniquely determined, otherwise what is determined uniquely is a certain equivalence class of laminations. Thurston conjectured that homeomorphism $f: \mathcal{M}(G) \rightarrow \mathcal{M}(H)$ between two hyperbolic manifolds is homotopic to an isometry if and only if corresponding ending laminations are the same, for the infinite ends, and the restriction of f to the boundary is homotopic to a conformal map $\partial\mathcal{M}(G) \rightarrow \partial\mathcal{M}(H)$ for the others. The recent confirmation of this Ending Lamination Conjecture was a *tour de force* due to Yair Minsky [48] together with Brock, Canary, and Masur, see [11], [12] and the references there.

On the other hand, Soma [58] completely described the topological structure of the geometric limits on the boundary of quasi-Fuchsian space of a closed surface. A similar analysis should work in the presence of parabolics and more generally for geometric limits at boundary cusps on the deformation space of any geometrically finite group (personal communication). Soma showed that “anything goes” as topologically allowed by the situation. At each cusp there is in general a huge range of different geometric limits, including many that are infinitely generated. It seems likely that the methods of Minsky et al will give additional insight into the geometric structure of the geometric limits.

6 Deformation of surface groups

In the analytic theory, perhaps the most useful realization of Teichmüller space $\mathcal{T}(R)$ is as the *Bers embedding* onto a certain open subset of a Banach space of quadratic differentials on R . Yet viewing $\mathcal{T}(R)$ in this way hides a lot of additional structure that goes with identifying the images in the Banach space with a space of quasi-Fuchsian groups.

Fix a hyperbolic Riemann surface $R = \mathbb{H}^2/G$. In this section we will present three methods for deforming the Fuchsian group G and surface R within $\mathfrak{A}(G)$. The first is analytic in character, using quasiconformal mappings and then projective structures. In this approach, $\mathcal{T}(R)$ will appear as a “Bers slice” in the quasiconformal deformation space $\mathfrak{T}(G)$. The second and third approaches are geometric; the hyperbolic structure on R will be deformed by earthquaking or more generally, by complex scaling.

6.1 Bers slices

We will obtain our “slice” by the method of *simultaneous uniformization* discovered by Bers.

Fix a Fuchsian group G acting in the upper halfplane (UHP) and of course in LHP as well. The anticonformal involution $J^*(z) = \bar{z}$ induces an anticonformal involution $J: R^{\text{top}} = \text{UHP}/G \leftrightarrow R_{\text{bot}} = \text{LHP}/G$. The involution J extends to a anticonformal involution of $\mathcal{M}(G)$ interchanging the two boundary components and inducing the identity automorphism of $\pi_1(\mathcal{M}(G)) \cong G$.

Suppose $\varphi: G \rightarrow H \in \mathfrak{T}(G)$ is induced by the quasiconformal map F of \mathbb{S}^2 . Denote its projection by $f: \partial\mathcal{M}(G) \rightarrow \partial\mathcal{M}(H)$. The f -images of the top and bottom components of $\partial\mathcal{M}(G)$ are called the *top and bottom* boundary components of $\partial\mathcal{M}(H)$. The quasiconformal group H carries a natural anti-quasiconformal involution J^* that exchanges the two components of $\Omega(H)$ while pointwise fixing $\Lambda(H)$. Its projection J to $\partial\mathcal{M}(H)$ (which extends to an involution of all $\mathcal{M}(H)$) interchanges the two boundary components while inducing the identity automorphism of $\pi_1(\mathcal{M}(H))$.

The natural involution J^* and its projection J , which are uniquely determined up to homotopy, will be denoted by J_H^* and J_H when we want to indicate the group.

Theorem 6.1 (Simultaneous uniformization [3]). *Suppose we are given two Riemann surfaces $(S^{\text{top}}, S_{\text{bot}})$ of finite hyperbolic area together with an anti-quasiconformal involution $J: S_{\text{bot}} \leftrightarrow S^{\text{top}}$. There exists a quasi-Fuchsian group H , uniquely determined up to Möbius equivalence, such that the top boundary component of $\mathcal{M}(H)$ is conformally equivalent to S^{top} , the bottom to S_{bot} , and J is homotopic to the natural involution J_H .*

The result H of simultaneous uniformization however is not a “marked” group, its relation to the basepoint G is not specified. A marking is specified by choosing a normalized quasiconformal map F of \mathbb{S}^2 that (i) projects to a map f of $\partial^{\text{top}}\mathcal{M}(R) \rightarrow \partial^{\text{top}}\mathcal{M}(H)$, and $\partial_{\text{bot}}\mathcal{M}(G) \rightarrow \partial_{\text{bot}}\mathcal{M}(H)$, and (ii) f satisfies $f \circ J_G = J_H \circ f$, up to homotopy. Via f , a given marking on $\partial_{\text{bot}}\mathcal{M}(G)$ determines a marking on S_{bot} and an isomorphism $\varphi: G \rightarrow H \in \mathfrak{T}(G)$.

Denote the marked quasi-Fuchsian group $H = ((S^{\text{top}}, S_{\text{bot}}; J_H), f)$. These groups fill out the quasi-Fuchsian space $\mathfrak{T}(G)$.

The *Bers slice* based on $R' = R_{\text{bot}} = \text{LHP}/G$ is defined as the set of points

$$\mathfrak{B}(R') = \mathfrak{B}(G) = \{H = ((S, R_{\text{bot}}; J), f)\} / \cong \subset \mathfrak{T}(G).$$

That is $\mathfrak{B}(G)$ consists of the normalized quasiconformal deformations F of G which are conformal on LHP, and which project to a quasiconformal map f of $R^{\text{top}} = \text{UHP}/G$ but a *conformal* map f on R_{bot} .

The slice is an analytic submanifold of $\mathfrak{T}(G)$ based on the origin G . Actually there is a Bers slice based on any point of $\mathfrak{T}(G)$, and even based on any point of $\partial\mathfrak{T}(G)$. Each of these slices is biholomorphically equivalent to the Teichmüller space $\mathcal{T}(R^{\text{top}})$.

The action of the mapping class group on the slice is as follows. If α is an automorphism of R_{bot} ,

$$\alpha : ((S, R_{\text{bot}}; J), f) \mapsto ((S, R_{\text{bot}}; J \circ \alpha), f).$$

The action of α does not change the conformal type of the bounding surfaces. What it changes is the topological relationship of the top to the bottom, the marking of the top relative to that of the bottom. As a result the quasi-Fuchsian group H will change to a new group H_1 in the slice, unless α represents a symmetry of H in that H_1 is conjugate to H – that is α fixes the corresponding Teichmüller point. The automorphism α is a biholomorphic automorphism of $\mathfrak{B}(G)$.

Now the compactness of the space of normalized conformal mappings of LHP implies that the closure $\overline{\mathfrak{B}(G)}$ in the representation variety $\mathfrak{R}(G)$ is compact. Therefore the Bers slice has a compact boundary $\partial\mathfrak{B}(R')$ which lies in $\mathfrak{R}_{\text{disc}}(G)$. Each boundary point corresponds to a Kleinian group H with $\partial_{\text{bot}}\mathcal{M}(H)$ conformally equivalent to R_{bot} , but the top boundary has been somehow degenerated.

The boundary, called the *Bers* or *analytic* boundary, is very interesting; it is analogous to the boundary $\mathbb{R} \cup \infty$ of the Teichmüller space of the torus. There is a certain countable dense set [38] of distinguished points on it called *cusps*. A cusp corresponds to a geometrically finite group H where R^{top} has become “pinched”. Thus if R^{top} is a closed surface of genus 2, the top component of $\partial\mathcal{M}(H)$ will be one twice-punctured torus, or two once-punctured tori, or four thrice-punctured spheres, etc. Moreover the augmented space $\mathfrak{B}(R) \cup \{\text{cusps}\}$ can be topologized by designating “horocyclic” neighborhoods of the cusps. Then its quotient with respect to the mapping class group is a compactification of moduli space. For an exposition see [43].

The other boundary points H correspond to groups with geometrically infinite top ends. If H has no new parabolics, then $\Omega(H)$ is connected and simply connected. The bottom surface $\Omega(H)/H$ is conformally equivalent to R_{bot} while the top surface has gone missing. Such a group is called *singly degenerate*.

For closed surfaces, the slice $\mathfrak{B}(G)$ is itself the intersection with $\mathfrak{T}(G)$ of a properly embedded analytic submanifold $\mathfrak{B}^*(G) \subset \mathfrak{R}(G)$ of the representation variety, called the *extended Bers slice* [30]. It is obtained by constructing all complex projective structures on R_{bot} . This means consider all Schwarzian differential equations $\{F; z\} = \phi$ on LHP where ϕ is the lift of a holomorphic quadratic differential on R_{bot} . There is a solution, uniquely determined up to postcomposition with a Möbius, which determines a homomorphism φ of G to a nonelementary subgroup of $\text{PSL}(2, \mathbb{C})$. In addition φ has the special property that it lifts to a homomorphism between corresponding matrix groups in $\text{SL}(2, \mathbb{C})$. The totality of all the homomorphisms make up

the extended slice; if the norm of ϕ not too large, the solution F will be a conformal map of LHP, thus determining a point in $\mathfrak{B}(R')$.

David Dumas has computed spectacular pictures of $\mathfrak{B}^*(G) \cap \text{Int}(\mathfrak{A}_{\text{disc}})(G)$ showing the Bers slice surrounded by an archipelago of islands. He did this in the quasi-Fuchsian space of a once-punctured torus (www.math.brown.edu/dumas).

6.2 Earthquakes

A geodesic lamination $\Lambda \subset \mathbb{H}^2$ is a closed set of mutually disjoint (hyperbolic) lines referred to as *leaves*. However two leaves may have a common end point.

If Λ is invariant under a Fuchsian group G , it projects to a lamination on the surface $R = \mathbb{H}^2/G$.

To introduce Thurston's theory of earthquakes [65] consider the upper halfplane model and a line ℓ which we may take to be the positive imaginary axis. Denote left and right quadrants determined by ℓ by A and B ; A and B have orientations inherited from \mathbb{C} . From the point of view of A , a *left earthquake* with fracture line ℓ is a discontinuous map which fixes A pointwise, and in B is an isometry moving B to the left with respect to A , that is, it moves B in the positive direction with respect to the positive orientation of ∂A . In B it therefore has the form $z \mapsto kz$, $k > 1$. It is uniquely determined once the displacement $\log k$ along ℓ is dictated.

On the other hand if we require that B be fixed, the left earthquake along ℓ moves A to the left from the point of view of a person standing in B . Namely in A it has the form $z \mapsto k^{-1}z$.

Next suppose we have a lamination with a finite number of leaves. Fix a gap σ – a component of $\mathbb{H}^2 \setminus \Lambda$ – as the base of operations. Suppose μ is a positive transverse measure: that is, to each leaf of the lamination is assigned a positive number as atomic measure. Normalize the earthquake to be the identity on σ . A transverse geodesic based in σ will cross a number of leaves. Carry out a sequence of left earthquakes in sequence along the various leaves, using the displacement assigned by μ .

Formal definition of earthquake. Suppose $\Lambda \subset \mathbb{H}^2$ is a geodesic lamination. A *left earthquake* is a possibly discontinuous injective and surjective map $E: \mathbb{H}^2 \rightarrow \mathbb{H}^2$ which is an isometry on each leaf of Λ and on each gap. Given two gaps and/or leaves $X \neq Y$, a line ℓ is said to be weakly separating if any path from a point of X to a point of Y intersects ℓ . Let E_X, E_Y denote the respective isometric restrictions of E . We require that (i) the *comparison isometry* $E_X^{-1} \circ E_Y$ be a loxodromic Möbius transformation T , (ii) the axis ℓ of T weakly separates X and Y , and (iii) T translates Y to the left, when viewed from X . The requirement (iii) means that the direction of translation along ℓ agrees with the orientation induced from $X \subset \mathbb{H}^2 \setminus \ell$. The case that one of X, Y is a line in the boundary of the other is exceptional in that the comparison map is the identity.

The earthquake maps Λ to another lamination Λ' . The inverse of a left earthquake is a right one.

If Λ has a finite number of leaves, left earthquakes are constructed as illustrated earlier. Thurston proves that these finite earthquakes are dense in all left earthquakes, in the topology of uniform convergence on compact sets.

A left earthquake between two Riemann surfaces is an injective, surjective map which lifts to a left earthquake of \mathbb{H}^2 . In particular Λ is invariant under the deck transformations. However if one or more leaves of Λ project to simple geodesics, lifts are determined only up to Dehn twists along the geodesics. To avoid this ambiguity one can associate the earthquake with the homotopy type of a homeomorphism between the surfaces. A more common way is to start with both an invariant lamination in \mathbb{H}^2 , and an invariant transverse measure (more of this below).

Theorem 6.2 (Earthquake Theorem [65]). *Every continuous orientation preserving map $\partial\mathbb{H}^2 \rightarrow \partial\mathbb{H}^2$ is the boundary values of a left earthquake E of \mathbb{H}^2 . The lamination Λ is uniquely determined. On Λ , E is uniquely determined except along those leaves ℓ on which it is discontinuous. For each such ℓ , there is a range of choices of translations ranging between the limiting values of E on the two sides; all the choices have same image in \mathbb{H}^2 .*

Suppose $R_i = \mathbb{H}^2/G_i$, $i = 1, 2$, are arbitrary Riemann surfaces with possible boundary contours ∂R_i coming from discontinuity of the action of G_i on maximal open intervals on $\partial\mathbb{H}^2$. Assume $h: R_1 \rightarrow R_2$ is an (orientation preserving) homeomorphism which extends to a continuous map $\partial R_1 \rightarrow \partial R_2$. Then the boundary values on $\partial\mathbb{H}^2$ of a lift of h are the boundary values of a left earthquake of \mathbb{H}^2 . It projects back to a left earthquake $E: R_1 \rightarrow R_2$. E has the same uniqueness indicated above.

This is a very general theorem. The second statement (which includes the first) follows from the first as lifts of h extend to continuous maps of $\partial\mathbb{H}^2$.

Associated to any left earthquake is a nonnegative Borel measure μ transverse to Λ . Two earthquakes corresponding to the same (Λ, μ) have isometric images. The measure is constructed by a process akin to Riemann integration (see [20]) based on the example of a finite number of leaves given above. If Λ is invariant under a Fuchsian group, μ will be as well.

Normally one only works with the restricted class of *locally bounded* earthquakes. These are the class of earthquakes whose transverse measures have the property that for some $K < \infty$, $\mu(\tau) < K$ for all transverse geodesic segments τ of unit length. If this is the case for (Λ, μ) invariant under a group G , the leaves on \mathbb{H}^2/G cannot end at a puncture, nor can they spiral in to a closed geodesic.

The boundary values on $\partial\mathbb{H}^2$ of locally bounded earthquakes are quasymmetric homeomorphisms, which means their boundary values have quasiconformal extensions to \mathbb{H}^2 (which are equivariant if (Λ, μ) is invariant under deck transformations). In the other direction, the boundary values of a quasiconformal mapping $\mathbb{H}^2 \rightarrow \mathbb{H}^2$, for example the boundary values of a lift of a quasiconformal map between surfaces, is also the boundary values of a locally bounded left earthquake as described in the Earthquake Theorem. For the details, consult [65].

The bottom line is that if we have Riemann surfaces R, S of finite hyperbolic area, given a quasiconformal map $f: R \rightarrow S$, there is a geodesic lamination and a locally bounded left earthquake E along it that also sends R to S . The homotopy class of f is essentially the same as that of E . Conversely every locally bounded earthquake is essentially homotopic to a quasiconformal map.

Earthquakes map one Fuchsian group G to another. Thus they give rise to the real Teichmüller space of deformations of Fuchsian groups. In §6.4 we will show how to complexify the technique so as to get a complex analytic structure.

6.3 Pleated surfaces

We have spoken of the structure of a convex hull boundary component, especially the dome over a simply connected region. Now consider the reverse process. That is, given a measured lamination (Λ, μ) in \mathbb{H}^2 , can we construct a generalized surface in \mathbb{H}^3 whose bending measure is μ ?

Let's start with the simplest cases. Take the equatorial plane \mathbb{H}^2 (the unit disk) in the ball model and fix a diameter ℓ . Bend \mathbb{H}^2 along ℓ with exterior bending angle $0 < \theta < \pi$. Here $\theta = 0$ corresponds to no bending at all. The other extreme $\theta = \pi$ corresponds to two situations: (i) folding \mathbb{H}^2 in half along ℓ , or (more commonly) (ii) pushing ℓ out to ∞ to become a single point ξ thereby forcing \mathbb{H}^2 in the limit to become two hyperbolic planes in \mathbb{H}^3 whose boundaries are tangent at ξ .

To normalize the direction of bending, bend so that the result lies in the upper half of the ball. The resulting "pleated surface" S bounds on one side a convex region whose floor on $\partial\mathbb{H}^3$ is bounded by two circular arcs with interior bending angle $\pi - \theta$. The dome has only one bending line.

The construction is easily generalized to a finite system of ordered, mutually disjoint hyperbolic lines, possibly with common endpoints, $\ell_1, \dots, \ell_k \subset \mathbb{H}^2$. Assign an exterior bending angle $0 < \theta_i < \pi$ to each line. Then systematically bend the plane \mathbb{H}^2 . For example we may assume that first bend along ℓ_1 results in $P_1 = P$ constructed above. Then in P_1 locate the copy of ℓ_2 , say it lies to the right of ℓ_1 . Then bend the halfplane in P_1 lying to the right of ℓ_1 along ℓ_2 with exterior angle θ_2 . And so on for all the lines. We end up with what is called a *pleated surface* P_k which is not necessarily embedded in \mathbb{H}^3 – it may well have self intersections. It has k bending lines, the images of the $\{\ell_i\}$. In any case there is a hyperbolic isometry $\Upsilon: \mathbb{H}^2 \rightarrow P_k$ – such that Υ^{-1} is just unbending. The finite measured lamination is carried to the bending lines and bending measure on P_k .

Given a geodesic lamination $\Lambda \subset \mathbb{H}^2$ and a transverse measure μ with values in $(-\infty, +\infty)$, a corresponding pleated surface can be constructed by using finite approximations [20]. The construction is such that if (Λ, μ) is invariant under a Fuchsian group G , then the construction of pleated surface automatically determines a deformation of G to a homomorphic image H . This is a group of Möbius transformations acting in \mathbb{H}^3 that map the pleated surface onto itself in a manner reflecting the action

of G in \mathbb{H}^2 , but H is unlikely to be discrete and the pleated surface may not be locally convex or even locally embedded.

Formally, a *pleated surface* is determined by a *pleating map* $f: R \rightarrow \mathcal{M}(G)$ of a hyperbolic surface R with the properties

- (i) f takes any rectifiable path in R to a path in $\mathcal{M}(G)$ of the same length;
- (ii) every $z \in R$ lies in an open geodesic arc which f maps to a geodesic arc in $\mathcal{M}(G)$;
- (iii) f sends cusps to cusps: the homomorphism $f_*: \pi_1(R) \rightarrow \pi_1(\mathcal{M}(G)) \cong G$ sends parabolics to parabolics.

Assumption (i) can be replaced by (i'): geodesic paths in S are sent to rectifiable paths of the same length in $\mathcal{M}(G)$. This apparently stronger definition is equivalent to (i) [16, I.5.2.6]. The pleated surface is called *incompressible* if $f_*: \pi_1(R) \rightarrow \pi_1(\mathcal{M}(G))$ is injective.

The *pleating locus* is the set $\Lambda \subset S$ consisting of those points $z \in R$ with the following property. There is one and only one open geodesic arc (up to inclusion) through z which f maps onto a geodesic arc in $\mathcal{M}(G)$. The pleating locus Λ is a closed subset of R and is in fact a geodesic lamination. The image $f(\Lambda)$ is often referred to as the *pleating locus* as well, or as the *bending lines*. The map f is an isometry of the complementary gaps onto in general infinitely sided polygons in $\mathcal{M}(G)$.

Given such a general pleated surface, there is likely to be a great deal of positive and negative bending. Yet by associating a transverse segment τ to the set of positive endpoints on $\partial\mathbb{H}^3$ of the oriented leaves through τ and then to a continuum in $\partial\mathbb{H}^3$, it is possible to construct a kind of bending measure called an “ \mathbb{R} -valued transverse cocycle”. This real valued measure, which is now only finitely additive, and the pleating locus $\Lambda \subset R$ characterize the pleated surface. It is very interesting that $\mathcal{T}(R)$ is filled out by the images under a rather explicit finite dimensional space of these cocycles. For the presentation of Bonahon’s theory of *shear cocycles* see [9], [8].

6.4 Complex scaling

This section is based on the exposition [22] where the reader will find additional details.

A crescent is a region bounded by two circular arcs with vertex angles $0 < \alpha < \pi$. It is conformally equivalent to the wedge $W_\alpha = \{z \in \mathbb{C} : 0 < \arg z < \alpha\}$. The *complex scaling* map with parameter $t \in \mathbb{C}$ is defined on the wedge W_α to be,

$$f_+(z) = f_+(re^{i\theta}) = re^{t\theta} e^{i\theta} = re^{u+i(1+v)\theta}, \quad t = u + iv, \quad z = re^{i\theta} \in W_\alpha.$$

In parallel with $f_+(z)$ define

$$f_-(z) = e^{-t\theta} z, \quad z \in W_\alpha.$$

We list the elementary properties for f_+ ; the map f_- has complementary properties.

- For all $-1 < \text{Im}(t) \leq +1$, the map f is a K_t -quasiconformal homeomorphism, and for all $-1 < \text{Im}(t) < \infty$, the map f is K_t -quasiregular where

$$K_t = \frac{1 + |\kappa(t)|}{1 - |\kappa(t)|}, \quad \kappa(t) = \frac{t}{2i - t} \cdot \frac{z}{\bar{z}}. \tag{6.1}$$

- The boundary values of f on each edge of W_α is the restriction of a Möbius transformation.
- f commutes with any Möbius transformation that maps W_α onto itself.

For larger angles, we must think of the image of W_α as spread over \mathbb{S}^2 .

Important special cases for f_+ are:

- (i) $\text{Re}(t) = u = 0$. Then $f(z) = e^{iv\theta}z$. This is a smoothed version of “grafting”. The wedge W_α of angle α is mapped onto the wedge $W_{(1+v)\alpha}$.
- (ii) $t = i$. Then $f(z) = re^{2i\theta}$ is called the *angle doubling* map. It is 2-quasiconformal.
- (iii) $v = \text{Im}(t) = 0$. Then $f(z) = e^{u\theta}z$. This is a smoothed version of an earthquake. The wedge W_α is mapped onto itself, but it is stretched (or contracted if $u < 0$) continuously from its right edge to its left.

Next, suppose we have a pair of symmetric wedges with respect to \mathbb{R} ,

$$W_+ = \{z : 0 \leq \beta < \arg z < \alpha + \beta < \pi\}, \quad W_- = \{z : -\alpha - \beta < \arg z < -\beta \leq 0\}.$$

Define f_+ in W_+ by conjugating W_+ to W_α via the rotation $z \mapsto e^{-i\beta}z$. Define f_- in W_- by conjugating it via the rotation $z \mapsto e^{i\beta}z$ to the action of f_- in the reflected image of W_α in \mathbb{R} .

Interpolating by the identity and a loxodromic transformation, we end up with the following *continuous* map $E_t(z) : \mathbb{C} \rightarrow \mathbb{C}$:

- $E_t(z) = f_+(z)$, if $z \in W_+$,
- $E_t(z) = f_-(z)$, if $z \in W_-$,
- $E_t(z) = z$, for z in the component of $\mathbb{C} \setminus (W_- \cup W_+)$ containing the positive real axis,
- $E_t(z) = e^{t\alpha}z$ for z in the component of $\mathbb{C} \setminus (W_- \cup W_+)$ containing the negative real axis.

Thus $E_0(z) = z$ while the limiting case E_{-i} collapses the crescents (vertex angles $\rightarrow 0$).

The mapping E_t is the basis of the angle scaling method.

The application is to a discrete G -invariant geodesic lamination $\Lambda \subset \mathbb{H}^2$, where G is Fuchsian; we can also take $G = \text{id}$. Typically we take Λ to be the set of lifts of a finite set of mutually disjoint simple closed geodesics on $R = \mathbb{H}^2/G$. In this case we can find mutually disjoint tubular neighborhoods of the geodesics on R . The totality of lifts to \mathbb{H}^2 of these neighborhoods forms a system of mutually disjoint crescents about the corresponding geodesics. We can take a maximal system of crescents in the

sense each boundary component of one crescent is tangent to a boundary component of another.

More generally we do not need to take the crescents to be centered on its associated geodesic or even to contain it. All we need is a set of crescents with the following three properties, expressed in the disk model \mathbb{D} :

- Corresponding to each leaf ℓ is exactly one crescent whose vertices are the endpoints of ℓ on $\partial\mathbb{D}$.
- The crescent interiors are mutually disjoint.
- The set of crescents is invariant under the action of G .

The reflection in $\partial\mathbb{D}$ gives a set of crescents in the exterior \mathbb{D}^{ext} of \mathbb{D} with the same vertex angles and associated with the reflection of Λ .

Denote by α_ℓ the vertex angle of the crescent associated with the leaf ℓ

With the given data consider the measured lamination

$$(\Lambda, \mu) = \{\ell, \alpha_\ell\}, \quad 0 < \alpha_\ell < \pi, \quad \ell \in \Lambda,$$

and its reflection in $\partial\mathbb{D}$. Apply the angle scaling map E_t to the infinite discrete system of pairs of crescents in $\mathbb{D} \cup \mathbb{D}^{\text{ext}}$. The map E_t is normalized by choosing a component C' of $\mathbb{D} \setminus \bigcup\{\text{crescents}\}$ and requiring E_t to be the identity on C' . Concentrating on the case that $\text{Im}(t) > -1$ we arrive at the following result. The term “holomorphic motion” will be explained in §7.3.

Theorem 6.3 ([22]). *Suppose (Λ, μ) is a G -invariant, discrete, measured lamination in \mathbb{D} associated with a system of crescent pairs in $\mathbb{D} \cup \mathbb{D}^{\text{ext}}$ as described above. There exists a G -equivariant complex scaling mapping $E_t(z) = E_t(z; \Lambda, \mu)$ for every $t \in \mathbb{C}$. It has the following properties.*

- (i) *For $\{-1 < \text{Im}(t) < +1\}$, $E_t: \mathbb{S}^2 \rightarrow \mathbb{S}^2$ is a K_t -quasiconformal homeomorphism. Furthermore $\{E_t\}$ is a holomorphic motion of \mathbb{S}^2 .*
- (ii) *For $\{-1 < \text{Im}(t)\}$, $E_t: \mathbb{D} \rightarrow \mathbb{S}^2$ is a K_t -quasiregular developing mapping. For each $z \in \overline{\mathbb{D}}$, $E_t(z)$ is defined and holomorphic in t .*
- (iii) *When $t = i$, the angle doubling map $E_i: \mathbb{D} \rightarrow \Omega_i = E_i(\mathbb{D})$ is a 2-quasiconformal homeomorphism which has a continuous extension to $\partial\mathbb{D}$. It collapses the crescents in \mathbb{D}^{ext} .*

The scaling map E_t can be decomposed as follows:

$$E_t = E_{iv} \circ E_u, \quad t = u + iv.$$

The map $E_u: \mathbb{S}^2 \rightarrow \mathbb{S}^2$ is a homeomorphism, its actions in \mathbb{D} and the exterior \mathbb{D}^{ext} being symmetric. Its action in \mathbb{D} is as a smoothed out earthquake sending Λ to a new lamination Λ_u ; we write $E_u: (\Lambda, \mu) \rightarrow (\Lambda_u, \mu)$ with the understanding that each leaf of Λ_u takes its measure from its preimage. Equally E_u sends a crescent system for (Λ, μ) to one for (Λ_u, μ) . The values of E_u on $\partial\mathbb{D}$ are exactly the boundary values of a classical earthquake. By reflection, E_u acts in \mathbb{D}^{ext} as well.

In contrast, the action of E_{iv} on Λ_u can be described as *bending*. We will describe this action on \mathbb{D} , there is a similar action on \mathbb{D}^{ext} . When $-1 < v < +1$, $E_{iv} : \overline{\mathbb{D}} \rightarrow \overline{E_{iv}(\mathbb{D})} = \overline{\Omega}_v$ is a quasiconformal homeomorphism sending \mathbb{D} to a new region Ω_v and the crescent C_ℓ associated with the leaf $\ell \subset \Lambda_u$ of vertex angle α_ℓ to a crescent C' in Ω_v of vertex angle $(1 + v)\alpha_\ell$. There is an isometry $E_{iv}^* : \mathbb{D} \rightarrow \text{Dome}(\Omega_v)$ that sends the leaf $\ell \subset \Lambda_u$ to the bending line ℓ' of exterior bending angle $v\alpha$. To complete the picture $r^{-1}(\ell') \subset \Omega_v$ is a crescent $C^* \subset \Omega_v$ of vertex angle $v\alpha_\ell$, where $r : \Omega_v \rightarrow \text{Dome}(\Omega_v)$ is the nearest point retraction. The two crescents C', C^* in Ω_v have the same vertices. There is a similar interpretation in terms of the pleated surface generated by (Λ, μ) for all $-1 < v$.

Thus, if $-1 < v < +1$, then $\Omega = E_t(\mathbb{D})$ is simply connected and $E_t^* = E_{iv}^* \circ E_u$ maps \mathbb{D} onto $\text{Dome}(\Omega_v)$. The maps E_t and E_t^* are related by the fact that they have the same boundary values on $\partial\mathbb{D}$.

In short, given the crescent system and $-1 < v < +1$, $E_t^*(\mathbb{D}) = E_{iv}^* \circ E_u(\mathbb{D})$ is the pleated surface $\text{Dome}(\Omega_v)$ in \mathbb{H}^3 . The interpretation can be extended to $-1 < v < \infty$ by using the pleated surface determined from Λ_u with bending angles given by $\{\alpha\}$.

We note that complex scaling can be done equally well in any quasidisk, given a system of crescents there.

6.5 Complex earthquakes

The range of complex scaling is \mathbb{S}^2 . In contrast, starting with $(\Lambda, \mu) \subset \mathbb{H}^2$ and $t = u + iv \in \mathbb{C}$, we showed in [20] how to construct what we called a “quakebend”. A quakebend consists of an earthquake in \mathbb{D} determined by the measure $(\Lambda, u\mu)$ followed by bending according to $(\Lambda, v\mu)$. In current terminology a quakebend is better called a *complex earthquake* and denoted by $\mathcal{CE}_t : \mathbb{D} \rightarrow \mathbb{H}^3$.

We are ultimately interested in exploring the set

$$X = \{t \in \mathbb{C} : \text{the image } \mathcal{CE}_t(\mathbb{D}) \text{ is the dome over a quasidisk}\}.$$

The set X is a kind of 1-dimensional slice through $\mathfrak{T}(G)$. We will use the generalization of complex scaling to show how to build the floor under such a dome as a mapping from \mathbb{D} .

7 Complex scaling for general laminations

The goal of this section is to generalize complex scaling from discrete laminations to general measured laminations. For a general measured lamination in \mathbb{D} typically has no isolated leaves, and therefore no associated crescent systems. We will first focus on generalizing the angle scaling construction to general measured laminations on \mathbb{D} without group actions. After explaining this we will bring group actions back in.

Normalize the measured lamination by requiring that

$$\|\mu\| = \sup_{\tau} \mu(\tau) = 1,$$

where τ ranges over all transverse open geodesic segments of unit length.

We begin by constructing a map Φ_t that extends the restriction of \mathcal{CE}_t to $\mathbb{S}^1 = \partial\mathbb{D}$ to a map of \mathbb{S}^2 . The result is displayed below as Proposition 7.1 and presented without proof. The statement uses the notion “holomorphic motion”. For the convenience of the reader unfamiliar with this important tool, we have appended an explanation in §7.3.

Proposition 7.1 ([23], [25]). *There exists a simply connected, strip-like region $\mathcal{T} \supset \mathbb{R}$, symmetric about \mathbb{R} , such that for $t \in \mathcal{T}$, $t = u + iv$ the following hold.*

- (i) *The complex earthquake $\mathcal{CE}_t: \mathbb{D} \rightarrow \mathbb{H}^3$ is a quasiisometry.*
- (ii) *\mathcal{CE}_t extends to $\partial\mathbb{D}$; its restriction to $\partial\mathbb{D}$ is a holomorphic motion of $\mathbb{S}^1 = \partial\mathbb{D}$ in \mathbb{S}^2 .*
- (iii) *The restriction of \mathcal{CE}_t to \mathbb{S}^1 is itself the restriction to \mathbb{S}^1 of a holomorphic motion $\Phi_t: \mathbb{S}^2 \rightarrow \mathbb{S}^2$ which is determined by $(\Lambda, t\mu)$.*
- (iv) *The restriction of Φ_t to \mathbb{D} is a quasiconformal map onto its image Ω_t .*
- (v) *The restriction of Φ_t to \mathbb{S}^1 is injective (from (ii)), thus mapping \mathbb{S}^1 to a Jordan curve bounding Ω_t . The bending measure of $\text{Dome}(\Omega_t)$ is $vt \cdot \mu^*$ where μ^* is the transfer of μ to the image $\Lambda_{u\mu}$ of Λ under the earthquake map $\mathcal{CE}_{u\mu}$.*
- (vi) *If (Λ, μ) is invariant under the group G , the holomorphic motion Φ_t can be chosen so that there is a homomorphism $\varphi_t: G \rightarrow H_t$ with*

$$\Phi_t(g(z)) = \varphi_t(g) \circ \Phi_t(z) \quad \text{for all } g \in G \text{ and } z \in \mathbb{D}.$$

Next we fix a point $t_0 = iv_0 \in \mathcal{T}$ and set $\Omega_0 = \Phi_{t_0}(\mathbb{D})$. We will now use the quasidisk Ω_0 as our base for scaling. The image of the complex earthquake $\mathcal{CE}_{t_0}: \mathbb{D} \rightarrow \mathbb{H}^3$ is $\text{Dome}(\Omega_0)$.

To set the stage, assume first that $\text{Dome}(\Omega_0)$ has a discrete set of bending lines. The inverse image of the bending lines under the nearest point retraction r is a set of associated crescents in Ω_0 . Let E_t denote the scaling map of Ω_0 with respect to this crescent system and its vertex angles. Normalize so it is the identity in a prescribed gap. By Theorem 6.3, for each $t \in \mathbb{C}$, E_t extends to a continuous map of the closure $\bar{\Omega}_0 \rightarrow \mathbb{S}^2$. Set

$$\Psi_t = E_{i(t-t_0)/t_0}, \quad t \in \mathcal{T}, \quad \text{so } \Psi_{t_0} = \text{id}, \quad \Psi_0 = E_{-i}.$$

Then

$$\Psi_t \circ \Phi_{t_0}(\zeta) = \Phi_t(\zeta) \quad \zeta \in \partial\mathbb{D}, \quad t \in \mathcal{T}. \tag{7.1}$$

We will construct for all nonnegative measured laminations (Λ, μ) a mapping Ψ_t that generalizes scaling on Ω_0 and continues to satisfy Equation (7.1). But for Ψ_t , its parameter t will be allowed to range over \mathbb{C} .

After a long chain of reasoning, the bottom line is this:

Theorem 7.2 ([25]). *Let (Λ, μ) be a nonnegative measured lamination normalized by $\|\mu\| = 1$. Fix $t_0 \in \mathcal{T}$ and set $\Omega_0 = \Phi_{t_0}(\mathbb{D})$. The map $\Psi_t: \Omega_0 \rightarrow \mathbb{S}^2$, $t \in \mathbb{C}$, has the following properties:*

- (i) $\Psi_{t_0} = \text{id}$.
- (ii) For each $z \in \Omega_0$, $\Psi_t(z)$ depends holomorphically on $t \in \mathbb{C}$.
- (iii) For each $t \in \mathcal{T}$, $\Psi_t(z)$ can be continuously extended to $\partial\Omega_0$ so that Equation (7.1) holds. In particular, the restriction $\Psi_0: \partial\Omega_0 \rightarrow \partial\mathbb{D}$ is inverse to $\Phi_{t_0}: \partial\mathbb{D} \rightarrow \partial\Omega_0$.
- (iv) For $t \in \mathcal{T} \cap \text{UHP}$, $\Psi_t: \Omega_0 \rightarrow \mathbb{S}^2$ is injective, and $\Psi_t(\Omega_0) = \Phi_t(\mathbb{D})$.
- (v) When $t \in \text{UHP}$, $\Psi_t: \Omega_0 \rightarrow \mathbb{S}^2$ is a locally injective quasiregular mapping.

In short, we have constructed a holomorphic family of generalized scaling maps $\Psi_t(z): \mathbb{C} \times \Omega_0 \rightarrow \mathbb{S}^2$ which are related to the boundary values of complex earthquakes $\mathcal{CE}_t: \partial\mathbb{D} \rightarrow \partial\text{Dome}(\Psi_t(\Omega_0))$, for $t \in \mathcal{T}$.

7.1 Sullivan’s theorem

We will digress in order to describe the application of Theorem 7.2 to Sullivan’s amazing theorem on the metric relation of a simply connected region to its dome. This states that there is a universal K , such that any simply connected region $\Omega \subset \mathbb{C}$, $\Omega \neq \mathbb{C}$, has the following property. There is a K -quasiconformal map $\Xi: \Omega \rightarrow \text{Dome}(\Omega)$ that extends continuously to the identity map $\partial\Omega = \partial\text{Dome}(\Omega)$. Moreover if Ω is invariant under a Möbius group so is Ξ . For the history and a proof see [25].

It was initially conjectured by Thurston that $K = 2$, but we now know that $K > 2$, even if one does not require equivariance, [21].

The proof of Sullivan’s theorem given in [25] proceeds as follows. First consider locally injective quasiregular (developing) mappings $g: \mathbb{D} \rightarrow \mathbb{S}^2$. Quasiregular maps can be factored $g = h \circ f$ where f is a quasiconformal (homeomorphism) and h is locally injective (because g is) and holomorphic on the image of f . Via g , the complex structure on \mathbb{S}^2 can be pulled back to \mathbb{D} , getting a new complex structure C_g on \mathbb{D} and a quasiconformal mapping $\hat{g}: \mathbb{D} \rightarrow C_g$. Now bringing in a Riemann mapping $R: C_g \rightarrow \mathbb{D}$, we end up with a quasiconformal map $R \circ \hat{g}: \mathbb{D} \rightarrow \mathbb{D}$. We can normalize R so that $\pm 1, i \in \partial\mathbb{D}$ are fixed by the extension $qs(g)$ to $\partial\mathbb{D}$ of $R \circ \hat{g}$. The conclusion is that the correspondence $g \mapsto qs(g)$ identifies each locally injective quasiregular map to a point in the universal Teichmüller space \mathcal{UT} – recall that \mathcal{UT} can be taken as the space of normalized quasisymmetric maps of \mathbb{S}^1 .

We will apply our construction to two maps: (i) to the locally injective quasiregular map

$$g_t = \Psi_f \circ \Phi_{t_0}: \mathbb{D} \rightarrow \Omega_0 \rightarrow \mathbb{S}^2, \quad t \in \text{UHP}.$$

and (ii) to the quasiconformal map

$$f_t = \Phi_{\Lambda, t\mu}: \mathbb{D} \rightarrow \Omega_t \in \mathbb{S}^2, \quad t \in \mathcal{T}.$$

Consider then the two maps

$$F: t \in \mathcal{T} \mapsto qs(f_t) \in \mathcal{UT}, \quad G: t \in \text{UHP} \mapsto qs(g_t) \in \mathcal{UT}.$$

At the end of a sequence of arguments [25] one establishes that

Theorem 7.3.

- F is holomorphic in t ,
- G is holomorphic in t ,
- $F = G$ on their common domain $\mathcal{T} \cap \text{UHP}$.

In short, F has a holomorphic extension to the simply connected $\mathcal{T} \cup \text{UHP}$.

Theorem 7.3 yields a new and shorter proof of Sullivan's Theorem along with a greatly improved estimate for K . It was with the idea of getting an improved estimate for K that led us to enlarge the domain for t as much as possible. The estimate we found is that $K \leq 13.88$ as compared to $K \leq 82.8$ in [20]. These estimates are for the equivariant case: if Ω is invariant under a Möbius group, the map $\Xi: \Omega \rightarrow \text{Dome}(\Omega)$ must be as well. If equivariance is not required to hold, Chris Bishop [6] showed that the estimate could be improved even further to 7.8.

7.2 Earthquake disks

We finally turn to the matter of earthquake disks. Given a measured lamination $(\Lambda, \mu) \subset \mathbb{H}^2$, and $t \in \mathbb{C}$, consider the pleated surface P_t given by the complex earthquake \mathcal{CE}_t . The bending measure of P_t is $\text{Im}(t) \cdot \mu^*$, where as usual μ^* is the carryforth of μ under the $\text{Re}(t) \cdot \mu$ earthquake of Λ .

Set

$$X = \{t \in \mathbb{C} : \mathcal{CE}_t(\mathbb{D}) \text{ is the dome of a simply connected region } \subset \mathbb{C}\}.$$

Theorem 7.4 (The Disk Theorem [25]). *The set $X = X_{\Lambda, \mu}$ has the following properties:*

- X is symmetric about the real t -axis.
- $X \subset \mathbb{C}$ is closed.
- $X \cap \{\text{Im}(t) > 0\} = \{t \in \text{UHP} : \Psi_t: \Omega_0 \rightarrow \mathbb{S}^2 \text{ is injective}\}$.
- $\mathbb{C} \setminus X$ has no bounded components; if $X_0 \subset X$ is a component, $\mathbb{C} \setminus X_0$ has no bounded components either.
- Every component of the interior of X is simply connected.

Note that X is invariant under translation $t \mapsto t + 2\pi i$. There are numerous open questions about X , in particular, whether it contains isolated points and whether it is locally connected.

Suppose that (Λ, μ) is invariant under a Fuchsian group G with $R = \mathbb{D}/G$ a Riemann surface of finite area. If $t \in X$ then the floor Ω_t under the dome P_t is invariant under a quasi-Fuchsian group. Consequently there is a holomorphic map of the interior $D: \text{Int}(X) \rightarrow \mathcal{T}(R)$.

If R is a once punctured torus, McMullen proved [40] that X is biholomorphically equivalent to $\mathcal{T}(R)$. It was in this paper that the term *earthquake disk* was coined for a properly embedded image, in a Teichmüller space, of the unit disk under a complex earthquake map. In the paper, bending in 3D is replaced by its analogue “grafting” in 2D (see §6.4).

7.3 Appendix: holomorphic motion

Let $B \subset \mathbb{S}^2$ be an arbitrary set containing at least three points. Let $\{f_\lambda(z)\}$ denote a family of functions $B \rightarrow \mathbb{S}^2$ with parameter $\lambda \in \mathbb{D}$. The family $\{f_\lambda\}$ is called a *holomorphic motion* of B if it has the following three properties:

- For each fixed $\lambda \in \mathbb{D}$, the map $f_\lambda: z \in B \mapsto f_\lambda(z) \in \mathbb{S}^2$ is injective.
- For each fixed $z \in B$, the map $\lambda \mapsto f_\lambda(z)$ is a holomorphic map $\mathbb{D} \rightarrow \mathbb{S}^2$.
- For each $z \in B$, $f_0(z) = z$.

What makes this notion significant is the following λ -Lemma as originally formulated by Mañé–Sad–Sullivan with the final statement in Sladkowski [57], see also [19].

Theorem 7.5 (λ -Lemma). *Let G be a group of Möbius transformations which preserve a subset $B \subset \mathbb{S}^2$. Suppose $\{f_\lambda\}$ is a holomorphic motion of B . Suppose further that for each $\lambda \in \mathbb{D}$, f_λ is G -equivariant: f_λ induces an isomorphism to a group of Möbius transformations $G \rightarrow G_\lambda$ that preserve $f_\lambda(B)$. Then*

- (i) $f_\lambda(z)$ is jointly continuous in $\lambda \in \mathbb{D}$ and $z \in B$;
- (ii) for fixed $\lambda \in \mathbb{D}$, $f_\lambda(z)$ is the restriction to B of a K_λ -quasiconformal, G -equivariant mapping $f_\lambda^*: \mathbb{S}^2 \rightarrow \mathbb{S}^2$, where

$$K_\lambda = \frac{1 + |\lambda|}{1 - |\lambda|};$$

- (iii) $\{f_\lambda^*\}$ is a holomorphic motion of \mathbb{S}^2 , $\lambda \in \mathbb{D}$.

Note that continuity in z is *not* assumed, it is a conclusion. In particular if $B = \mathbb{D}$, then $f_\lambda(\mathbb{D})$ is a quasidisk.

In our application, $\lambda = t \in \mathcal{T}$, where \mathcal{T} is equivalent to the unit disk. We set $B = \partial\mathbb{D}$, for then $\mathcal{C}\mathcal{E}_t$ is injective on B for $t \in \mathcal{T}$, Proposition 7.1(ii).

References

- [1] I. Agol, Tameness of hyperbolic 3-manifolds. Preprint, 2004; arXiv:math.GT/0405568. [415](#)
- [2] L. V. Ahlfors, Finitely generated Kleinian groups. *Amer. J. Math.* 86 (1964) 413–429; correction *ibid* 87 (1965), 759. [413](#)
- [3] L. Bers, Simultaneous uniformization. *Bull. Amer. Math. Soc.* 66 (1960), 94–97. [424](#), [430](#)
- [4] L. Bers, Spaces of Kleinian groups. In *Several Complex Variables, I*, Lecture Notes in Math. 155, Springer-Verlag, Berlin 1970, 9–34.
- [5] L. Bers and L. Greenberg, Isomorphisms between Teichmüller spaces. In *Advances in the Theory of Riemann Surfaces*, Ann. of Math. Stud. 66, Princeton University Press, Princeton, N.J., 1971. [425](#)
- [6] C. J. Bishop, An explicit constant for Sullivan’s convex hull theorem. In *In the tradition of Ahlfors and Bers, III*, Contemp. Math. 355, Amer. Math. Soc., Providence, RI, 2004, 41–69. [441](#)
- [7] C. J. Bishop, A fast approximation to the Riemann map. Preprint. [420](#)
- [8] F. Bonahon, Shearing hyperbolic surfaces, bending pleated surfaces, and Thurston’s symplectic form. *Ann. Fac. Sci. Toulouse* 5 (1996), 233–297. [435](#)
- [9] F. Bonahon, Geodesic laminations on surfaces. In *Laminations and Foliations in Dynamics, Geometry, and Topology* (M. Lyubich et al., eds.), Contemp. Math. 269, Amer. Math. Soc., Providence, RI, 2001, 1–37. [435](#)
- [10] F. Bonahon and J.-P. Otal, Lamination mesurées de plissage des variétés hyperboliques de dimension 3. *Ann. of Math.* 160 (2004), 1013–1055. [427](#)
- [11] J. F. Brock, R. D. Canary, and Y. Minsky, Classification of Kleinian surface groups II: The ending lamination conjecture. Preprint, 2004; arXiv:math.GT/0412006. [414](#), [415](#), [429](#)
- [12] J. F. Brock, R. D. Canary, and Y. Minsky, The classification of finitely-generated Kleinian groups. In preparation. [414](#), [415](#), [429](#)
- [13] D. Calegari and D. Gabai, Shrinkwrapping and the taming of hyperbolic 3-manifolds. *J. Amer. Math. Soc.* 19 (2006), 385–446. [415](#)
- [14] R. D. Canary, A covering theorem for hyperbolic 3-manifolds and its applications. *Topology* 35 (1996), 751–778. [428](#)
- [15] R. D. Canary and D. McCullough, Homotopy Equivalence of 3-Manifolds and Deformation Theory of Finitely Generated Kleinian groups. *Mem. Amer. Math. Soc.* 172 (812), (2004). [426](#)
- [16] R. D. Canary, D. B. A Epstein, P. Green, Notes on notes of Thurston. In *Fundamentals of Hyperbolic Geometry: Selected Expositions* (R. D. Canary, D. B. A. Epstein, A. Marden, eds.), London Math. Soc. Lecture Note Ser. 328, Cambridge University Press, Cambridge 2006, 1–115. [435](#)
- [17] Y. Choi and C. Series, Lengths are coordinates for convex structures. *J. Differential Geom.* 73 (1) (2006), 75–117. [427](#)

- [18] A. Douady and C. J. Earle, Conformally natural extensions of homeomorphisms of the circle. *Acta Math.* 157 (1986), 23–48. [418](#)
- [19] C. J. Earle, I. Kra, and S. L. Krushkal, Holomorphic motions and Teichmüller spaces. *Trans. Amer. Math. Soc.* 343 (1994), 927–948. [442](#)
- [20] 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* (R. D. Canary, D. B. A. Epstein, A. Marden, eds.), London Math. Soc. Lecture Note Ser. 328, Cambridge University Press, Cambridge 2005, 117–266. [419](#), [433](#), [434](#), [438](#), [441](#)
- [21] D. B. A. Epstein and V. Markovic, The logarithmic spiral: A counterexample to the $K = 2$ conjecture. *Ann. of Math.* 161 (2005), 925–957. [440](#)
- [22] D. B. A. Epstein, A. Marden, and V. Markovic, Complex angle scaling. In *Kleinian Groups and Hyperbolic 3-Manifolds* (Y. Komari, V. Markovic, C. Series, eds.), London Math. Soc. Lecture Note Ser. 299, Cambridge University Press, Cambridge 2003, 343–362. [435](#), [437](#)
- [23] D. B. A. Epstein, A. Marden, and V. Markovic, Quasiconformal homeomorphisms and the convex hull boundary. *Ann. of Math.* 159 (2004), 305–336. [420](#), [439](#)
- [24] D. B. A. Epstein, A. Marden, and V. Markovic, Convex regions in the plane and their domes. *Proc. London Math. Soc.* (3) 92 (3) (2006), 624–654. [420](#)
- [25] D. B. A. Epstein, A. Marden, and V. Markovic, Complex earthquakes and deformations of the unit disk. Preprint, 2004. [439](#), [440](#), [441](#)
- [26] R. A. Evans, Weakly type-preserving sequences and strong convergence. *Geom. Dedicata* 108 (2004), 71–92. [428](#)
- [27] M. Feighn and D. McCullough, Finiteness conditions for 3-manifolds with boundary. *Amer. J. Math.* 109 (1987), 1155–1169. [414](#)
- [28] J. Gilman and P. Waterman, Classical two-parabolic T -Schottky groups. *J. Anal. Math.* 98 (2006), 1–42. [419](#)
- [29] M. Feighn and G. Mess, Conjugacy classes of finite subgroups in Kleinian groups. *Amer. J. Math.* 113 (1991), 179–188. [414](#)
- [30] D. Gallo, M. Kapovich, and A. Marden, The monodromy groups of Schwarzian equations on closed Riemann surfaces. *Ann. of Math.* 151 (2000), 625–704. [431](#)
- [31] J. Hempel *3-Manifolds*. Ann. of Math. Studies 86, Princeton University Press, Princeton, N.J., 1976. [416](#)
- [32] T. Jørgensen and P. Klein, Algebraic convergence of finitely generated Kleinian groups. *Quart. J. Math.* 33 (1982), 325–332. [426](#), [428](#)
- [33] T. Jørgensen and A. Marden, Algebraic and geometric convergence of Kleinian groups. *Math. Scand.* 66 (1990), 47–72. [428](#)
- [34] M. Kapovich, *Hyperbolic Manifolds and Discrete Groups*. Progr. Math. 183, Birkhäuser, Boston, MA, 2001. [412](#), [421](#), [427](#)
- [35] I. Kra, On spaces of Kleinian groups. *Comment. Math. Helv.* 47 (1972), 53–69. [424](#), [425](#)
- [36] D. McCullough, Compact submanifolds of 3-manifolds with boundary. *Quart. J. Math.* 37 (1986), 299–307. [415](#)

- [37] D. McCullough and A. Miller, Homeomorphisms of 3-manifolds with compressible boundary. *Mem. Amer. Math. Soc.* 61 (344) (1986).
- [38] C. McMullen, Cusps are dense. *Ann. of Math.* 13 (1991), 217–247. [431](#)
- [39] C. McMullen, *Renormalization and 3-Manifolds which Fiber over the Circle*. Ann. of Math. Stud. 142, Princeton University Press, Princeton, N.J., 1996. [418](#)
- [40] C. McMullen, Complex earthquakes and Teichmüller theory, *J. Amer. Math. Soc.* 11 (1998), 283–320. [427](#), [442](#)
- [41] A. Marden, On homotopic mappings of Riemann surfaces. *Ann. of Math.* 90 (1969), 1–8. [425](#)
- [42] A. Marden, The geometry of finitely generated kleinian groups, *Ann. of Math.* 99 (1974), 383–462. [421](#), [426](#)
- [43] A. Marden, Geometric relations between homeomorphic Riemann surfaces, *Bull. Amer. Math. Soc.* 3 (1980), 1007–1017. [431](#)
- [44] A. Marden, A proof of Ahlfors Finiteness Theorem. In *Spaces of Kleinian Groups* (Y. Minsky, M. Sakuma, C. Series, eds.), London Math. Soc. Lecture Note Ser. 329, Cambridge University Press, Cambridge 2005, 247–257.
- [45] A. Marden, *Outer Circles: An Introduction to Hyperbolic 3-Manifolds*. Cambridge University Press, Cambridge 2007, to appear. [412](#)
- [46] B. Maskit, Self-maps of Kleinian groups. *Amer. J. Math.* 93 (1971), 840–856. [423](#), [424](#), [425](#)
- [47] K. Matsuzaki and M. Taniguchi, *Hyperbolic Manifolds and Kleinian Groups*. Oxford Math. Monogr., Oxford University Press, New York 1998. [412](#)
- [48] Y. Minsky, Classification of Kleinian surface groups I: models and bounds. Preprint, 2004; arXiv:math.GT/0302208. [415](#), [429](#)
- [49] G. D. Mostow, *Strong Rigidity of Locally Symmetric Spaces*, Ann. of Math. Stud. 78, Princeton University Press, Princeton, N.J., 1973. [421](#)
- [50] J.R Munkres, Obstructions to smoothing piecewise-differentiable homeomorphisms. *Ann. of Math.* 72 (1960), 521–554. [416](#)
- [51] Ch. Pommerenke, Uniformly perfect sets and Fuchsian groups. *Analysis* 4 (1984), 299–321. [417](#)
- [52] J.-P. Otal, Thurston’s hyperbolization of Haken manifolds. In *Surveys of Differential Geometry*, Vol. III, International Press, Boston, MA, 1998, 77–194. [421](#)
- [53] 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. [421](#)
- [54] G. Prasad, Strong rigidity of \mathbb{Q} -rank 1 lattices. *Invent. Math.* 21 (1973), 255–286. [421](#)
- [55] G. P. Scott, Finitely generated 3-manifolds are finitely presented, *J. London Math. Soc.* 6 (1973), 437–440. [412](#), [414](#)
- [56] G. P. Scott, Compact submanifolds of 3-manifolds. *J. London Math. Soc.* 7 (1973), 246–250. [414](#)
- [57] Z. Slodkowski, Holomorphic motions and polynomial hulls. *Proc. Amer. Math. Soc.* 111 (1991), 347–355; *Abstracts Amer. Math. Soc.* 13 (1992), 259. [442](#)

- [58] T. Soma, Geometric limits of quasifuchsian groups. Preprint. [429](#)
- [59] D. Sullivan, On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions. In *Riemann Surfaces and Related Topics* (I. Kra, B. Maskit, eds.), Ann. of Math. Stud. 97, Princeton University Press, Princeton, N.J., 1981, 465–496.
- [60] D. Sullivan, Quasiconformal homeomorphisms and dynamics II: Structural stability implies hyperbolicity for Kleinian groups. *Acta Math.* 155 (1987), 243–260. [427](#)
- [61] G. A. Swarup, Two finiteness properties in 3-manifolds. *Bull. London Math. Soc.* 12 (1980) 296–302. [426](#)
- [62] W. P. Thurston, *The Geometry and Topology of Three-Manifolds*. Princeton University Lecture Notes, 1979; <http://msri.org/publications/books/gt3m/>. [412](#)
- [63] W. P. Thurston, Three dimensional manifolds, Kleinian groups and hyperbolic geometry. *Bull. Amer. Math. Soc.* 6 (1982), 357–381. [421](#)
- [64] W. P. Thurston, Hyperbolic structures on 3-manifolds I: Deformation of acylindrical manifolds. *Ann. of Math.* 124 (1986), 203–246. [427](#)
- [65] W. P. Thurston, Earthquakes in two-dimensional hyperbolic geometry. In *Fundamentals of Hyperbolic Geometry: Selected Expositions*, (R. D. Canary, D. B. A. Epstein, A. Marden, eds.), London Math. Soc. Lecture Note Ser. 328, Cambridge University Press, Cambridge 2005, 267–289. [432](#), [433](#)
- [66] P. Tukia, Quasiconformal extension of quasisymmetric mappings compatible with a Möbius group. *Acta Math.* 154 (1985), 153–193. [418](#)
- [67] P. Tukia, personal communication. [418](#)

Geometry of the complex of curves and of Teichmüller space

Ursula Hamenstädt

Contents

1	Introduction	447
2	Train tracks and geodesic laminations	449
3	Hyperbolicity of the complex of curves	454
4	Geometry of Teichmüller space	462
	References	465

1 Introduction

Consider a compact oriented surface S of genus $g \geq 0$ from which $m \geq 0$ points, so-called *punctures*, have been deleted. We assume that S is non-exceptional, i.e. that $3g - 3 + m \geq 2$; this rules out a sphere with at most four punctures and a torus with at most one puncture.

In [11], Harvey associates to such a surface the following simplicial complex.

Definition 1.1. The *complex of curves* $\mathcal{C}(S)$ for the surface S is the simplicial complex whose vertices are the free homotopy classes of essential simple closed curves on S and whose simplices are spanned by collections of such curves which can be realized disjointly.

Here we mean by an *essential* simple closed curve a simple closed curve which is not contractible nor homotopic into a puncture. Since $3g - 3 + m$ is the number of curves in a *pants decomposition* of S , i.e. a maximal collection of disjoint mutually not freely homotopic essential simple closed curves which decompose S into $2g - 2 + m$ open subsurfaces homeomorphic to a thrice punctured sphere, the dimension of $\mathcal{C}(S)$ equals $3g - 4 + m$.

In the sequel we restrict our attention to the one-skeleton of the complex of curves which is usually called the *curve graph*; by abuse of notation, we denote it again by $\mathcal{C}(S)$. Since $3g - 3 + m \geq 2$ by assumption, $\mathcal{C}(S)$ is a nontrivial graph which moreover is connected [11]. However, this graph is locally infinite. Namely, for every simple closed curve α on S the surface $S - \alpha$ which we obtain by cutting S open along

α contains at least one connected component of Euler characteristic at most -2 , and such a component contains infinitely many distinct free homotopy classes of simple closed curves which viewed as curves in S are disjoint from α .

Providing each edge in $\mathcal{C}(S)$ with the standard euclidean metric of diameter 1 equips the curve graph with the structure of a geodesic metric space. Since $\mathcal{C}(S)$ is not locally finite, this metric space $(\mathcal{C}(S), d)$ is not locally compact. Masur and Minsky [19] showed that nevertheless its geometry can be understood quite explicitly. Namely, $\mathcal{C}(S)$ is hyperbolic of infinite diameter. Here for some $\delta > 0$ a geodesic metric space is called δ -hyperbolic in the sense of Gromov if it satisfies the δ -thin triangle condition: For every geodesic triangle with sides a, b, c the side c is contained in the δ -neighborhood of $a \cup b$. Later Bowditch [2] gave a simplified proof of the result of Masur and Minsky which can also be used to compute explicit bounds for the hyperbolicity constant δ .

Since the Euler characteristic of S is negative, the surface S admits a complete hyperbolic metric of finite volume. The group of diffeomorphisms of S which are isotopic to the identity acts on the space of such metrics. The quotient space under this action is the Teichmüller space $\mathcal{T}_{g,m}$ for S of all marked isometry classes of complete hyperbolic metrics on S of finite volume, or, equivalently, the space of all marked complex structures on S of finite type. The Teichmüller space can be equipped with a natural topology, and with this topology it is homeomorphic to $\mathbb{R}^{6g-6+2m}$. The mapping class group $\mathcal{M}_{g,m}$ of all isotopy classes of orientation preserving diffeomorphisms of S acts properly discontinuously as a group of diffeomorphisms of Teichmüller space preserving a complete Finsler metric, the so-called Teichmüller metric. The quotient orbifold is the moduli space $\text{Mod}(S)$ of S of all isometry classes of complete hyperbolic metrics of finite volume on S (for all this see [12]).

The significance of the curve graph for the geometry of Teichmüller space comes from the obvious fact that the mapping class group acts on $\mathcal{C}(S)$ as a group of simplicial isometries. Even more is true: If S is not a twice punctured torus or a closed surface of genus 2, then the extended mapping class group of isotopy classes of all diffeomorphisms of S coincides precisely with the group of simplicial isometries of $\mathcal{C}(S)$; for a closed surface of genus 2, the group of simplicial isometries of $\mathcal{C}(S)$ is the quotient of the extended mapping class group under the hyperelliptic involution which acts trivially on $\mathcal{C}(S)$ (see [13] for an overview on this and related results). Moreover, there is a natural map $\Psi: \mathcal{T}_{g,m} \rightarrow \mathcal{C}(S)$ which is coarsely $\mathcal{M}_{g,m}$ -equivariant and coarsely Lipschitz with respect to the Teichmüller metric on $\mathcal{T}_{g,m}$. By this we mean that there is a number $a > 1$ such that $d(\Psi(\phi h), \phi(\Psi h)) \leq a$ for all $h \in \mathcal{T}_{g,m}$ and all $\phi \in \mathcal{M}_{g,m}$ and that moreover $d(\Psi h, \Psi h') \leq a d_{\mathcal{T}}(h, h') + a$ for all $h, h' \in \mathcal{T}_{g,m}$ where $d_{\mathcal{T}}$ denotes the distance function on $\mathcal{T}_{g,m}$ induced by the Teichmüller metric (see Section 4).

As a consequence, the geometry of $\mathcal{C}(S)$ is related to the large-scale geometry of the Teichmüller space and the mapping class group. We discuss this relation in Section 4. In Section 3 we give a proof of the hyperbolicity of the curve graph using train tracks and splitting sequences of train tracks in a consistent way as the main tool.

Section 2 introduces train tracks, geodesic laminations and quadratic differentials and summarizes some of their properties.

This work was partially supported by SFB 611.

2 Train tracks and geodesic laminations

Let S be a nonexceptional surface of finite type and choose a complete hyperbolic metric on S of finite volume. With respect to this metric, every essential free homotopy class of loops can be represented by a closed geodesic which is unique up to parametrization. This geodesic is *simple*, i.e. without self-intersection, if and only if the free homotopy class has a simple representative (see [4]). In other words, there is a one-to-one correspondence between vertices of the curve graph and simple closed geodesics on S . Moreover, there is a fixed *compact* subset S_0 of S containing all simple closed geodesics.

The *Hausdorff distance* between two closed bounded subsets A, B of a metric space X is defined to be the infimum of all numbers $\varepsilon > 0$ such that A is contained in the ε -neighborhood of B and B is contained in the ε -neighborhood of A . This defines indeed a distance and hence a topology on the space of closed bounded subsets of X ; this topology is called the *Hausdorff topology*. If X is compact then the space of closed subsets of X is compact as well. In particular, for the distance on S induced by a complete hyperbolic metric of finite volume, the space of closed subsets of the compact set $S_0 \subset S$ is compact with respect to the Hausdorff topology.

The collection of all simple closed geodesics on S is *not* a closed set with respect to the Hausdorff topology, but a point in its closure can be described as follows.

Definition 2.1. A *geodesic lamination* for a complete hyperbolic structure of finite volume on S is a compact subset of S which is foliated into simple geodesics.

Thus every simple closed geodesic is a geodesic lamination which consists of a single leaf. The space of geodesic laminations on S equipped with the Hausdorff topology is compact, and it contains the closure of the set of simple closed geodesics as a *proper* subset. Note that every lamination in this closure is necessarily connected.

To describe the structure of the space of geodesic laminations more explicitly we introduce some more terminology.

Definition 2.2. A geodesic lamination λ is called *minimal* if each of its half-leaves is dense in λ . A geodesic lamination λ is *maximal* if all its complementary components are ideal triangles or once punctured monogons. A geodesic lamination is called *complete* if it is maximal and can be approximated in the Hausdorff topology for compact subsets of S by simple closed geodesics.

As an example, a simple closed geodesic is a minimal geodesic lamination. A minimal geodesic lamination with more than one leaf has uncountably many leaves.

Every minimal geodesic lamination can be approximated in the Hausdorff topology by simple closed geodesics [5]. Moreover, a minimal geodesic lamination λ is a *sublamination* of a complete geodesic lamination [9], i.e. there is a complete geodesic lamination which contains λ as a closed subset. In particular, every simple closed geodesic on S is a sublamination of a complete geodesic lamination. Every geodesic lamination λ is a disjoint union of finitely many minimal components and a finite number of isolated leaves. Each of the isolated leaves of λ either is an isolated closed geodesic and hence a minimal component, or it *spirals* about one or two minimal components [1], [5], [24]. This means that the set of accumulation points of an isolated half-leaf of λ is a minimal component of λ .

Geodesic laminations which are disjoint unions of minimal components can be equipped with the following additional structure.

Definition 2.3. A *measured geodesic lamination* is a geodesic lamination together with a translation invariant *transverse measure*.

A transverse measure for a geodesic lamination λ assigns to every smooth compact arc c on S with endpoints in the complement of λ and which intersects λ transversely a finite Borel measure on c supported in $c \cap \lambda$. These measures transform in the natural way under homotopies of c by smooth arcs transverse to λ which move the endpoints of the arc c within fixed complementary components of λ . The support of the measure is the smallest sublamination ν of λ such that the measure on any such arc c which does not intersect ν is trivial. This support is necessarily a union of minimal components of λ . An example for a measured geodesic lamination is a *weighted simple closed geodesic* which consists of a simple closed geodesic α and a positive weight $a > 0$. The measure disposed on a transverse arc c is then the sum of the Dirac masses on the intersection points between c and α multiplied with the weight a .

The space \mathcal{ML} of measured geodesic laminations on S can naturally be equipped with the *weak*-topology*. This topology restricts to the weak*-topology on the space of measures on a given arc c which is transverse to each lamination from an open subset of lamination space. The natural action of the group $(0, \infty)$ by scaling is continuous with respect to this topology, and the quotient is the space \mathcal{PML} of *projective measured geodesic laminations*. This space is homeomorphic to a sphere of dimension $6g - 7 + 2m$ (see [5], [6], [25]).

The *intersection number* $i(\gamma, \delta)$ between two simple closed curves $\gamma, \delta \in \mathcal{C}(S)$ equals the minimal number of intersection points between representatives of the free homotopy classes of γ, δ . This intersection function extends to a continuous pairing $i: \mathcal{ML} \times \mathcal{ML} \rightarrow [0, \infty)$, called the *intersection form*.

Measured geodesic laminations are intimately related to more classical objects associated to Riemann surfaces, namely *holomorphic quadratic differentials*. A holomorphic quadratic differential q on a Riemann surface S assigns to each complex coordinate z an expression of the form $q(z)dz^2$ where $q(z)$ is a holomorphic function on the domain of the coordinate system, and $q(z)(dz/dw)^2 = q(w)$ for overlapping coordinates z, w . We require that q has at most a simple pole at each puncture of S . If q

does not vanish identically, then its zeros are isolated and independent of the choice of a complex coordinate. If $p \in S$ is not a zero for q then there is a coordinate z near p , unique up to multiplication with ± 1 , such that p corresponds to the origin and that $q(z) \equiv 1$. Writing $z = x + iy$ for this coordinate, the euclidean metric $dx^2 + dy^2$ is uniquely determined by q . The arcs parallel to the x -axis (or y -axis, respectively) define a foliation \mathcal{F}_h (or \mathcal{F}_v) on the set of regular points of q called the *horizontal* (or *vertical*) foliation. The vertical length $|dy|$ defines a *transverse measure* for the horizontal foliation, and the horizontal length $|dx|$ defines a transverse measure for the vertical foliation. The foliations $\mathcal{F}_h, \mathcal{F}_v$ have singularities of the same type at the zeros of q and at the punctures of S (see [27] for more on quadratic differentials and measured foliations).

There is a one-to-one correspondence between measured geodesic laminations and (equivalence classes of) measured foliations on S (see [16] for a precise statement). The pair of measured foliations defined by a quadratic differential q corresponds under this identification to a pair of measured geodesic laminations $\lambda \neq \mu \in \mathcal{ML}$ which *jointly fill up* S . This means that for every $\eta \in \mathcal{ML}$ we have $i(\lambda, \eta) + i(\mu, \eta) > 0$.

Vice versa, every pair of measured geodesic laminations $\lambda \neq \mu \in \mathcal{ML}$ which jointly fill up S defines a unique complex structure of finite type on S together with a holomorphic quadratic differential $q(\lambda, \mu)$ (see [14] and the references given there) whose *area*, i.e. the area of the singular euclidean metric defined by $q(\lambda, \mu)$, equals $i(\lambda, \mu)$. If α, β are *simple multi-curves* on S , which means that α and β consist of collections $c = c_1 \cup \dots \cup c_\ell \subset \mathcal{C}(S)$ of free homotopy classes of simple closed curves which can be realized disjointly, and if α, β jointly fill up S , then for all $a > 0, b > 0$ the quadratic differential $q(a\alpha, b\beta)$ defined by the measured geodesic laminations $a\alpha, b\beta$ can explicitly be constructed as follows. Choose smooth representatives of α, β , again denoted by α, β , which intersect transversely in precisely $i(\alpha, \beta)$ points; for example, the geodesic representatives of α, β with respect to any complete hyperbolic metric on S of finite volume have this property. For each intersection point between α, β choose a closed rectangle in S with piecewise smooth boundary containing this point in its interior and which does not contain any other intersection point between α, β . We allow that some of the vertices of such a rectangle are punctures of S . These rectangles can be chosen in such a way that they provide S with the structure of a cubical complex: The boundary of each component D of $S - \alpha - \beta$ is a polygon with an even number of sides which are subarcs of α, β in alternating order. If D does not contain a puncture, then its boundary has at least four sides. Thus we can construct the rectangles in such a way that their union is all of S and that the intersection between any two distinct such rectangles either is a common side or a common vertex. Each rectangle from this cubical complex has two sides which are “parallel” to α and two sides “parallel” to β (see Section 4 of [14] for a detailed discussion of this construction).

Equip each rectangle with an euclidean metric such that the sides parallel to α are of length b , the sides parallel to β are of length a and such that the metrics on two rectangles coincide on a common boundary arc. These metrics define a piecewise euclidean metric on S with a singularity of cone angle $k\pi \geq 3\pi$ in the interior of each

disc component of $S - \alpha - \beta$ whose boundary consists of $2k \geq 6$ sides. The metric also has a singularity of cone angle π at each puncture of S which is contained in a punctured disc component with two sides. Since there are precisely $i(\alpha, \beta)$ rectangles, the area of this singular euclidean metric on S equals $abi(\alpha, \beta)$. The line segments of this singular euclidean metric which are parallel to α and β define singular foliations $\mathcal{F}_\alpha, \mathcal{F}_\beta$ on S with transverse measures induced by the singular metric. The metric defines a complex structure on S and a quadratic differential $q(a\alpha, b\beta)$ which is holomorphic for this structure and whose horizontal and vertical foliations are just $\mathcal{F}_\alpha, \mathcal{F}_\beta$, with transverse measures determined by the weights a and b . The assignment which associates to $a > 0$ the Riemann surface structure determined by $q(a\alpha, \beta)$ is up to parametrization the geodesic in the Teichmüller space with respect to the *Teichmüller metric* whose cotangent bundle contains the differentials $q(a\alpha, \beta)$ (compare [12], [14]).

A *geodesic segment* for a quadratic differential q is a path in S not containing any singularities in its interior and which is a geodesic in the local euclidean structure defined by q . A closed geodesic is composed of a finite number of such geodesic segments which meet at singular points of q and make an angle at least π on either side. Every essential closed curve c on S is freely homotopic to a closed geodesic with respect to q , and the length of such a geodesic η is the infimum of the q -lengths of any curve freely homotopic to η (compare [26] for a detailed discussion of the technical difficulties caused by the punctures of S) and will be called the q -length of our closed curve c . If $q = q(\lambda, \nu)$ for $\lambda, \nu \in \mathcal{ML}$ then this q -length is bounded from above by $2i(\lambda, c) + 2i(\nu, c)$ (see [26]).

Thurston invented a way to understand the structure of the space of geodesic laminations by squeezing almost parallel strands of such a lamination to a simple arc and analyzing the resulting graph. The structure of such a graph is as follows.

Definition 2.4. A *train track* on the surface S is an embedded 1-complex $\tau \subset S$ whose edges (called *branches*) are smooth arcs with well-defined tangent vectors at the endpoints. At any vertex (called a *switch*) the incident edges are mutually tangent. Through each switch there is a path of class C^1 which is embedded in τ and contains the switch in its interior. In particular, the branches which are incident on a fixed switch are divided into “incoming” and “outgoing” branches according to their inward pointing tangent at the switch. Each closed curve component of τ has a unique bivalent switch, and all other switches are at least trivalent. The complementary regions of the train track have negative Euler characteristic, which means that they are different from discs with 0, 1 or 2 cusps at the boundary and different from annuli and once-punctured discs with no cusps at the boundary. A train track is called *generic* if all switches are at most trivalent.

In the sequel we only consider generic train tracks. For such a train track τ , every complementary component is a bordered subsurface of S whose boundary consists of a finite number of arcs of class C^1 which come together at a finite number of *cusps*. Moreover, for every switch of τ there is precisely one complementary component

containing the switch in its closure which has a cusp at the switch. A detailed account on train tracks can be found in [25] and [23].

A geodesic lamination or a train track λ is *carried* by a train track τ if there is a map $F: S \rightarrow S$ of class C^1 which is isotopic to the identity and maps λ to τ in such a way that the restriction of its differential dF to every tangent line of λ is non-singular. Note that this makes sense since a train track has a tangent line everywhere.

If c is a simple closed curve carried by τ with carrying map $F: c \rightarrow \tau$ then c defines a *counting measure* μ_c on τ . This counting measure is the non-negative weight function on the branches of τ which associates to an open branch b of τ the number of connected components of $F^{-1}(b)$. A counting measure is an example for a *transverse measure* on τ which is defined to be a nonnegative weight function μ on the branches of τ satisfying the *switch condition*: for every switch s of τ , the sum of the weights over all incoming branches at s is required to coincide with the sum of the weights over all outgoing branches at s . The set $V(\tau)$ of all transverse measures on τ is a closed convex cone in a linear space and hence topologically it is a closed cell. More generally, every measured geodesic lamination λ on S which is carried by τ via a carrying map $F: \lambda \rightarrow \tau$ defines a transverse measure on τ by assigning to a branch b the total mass of the pre-image of b under F ; the resulting weight function is independent of the particular choice of F . Moreover, every transverse measure for τ can be obtained in this way (see [25]).

Definition 2.5. A train track is called *recurrent* if it admits a transverse measure which is positive on every branch. A train track τ is called *transversely recurrent* if every branch b of τ is intersected by an embedded simple closed curve $c = c(b) \subset S$ which intersects τ transversely and is such that $S - \tau - c$ does not contain an embedded *bigon*, i.e. a disc with two corners at the boundary. A recurrent and transversely recurrent train track is called *birecurrent*. A generic transversely recurrent train track which carries a complete geodesic lamination is called *complete*.

For every recurrent train track τ , measures which are positive on every branch define the interior of the convex cone $V(\tau)$ of all transverse measures. A complete train track is birecurrent [9].

A half-branch \tilde{b} in a generic train track τ incident on a switch v is called *large* if the switch v is trivalent and if every arc $\rho: (-\varepsilon, \varepsilon) \rightarrow \tau$ of class C^1 which passes through v meets the interior of \tilde{b} . A branch b in τ is called *large* if each of its two half-branches is large; in this case b is necessarily incident on two distinct switches (for all this, see [25]).

There is a simple way to modify a transversely recurrent train track τ to another transversely recurrent train track. Namely, if e is a large branch of τ then we can perform a right or left *split* of τ at e as shown in Figure A below. The split τ' of a train track τ is carried by τ . If τ is complete and if the complete geodesic lamination λ is carried by τ , then for every large branch e of τ there is a unique choice of a right or left split of τ at e with the property that the split track τ' carries λ , and τ' is complete. In particular, a complete train track τ can always be split at any large branch e to a

complete train track τ' ; however there may be a choice of a right or left split at e such that the resulting train track is not complete any more (compare p. 120 in [25]).

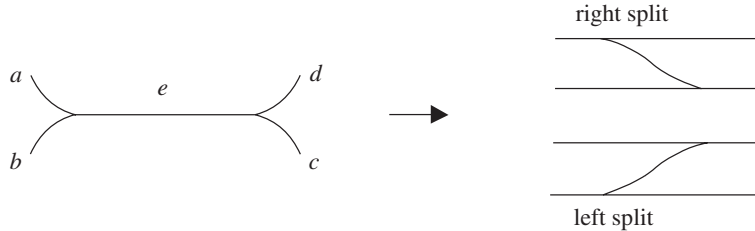


Figure A

In the sequel we denote by $\mathcal{T}T$ the collection of all isotopy classes of complete train tracks on S . A sequence $(\tau_i) \subset \mathcal{T}T$ of complete train tracks is called a *splitting sequence* if τ_{i+1} can be obtained from τ_i by a single split at some large branch e .

3 Hyperbolicity of the complex of curves

In this section we present a proof of hyperbolicity of the curve graph using the main strategy of Masur and Minsky [19] and Bowditch [2] in a modified form. The first step consists in guessing a family of uniform *quasi-geodesics* in the curve graph connecting any two points. Here a p -quasi-geodesic for some $p > 1$ is a curve $c: [a, b] \rightarrow \mathcal{C}(S)$ which satisfies

$$d(c(s), c(t))/p - p \leq |s - t| \leq pd(c(s), c(t)) + p \quad \text{for all } s, t \in [a, b].$$

Note that a quasi-geodesic does not have to be continuous. In a hyperbolic geodesic metric space, every p -quasi-geodesic is contained in a fixed tubular neighborhood of any geodesic joining the same endpoints, so the δ -thin triangle condition also holds for triangles whose sides are uniform quasi-geodesics [3]. As a consequence, for every triangle in a hyperbolic geodesic metric space with uniform quasi-geodesic sides there is a “midpoint” whose distance to each side of the triangle is bounded from above by a universal constant. The second step of the proof consists in finding such a midpoint for triangles whose sides are curves of the distinguished curve family. This is then used in a third step to establish the δ -thin triangle condition for the distinguished family of curves and derive from this hyperbolicity of $\mathcal{C}(S)$. By abuse of notation, in the sequel we simply write $\alpha \in \mathcal{C}(S)$ if α is a free homotopy class of an essential simple closed curve on S , i.e. if α is a vertex of $\mathcal{C}(S)$.

We begin with defining a map from the set $\mathcal{T}T$ of complete train tracks on S into $\mathcal{C}(S)$. For this we call a transverse measure μ for a complete train track τ a *vertex cycle* [19] if μ spans an extreme ray in the convex cone $V(\tau)$ of all transverse measures on τ .

Up to scaling, every vertex cycle μ is a counting measure of a simple closed curve c which is carried by τ [19]. Namely, the switch conditions are a family of linear equations with integer coefficients for the transverse measures on τ . Thus an extreme ray is spanned by a nonnegative *rational* solution which can be scaled to a nonnegative integral solution. From every integral transverse measure μ for τ we can construct a unique *simple weighted multi-curve*, i.e. a simple multi-curve together with a family of weights for each of its components, which is carried by τ and whose counting measure coincides with μ as follows. For each branch b of τ draw $\mu(b)$ disjoint arcs parallel to b . By the switch condition, the endpoints of these arcs can be connected near the switches in a unique way so that the resulting family of arcs does not have self-intersections. Let c be the simple multi-curve consisting of the free homotopy classes of the connected components of the resulting curve \tilde{c} . To each such homotopy class associate the number of components of \tilde{c} in this class as a weight. The resulting simple weighted multi-curve is carried by τ , and its counting measure equals μ . Thus if there are at least two components of \tilde{c} which are not freely homotopic then the weighted counting measures of these components determine a decomposition of μ into transverse measures for τ which are not multiples of μ . This is impossible if μ is a vertex cycle. Hence c consists of a single component and up to scaling, μ is the counting measure of a simple closed curve on S .

A simple closed curve which is carried by τ , with carrying map $F: c \rightarrow \tau$, defines a vertex cycle for τ only if $F(c)$ passes through every branch of τ at most twice, with different orientation (Lemma 2.2 of [7]). In particular, the counting measure μ_c of a simple closed curve c which defines a vertex cycle for τ satisfies $\mu_c(b) \leq 2$ for every branch b of τ .

In the sequel we mean by a vertex cycle of a complete train track τ an *integral* transverse measure on τ which is the counting measure of a simple closed curve c on S carried by τ and which spans an extreme ray of $V(\tau)$; we also use the notion vertex cycle for the simple closed curve c . Since the number of branches of a complete train track on S only depends on the topological type of S , the number of vertex cycles for a complete train track on S is bounded by a universal constant (see [19] and [7]).

The following observation of Penner and Harer [25] is essential for all what follows. Denote by $\mathcal{MC}(S)$ the space of all simple multi-curves on S . Let $P = \bigcup_{i=1}^{3g-3+m} \gamma_i \in \mathcal{MC}(S)$ be a pants decomposition for S , i.e. a simple multi-curve with the maximal number of components. Then there is a special family of complete train tracks with the property that each pants curve γ_i admits a closed neighborhood A diffeomorphic to an annulus and such that $\tau \cap A$ is diffeomorphic to a *standard twist connector* depicted in Figure B. Such a train track clearly carries each pants curve from the pants decomposition P as a vertex cycle; we call it *adapted* to P . For every complete geodesic lamination λ there is a train track τ adapted to P which carries λ ([25], see also [9], [7]).

Since every simple multi-curve is a subset of a pants decomposition of S , we can conclude.

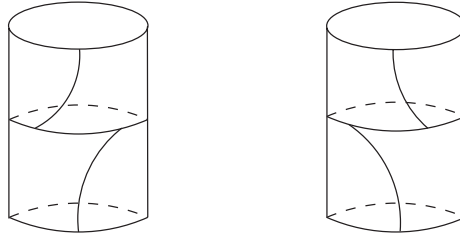


Figure B

Lemma 3.1 ([10]). *For every pair $(\alpha, \beta) \in \mathcal{MC}(S) \times \mathcal{MC}(S)$ there is a splitting sequence $(\tau_i)_{0 \leq i \leq m} \subset \mathcal{TT}$ of complete train tracks with the property that τ_0 is adapted to a pants decomposition $P_\alpha \supset \alpha$ and that each component of β is a vertex cycle for τ_m .*

We call a splitting sequence as in the lemma an $\alpha \rightarrow \beta$ -splitting sequence. Note that such a sequence is by no means unique.

The distance in $\mathcal{C}(S)$ between two simple closed curves α, β is bounded from above by $i(\alpha, \beta) + 1$ (Lemma 1.1 of [2] and Lemma 2.1 of [19]). In particular, there is a number $D_0 > 0$ with the following property. Let $\tau, \tau' \in \mathcal{TT}$ and assume that τ' is obtained from τ by at most one split. Then the distance in $\mathcal{C}(S)$ between any vertex cycle of τ and any vertex cycle of τ' is at most D_0 (see [19] and the discussion following Corollary 2.3 in [7]).

Define a map $\Phi: \mathcal{TT} \rightarrow \mathcal{C}(S)$ by assigning to a train track $\tau \in \mathcal{TT}$ a vertex cycle $\Phi(\tau)$ for τ . By our above discussion, for any two choices Φ, Φ' of such a map we have $d(\Phi(\tau), \Phi'(\tau)) \leq D_0$ for all $\tau \in \mathcal{TT}$. Images under the map Φ of splitting sequences then define a family of curves in $\mathcal{C}(S)$ which connect any pair of points in a D_0 -dense subset of $\mathcal{C}(S) \times \mathcal{C}(S)$, equipped with the product metric. As a consequence, we can use such images of splitting sequences as our guesses for uniform quasi-geodesics. It turns out that up to parametrization, these curves are indeed p -quasi-geodesics in $\mathcal{C}(S)$ for a universal number $p > 0$ only depending on the topological type of the surface S ([20], see also [7]).

To explain this fact we use the following construction of Bowditch [2]. For multicurves $\alpha, \beta \in \mathcal{MC}(S)$ which jointly fill up S , i.e. which cut S into components which are homeomorphic to discs and once punctured discs, and for a number $a > 0$ let $q(a\alpha, \beta/ai(\alpha, \beta))$ be the area one quadratic differential whose horizontal foliation corresponds to the measured geodesic lamination $a\alpha$ and whose vertical measured foliation corresponds to the measured geodesic lamination $\beta/ai(\alpha, \beta)$. For $r > 0$ define

$$L_a(\alpha, \beta, r) = \{\gamma \in \mathcal{C}(S) \mid \max\{ai(\gamma, \alpha), i(\gamma, \beta)/ai(\alpha, \beta)\} \leq r\}.$$

Then $L_a(\alpha, \beta, r)$ is contained in the set of all simple closed curves on S whose $q(a\alpha, \beta/ai(\alpha, \beta))$ -length does not exceed $2r$. Note that we have $L_a(\alpha, \beta, r) =$

$L_{1/ai(\alpha,\beta)}(\beta, \alpha, r)$ for all $r > 0$, moreover $\alpha' \in L_a(\alpha, \beta, r)$ for every component α' of α and every sufficiently large $a > 0$, and $\beta' \in L_a(\alpha, \beta, r)$ for every component β' of β and every sufficiently small $a > 0$. Thus for fixed $r > 0$ we can think of a suitably chosen assignment which associates to a number $s > 0$ a point in $L_s(\alpha, \beta, r)$ as a curve in $\mathcal{C}(S)$ connecting a component of β to a component of α (provided, of course, that the sets $L_s(\alpha, \beta, r)$ are non-empty). Lemma 2.5 of [7] links such curves to splitting sequences.

Lemma 3.2 ([7]). *There is a number $k_0 \geq 1$ with the following property. Let P be a pants decomposition of S , let $\alpha \in \mathcal{MC}(S)$ be such that α and P jointly fill up S and let $(\tau_i)_{0 \leq i \leq m} \subset \mathcal{TT}$ be a $P \rightarrow \alpha$ -splitting sequence. Then there is a non-decreasing surjective function $\kappa: (0, \infty) \rightarrow \{0, \dots, m\}$ such that $\kappa(s) = 0$ for all sufficiently small $s > 0$, $\kappa(s) = m$ for all sufficiently large $s > 0$ and that for all $s \in (0, \infty)$ there is a vertex cycle of $\tau_{\kappa(s)}$ which is contained in $L_s(\alpha, P, k_0)$.*

Since for every multi-curve $\alpha \in \mathcal{MC}(S)$ and every pants decomposition P of S there is a $P \rightarrow \alpha$ -splitting sequence, we conclude that for every $k \geq k_0$ and every $s > 0$ the set $L_s(\alpha, P, k)$ is non-empty. To obtain a control of the size of these sets, Bowditch [2] uses the following observation (Lemma 4.1 in [2]) whose first part was earlier shown by Masur and Minsky (Lemma 5.1 of [19]).

Lemma 3.3 ([2]). *There is a number $k_1 \geq k_0$ with the following property. For all $\alpha, \beta \in \mathcal{MC}(S)$ which jointly fill up S and every $a \in (0, \infty)$ there is a $\delta \in L_a(\alpha, \beta, k_1)$ such that for every $\gamma \in \mathcal{MC}(S)$ we have $i(\delta, \gamma) \leq k_1 \max\{ai(\alpha, \gamma), i(\gamma, \beta)/ai(\alpha, \beta)\}$. In particular, for every $R > 0$, for all $\alpha, \beta \in \mathcal{MC}(S)$ and for every $a > 0$ the diameter of the set $L_a(\alpha, \beta, R)$ is not bigger than $2k_1R + 1$.*

Proof. In [19], [2] it is shown that there is a number $\nu > 0$ only depending on the topological type of S and there is an embedded essential annulus in S whose width with respect to the piecewise euclidean metric defined by the quadratic differential $q(a\alpha, \beta/ai(\alpha, \beta))$ is at least ν . This means that the distance between the boundary circles of the annulus is at least ν . Assuming the existence of such an annulus, let δ be its core-curve. Then for every simple closed curve γ on S and for every essential intersection of γ with δ there is a subarc of γ which crosses through this annulus and hence whose length is at least ν ; moreover, different subarcs of γ corresponding to different essential intersections between γ and δ are disjoint. Thus the length with respect to the singular euclidean metric on S of any simple closed curve γ on S is at least $\nu i(\gamma, \delta)$. On the other hand, by construction the minimal length with respect to this metric of a curve in the free homotopy class of γ is bounded from above by $2 \max\{ai(\alpha, \gamma), i(\beta, \gamma)/ai(\alpha, \beta)\}$ and therefore the core curve δ of the annulus has the properties stated in the first part of our lemma (see [2]).

The second part of the lemma is immediate from the first. Namely, let $\alpha, \beta \in \mathcal{MC}(S)$ and let $a > 0$. Choose $\delta \in L_a(\alpha, \beta, k_1)$ which satisfies the properties stated

in the first part of the lemma. If $\gamma \in L_a(\alpha, \beta, R)$ for some $R > 0$ then we have $i(\gamma, \delta) \leq k_1 R$ and hence $d(\gamma, \delta) \leq k_1 R + 1$. \square

As an immediate consequence of Lemma 3.2 and Lemma 3.3 we observe that there is a universal number $D_1 > 0$ with the following property. Let P be a pants decomposition for S , let $\beta \in \mathcal{MC}(S)$ and let $(\tau_i)_{0 \leq i \leq m} \subset \mathcal{TT}$ be any $P \rightarrow \beta$ -splitting sequence; then the Hausdorff distance in $\mathcal{C}(S)$ between the sets $\{\Phi(\tau_i) \mid 0 \leq i \leq m\}$ and $\bigcup_{a>0} L_a(\beta, P, k_1)$ is at most $D_1/16$. If $c > 0$ and if $j \leq m$ is such that there is a vertex cycle γ for $\tau(j)$ which is contained in $L_c(\beta, P, k_1)$ then the splitting sequence $(\tau_i)_{0 \leq i \leq j}$ is a $P \rightarrow \gamma$ -splitting sequence and hence the Hausdorff distance between $\bigcup_{a>0} L_a(\gamma, P, k_1)$ and $\bigcup_{a \geq c} L_a(\beta, P, k_1)$ is at most $D_1/8$. Moreover, for every $\beta \in \mathcal{C}(S)$ and every simple multi-curve Q containing β as a component the Hausdorff-distance between the sets $\bigcup_{a>0} L_a(\beta, P, k_1)$ and $\bigcup_{a>0} L_a(Q, P, k_1)$ is at most $D_1/8$. Thus if Q, Q' are pants decompositions for S containing a common curve $\beta \in \mathcal{C}(S)$ then the Hausdorff distance between $\bigcup_a L_a(Q, P, k_1)$ and $\bigcup_a L_a(Q', P, k_1)$ is at most $D_1/4$.

On the other hand, for multi-curves $P, Q \in \mathcal{MC}(S)$ we have

$$\bigcup_{a>0} L_a(P, Q, k_1) = \bigcup_{a>0} L_a(Q, P, k_1).$$

Therefore from two applications of our above consideration we obtain the following. Let $\alpha, \beta \in \mathcal{C}(S)$ and let P, P', Q, Q' be any pants decompositions for S containing α, β ; then the Hausdorff distance between $\bigcup_{a>0} L_a(P, Q, k_1)$ and $\bigcup_{a>0} L_a(P', Q', k_1)$ is not bigger than $D_1/2$. By our choice of D_1 this implies that the Hausdorff distance between the images under Φ of any $\alpha \rightarrow \beta$ - or $\beta \rightarrow \alpha$ -splitting sequences is bounded from above by D_1 .

Now let $\alpha, \beta, \gamma \in \mathcal{C}(S)$ be such that their pairwise distance in $\mathcal{C}(S)$ is at least 3; then any two of these curves jointly fill up S . Choose pants decompositions $P_\alpha, P_\beta, P_\gamma$ containing α, β, γ . Then there are unique numbers $a, b, c > 0$ such that $abi(P_\alpha, P_\beta) = bci(P_\beta, P_\gamma) = aci(P_\gamma, P_\alpha) = 1$. By construction, we have

$$L_a(P_\alpha, P_\beta, k_1) = L_b(P_\beta, P_\alpha, k_1), \quad L_b(P_\beta, P_\gamma, k_1) = L_c(P_\gamma, P_\beta, k_1)$$

and $L_c(P_\gamma, P_\alpha, k_1) = L_a(P_\alpha, P_\gamma, k_1)$. Choose a point $\delta \in L_a(P_\alpha, P_\beta, k_1)$ such that for every ζ in $\mathcal{MC}(S)$ we have

$$i(\delta, \zeta) \leq k_1 \max\{ai(P_\alpha, \zeta), i(\zeta, P_\beta)/ai(P_\alpha, P_\beta)\};$$

such a point exists by Lemma 3.3. Applying this inequality to $\zeta = P_\gamma$ yields $ci(\delta, P_\gamma) \leq k_1$. For $\zeta = P_\beta$ we obtain

$$i(\delta, P_\beta)/ci(P_\gamma, P_\beta) \leq ai(P_\alpha, P_\beta)/ci(P_\gamma, P_\beta) = 1,$$

and for $\zeta = P_\alpha$ we obtain

$$i(\delta, P_\alpha)/ci(P_\gamma, P_\alpha) \leq 1/aci(P_\gamma, P_\alpha) = 1.$$

Therefore we have $\delta \in L_c(P_\gamma, P_\beta, k_1) \cap L_a(P_\alpha, P_\beta, k_1) \cap L_c(P_\gamma, P_\alpha, k_1)$. Together with Lemma 3.2 and our above remark we conclude that there is a universal constant $D_2 > 0$ such that the distance between $\phi(\alpha, \beta, \gamma) = \delta$ and the image under Φ of any $\alpha \rightarrow \beta$ -splitting sequence, any $\alpha \rightarrow \gamma$ -splitting sequence and any $\gamma \rightarrow \beta$ -splitting sequence is bounded from above by D_2 .

We use the map ϕ to derive the δ -thin triangle condition for triangles whose sides are images under the map Φ of splitting sequences in $\mathcal{T}T$.

Lemma 3.4. *There is a number $D_3 > 0$ with the following property. Let $\alpha, \beta, \gamma \in \mathcal{C}(S)$ and let a, b, c be the image under Φ of a $\beta \rightarrow \gamma, \gamma \rightarrow \alpha, \alpha \rightarrow \beta$ -splitting sequence. Then the D_3 -neighborhood of $a \cup b$ contains c .*

Proof. Let $\alpha, \beta, \gamma \in \mathcal{C}(S)$ and assume that $d(\beta, \gamma) \leq p$ for some $p > 0$. Let $(\tau_i)_{0 \leq i \leq m}$ be an $\alpha \rightarrow \beta$ -splitting sequence and let $(\eta_j)_{0 \leq j \leq \ell}$ be an $\alpha \rightarrow \gamma$ -splitting sequence; if $D_1 > 0$ is as above then the Hausdorff distance between $\{\Phi(\tau_i) \mid 0 \leq i \leq m\}$ and $\{\Phi(\eta_j) \mid 0 \leq j \leq \ell\}$ is at most $2pD_1$. Namely, we observed that the Hausdorff distance between the image under Φ of any two $\alpha \rightarrow \beta$ -splitting sequences is bounded from above by D_1 . Moreover, if $d(\beta, \gamma) = 1$ then $\beta \cup \gamma \in \mathcal{MC}(S)$ and hence there is an $\alpha \rightarrow \beta$ -splitting sequence which also is an $\alpha \rightarrow \gamma$ -splitting sequence. Thus the statement of the corollary holds for $p = 1$, and the general case follows from a successive application of this fact for the points on a geodesic in $\mathcal{C}(S)$ connecting β to γ .

Now let $\alpha, \beta, \gamma \in \mathcal{C}(S)$ be arbitrary points whose pairwise distance is at least 3. Let again $(\tau_i)_{0 \leq i \leq m}$ be an $\alpha \rightarrow \beta$ -splitting sequence. By the definition of ϕ and the choice of the constant $D_2 > 0$ above there is some $i_0 \leq m$ such that the distance between $\Phi(\tau_{i_0})$ and $\phi(\alpha, \beta, \gamma)$ is at most D_2 . Let $(\eta_j)_{0 \leq j \leq \ell}$ be any $\alpha \rightarrow \phi(\alpha, \beta, \gamma)$ -splitting sequence. By our above consideration, the Hausdorff distance between $\{\Phi(\tau_i) \mid 0 \leq i \leq i_0\}$ and $\{\Phi(\eta_j) \mid 0 \leq j \leq \ell\}$ is at most $2D_1D_2$. Similarly, let $(\zeta_j)_{0 \leq j \leq n}$ be an $\alpha \rightarrow \gamma$ -splitting sequence. Then there is some $j_0 > 0$ such that $d(\Phi(\zeta_{j_0}), \phi(\alpha, \beta, \gamma)) \leq D_2$. By our above argument, the Hausdorff distance between the sets $\{\Phi(\tau_i) \mid 0 \leq i \leq i_0\}$ and $\{\Phi(\zeta_j) \mid 0 \leq j \leq j_0\}$ is at most $4D_1D_2$.

As a consequence, there are numbers $a(\alpha, \beta) > 0$ and $a(\alpha, \gamma) > 0$ such that $\phi(\alpha, \beta, \gamma) \in L_{a(\alpha, \beta)}(\beta, \alpha, k_1) \cap L_{a(\alpha, \gamma)}(\gamma, \alpha, k_1)$ and that the Hausdorff distance between $\bigcup_{a \geq a(\alpha, \beta)} L_a(\beta, \alpha, k_1)$ and $\bigcup_{a \geq a(\alpha, \gamma)} L_a(\gamma, \alpha, k_1)$ is at most $6D_1D_2$. The same argument, applied to a $\beta \rightarrow \alpha$ -splitting sequence and a $\beta \rightarrow \gamma$ -splitting sequence, shows that $\bigcup_{a \leq a(\alpha, \beta)} L_a(\alpha, \beta, k_1)$ is contained in the $6D_1D_2$ -neighborhood of $\bigcup_a L_a(\beta, \gamma, k_1)$. Then $\{\Phi(\tau_i) \mid 0 \leq i \leq m\}$ is contained in the $12D_1D_2$ -neighborhood of the union of the image under Φ of a $\gamma \rightarrow \alpha$ -splitting sequence and a $\beta \rightarrow \gamma$ -splitting sequence. This shows the lemma. \square

Hyperbolicity of the curve graph now follows from Lemma 3.4 and the following criterion.

Proposition 3.5. *Let (X, d) be a geodesic metric space. Assume that there is a number $D > 0$ and for every pair of points $x, y \in X$ there is an arc $\eta(x, y): [0, 1] \rightarrow X$ connecting $\eta(x, y)(0) = x$ to $\eta(x, y)(1) = y$ so that the following conditions are satisfied.*

- (1) *If $d(x, y) \leq 1$ then the diameter of $\eta(x, y)[0, 1]$ is at most D .*
- (2) *For $x, y \in X$ and $0 \leq s \leq t \leq 1$, the Hausdorff distance between $\eta(x, y)[s, t]$ and $\eta(\eta(x, y)(s), \eta(x, y)(t))[0, 1]$ is at most D .*
- (3) *For any $x, y, z \in X$ the set $\eta(x, y)[0, 1]$ is contained in the D -neighborhood of $\eta(x, z)[0, 1] \cup \eta(z, y)[0, 1]$.*

Then (X, d) is δ -hyperbolic for a number $\delta > 0$ only depending on D .

Proof. Let (X, d) be a geodesic metric space. Assume that there is a number $D > 0$ and there is a family of paths $\eta(x, y): [0, 1] \rightarrow X$, one for every pair of points $x, y \in X$, which satisfy the hypotheses in the statement of the proposition. To show hyperbolicity for X it is then enough to show the existence of a constant $\kappa > 0$ such that for all $x, y \in X$ and every geodesic $\nu: [0, \ell] \rightarrow X$ connecting x to y , the Hausdorff-distance between $\nu[0, \ell]$ and $\eta(x, y)[0, 1]$ is at most κ . Namely, if this is the case then for every geodesic triangle with sides a, b, c the side a is contained in the $3\kappa + D$ -neighborhood of $b \cup c$.

To show the existence of such a constant $\kappa > 0$, let $x, y \in X$ and let $c: [0, 2^k] \rightarrow X$ be any path of length $\ell(c) = 2^k$ parametrized by arc length connecting x to y . Write $\eta_1 = \eta(c(0), c(2^{k-1}))$ and write $\eta_2 = \eta(c(2^{k-1}), c(2^k))$. By our assumption, the D -neighborhood of $\eta_1 \cup \eta_2$ contains $\eta(c(0), c(2^k))$. Repeat this construction with the points $c(2^{k-2}), c(3 \cdot 2^{k-2})$ and the arcs η_1, η_2 . Inductively we conclude that the path $\eta(c(0), c(2^k))$ is contained in the $(\log_2 \ell(c))D$ -neighborhood of a path $\tilde{c}: [0, 2^k] \rightarrow X$ whose restriction to each interval $[m-1, m]$ ($m \leq 2^k$) equals up to parametrization the arc $\eta(c(m-1), c(m))$. Since $d(c(m-1), c(m)) \leq 1$, by assumption the diameter of each of the sets $\eta(c(m-1), c(m))[0, 1]$ is bounded from above by D and therefore the arc $\eta(c(0), c(2^k))$ is contained in the $(\log_2 \ell(c))D + D$ -neighborhood of $c[0, 2^k]$.

Now let $c: [0, k] \rightarrow X$ be a geodesic connecting $c(0) = x$ to $c(k) = y$ which is parametrized by arc length. Let $t > 0$ be such that $\eta(x, y)(t)$ has maximal distance, say χ , to $c[0, k]$. Choose some $s > 0$ such that $d(c(s), \eta(x, y)(t)) = \chi$ and let $t_1 < t < t_2$ be such that $d(\eta(x, y)(t), \eta(x, y)(t_u)) = 2\chi$ ($u = 1, 2$). In the case that there is no $t_1 \in [0, t)$ (or $t_2 \in (t, 1]$) with $d(\eta(x, y)(t), \eta(x, y)(t_1)) \geq 2\chi$ (or $d(\eta(x, y)(t), \eta(x, y)(t_2)) \geq 2\chi$) we choose $t_1 = 0$ (or $t_2 = 1$). By our choice of χ , there are numbers $s_u \in [0, k]$ such that $d(c(s_u), \eta(x, y)(t_u)) \leq \chi$ ($u = 1, 2$). Then the distance between $c(s_1)$ and $c(s_2)$ is at most 6χ . Compose the subarc $c[s_1, s_2]$ of c with a geodesic connecting $\eta(x, y)(t_1)$ to $c(s_1)$ and a geodesic connecting $c(s_2)$ to $\eta(x, y)(t_2)$. We obtain a curve ν of length at most 8χ . By our above observation, the $(\log_2(8\chi))D + D$ -neighborhood of this curve contains the arc $\eta(\eta(x, y)(t_1), \eta(x, y)(t_2))$. However, the Hausdorff distance between $\eta(x, y)[t_1, t_2]$

and $\eta(\eta(x, y)(t_1), \eta(x, t)(t_2))$ is at most D and therefore the $(\log_2(8\chi))D + 2D$ -neighborhood of the arc ν contains $\eta(x, y)[t_1, t_2]$. But the distance between $\eta(t)$ and our curve ν equals χ by construction and hence we have $\chi \leq (\log_2(8\chi))D + 2D$. In other words, χ is bounded from above by a universal constant $\kappa_1 > 0$, and $\eta(x, y)$ is contained in the κ_1 -neighborhood of the geodesic c .

A similar argument also shows that the $3\kappa_1$ -neighborhood of $\eta(x, y)$ contains $c[0, k]$. Namely, by the above consideration, for every $t \leq 1$ the set $A(t) = \{s \in [0, k] \mid d(c(s), \eta(x, y)(t)) \leq \kappa_1\}$ is a non-empty closed subset of $[0, k]$. The diameter of the sets $A(t)$ is bounded from above by $2\kappa_1$. Assume to the contrary that $c[0, k]$ is not contained in the $3\kappa_1$ -neighborhood of $\eta(x, y)$. Then there is a subinterval $[a_1, a_2] \subset [0, k]$ of length $a_2 - a_1 \geq 4\kappa_1$ such that $d(c(s), \eta(x, y)[0, 1]) > \kappa_1$ for every $s \in (a_1, a_2)$. Since $d(c(s), c(t)) = |s - t|$ for all s, t we conclude that for every $t \in [0, 1]$ the set $A(t)$ either is entirely contained in $[0, a_1]$ or it is entirely contained in $[a_2, k]$. Define $C_1 = \{t \in [0, 1] \mid A(t) \subset [0, a_1]\}$ and $C_2 = \{t \in [0, 1] \mid A(t) \subset [a_2, k]\}$. Then the sets C_1, C_2 are disjoint and their union equals $[0, 1]$; moreover, we have $0 \in C_1$ and $1 \in C_2$. On the other hand, the sets C_i are closed. Namely, let $(t_i) \subset C_1$ be a sequence converging to some $t \in [0, 1]$. Let $s_i \in A(t_i)$ and assume after passing to a subsequence that $s_i \rightarrow s \in [0, a_1]$. Now $\kappa_1 \geq d(c(s_i), \eta(x, y)(t_i)) \rightarrow d(c(s), \eta(x, y)(t))$ and therefore $s \in A(t)$ and hence $t \in C_1$. However, $[0, 1]$ is connected and hence we arrive at a contradiction. In other words, the geodesic c is contained in the $3\kappa_1$ -neighborhood of $\eta(x, y)$. This completes the proof of the proposition. \square

As an immediate corollary, we obtain the following result.

Theorem 3.6 ([19], [2]). *The curve graph is hyperbolic.*

Proof. Let $\Phi: \mathcal{T}\mathcal{T} \rightarrow \mathcal{C}(S)$ be as before. For $\alpha, \beta \in \mathcal{C}(S)$ choose an $\alpha \rightarrow \beta$ -splitting sequence $(\tau_i)_{0 \leq i \leq m}$. Define an arc $\eta(\alpha, \beta): [0, 1] \rightarrow \mathcal{C}(S)$ by requiring that for $1 \leq i \leq m$ the restriction of $\eta(\alpha, \beta)$ to the interval $[\frac{i}{m+2}, \frac{i+1}{m+2}]$ is a geodesic connecting $\Phi(\tau_{i-1})$ to $\Phi(\tau_i)$ and that the restriction of $\eta(\alpha, \beta)$ to $[0, \frac{1}{m+2}]$ (or $[\frac{m-1}{m-2}, 1]$) is a geodesic connecting α to $\Phi(\tau_0)$ (or $\Phi(\tau_m)$ to β).

We claim that this family of arcs satisfy the assumptions in Proposition 3.5. Namely, we observed before that for all $\alpha, \beta \in \mathcal{C}(S)$, the Hausdorff distance between the image under Φ of any two $\alpha \rightarrow \beta$ -splitting sequences is bounded from above by a universal constant. Now if $(\tau_i)_{0 \leq i \leq m}$ is any splitting sequence, then for all $0 \leq k \leq \ell \leq m$ the sequence $(\tau_i)_{k \leq i \leq \ell}$ is a $\Phi(\tau_k) \rightarrow \Phi(\tau_\ell)$ -splitting sequence and hence our curve system satisfies the second condition in Proposition 3.5.

Moreover, curves $\alpha, \beta \in \mathcal{C}(S)$ with $d(\alpha, \beta) = 1$ can be realized disjointly, and $\alpha \cup \beta$ is a multi-curve. For such a pair of curves we can choose a constant $\alpha \rightarrow \beta$ -splitting sequence; hence our curve system also satisfies the first condition stated in Proposition 3.5. Finally, the third condition was shown to hold in Lemma 3.4.

Now hyperbolicity of the curve graph follows from Proposition 3.5. \square

A curve $c: [0, m] \rightarrow \mathcal{C}(S)$ is called an *unparametrized p -quasi-geodesic* for some $p > 1$ if there is a homeomorphism $\rho: [0, u] \rightarrow [0, m]$ for some $u > 0$ such that

$$d(c(\rho(s)), c(\rho(t)))/p - p \leq |s - t| \leq pd(c(\rho(s)), c(\rho(t))) + p$$

for all $s, t \in [0, u]$. We define a map $c: \{0, \dots, m\} \rightarrow \mathcal{C}(S)$ to be an unparametrized q -quasi-geodesic if this is the case for the curve \tilde{c} whose restriction to each interval $[i, i + 1)$ coincides with $c(i)$. The following observation is immediate from Proposition 3.5 and its proof.

Corollary 3.7 ([20], [7]). *There is a number $p > 0$ such that the image under Φ of an arbitrary splitting sequence is an unparametrized p -quasi-geodesic.*

Proof. By Proposition 3.5, Theorem 3.6 and their proofs, there is a universal number $D > 0$ with the property that for every splitting sequence $(\tau_i)_{0 \leq i \leq m}$ and every geodesic $c: [0, s] \rightarrow \mathcal{C}(S)$ connecting $c(0) = \Phi(\tau_0)$ to $c(s) = \Phi(\tau_m)$, the Hausdorff distance between the sets $\{\Phi(\tau_i) \mid 0 \leq i \leq m\}$ and $c[0, s]$ is at most D . From this the corollary is immediate. □

4 Geometry of Teichmüller space

In this section, we relate the geometry of the curve graph to the geometry of Teichmüller space equipped with the Teichmüller metric. For this we first define a map $\Psi: \mathcal{T}_{g,m} \rightarrow \mathcal{C}(S)$ as follows. By a well-known result of Bers (see [4]) there is a number $\chi > 0$ only depending on the topological type of S such that for every complete hyperbolic metric on S of finite volume there is a pants decomposition P for S which consists of simple closed geodesics of length at most χ . Since the distance between any two points $\alpha, \beta \in \mathcal{C}(S)$ is bounded from above by $i(\alpha, \beta) + 1$, the collar lemma for hyperbolic surfaces (see [4]) implies that the diameter in $\mathcal{C}(S)$ of the set of simple closed curves whose length with respect to the fixed metric is at most χ is bounded from above by a universal constant $D > 0$. Define $\Psi: \mathcal{T}_{g,m} \rightarrow \mathcal{C}(S)$ by assigning to a finite volume hyperbolic metric h on S a simple closed curve $\Psi(h)$ whose h -length is at most χ . Then for any two maps Ψ, Ψ' with this property and every $h \in \mathcal{T}_{g,m}$ the distance in $\mathcal{C}(S)$ between $\Psi(h)$ and $\Psi'(h)$ is at most D . Moreover, the map Ψ is coarsely equivariant with respect to the action of the mapping class group $\mathcal{M}_{g,m}$ on $\mathcal{T}_{g,m}$ and $\mathcal{C}(S)$: For every $h \in \mathcal{T}_{g,m}$ and every $\phi \in \mathcal{M}_{g,m}$ we have $d(\Psi(\phi(h)), \phi(\Psi(h))) \leq D$.

The following result is due to Masur and Minsky (Theorem 2.6 and Theorem 2.3 of [19]). For its formulation, let d_T be the distance function on $\mathcal{T}_{g,m}$ induced by the Teichmüller metric.

Theorem 4.1 ([19]). (1) *There is a number $a > 0$ such that*

$$d(\Psi h, \Psi h') \leq a d_T(h, h') + a$$

for all $h, h' \in \mathcal{T}_{g,m}$.

(2) *There is a number $\tilde{p} > 0$ with the following property. Let $\gamma : (-\infty, \infty) \rightarrow \mathcal{T}_{g,m}$ be any Teichmüller geodesic; then the assignment $t \rightarrow \Psi(\gamma(t))$ is an unparametrized \tilde{p} -quasi-geodesic in $\mathcal{C}(S)$.*

Proof. Let $\gamma : (-\infty, \infty) \rightarrow \mathcal{T}_{g,m}$ be any Teichmüller geodesic parametrized by arc length. Then the cotangent of γ at $t = 0$ is a quadratic differential q of area one defined by a pair $(\lambda, \mu) \in \mathcal{ML} \times \mathcal{ML}$ of measured geodesic laminations which jointly fill up S . The cotangent of γ at t is given by the quadratic differential $q(t)$ defined by the pair $(e^t\lambda, e^{-t}\mu)$. For $k_1 > 0$ as in Lemma 3.3 and $t \in \mathbb{R}$ let $\zeta(t) \in \mathcal{C}(S)$ be a curve whose $q(t)$ -length is at most $2k_1$. For every $\beta \in [0, 1]$ the $q(t + \beta)$ -length of $\zeta(t)$ is bounded from above by $2ek_1$ and therefore by Lemma 3.3, for every t the distance in $\mathcal{C}(S)$ between $\zeta(t)$ and $\zeta(t + \beta)$ is bounded from above by a universal constant $k_2 > 0$. In particular, the assignment $t \rightarrow \zeta(t)$ satisfies $d(\zeta(s), \zeta(t)) \leq k_2|s - t| + k_2$. Hence for the proof of our lemma, we only have to show that there is a constant $k_3 > 0$ such that for every $h \in \mathcal{T}_{g,m}$ and every holomorphic quadratic differential q of area one for h , the distance between $\Psi(h)$ and a curve on S whose q -length is bounded from above by $2k_1$ is uniformly bounded.

Thus let h be a complete hyperbolic metric of finite volume and let q be a holomorphic quadratic differential for h of area one. By the collar lemma of hyperbolic geometry, a simple closed geodesic c for h whose length is bounded from above by χ is the core curve of an embedded annulus A whose *modulus* is bounded from below by a universal constant $\varepsilon > 0$; we refer to [27] for a definition of the modulus of an annulus and its properties. Then the *extremal length* of the core curve of A is bounded from above by a universal constant $m > 0$. Now the area of q equals one and therefore the q -length of the core curve c does not exceed \sqrt{m} by the definition of extremal length (see e.g. [22]). In other words, the q -length of the curve $\Psi(h)$ is uniformly bounded which together with Lemma 3.3 implies our claim. The theorem follows. □

There are also Teichmüller geodesics in Teichmüller space which are mapped by Ψ to *parametrized* quasi-geodesics in $\mathcal{C}(S)$. For their characterization, denote for $\varepsilon > 0$ by $\mathcal{T}_{g,m}^\varepsilon$ the subset of Teichmüller space consisting of all marked hyperbolic metrics for which the length of the shortest closed geodesic is at least ε . The set $\mathcal{T}_{g,m}^\varepsilon$ is invariant under the action of the mapping class group and projects to a *compact* subset of moduli space. Moreover, every compact subset of moduli space is contained in the projection of $\mathcal{T}_{g,m}^\varepsilon$ for some $\varepsilon > 0$.

Cobounded Teichmüller geodesics, i.e. Teichmüller geodesics which project to a compact subset of moduli space, relate the geometry of Teichmüller space to the geometry of the curve graph. We have

Proposition 4.2 ([8]). *The image under Ψ of a Teichmüller geodesic $\gamma : \mathbb{R} \rightarrow \mathcal{T}_{g,m}$ is a parametrized quasi-geodesic in $\mathcal{C}(S)$ if and only if there is some $\varepsilon > 0$ such that $\gamma(\mathbb{R}) \subset \mathcal{T}_{g,m}^\varepsilon$.*

Minsky [21] discovered earlier that the Teichmüller metric near a cobounded geodesic line has properties similar to properties of a hyperbolic geodesic metric

space. Namely, for a Teichmüller geodesic $\gamma: \mathbb{R} \rightarrow \mathcal{T}_{g,m}^\varepsilon$ the map which associates to a point $h \in \mathcal{T}_{g,m}$ a point on $\gamma(\mathbb{R})$ which minimizes the Teichmüller distance is coarsely Lipschitz and contracts distances in a way which is similar to the contraction property of the closest point projection from a δ -hyperbolic geodesic metric space to any of its bi-infinite geodesics.

A hyperbolic geodesic metric space X admits a *Gromov boundary* which is defined as follows. Fix a point $p \in X$ and for two points $x, y \in X$ define the *Gromov product* $(x, y)_p = \frac{1}{2}(d(x, p) + d(y, p) - d(x, y))$. Call a sequence $(x_i) \subset X$ *admissible* if $(x_i, x_j)_p \rightarrow \infty$ ($i, j \rightarrow \infty$). We define two admissible sequences $(x_i), (y_i) \subset X$ to be *equivalent* if $(x_i, y_i)_p \rightarrow \infty$. Since X is hyperbolic, this defines indeed an equivalence relation (see [3]). The Gromov boundary ∂X of X is the set of equivalence classes of admissible sequences $(x_i) \subset X$. It carries a natural Hausdorff topology. For the curve graph, the Gromov boundary was determined by Klarreich [15] (see also [7]).

For the formulation of Klarreich's result, we say that a minimal geodesic lamination λ *fills up* S if every simple closed geodesic on S intersects λ transversely, i.e. if every complementary component of λ is an ideal polygon or a once punctured ideal polygon with geodesic boundary [5]. For any minimal geodesic lamination λ which fills up S , the number of geodesic laminations μ which contain λ as a sublamination is bounded by a universal constant only depending on the topological type of the surface S . Namely, each such lamination μ can be obtained from λ by successively subdividing complementary components P of λ which are different from an ideal triangle or a once punctured monogon by adding a simple geodesic line which either connects two non-adjacent cusps or goes around a puncture. Note that every leaf of μ which is not contained in λ is necessarily isolated in μ .

Recall that the space \mathcal{L} of geodesic laminations on S equipped with the restriction of the Hausdorff topology for compact subsets of S is compact and metrizable. It contains the set \mathcal{B} of minimal geodesic laminations which fill up S as a subset which is neither closed nor dense. We define on \mathcal{B} a new topology which is coarser than the restriction of the Hausdorff topology as follows. Say that a sequence $(\lambda_i) \subset \mathcal{L}$ *converges in the coarse Hausdorff topology* to a minimal geodesic lamination μ which fills up S if every accumulation point of (λ_i) with respect to the Hausdorff topology contains μ as a sublamination. Define a subset A of \mathcal{B} to be closed if and only if for every sequence $(\lambda_i) \subset A$ which converges in the coarse Hausdorff topology to a lamination $\lambda \in \mathcal{B}$ we have $\lambda \in A$. We call the resulting topology on \mathcal{B} the *coarse Hausdorff topology*. The space \mathcal{B} is not locally compact. Using this terminology, Klarreich's result [15] can be formulated as follows.

Theorem 4.3 ([15], [7]). (1) *There is a natural homeomorphism Λ of \mathcal{B} equipped with the coarse Hausdorff topology onto the Gromov boundary $\partial\mathcal{C}(S)$ of the complex of curves $\mathcal{C}(S)$ for S .*

(2) *For $\mu \in \mathcal{B}$ a sequence $(c_i) \subset \mathcal{C}(S)$ is admissible and defines the point $\Lambda(\mu) \in \partial\mathcal{C}(S)$ if and only if (c_i) converges in the coarse Hausdorff topology to μ .*

Recall that every Teichmüller geodesic in $\mathcal{T}_{g,m}$ is uniquely determined by a pair of projective measured laminations which jointly fill up S . The following corollary is immediate from Theorem 4.1 and Theorem 4.3 with $\tilde{p} > 0$ as in Theorem 4.1.

Corollary 4.4. *Let $\lambda, \mu \in \mathcal{PML}$ be such that their supports are minimal and fill up S . Then the image under Ψ of the unique Teichmüller geodesic in $\mathcal{T}_{g,m}$ determined by λ and μ is a bi-infinite unparametrized \tilde{p} -quasi-geodesic in $\mathcal{C}(S)$, and every bi-infinite unparametrized \tilde{p} -quasi-geodesic in $\mathcal{C}(S)$ is contained in a uniformly bounded neighborhood of a curve of this form.*

Our above discussion also gives information on images under Ψ of a convergent sequence of geodesic lines in Teichmüller space. Namely, if (γ_i) is such a sequence of Teichmüller geodesic lines converging to a Teichmüller geodesic which is determined by a pair of projective measured geodesic laminations (α, β) so that the support of α does not fill up S then there is a curve $\zeta \in \mathcal{C}(S)$, a number $m > 0$ and a sequence $j(i) \rightarrow \infty$ such that $d(\Psi(\gamma_i[0, j(i)]), \zeta) \leq m$.

On the other hand, the image under Ψ of “most” Teichmüller geodesics are unparametrized quasi-geodesics of infinite diameter which are not *parametrized* quasi-geodesics. For example, let $\lambda \in \mathcal{PML}$ be a projective measured geodesic lamination whose support λ_0 is minimal and fills up S but is not *uniquely ergodic*. This means that the dimension of the space of transverse measures supported in λ_0 is at least 2. Let $\gamma: [0, \infty) \rightarrow \mathcal{T}_{g,m}$ be a Teichmüller geodesic ray determined by a quadratic differential whose horizontal foliation corresponds to λ . By a result of Masur [17], the projection of γ to moduli space eventually leaves every compact set. On the other hand, since λ_0 is minimal and fills up S the points $\gamma(t)$ converge as $t \rightarrow \infty$ to λ viewed as a point in the *Thurston boundary* of the Thurston compactification of Teichmüller space [18] (compare also [6] for the construction of the Thurston compactification). By the definition of the map Ψ , the projective measured geodesic laminations defined by the curves $\Psi\gamma(t)$ converge as $t \rightarrow \infty$ to λ and therefore the curves $\Psi\gamma(t)$ converge in the coarse Hausdorff topology to λ_0 . By Theorem 4.3, this implies that the diameter of $\Psi\gamma[0, \infty)$ is infinite.

References

- [1] 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. 450
- [2] B. Bowditch, Intersection numbers and the hyperbolicity of the curve complex. *J. Reine Angew. Math.* 598 (2006), 105–129. 448, 454, 456, 457, 461
- [3] M. Bridson, A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren Math. Wiss. 319, Springer-Verlag, Berlin 1999. 454, 464

- [4] P. Buser, *Geometry and spectra of compact Riemann surfaces*. Progr. Math. 106, Birkhäuser, Boston, MA, 1992. [449](#), [462](#)
- [5] R. Canary, D. Epstein, P. Green, Notes on notes of Thurston. In *Analytical and geometric aspects of hyperbolic space* (D. Epstein, ed.), London Math. Soc. Lecture Note Ser. 111, Cambridge University Press, Cambridge 1987, 3–92. [450](#), [464](#)
- [6] A. Fathi, F. Laudenbach, and V. Poénaru (eds.), Travaux de Thurston sur les surfaces. *Astérisque* 66–67 (1979). [450](#), [465](#)
- [7] U. Hamenstädt, Train tracks and the Gromov boundary of the complex of curves. In *Spaces of Kleinian groups* (Y. Minsky, M. Sakuma, C. Series, eds.), London Math. Soc. Lecture Note Ser. 329, Cambridge University Press, Cambridge 2006, 187–207. [455](#), [456](#), [457](#), [462](#), [464](#)
- [8] U. Hamenstädt, Word hyperbolic extensions of surface groups. Preprint, 2005; arXiv:math.GT/0505244. [463](#)
- [9] U. Hamenstädt, Geometry of the mapping class groups I: Boundary amenability. Preprint, 2006; arXiv:math.GR/0510116. [450](#), [453](#), [455](#)
- [10] U. Hamenstädt, Geometry of the mapping class groups II: (Quasi)-geodesics. Preprint, 2006; arXiv:math.GR/0511349. [456](#)
- [11] W. J. Harvey, Boundary structure of the modular group. In *Riemann Surfaces and Related topics: Proceedings of the 1978 Stony Brook Conference* (I. Kra and B. Maskit, eds.), Ann. of Math. Stud. 97, Princeton, Princeton, N.J., 1981, 245–251. [447](#)
- [12] Y. Imayoshi, M. Taniguchi, *An introduction to Teichmüller spaces*. Springer-Verlag, Tokyo 1992. [448](#), [452](#)
- [13] N. V. Ivanov, Mapping class groups. Chapter 12 in *Handbook of geometric topology* (R. J. Daverman and R. B. Sher, eds.), Elsevier Science, Amsterdam 2002, 523–633. [448](#)
- [14] S. Kerckhoff, Lines of minima in Teichmüller space. *Duke Math. J.* 65 (1992), 187–213. [451](#), [452](#)
- [15] E. Klarreich, The boundary at infinity of the curve complex and the relative Teichmüller space. Unpublished manuscript, 1999. [464](#)
- [16] G. Levitt, Foliations and laminations on hyperbolic surfaces. *Topology* 22 (1983), 119–135. [451](#)
- [17] H. Masur, Interval exchange transformations and measured foliations. *Ann. of Math.* 155 (1982), 169–200. [465](#)
- [18] H. Masur, Two boundaries of Teichmüller space. *Duke Math. J.* 49 (1982), 183–190. [465](#)
- [19] H. Masur, Y. Minsky, Geometry of the complex of curves I: Hyperbolicity. *Invent. Math.* 138 (1999), 103–149. [448](#), [454](#), [455](#), [456](#), [457](#), [461](#), [462](#)
- [20] H. Masur, Y. Minsky, Quasiconvexity in the curve complex. In *In the tradition of Ahlfors and Bers, III, Contemp. Math.* 355, Amer. Math. Soc., Providence, RI, 2004, 309–320. [456](#), [462](#)
- [21] Y. Minsky, Quasi-projections in Teichmüller space. *J. Reine Angew. Math.* 473 (1996), 121–136. [463](#)
- [22] Y. Minsky, Extremal length estimates and product regions in Teichmüller space. *Duke Math. J.* 83 (1996), 249–286. [463](#)

- [23] L. Mosher, Train track expansions of measured foliations. Unpublished manuscript. [453](#)
- [24] J. P. Otal, Le Théorème d'hyperbolisation pour les variétés fibrées de dimension 3. *Astérisque* 235 (1996). [450](#)
- [25] R. C. Penner, J. L. Harer, *Combinatorics of train tracks*. Ann. Math. Studies 125, Princeton University Press, Princeton, Princeton, N.J., 1992. [450](#), [453](#), [454](#), [455](#)
- [26] K. Rafi, A characterization of short curves of a Teichmüller geodesic. *Geom. Topol.* 9 (2005), 179–202. [452](#)
- [27] K. Strebel, *Quadratic differentials*. *Ergeb. Math. Grenzgeb.* (3) 5, Springer-Verlag, Berlin 1984. [451](#), [463](#)

Part C

Surfaces with singularities and discrete Riemann surfaces

Chapter 11

Parameters for generalized Teichmüller spaces

Charalampos Charitos and Ioannis Papadoperakis

Contents

1	Introduction	471
2	The Teichmüller space of a surface	472
3	Hyperbolic surfaces with a conical singularity	476
3.1	Preliminaries	476
3.2	The geometry of a hyperbolic pair of pants with a conical singularity	483
3.3	The Teichmüller space of a hyperbolic pair of pants with a conical singularity	490
4	The Teichmüller space of an ideal 2-dimensional simplicial complex	493
4.1	Definitions and preliminaries	493
4.2	The completeness of ideal hyperbolic structures	497
4.3	Parametrization of the Teichmüller space	497
	References	506

1 Introduction

In this chapter, we consider spaces of hyperbolic metrics on a space X which is homeomorphic to either a surface or a 2-dimensional simplicial complex with its vertices deleted.¹ In the case where X is a surface, the hyperbolic metrics have conical singularities. In the case where X is a 2-dimensional simplicial complex with its vertices deleted, by a hyperbolic metric, we mean a length metric in which every 2-dimensional simplex with its vertices deleted is isometric to an ideal triangle. In each case, the Teichmüller space $\mathcal{T}(X)$ of X is defined as the quotient space of the set of hyperbolic metrics on X by the group of homeomorphisms of X which are homotopic to the identity.

The first section concerns hyperbolic metrics on a surface. We consider in this section surfaces of finite type and we recall Thurston's parametrization of the Teichmüller space by the lengths of finitely many simple closed geodesics.

In the second section we consider hyperbolic surfaces with conical singularities. We parametrize the Teichmüller space of X in the case where X is a hyperbolic

¹Note that in this chapter, a 2-dimensional simplicial complex will be defined as a topological space equipped with a certain decomposition by simplices; this is not the usual definition of a simplicial complex.

pair of pants with a single singularity. This parametrization allows us to show that the Teichmüller space of a hyperbolic pair of pants with a single singularity is contractible.

In the third section we consider a class of 2-dimensional simplicial complexes X which admit a decomposition into topological triangles such that every two triangles have at most one edge in common. We define the Teichmüller space $\mathcal{T}(X)$ of X and, as in the classical case presented in Section 2, we find a finite number of essentially simple closed geodesics whose lengths parametrize $\mathcal{T}(X)$.

We would like to thank G. Tsapogas for his helpful remarks and corrections.

2 The Teichmüller space of a surface

Consider a compact connected oriented differentiable surface X of genus $g \geq 0$ with $k \geq 0$ points removed, and m compact boundary components ($m \geq 0$). We assume that $2g - 2 + m + k \geq 1$. Throughout this section “a hyperbolic structure on X ” will always mean a complete, hyperbolic Riemannian metric on X which makes the boundary ∂X geodesic and the volume of X finite. The condition $2g - 2 + m + k \geq 1$ guarantees that such a structure always exists on X .

Definition 2.1. The Teichmüller space $\mathcal{T}(X)$ of X is the quotient of the set $\mathcal{H}(X)$ of hyperbolic structures on X by $\text{Diff}_0 X$, where $\text{Diff}_0 X$ is the group of diffeomorphisms of X which are homotopic to the identity by a homotopy h_t that respects the boundary for each value of the parameter t .

If h is a hyperbolic structure on X , then, by abuse of language, we also denote by h the equivalence class of h in $\mathcal{T}(X)$.

Given an element $h \in \mathcal{T}(X)$ we shall describe briefly a finite number of simple closed geodesics whose lengths determine h . This well-known result (see for example [9] for the compact case) will be used in an essential way in Section 4.

The main object of this section is an exposition of the following well-known theorem. The next sections contain an exposition of results with a similar flavour for surfaces with singularities.

Theorem 2.2. *For each element $h \in \mathcal{T}(X)$ there exists a finite number of simple closed geodesics in X whose lengths determine h .*

The idea for the proof of Theorem 2.2 is as follows:

Consider a hyperbolic structure h on X . Since X is of finite volume, it has a finite number of boundary components and finitely many cusps (see Theorem 9.8.6 in [16]). The next two properties are well known (see for example Paragraph 1.6 in [4]):

- (P1) Let b be a simple closed curve in X . If b is not contractible to a point and if b is not homotopic to a non-contractible closed curve contained entirely in a cusp, then there is a unique simple closed geodesic c homotopic to b .

(P2) Let b_1, b_2 be two curves satisfying (P1) and, in addition, $b_1 \cap b_2 = \emptyset$. If b_1 is not homotopic to b_2 , then the corresponding simple geodesics c_1 and c_2 respectively, satisfy $c_1 \cap c_2 = \emptyset$.

Based on Properties (P1) and (P2) we may decompose X into surfaces $P_k, k = 1, \dots, n$ of genus 0, with geodesic boundary. Each P_k is homeomorphic to either a disc with two holes i.e. a disc with two open discs removed or, an annulus minus one point or, a disc minus two points. Furthermore, in the case where X is not a torus with one cusp, we may assume that each boundary component of each P_k is adjacent to exactly two distinct surfaces of the decomposition. This geometrical decomposition can be easily established; we first find simple closed curves which give a topological decomposition of X into surfaces with the above mentioned properties. We then replace each such curve by the corresponding simple closed geodesic in its free homotopy class. In the case where X is a torus with a cusp we just cut X along a simple closed geodesic c and we obtain a surface homeomorphic to an annulus with one puncture.

A surface $P_k, k \in \{1, 2, \dots, n\}$ having

three boundary components (i.e. homeomorphic to a disc with two holes) will be called *hyperbolic pair of pants*;

two boundary components (i.e. homeomorphic to an annulus minus one point) will be called *cusped annulus*;

one boundary component (i.e. homeomorphic to a disc minus two points) will be called *double cusped disc*.

In Figure 1, an example of a decomposition of a hyperbolic surface with $g = 2, k = 3$ and $m = 2$ into surfaces of genus 0 is shown.

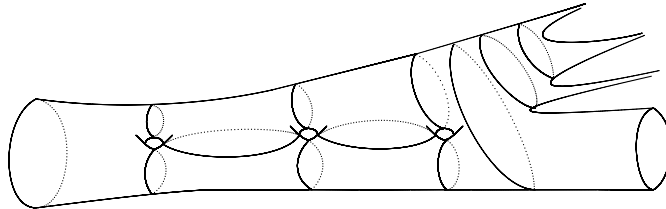


Figure 1

Let $P_k, k \in \{1, 2, \dots, n\}$ be a surface of the decomposition of X . The restriction of the hyperbolic structure h on P_k will be denoted by $h|_{P_k}$. If P_k is a pair of pants, we denote by $\partial_i P_k, i = 1, 2, 3$, the boundary components of P_k and by $l_h(\partial_i P_k)$ the length of $\partial_i P_k$. If P_k has a cusp, then, by abuse of language, we denote by $\partial_i P_k$ either a boundary component of P_k or a cusp of P_k . In the case $\partial_i P_k$ is a cusp we set $l_h(\partial_i P_k) = 0$.

If $\mathcal{T}(P_k)$ denotes the Teichmüller space of P_k and if ρ is an element of $\mathcal{T}(P_k)$ we consider the map

$$L: \mathcal{T}(P_k) \rightarrow [0, \infty)^3$$

defined by

$$L(\rho) = (l_\rho(\partial_1 P), l_\rho(\partial_2 P), l_\rho(\partial_3 P)).$$

It is well known that the map L is a homeomorphism (see Exposé 3, Theorem 5 of [9] for a proof in the case where P_k is a pair of pants). This implies that the lengths $l_h(c_i)$ of the geodesics c_i determine the geometry of each P_k . In other words, the lengths $l_h(c_i)$ parametrize the induced hyperbolic structure $h|_{P_k} \in \mathcal{T}(P_k)$.

Now, the surfaces $P_k, k \in \{1, 2, \dots, n\}$ are the building-blocks of X in the sense that every complete hyperbolic surface X of finite volume can be constructed by considering a finite number of surfaces of genus 0, each being either a hyperbolic pair of pants or, a cusped annulus or, a double cusped disc, and gluing them along their boundaries.

Therefore, we consider the following situation: Let P_1, P_2 be two hyperbolic surfaces each being isometric to either a hyperbolic pair of pants or a cusped annulus or a double cusped disc. Let $h_1 \in \mathcal{T}(P_1)$ and $h_2 \in \mathcal{T}(P_2)$ and assume that P_1, P_2 have boundary components c_1, c_2 respectively such that $l_{h_1}(c_1) = l_{h_2}(c_2) = l \neq 0$. This last condition permits us to glue P_1 and P_2 , by identifying c_1 and c_2 via an isometry, in order to form a surface, say P .

We denote by c the geodesic which is the image of c_1 and c_2 in P by the natural inclusion map. Obviously, there is a degree of freedom when the gluing is performed: any two gluings differ by an amount of rotation $t \in \mathbb{R}$ (see Figure 2). The parameter t is called the *twist parameter along c* and it represents a signed distance measured in multiples of the geodesic length l . (The twist parameter takes values in \mathbb{R} rather than in \mathbb{S}^1 since we want to distinguish isometric hyperbolic structures on P which do not represent the same element in $\mathcal{T}(P)$.)

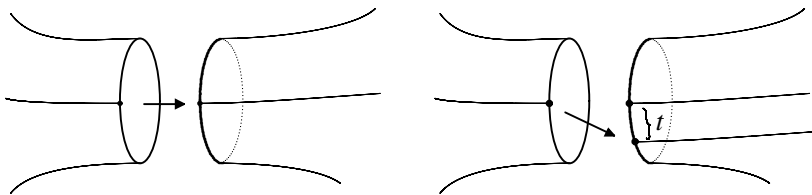


Figure 2

More precisely, assume that a surface P is obtained by gluing P_1 with P_2 and let h be the resulting hyperbolic structure on P . Let a be a simple closed geodesic of P which intersects c in two essential points and thus a is not isotopic to a curve disjoint from c . Denote by a' the geodesic of P which is isotopic to a curve obtained by performing a right Dehn twist on a about c ; such geodesics a and a' are shown in Figure 3.

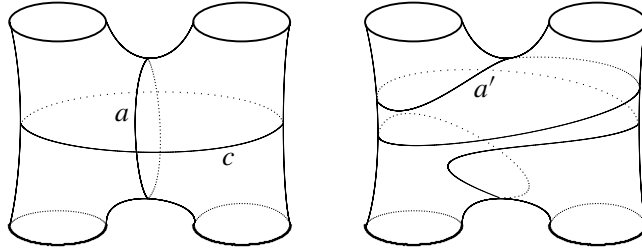


Figure 3

We consider a collar of c in P , homeomorphic to $c \times [0, 1]$ such that $c = c \times \{0\}$. Then, for each $t \in \mathbb{R}$, there exists a diffeomorphism ϕ_t of $c \times [0, 1]$ onto itself such that

- (1) ϕ_t is the identity in a neighborhood of $c \times \{1\}$;
- (2) ϕ_t is an isometry in a neighborhood of $c \times \{0\}$;
- (3) the lift of ϕ_t to the universal covering $\mathbb{R} \times [0, 1]$ is a translation of the form $x \mapsto x + lt$ on $\mathbb{R} \times \{0\}$.

We define a new hyperbolic structure h_t on P as follows: h_t is equal to $\phi_t^*(h)$ on $c \times [0, 1]$, where ϕ_t^* denotes the push-forward and $h_t = h$ outside $c \times [0, 1]$. Obviously $h_0 = h$. We shall be saying that h_t results from h by performing a rotation of angle t along c .

Denote by a_t (respectively a'_t) the geodesic of P with respect to the hyperbolic structure h_t , which is freely homotopic to a (resp. to a'). If we define the function $f: \mathbb{R} \rightarrow \mathbb{R}$ by $f(t) = l_{h_t}(a_t)$ then we can prove that f is strictly convex and takes its minimum at zero (see Exposé 7, Proposition 2, p. 130 in [9]). Similarly for the function $f'(t) = l_{h_t}(a'_t)$, and we may verify easily that $f'(t + 1) = f(t)$ for all t .

Finally, by using the strict convexity of f and f' , as well as the relation $f'(t + 1) = f(t)$, we deduce easily that if $f(t) = f(s)$ and $f'(t) = f'(s)$ then $t = s$. Therefore the lengths $l_{h_t}(a_t)$ and $l_{h_t}(a'_t)$ determine the twist parameter t and hence the hyperbolic structure h_t .

As explained above, to each gluing of pair of pants P_1 and P_2 along boundary components c_1 and c_2 , there corresponds a twist parameter t and geodesics a and a' . If X is not a torus with one cusp then each geodesic c_i of X belongs to two distinct pair of pants, say P_k and P_m , so we may assign to each c_i a twist parameter t_i as well as geodesics a_i and a'_i . From the discussion above it follows that each element $h \in \mathcal{T}(X)$ is determined by either the parameters $\{l_h(c_i), t_i\}_{i=1,2,\dots,n}$ or the parameters $\{l_h(c_i), l_h(a_i), l_h(a'_i)\}_{i=1,2,\dots,n}$. The first family of parameters are known as *Fenchel–Nielsen parameters*.

If X is a cusped torus we consider a cusped annulus by cutting X along a simple closed geodesic c . Let a be a simple closed geodesic which intersects c essentially in two points and let a' be the geodesic which is isotopic to the curve a twisted along c .

In a similar way we may prove that each element $h \in \mathcal{T}(X)$ is determined by the parameters $\{l_h(c), l_h(a), l_h(a')\}$.

Remark. If X is a closed surface of genus $g \geq 2$, U. Hamenstädt in [12] parametrized $\mathcal{T}(X)$ by the lengths of $6g - 5$ simple closed geodesics. These curves are defined by decomposing X into $2g - 2$ pair of pants with the aid of $3g - 3$ geodesics $\{c_i\}$ and it is proven that the $3g - 3$ twist parameters along c_i defined above, are determined by the lengths of $3g - 2$ simple closed geodesics which intersect each c_i once.

3 Hyperbolic surfaces with a conical singularity

This section is based on results published by the authors in [8]. We consider compact hyperbolic surfaces X with a finite number of conical singularities $\{s_1, \dots, s_n\}$. Every point $p \in \text{Int}(X) - \{s_1, \dots, s_n\}$ has a neighborhood isometric to an open subset of the hyperbolic plane \mathbb{H}^2 and the angle at each conical singularity (see Definition 3.1 below) is greater than 2π . Such surfaces X will be called *hyperbolic surfaces with conical singularities* (h.s.c.s.).

In the first paragraph we present examples of h.s.c.s. and describe general properties of them, resulting from the fact that h.s.c.s. are spaces of curvature less than or equal to -1 (see Paragraph 2.1 below for a precise definition). In this way, we shall see that the geometry of h.s.c.s. is quite different from the geometry of hyperbolic surfaces without singularities.

In the second paragraph we consider a pair of pants with a single singularity. We introduce two systems of real parameters which determine the geometry of a pair of pants with a single singularity. These parameters are of two kinds: those depending on the position of the conical singularity and those that do not.

In the third paragraph we state the main result of this section which asserts that the Teichmüller space $\mathcal{T}(X, s)$ of a hyperbolic pair of pants X with a single singularity s , is contractible.

3.1 Preliminaries

We summarize here some standard definitions and statements that we shall need in the sequel (for more details see, for example, [15]). A *geodesic segment* (resp. *geodesic ray*, *geodesic line*) in a metric space $(Y, |\cdot|)$ is an isometric map $f: I = [a, b] \rightarrow Y$ (resp. $I = [0, \infty)$, $I = \mathbb{R}$). A *local geodesic* in Y is a path $f: I \rightarrow Y$ such that for each $t \in I$, there exists an interval $I(t) \subset I$ which is a neighborhood of t in I , such that the restriction of f to $I(t)$ is a geodesic. A *closed geodesic* is a local geodesic $g: (-\infty, +\infty) \rightarrow Y$ which is a periodic map. The metric space Y is said to be *geodesic* if for each two points x, y there is a geodesic segment $f: [a, b] \rightarrow Y$ with $f(a) = x$, $f(b) = y$. We write $[x, y]$ for a geodesic segment joining x with y .

A triangle $T = (xyz)$ in a geodesic metric space Y is a triple x, y, z of points in Y which are called the vertices of T , together with a triple $[x, y], [x, z], [y, z]$ of geodesic segments joining them. These geodesic segments are called the sides of T .

Let \mathbb{H}^2 be the hyperbolic plane. If T is a triangle in Y , then a *comparison triangle for T in \mathbb{H}^2* is a triangle T^* whose sides have the same length as the sides of T , together with a map f from the disjoint union of the sides of T onto the sides of T^* , such that f is consistent on the vertices on T and such that the restriction of f to each side of T is an isometry. The map f is called a *comparison map* for T . We note that the comparison triangle T^* together with its associated comparison map f are unique up to composition by an isometry of \mathbb{H}^2 .

A triangle T in Y is said to satisfy the *CAT(-1)-inequality* if its associated comparison map is distance non-decreasing, that is, for every x, y in T ,

$$|x - y|_Y \leq |f(x) - f(y)|_{\mathbb{H}^2}.$$

A geodesic metric space Y is said to be a *CAT(-1)-space* if every triangle in Y satisfies the *CAT(-1)-inequality*. If a metric space Y satisfies locally the *CAT(-1)-inequality*, i.e. if every point in Y has an open neighborhood which is a *CAT(-1)-space* with respect to the induced metric, then we say that Y is of *curvature less than or equal to -1*.

Assuming that the geodesic metric space Y is locally compact and complete it follows that every bounded subset of Y is relatively compact i.e. Y is proper. If we further assume that Y satisfies the *CAT(-1)-inequality*, then Y is a hyperbolic space in the sense of Gromov (see for example [5] Chapter 1, Theorem 5.1). Now, since Y is geodesic proper and hyperbolic in the sense of Gromov, the (visual) *boundary* $\text{bd}(Y)$ of Y can be defined as the space of equivalence classes of asymptotic geodesic rays starting at a fixed point [5]. On the other hand, it follows from the *CAT(-1)-inequality*, that for every two points x, y in $Y \cup \text{bd}(Y)$ there is a unique (up to parametrization) geodesic joining them.

Finally, if Y is of curvature less than or equal to -1 then its universal covering \tilde{Y} satisfies the *CAT(-1)-inequality*. This follows from a theorem of Cartan–Hadamard (see for example [3], Theorem 4.1). If furthermore Y is compact then for every closed curve a of Y which is not nullhomotopic, there is a unique geodesic in the free homotopy class of a .

We give now the precise definition of hyperbolic surfaces with conical singularities.

Definition 3.1. Let X be a compact orientable surface with or without boundary equipped with a geodesic metric distance $\rho(\cdot, \cdot)$. We say that X has a hyperbolic structure with finitely many conical singularities $\Sigma = \{s_1, \dots, s_n\}$, if there exists a triangulation \mathcal{D} of X by geodesic triangles such that:

- (1) Each triangle of \mathcal{D} , with the induced metric, is isometric to a triangle in the hyperbolic space \mathbb{H}^2 .
- (2) Each $s_i, i = 1, 2, \dots, n$ is a vertex of the triangulation \mathcal{D} .

- (3) If v is a vertex of \mathcal{D} and if we denote by $\theta_{\mathcal{D}}(v)$ the sum of all angles formed by the edges of \mathcal{D} which have v as vertex, then

$$\begin{aligned} \theta_{\mathcal{D}}(v) &= 2\pi \text{ if } v \in \text{Int}(X) - \Sigma \quad \text{and} \quad \theta_{\mathcal{D}}(v) = \pi \text{ if } v \in \partial X - \Sigma, \\ \theta_{\mathcal{D}}(v) &> 2\pi \text{ if } s_i \in \text{Int}(X) \cap \Sigma \quad \text{and} \quad \theta_{\mathcal{D}}(v) > \pi \text{ if } s_i \in \partial X \cap \Sigma \end{aligned}$$

where ∂X is the boundary of X .

The metric $\rho(\cdot, \cdot)$ is called a hyperbolic structure on X and X is called a *hyperbolic surface with conical singularities (h.s.c.s.)*. The angle $\theta_{\mathcal{D}}(s_i)$ will be called the *opening* of the conical singularity s_i with respect to \mathcal{D} .

We note here that the definition is independent of the triangulation \mathcal{D} , i.e., if \mathcal{D}' is another triangulation of X which satisfies conditions (1) and (2) above, and if v is a vertex of \mathcal{D}' , then

$$\begin{aligned} \theta_{\mathcal{D}'}(v) &= 2\pi \text{ if } v \in \text{Int}(X) - \Sigma \quad \text{and} \quad \theta_{\mathcal{D}'}(v) = \pi \text{ if } v \in \partial X - \Sigma, \\ \theta_{\mathcal{D}'}(s_i) &= \theta_{\mathcal{D}}(s_i) \quad \text{for each } i = 1, 2, \dots, n. \end{aligned}$$

Since the opening of a conical singularity is independent of the triangulation, we may talk about the opening $\theta(s_i)$ of a conical singularity without reference to the triangulation considered. In what follows we shall denote by $\mathcal{H}(X)$ the space of all hyperbolic structures on X equipped with the compact-open topology. In [13], [14] and [18], h.s.c.s. are defined in an equivalent way by using conformal metrics.

Let now X be a compact h.s.c.s. Then X is a space of curvature less than or equal to -1 . This follows from a theorem of Gromov (see for example [11], 4.2.D or [1], Chapter 10, Theorem 15) which asserts, in dimension 2, that a locally finite polyhedron K obtained by gluing together geodesic triangles of the hyperbolic space \mathbb{H}^2 , is of curvature less than or equal to -1 , if and only if for each $x \in K$ the direction space at x satisfies the CAT(1)-inequality, equivalently $\theta(v) \geq 2\pi$ for each vertex $v \in K$.

If we denote by \tilde{X} the universal covering of X , then \tilde{X} is simply connected, geodesic, complete and locally compact. Therefore \tilde{X} satisfies the CAT(-1)-inequality globally, as remarked above.

If X is a closed h.s.c.s. then \tilde{X} is homeomorphic to \mathbb{R}^2 since X cannot be homeomorphic to \mathbb{S}^2 . On the other hand $X = \tilde{X}/G$ where G is a discrete group of isometries of \tilde{X} which acts freely and cocompactly on \tilde{X} . It is well known that G is a hyperbolic group in the sense of Gromov and $\text{bd}(\tilde{X}) = \text{bd}(G)$ (see for example Theorem 4.1 of [5]). But since G is a surface group, $\text{bd}(G) = \mathbb{S}^1$. From this we easily deduce that the compactification $\tilde{X} \cup \text{bd}(\tilde{X})$ of \tilde{X} is homeomorphic to the closed unit 2-disc D .

Definition 3.2. A *polygon* $P = (x_1 x_2 \dots x_n)$ in \tilde{X} is a subset of \tilde{X} homeomorphic to the closed unit 2-disc D such that the subarc of ∂P between two consecutive points x_i, x_{i+1} is a geodesic segment of \tilde{X} . This geodesic segment $[x_i, x_{i+1}]$ will be called an edge of P . The points x_i will be called vertices of P .

We say that a polygon P of \tilde{X} is *realizable* in \mathbb{H}^2 if there is a polygon in \mathbb{H}^2 which is isometric to P .

We next present examples of *h.s.c.s.*

Example 1. We consider a regular hyperbolic octagon P in \mathbb{H}^2 such that the sum of the angles of P is greater than 2π . By gluing together the sides of P , as indicated in Figure 4, we obtain a hyperbolic surface of genus 2 with one conical singularity.

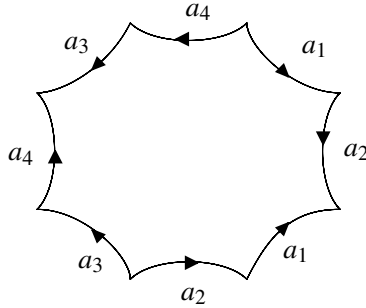


Figure 4

Similarly we can construct a hyperbolic structure with one conical singularity on any closed surface of genus $g > 2$.

Example 2. Let P be a surface homeomorphic to a disc with two holes, equipped with a hyperbolic structure with one conical singularity p in its interior and such that the boundary $\partial P = c_1 \cup c_2 \cup c_3$ consists of closed geodesics (without singularities), i.e. P is a hyperbolic pair of pants with one conical singularity in its interior. Let d_i , $i = 1, 2, 3$ be geodesic arcs realizing the distance between p and c_i respectively. For each $i = 1, 2, 3$ we consider the geodesic loop a_i with base point p (i.e. a_i is a local geodesic arc whose initial and terminal point is p) which is freely homotopic to c_i and we may assume, as one can see from the construction below, that $a_i \cap a_j = \{p\}$ for $i \neq j$ (see Figure 5 (a)). We denote by $l(c_i)$, $l(d_i)$, $l(a_i)$ the lengths of c_i , d_i , a_i respectively. We cut P along the arcs d_i , $i = 1, 2, 3$. Then we obtain a connected surface Q which can be decomposed into one hyperbolic triangle whose sides have lengths $l(a_i)$ and three quadrilaterals, each with two right angles. The sides of these quadrilaterals are shown in Figure 5 (b). Therefore, by following the inverse procedure, we consider a hyperbolic triangle $T = (xyz)$ and three hyperbolic quadrilaterals Q_1 , Q_2 , Q_3 , each one with two right angles, which can be glued together to form T , as shown in Figure 5 (b). Denote by $\angle(x, T)$ the angle of T at x and by $\angle(x, Q_i)$ the angle of Q_i at x and we define $\Phi(x) = \angle(x, Q_1) + \angle(x, Q_2) + \angle(x, T)$. Similarly we define $\Phi(y)$ and $\Phi(z)$. We can assume that $\Phi(x) + \Phi(y) + \Phi(z) > 2\pi$. Now, by identifying the equal sides of each quadrilateral Q_i , we obtain a hyperbolic pair of pants with one conical singularity.

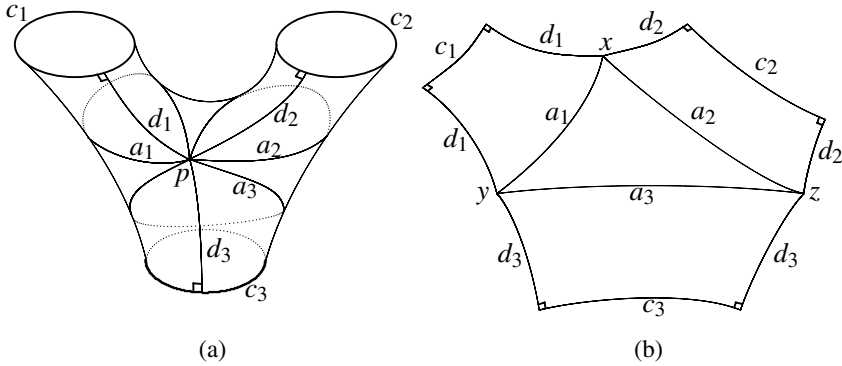


Figure 5

Finally by gluing by isometries along their boundaries $2g - 2$ hyperbolic pair of pants, each one with a single singularity, we obtain a closed h.s.c.s. of genus g with $2g - 2$ conical singularities.

Let now s_1, \dots, s_n be conical singularities of X and for $i = 1, \dots, n$ let θ_i be the opening of s_i . Then we have the following proposition which connects the number of singularities with their opening and the genus of X .

Proposition 3.3. *If X is a closed h.s.c.s. of genus g , then $\sum_{i=1}^n (\theta_i - 2\pi) < (4g - 4)\pi$.*

Proof. We consider a triangulation \mathcal{D} of X by hyperbolic triangles such that the points s_i are vertices of \mathcal{D} . Let T, E, V be the number of triangles, edges and vertices of \mathcal{D} respectively. Then the Euler characteristic is $V - E + T = 2 - 2g$. On the other hand, $3T = 2E$, since each triangle has three edges and each edge is adjacent to exactly two triangles. The sum of the angles of all triangles is smaller than $T\pi$. Therefore, if we take the sum of the angles around each vertex, we have

$$\sum_{i=1}^n \theta_i + (V - n)2\pi < T\pi.$$

Thus,

$$\sum_{i=1}^n (\theta_i - 2\pi) + 2n\pi + 2(V - n)\pi < (V + 4g - 4)\pi,$$

and so

$$\sum_{i=1}^n (\theta_i - 2\pi) < (4g - 4)\pi. \quad \square$$

We deduce the following

Corollary 3.4. *Let X be a h.s.c.s. of genus 0 with geodesic boundary and let m be the number of boundary components, $m \geq 3$. Then $\sum_{i=1}^n (\theta_i - 2\pi) < (2m - 4)\pi$.*

Proof. By taking the double of X we obtain a closed h.s.c.s. of genus $g = m - 1$. Now the corollary follows by Proposition 3.3. \square

We exhibit a few properties of h.s.c.s. in order to further understand their geometry.

We begin with an example which shows that if γ is a simple closed non-contractible curve on a h.s.c.s. X then the unique closed geodesic which exists in the free homotopy class of γ is not necessarily simple, even if X has a single singularity.

Example 3. We glue eight hyperbolic triangles $T_i = (sx_i x_{i+1})$, $i = 1, \dots, 7$ and $T_8 = (sx_8 x_1)$ as it is shown in Figure 6 and denote by $P = (x_1 \dots x_8)$ the resulting space. The triangles T_i are chosen so that

- each T_i is isosceles;
- T_i is isometric to T_j for all i, j ;
- the sum of the angles over all vertices x_1, \dots, x_8 of P is equal to 2π ;
- there is a conical singularity of angle 5π at the center s of P .

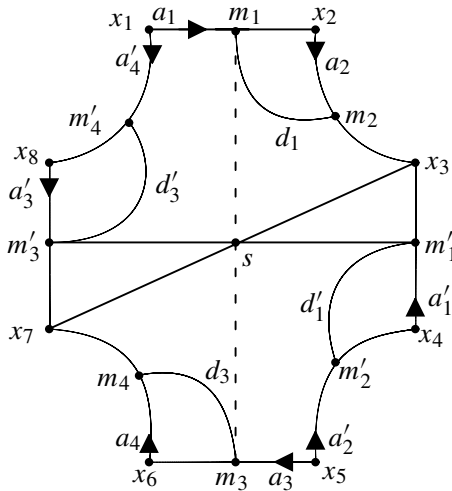


Figure 6

So P is a regular hyperbolic octagon. If we glue the sides a_1, a_2, a_3, a_4 of P , as indicated in Figure 4 we obtain a closed surface X of genus 2 with one conical singularity at s . Let m_1, m'_1, m_3, m'_3 be the midpoints of sides $[x_1, x_2], [x_3, x_4], [x_5, x_6], [x_7, x_8]$ respectively. We consider the union $\gamma = [m_1, s] \cup [s, m'_3] \cup [m_3, s] \cup$

$[s, m'_1]$ of the oriented geodesic segments (see Figure 6). Obviously, γ is a closed geodesic in X which has a double point. On the other hand, there is a simple closed curve on X which is freely homotopic to γ .

Assume that X is a compact h.s.c.s. which has only one conical singularity. In this case we may find simple closed geodesics on X which cut X into a surface of genus 0. More precisely we have:

Proposition 3.5. *Let X be a closed h.s.c.s. of genus g with a single conical singularity s_1 . Then there are g simple closed geodesics of X which do not contain s_1 and which cut X into a surface N of genus 0.*

Proof. Let $X = \tilde{X}/G$, where \tilde{X} is the universal covering of X and G is a discrete group of isometries of \tilde{X} which act freely on \tilde{X} . If $p: \tilde{X} \rightarrow X$ is the covering projection we consider a point $s \in p^{-1}(s_1)$. We consider also the Dirichlet polygon $\overline{D(s)}$ for G , with center s (see for example [16], p. 243). For $h \in G$, if $H_h(s) = \{x \in \tilde{X} : \text{dist}(x, s) < \text{dist}(x, h(s))\}$, then $D(s) = \bigcap \{H_h(s) : h \neq 1 \text{ in } G\}$. By standard methods we may prove that $\overline{D(s)} = (x_1x_2 \dots x_{4g})$ is a $4g$ -polygon in \tilde{X} with sides $a_i, a'_i, i = 1, \dots, 2g$. The surface X is obtained by gluing in an appropriate way the sides of $\overline{D(s)}$ (for example the case $g = 2$ is shown in Figure 6).

Now let d_i (resp. d'_i), $i = 1, 3, \dots, 2g - 1$, be the geodesic segments in $\overline{D(s)}$ which join the midpoints m_i, m_{i+1} of a_i and a_{i+1} (resp. the midpoints points m'_i, m'_{i+1} of a'_i and a'_{i+1}). Obviously $[s, m_i]$ and $[s, m'_i]$ are perpendicular to a_i and a'_i respectively. The quadrilateral $(sm_1x_2m_2)$ has two right angles and, therefore, the angle at s of $[s, m_1]$ and $[s, m_2]$ is less than π . Thus, the geodesic segment d_1 does not contain s . Similarly, we may prove that d_i, d'_i do not contain s . Now, we can easily see that d_i, d'_i fit together so that the angles formed at $m_i, m_{i+1}, m'_i, m'_{i+1}$ are equal to π . Therefore, if we set $c_i = d_i \cup d'_i, i = 1, 3, \dots, 2g - 1$, we deduce that $p(c_i)$ are simple closed geodesics in X which do not contain s_1 . Furthermore if we cut X along $p(c_i)$ we obtain a h.s.c.s. N of genus 0 with geodesic boundary and one conical singularity in its interior. □

Now the question that arises naturally is whether the surface N of Proposition 3.5 can be decomposed into pairs of pants by cutting along simple closed geodesics.

We have the following proposition:

Proposition 3.6. *Assume that N has m boundary components, $m \geq 5$. If $\theta(s_1) \geq 2(m - 2)\pi$ then N cannot be decomposed into pairs of pants by cutting it along simple closed geodesics.*

Proof. From Corollary 3.4 we know that $\theta(s_1) < 2(m - 1)\pi$.

If the proposition were not true then we would be able to find a simple closed geodesic in N which does not contain s_1 . But in this case s_1 is contained in the interior of a subsurface N' with $m - 1$ boundary components. Therefore we would have that $\theta(s_1) < 2(m - 2)\pi$, which contradicts our hypothesis. □

We consider now the special case where X is homeomorphic to a closed surface of genus two. Then we have

Proposition 3.7. *A closed surface of genus two can be decomposed into two hyperbolic pair of pants by cutting it along three simple closed geodesics.*

Proof. From Proposition 3.5 we may find two simple closed geodesics which cut X into a surface N of genus 0. Now let $P = \overline{D(s)} = (x_1 \dots x_8)$ be the Dirichlet polygon in \tilde{X} considered in Proposition 3.5.

We consider the piecewise geodesic of $\gamma = [x_3, s] \cup [s, x_7]$ in P . We show that γ projects to a closed geodesic in N and thus γ separates N into two hyperbolic pairs of pants. In fact, the two angles formed by the geodesic segments $[x_3, s]$, $[s, x_7]$ at s are equal, and greater than π . On the other hand, by construction it follows that

$$\angle_P(x_3; s, x_2) + \angle_P(x_4; x_5, x_3) + \angle_P(x_1; x_2, x_8) + \angle_P(x_6; x_7, x_5) + \angle_P(x_7; x_8, s) = \pi$$

and

$$\angle_P(x_3; s, x_4) + \angle_P(x_2; x_1, x_3) + \angle_P(x_5; x_4, x_6) + \angle_P(x_8; x_7, x_1) + \angle_P(x_7; x_6, s) = \pi.$$

From these equalities we deduce immediately that the simple closed curve γ is a geodesic. \square

3.2 The geometry of a hyperbolic pair of pants with a conical singularity

In this section we assume that X is a h.s.c.s. which is homeomorphic to a disc with two holes. We assume further that X has only one conical singularity s . We call such a surface X a *hyperbolic pair of pants with one singularity*. The aim of this section is to find a collection of real parameters with the smallest possible cardinality that determine the geometry of X . These parameters are of two kinds: those depending on the position of the conical singularity and those which do not.

The disc with two holes X is the simplest compact planar surface which admits a hyperbolic structure with conical singularities and geodesic boundary. If X has more than one conical singularity, it is an interesting but harder problem to find parameters which determine the geometry of X .

Let $\partial X = c_1 \cup c_2 \cup c_3$ and we denote by $\rho(\cdot, \cdot)$ the distance on X . Let d_i be the geodesic arc that realizes the distance $\rho(s, c_i)$ between the singularity s and the curve c_i . We denote by l_i the length of c_i and by r_i the length of d_i . Obviously the parameters r_i depend on the position of s . We have the following proposition:

Proposition 3.8. *Let l_i be the lengths of the boundary components c_i and r_i be the lengths of the geodesic arcs d_i between s and c_i for $i = 1, 2, 3$.*

1) *If $s \in \text{Int}(X)$ then the real parameters $l_i, r_i, i = 1, 2, 3$ determine at most four non-isometric hyperbolic structures on X .*

2) *If $s \in \partial X$ then the real parameters $l_i, r_i, i = 1, 2, 3$ determine at most three non-isometric hyperbolic structures on X .*

Proof. First we assume that $s \in \text{Int}(X)$ (see Figure 7 (a)). We cut X along $d_1 \cup d_2 \cup d_3$ to obtain a surface denoted by P . Let $p: \tilde{X} \rightarrow X$ be the covering projection. We may

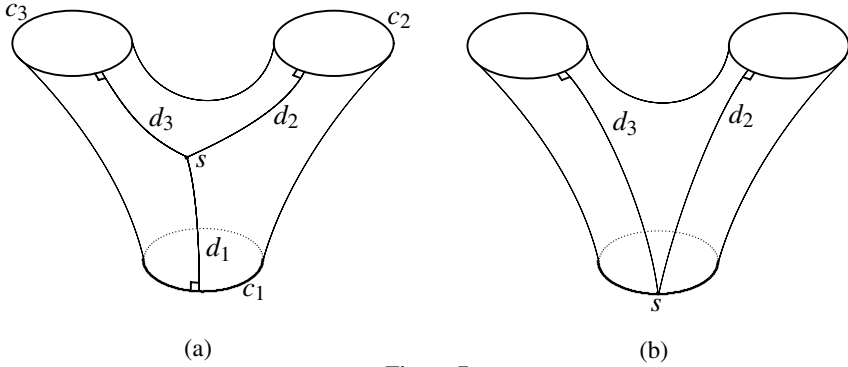


Figure 7

realize P as a polygon, say $P = (qps_1xys_2zws_3)$, isometrically embedded in \tilde{X} , where $s_1, s_2, s_3 \in p^{-1}(s)$. P with the induced geometry satisfies the $\text{CAT}(-1)$ -inequality. We consider the following cases:

Case A_0 : Every geodesic segment $[s_i, s_j]$, $i, j \in \{1, 2, 3\}$, intersects ∂P only at its endpoints. Then the triangle $T = (s_1s_2s_3)$ of P is non-degenerate and therefore it has, by $\text{CAT}(-1)$ inequality, all its angles smaller than π . The quadrilateral (s_1xys_2) is realizable in \mathbb{H}^2 and therefore is uniquely determined up to isometry of \mathbb{H}^2 , by the right angles at x and y and by the lengths $r_2 = \text{length}([s_1, x]) = \text{length}([s_2, y])$ and $l_2 = \text{length}([x, y])$. Therefore the length of $[s_1, s_2]$ is determined by the parameters r_2 and l_2 .

Similarly the quadrilaterals (s_1ps_3) and (s_2zws_3) are realizable in \mathbb{H}^2 and the lengths of $[s_2, s_3]$, $[s_1, s_3]$ are known. From this analysis we can easily construct X in the case A_0 as follows: The triangle $T = (s_1s_2s_3)$ and the quadrilaterals $Q_1 = (s_1xys_2)$, $Q_2 = (s_2zws_3)$, $Q_3 = (s_3qps_1)$ are realized as polygons of \mathbb{H}^2 and they are unique up to isometry. By gluing these polygons along their common sides we get the surface X .

If we denote by $\angle_P(s_i; s_j, s_k)$ the angle in P formed by the geodesic segments $[s_i, s_j]$ and $[s_i, s_k]$ at s_i , then we define the set A_0 of all hyperbolic structures on X such that $\angle_P(s_i; s_j, s_k) < \pi$ for every $i, j, k \in \{1, 2, 3\}$. Obviously to each such hyperbolic structure we may correspond a unique configuration of \mathbb{H}^2 , say of type A_0 , which consists of a triangle T and three quadrilaterals Q_i of \mathbb{H}^2 , glued as in Figure 8 (a). Conversely, to each configuration of type A_0 , we may correspond a unique hyperbolic structure on X .

Finally we remark that in the case A_0 , the whole polygon P is realizable in \mathbb{H}^2 .

Case $A_{0,1}$: We assume that $\angle_P(s_1; s_2, s_3) = \pi$. In this case the triangle $T = (s_1s_2s_3)$ is degenerate and coincides with the geodesic segment $[s_2, s_3]$ (see Figure 8 (b)). We

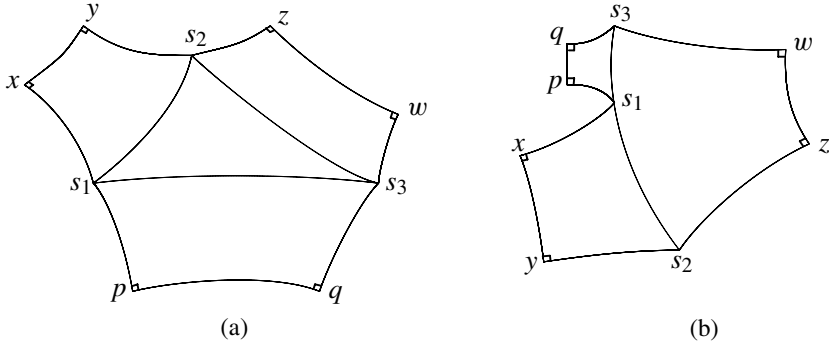


Figure 8

define by $A_{0,1}$ the set of all hyperbolic structures on X such that $\angle_P(s_1; s_2, s_3) = \pi$. Therefore, to each hyperbolic structure in $A_{0,1}$ corresponds a unique configuration of \mathbb{H}^2 , say of type $A_{0,1}$, which consists of three quadrilateral Q_i of \mathbb{H}^2 , glued as in Figure 8 (b). Conversely to each configuration of type $A_{0,1}$, we may assign a unique hyperbolic structure on X .

Similarly to the Case $A_{0,1}$ we define the cases $A_{0,2}$ and $A_{0,3}$ when

$$\angle_P(s_2; s_1, s_3) = \pi \quad \text{and} \quad \angle_P(s_3; s_1, s_2) = \pi$$

respectively. Analogously, we define the subsets $A_{0,2}$ and $A_{0,3}$ of hyperbolic structures on X .

Case A_1 : The geodesic in P which joins the points s_2, s_3 passes through s_1 and $\angle_P(s_1; s_2, s_3) > \pi$. This implies that $[s_2, s_3] = [s_2, s_1] \cup [s_1, s_3]$. Now, if we consider the pentagon $R = (s_1 s_2 z w s_3)$ it then follows that $\angle_R(s_1; s_2, s_3) > \pi$ and R is realizable in \mathbb{H}^2 .

The quadrilaterals $(s_3 p q s_1)$ and $(s_1 x y s_2)$ are uniquely determined up to isometry of \mathbb{H}^2 , similarly to the case A_0 . So the lengths of $[s_1, s_2]$ and $[s_1, s_3]$ are determined. Therefore the pentagon R can be constructed in \mathbb{H}^2 in a unique way, since we know all its edges and that the angles at z, w are right, and that the angle at s_1 is greater than π . Thus, under the assumption $\angle_P(s_1; s_2, s_3) > \pi$, P is uniquely determined, and the surface X as well.

We define by A_1 the set of all hyperbolic structures on X such that $\angle_P(s_1; s_2, s_3) > \pi$. Obviously to each hyperbolic structure in A_1 we may assign a unique configuration, say of type A_1 , which consists of a pentagon R and two quadrilaterals Q_2, Q_3 of \mathbb{H}^2 , glued as in Figure 9. Conversely, to each configuration of type A_1 , we may assign a unique hyperbolic structure on X . The triangle $T = (s_1 s_2 s_3)$ appearing in the configuration of Figure 9 is called *imaginary*, because it does not actually exists in X .

We remark that P is not necessarily realizable as a polygon of \mathbb{H}^2 ; this is the reason that in Figure 9 in \mathbb{H}^2 , the quadrilaterals Q_2 and Q_3 intersect each other.

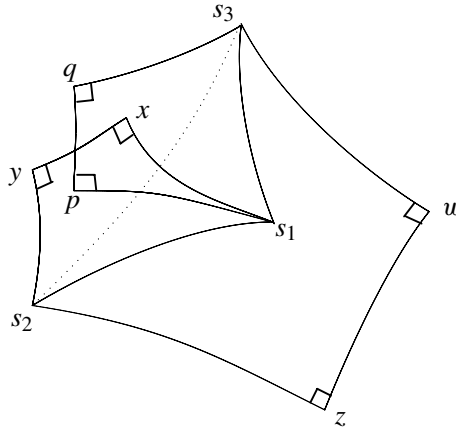


Figure 9

Similarly to the case A_1 , we consider the cases A_2 and A_3 where we have

$$\angle_P(s_2; s_1, s_3) > \pi \quad \text{or} \quad \angle_P(s_3; s_1, s_2) > \pi,$$

respectively. In these cases we also define the corresponding sets A_2 and A_3 and we set

$$\bar{A}_0 = A_0 \cup A_{0,1} \cup A_{0,2} \cup A_{0,3} \quad \text{and} \quad \bar{A}_k = A_k \cup A_{0,k}, \quad k = 1, 2, 3.$$

We assume in the following that $s \in \partial X$ (see Figure 7 (b)).

We also assume, without loss of generality, that $s \in c_1$. Then $r_1 = 0$. If we cut and open X along $d_1 \cup d_2$ then we obtain a 7-gon $P = (s_1 x y s_2 s_3 q p)$. We first claim that $\angle_P(s_1; s_2, s_3) < \pi$. Indeed, we denote by γ_i the geodesic loop in X based at s which is freely homotopic to the boundary c_i for $i = 1, 2, 3$. If $s \in c_1$ then $c_1 = \gamma_1$. Moreover if $\angle_P(s_1; s_2, s_3) \geq \pi$ then it follows that γ_1 is the juxtaposition of γ_2 and γ_3 , which is impossible. Now we distinguish the following cases:

Case $\partial_1 B_0$: If $\angle_P(s_2; s_1, s_3) < \pi$ and $\angle_P(s_3; s_1, s_2) < \pi$ then we define the corresponding subset $\partial_1 B_0$ of hyperbolic structures on X . At every hyperbolic structure in $\partial_1 B_0$ we may assign a unique configuration in \mathbb{H}^2 , say of type $\partial_1 B_0$, (see Figure 10 (a)) and conversely. Obviously the parameters r_i and l_i determine, up to isometry, each element of $\partial_1 B_0$.

Case $\partial_1 B_{0,2}$: $\angle_P(s_2; s_1, s_3) = \pi$ (see Figure 10 (b)) and we define the corresponding subset $\partial_1 B_{0,2}$ of hyperbolic structures on X .

Case $\partial_1 B_{0,3}$: $\angle_P(s_3; s_1, s_2) = \pi$ and we define also the subset $\partial_1 B_{0,3}$ of hyperbolic structures on X .

$$\text{We set } \overline{\partial_1 B_0} = \partial_1 B_0 \cup \partial_1 B_{0,2} \cup \partial_1 B_{0,3}.$$

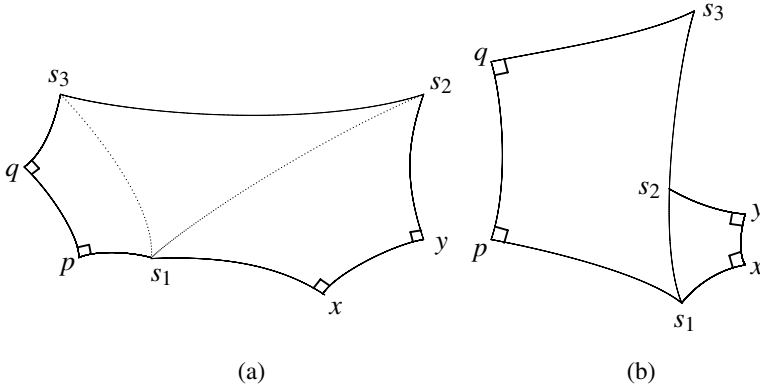


Figure 10

Case $\partial_1 B_2$: $\angle_P(s_2; s_1, s_3) > \pi$ (see Figure 11) and we define the corresponding subset $\partial_1 B_2$ of hyperbolic structures on X . We set $\overline{\partial_1 B_2} = \partial_1 B_2 \cup \partial_1 B_{0,2}$.

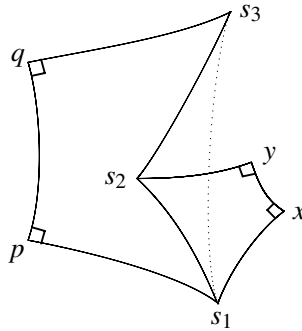


Figure 11

Case $\partial_1 B_3$: $\angle_P(s_3; s_1, s_2) > \pi$ and we also define the subset $\partial_1 B_3$ of hyperbolic structures on X . We set $\overline{\partial_1 B_3} = \partial_1 B_3 \cup \partial_1 B_{0,3}$.

If $s \in c_2$ we define sequentially the sets $\partial_2 B_0, \partial_2 B_{0,1}, \partial_2 B_{0,3}, \partial_2 B_1, \partial_2 B_3$ and if $s \in c_3$ we define the sets $\partial_3 B_0, \partial_3 B_{0,1}, \partial_3 B_{0,2}, \partial_3 B_1, \partial_3 B_2$. In the sequel we similarly define the sets $\overline{\partial_i B_j}, i = 2, 3, j = 0, 1, 2, 3$ and $i \neq j$.

Finally we set

$$\overline{\overline{A_0}} = \overline{A_0} \cup \overline{\partial_1 B_0} \cup \overline{\partial_2 B_0} \cup \overline{\partial_3 B_0}, \overline{\overline{A_k}} = \overline{A_k} \cup \overline{\partial_i B_k} \cup \overline{\partial_j B_k}, k = 1, 2, 3, i, j \in \{1, 2, 3\} - \{k\}.$$

From the analysis above we deduce that each hyperbolic structure is uniquely determined, up to isometry, from the parameters l_i and r_i .

On the other hand, given any six parameters $l_i, r_i, i = 1, 2, 3$ which determine a hyperbolic structure of type A_0 , we cannot deduce that the same l_i, r_i determine a

hyperbolic structure in A_k , $k = 1, 2, 3$. For example in the case A_1 , the lengths of $[s_i, s_j]$, which are determined by l_i, r_i , must satisfy certain compatibility conditions so that $s_1 \in \text{Int}(Q_2)$. Therefore, if $s \in \text{Int}(X)$, l_i, r_i determine at most four non-isometric hyperbolic structures, while if $s \in \partial X$, the parameters l_i, r_i determine at most three non-isometric hyperbolic structures. \square

Remark. Let X be a hyperbolic surface with one singularity s , which is homeomorphic to a disc with n holes, $n \geq 3$. We denote again by c_i the boundary components of X and by d_i the geodesic segments from s to c_i . If we set $l_i = \text{length}(c_i)$ and $r_i = \text{length}(d_i)$ then the parameters l_i, r_i do not characterize X in the sense of Proposition 3.8. This follows from the fact that if we cut and open X along $d_1 \cup \dots \cup d_n$ then instead of a triangle $(s_1 s_2 s_3)$ (see for example the case A_0 in Proposition 3.8) we get a hyperbolic n -gon which, of course, is not determined uniquely by the length of its edges.

In what follows, we introduce a set of real parameters which are defined independently of the position of the singularity s . We shall prove that these parameters determine, up to isometry, the surface X .

Let $\partial X = c_1 \cup c_2 \cup c_3$ and let d_i be a geodesic segment which realizes the distance between c_i and $c_{i+1(\text{mod } 3)}$, $i = 1, 2, 3$. It is not hard to prove, by using the CAT(-1) inequality, that d_i is unique for each i . Assume $d_1 \cap d_2 \cap d_3 \neq \{s\}$. Then if we cut and open X along $d = d_1 \cup d_2 \cup d_3$ we obtain two connected components P, Q . Each P, Q is a polygon isometrically embedded in \tilde{X} . We denote by a_i the subarc $P \cap c_i$ of c_i and by b_i the subarc $Q \cap c_i$. Set $l_i = \text{length}(c_i)$, $r_i = \text{length}(d_i)$ and $k_i = \text{length}(a_i)$. We remark that $r_i, i = 1, 2, 3$ are the distances between the sides of the polygon P .

Definition 3.9. The six non-negative parameters $r_i, k_i, i = 1, 2, 3$ will be called the *distance parameters* of X .

We have the following theorem:

Theorem 3.10. *The parameters $l_i, r_i, k_i, i = 1, 2, 3$ determine a unique, up to isometry, hyperbolic pair of pants X with one singularity.*

Proof. To prove the theorem we distinguish the following cases:

Case I: $k_1 \cdot k_2 \cdot k_3 \neq 0$. This implies that we can consider P and Q as hexagons isometrically embedded in \tilde{X} . From the study of the parameters l_i, r_i, k_i we can determine whether $s \in P$ and $s \notin Q$ or $s \in Q$ and $s \notin P$. Indeed, if $s \in P$ and $s \notin Q$ then Q is a right-angled hexagon of \mathbb{H}^2 . Therefore, if we set $k'_i = l_i - k_i$, the parameters k'_i, r_i satisfy certain compatibility conditions (see for example [2], Section 7.19, Theorem 7.19.2). Conversely, if k'_i, r_i satisfy these conditions we deduce that $s \in P$ and $s \notin Q$. Similarly we may study the case $s \in Q$ and $s \notin P$. In conclusion the study of parameters l_i, r_i, k_i allows one to tell whether $s \in P$ and $s \notin Q$ or $s \in Q$ and $s \notin P$ or $s \in P \cap Q$.

We assume now, without loss the generality, that $s \in P$ and $s \notin Q$ (similarly we treat the case $s \in Q$ and $s \notin P$). Hence P is an hexagon in \tilde{X} which has all its angles equal to $\frac{\pi}{2}$ and Q is a right-angled hexagon in \mathbb{H}^2 . Therefore we deduce that P, Q , and hence, the pair of pants X , is determined uniquely by the parameters l_i, r_i, k_i .

We assume finally that $s \in P \cap Q$. In this case s is a common vertex of P and Q . Then the angles $\angle(s, P)$ and $\angle(s, Q)$ are not necessarily right but are both smaller than 2π and so P and Q are realizable as 6-gons in \mathbb{H}^2 . Therefore, P, Q and hence X are uniquely determined by the parameters l_i, r_i, k_i and the theorem is proven in Case I.

Case II: $k_1 \cdot k_2 \neq 0, k_3 = 0$. If the singularity s lies in the interior of X , then the geodesic segments d_2 and d_3 have a common part in X (see Figure 12 (a)). Now the polygon P in \tilde{X} is a 5-gon $(x_1x_2sx_3x_4)$ which is realizable in \mathbb{H}^2 . In P we know that the angles at x_i for $i = 1, 2, 3, 4$ are right and we also know the distances $|x_1 - x_2|_P = k_1, |x_3 - x_4|_P = k_2, |x_1 - x_4|_P = r_1$. Therefore P is unique, up to isometry, and the distances $|s - x_2|_P$ and $|s - x_3|_P$ can be expressed as a function of k_1, k_2, r_1 (see Figure (12b)). This analysis determines whether s lies in the component c_3 of ∂X or, not. Indeed, when $s \in c_3$ then $|s - x_2|_P = r_2$ and $|s - x_3|_P = r_3$. Therefore, if r_2, r_3 satisfy certain compatibility conditions with k_1, k_2, r_1 then there exists a 5-gon in \mathbb{H}^2 with four right angles and whose sides have lengths k_1, k_2, r_1, r_2, r_3 . The existence of a such 5-gon implies that $s \in c_3$.

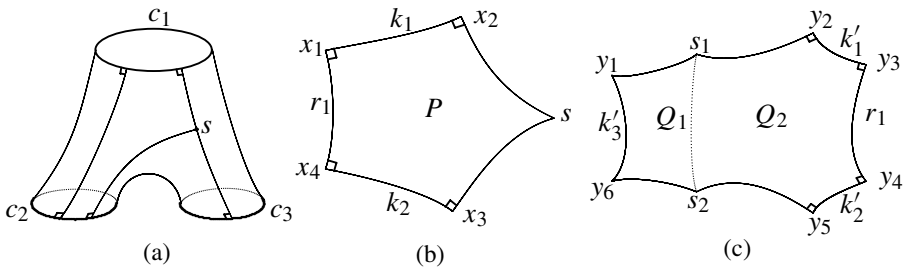


Figure 12

Therefore we distinguish two subcases:

First we assume that the relations between the parameters k_1, k_2, r_1, r_2, r_3 impose that $s \in \text{Int}(X)$ (see Figure 12). We have seen that P is realizable in \mathbb{H}^2 and it is unique. $Q = (y_1s_1y_2y_3y_4y_5s_2y_6)$ is also realizable in \mathbb{H}^2 . Indeed, the geodesic segment $[s_1, s_2]$ is contained in Q and it divides Q into two sub-polygons $Q_1 = (y_1s_1s_2y_6)$, $Q_2 = (s_1y_2y_3y_4y_5s_2)$, see Figure 12(c).

For Q_1 we know: $|y_1 - y_6|_{Q_1} = k'_3, \angle(y_1, Q_1) = \angle(y_6, Q_1) = \frac{\pi}{2}, |y_1 - s_1|_{Q_1} = r_3 - |x_2 - s|_P, |y_6 - s_2|_{Q_1} = r_2 - |x_3 - s|_P$.

Now the lengths $|x_2 - s|_P, |x_3 - s|_P$ are uniquely determined from the construction of P . Therefore Q_1 is uniquely determined up to isometry. Analogously, we show

that Q_2 is uniquely determined. Finally we conclude that X is uniquely determined by the parameters l_i, r_i, k_i .

Second we assume that the relations between the parameters k_1, k_2, r_1, r_2, r_3 impose that $s \in c_3 \subset \partial X$ (see Figure 13). Then P is a 5-gon and Q a 6-gon realizable in \mathbb{H}^2 and we proceed as above.

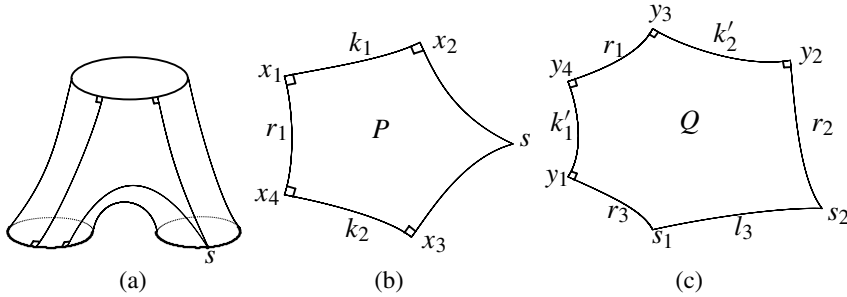


Figure 13

Finally we treat the cases $k_2 \cdot k_3 \neq 0, k_1 = 0$ and $k_1 \cdot k_3 \neq 0, k_2 = 0$ similarly. The case $k_i = k_j = 0$ and $k_m \neq 0$ with $i, j, m \in \{1, 2, 3\}$, cannot occur.

Case III: $k_1 = k_2 = k_3 = 0$. In this case we deduce that $d_1 \cap d_2 \cap d_3 = \{s\}$ and, therefore, we fall into Proposition 3.8. We distinguish two subcases:

Subcase III_a: $s \in \text{Int}(X)$. Then the angles formed by the geodesic segments $[s, x_1], [s, x_2], [s, x_3]$ at s are all greater than or equal to π . Thus the parameters l_i, r_i determine a hyperbolic structure of type A_0 (see Proposition 3.8), since in all other types of hyperbolic structures, we can verify easily, that the above mentioned angles around s are smaller than π .

Subcase III_b: $s \in \partial X$. Let $x_i = c_i \cap (d_1 \cup d_2 \cup d_3)$. Then from the relation that relates the parameters r_i we can deduce in which component of ∂X the singularity s belongs. For example if $r_2 = r_1 + r_3$ then $s \in c_1$. So we assume without loss of generality, that $s \in c_1$. Then the angle formed by the geodesics d_1, d_3 at s is greater than or equal to π . Therefore the parameters l_i, r_i determine a unique hyperbolic structure of type $\overline{\partial_1 B_0}$ since in all other types of hyperbolic structures, the angle at s formed by the segments d_1 and d_3 is smaller than π . □

3.3 The Teichmüller space of a hyperbolic pair of pants with a conical singularity

In this section we also assume that X is a hyperbolic pair of pants with one conical singularity, say s . We shall define the Teichmüller space $\mathcal{T}(X, s)$ of X and a parameter space \mathcal{B} for $\mathcal{T}(X, s)$.

We fix an orientation on X and let $\text{Homeo}^+(X, \partial)$ be the set of homeomorphisms of X which preserve the orientation and each boundary component of X . It is well

known that each element of $\text{Homeo}^+(X, \partial)$ is isotopic to the identity (see Exposé 2 in [9]). The space $\text{Homeo}^+(X, \partial)$ acts on $\mathcal{H}(X)$ as follows: If $f \in \text{Homeo}^+(X, \partial)$ and $\rho \in \mathcal{H}(X)$ then $(f, \rho) \rightarrow f^*\rho$ where $f^*\rho(x, y) = \rho(f(x), f(y))$.

Definition 3.11. We define the Teichmüller space $\mathcal{T}(X, s)$ of X with a conical singularity s as the quotient $\mathcal{H}(X)/\text{Homeo}^+(X, \partial)$.

Obviously $\mathcal{T}(X, s)$ consists of all hyperbolic structures which belong to one of the sets $\bar{A}_k, k = 0, 1, 2, 3$ defined in Proposition 3.8.

To build a parameter space for $\mathcal{T}(X, s)$ which will allow us to understand the topology of $\mathcal{T}(X, s)$ we proceed as follows:

We consider a convex subset \mathcal{C}_0 of \mathbb{R}^6 consisting of all $(a, b, c, r_1, r_2, r_3) \in \mathbb{R}^6$:

$$a, b, c > 0 \tag{1}$$

$$a \leq b + c \tag{2}$$

$$b \leq a + c \tag{3}$$

$$c \leq a + b \tag{4}$$

$$r_1, r_2, r_3 \geq 0 \tag{5}$$

$$r_1 + r_2 > 0 \tag{6}$$

$$r_2 + r_3 > 0 \tag{7}$$

$$r_1 + r_3 > 0. \tag{8}$$

Obviously \mathcal{C}_0 is a non-bounded convex subset of \mathbb{R}^6 . Let $\partial_a\mathcal{C}_0$ (resp. $\partial_b\mathcal{C}_0, \partial_c\mathcal{C}_0$) be the subset of \mathcal{C}_0 obtained by replacing inequality (2) (resp. (3), (4)) above with the equality $a = b + c$. The sets $\partial_a\mathcal{C}_0, \partial_b\mathcal{C}_0, \partial_c\mathcal{C}_0$ are pairwise disjoint subsets of the boundary $\partial\mathcal{C}_0$ of \mathcal{C}_0 . On the other hand, we define the subset $\partial_1\mathcal{C}_0$ (resp. $\partial_2\mathcal{C}_0, \partial_3\mathcal{C}_0$) of \mathcal{C}_0 which is defined by the equation $r_1 = 0$ (resp. $r_2 = 0, r_3 = 0$). The sets $\partial_1\mathcal{C}_0, \partial_2\mathcal{C}_0, \partial_3\mathcal{C}_0$ are also pairwise disjoint subsets of $\partial\mathcal{C}_0$. However each $\partial_i\mathcal{C}_0$ intersects $\partial_a\mathcal{C}_0$ and let $\partial_{i,a}\mathcal{C}_0 = \partial_a\mathcal{C}_0 \cap \partial_i\mathcal{C}_0$. Obviously each $\partial_{i,a}\mathcal{C}_0$ is homeomorphic to \mathbb{R}^4 and is a convex connected component of the boundary $\partial(\partial_a\mathcal{C}_0)$ of $\partial_a\mathcal{C}_0$. More precisely we have $\partial(\partial_a\mathcal{C}_0) = \partial_{1,a}\mathcal{C}_0 \cup \partial_{2,a}\mathcal{C}_0 \cup \partial_{3,a}\mathcal{C}_0$.

On the other hand, for each i , the set $\partial_{i,a}\mathcal{C}_0$ is a convex connected component of the boundary $\partial(\partial_i\mathcal{C}_0)$ of $\partial_i\mathcal{C}_0$. So if we define, in a similar fashion, the subsets $\partial_{i,b}\mathcal{C}_0 = \partial_b\mathcal{C}_0 \cap \partial_i\mathcal{C}_0$ and $\partial_{i,c}\mathcal{C}_0 = \partial_c\mathcal{C}_0 \cap \partial_i\mathcal{C}_0$, then we have that $\partial(\partial_i\mathcal{C}_0) = \partial_{i,a}\mathcal{C}_0 \cup \partial_{i,b}\mathcal{C}_0 \cup \partial_{i,c}\mathcal{C}_0$.

Now we consider three copies of \mathcal{C}_0 , say $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$. On the disjoint union $\mathcal{C}_0 \cup \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$ we identify \mathcal{C}_0 with \mathcal{C}_1 along $\partial_a\mathcal{C}$, \mathcal{C}_0 with \mathcal{C}_2 along $\partial_b\mathcal{C}$ and \mathcal{C}_0 with \mathcal{C}_3 along $\partial_c\mathcal{C}$. Let \mathcal{C} be the resulting space. From the convexity of the sets $\partial_a\mathcal{C}, \partial_b\mathcal{C}$ and $\partial_c\mathcal{C}$ it follows that \mathcal{C} is a non-compact manifold embedded in \mathbb{R}^6 , with three boundary components. The interior of \mathcal{C} is homeomorphic to \mathbb{R}^6 and each boundary component $\partial_i\mathcal{C}$ of $\mathcal{C}, i = 1, 2, 3$, is homeomorphic to \mathbb{R}^5 . We remark that the restriction of $\partial_i\mathcal{C}$ on every \mathcal{C}_i (considered as subset of \mathcal{C}) is defined by the relation $r_i = 0$.

Now we construct an embedding $\Phi: \mathcal{T}(X, s) \rightarrow \mathcal{C}$. To do this, we consider a hyperbolic structure $h \in \mathcal{T}(X, s)$. As remarked above, the hyperbolic structure h belongs to a set $\bar{\bar{A}}_k$, $k = 0, 1, 2, 3$. Therefore, to each $h \in \mathcal{T}(X, s)$ we may correspond to h a unique configuration of \mathbb{H}^2 , as it is described in Figures 8–11. Let $a = |s_2 - s_3|$, $b = |s_1 - s_3|$, $c = |s_1 - s_2|$, where $(s_1 s_2 s_3)$ is the triangle appearing in each configuration which determines h and (r_1, r_2, r_3) are the distances defined in Proposition 3.8; notice here that $(s_1 s_2 s_3)$ may be an imaginary triangle.

Now, we define a 1-1 mapping

$$\Phi_0: \bar{\bar{A}}_0 \rightarrow \mathcal{C}_0$$

such that to each $h \in \bar{\bar{A}}_0$ we assign the unique parameters (a, b, c, r_1, r_2, r_3) of \mathcal{C}_0 which determine h . Obviously we have the relations (*) below:

$$\begin{aligned} \Phi_0(A_{0,1}) \subset \partial_a \mathcal{C}_0, \quad \Phi_0(A_{0,2}) \subset \partial_b \mathcal{C}_0, \quad \Phi_0(A_{0,3}) \subset \partial_c \mathcal{C}_0 \\ \Phi_0(\partial_i \bar{B}_0) \subset \partial_i \mathcal{C}_0, \quad i = 1, 2, 3 \end{aligned} \tag{*}$$

We define similarly 1 – 1 mappings

$$\Phi_k: \bar{\bar{A}}_k \rightarrow \mathcal{C}_k, \quad k = 1, 2, 3$$

which satisfy the relations (**) below:

$$\begin{aligned} \Phi_1(A_{0,1}) \subset \partial_a \mathcal{C}_0, \quad \Phi_2(A_{0,2}) \subset \partial_b \mathcal{C}_0, \quad \Phi_3(A_{0,3}) \subset \partial_c \mathcal{C}_0 \\ \Phi_k(\partial_i \bar{B}_k) \subset \partial_i \mathcal{C}_k, \quad k \in \{1, 2, 3\}, \quad i \in \{1, 2, 3\} - \{k\}. \end{aligned} \tag{**}$$

From the relations (*) and (**) above we have that $\Phi_0|_{\bar{\bar{A}}_0 \cap \bar{\bar{A}}_0}$ is equal to $\Phi_k|_{\bar{\bar{A}}_0 \cap \bar{\bar{A}}_k}$, so Φ_0 and Φ_k fit properly and give a mapping $\Phi: \mathcal{T}(X, s) \rightarrow \mathcal{C}$ which is 1-1. We set $\Phi(\mathcal{T}(X, s)) = \mathcal{B}$.

Proposition 3.12. *The mapping Φ is continuous and open, thus $\Phi: \mathcal{T}(X, s) \rightarrow \mathcal{B}$ is a homeomorphism.*

Proof. We first we show that \mathcal{B} is an open subset of \mathcal{C} . Indeed, let

$$(a, b, c, r_1, r_2, r_3) = \Phi(h)$$

for some $h \in \mathcal{T}(X, s)$. As explained in Proposition 3.8, a hyperbolic pair of pants X equipped with the hyperbolic structure h , can be constructed by gluing either a triangle and three quadrilaterals of \mathbb{H}^2 or a 5-gon and two quadrilaterals of \mathbb{H}^2 . The lengths of the sides of these polygons take values in the set $\{a, b, c, r_1, r_2, r_3\}$. More precisely if $(a, b, c, r_1, r_2, r_3) \in \mathcal{C}_k$ then (a, b, c, r_1, r_2, r_3) determine a configuration of type $\bar{\bar{A}}_k$ and thus we construct (X, h) i.e. the hyperbolic structure h on X . It is straightforward to see that if we slightly change the lengths of the sides of the polygons that we glued in order to construct (X, h) , then with the new polygons we may construct a new hyperbolic structure h on X . Therefore there is an $\varepsilon > 0$ such

that if $(a', b', c', r'_1, r'_2, r'_3) \in \mathcal{C}$ and it is ε -near to (a, b, c, r_1, r_2, r_3) then there is $h' \in \mathcal{T}(X, s)$ such that $\Phi(h') = (a', b', c', r'_1, r'_2, r'_3)$.

Now, to finish the proof of the proposition it is sufficient to show that the mapping $\Phi^{-1}: \mathcal{B} \rightarrow \mathcal{T}(X, s)$ is continuous. But it is obvious that there is an $\varepsilon > 0$ such that if $(a', b', c', r'_1, r'_2, r'_3) \in \mathcal{B}$ is ε -near to $(a, b, c, r_1, r_2, r_3) \in \mathcal{B}$ then for every $x, y \in X$ the values $h(x, y)$ and $h'(x, y)$ are arbitrary close, where h, h' are the images of (a, b, c, r_1, r_2, r_3) and $(a', b', c', r'_1, r'_2, r'_3)$ by Φ^{-1} . \square

The following corollary follows immediately from the above proposition.

Corollary 3.13. $\mathcal{T}(X, s)$ is homeomorphic to a non-compact submanifold of \mathbb{R}^6 of dimension 6, with three boundary components.

We state now the main theorem of this chapter:

Theorem 3.14. $\mathcal{T}(X, s)$ as well as each boundary component of $\partial(\mathcal{T}(X, s))$ are contractible spaces.

For the proof of this theorem, which is quite long and technical, we construct step by step a homotopy $H: \mathcal{T}(X, s) \times [0, 1] \rightarrow \mathcal{T}(X, s)$ such that $H_{|\mathcal{T}(X, s) \times \{0\}} = \text{id}$ and $H_{|\mathcal{T}(X, s) \times \{1\}} = \text{constant map}$. For details, see [8].

4 The Teichmüller space of an ideal 2-dimensional simplicial complex

In this section we give the definition of 2-dimensional ideal simplicial complexes X which is slightly different from the corresponding definition of [6], we consider the Teichmüller space $\mathcal{T}(X)$ of X and we introduce the notion of *essentially simple closed curves* in X .

The main result is:

For each element $h \in \mathcal{T}(X)$ there exists a finite number of essentially simple closed geodesics in X whose lengths determine h .

4.1 Definitions and preliminaries

A 2-dimensional simplicial complex is a topological space X equipped with two (finite or infinite) sets \mathcal{C} and \mathcal{F} that satisfy the following properties:

(1) Each element $T \in \mathcal{C}$ is a *topological triangle*, that is, a topological space homeomorphic to a 2-dimensional closed disc with three distinguished distinct points on its boundary ∂T . The distinguished points are called the vertices of T . The closures in ∂T of the connected components of the complement of the vertices are called the *edges* of T .

(2) Each element $f \in \mathcal{F}$ is a homeomorphism $f: A \rightarrow B$, where A and B are distinct edges of two distinct triangles T and T' belonging to the collection \mathcal{C} . For each edge A of a triangle $T \in \mathcal{C}$, there exists an $f \in \mathcal{F}$ and an edge B of some triangle $T' \in \mathcal{C}$, with f sending A to B . The elements of \mathcal{F} are called *gluing maps*.

(3) As a topological space, the simplicial complex X is the quotient of the disjoint union of the elements of \mathcal{C} by the equivalence relation which identifies any two points x and y belonging respectively to edges A and B of triangles T and T' in \mathcal{C} , which are related by a map $f: A \rightarrow B$ in \mathcal{F} satisfying $f(x) = y$. Let $\pi: \mathcal{C} \rightarrow X$ denote the quotient map. We make the assumption that π is injective on each edge of a triangle in \mathcal{C} . The image by π of a triangle $T \in \mathcal{C}$ is called a *face* of X , the image by π of an edge of T is called an *edge* of X and the image by π of a vertex of T is called a *vertex* of X .

(4) Every two faces T_1, T_2 of X with $T_1 \cap T_2 \neq \emptyset$ intersect in either a single edge of X or a single vertex of X .

A face T of X with vertices x, y, z is denoted by (xyz) .

(5) The space X is arcwise connected.

(6) The space X is *locally finite*, which means that each edge and each vertex of X is contained in finitely many faces.

We shall say that a 2-dimensional simplicial complex is *finite* if the number of its faces is finite.

Let X be a 2-dimensional complex and let Λ be the set of edges of X . There is a subset $\Lambda' \subset \Lambda$, whose elements, called *singular edges*, are those edges of X whose inverse image by π consists of three or more distinct edges of triangles of \mathcal{C} . An edge of X which is in $\Lambda - \Lambda'$ is called a *non-singular edge*. The set of vertices of X will be denoted by S .

We shall be interested in metrics on $X' = X - S$ which are obtained in the following manner: for each topological triangle $T \in \mathcal{C}$, we delete the vertices of T and we equip this triangle with its vertices deleted with a metric which makes it isometric to an ideal hyperbolic triangle. In this way each edge of X with its vertices deleted is equipped with an induced metric and becomes isometric to the infinite line. We now take each map $f: A \rightarrow B$ in \mathcal{F} to be an isometry, with respect to these metrics. In this way, each face of X , with its vertices deleted, becomes isometric to an ideal hyperbolic triangle. The length $l_X(\gamma)$ of a path $\gamma: I \rightarrow X'$ is now defined as the sum of the lengths of the components of the intersection of γ with the faces of X . Let $h: X' \times X' \rightarrow \mathbb{R}$ be the function defined in the following way: for $(x, y) \in X' \times X'$, $h(x, y)$ is equal to the infimum of the set of lengths of paths in X' joining these points.

From Proposition 1.2 of [6] it follows that the function $h: X' \times X' \rightarrow \mathbb{R}$ is a metric on X' and the metric space (X', h) is a length space. Any such metric h on X' will be called an *ideal hyperbolic metric* on X . Furthermore, if h is complete, h is called an *ideal hyperbolic structure* on X . We also say that the space X' equipped with an ideal hyperbolic structure h is an *ideal 2-dimensional simplicial complex*. The faces of X' are ideal hyperbolic triangles and the edges of X' are the edges of these

ideal triangles. In Proposition 4.3 below we shall show that there always exist ideal hyperbolic structures on X .

In what follows, we shall always assume that the metric h on X' is complete. Although the metric h is defined on X' , we shall be speaking about a *metric* on X . By the term *geodesic* (resp. *local geodesic*) in X we mean a *geodesic* (resp. a *local geodesic*) in X' with respect to the metric h , see Paragraph 2.1 for the definitions.

Every ideal hyperbolic structure h on X is a geodesic metric since X' with h is a complete and locally compact length space (see [10], Chapter 1, Theorem 1.10). As in Paragraph 2.1, we write $[x, y]$ for a geodesic segment joining the points $x, y \in X'$.

The space X' is a local CAT(-1)-space. This follows from the fact that each point in X' has a neighborhood which is isometric to the union of a finite number of half-discs in \mathbb{H}^2 whose diameters are of equal length, the half discs being glued along their diameters, see Corollary 1.5 of [6].

The universal covering \tilde{X} of X inherits the structure of a 2-dimensional simplicial complex. Obviously, if we consider an ideal hyperbolic structure h on X then h can be lifted, via the covering projection $p: \tilde{X} \rightarrow X$, to an ideal hyperbolic structure on \tilde{X} , i.e. to a complete length metric \tilde{h} on $\tilde{X} - p^{-1}(S)$.

Definition 4.1. Let X be a 2-dimensional simplicial complex, \mathcal{C} the set of its triangles and \mathcal{F} the associated set of gluing maps. Two ideal hyperbolic structures h and h' on X are said to be equivalent if there exists a homeomorphism $F: X \rightarrow X$ which induces the identity map on the edges of X and which satisfies $F^*(h) = h'$, where by $F^*(h)$ we denote the pull-back of the metric h via F . We denote by $\mathcal{T}(X)$ the set of equivalence classes of ideal hyperbolic structures on X and we call this set the *Teichmüller space* of X .

If h is an ideal hyperbolic structure on X , we shall also denote by h its equivalence class in $\mathcal{T}(X)$.

Let T be an ideal hyperbolic triangle. Then T has a *distinguished* point which is the barycentre of T . Furthermore, each edge of T is equipped with a distinguished point, namely, the foot of the perpendicular drawn from the barycentre of T to that edge. We shall call this point the *centre* of the edge.

There are several ways of describing the topology of $\mathcal{T}(X)$, and we shall use here the *shift* parameters. Let X be a 2-dimensional simplicial complex equipped with an ideal hyperbolic structure h and let S be the set of vertices of X . In order to describe the shift parameters, we start by choosing once and for all an orientation on each edge of X . If T, T' are two faces of X with $T \cap T' \neq \emptyset$ then, by hypothesis, they have only one edge in common, say E . We shall use the same letters T, T' to indicate the corresponding ideal hyperbolic triangles in X' . We define the quantity $x_h(T, T')$ (or $x_h(T, T', E)$, if we want to specify the edge E) as the algebraic distance on E from the distinguished point associated to T to the distinguished point associated to T' , the sign being specified by the chosen orientation on E . We call $x_h(T, T')$ the *shift parameter* on the ordered pair (T, T') .

Let \mathcal{B} be the set of ordered pairs (T, T') where T, T' are triangles of X . The shift parameters define a map $\mathcal{I} : \mathcal{T}(X) \rightarrow \mathbb{R}^{\mathcal{B}}$, by the formula

$$\mathcal{I}(h)(T, T') = x_h(T, T').$$

The map \mathcal{I} is clearly injective, since each element of $\mathcal{T}X$ is determined by the associated set of gluing maps, which are isometries between edges of ideal hyperbolic triangles.

Equip \mathcal{B} with the discrete topology, $\mathbb{R}^{\mathcal{B}}$ with the product topology and $\mathcal{T}(X)$ with the topology induced from the embedding $\mathcal{I} : \mathcal{T}(X) \rightarrow \mathbb{R}^{\mathcal{B}}$.

Note finally that, for reasons explained in [7], the map $\mathcal{I} : \mathcal{T}(X) \rightarrow \mathbb{R}^{\mathcal{B}}$ is not onto.

Remark 4.2. In the definition of a simplicial complex above, we do not identify two edges belonging to a single triangle in \mathcal{C} and every two distinct triangles have at most one edge in common. This results from condition (4). Example 4 below shows that without this condition the lengths of essentially simple closed geodesics cannot determine a hyperbolic structure in $\mathcal{T}(X)$.

Example 4. We consider a space Z consisting of two ideal triangles T, T' whose edges are matched as shown in Figure 14. Let E be the edge of Z equipped with an orientation, see Figure 14. Obviously the ideal triangles T, T' have more than one common edge. It not difficult to show that $\mathcal{T}(Z)$ may be parametrized by the

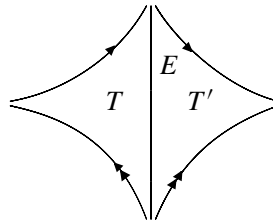


Figure 14

shift parameter $x_h(T, T', E)$. Indeed, the space Z has two cusps and it is obtained by performing three gluings along the sides of T and T' . So the dimension of $\mathcal{T}(Z)$ is 1 (see Theorem 3.3 in [6]). Now, in order to determine $x_h(T, T', E)$ we need the lengths of two closed geodesics; one which determines $|x_h(T, T', E)|$ and a second one which determines the sign of $x_h(T, T', E)$. But, it is easy to see that there exists only one simple closed curve in X which is not nullhomotopic. Therefore, there is only one essentially simple closed geodesic in X (see Definition 4.4 below).

4.2 The completeness of ideal hyperbolic structures

Let v be a vertex of X . The *link* $L(v)$ of v is a simplicial graph embedded in X , which is defined by taking one vertex on each edge of X ending on v and then joining two such vertices by an edge contained in a face of X if and only if these vertices correspond to edges which are on the boundary of that face. We note that if X is locally finite, the graph $L(v)$ is finite and it is easy to see that the link $L(v)$ is connected. We note also that if V is a small enough closed ball in X centered at v (with respect to some simplicial metric on X) then V has a natural structure of a geometric cone of the form $v \cdot L(v)$, that is, the topological quotient space $[0, 1] \times L(v)/R$, where R is the equivalence relation $(0, x) \sim (0, x')$ for all $x, x' \in L(v)$.

Let us fix a hyperbolic metric h on X . If v is a vertex of X , then v will be a vertex (at infinity) of finitely many ideal triangles of X' and we call v a *cuspid* of X' . Obviously, a “neighborhood” of the cusp v in X' is homeomorphic to $v \cdot L(v) - \{v\}$. We can associate a natural foliation on such a neighborhood of v , which is well defined up to restriction to a smaller neighborhood. The definition is as follows. Consider an ideal triangle in $X' = X - S$, having v as one of its ideal vertices. Consider a foliation of a horoball neighborhood of v in the ideal triangle, whose leaves are pieces of horocycles which are centered at the ideal vertex v . Considering now all ideal triangles of X' admitting v , as a vertex, we can glue together these foliations, and define a foliation of a neighborhood of the cusp v in X' . (Note that the foliation is singular with singular locus contained in the singular edges of X). Such a foliation is called a *horocyclic foliation* around v . From Proposition 3.4.18 in [17] it follows that: the metric space (X', h) is complete if and only if, for each cusp v of X' and for each horocyclic foliation of a neighborhood of v , the restriction of the foliation to a smaller neighborhood of v is a product foliation on a space of the form $L(v) \times [0, \infty)$, the leaves of which are the fibers $L(v) \times \{t\}$. Proposition 3.4.18 in [17] concerns the case where X' is a hyperbolic surface, but the case considered here follows in the same way.

Now, if for every two adjacent ideal triangles T, T' of X' , $x_h(T, T') = 0$, then h is obviously a complete metric i.e. an ideal hyperbolic structure. Therefore, we immediately deduce the following:

Proposition 4.3. *Every 2-dimensional simplicial complex X possesses an ideal hyperbolic structure.*

4.3 Parametrization of the Teichmüller space

In this paragraph we shall find real parameters for the Teichmüller space $\mathcal{T}(X)$ which are lengths of closed geodesics in X . These geodesics are not necessarily simple but each one contains in its free homotopy class a simple closed curve without back and forth.

For the rest of this paragraph we fix an ideal hyperbolic structure h on X .

We say that a closed curve $c: [0, 1] \rightarrow X'$ has a *back and forth* if there are $t_1, t_2 \in [0, 1]$, $t_1 < t_2$, and an ideal triangle T of X' such that $c(t_1), c(t_2)$ belong to the same edge of T and $c([t_1, t_2]) \subset T$.

Now, we give the following definition

Definition 4.4. Let γ be a closed geodesic in X . We say that γ is *essentially simple* if there exists a homotopy $h_t: [0, 1] \rightarrow X'$ sending γ to a simple closed curve c such that the closed curve h_t does not have back and forth for each value of the parameter t .

In Figure 15 (a) the geodesic γ_1 , which has no back and forth, can be homotoped to a simple curve c_1 without back and forth for all times t , while in Figure 15 (b) the geodesic γ_2 , which has no back and forth too, can be homotoped to a simple curve c_2 but the new curve c_2 has back and forth. So γ_1 is an essentially simple geodesic but not γ_2 .

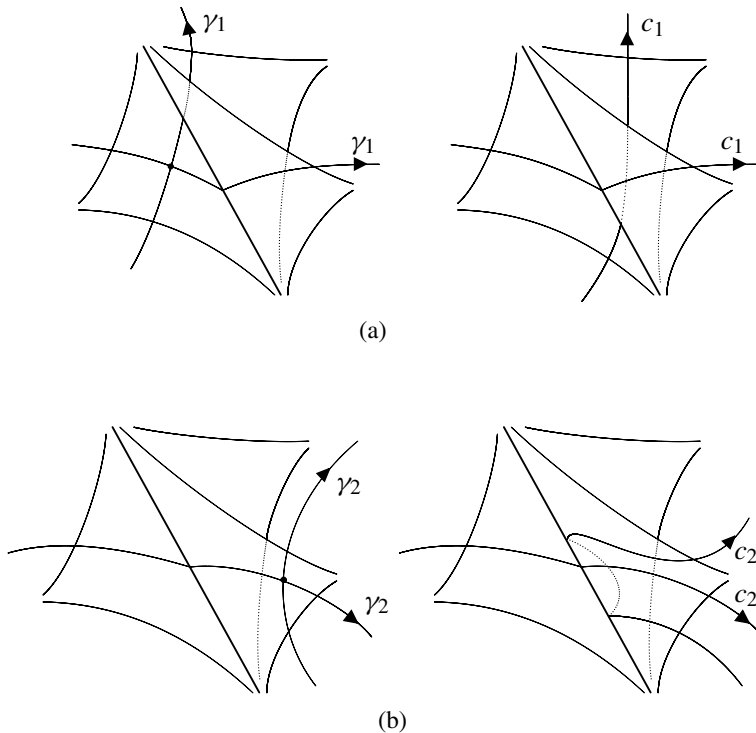


Figure 15

It is easy to see that if γ is an essentially simple closed geodesic, then the images $h_t([0, 1])$ remain in the same ideal triangles of X' as γ , for all t . Therefore, if there

exists a surface S embedded in X' with $\gamma \subset S$, for example this always happens if γ does not intersect a singular edge of X' , then γ is a simple closed geodesic. Indeed, if γ is not simple then we can find a simple closed curve c which is homotopic to γ in S . But this is impossible.

The main theorem of this section is the following:

Theorem 4.5. *The element h of $\mathcal{T}(X)$ can be determined by the lengths of a finite number of essentially simple closed geodesics in X .*

Proof. Let $T = (abc)$ and $T' = (acd)$ be two faces of X which share a common edge. The corresponding ideal triangles of X' are also denoted by T and T' . The edge of X' joining the cusps a and c is denoted by $E(a, c)$.

In order to prove the theorem it suffices to find a finite number of essentially simple closed geodesics of X whose lengths determine the shift parameter $x_h(T, T')$.

We may assume, without loss of generality, that each edge of X' is adjacent to at most three ideal triangles. Indeed, we denote by k_E the number of ideal triangles which are adjacent to an edge E and we call k_E the *index* of E . Assume E is an edge of index $k_E \geq 4$. Then there is an ideal simplicial complex Y (not necessarily connected) and two edges E_1, E_2 of Y such that

- $k_{E_1} < k_E, k_{E_2} < k_E$ and $k_{E_1} + k_{E_2} = k_E$,
- there exists an isometry $E_1 \rightarrow E_2$ such that X' is obtained from Y by identifying E_1 with E_2 via this isometry,
- T, T' belong to the same connected component of Y and share the common edge $E(a, c)$.

By repeated application of this procedure, we obtain an ideal simplicial complex Z which contains T and T' glued along $E(a, c)$ and the index of edges of Z is at most three.

Figure 16 displays how the edge E of X' gives rise to two edges E_1, E_2 in Y with $k_{E_1} = 2$ and $k_{E_2} = 3$.

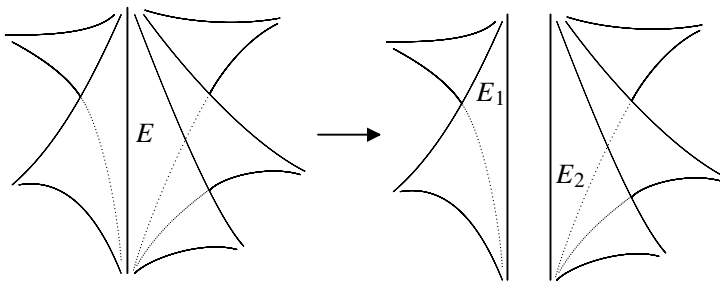


Figure 16

Since the metric space (X', h) is complete, there is a neighborhood V of the cusp a (resp. of the cusps b, c, d) in X' which is homeomorphic to $L(a) \times [0, \infty)$ and foliated by a horocyclic foliation whose leaves are the fibers $L(a) \times \{t\}$.

We pick an arbitrary leaf, say H_a , of this horocyclic foliation around a . H_a forms a closed graph whose edges are the pieces of horocycles lying in ideal triangles and the vertices are the intersection points of H_a with the edges of X' abutting at a . Let $a_b = H_a \cap E(a, b)$, $a_c = H_a \cap E(a, c)$, $a_d = H_a \cap E(a, d)$ and denote by $H[a_b, a_c]$ (resp. $H[a_c, a_d]$) the subarc of H_a contained in (abc) (resp. in (acd)) with endpoints a_b, a_c (resp. a_c, a_d), see Figure 17. With respect to the order of inclusion, we consider

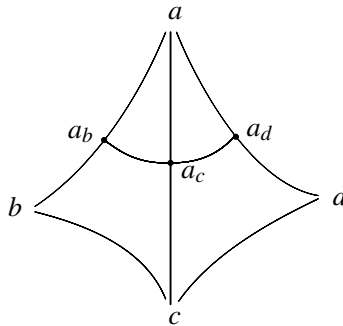


Figure 17

a minimal subgraph H_a^0 of H_a which contains $H[a_b, a_c]$ and $H[a_c, a_d]$. We have the following claim:

Claim 1. The graph H_a^0 is homeomorphic to either a circle or the configuration shown in Figure 18 (a); in the latter case the graph will be called of type \mathcal{G}_1 .

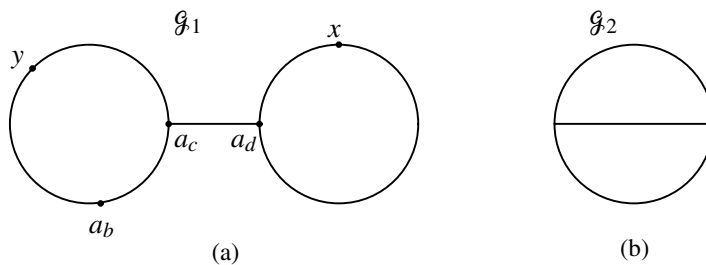


Figure 18

Proof of Claim 1. There exists an arc $\gamma_1 : [0, 1] \rightarrow H_a$ such that

- (i) γ_1 is 1-1 for each $t \in (0, 1)$;
- (ii) $\gamma_1(0) = a_b$ and $\gamma_1(1) \in H[a_b, a_c] \cup H[a_c, a_d]$;
- (iii) $\gamma_1([0, 1]) \cap (H[a_b, a_c] \cup H[a_c, a_d]) = \{\gamma_1(0), \gamma_1(1)\}$.

Obviously, $\gamma_1(1) \in \{a_b, a_c, a_d\}$. Therefore, either $\gamma_1(1) = \{a_d\}$ or, $\gamma_1(1) = \{a_c\}$ or, $\gamma_1(1) = \{a_b\}$. We denote these cases by (I₁), (I₂), (I₃) respectively.

Similarly, there exists an arc $\gamma_2: [0, 1] \rightarrow H_a$ such that

- (i) γ_2 is 1-1 for each $t \in (0, 1)$;
- (ii) $\gamma_2(0) = a_d$ and $\gamma_1(1) \in H[a_b, a_c] \cup H[a_c, a_d]$;
- (iii) $\gamma_2([0, 1]) \cap (H[a_b, a_c] \cup H[a_c, a_d]) = \{\gamma_2(0), \gamma_2(1)\}$.

Obviously, $\gamma_2(1) \in \{a_b, a_c, a_d\}$. Therefore, either $\gamma_2(1) = \{a_b\}$ or, $\gamma_2(1) = \{a_c\}$ or, $\gamma_2(1) = \{a_d\}$. We denote these cases by (II₁), (II₂), (II₃) respectively.

By combining these cases for the arcs γ_1 and γ_2 , and given that each vertex of H_a^0 is of index 3, we conclude the proof of the claim. More precisely: In the case (I₁) or (II₁) we may choose H_a^0 to be homeomorphic to \mathbb{S}^1 . In the cases (I₃) and (II₃) or, (I₃) and (II₂) or, (I₂) and (II₃) we may choose H_a^0 to be homeomorphic to \mathcal{G}_1 .

Now, we consider a closed curve $r_0: [0, 1] \rightarrow H_a^0$ such that

- (1) r_0 is locally 1-1 and onto;
- (2) there are points $t_1 < t_2$ in $[0, 1]$ such that $r_0(t_1) = a_b$, $r_0(t_2) = a_d$ and $r_0([t_1, t_2]) \cap H_a^0 = H[a_b, a_c] \cup H[a_c, a_d]$;
- (3) the number of times that r_0 passes through each vertex of H_a^0 is minimum among all curves $r: [0, 1] \rightarrow H_a^0$ satisfying properties (1) and (2) above.

For example, in Figure 18 (a) we consider H_a^0 to be of type \mathcal{G}_1 . If we choose a curve $r_0: [0, 1] \rightarrow H_a^0$ such that $r_0(0) = a_b$, $r_0(t_1) = a_c$, $r_0(t_2) = a_d$, $r_0(t_3) = x$, $r_0(t_4) = a_d$, $r_0(t_5) = a_c$, $r_0(t_6) = y$, $r_0(1) = a_b$, for $0 < t_1 < t_2 < t_3 < t_4 < t_5 < t_6 < 1$, then we can see that r_0 satisfies the requirements (1), (2), (3) above.

Now, for the curve r_0 , we consider a partition $0 = t_1 < t_2 < \dots < t_n = 1$ of $[0, 1]$ such that $r_0([t_i, t_{i+1}]) \subset T_i$ for all $i = 1, \dots, n - 1$, where T_i is an ideal triangle of X' and $r_0(t_i) \in E_i$, $r_0(t_{i+1}) \in E_{i+1}$, where E_i and E_{i+1} are edges of T_i . We shall say that the family $\{T_i\}$ covers the curve r_0 and we remark that the triangles T_i are not necessarily distinct. By taking the triangles T_i with an orientation on their sides induced from the orientation of the edges of X' , we glue them successively and we construct a surface Y_0 . More precisely, for each $i = 1, \dots, n - 1$, we glue T_i with T_{i+1} along their common side E_{i+1} with shift parameter equal to $x_h(T_i, T_{i+1})$ and T_n with T_1 along E_{n+1} with shift parameter equal to $x_h(T_n, T_1)$. In this way we build a new surface Y_0 with a cusp (Figure 19). All edges E_i of Y_0 are ending at this cusp, which is denoted again by a . Notice that (abc) and (acd) appear as ideal triangles of Y_0 . The surface Y_0 is called *the development of r_0 by the ideal triangles $\{T_i\}$* . The edge $E(a, b)$ of Y_0 is adjacent to (abc) and to another ideal triangle of Y_0 , say (abe) . Obviously, (abe) is one of the ideal triangles T_i which cover r_0 , so (abe) is also a face of X' .

For the cusp b of X' we work similarly considering the triangles (abc) , (abe) instead of the triangles (abc) and (acd) . This means that, we consider a horocyclic foliation around b , we choose a leaf H_b of it, as well as a minimal subgraph H_b^0 of H_b containing $H[b_a, b_c]$ and $H[b_a, b_e]$, where $H[b_a, b_c] = H_b \cap (abc)$ and $H[b_a, b_e] =$

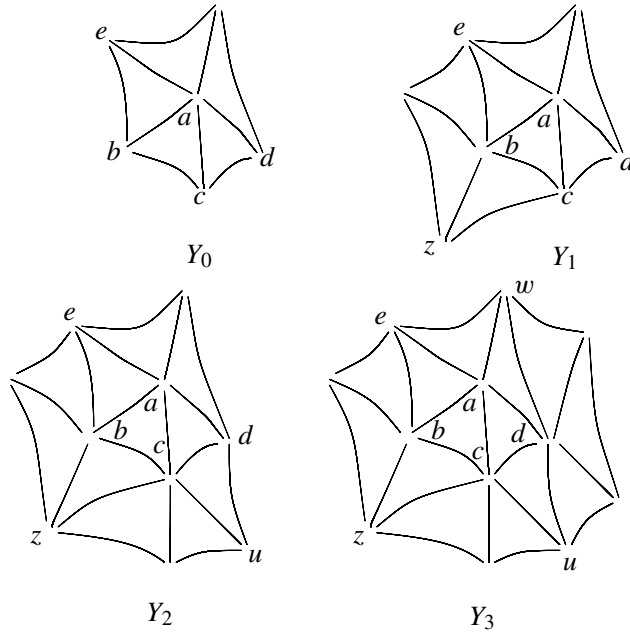


Figure 19

$H_b \cap (abe)$. Similarly to r_0 , we consider a curve $r_1: [0, 1] \rightarrow H_b^0$ satisfying properties (1), (2), (3) above and which is covered by a finite family of ideal triangles, say $\{T_i^1\}$, of X' . We “develop” further the surface Y_0 by gluing to Y_0 all the triangles T_i^1 . So, Y_1 contains Y_0 and Y_1 also contains the development of r_1 by the triangles $\{T_i^1\}$. In this way we build a surface Y_1 with two cusps a and b (Figure 19). Notice that every two adjacent ideal triangles of Y_1 have the same shift parameter as ideal triangles of X' . The edge $E(b, c)$ of Y_1 is adjacent to (abc) and to another ideal triangle of Y_1 , say (bcz) .

We consider now the cusp c of X' and a leaf H_c around c as before. We set $H[c_z, c_b] = H_c \cap (zcb)$, $H[c_b, c_a] = H_c \cap (cba)$, $H[c_d, c_a] = H_c \cap (acd)$ and we choose a minimal subgraph of H_c^0 of H_c containing $H[c_z, c_b]$, $H[c_b, c_a]$ and $H[c_d, c_a]$. We have the following claim:

Claim 2. The graph H_c^0 is homeomorphic to either a circle or to a graph of type \mathcal{G}_1 or to a graph of type \mathcal{G}_2 (see Figure 18).

The proof of Claim 2 is similar to the proof that of Claim 1. By combining all possible ways of going from point c_d to point c_z in H_c we form a subgraph which, if it is not a circle, it is either of type \mathcal{G}_1 or of type \mathcal{G}_2 .

Now, we consider a closed curve $r_2: [0, 1] \rightarrow H_c^0$ which satisfies properties (1) and (3) above, while property (2) is replaced by the following property (2').

(2') There are points $t_1 < t_2 < t_3 < t_4$ in $[0, 1]$ such that $r_2(t_1) = c_z, r_2(t_2) = c_b, r_2(t_3) = c_a, r_2(t_4) = c_d$ and $r_2((t_i, t_{i+1})) \cap \{c_z, c_b, c_d, c_a\} = \emptyset$ for each $i = 1, 2, 3, 4$.

As before, there exists a family $\{T_i^2\}$ of ideal triangles of X' which cover r_2 . So, we build a surface Y_2 which contains Y_1 and also contains the development of r_2 by the ideal triangles $\{T_i^2\}$. Therefore Y_2 is a new surface which contains the cusps a, b, c (Figure 19). The edges $E(c, d)$ and $E(a, d)$ of Y_2 which belong to the triangle (acd) are adjacent to two ideal triangles of Y_2 , say (cdu) and (adw) respectively.

Finally, by considering the cusp d of X' we repeat the same procedure with d in place of c . That is, we consider the three pieces of horocycle $H[d_u, d_c] = H_d \cap (duc), H[d_c, d_a] = H_d \cap (dca), H[d_a, d_w] = H_d \cap (daw)$, where H_d is a leaf of the horocyclic foliation around d , we next consider the minimal subgraph H_d^0 of H_d which contains $H[d_u, d_c], H[d_c, d_a], H[d_a, d_w]$ and so on. In this way we construct a surface Y_3 which contains the cusps a, b, c, d (Figure 19).

From the construction of Y_3 it follows that there exists a natural projection $\pi : Y_3 \rightarrow X'$ which is a local isometry in the sense that for every $y \in Y_3$ there exists a neighborhood U of y in Y_3 which is mapped isometrically by π onto a subset of X' . It is immediate to see that for any curve $\rho : [0, 1] \rightarrow \pi(Y_3)$ there exists a lifting $r : [0, 1] \rightarrow Y_3$ of ρ , i.e. $\pi \circ r = \rho$.

Now we shall prove that the shift parameters of all pairs of ideal triangles which are glued in Y_3 , and hence $x_h(T, T')$, are determined by the lengths of simple closed geodesics of Y_3 which are projected to essentially simple geodesics in X' .

We need the following

Claim 3. In a neighborhood of the cusp a (resp. of b, c, d) of Y_3 we may find a simple closed curve r_a (resp. r_b, r_c, r_d) (see Figure 20) such that the following holds:

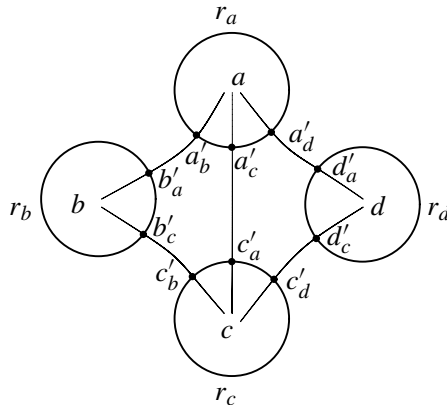


Figure 20

- r_a, r_b, r_c, r_d are projected onto a simple closed curve of X' .
- If we set $a'_b = r_a \cap E(a, b), a'_c = r_a \cap E(a, c), a'_d = r_a \cap E(a, d)$ and define, similarly, the points $b'_a, b'_c \in r_b, c'_a, c'_b, c'_d \in r_c$ and $d'_a, d'_c \in r_d$, then the

projection $\pi : Y_3 \rightarrow X'$ is 1-1 on the graph $G = r_a \cup r_b \cup r_c \cup r_d \cup E[a'_b, b'_a] \cup E[b'_c, c'_b] \cup E[a'_c, c'_a] \cup E[a'_d, d'_a] \cup E[d'_c, c'_d]$, where $E[a'_b, b'_a] \subset E(a, b)$ is the arc with endpoints a'_b, b'_a . Similarly the arcs $E[b'_c, c'_b]$, $E[a'_c, c'_a]$, $E[a'_d, d'_a]$, $E[d'_c, c'_d]$ in the corresponding edges of Y_3 are defined.

Proof of Claim 3. If the graph H_x^0 for $x \in \{a, b, c, d\}$ is homeomorphic to a circle, then there are neighborhoods of the cusp x in Y_3 and X' respectively, which are isometric via π . Therefore, we may choose an arbitrary simple closed curve r_x in a neighborhood of x in Y_3 .

If the graph H_x^0 is of type \mathcal{G}_1 for $x \in \{a, b, c, d\}$ or, if H_x^0 is of type \mathcal{G}_2 for $x \in \{c, d\}$, then we choose a simple closed curve ρ_x in a neighborhood of the cusp x in X' , and we denote by r_x a lifting of ρ_x in Y_3 . Obviously r_x is a simple closed curve. Furthermore, we must choose ρ_x , and hence r_x , properly so that the projection $\pi : Y_3 \rightarrow X'$ restricted on the graph G is 1-1.

Without loss of generality, we assume that H_x^0 is of type \mathcal{G}_1 for $x = a$ and H_x^0 is of type \mathcal{G}_2 for $x = c$. We consider points a_b, a_c, a_d placed on H_a^0 and points c_a, c_b, c_d , placed on H_c^0 . Let ρ_a and ρ_c be simple closed curves contained in a neighborhood of cusps a and c of X' respectively.

For $y \in \{b, c, d\}$, if the edge $E(a, y)$ intersects ρ_a in more than one point then we define the points $a''_y = E(a, y) \cap \rho_a$ so that $E[a''_y, y] \cap \rho_a = a''_y$, where $E[a''_y, y]$ is the subset of $E(a, y)$ from a''_y to y which is homeomorphic to the interval $[0, +\infty)$. We set $a'_y = \pi^{-1}(a''_y) \cap r_a$. Similarly we define $c''_y = E(c, y) \cap \rho_c$ and $c'_y = \pi^{-1}(c''_y) \cap r_c$ where $y \in \{a, b, d\}$. Now, with this choice of curves and points, the graph G is defined and we may easily verify that $\pi : G \rightarrow X'$ is 1-1. In Figures 21 and 22 we indicate how to choose the curves ρ_a and ρ_c as well as the points a''_y and c''_y in two typical cases.

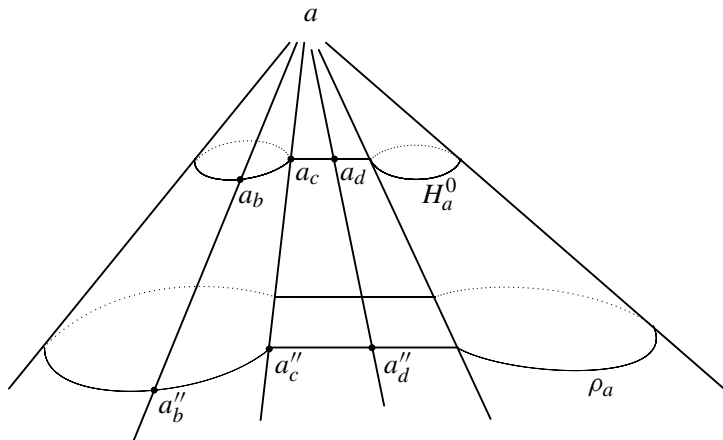


Figure 21

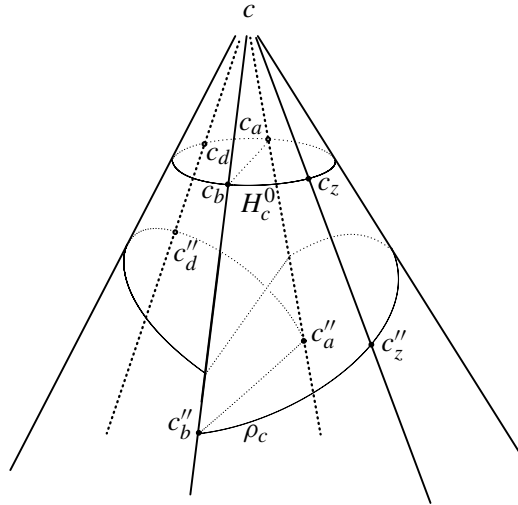


Figure 22

Now, all the simple closed geodesics in Y_3 which are used to determine the shift parameter $\chi_h(T, T')$, can be homotoped to simple closed curves which do not have back and forth and which lie in an arbitrarily small neighborhood \mathcal{U} of graph G . A number of such geodesics are drawn in bold in Figure 23. On the other hand,

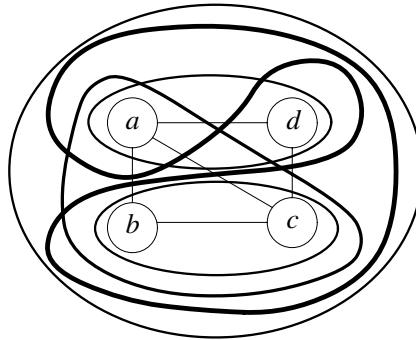


Figure 23

these geodesics project to essentially simple closed geodesics in X' because they are homotopic in Y_3 to simple closed curves lying in \mathcal{U} . This proves the theorem. \square

References

- [1] E. Ghys et P. de La Harpe (eds.), *Sur les groupes hyperboliques d'après M. Gromov*. Progr. Math. 83, Birkhäuser, Boston, MA, 1990. [478](#)
- [2] A. F. Beardon, *The geometry of discrete groups*. Grad. Texts in Math. 91, Springer-Verlag, New York 1983. [488](#)
- [3] M. Bridson, A. Haefliger, *Metric spaces of non-positive curvature*. Grundlehren Math. Wiss. 319, Springer-Verlag, Berlin 1999. [477](#)
- [4] P. Buser, *Geometry and spectra of compact Riemann surfaces*. Progr. Math. 106, Birkhäuser, Boston, MA, 1992. [472](#)
- [5] M. Coornaert, T. Delzant, A. Papadopoulos, *Géométrie et théorie des groupes*. Lecture Notes in Math. 1441, Springer-Verlag, Berlin 1990. [477](#), [478](#)
- [6] C. Charitos, A. Papadopoulos, The geometry of ideal 2-dimensional simplicial complexes. *Glasgow Math. J.* 43 (2001), 39–66. [493](#), [494](#), [495](#), [496](#)
- [7] C. Charitos, A. Papadopoulos, Hyperbolic structures and measured foliations on 2-dimensional complexes. *Monatsh. Math.* 139 (2003), 1–17. [496](#)
- [8] C. Charitos, I. Papadoperakis, On the geometry of hyperbolic surfaces with a conical singularity. *Ann. Global Anal. Geom.* 23 (2003), 323–357. [476](#), [493](#)
- [9] A. Fathi, F. Laudenbach, and V. Poénaru (eds.), *Travaux de Thurston sur les surfaces*. *Astérisque* 66–67 (1979). [472](#), [474](#), [475](#), [491](#)
- [10] M. Gromov, *Structures métriques pour les variétés riemanniennes*. Rédigé par J. Lafontaine et P. Pansu, Fernand Nathan, Paris 1981. [495](#)
- [11] M. Gromov, Hyperbolic groups. In *Essays in group theory* (S. Gersten, ed.), Math. Sci. Res. Inst. Publ. 8, Springer-Verlag, New York 1987, 75–263. [478](#)
- [12] U. Hamenstädt, Parametrizations of Teichmüller space and its Thurston boundary. In *Geometric analysis and nonlinear partial differential equations*, Springer-Verlag, Berlin 2003, 81–88. [476](#)
- [13] R. McOwen, Point singularities and conformal metrics on Riemann Surfaces. *Proc. Amer. Math. Soc.* 103 (1) (1988), 222–224. [478](#)
- [14] R. McOwen, Conformal metrics with singularities and finite negative total curvature on Riemann surfaces. In *Geometry and nonlinear partial differential equations*, Contemp. Math. 127, Amer. Math. Soc., Providence, RI, 1992, 101–107. [478](#)
- [15] F. Paulin, Constructions of hyperbolic groups via hyperbolization of polyhedra. In *Group theory from a geometrical viewpoint* ((E. Ghys, A. Haefliger, A. Verjovsky, eds.), ICTP (Trieste, Italy, 1990), World Scientific Publishing Co., Inc., River Edge, NJ, 1991. [476](#)
- [16] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, Grad. Texts Math., Springer-Verlag, 1994. [472](#), [482](#)
- [17] W. P. Thurston, *Three-dimensional geometry and topology*. Princeton Math. Ser. 35, Princeton University Press, Princeton, NJ, 1997. [497](#)
- [18] M. Troyanov, Prescribing curvature on compact surfaces with conical singularities. *Trans. Amer. Math. Soc.* 324 (1991), 793–821. [478](#)

Chapter 12

On the moduli space of singular euclidean surfaces

Marc Troyanov

Contents

1	Introduction	508
2	Piecewise flat surfaces	510
	2.1 Euclidean triangulation on a surface	510
	2.2 The universal branched cover of a piecewise flat surface	511
	2.3 The developing map of a piecewise flat surface	512
	2.4 The holonomy of a piecewise flat surface	514
	2.5 The developing map near a singularity	515
	2.6 Geometric equivalence of euclidean triangulations	516
	2.7 Flat metrics with conical singularities	516
	2.8 Relation with Riemann surfaces	517
3	Punctured surfaces	518
	3.1 Punctured surfaces and their fundamental groups	518
	3.2 Uniformization of a punctured Riemann surface	519
	3.3 Some groups of diffeomorphisms of a punctured surface	521
	3.4 Outer automorphisms	521
	3.5 Lifting the group $\text{Diff}^0(\Sigma')$ on \mathbb{U}	522
4	The representation variety of a finitely generated group in $\text{SE}(2)$	524
	4.1 On the cohomology of groups	524
	4.2 Abelian representations	526
	4.3 Representations in $\text{SE}(2)$	527
	4.4 Conjugation by similarities	529
5	Deformation theory	530
	5.1 The moduli and Teichmüller spaces	530
	5.2 The deformation space of piecewise flat metrics	532
	5.3 Revisiting the developing map and the holonomy	532
6	The main theorem	535
	6.1 The case of the sphere	537
	References	538

1 Introduction

The Teichmüller space of a punctured surface is the space of hyperbolic metrics with cusps up to isotopy on that surface, however, it can also be seen as the space of flat metrics with conical singularities of prescribed angles at the punctures up to isotopy and rescaling. The aim of the present chapter is to use this fact and show how the theory of piecewise flat surfaces and their deformations leads to new geometric structures in Teichmüller theory.

In the first section, which is rather elementary, we describe the geometry of piecewise flat surfaces. The second section describes the topology of punctured surfaces and their diffeomorphism groups. In the third section we discuss the representation space of a finitely generated group π into the group $SE(2)$ of rigid motions in the euclidean plane. In the last two sections, we apply the previous results to construct a new geometric structure on the Riemann moduli space $\mathcal{M}_{g,n}$ of a surface Σ of genus g with n punctures. More specifically, we show that this moduli space is a good orbifold¹ which admits a family of geometric structures locally modeled on the homogeneous spaces $\Xi = \mathbb{T}^{2g} \times \mathbb{C}\mathbb{P}^{2g+n-3}$.

We now discuss our main result. We first define a *punctured surface* $\Sigma_{g,n}$ of type (g, n) to be a fixed connected closed orientable surface S of genus g together with a distinguished set of n pairwise distinct points $p_1, p_2, \dots, p_n \in \Sigma_{g,n}$. The *Teichmüller space* $\mathcal{T}_{g,n}$ of $\Sigma_{g,n}$ is the set of conformal structures on $\Sigma_{g,n}$ modulo isotopies fixing the punctures (see Section 5 for a precise definition). This space is a real analytic variety in a natural way; if $2g - 2 + n > 0$, then it is isomorphic to $\mathbb{R}^{6g-6+2n}$. The group of isotopy classes of orientation-preserving diffeomorphisms of $\Sigma_{g,n}$ fixing the punctures is called the *pure mapping class group* and denoted by $\text{PMod}_{g,n}$. It acts in a natural way on the Teichmüller space $\mathcal{T}_{g,n}$.

We are now in a position to state the main result:

Theorem. *Given a punctured surface $\Sigma_{g,n}$ such that $2g + n - 2 > 0$, we can construct a group homomorphism*

$$\Phi: \text{PMod}_{g,n} \rightarrow \mathcal{G} = \text{Aut}(\mathbb{T}^{2g}) \times \text{PGL}_{2g+n-2}(\mathbb{C})$$

and a Φ -equivariant local homeomorphism

$$\mathcal{H}: \mathcal{T}_{g,n} \rightarrow \Xi = \mathbb{T}^{2g} \times \mathbb{C}\mathbb{P}^{2g+n-3}.$$

To say that \mathcal{H} is Φ -equivariant means that $\mathcal{H}(A\mu) = \Phi(A) \cdot \mathcal{H}(\mu)$ for all $A \in \text{PMod}_{g,n}$ and $\mu \in \mathcal{T}_{g,n}$.

The pair (\mathcal{H}, Φ) depends on n parameters $\beta_1, \beta_2, \dots, \beta_n \in (-1, \infty)$ such that $\sum_{j=1}^n \beta_j = 2g - 2$ and no $\beta_i \in \mathbb{Z}$.

¹Recall that an *orbifold* is a space which is locally the quotient of a manifold by a finite group. A *good orbifold* is globally the quotient of a manifold by a group acting properly and discontinuously (but in general not freely).

The *moduli space* $\mathcal{M}_{g,n}$ of $\Sigma_{g,n}$ is the set of conformal structures on $\Sigma_{g,n}$ modulo diffeomorphisms fixing the punctures. It is the quotient of the Teichmüller space by the pure mapping class group of $\Sigma_{g,n}$; in other words $\mathcal{M}_{g,n}$ is a good orbifold whose universal cover is $\mathcal{T}_{g,n}$ and fundamental group is $\text{PMod}_{g,n}$. In the geometric language of (G, X) -structures on manifolds and orbifolds (see [7], [9], [14], [29]), this theorem says that *we have constructed a family of geometric structures on the orbifold $\mathcal{M}_{g,n}$ which is modeled on the homogeneous space $\Xi = \mathbb{T}^{2g} \times \mathbb{C}\mathbb{P}^{2g+n-3}$. This family is parametrized by the β_i 's.*

The composition of the map \mathcal{H} in the theorem with the projection on the torus \mathbb{T}^{2g} gives us a map $\rho: \mathcal{T}_{g,n} \rightarrow \mathbb{T}^{2g}$ called the *character map*. It was proved by W. A. Veech that this map is a real analytic submersion. Its fibers describe a foliation whose leaves carry a geometric structure locally modelled on the complex projective space $\mathbb{C}\mathbb{P}^{2g+n-3}$, see [35] for proofs of these facts and a discussion of other related geometric structures on $\mathcal{T}_{g,n}$.

The proof of this theorem is based on the following strategy: we first show that the Teichmüller space can be seen as a deformation space of *flat metrics* on $\Sigma_{g,n}$ having *conical singularities* of prescribed angles at the punctures. We associate to such a metric a homomorphism, called the *holonomy* of the metric, from the fundamental group of the surface to the group $\text{SE}(2)$ of direct isometries of the euclidean plane. We then show that such a homomorphism can be seen as a point in the variety Ξ . In brief, $\mathcal{H}: \mathcal{T}_{g,n} \rightarrow \Xi$ maps the isotopy class of a singular flat metric to its holonomy representation.

In the special case of the punctured sphere, a stronger form of this theorem has been obtained by Deligne and Mostow [10] using some techniques from algebraic geometry and by Thurston [30] using an approach closer to ours.

To conclude this introduction, let us stress that the importance of piecewise flat metrics in Teichmüller theory is illustrated by the large number of papers dedicated to this subject. In addition to the work of Veech and Thurston already quoted, let us mention the contributions of Rivin [27], Bowditch [5], Epstein and Penner [11] to name a few. Piecewise euclidean metrics also appear in quantum gravity and in topological quantum field theory, see [2], [8] and the references therein. Although the present chapter starts with elementary considerations, the reader ought not to consider it as a global survey of this vast subject.

Acknowledgments. I would like to thank Babak Modami and François Fillastre for having carefully read the manuscript and for their comments. Finally, this exposition is dedicated to the memory of Michel Matthey.

2 Piecewise flat surfaces

2.1 Euclidean triangulation on a surface

A *piecewise flat surface* is a metric space obtained by properly gluing a stock of euclidean triangles in such a way that whenever two triangles meet along an edge, they are glued by an isometry along that edge. More precisely:

Definition 2.1. A *euclidean triangulation* of a surface Σ is a set of pairs $\mathcal{T} = \{(T_\alpha, f_\alpha)\}_{\alpha \in A}$ where each T_α is a compact subset of Σ and $f_\alpha: T_\alpha \rightarrow \mathbb{R}^2$ is a homeomorphism onto a non degenerate triangle $f_\alpha(T_\alpha)$ in the euclidean plane \mathbb{R}^2 . A subset e of T_α is an *edge* if $f_\alpha(e)$ is an edge of $f_\alpha(T_\alpha)$ and a point p of T_α is a *vertex* if its image under f_α is a vertex of $f_\alpha(T_\alpha)$.

The euclidean triangulation \mathcal{T} is subject to the following conditions:

- (i) The triangles cover the surface: $\Sigma = \bigcup_\alpha T_\alpha$.
- (ii) If $\alpha \neq \beta$, then the intersection $T_\alpha \cap T_\beta$ is either empty, or an edge or a vertex.
- (iii) If $T_\alpha \cap T_\beta \neq \emptyset$, then there is an element $g_{\alpha\beta} \in E(2)$ (the group of isometries of the euclidean plane) such that $f_\alpha = g_{\alpha\beta} f_\beta$ on that intersection.

An element $(T_\alpha, f_\alpha) \in \mathcal{T}$ is called a *triangle* or a *2-simplex* of the triangulations, we often just denote it by T_α . The vertices and edges are called 0- and 1-simplices respectively.

Two euclidean triangulations $\mathcal{T} = \{(T_\alpha, f_\alpha)\}_{\alpha \in A}$ and $\mathcal{T}' = \{(T_\alpha, f'_\alpha)\}_{\alpha \in A}$ of the same surface Σ are considered to be equal if they have the same simplices and, for any $\alpha \in A$, there is an isometry $g_\alpha \in E(2)$ such that $f'_\alpha = g_\alpha f_\alpha$.

Definition 2.2. A *piecewise flat surface* (Σ, \mathcal{T}) is a surface together with a euclidean triangulation.

A piecewise flat surface (Σ, \mathcal{T}) comes with a number of additional structures. In particular there is a well defined *area measure* which coincides with the 2-dimensional Lebesgue measure on each euclidean triangle T . We can also define the length $\ell(c)$ of an arbitrary curve $c: [0, 1] \rightarrow \Sigma$ by the following axioms:

- (i) if c is contained in a triangle T of \mathcal{T} , then $\ell(c)$ is the euclidean length.
- (ii) ℓ is additive: if c is the concatenation of two curves $c_1 c_2$, then $\ell(c) = \ell(c_1) + \ell(c_2)$.

The piecewise flat surface is thus a length space (see [6] for this notion). If the surface is connected, then it is also a metric space for the distance given by the infimum of the lengths of all curves joining two given points.

There is one more structure, called the *singularity order* and which is defined as the angular excess at the vertices counted in number of turns. It tells us how singular each vertex is compared to an ordinary point; the precise definition is the following:

Definition 2.3. The vertex $p \in \Sigma$ is said to be a *conical point of total angle* θ if

$$\theta = \sum_{j=1}^k \varphi_j$$

where $\varphi_1, \dots, \varphi_k$ are the angles of all the triangles in \mathcal{T} which are incident to p . The *singularity order* $\beta(p)$ of a vertex p is the angular excess at p measured in number of turns:

$$\beta(p) = \frac{\theta}{2\pi} - 1.$$

We extend the function β to all points of Σ by setting $\beta(x) = 0$ if $x \in \Sigma$ is not a vertex. The point x is then termed *singular* if $\beta(x) \neq 0$ (i.e. if $\theta \neq 2\pi$) and *regular* otherwise.

Proposition 2.4 (Gauss–Bonnet formula). *For any euclidean triangulation on a compact surface without boundary Σ , we have*

$$\chi(\Sigma) + \sum_{x \in \Sigma} \beta(x) = 0, \quad (2.1)$$

where $\chi(\Sigma)$ is the Euler characteristic of the surface.

The proof is a direct counting argument based on the definition of the Euler characteristic and the fact that the three internal angles of a euclidean triangle add up to π , see [31].

2.2 The universal branched cover of a piecewise flat surface

If (Σ, \mathcal{T}) is a piecewise flat surface, we denote by $\Sigma' = \Sigma \setminus \{p_1, \dots, p_n\}$ the open surface obtained by removing the singular vertices $p_1, \dots, p_n \in \Sigma$.

Definition 2.5. A path $c: [0, 1] \rightarrow \Sigma$ is *admissible* if it has finitely many intersections with the edges of the triangulation and if $c(s) \in \Sigma'$ for any $0 < s < 1$. A homotopy c_t is an *admissible homotopy* if $s \mapsto c_t(s) \in \Sigma$ is an admissible path for any $0 \leq t \leq 1$.

Let us choose a fixed triangle $T_0 \in \mathcal{T}$ and call it *home* (or the *base triangle*). We also choose a base point x_0 in the interior of T_0 .

Definition 2.6. The *universal branched cover* of (Σ, \mathcal{T}) is the euclidean two-dimensional complex $\hat{\mathcal{T}}$ obtained as follows: a k -simplex $\hat{\sigma}$ of $\hat{\mathcal{T}}$, where $k = 0, 1$ or 2 , is a pair $(\sigma, [c])$ where σ is a k -simplex of \mathcal{T} and $[c]$ is an admissible homotopy class of paths joining T_0 to a point in σ .

The universal branched cover $\hat{\mathcal{T}}$ is a simplicial complex (which is not locally finite) and there is an obvious simplicial map $\hat{\mathcal{T}} \rightarrow \mathcal{T}$ sending $(\sigma, [c])$ to σ .

We denote by $\hat{\Sigma}$ the geometric realization of $\hat{\mathcal{T}}$. This is a triangulated topological space and it comes with a continuous surjective map $P: \hat{\Sigma} \rightarrow \Sigma$ sending each simplex of $\hat{\mathcal{T}}$ homeomorphically onto the corresponding simplex in \mathcal{T} . We turn $\hat{\Sigma}$ into a metric space (in fact a length space) by requiring P to be an isometry on each simplex (concretely, we give to each simplex $\hat{\sigma} = (\sigma, [c])$ in $\hat{\mathcal{T}}$ the geometry of the euclidean simplex σ in \mathcal{T}).

Another way to understand $\hat{\Sigma}$ is the following: let $\tilde{\Sigma}'$ be the universal cover of Σ' . It is naturally a length space (in fact a flat Riemannian surface) and $\hat{\Sigma}$ is its metric completion.

2.3 The developing map of a piecewise flat surface

Definition 2.7. An edge of the piecewise flat surface (Σ, \mathcal{T}) is said to be *interior* if it is not contained in the boundary of Σ . The *hinge* of an interior edge e is the unique pair of triangles $T_1, T_2 \in \mathcal{T}$ which are incident with e .

Given an interior edge e with hinge (T_1, T_2) and an isometry $f_1: T_1 \rightarrow \mathbb{R}^2$, there exists a unique isometry $f_2: T_2 \rightarrow \mathbb{R}^2$ such that $f_1(T_1)$ and $f_2(T_2)$ have disjoint interiors and $f_1(e) = f_2(e)$. By juxtaposing these maps, we obtain a map

$$f_e = f_1 \cup f_2: T_1 \cup T_2 \rightarrow \mathbb{R}^2$$

which is an isometry of the hinge onto a quadrilateral in the euclidean plane. The map f_e just described is called an *unfolding* of the hinge. One also says that f_2 is the *continuation* of f_1 across the edge e .

The notions of hinge, unfolding and continuation of an isometry across an edge are similarly defined on the universal branched cover $\hat{\Sigma}$.

Proposition 2.8. *Let (Σ, \mathcal{T}) be a piecewise flat surface with home triangle T_0 and choose an isometry f_0 from T_0 onto a triangle in \mathbb{R}^2 . Then there exists a unique map $f: \hat{\Sigma} \rightarrow \mathbb{R}^2$ such that f coincides with f_0 on T_0 and f maps every hinge onto a quadrilateral in \mathbb{R}^2 .*

Proof. Let \hat{x} be a point in $\hat{\Sigma}$. This point belongs to a simplex $\hat{\sigma} = (\sigma, [c])$ in $\hat{\mathcal{T}}$. Choose an admissible arc c connecting the base point $x_0 \in T_0$ to σ . Because c is admissible, it crosses only finitely many edges e_1, e_2, \dots, e_m in that order (repetitions may occur). We associate to the path c a sequence of triangles $T_1, T_2, \dots, T_m \in \mathcal{T}$ by requiring that (T_0, T_1) be the hinge of e_1 , (T_1, T_2) be the hinge of e_2 and so on. We then define $f_j: T_j \rightarrow \mathbb{R}^2$ to be the continuation of f_{j-1} across the edge e_j (for $1 \leq j \leq m$) and we finally set $f(\hat{x}) = f_m(P(\hat{x}))$. The point $f(\hat{x}) \in \mathbb{R}^2$ only depends on the homotopy class $[c]$ and not on the representative path c .

It is clear from the construction that $f: \hat{\Sigma} \rightarrow \mathbb{R}^2$ maps every hinge onto a quadrilateral in \mathbb{R}^2 . Since f extends f_0 , the proof is complete. \square

Definition 2.9. The map $f: \hat{\Sigma} \rightarrow \mathbb{R}^2$ is the *developing map* of the piecewise flat surface.

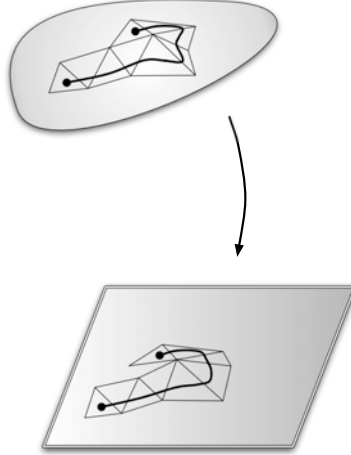


Figure 1. The developing map.

If $f': \hat{\Sigma} \rightarrow \mathbb{R}^2$ is another developing map, then, clearly, $f' = g \circ f$ where $g: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is the unique isometry of the plane such that $g(f(T_0)) = f'(T_0)$.

When Σ is the boundary surface of a convex polyhedron in \mathbb{R}^3 , the developing map is a very concrete operation. It is obtained by first placing the initial face (home) somewhere on the plane and then rolling without slipping the polyhedron, face after face, following an admissible path. Observe in particular that we can move our polyhedron toward any point in the plane. This is a general fact:

Proposition 2.10. *Let Σ be a compact piecewise flat surface without boundary. Then any developing map $f: \hat{\Sigma} \rightarrow \mathbb{R}^2$ is surjective.*

Proof. Observe first that f is a closed map (because it is an isometry on each triangle). Suppose that $\mathbb{R}^2 \setminus f(\hat{\Sigma}) \neq \emptyset$, then this set is open and we can find a point $y \in \mathbb{R}^2$ which lies on the boundary of $f(\hat{\Sigma})$. Because f is closed, we can find $\hat{x} \in \hat{\Sigma}$ with $f(\hat{x}) = y$. Let $x = P(\hat{x}) \in \Sigma$. This point cannot be in the interior of any triangle of the triangulation, thus x lies on an edge e . Moving slightly the point y if necessary, we can assume that x lies in the interior e (i.e. that x is not a vertex).

Since Σ has no boundary, e is an interior edge; the developing map f sends the hinge of e onto a quadrilateral Q in \mathbb{R}^2 . The interior of e is sent in the interior of $Q \subset f(\hat{\Sigma})$. This contradicts the point y lying on the boundary of $f(\hat{\Sigma})$. \square

2.4 The holonomy of a piecewise flat surface

The set of all admissible homotopy classes in the piecewise flat surface (Σ, \mathcal{T}) which start and end at the base point x_0 form a group π with respect to the concatenation. This group coincides with the fundamental group $\pi_1(\Sigma', x_0)$.

If $[a] \in \pi$ and $(\sigma, [b]) \in \tilde{\mathcal{T}}$, then $(\sigma, [ba])$ is well defined, and this gives a simplicial action of π on $\tilde{\mathcal{T}}$.

Corresponding to this simplicial action, there is an action of π on $\hat{\Sigma}$ by isometries; the orbits space of this action coincides with the surface itself. i.e. $\Sigma = \hat{\Sigma}/\pi$.

If $f: \hat{\Sigma} \rightarrow \mathbb{R}^2$ is a developing map of (Σ, \mathcal{T}) and $\gamma = [c] \in \pi$, then there is a unique isometry $g: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that the $g(f(T_0)) = f(T_0, [c])$. We denote this isometry by $g = \varphi(\gamma)$.

Proposition 2.11. *The map $\varphi: \pi \rightarrow E(2)$ (the group of isometries of the euclidean plane) is a group homomorphism.*

Proof. This easily follows from the construction of the developing map. \square

Definition 2.12. The homomorphism $\varphi: \pi \rightarrow E(2)$ is called the *holonomy* associated to the developing map f .

Proposition 2.13. *If Σ is compact without boundary, then the group $H = \varphi(\pi) \subset E(2)$ has no bounded orbit (in particular it has no fixed point).*

An obvious but important consequence is the fact that H is not conjugate to a subgroup of $O(2)$.

Proof. Suppose that there is a point $y \in \mathbb{R}^2$ such that $H \cdot y$ is bounded. Since the developing map is surjective, there exists a point $\hat{x} \in \hat{\Sigma}$ such that $f(\hat{x}) = y$. Observe that $H \cdot y = f(\pi \cdot \hat{x})$. Any point in the surface Σ can be connected to $x = P(\hat{x})$ by a path of length at most $D = \text{diam}(\Sigma)$, hence any point in $\hat{\Sigma}$ can be connected to a point in the orbit $\pi \cdot \hat{x}$ by a path of length at most D .

Since f preserves the length of all paths, it follows that any point in the image $f(\hat{\Sigma})$ can be connected to a point in the orbit $H \cdot y$ by a path of length at most D . The last assertion contradicts the surjectivity of f . \square

Recall that the developing map of a piecewise flat surface is not unique, it depends on the choice of an isometry of the home triangle into \mathbb{R}^2 . However the holonomy is well defined up to conjugacy:

Proposition 2.14. *Let $f, f': \hat{\Sigma} \rightarrow \mathbb{R}^2$ be two developing maps of the piecewise flat surface (Σ, \mathcal{T}) , and let $\varphi, \varphi': \pi \rightarrow E(2)$ be the corresponding holonomies. Then $\varphi'(\gamma) = g\varphi(\gamma)g^{-1}$ where $g \in E(2)$ is the unique isometry such that $f' = g \circ f$.*

Proof. The holonomy is defined by the condition $f(T_0, \gamma) = \varphi(\gamma)(f(T_0))$, hence

$$\begin{aligned} \varphi'(\gamma)(f(T_0)) &= f'(T_0, \gamma) = g \circ f(T_0, \gamma) = g \circ \varphi(\gamma)(f(T_0)) \\ &= g \circ \varphi(\gamma) \circ g^{-1}(f'(T_0)). \quad \square \end{aligned}$$

2.5 The developing map near a singularity

The previous notions can be clearly visualized if one restricts one's attention to a simply connected region $\Omega \subset \Sigma$ which is a union of triangles and which contains exactly one singular vertex p of order $\beta = \beta(p) \neq 0$.

Suppose that the base point x_0 sits in Ω and choose a loop c in $\Omega' = \Omega \setminus \{p\}$, based at x_0 and surrounding the point p once (so that $[c]$ is a generator of $\pi_1(\Omega', x_0) \cong \mathbb{Z}$).

Choose a connected component $\hat{\Omega}$ of the inverse image $P^{-1}(\Omega) \subset \hat{\Sigma}$ and still denote by P the (restriction of the) projection $P: \hat{\Omega} \rightarrow \Omega$.

We want to describe the geometry of $\hat{\Omega}$, of the map P as well as the developing map and holonomy restricted to $\hat{\Omega}$.

It is enough to consider the case where Ω is the “star” of the vertex p , i.e. the union of all triangles incident with p (if Ω is a larger region, the other triangles will simply appear as an appendix glued to the star of p).

The space $\hat{\Omega}$ is the geometric realization of a simplicial complex whose simplices are simplices in Ω together with an admissible homotopy class of curve connecting the base point to the given simplex.

Let us denote by T_1, T_2, \dots, T_k the list of all triangles (i.e. 2-simplices) incident with p and assume that $x_0 \in T_1$. Assume also that T_i has a common edge with T_{i+1} and T_k has a common edge with T_1 . Then a triangle in $\hat{\Omega}$ is given by a pair $(T_i, [a])$ where $[a]$ is the homotopy class of a curve a in Ω' from x_0 to T_i . This homotopy class is parametrized by a single integer $d \in \mathbb{Z}$ (the degree of a) which counts the number of times a turns around the point p . In other words, $\hat{\Omega}$ is an infinite strip made out of countably many copies of each triangle T_1, T_2, \dots, T_k each indexed by the degree $d \in \mathbb{Z}$

$$\hat{\Omega} = \bigcup_{d \in \mathbb{Z}} (T_{1,d} \cup T_{2,d} \cup \dots \cup T_{k,d}).$$

To develop $\hat{\Omega}$, start with an isometry f_1 from T_1 to a triangle in the euclidean plane and continue this isometry by unfolding each hinge in $\hat{\Omega}$. The developing map f then clearly satisfies

$$f(T_{i,d}) = R^d f(T_{i,0})$$

where R is a rotation of angle θ (= the sum of the angles at p of the triangles T_1, T_2, \dots, T_k) around the point $q = f_1(p)$. The rotation $R \in E(2)$ is clearly the holonomy of the generator $[c]$ of $\pi_1(\Omega', x_0)$.

We collect in the next proposition, some of the conclusions of the previous discussion:

Proposition 2.15. (1) The inverse image $P^{-1}(p)$ of p in $\hat{\Omega}$ contains exactly one point \hat{p} .

(2) The holonomy $\varphi(c)$ of $[c]$ is a rotation of angle $\theta = 2\pi(\beta + 1)$.

(3) If β is not an integer, then $q = f(\hat{p})$ is the unique fixed point of the rotation $\varphi(c)$. \square

2.6 Geometric equivalence of euclidean triangulations

Let (Σ, \mathcal{T}) be a piecewise flat surface. Choose a triangle $T_{\alpha_0} \in \mathcal{T}$ and a point q in the interior of an edge of T_{α_0} . If one connects the point q to the opposite vertex in T_{α_0} by a euclidean segment, one obtains two subtriangles $T'_{\alpha_0}, T''_{\alpha_0}$ whose union is T_{α_0} .

If one replaces the triangle T_{α_0} with T'_{α_0} and T''_{α_0} in the triangulation \mathcal{T}_q , one obtains a new triangulation \mathcal{T}_q .

Definition 2.16. a) The triangulation \mathcal{T}_q is said to be obtained from \mathcal{T} by an *elementary subdivision*.

b) The *geometric equivalence* is the equivalence relation on the set of euclidean triangulations on a surface which is generated by elementary subdivisions.

In other words, two euclidean triangulations $\mathcal{T}_1, \mathcal{T}_2$ on the surface Σ are geometrically equivalent if there is a common subdivision \mathcal{T} which is also a euclidean triangulation.

Proposition 2.17. The area measure dA , the length structure ℓ , the singularity order β , the developing map and holonomy are invariants of this equivalence relation.

Proof. The statement is obvious for dA , ℓ and β . Observe now that if T is a triangle of \mathcal{T} and T', T'' is an elementary subdivision of T , then the pair (T', T'') is the hinge of their common edge $e \subset T$. Observe also that if $f: T \rightarrow \mathbb{R}^2$ is an isometry, then f is an unfolding of that hinge.

This argument shows that the developing map remains unchanged when subdividing the triangulation. Since the developing map is invariant, so is the holonomy. \square

2.7 Flat metrics with conical singularities

If (Σ, \mathcal{T}) is a piecewise flat surface, then Σ' carries a well defined Riemannian metric m ; this metric is flat (i.e. it has no curvature) and in the neighbourhood of a conical singularity of total angle θ , we can introduce polar coordinates (r, φ) , where $r \geq 0$ is the distance to p and $\varphi \in \mathbb{R}/(\theta\mathbb{Z})$ is the angular variable (it is defined modulo θ). In these coordinates, the metric reads

$$m = dr^2 + r^2 d\varphi^2.$$

A calculation shows that this metric can be written as

$$m_\beta = |z|^{2\beta} |dz|^2, \quad (2.2)$$

where $z = \frac{1}{\beta+1} (r e^{i\varphi})^{\beta+1}$ (see [31]).

Definition 2.18. A *flat surface with conical singularities* (Σ, m) is a surface Σ together with a singular Riemannian metric m which is isometric to the metric m_β in (2.2) in the neighbourhood of every point $p \in \Sigma$, where $\beta = \beta(p) \in (-1, \infty)$.

One says that $\beta(p)$ is the *singularity order* of p and p is a conical singularity if $\beta(p) \neq 0$. The singular points form a discrete set and the formal sum (with discrete support) $\sum \beta(p) p$ is called the *divisor* of the singular metric m . One sometimes also says that m *represents* this divisor.

Proposition 2.19. *Any compact flat surface with conical singularities (Σ, m) can be geodesically triangulated. The resulting triangulation is a euclidean triangulation on Σ and the associated length structure coincides with the length in the metric m . \square*

A proof can be found in [31] and in [30]. See also [18] and [27] for further discussions on triangulations of piecewise flat surfaces.

Proposition 2.20. *Two euclidean triangulations $\mathcal{T}, \mathcal{T}'$ on a compact surface Σ are geometrically equivalent if and only if they give rise to the same flat surface with conical singularities m on Σ .*

Proof. It is clear from Proposition 2.17, that two triangulations which are geometrically equivalent give rise to the same singular flat metric. Conversely, suppose that the triangulations \mathcal{T} and \mathcal{T}' define the same metric, then each triangle of \mathcal{T} is decomposed by \mathcal{T}' in a finite number of polygonal regions. We may then further decompose these polygons in euclidean triangles, and we thus obtain a new euclidean triangulation of Σ which is a subdivision of both \mathcal{T} and \mathcal{T}' . \square

2.8 Relation with Riemann surfaces

If (Σ, m) is an oriented flat surface with conical singularities, then it is covered by charts $\{(U_j, z_j)\}$ such that the metric m takes the form (2.2) in each U_j . The transition from one such coordinate z_j to another one is given by a conformal transformation. Thus Σ is a Riemann surface with a holomorphic atlas given by $\{(U_j, z_j)\}$.

Remark 2.21. The reader should observe here that the conical singularities are invisible from the conformal viewpoint. This is a consequence of the formula (2.2) which shows that the singular metric is conformal to a smooth metric. It can also be seen as a consequence of the theorem of removability of singularities of locally bounded meromorphic functions.

In the converse direction, we can start with a closed Riemann surface with a divisor and ask whether there is a conformal flat metric representing this divisor. The answer is positive and the following theorem classifies all compact euclidean surfaces with conical singularities.

Theorem 2.22. *Let Σ be a compact connected Riemann surface without boundary. Fix n distinct points $p_1, p_2, \dots, p_n \in \Sigma$ and n real numbers $\beta_1, \beta_2, \dots, \beta_n \in (-1, \infty)$.*

There exists a conformal flat metric m on Σ having a conical singularity of order β_j at p_j ($j = 1, \dots, n$) if and only if the Gauss–Bonnet condition $\chi(S) + \sum_{j=1}^n \beta_j = 0$ holds. This metric is unique up to homothety. \square

See [31], a shorter proof can be found in [33, §IV].

Remark 2.23. A careful examination of the proof shows that the metric m depends continuously on all the parameters: The conformal structure, the points p_j and the orders β_j .

There is a similar theorem for the case of hyperbolic metrics with conical singularities, see [15], [23], [26], [32]. There are also various other extensions (non constant curvature, non orientable surfaces, non compact surfaces, and surfaces with boundary, see [17], [32], [33]). The case of spherical metric is more delicate, see [12], [34] for a study of spherical metric with three conical singularities on the 2-sphere.

Theorem 2.24. *Given a compact oriented surface Σ , there are natural bijections between the following three sets:*

- (1) *The set of geometric equivalence classes of euclidean triangulations on Σ up to homothety;*
- (2) *the set of flat metrics m on Σ with conical singularities up to homothety;*
- (3) *the set of conformal structures on Σ together with a finite real divisor $\sum_i \beta_i p_i$ such that $\beta_i > -1$ and the Gauss–Bonnet condition (2.1) is satisfied.*

Proof. Theorem 2.22 says precisely that there is a bijection between sets (2) and (3). Proposition 2.20 shows that there is a natural injection from (1) to (2), this injection is surjective by Proposition 2.19. \square

3 Punctured surfaces

3.1 Punctured surfaces and their fundamental groups

Definition 3.1. We define a *punctured surface* $\Sigma_{g,n}$ to be an oriented, closed connected surface Σ of genus g together with a distinguished set of n pairwise distinct points $p_1, p_2, \dots, p_n \in \Sigma_{g,n}$.

The points p_1, \dots, p_n are considered to be special places (with some geometric significance) on the surface. We call them the *punctures* and we denote by $\Sigma'_{g,n}$ the surface obtained by removing them:

$$\Sigma'_{g,n} = \Sigma_{g,n} \setminus \{p_1, p_2, \dots, p_n\}.$$

The connected sum of two punctured surfaces is defined by removing a disk containing no puncture in each surface and then gluing them along their boundary. The resulting surface is again a punctured surface. In fact we have

$$\Sigma_{g_1, n_1} \# \Sigma_{g_2, n_2} = \Sigma_{g_1+g_2, n_1+n_2},$$

where the symbol $\#$ means the connected sum. In particular

$$\Sigma_{g,n} = \Sigma_{g,0} \# \Sigma_{0,n}. \quad (3.1)$$

We easily deduce from this observation that the Euler characteristic of $\Sigma'_{g,n}$ is given by

$$\chi(\Sigma'_{g,n}) = 2 - 2g - n. \quad (3.2)$$

If $n > 0$, then $\Sigma'_{g,n}$ can be homotopically retracted onto a bouquet of $2g + n - 1$ circles and the fundamental group $\pi_{g,n}$ of $\Sigma'_{g,n}$ is thus a free group on $2g + n - 1$ generators.

Note that $\pi_{g,n}$ also admits the following presentation with $2g + n$ generators and one relation:

$$\pi_{g,n} = \langle a_1, \dots, a_g, b_1, \dots, b_g, c_1, \dots, c_n \mid \prod [a_i, b_i] = \prod c_j \rangle, \quad (3.3)$$

this presentation is a consequence of the identity (3.1) and Van Kampen's theorem.

3.2 Uniformization of a punctured Riemann surface

Let us fix a conformal structure $[m]$ on $\Sigma_{g,n}$. Assuming that $2 - 2g - n < 0$, the uniformization theorem states that (Σ', m) is conformally equivalent to \mathbb{U}/Γ where $\mathbb{U} = \{z \in \mathbb{C} \mid \operatorname{Re}(z) > 0\}$ is the upper-half plane, and $\Gamma \subset \operatorname{PSL}_2(\mathbb{R})$ is a Fuchsian group of the first kind² isomorphic to $\pi_{g,n}$.

The isomorphism $\pi_{g,n} \rightarrow \Gamma$ is compatible with the punctures in the sense that the generator c_i is sent to a parabolic element of Γ and the generators a_i, b_i are sent to hyperbolic elements (here, the letters a_i, b_i, c_j refer to the presentation (3.3)).

Let us denote by $\Upsilon \subset \partial\mathbb{U} = \mathbb{R} \cup \{\infty\}$ the set *cusp points* of Γ , i.e. the set of fixed points of all parabolic elements in Γ . Following [28, page 10], we define a topology on $\hat{\mathbb{U}} = \mathbb{U} \cup \Upsilon$ as follows: for a point $z \in \mathbb{U}$, the family of hyperbolic disks $D(z, \rho)$ is a fundamental system of neighborhoods of z . For a point $y \in \Upsilon$ the family of horodisks centered at y is a fundamental system of neighborhoods of y . With this topology, $\hat{\mathbb{U}}$ is a Hausdorff space and Γ acts by homeomorphisms. The space is not

²Recall that a *Fuchsian group* is a discrete subgroup of $\operatorname{PSL}_2(\mathbb{R})$, it is of the *first kind* if there is a fundamental domain $D \subset \mathbb{U}$ of finite hyperbolic area.

locally compact and Υ is topologically a discrete space. Standard arguments from hyperbolic geometry (see e.g. [28]) show that the projection map $P : \mathbb{U} \rightarrow \mathbb{U}/\Gamma = \Sigma'$ extends to a surjective continuous map

$$\hat{P} : \hat{\mathbb{U}} \rightarrow \Sigma \tag{3.4}$$

where $\hat{\mathbb{U}} = \mathbb{U} \cup \Upsilon$. This extension maps Υ to the punctures $\{p_i\} \subset \Sigma$.

Remarks. 1) The previous considerations show that there exists a unique metric m_{-1} on Σ' of constant curvature -1 which is complete, has finite volume and belongs to the conformal structure $[m]$. This metric has a cusp at each puncture p_i , it is the unique metric such that P^*m_{-1} is the Poincaré metric on \mathbb{U} ; its existence can also be proved by directly solving the prescribed curvature equation (see [16], [17]).

2) We know from Theorem 2.22 that the conformal class $[m]$ also contains a metric m_0 on Σ' , unique up to homothety, which is flat and has a conical singularity of order β_j at p_j ($j = 1, \dots, n$) provided (2.1) holds.

This flat metric lifts as a flat conformal metric $\tilde{m}_0 = P^*(m_0)$ on \mathbb{U} . For this metric, \mathbb{U} is not complete and its completion is given by $\hat{\mathbb{U}} = \mathbb{U}$. The map $\hat{P} : \hat{\mathbb{U}} \rightarrow \Sigma$ is thus a concrete model of the universal branched covering introduced earlier.

We identify the set Υ as a subset of Γ as follows: we first fix a base point $\tilde{z}_0 \in \mathbb{U}$ and let $z_0 = P(\tilde{z}_0) \in \Sigma$. For $y \in \Upsilon$, let us denote by $\tilde{\gamma}_y$ the hyperbolic ray in \mathbb{U} starting at \tilde{z}_0 and asymptotic to the point y , and let $\gamma_y = P(\tilde{\gamma}_y)$, this is a path joining z_0 to a puncture $p_i = P(y)$. Now let $D_i \subset \Sigma$ be a small disk around p_i containing no other puncture, and let $\gamma'_y = \gamma_y \setminus D_i$.

We now define $c_y \in \pi_1(\Sigma', z_0)$ to be the homotopy class of the path obtained by following γ'_y , then ∂D_i (in the positive direction) and then $(\gamma'_y)^{-1}$.

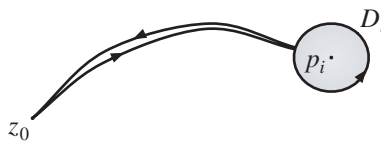


Figure 2. The homotopy class c_y .

Recall that we have a canonical isomorphism, $\Gamma \cong \pi_1(\Sigma', z_0) = \pi_{g,n}$, we have thus constructed a map

$$\begin{aligned} \Upsilon &\rightarrow \Gamma, \\ y &\mapsto c_y. \end{aligned} \tag{3.5}$$

It is clear that $c_y \in \Gamma$ is a parabolic element fixing y , in particular, the map $\Upsilon \rightarrow \Gamma$ is injective.

3.3 Some groups of diffeomorphisms of a punctured surface

Given a punctured surface $\Sigma = \Sigma_{g,n}$, we define $\text{Diff}_{g,n}$ to be the group of diffeomorphisms $h: \Sigma \rightarrow \Sigma$ which leaves the set $\{p_1, \dots, p_n\}$ of punctures invariant. We also introduce the following subgroups: $\text{Diff}_{g,n}^+ \subset \text{Diff}_{g,n}$ is the subgroup of orientation preserving diffeomorphisms, $\text{PDiff}_{g,n}$ is the subgroup of pure diffeomorphisms, i.e. diffeomorphisms fixing each puncture p_i individually and $\text{PDiff}_{g,n}^+ = \text{PDiff}_{g,n} \cap \text{Diff}_{g,n}^+$.

Every element $h \in \text{Diff}_{g,n}^+$ permutes the punctures and we have an exact sequence

$$1 \rightarrow \text{PDiff}_{g,n}^+ \rightarrow \text{Diff}_{g,n}^+ \rightarrow \text{Sym}(n) \rightarrow 1.$$

where $\text{Sym}(n)$ is the permutation group of $\{p_1, \dots, p_n\}$.

We also define $\text{Diff}_{g,n}^0 \subset \text{PDiff}_{g,n}^+$ to be the group of diffeomorphisms which are isotopic to the identity through an isotopy fixing the punctures. The quotient

$$\text{Mod}_{g,n} = \pi_0(\text{Diff}_{g,n}^+) = \text{Diff}_{g,n}^+ / \text{Diff}_{g,n}^0,$$

is called the *mapping class group* or the *modular group* of the punctured surface $\Sigma_{g,n}$, and

$$\text{PMod}_{g,n} = \pi_0(\text{PDiff}_{g,n}^+) = \text{PDiff}_{g,n}^+ / \text{Diff}_{g,n}^0,$$

is the *pure mapping class group*.

These groups have been intensely studied since the pioneer work of Dehn and Nielsen. We refer to [4], [19], [24], [36] among many other papers for more information.

3.4 Outer automorphisms

The mapping class group is related to the group of outer automorphisms of the fundamental group of Σ' . Let us recall this purely algebraic notion: If π is an arbitrary group, we denote by $\text{Aut}(\pi)$ the group of all its automorphisms and by $\text{Inn}(\pi) \subset \text{Aut}(\pi)$ the subgroup of inner automorphisms (i.e. conjugations $\gamma \rightarrow \alpha\gamma\alpha^{-1}$). This is a normal subgroup.

Definition 3.2. The group of *outer automorphisms* of π is the quotient

$$\text{Out}(\pi) = \text{Aut}(\pi) / \text{Inn}(\pi).$$

An outer automorphism is thus an automorphism of π defined up to conjugacy.

Lemma 3.3. *There is a naturally defined group homomorphism*

$$\text{Mod}_{g,n} \rightarrow \text{Out}(\pi_{g,n}).$$

Proof. This homomorphism is defined as follows. Let $h \in \text{Diff}(\Sigma')$ be an arbitrary diffeomorphism and fix a base point $*$ and a path δ in Σ' connecting $*$ to $h(*)$. If γ

is a loop in Σ' based at $*$, then we set

$$h_\delta(\gamma) = \delta^{-1}(h \circ \gamma)\delta.$$

This defines an automorphism $h_{\delta,\#} \in \text{Aut}(\pi_{g,n})$.

If δ' is another path connecting $*$ to $h(*)$, then $h_{\delta,\#}$ and $h_{\delta',\#}$ are conjugate by $\delta^{-1}\delta'$. The outer automorphism $h_\# \in \text{Out}(\pi_{g,n})$ is thus well defined independently of the choice of the path δ and it is clear that if h is homotopic to the identity, then it acts trivially on $\pi_{g,n}$, i.e. we have defined a map $\text{Mod}_{g,n} \rightarrow \text{Out}(\pi_{g,n})$. It is routine to check that it is a group homomorphism. \square

Introducing the group $\text{POut}(\pi_{g,n}) \subset \text{Out}(\pi_{g,n})$ of all outer automorphisms preserving the conjugacy class of each c_i ($i = 1, \dots, n$) in the presentation (3.3), we have the following deep result:

Theorem 3.4. *If $g > 0$ and $n > 0$, then the homomorphism defined in the previous lemma induces an isomorphism*

$$\Phi: \text{PMod}_{g,n} \xrightarrow{\sim} \text{POut}(\pi_{g,n}). \tag{3.6}$$

\square

This is the so-called *Dehn–Nielsen–Baer Theorem*, see [19], [36] for a proof.

3.5 Lifting the group $\text{Diff}^0(\Sigma')$ on \mathbb{U}

Using the notations of Section 3.2, one writes the universal branched covering of $\Sigma_{g,n}$ as $\hat{P}: \hat{\mathbb{U}} \rightarrow \Sigma$, where $\hat{\mathbb{U}} = \mathbb{U} \cup \Upsilon$ (we still assume $2 - 2g - n < 0$).

We denote by $\text{Diff}^+(\mathbb{U})$ the group of orientation preserving diffeomorphisms of $\mathbb{U} = \tilde{\Sigma}'$ and we define the *normalizer* $N(\Gamma)$ and the *centralizer* $C(\Gamma)$ of Γ in $\text{Diff}^+(\mathbb{U})$ by

$$N(\Gamma) = \{h \in \text{Diff}^+(\mathbb{U}) \mid h\Gamma = \Gamma h\}.$$

and

$$C(\Gamma) = \{h \in \text{Diff}^+(\mathbb{U}) \mid h \circ \gamma = \gamma \circ h \text{ for all } \gamma \in \Gamma\}.$$

Observe that $C(\Gamma) = \ker(\psi)$, where $\psi: N(\Gamma) \rightarrow \text{Aut}(\Gamma)$ is defined by $\psi(h): \gamma \rightarrow h\gamma h^{-1}$.

The *center* of Γ is the intersection $Z(\Gamma) = \Gamma \cap C(\Gamma)$; it is the largest abelian subgroup of Γ .

Lemma 3.5. *Let Γ be an arbitrary Fuchsian group, then $Z(\Gamma)$ is trivial unless Γ is cyclic.*

Proof. This follows from classical Fuchsian group theory. Indeed, it is well known that if γ_1, γ_2 are non-trivial elements in $\text{PSL}_2(\mathbb{R})$, then they commute if and only if they have the same fixed points (see e.g. [20, Theorem 2.3.2]). So if $Z(\Gamma)$ contains a

non-trivial element γ_0 , then any $\gamma \in \Gamma \setminus \{\text{id}\}$ must have the same fixed points as γ_0 and it follows from [20, theorem 2.3.5]) that Γ is cyclic. \square

Recall the projection $P: \mathbb{U} \rightarrow \Sigma' = \mathbb{U}/\Gamma$. For any element $h \in N(\Gamma)$, we define $P_*h: \Sigma' \rightarrow \Sigma'$ by $P_*h(x) = P(h(\tilde{x}))$ where $\tilde{x} \in \mathbb{U}$ is an arbitrary point in $P^{-1}(x)$. This map is well-defined, because the condition $h\Gamma = \Gamma h$ means precisely that h maps Γ -orbits in U to Γ -orbits, and it is clearly a diffeomorphism. We thus have defined a map

$$P_*: N(\Gamma) \rightarrow \text{Diff}^+(\Sigma'),$$

and it is obviously a group homomorphism.

Proposition 3.6. Γ is a normal subgroup in $N(\Gamma)$ and P_* defines an isomorphism from $N(\Gamma)/\Gamma$ to $\text{Diff}^+(\Sigma')$.

Proof. It is obvious that $\Gamma \subset N(\Gamma)$ is normal and that $P_*(\Gamma) = \{\text{id}\}$. In particular P_* factors through a well defined homomorphism $N(\Gamma)/\Gamma \rightarrow \text{Diff}^+(\Sigma')$. This homomorphism is surjective since every diffeomorphism of Σ' lifts to the universal cover \mathbb{U} of Σ' .

Suppose now that $P_*h = \text{id}$. Then $h(x) \in \Gamma \cdot x$ for all $x \in U$. This means that there exists a map $U \rightarrow \Gamma$, $x \mapsto \gamma_x$ such that $h(x) = \gamma_x x$ for all $x \in U$. Since h is continuous, so is this map, but this implies that $x \mapsto \gamma_x$ is constant because Γ is a discrete group. It follows that $h \in \Gamma$ and we have shown that $P_*: N(\Gamma)/\Gamma \rightarrow \text{Diff}^+(\Sigma')$ is also injective. \square

Lemma 3.7. P_* maps $C(\Gamma)$ isomorphically onto $\text{Diff}^0(\Sigma')$.

Proof. Suppose that $P_*h \in \text{Diff}^0(\Sigma')$. Then there exists an isotopy $h_t \in N(\Gamma)$ such that $h_0 = \text{id}$ and $h_1 = h$. Hence $\psi(h_t) \in \text{Aut}(\Gamma)$ is constant by continuity. Because $\psi(h_0) = \psi(\text{id}) \in \text{Aut}(\Gamma)$ is the trivial element, we have $h \in \ker \psi = C(\Gamma)$.

In the reverse direction, we use an argument going back to Nielsen: Suppose that $h \in \ker \psi = C(\Gamma)$ and define $h_t(x) \in U$ to be the point on the hyperbolic segment $[x, h(x)]$ such that $d(x, h_t(x)) = td(x, h(x))$ (where d is the hyperbolic distance in \mathbb{U}). Since $h \in \ker \psi$ and Γ preserves the hyperbolic distance in \mathbb{U} , the segment $[\gamma x, h(\gamma x)]$ coincides with $[\gamma h_t(x), \gamma h(x)]$ for any $x \in \mathbb{U}$ and any $\gamma \in \Gamma$. Therefore we have $h_t(\gamma x) = \gamma h_t(x)$, i.e. $h_t \in C(\Gamma) \subset N(\Gamma)$. The path $P_*h_t \in \text{Diff}(\Sigma')$ is an isotopy from P_*h to the identity and we conclude that $P_*h \in \text{Diff}^0(\Sigma')$.

We have proved that $P_*^{-1}(\text{Diff}^0(\Sigma')) = C(\Gamma)$. It is now clear that $P_*: C(\Gamma) \rightarrow \text{Diff}^0(\Sigma')$ is an isomorphism since its kernel is $C(\Gamma) \cap \Gamma = Z(\Gamma) = \{\text{id}\}$. \square

Corollary 3.8. P_* induces an isomorphism from $N(\Gamma)/(\Gamma \times C(\Gamma))$ to the modular group $\text{Mod}_{g,n}$.

4 The representation variety of a finitely generated group in SE(2)

Given a finitely generated group π and an algebraic Lie group G , it is easy to see that the set $\text{Hom}(\pi, G)$ is an algebraic set. The group G itself acts on $\text{Hom}(\pi, G)$ by conjugation: $g \cdot \varphi(\gamma) = g^{-1}\varphi(\gamma)g$. The quotient space is called the *representation variety* of π in G and denoted by

$$\mathcal{R}(\pi, G) = \text{Hom}(\pi, G)/G.$$

This variety plays an important role in the study of geometric structures on manifolds, see e.g. [14].

The discussion in Section 2.4 shows that an element of the representation variety $\mathcal{R}(\pi, \text{E}(2))$ is associated to any piecewise flat surface (Σ, \mathcal{T}) (where $\pi = \pi_1(\Sigma', x_0)$). In the present section, we investigate the structure of $\mathcal{R}(\pi, \text{E}(2))$ (in fact, for convenience, we shall restrict ourselves to the subgroup $\text{SE}(2) \subset \text{E}(2)$ of orientation preserving isometries of the euclidean plane, this is a subgroup of index 2).

4.1 On the cohomology of groups

We will need some elementary results from group cohomology; here we recall a few basic definitions and facts.

Let π be an arbitrary group and A be a π -module, i.e. an abelian group with a representation $\rho: \pi \rightarrow \text{Aut}(A)$.

Definition 4.1. (1) A 1-cocycle in A is a map $\sigma: \pi \rightarrow A$ such that

$$\sigma(\gamma_1\gamma_2) = \sigma(\gamma_1) + \rho(\gamma_1) \cdot \sigma(\gamma_2)$$

for any $\gamma_1, \gamma_2 \in \pi$. The set of 1-cocycles is an abelian group denoted by $Z^1(\pi, A)$.

(2) The 1-cocycle $\sigma \in Z^1(\pi, A)$ is a 1-coboundary if it can be written as

$$\sigma = \delta_a(\gamma) = \rho(\gamma) \cdot a - a$$

for some element $a \in A$. The set of 1-coboundaries is a subgroup of $Z^1(\pi, A)$ denoted by $B^1(\pi, A)$.

(3) The quotient

$$H^1(\pi, A) = Z^1(\pi, A)/B^1(\pi, A)$$

is the *first cohomology group of π with values in A* .

Example. Let us compute the first cohomology group when $A = k$ is a field and π is a finitely generated group. We denote by k_ρ the π -module k with the representation $\rho: \pi \rightarrow \text{Aut}(k)$.

Assume first that the representation $\rho: \pi \rightarrow \text{Aut}(k)$ is a scalar representation, i.e. $\rho: \pi \rightarrow k^* \subset \text{Aut}(k)$ and that $\pi = F_s = \langle a_1, a_2, \dots, a_s \rangle$ is a free group on s generators.

Since π is free, the homomorphism $\rho: \pi \rightarrow k^*$ is completely determined by the vector $r = (\rho(a_1), \rho(a_2), \dots, \rho(a_s)) \in (k^*)^s$. Likewise, a cocycle is given by the vector

$$t = (\tau(a_1), \tau(a_2), \dots, \tau(a_s)) \in k^s.$$

There is no restriction on the vector $t \in k^s$ (again because π is free) and thus

$$Z^1(\pi, k_\rho) \cong k^s. \tag{4.1}$$

An element $\sigma \in Z^1(\pi, k_\rho)$ is a coboundary if $\sigma = u(\text{id} - \rho)$ for some $u \in k$, thus

$$B^1(\pi, k_\rho) \cong k \cdot (1 - \rho(a_1), 1 - \rho(a_2), \dots, 1 - \rho(a_s)) \in k^s.$$

Let us choose a linear form $\mu: k^s \rightarrow k$ such that $\mu \equiv 0$ if ρ is trivial and

$$\mu(1 - \rho(a_1), 1 - \rho(a_2), \dots, 1 - \rho(a_s)) \neq 0$$

else. It is easy to check that

$$B^1(\pi, k_\rho) \oplus \ker \mu = k^s = Z^1(\pi, k_\rho)$$

in k^s and we thus obtain the following

Proposition 4.2. *For any free group on s generators, we have*

$$H^1(\pi, k_\rho) = Z^1(\pi, k_\rho) / B^1(\pi, k_\rho) = \ker \mu \cong \begin{cases} k^s & \text{if } \rho \text{ is trivial,} \\ k^{s-1} & \text{otherwise.} \end{cases} \quad \square$$

General case. Let us compute the first cohomology group when $A = k$ is a field and π is a finitely presented group with presentation

$$\pi = \langle S \mid R \rangle.$$

Here $S = \{a_1, a_2, \dots, a_s\} \subset \pi$ denotes a finite set generating the group and $R = \{r_1, r_2, \dots, r_m\} \subset F(S)$ (the free group on S) is a finite set of words in S defining all the relations among the elements of S . We denote by k_ρ the π -module k with the representation $\rho: \pi \rightarrow \text{Aut}(k)$.

For any relation $r = a_{i_1} a_{i_2} \dots a_{i_p} \in R$, we introduce the linear form $\lambda_r: k^s \rightarrow k$ defined by

$$\lambda_r(\sigma) = \sum_{\mu=1}^p \left(\prod_{v < \mu} \rho(a_{i_v}) \right) \sigma(a_{i_\mu}) \tag{4.2}$$

and we define $\Lambda: k^s \rightarrow k^m$, by

$$\Lambda(\sigma) = (\lambda_{r_1}(\sigma), \dots, \lambda_{r_m}(\sigma)). \tag{4.3}$$

Lemma 4.3. *The space of 1-cocycles in k_ρ is given by*

$$Z^1(\pi, k_\rho) = \ker \Lambda = \bigcap_{r \in R} \ker \lambda_r \subset k^s.$$

Proof. If $\sigma \in Z^1(\pi, k_\rho)$ and $r = a_{i_1} a_{i_2} \dots a_{i_p} \in R$, then we deduce from the cocycle relation that

$$\begin{aligned} 0 &= \sigma(r) = \sigma(a_{i_1} a_{i_2} \dots a_{i_p}) = \sigma(a_{i_1}) + \rho(a_{i_1}) \sigma(a_{i_2} \dots a_{i_p}) \\ &= \sigma(a_{i_1}) + \rho(a_{i_1}) \sigma(a_{i_2}) + \rho(a_{i_1}) \rho(a_{i_2}) \sigma(a_{i_3} \dots a_{i_p}) \\ &= \sum_{\mu=1}^p \left(\prod_{\nu < \mu} \rho(a_{i_\nu}) \right) \sigma(a_{i_\mu}). \end{aligned} \quad \square$$

On the other hand, since any 1-coboundary in k_ρ is a multiple of $\rho - 1$, we have

$$B^1(\pi, k_\rho) = k \cdot (\rho - 1) \subset k^s.$$

We have proved the following

Proposition 4.4. *The first cohomology group of the finitely presented group $\pi = \langle S \mid R \rangle$ with value in k_ρ is given by*

$$H^1(\pi, k_\rho) = \ker \Lambda / (k(\rho - 1)). \quad \square$$

In particular, if π has exactly one non trivial relation, then

$$H^1(\pi, k) \cong \begin{cases} k^{s-1} & \text{if } \rho \text{ is trivial,} \\ k^{s-2} & \text{otherwise,} \end{cases}$$

where $s = \text{Card}(S)$ is the number of generators.

4.2 Abelian representations

Representations of a finitely presented group π in an abelian Lie group are easy to describe:

Lemma 4.5. *If G is an abelian group, then $\mathcal{R}(\pi, G) = \text{Hom}(\pi, G)$. This set is itself an abelian topological group.*

Proof. Since there are no non trivial inner automorphisms in an abelian group, it is clear that $\mathcal{R}(\pi, G) = \text{Hom}(\pi, G)$.

We endow $\text{Hom}(\pi, G)$ with the compact open topology and we define a product on this space by

$$(\varphi_1 \varphi_2)(\gamma) = \varphi_1(\gamma) \varphi_2(\gamma)$$

for $\varphi_1, \varphi_2 \in \text{Hom}(\pi, G)$ and $\gamma \in \pi$. The following calculation shows that $\text{Hom}(\pi, G)$ is a group for this multiplication:

$$\begin{aligned} (\varphi_1\varphi_2)(\gamma_1\gamma_2) &= \varphi_1(\gamma_1\gamma_2)\varphi_2(\gamma_1\gamma_2) \\ &= \varphi_1(\gamma_1)\varphi_1(\gamma_2)\varphi_2(\gamma_1)\varphi_2(\gamma_2) \\ &= \varphi_1(\gamma_1)\varphi_2(\gamma_1)\varphi_1(\gamma_2)\varphi_2(\gamma_2) \\ &= (\varphi_1\varphi_2)(\gamma_1)(\varphi_1\varphi_2)(\gamma_2). \end{aligned}$$

The identity e in $\text{Hom}(\pi, G)$ is the trivial representation. Observe finally that this group is abelian since $\varphi_1(\gamma)\varphi_2(\gamma) = \varphi_2(\gamma)\varphi_1(\gamma)$. \square

Recall that the abelianized group of π is the abelian group

$$\text{Ab}(\pi) = \pi/[\pi, \pi].$$

Another useful remark is that if G is abelian, then

$$\text{Hom}(\pi' \times \pi'', G) = \text{Hom}(\pi', G) \times \text{Hom}(\pi'', G)$$

for any groups π', π'' .

Assume now that π is a finitely generated group. $\text{Ab}(\pi)$ is then an abelian group of finite type, hence

$$\text{Ab}(\pi) = \pi/[\pi, \pi] = \mathbb{Z}^r \oplus F$$

where F is a finite abelian group (the torsion) and $r \in \mathbb{N}$ is the *abelian rank* of π .

We obviously have $\text{Hom}(\pi, G) = \text{Hom}(\text{Ab}(\pi), G)$ and it is clear that all representation varieties of a finitely generated group π in an abelian Lie group G can be deduced from the following special cases:

- (1) $\text{Hom}(\mathbb{Z}, \mathbb{R}) = \mathbb{R}$;
- (2) $\text{Hom}(\mathbb{Z}, U(1)) = U(1)$;
- (3) $\text{Hom}(\mathbb{Z}/m\mathbb{Z}, \mathbb{R}) = 0$;
- (4) $\text{Hom}(\mathbb{Z}/m\mathbb{Z}, U(1)) = \{z \in \mathbb{C} \mid z^m = 1\}$.

For instance, if π is the free group on s generators, then $\text{Ab}(\pi) = \mathbb{Z}^s$. Thus $\text{Hom}(\pi, \mathbb{R}) = \mathbb{R}^s$ and $\text{Hom}(\pi, U(1)) = \mathbb{T}^s$.

Another simple example, with torsion, is the group $\pi' = \langle a, b, c \mid [a, b] = c^m \rangle$. We have $\text{Ab}(\pi') = \mathbb{Z}^2 \oplus \mathbb{Z}/m\mathbb{Z}$, therefore $\text{Hom}(\pi', \mathbb{R}) = \mathbb{R}^2$ and

$$\text{Hom}(\pi', U(1)) = \mathbb{T}^2 \oplus \{e^{2ki\pi/m} \mid m = 0, 1, \dots, m-1\}.$$

4.3 Representations in $\text{SE}(2)$

We denote by $\text{SE}(2) = \text{Iso}^+(\mathbb{R}^2)$ the group of orientation preserving isometries of the euclidean plane.

We may identify the euclidean plane with the complex line \mathbb{C} : any $g \in \text{SE}(2)$ can then be written as $g(z) = u \cdot z + v$ where $u \in U(1) \subset \mathbb{C}^*$ and $v \in \mathbb{C}$. We thus identify $\text{SE}(2)$ with the subgroup of $\text{GL}_2(\mathbb{C})$ consisting of matrices of the form

$$\text{SE}(2) = \left\{ \begin{pmatrix} u & v \\ 0 & 1 \end{pmatrix} \mid u, v \in \mathbb{C}, |u| = 1 \right\}.$$

In particular $\text{SE}(2)$ is a semidirect product $U(1) \rtimes \mathbb{C}$ and any representation $\varphi \in \text{Hom}(\pi, \text{SE}(2))$ can be written as

$$\varphi = \begin{pmatrix} \rho_\varphi & \tau_\varphi \\ 0 & 1 \end{pmatrix} \quad (4.4)$$

where $\rho_\varphi: \pi \rightarrow U(1)$ and $\tau_\varphi: \pi \rightarrow \mathbb{C}$. Observe the following:

Lemma 4.6. *The map $\rho_\varphi: \pi \rightarrow U(1)$ is a group homomorphism. It only depends on the conjugacy class of φ .*

The proof is elementary.

Definition 4.7. The homomorphism $\rho_\varphi: \pi \rightarrow U(1)$ is the *character* of the representation class $\varphi \in \text{Hom}(\pi, \text{SE}(2))$.

Remark. In the literature on group representations, the character $\chi_\varphi: \pi \rightarrow K$ of a representation $\varphi \in \text{GL}_n(K)$ is classically defined to be the trace of the representation. The two notions of characters are equivalent as shown by the formula

$$\chi_\varphi = \text{Tr } \varphi = 1 + \rho_\varphi.$$

Any homomorphism $\rho \in \text{Hom}(\pi, U(1))$ defines a structure of π -module on \mathbb{C} . We will denote by \mathbb{C}_ρ this π -module, and we have:

Proposition 4.8. *Given any pair of maps $\rho: \pi \rightarrow U(1)$ and $\tau: \pi \rightarrow \mathbb{C}$, the map $\varphi: \pi \rightarrow \text{SE}(2)$ given by (4.4) is a group homomorphism if and only if $\rho \in \text{Hom}(\pi, U(1))$ and $\tau \in Z^1(\pi, \mathbb{C}_\rho)$.*

Proof. Suppose that $\varphi: \pi \rightarrow \text{SE}(2)$ is given by (4.4). Then we have

$$\varphi(\gamma_1\gamma_2) = \begin{pmatrix} \rho(\gamma_1\gamma_2) & \tau(\gamma_1\gamma_2) \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{aligned} \varphi(\gamma_1)\varphi(\gamma_2) &= \begin{pmatrix} \rho(\gamma_1) & \tau(\gamma_1) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \rho(\gamma_2) & \tau(\gamma_2) \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \rho(\gamma_1)\rho(\gamma_2) & \tau(\gamma_1) + \rho(\gamma_1)\tau(\gamma_2) \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

It follows that φ is a group homomorphism (i.e. $\varphi(\gamma_1\gamma_2) = \varphi(\gamma_1)\varphi(\gamma_2)$) if and only if

$$\rho(\gamma_1\gamma_2) = \rho(\gamma_1)\rho(\gamma_2)$$

and

$$\tau(\gamma_1\gamma_2) = \tau(\gamma_1) + \rho(\gamma_1)\tau(\gamma_2).$$

In other words φ is a homomorphism if and only if $\rho: \pi \rightarrow U(1)$ is a homomorphism and τ is a 1-cocycle in the corresponding π -module \mathbb{C}_ρ . \square

This proposition says that the map from $\text{Hom}(\pi, \text{SE}(2))$ to the set

$$\{(\rho, \tau) \mid \rho \in \text{Hom}(\pi, U(1)) \text{ and } \tau \in Z^1(\pi, \mathbb{C}_\rho)\}$$

given by $\varphi \rightarrow (\rho_\varphi, \tau_\varphi)$, is a bijection. In particular we have

Corollary 4.9. *If π is the free group on s generators, then*

$$\text{Hom}(\pi, \text{SE}(2)) \simeq \mathbb{T}^s \times \mathbb{C}^s.$$

Proof. This follows from equation (4.1) and the fact that $\text{Hom}(\pi, U(1)) = \mathbb{T}^s$. \square

4.4 Conjugation by similarities

Recall that a *similarity* in the plane is the composition of an isometry with a homothety.

Once we identify the euclidean plane with the complex line \mathbb{C} , any similarity $g \in \text{Sim}(2)$ can be written as $g(z) = a \cdot z + b$ where $a \in \mathbb{C}^*$ and $b \in \mathbb{C}$. We thus identify $\text{Sim}(2)$ with the following subgroup of $\text{GL}_2(\mathbb{C})$:

$$\text{Sim}(2) = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \mid a, b \in \mathbb{C}, a \neq 0 \right\}.$$

In particular we have

$$\text{Sim}(2) = \mathbb{R}_+ \rtimes \text{SE}(2) = \mathbb{C}^* \rtimes \mathbb{C}.$$

Definition 4.10. Two representations $\varphi_1, \varphi_2: \pi \rightarrow \text{SE}(2)$ are *similar* if they are conjugate modulo a similarity.

Proposition 4.11. *Given a homomorphism $\rho: \pi \rightarrow U(1)$ and two cocycles $\tau_1, \tau_2 \in Z^1(\pi, \mathbb{C}_\rho)$, then the representations*

$$\varphi_1 = \begin{pmatrix} \rho & \tau_1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \varphi_2 = \begin{pmatrix} \rho & \tau_2 \\ 0 & 1 \end{pmatrix} \quad (4.5)$$

are similar if and only if there exists a complex number $a \in \mathbb{C}^$ such that*

$$\tau_2 = a\tau_1 \in H^1(\pi, \mathbb{C}_\rho).$$

Proof. The homomorphisms φ_1 and φ_2 are similar if and only if there exists

$$g = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \in \text{Sim}(2)$$

such that $\varphi_2 = g\varphi_1g^{-1}$, i.e.

$$\begin{aligned} \begin{pmatrix} \rho & \tau_2 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \rho & \tau_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/a & -b/a \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \rho & a\tau_1 + b - \rho b \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

This shows that

$$\tau_2 - a\tau_1 = b(1 - \rho) \in B^1(\pi, \mathbb{C}_\rho). \quad \square$$

For any homomorphism $\varphi: \pi \rightarrow \text{SE}(2)$ and any $\lambda \in \mathbb{R}_+$, we can define a new homomorphism $\lambda \cdot \varphi: \pi \rightarrow \text{SE}(2)$ by

$$\lambda \cdot \varphi = \begin{pmatrix} \rho & \lambda\tau \\ 0 & 1 \end{pmatrix},$$

This formula defines an action of the multiplicative group \mathbb{R}_+ on $\mathcal{R}(\pi, \text{SE}(2))$, and we denote the quotient by

$$\mathcal{R}(\pi, \text{SE}(2)) = \mathcal{R}(\pi, \text{SE}(2))/\mathbb{R}_+.$$

It follows directly from the definition that

$$\mathcal{R}(\pi, \text{SE}(2)) = \text{Hom}(\pi, \text{SE}(2))/\text{Sim}(2).$$

where $\text{Sim}(2)$ acts by conjugation on $\text{Hom}(\pi, \text{SE}(2))$.

Let us also define

$$\mathcal{R}^{\text{reg}} = \{[\varphi] = [\rho, \tau] \in \mathcal{R}(\pi, \text{SE}(2)) \mid \rho_\varphi \neq \text{id and } \tau \neq 0.\}$$

Corollary 4.12. *If π is a free group on s generators, then*

$$\mathcal{R}^{\text{reg}} \simeq (\mathbb{T}^s \setminus \{\text{id}\}) \times \mathbb{C}\mathbb{P}^{s-2}.$$

Proof. This is an immediate consequence of the previous results, in particular Proposition 4.2 and 4.11. □

5 Deformation theory

5.1 The moduli and Teichmüller spaces

The moduli space of $\Sigma_{g,n}$ is the quotient of the space of conformal structures on $\Sigma_{g,n}$ by the pure diffeomorphism group. Let us be more specific: recall first that a conformal

structure is an equivalence class of smooth Riemannian metric m on $\Sigma = \Sigma_{g,n}$, where two Riemannian metrics m_1, m_2 are equivalent if and only if there exists a function $u: \Sigma \rightarrow \mathbb{R}$ such that

$$m_2 = e^{2u} m_1.$$

We denote by $\text{Met}(\Sigma)$ the space of all smooth Riemannian metrics on Σ endowed with its natural C^∞ topology and by

$$\text{Conf}(\Sigma) = \text{Met}(\Sigma) / C^\infty(\Sigma)$$

the space of conformal structures. We then define the *moduli space* of $\Sigma_{g,n}$ to be the quotient

$$\mathcal{M}_{g,n} = \text{Conf}(\Sigma) / \text{PDiff}_{g,n}^+.$$

A point $\mu \in \mathcal{M}_{g,n}$ is concretely represented by a Riemannian metric m on Σ , and two Riemannian metrics m_1, m_2 represent the same modulus point μ if and only if there exists a smooth function $u: \Sigma \rightarrow \mathbb{R}$ and a diffeomorphism $h \in \text{PDiff}_{g,n}^+$ such that $m_2 = e^{2u} h^*(m_1)$.

A remark about the smoothness: By definition a point μ in the moduli space is represented by a smooth metric. In particular, the punctures play no role in the definition of the spaces $\text{Met}(\Sigma)$ and $\text{Conf}(\Sigma)$ (but they do in the definition of the moduli space $\mathcal{M}_{g,n}$). However, since only the conformal class of the metric matters, one may also represent μ by a singular metric m as long as this metric is conformally equivalent to a smooth one. In particular we can (and will) represent a point in $\mathcal{M}_{g,n}$ by a metric having conical singularities at the punctures of $\Sigma_{g,n}$, see Remark 2.21

The moduli space is a complicated object, and it is useful to also introduce the simpler space obtained by considering isotopy classes of conformal structures on $\Sigma_{g,n}$ instead of isomorphism classes: this is the *Teichmüller space* defined as

$$\mathcal{T}_{g,n} = \text{Conf}(\Sigma_{g,n}) / \text{Diff}_{g,n}^0.$$

Let us list some of the basic facts about these spaces:

- (1) The Teichmüller space $\mathcal{T}_{g,n}$ is a real analytic variety in a natural way. If $3g-3+n > 0$, then it is isomorphic to $\mathbb{R}^{6g-6+2n}$. This space has also a natural complex structure.
- (2) The pure mapping class group $\text{PMod}_{g,n}^+$ acts properly and discontinuously on $\mathcal{T}_{g,n}$.
- (3) The moduli space is the quotient $\mathcal{M}_{g,n} = \mathcal{T}_{g,n} / \text{PMod}_{g,n}^+$. It is thus a good orbifold of dimension $6g - 6 + 2n$ with fundamental group $\pi_1(\mathcal{M}_{g,n}) = \text{PMod}_{g,n}$.
- (4) There exists a torsion free subgroup $M_0 \subset \text{PMod}_{g,n}^+$ of finite index acting freely on $\mathcal{T}_{g,n}$. The quotient map $\mathcal{T}_{g,n} / M_0$ is a non singular analytic manifold which is a finite cover of the orbifold $\mathcal{M}_{g,n}$.

Statement (1) is explained in any textbook on Teichmüller theory such as [1]. Statement (2) was first proved by S. Kravetz [22], see also [1]. (3) is a consequence of (1) and (2) and the last statement is discussed in [19, §5.4].

5.2 The deformation space of piecewise flat metrics

Let us denote by $\mathcal{E}_{g,n}$ the set of all flat metrics on $\Sigma_{g,n}$ with possible conical singularities at the punctures (it is not empty since we have assumed $2g + n - 2 > 0$). To any flat metric $m \in \mathcal{E}_{g,n}$, we associate the following basic invariants: Its conformal class $[m] \in \text{Conf}(\Sigma)$, its area $A = A(m) > 0$ and the order $\beta_i > -1$ of m at the point p_i .

Theorem 5.1. *The map*

$$\begin{aligned} \mathcal{E}_{g,n} &\rightarrow \text{Conf}(\Sigma) \times \mathbb{R}_+^n, \\ m &\mapsto ([m], (s_1, \dots, s_n)), \end{aligned}$$

where $s_i = A(1 + \beta_i) > 0$, is a bijection.

Proof. This is just a reformulation of Theorem 2.22. □

Definition 5.2. We will endow the set $\mathcal{E}_{g,n}$ with the topology for which this map is a homeomorphism.

A metric $m_2 \in \mathcal{E}_{g,n}$ is said to be a *deformation* of the metric $m_1 \in \mathcal{E}_{g,n}$ if the two metrics differ by a homothety and an isotopy fixing the punctures, i.e. if there exists $h \in \text{PDiff}_{g,n}^0$ and $\lambda > 0$ such that $m_2 = \lambda h^*(m_1)$. We denote by $\mathcal{D}\mathcal{E}_{g,n}$ the deformation space of flat metrics on $\Sigma_{g,n}$ with possible conical singularities at the punctures:

$$\mathcal{D}\mathcal{E}_{g,n} = \mathcal{E}_{g,n} / (\mathbb{R}_+ \times \text{PDiff}_{g,n}^0).$$

Corollary 5.3. *This space is homeomorphic to $\mathbb{R}^{6g+3n-7}$. In fact we have the following canonical identification:*

$$\mathcal{D}\mathcal{E}_{g,n} = \mathcal{T}_{g,n} \times \Delta,$$

where $\mathcal{T}_{g,n}$ is the Teichmüller space and $\Delta \subset \mathbb{R}^n$ is defined by

$$\Delta = \{ \vec{\beta} = (\beta_1, \dots, \beta_n) \in \mathbb{R}^n \mid \beta_i > -1 \text{ and } \sum_i \beta_i = 2g - 2 \}.$$

Let us fix an element $\vec{\beta} = (\beta_1, \dots, \beta_n) \in \Delta$ and denote by $\mathcal{E}_{g,n}(\vec{\beta})$ the space of singular flat metrics with a conical singularity of order β_i at p_i ($i = 1, \dots, n$). We also introduce the corresponding deformation space: $\mathcal{D}\mathcal{E}_{g,n}(\vec{\beta}) = \mathcal{E}_{g,n}(\vec{\beta}) / (\mathbb{R}_+ \times \text{PDiff}_{g,n})$. The previous corollary gives us the identification

$$\mathcal{D}\mathcal{E}_{g,n}(\vec{\beta}) = \mathcal{T}_{g,n}.$$

5.3 Revisiting the developing map and the holonomy

Consider the punctured surface $\Sigma_{g,n} = \hat{U} / \Gamma$ as in Section 3.2, and fix a flat metric m_0 with conical singularity of order β_j at p_j ($j = 1, \dots, n$). If f_0 is a germ of an

isometry near a point \tilde{z}_0 , to the euclidean plane (identified with \mathbb{C}), then we obtain a map $f: \mathbb{U} \rightarrow \mathbb{C}$ by analytic continuation from f_0 . This map is a local isometry for the metric m_0 on \mathbb{U} and the canonical metric on \mathbb{C} (indeed, the set of points where a map f between two flat surfaces is an isometry is easily seen to be both open and closed). The map f extends by continuity to $\hat{\mathbb{U}}$. The resulting map $f: \hat{\mathbb{U}} \rightarrow \mathbb{C}$ is the developing map which we already met in Section 2.3. The associated holonomy is the unique homomorphism $\varphi: \Gamma \rightarrow \text{SE}(2)$ such that

$$f(\gamma u) = \varphi(\gamma) f(u).$$

Theorem 5.4. *The following properties of the developing map and its holonomy are satisfied:*

- (1) $f: \hat{\mathbb{U}} \rightarrow \mathbb{C}$ is surjective;
- (2) $f(\gamma u) = \varphi(\gamma) f(u)$ for all $\gamma \in \Gamma$;
- (3) for any $y \in P^{-1}(p_i) \subset \Upsilon$, the isometry $\varphi(c_y) \in \text{SE}(2)$ is a rotation of angle $\theta_i = 2\pi(\beta_i + 1)$;
- (4) if β_i is not an integer, then $f(y)$ is the unique fixed point of $\varphi(c_y)$.

Proof. The first assertion has been proved in Proposition 2.10, the second is the definition of the holonomy and the last two assertions are contained in Proposition 2.15. \square

Corollary 5.5. *If $\beta_i \notin \mathbb{Z}$ for any $i = 1, \dots, n$, then the restriction of f to the set Υ is determined by the holonomy.*

Proof. Fix $y \in \Upsilon$ and let $c_y \in \Gamma$ be the corresponding group element given by the map (3.5). Then $f(y) \in \mathbb{C}$ is the fixed point of the rotation $\varphi(c_y) \in \text{SE}(2)$. This fixed point is given explicitly by

$$f(y) = \frac{\tau_y}{(1 - \rho_y)}, \quad (5.1)$$

where $\rho_y \in U(1)$ is the rotation part and $\tau_y \in \mathbb{C}$ is the translation part of $\varphi(c_y)$. \square

Theorem 5.6. (A) *There is a well-defined map*

$$\text{hol}: \mathcal{DE}_{g,n} \rightarrow \mathcal{R}(\pi_{g,n}, \text{SE}(2)),$$

such that $\text{hol}([m])$ is the conjugacy class of the holonomy homomorphism $\varphi_m: \pi_{g,n} \rightarrow \text{SE}(2)$.

(B) *The map $\text{hol}: \mathcal{DE}_{g,n} \rightarrow \mathcal{R}(\pi_{g,n}, \text{SE}(2))$ is continuous.*

(C) *There are natural actions of $\text{PMod}_{g,n}$ on $\mathcal{DE}_{g,n}$ and of $\text{POut}(\pi_{g,n})$ on $\mathcal{R}(\pi, \text{SE}(2))$, and the map hol is Φ -equivariant where Φ is the Dehn–Nielsen–Baer isomorphism (3.6).*

(D) *The map hol is locally injective.*

Remarks. 1.) The map $\text{hol}: \mathcal{DE}_{g,n} \rightarrow \mathcal{R}(\pi_{g,n}, \text{SE}(2))$ is called the *holonomy mapping*.

2.) A more elaborate investigation would show that the holonomy mapping is in fact real analytic, see [35]. The proof below is perhaps not optimal from the point of view of rigour, but we have tried to emphasize the geometric point of view.

Proof of Theorem 5.6. (A) To any flat metric m on $\Sigma_{g,n}$ with conical singularities at the punctures, we have associated a holonomy homomorphism $\varphi_m: \pi_{g,n} \rightarrow \text{SE}(2)$ which depends on the choice of a developing map f_m , but changing the developing map does not affect the conjugacy class of φ_m (see Proposition 2.14). On the other hand, it is clear that if two flat metrics m, m' on $\Sigma_{g,n}$ are similar, the associated holonomies $\varphi_m, \varphi_{m'}$ are also similar. In short, to any deformation class of flat metric $[m] \in \mathcal{DE}_{g,n}$ with conical singularities on $\Sigma_{g,n}$ we associate a well defined element $[\varphi_m] = \text{hol}(m) \in \mathcal{R}(\pi_{g,n}, \text{SE}(2))$.

(B) The developing map f_m is not uniquely associated to a flat metric m , but it is well defined modulo $\text{SE}(2)$ (two developing maps for the same metric differ by postcomposition with an isometry). The $\text{SE}(2)$ orbit of the developing map f_m varies continuously with the metric m and therefore it is also the case for the associated holonomy class. Hence the map $\text{hol}: \mathcal{DE}_{g,n} \rightarrow \mathcal{R}(\pi_{g,n}, \text{SE}(2))$ is continuous.

(C) Any diffeomorphism h of $\Sigma_{g,n}$ fixing the punctures acts on $\mathcal{E}_{g,n}$ by pulling back the metric ($m \mapsto h^*m$). If h is isotopic to the identity, it acts trivially on $\mathcal{DE}_{g,n}$; we thus have a well defined action of $\text{PMod}_{g,n}$ on $\mathcal{DE}_{g,n}$.

Similarly, any automorphism of $\pi_{g,n}$ acts on $\text{Hom}(\pi_{g,n}, \text{SE}(2))$, and inner automorphisms act trivially on the two representation spaces $\mathcal{R}(\pi_{g,n}, \text{SE}(2))$ and $\mathcal{R}(\pi_{g,n}, \text{SE}(2))$. We thus have a natural action of $\text{POut}(\pi_{g,n})$ on these spaces. It is clear from the construction of the isomorphism $\Phi: \text{PMod}_{g,n} \xrightarrow{\sim} \text{POut}(\pi_{g,n})$ (see the proof of Lemma 3.3) that the map hol is equivariant.

(D) To prove the local injectivity of hol , we consider two nearby flat metrics m, m' with conical singularities on $\Sigma_{g,n} = \hat{\mathcal{U}}/\Gamma$ and we assume that they have the same holonomy φ . Since the holonomy around a conical singularity p_i is a rotation of angle $\theta_i = 2\pi(\beta_i + 1)$, it is clear that both metrics m and m' have the same singularity order (the holonomy only controls the cone angle modulo 2π , but since m and m' are nearby metrics, they actually have equal cone angles).

It follows that both metrics are isometric near the singularities: we can thus find an isotopy h_1 of the surface such that $m = h_1^*m'$ near the singularities. Hence we can simply assume without loss of generality that $m = m'$ near the singularities; it is therefore possible to divide the surface in $n + 1$ parts

$$\Sigma_{g,n} = D \cup E_1 \cup \dots \cup E_n,$$

where $D \subset \Sigma'$ is a compact region and E_i is a neighbourhood of the puncture p_i such that $m = m'$ on E_i . We also assume that the E_i are pairwise disjoint disks. We denote by $\hat{E}_i = P^{-1}(E_i) \subset \hat{\mathcal{U}}$ and $\hat{D} = P^{-1}(D) \subset \hat{\mathcal{U}}$ the lifts of E_i and D on the

universal branched cover $P: \hat{U} \rightarrow \Sigma_{g,n}$. The set $\hat{E} = \cup_i \hat{E}_i \subset \hat{U}$ is a neighbourhood of $\Upsilon = P^{-1}(\{\text{punctures}\})$.

Let f_m and $f_{m'}$ be the developing maps of m and m' . By Corollary 5.5, the two maps coincide on Υ . Because the two metrics coincide on E_i , the map f_m and $f_{m'}$ coincide up to a rotation on each component of \hat{E}_i ; we can thus find a second isotopy h_2 of Σ , which is a rotation near the punctures and is the identity on D and such that $f_{m'} \circ \hat{h}_2 = f_m$ on \hat{E} .

Replacing m' with h_2^*m' , we can thus assume that both developing maps coincide on \hat{E} .

To any point $x \in U$, we associate the set

$$\Lambda(m, m', x) = f_m^{-1}(f_{m'}(x)) \subset \hat{U}.$$

Since f_m is a local diffeomorphism, $\Lambda(m, m', x)$ is a discrete set. It varies continuously with m, m' .

Claim. If m is close enough to m' , then for any point $x \in U$, there exists a unique point $y = Q(x) \in \Lambda(m, m', x)$ which is the nearest point for the hyperbolic distance. The map $x \mapsto Q(x)$ is Γ -equivariant.

Indeed, if $x \in \hat{E}$, then the claim is clear: since $f_m(x) = f_{m'}(x)$, we have $Q(x) = x$. For any point, the claim is clear if $m = m'$ (and in this case $Q(x) = x$). For points in \hat{D} , and m' close to m the claim follows from the compactness of D and the discreteness and continuity of $\Lambda(x, m, m')$.

For $t \in [0, 1]$, we denote by $Q_t(x)$ the point on the hyperbolic segment $[x, Q(x)]$ such that $d_H(x, Q_t(x)) = t d_H(x, Q(x))$ (observe that if $x \in \hat{E}$, then $Q_t(x) = x$ for any t). This is a Γ -equivariant isotopy of U from the identity to Q . It extends as the identity on Υ .

We now define an isotopy $h_t: \Sigma \rightarrow \Sigma$ by $h_t(x) = P(Q_t(P^{-1}(x)))$. It is a well defined isotopy such that $h_1^*m' = m$, since we clearly have $f_m = f_{m'} \circ Q$.

We thus have proved that two metrics with the same holonomy are isotopic provided they are close enough. In other words, the map hol is locally injective. \square

6 The main theorem

We are now in position to prove the main result. First recall the statement:

Theorem 6.1. *Given a punctured surface $\Sigma_{g,n}$ such that $2g + n - 2 > 0$ and $\vec{\beta} \in \Delta$ such that no β_i is an integer, there is a well defined group homomorphism*

$$\Phi: \text{PMod}_{g,n} \rightarrow \mathfrak{G} = \text{Aut}(\mathbb{T}^{2g}) \times \text{PGL}_{2g+n-2} \mathbb{C},$$

and a Φ -equivariant local homeomorphism

$$\mathcal{H}: \mathcal{T}_{g,n} \rightarrow \mathfrak{E} = \mathbb{T}^{2g} \times \mathbb{C}\mathbb{P}^{2g+n-3}.$$

The theorem says that $\mathcal{M}_{g,n} = \mathcal{T}_{g,n}/\text{PMod}_{g,n}$ is a good orbifold with a (\mathcal{G}, Ξ) -structure.

Proof. The group homomorphism Φ is given by the Dehn–Nielsen–Baer isomorphism and the map \mathcal{H} is essentially given by the holonomy mapping of the previous theorem. We divide the proof of the theorem in five steps:

Recall that holonomy splits in a rotation part $\rho_m: \pi \rightarrow U(1)$ (the character) and a translation part τ_m . The character depends only on the conjugacy class of φ_m .

Step 1. τ_m is not identically zero.

Indeed, if $\tau_m \equiv 0$, then the holonomy group $\varphi_m(\pi_{g,n})$ is a pure rotation group in the plane. This is impossible by Proposition 2.13.

Step 2. There is a canonical isomorphism

$$\begin{aligned} \text{Hom}(\pi_{g,n}, U(1)) &\simeq \text{Hom}(\pi_{g,0}, U(1)) \times \text{Hom}(\pi_{0,n}, U(1)), \\ \rho_m &\mapsto (\rho', \rho''). \end{aligned}$$

Furthermore $\rho'' \in \text{Hom}(\pi_{0,n}, U(1))$ is given by

$$\rho''(c_i) = e^{\theta_i},$$

where c_i is the homotopy class of a loop traveling once around the puncture p_i and $\theta_i = 2\pi(\beta_i + 1)$ is the total angle at the cone point p_i .

This splitting easily follows from the identity (3.1) and the fact that $U(1)$ is abelian.

Let us now fix an element $\vec{\beta} = (\beta_1, \dots, \beta_n) \in \Delta$ and set

$$\mathcal{R}_{\vec{\beta}}(\pi_{g,n}, \text{SE}(2)) = \{ \varphi \in \mathcal{R}(\pi_{g,n}, \text{SE}(2)) \mid \rho''(c_i) = e^{\theta_i}, i = 1, \dots, n \}$$

and

$$\mathcal{R}_{\vec{\beta}}^{\text{reg}} = \mathcal{R}^{\text{reg}} \cap \mathcal{R}_{\vec{\beta}}.$$

Step 3. If at least one β_i is not an integer, then we have

$$\mathcal{R}_{\vec{\beta}}^{\text{reg}}(\pi_{g,n}, \text{SE}(2)) \simeq \Xi = \mathbb{T}^{2g} \times \mathbb{C}\mathbb{P}^{2g+n-3}.$$

Indeed, it follows from Step 2 and the results of Section 4 that any $\varphi \in \mathcal{R}_{\vec{\beta}}^{\text{reg}}$ is characterized by $\rho' \in \text{Hom}(\pi_{g,0}, U(1)) \simeq \mathbb{T}^{2g}$ and the projective class of $\tau \in H^1(\pi_{g,n}, \mathbb{C}_\rho) \simeq \mathbb{C}^{2g+n-2}$ (because $\pi_{g,n}$ is isomorphic to the free group on $s = 2g + n - 1$ generators).

Step 4. The group $\text{POut}(\pi_{g,n})$ acts naturally on Ξ and thus we have a natural homomorphism $\Phi: \text{PMod}_{g,n} \rightarrow \mathcal{G} = \text{Aut}(\mathbb{T}^{2g}) \times \text{PGL}_{2g+n-2} \mathbb{C}$.

This is clear from Step 3 and Part (C) of Theorem 5.6.

Step 5. The map \mathcal{H} given by the composition:

$$\mathcal{T}_{g,n} \xrightarrow{\sim} \mathcal{DE}_{g,n}(\vec{\beta}) \xrightarrow{\text{hol}} \mathcal{R}_{\vec{\beta}}^{\text{reg}} \xrightarrow{\sim} \Xi$$

is well defined, continuous, locally injective and Φ -equivariant.

Indeed, the fact that no β_i is integer, taken together with Step 1, implies that $\text{hol}: \mathcal{DE}_{g,n} \rightarrow \mathcal{SR}(\pi, \text{SE}(2))$ maps $\mathcal{DE}_{g,n}(\vec{\beta})$ into $\mathcal{SR}_{\vec{\beta}}^{\text{reg}}$. The map $\mathcal{H}: \mathcal{T}_{g,n} \rightarrow \Xi$ is therefore well defined. It follows from Theorem 5.6 that \mathcal{H} is continuous, locally injective and Φ -equivariant.

It remains only to show that \mathcal{H} is a local homeomorphism, but since $\mathcal{T}_{g,n}$ and Ξ are both manifolds of dimension $6g - 6 + 2n$, the conclusion follows from Brouwer's theorem on invariance of dimension. \square

6.1 The case of the sphere

Suppose that $g = 0$, i.e. Σ is a sphere, choose n numbers ($n \geq 2$) $\beta_1, \beta_2, \dots, \beta_n$ such that $2 + \sum_i \beta_i = 0$, and denote by \mathcal{M} the space of flat metrics on S^2 having n conical singularities of order $\beta_1, \beta_2, \dots, \beta_n$.

Such a metric $m \in \mathcal{M}$ can be uniformized as follows: identify Σ with $\mathbb{C} \cup \infty$, and write m as

$$m = C \cdot \prod_{i=1}^n |z - p_i|^{2\beta_i} |dz|^2,$$

where p_1, p_2, \dots, p_n is the set of conical singularities and C is a positive constant representing a dilation factor. It is easy to see that \mathcal{M} is homeomorphic to the quotient

$$\{(p_1, p_2, \dots, p_n) \in (\mathbb{C} \cup \infty)^n : p_i \neq p_j \text{ if } i \neq j\} / \text{PSL}_2(\mathbb{C}),$$

\mathcal{M} is thus a complex manifold of dimension $n - 3$, its fundamental group is the pure braid group PB_n .

Applying the main theorem, we obtain a representation

$$\Phi: \text{PMod}_{0,n} = \text{PB}_n \rightarrow \text{PGL}_{n-2}(\mathbb{C})$$

and a Φ equivariant, local homeomorphism

$$\mathcal{H}: \mathcal{T}_{0,n} \rightarrow \mathbb{CP}^{n-3}.$$

In fact, a finer analysis shows that the image of Φ is contained in $\text{PU}(1, n - 3) \subset \text{PGL}_{n-2}(\mathbb{C})$. Furthermore, when the orders satisfy some arithmetical conditions, the image of Φ is a lattice in $\text{PU}(1, n - 3)$:

Theorem 6.2. *Assume that $-1 < \beta_1, \beta_2, \dots, \beta_n < 0$, $\sum_i \beta_i = -2$ and suppose that*

$$\beta_i + \beta_j > -1 \Rightarrow (1 + \beta_j + \beta_i)^{-1} \in \mathbb{N}, \quad (6.1)$$

then $\Phi(\text{PB}_n)$ is a lattice in $\text{PU}(1, n - 3)$.

These lattices are quotients of the braid group. Some of them are non arithmetic.

This theorem was first proved by Schwartz (1873) for $n = 4$ and by Picard (1888) for $n = 5$ in their study of the monodromy of the hypergeometric equations. It has been generalized for any n by P. Deligne and G. Mostow in 1986, see [10]. These

authors use the cohomology with coefficients in flat vector bundle on an algebraic curve.

In the paper [30], W. Thurston obtained the same result by studying a deformation space of piecewise flat triangulations on the sphere (this nice paper is a 1987 preprint of W. Thurston, which has been rewritten and appeared in electronic form in 1998). It is worthwhile to quote also the related papers [3], [13], [25], [21].

Our approach can be seen as a bridge between the approach of Thurston and that of Deligne–Mostow.

Observe that the moduli space $\mathcal{M} = \mathcal{T}_{0,n}/\text{PB}_n$ carries a complex hyperbolic metric (depending upon the choice of the β_i 's). It is not complete as a Riemannian manifold and it carries a natural completion $\bar{\mathcal{M}}$. Thurston shows that $\bar{\mathcal{M}}$ is a complex hyperbolic manifold with singularities of conical type. This cone-manifold has finite volume.

Furthermore, when the β_i 's satisfy the condition (6.1), then $\bar{\mathcal{M}}$ is an orbifold. It is thus possible to construct complete complex hyperbolic orbifolds $\bar{\mathcal{M}}$ of finite volume.

References

- [1] W. Abikoff, W. Abikoff, *The real analytic theory of Teichmüller space*. Lecture Notes in Math. 820, Springer, Berlin 1980. [531](#)
- [2] J. Ambjorn, M. Carfora, A. Marzuoli, *The geometry of dynamical triangulations*. Lecture Notes in Phys. New Ser. m Monogr. 50, Springer-Verlag, Berlin, 1997. [509](#)
- [3] C. Bavard and E. Ghys, Polygones du plan et polyèdres hyperboliques. *Geom. Dedicata* 43 (2) (1992) 207–224. [538](#)
- [4] J. S. Birman, *Braids, links and mapping class groups*. Ann. of Math. Stud. 82, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo 1974. [521](#)
- [5] B. H. Bowditch. Singular euclidean structures on surfaces. *J. London Math. Soc.* 44 (1991), 553–565. [509](#)
- [6] D. Burago, Yu. Burago and S. Ivanov, *A course in metric geometry*. Grad. Stud. Math. 33, Amer. Math. Soc., Providence, RI, 2001 [510](#)
- [7] R. D. Canary, D. B. A. 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. [509](#)
- [8] M. Carfora, C. Dappiaggi, and A. Marzuoli, The modular geometry of random Regge triangulations. *Classical Quantum Gravity* 19 (20) (2002), 5195–5220. [509](#)
- [9] S. Choi, Geometric structures on orbifolds and holonomy representations. *Geom. Dedicata* 104 (2004), 161–199. [509](#)
- [10] P. Deligne and G. Mostow, Monodromy of hypergeometric functions and non lattice integral monodromy. *Inst. Hautes Études Sci. Publ. Math.* 63 (1986), 5–89. [509](#), [537](#)
- [11] D. B. A. Epstein and R. C. Penner, Euclidean decompositions of noncompact hyperbolic manifolds. *J. Differential Geom.* 27 (1) (1988), 67–80. [509](#)

- [12] A. Eremenko, Metrics of positive curvature with conic singularities on the sphere. *Proc. Amer. Math. Soc.* 132 (11) (2004), 3349–3355 [518](#)
- [13] F. Fillastre, From spaces of polygons to spaces of polyhedra following Bavard, Ghys and Thurston. Paper on arXiv:math.MG/0308187 [538](#)
- [14] 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. [509](#), [524](#)
- [15] M. Heins, On a class of conformal metrics. *Nagoya Math. J.* 21 (1962), 1–60. [518](#)
- [16] D. Hulin and M. Troyanov, Sur la courbure des surfaces ouvertes. *C. R. Acad. Sci. Paris* 310 (1990) 203–206. [520](#)
- [17] D. Hulin and M. Troyanov, Prescribing curvature on open surfaces. *Math. Ann.* 293 (1992), 277–315. [518](#), [520](#)
- [18] C. Indermitte, T. M. Lieblich, H. Clémençon and Marc Troyanov, Voronoi Diagrams on Piecewise flat surfaces and an application to biological growth. *Theoret. Comput. Sci.* 263 (2001) 263–274. [517](#)
- [19] N. V. Ivanov, Mapping class groups. In *Handbook of geometric topology* (R. J. Daverman and R. B. Sher, eds.) North-Holland, Amsterdam 2002, 523–633. [521](#), [522](#), [531](#)
- [20] S. Katok, *Fuchsian groups*. Chicago Lectures in Math., University of Chicago Press, Chicago, IL, 1992. [522](#), [523](#)
- [21] S. Kojima, Complex hyperbolic cone structures on the configuration spaces. *Rend. Istit. Mat. Univ. Trieste* 32 (2001), suppl. 1 149–163. [538](#)
- [22] S. Kravetz, On the geometry of Teichmüller spaces and the structure of their modular groups. *Ann. Acad. Sci. Fenn. Ser. A1* 278, (1959) [531](#)
- [23] R. McOwen, Point singularities and conformal metrics on Riemann surfaces. *Proc. Amer. Math. Soc.* 103 (1988), 222–224. [518](#)
- [24] 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. [521](#)
- [25] G. Mostow, Braids, hypergeometric functions, and lattices. *Bull. Amer. Math. Soc.* 16 (2) (1987), 225–246. [538](#)
- [26] É. Picard, De l'intégration de l'équation $\Delta u = e^u$ sur une surface de Riemann fermée. *J. Reine Angew. Math.*, 130 (1905), 243–258. [518](#)
- [27] I. Rivin, Combinatorial optimization in geometry. *Adv. in Appl. Math.* (1) 31 (2003), 242–271. [509](#), [517](#)
- [28] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*. Publications of the Mathematical Society of Japan 11, Iwanami Shoten Publishers, Tokyo; Princeton University Press, Princeton, N.J., 1971. [519](#), [520](#)
- [29] W. Thurston, *Three-dimensional geometry and topology*. Vol. 1. Princeton University Press, Princeton, NJ, 1997 [509](#)
- [30] W. Thurston, *Shapes of polyhedra and triangulations of the sphere*, The Epstein birthday schrift, Geom. Topol. Monogr. 1, Geom. Topol. Publ., Coventry 1998, 511–549 (electronic). [509](#), [517](#), [538](#)

- [31] M. Troyanov, Les surfaces euclidiennes à singularités coniques. *Enseign. Math.* (2) 32 (1986), 79–94. [511](#), [517](#), [518](#)
- [32] M. Troyanov, Prescribing curvature on compact surfaces with conical singularities. *Trans. Amer. Math. Soc.* 324 (1991), 793–821. [518](#)
- [33] M. Troyanov, Surfaces Riemanniennes à singularités simples. *Proc. Sympos. Pure Math.* 54 (Part 2) (1993), 619–628. [518](#)
- [34] M. Umehara and K. Yamada, Metrics of constant curvature 1 with three conical singularities on 2-sphere. *Illinois J. Math.* 44 (2000), 72–94. [518](#)
- [35] W. A. Veech, Flat surfaces. *Amer. J. Math.* 115 (3) (1993), 589–689. [509](#), [534](#)
- [36] H. Zieschang, E. Vogt, H.-D. Coldeway, *Surfaces and planar discontinuous groups*. Lecture Notes in Math. 835, Springer-Verlag, Berlin 1980. [521](#), [522](#)

Chapter 13

Discrete Riemann surfaces

Christian Mercat

Contents

1	Introduction	541
2	Discrete Riemann surfaces	542
2.1	Discrete Hodge theory	542
2.2	Wedge product	545
2.3	Energies	547
3	Period matrix	548
3.1	Intersection number, on Λ and on \diamond diamondsuit	548
3.2	Canonical dissection, fundamental polygon	548
3.3	Bilinear relations	552
3.4	Basis of harmonic forms, basis of holomorphic forms	552
3.5	Period matrix	555
4	Criticality and integrable system	556
4.1	Criticality	556
4.2	Period matrix	557
4.3	Genus one case	559
4.4	$\bar{\partial}$ operator	560
4.5	Geometrical interpretation	561
4.6	Quasiconformal maps	562
4.7	Discrete exponential	563
4.8	Series	564
4.9	Basis	565
4.10	Continuous limit	566
4.11	Cross-ratio preserving maps	569
4.12	Bäcklund transformation	571
	References	574

1 Introduction

Riemann surfaces theory was a major achievement of XIXth century mathematics, setting the framework where modern complex analysis bloomed. Nowadays, surfaces are intensively used in computer science for numeric computations, ranging from

visualization to pattern recognition and approximation of partial differential equations. A lot of these computations involve, at the continuous level, analytic functions. But very few algorithms care about this specificity, although analytic functions form a relatively small vector space among the space of functions, problems are usually crudely discretized in a way that does not take advantage of analyticity.

The theory of discrete Riemann surfaces aims at filling this gap and setting the theoretical framework in which the notion of discrete analyticity is set on solid grounds.

Most of the results in this chapter are a straightforward application of the continuous theory [7], [14] together with the results in [10], [11], [12], to which we refer for details. We define the discrete period matrix, which is twice as large as in the continuous case: the periods of a holomorphic form on the graph and on its dual are in general different, but the continuous limit theorem, given a refining sequence of critical maps, ensures that they converge to the same value. The last section deals with criticality, defining the framework in which the continuous limit theorem takes place, and integrability, defined as a 3D-consistency of the equations ruling not only the linear discrete holomorphy but a quadratic version as well.

2 Discrete Riemann surfaces

2.1 Discrete Hodge theory

We recall in this section basic definitions and results from [11] where the notion of discrete Riemann surfaces was defined. We are interested in discrete oriented surfaces given by a cellular decomposition \diamond of dimension two, where all faces are *quadrilaterals* (a quad-graph [8], [2], [1]). We will assume that \diamond is bipartite with black vertices connected to white vertices. The diagonals of \diamond connect vertices of

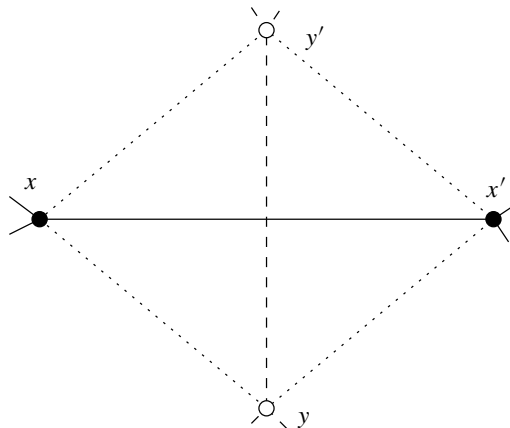
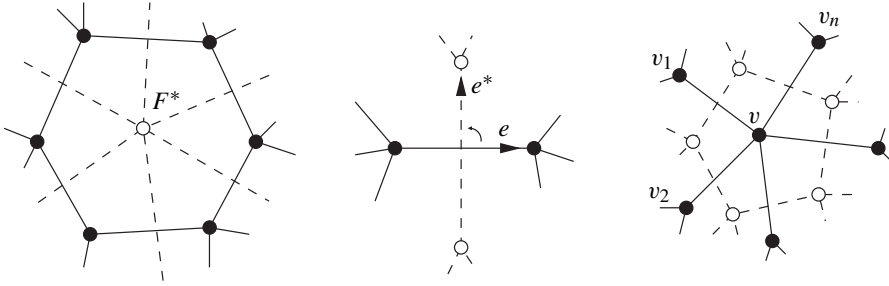


Figure 1. The vertices and diagonals of a quadrilateral define a pair of dual edges.

the same type and define, up to homotopy and away from the boundary, two dual cellular decompositions Γ and Γ^* : The edges in Γ^* are dual to edges in Γ , faces in



1. The vertex F^* dual to a face F . 2. Dual edges. 3. The face v^* dual to a vertex v .

Figure 2. Duality.

Γ_2^* are dual to vertices in Γ_0 and vice-versa. This correspondance shows that there is no lack of generality in considering quad-graphs. Their disjoint union is denoted the *double* $\Lambda = \Gamma \sqcup \Gamma^*$. A *discrete conformal structure* on Λ is a real positive function ρ on the unoriented edges satisfying $\rho(e^*) = 1/\rho(e)$. It defines a genuine Riemann surface structure on the discrete surface: Choose a length δ and realize each quadrilateral by a lozenge of side δ , whose diagonals have a length ratio given by ρ . Gluing them together provides a flat Riemannian metric with conic singularities at the vertices, hence a conformal structure [16]. Such a discrete conformal structure leads to a straightforward discrete version of the *Cauchy–Riemann equation*. A function on the vertices of \diamond is *discrete holomorphic* iff for every quadrilateral $(x, y, x', y') \in \diamond_2$,

$$f(y') - f(y) = i \rho(x, x') (f(x') - f(x)). \tag{2.1}$$

We recall elements of de-Rham cohomology, doubled in our context: The complex of *chains* $C(\Lambda) = C_0(\Lambda) \oplus C_1(\Lambda) \oplus C_2(\Lambda)$ is the vector space (over \mathbb{R} the field of reals) spanned by vertices, edges and faces. It is equipped with a *boundary* operator $\partial: C_k(\Lambda) \rightarrow C_{k-1}(\Lambda)$, null on vertices and fulfilling $\partial^2 = 0$. The kernel $\text{Ker } \partial := Z_\bullet(\Lambda)$ of the boundary operator is the set of closed chains or *cycles*. The image of the boundary operator are the *exact* chains. The dual spaces of forms are called *cochains*, $C^k(\Lambda) := \text{Hom}(C_k(\Lambda), \mathbb{R})$. Coupling is denoted by functional and integral notation. The dual of the boundary operator is called the *coboundary* operator $d: C^k(\Lambda) \rightarrow C^{k+1}(\Lambda)$, defined by Stokes formula:

$$\int_{(x,x')} df := f(\partial(x, x')) = f(x') - f(x), \quad \iint_F d\alpha := \oint_{\partial F} \alpha.$$

A *cocycle* is a closed cochain and we note $\alpha \in Z^k(\Lambda)$.

These spaces are equipped with the canonical scalar product, weighed according to ρ on edges and averaged on the graph and its dual:

$$(\alpha, \beta)_{\mathbb{R}} := \frac{1}{2} \sum_{e \in \Lambda_1} \rho(e) \left(\int_e \alpha \right) \left(\int_e \beta \right). \tag{2.2}$$

Duality of complexes allows us to define a *Hodge operator* $*$ on forms by

$$\begin{aligned} * : C^k(\Lambda) &\rightarrow C^{2-k}(\Lambda); \\ C^0(\Lambda) \ni f &\mapsto *f : \iint_F *f := f(F^*), \\ C^1(\Lambda) \ni \alpha &\mapsto *\alpha : \int_e *\alpha := -\rho(e^*) \int_{e^*} \alpha, \\ C^2(\Lambda) \ni \omega &\mapsto *\omega : (*\omega)(x) := \iint_{x^*} \omega. \end{aligned} \tag{2.3}$$

It fulfills $*^2 = (-\text{Id}_{C^k})^k$. The endomorphism $\Delta := -d * d * - * d * d$ is the usual discrete *Laplacian*: It decomposes onto the two independent graphs and its formula on a function at a vertex $x \in \Gamma_0$ with neighbours $x_1, \dots, x_V \in \Gamma_0$ is the usual weighted averaged difference:

$$(\Delta(f))(x) = \sum_{k=1}^V \rho(x, x_k) (f(x) - f(x_k)).$$

The space of *harmonic forms* is defined as its kernel $\partial * = \text{Ker} \Delta$.

The Hodge star and the Laplacian are real operators. Since $*^2 = -\text{Id}$ on functions, it is natural to consider them on complexified cochains, equipped with the hermitian version of the scalar product (2.2). The discrete holomorphic forms are special complex harmonic forms: a 1-form

$$\alpha \in C^1(\Lambda) \text{ is holomorphic iff } d\alpha = 0 \text{ and } *\alpha = -i\alpha, \tag{2.4}$$

that is to say if it is closed and of type $(1, 0)$. Let d' , resp. d'' the compositions of the exterior derivative with the projection on the space of $(1, 0)$, resp. $(0, 1)$ -forms,

$$d' := \pi_{(1,0)} \circ d : C^0(\Lambda) \rightarrow C^1(\Lambda), \quad d' := d \circ \pi_{(1,0)} : C^1(\Lambda) \rightarrow C^2(\Lambda)$$

and similarly for d'' and $\pi_{(0,1)}$.

Definition (2.4) is equivalent to $d'\alpha = 0$. We will note $\alpha \in \Omega^1(\Lambda)$. A function $f : \Lambda_0 \rightarrow \mathbb{C}$ is *holomorphic* iff df is holomorphic, which is equivalent to (2.1) and we note $f \in \Omega^0(\Lambda)$.

In the compact case, $- * d *$ is the adjoint d^* of the coboundary operator d and the Hodge theorem orthogonally decomposes forms into exact, coexact and harmonic,

$$C^k(\Lambda) = \text{Im } d \oplus^\perp \text{Im } d^* \oplus^\perp \text{Ker } \Delta,$$

harmonic forms are the closed and co-closed ones, and harmonic 1-form are the orthogonal sum of holomorphic and anti-holomorphic ones:

$$\text{Ker } \Delta = \text{Ker } d \cap \text{Ker } d^* = \text{Ker } d' \oplus^\perp \text{Ker } d''.$$

2.2 Wedge product

We consider similarly $C^\bullet(\diamond)$ cochains on the graph \diamond . We construct a wedge product on \diamond such that

- the canonical weighted hermitian scalar product reads as expected

$$(\alpha, \beta) = \iint \alpha \wedge * \bar{\beta},$$

- and the coboundary operator d_\diamond on \diamond , is a derivation for this product $\wedge: C^k(\diamond) \times C^l(\diamond) \rightarrow C^{k+l}(\diamond)$.

It is defined by the following formulae, for $f, g \in C^0(\diamond)$, $\alpha, \beta \in C^1(\diamond)$ and $\omega \in C^2(\diamond)$:

$$(f \cdot g)(x) := f(x) \cdot g(x) \quad \text{for } x \in \diamond_0, \tag{2.5}$$

$$\int_{(x,y)} f \cdot \alpha := \frac{f(x) + f(y)}{2} \int_{(x,y)} \alpha \quad \text{for } (x, y) \in \diamond_1, \tag{2.6}$$

$$\iint_{(x_1,x_2,x_3,x_4)} \alpha \wedge \beta := \frac{1}{4} \sum_{k=1}^4 \int_{(x_{k-1},x_k)} \alpha \int_{(x_k,x_{k+1})} \beta - \int_{(x_{k+1},x_k)} \alpha \int_{(x_k,x_{k-1})} \beta, \tag{2.7}$$

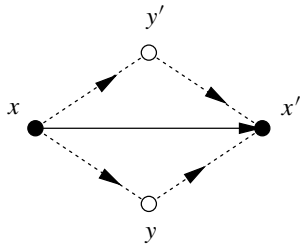
$$\iint_{(x_1,x_2,x_3,x_4)} f \cdot \omega := \frac{f(x_1)+f(x_2)+f(x_3)+f(x_4)}{4} \iint_{(x_1,x_2,x_3,x_4)} \omega \quad \text{for } (x_1, x_2, x_3, x_4) \in \diamond_2. \tag{2.8}$$

A form on \diamond can be *averaged* into a form on Λ : This map A from $C^\bullet(\diamond)$ to $C^\bullet(\Lambda)$ is the identity for functions and defined by the following formulae for 1 and 2-forms:

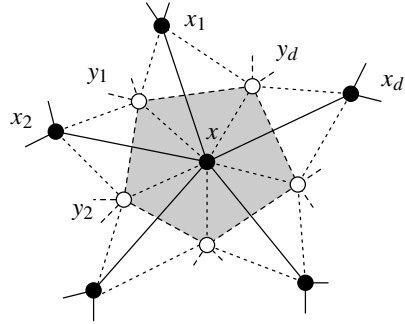
$$\int_{(x,x')} A(\alpha_\diamond) := \frac{1}{2} \left(\int_{(x,y)} + \int_{(y,x')} + \int_{(x,y')} + \int_{(y',x')} \right) \alpha_\diamond, \tag{2.9}$$

$$\iint_{x^*} A(\omega_\diamond) := \frac{1}{2} \sum_{k=1}^d \iint_{(x_k,y_k,x,y_{k-1})} \omega_\diamond, \tag{2.10}$$

where notations are made clear in Figure 3. The map A is neither injective nor surjective in the non simply-connected case, so we can neither define a Hodge star on \diamond nor a wedge product on Λ . Its kernel on 1-forms is the line spanned by the coboundary of the bi-constant: $\text{Ker}(A) = \text{Span}(d_\diamond \varepsilon)$, where ε takes the value $+1$ on the vertices of Γ and -1 on the vertices of Γ^* . But $d_\Lambda A = A d_\diamond$ so it carries



(2.9)



(2.10)

Figure 3. Notations.

cocycles on \diamond to cocycles on Λ . Its image are these cocycles of Λ verifying that their holonomies along cycles of Λ only depend on their homology on the combinatorial surface.

Given a 1-cocycle $\mu \in Z^1(\Lambda)$ with such a property, a corresponding 1-cocycle $v \in Z^1(\diamond)$ is built in the following way: Choose an edge $(x_0, y_0) \in \diamond_1$; for an edge $(x, y) \in \diamond_1$ with x and x_0 on the same leaf of Λ , choose two paths λ_{x,x_0} and $\lambda_{y_0,y}$ on the double graph Λ , from x to x_0 and y_0 to y respectively, and define

$$\int_{(x,y)} v := \int_{\lambda_{x,x_0}} \mu + \int_{\lambda_{y_0,y}} \mu - \oint_{[\gamma]} \mu \tag{2.11}$$

where $[\gamma] = [\lambda_{x,x_0} + (x_0, y_0) + \lambda_{y_0,y} + (y, x)]$ is the class of the full cycle in the homology of the surface. Changing the base points change μ by a multiple of $d_\diamond \varepsilon$.

It follows in the compact case that the dimensions of the harmonic forms on \diamond (the kernel of ΔA) modulo $d\varepsilon$, as well as the harmonic forms on Λ with same holonomies on the graph and on its dual, are twice the genus of the surface, as expected. Unfortunately, the space $\text{Im } A = \mathcal{H}^\perp \oplus \text{Im } d$ is not stable by the Hodge star $*$. We could nevertheless define holomorphic 1-forms on \diamond but their dimension would be much smaller than in the continuous theory, namely the genus of the surface. Criticality, defined in Section 4 provides conditions which ensure that the space $* \text{Im } A$ is “close” to $\text{Im } A$.

We construct an *heterogeneous* wedge product for 1-forms: with $\alpha, \beta \in C^1(\Lambda)$, define $\alpha \wedge \beta \in C^1(\diamond)$ by

$$\iint_{(x,y,x',y')} \alpha \wedge \beta := \frac{1}{2} \left(\int_{(x,x')} \alpha \int_{(y,y')} \beta + \int_{(y,y')} \alpha \int_{(x',x)} \beta \right). \tag{2.12}$$

It verifies $A(\alpha_\diamond) \wedge A(\beta_\diamond) = \alpha_\diamond \wedge \beta_\diamond$, the first wedge product being between 1-forms on Λ and the second between forms on \diamond . The usual scalar product on

compactly supported forms on Λ reads as expected:

$$(\alpha, \beta) = \frac{1}{2} \sum_{e \in \Lambda_1} \rho(e) \left(\int_e \alpha \right) \left(\int_e \bar{\beta} \right) = \iint_{\diamond_2} \alpha \wedge * \bar{\beta}. \quad (2.13)$$

2.3 Energies

The L^2 norm of the 1-form df , called the *Dirichlet energy* of the function f , is the mean of the usual Dirichlet energies on each independent graph:

$$\begin{aligned} E_D(f) &:= \frac{1}{2} \|df\|^2 = \frac{1}{2} (df, df) = \frac{1}{4} \sum_{(x, x') \in \Lambda_1} \rho(x, x') |f(x') - f(x)|^2 \\ &= \frac{E_D(f|\Gamma) + E_D(f|\Gamma^*)}{2}. \end{aligned} \quad (2.14)$$

Harmonic maps minimize this energy among functions fulfilling certain boundary conditions.

The *conformal energy* of a map measures its conformality defect, it is null on holomorphic functions:

$$E_C(f) := \frac{1}{4} \|df - i * df\|^2. \quad (2.15)$$

It is related to the Dirichlet energy through the same formula as in the continuous case:

$$\begin{aligned} E_C(f) &= \frac{1}{4} (df - i * df, df - i * df) \\ &= \frac{1}{4} \|df\|^2 + \frac{1}{4} \|-i * df\|^2 + \frac{1}{2} \operatorname{Re}(df, -i * df) \\ &= \frac{1}{2} \|df\|^2 + \frac{1}{2} \operatorname{Im} \iint_{\diamond_2} df \wedge \bar{df} \\ &= E_D(f) - \mathcal{A}(f) \end{aligned} \quad (2.16)$$

where the *area of the image* of the application f in the complex plane has the same formula

$$\mathcal{A}(f) = \frac{i}{2} \iint_{\diamond_2} df \wedge \bar{df} \quad (2.17)$$

as in the continuous case since, for a face $(x, y, x', y') \in \diamond_2$, the algebraic area of the oriented quadrilateral $(f(x), f(x'), f(y), f(y'))$ is given by

$$\begin{aligned} \iint_{(x, y, x', y')} df \wedge \bar{df} &= i \operatorname{Im} \left((f(x') - f(x)) \overline{(f(y') - f(y))} \right) \\ &= -2i \mathcal{A}(f(x), f(x'), f(y), f(y')). \end{aligned}$$

3 Period matrix

We use the convention of Farkas and Kra [7], chapter III, to which we refer for details. Consider (\diamond, ρ) a discrete compact Riemann surface.

3.1 Intersection number, on Λ and on \diamond

For a given simple (real) cycle $C \in Z_1(\Lambda)$, we construct a harmonic 1-form η_C such that $\int_A \eta_C$ counts the algebraic number of times A contains an edge dual to an edge of C : It is the solution of a Neumann problem on the surface cut open along C (see [10] for details). It is the minimum of the Dirichlet energy, which is convex on the space of closed 1-forms having the desired holonomies. It follows from standard homology technique that η_C depends only on the homology class of C (all the cycles which differ from C by an exact cycle ∂A) and can be extended linearly to all cycles as $\eta_\bullet: H_1(\Lambda) \rightarrow C^1(\Lambda)$; it fulfills, for a closed form θ ,

$$\oint_C \theta = \iint_{\diamond} \eta_C \wedge \theta, \quad (3.1)$$

and a basis of the homology provides a dual basis of harmonic forms on Λ . Beware that if the cycle $C \in Z_1(\Gamma)$ is purely on Γ , then this form $\eta_{C|_{\Gamma}} = 0$ is null on Γ .

The *intersection number* between two cycles $A, B \in Z_1(\Lambda)$ is defined as

$$A \cdot B := \iint_{\diamond} \eta_A \wedge \eta_B. \quad (3.2)$$

It is obviously linear and antisymmetric, it is an integer number for integer cycles. Let us stress again that the intersection of a cycle on Γ with another cycle on Γ is always null. A cycle $C \in Z_1(\diamond)$ defines a pair of cycles on each graph $C_{\Gamma} \in Z_1(\Gamma)$, $C_{\Gamma^*} \in Z_1(\Gamma^*)$ which are homologous to C on the surface, composed of portions of the boundary of the faces on Λ dual to the vertices of C . They are uniquely defined if we require that they lie “to the left” of C as shown in Figure 4. By the procedure (2.11) applied to $\eta_{C_{\Gamma}} + \eta_{C_{\Gamma^*}}$, we construct a 1-cocycle $\eta_C \in Z^1(\diamond)$ unique up to $d\varepsilon$, and since $d\varepsilon \wedge \theta = 0$ for all θ , Equation (3.2) defines an intersection number on $Z_1(\diamond)$. Unlike the intersection number on Λ , this one has all the usual expected properties. In particular Equation (3.2) holds for $A, B \in Z_1(\diamond)$.

3.2 Canonical dissection, fundamental polygon

The complex \diamond being connected, consider a maximal tree $T \subset \diamond_1$, that is to say T is simply connected and every edge added to T forms a cycle. A *canonical dissection* or cut-system \aleph of the genus g discrete Riemann surface \diamond is given by a set of oriented edges $(e_k)_{1 \leq k \leq 2g}$ such that the cycles $\aleph \subset (T \cup e_k)$ form a basis of the homology

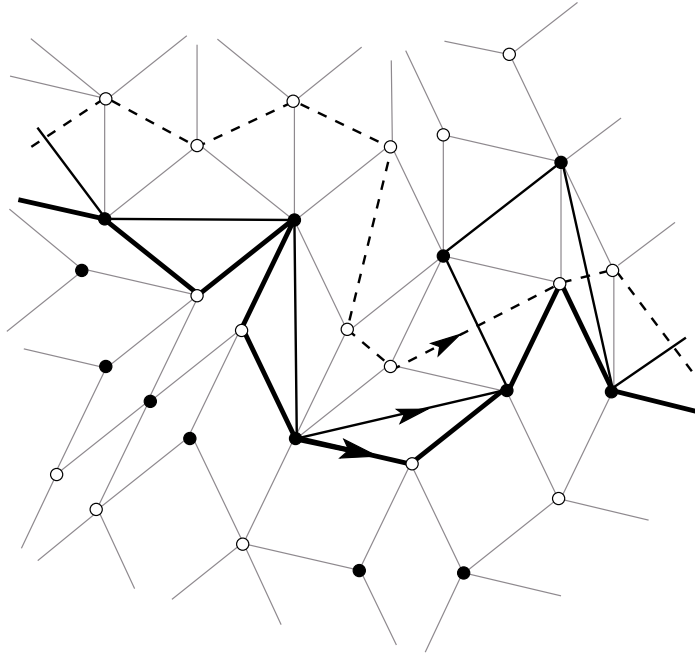


Figure 4. A path C on \diamond defines a pair of paths C_Γ and C_{Γ^*} on its left.

group $H_1(\diamond)$ verifying, for $1 \leq k, \ell \leq g$

$$\mathfrak{N}_k \cdot \mathfrak{N}_\ell = 0, \quad \mathfrak{N}_{k+g} \cdot \mathfrak{N}_{\ell+g} = 0, \quad \mathfrak{N}_k \cdot \mathfrak{N}_{\ell+g} = \delta_{k,\ell}. \quad (3.3)$$

They actually form a basis of the fundamental group $\pi_1(\diamond)$ and the defining relation among them is (noted multiplicatively)

$$\prod_{k=1}^g \mathfrak{N}_k \mathfrak{N}_{k+g} \mathfrak{N}_k^{-1} \mathfrak{N}_{k+g}^{-1} = 1. \quad (3.4)$$

The construction of such a basis is standard and we shall not repeat the procedure. What is less standard is the interpretation of Equation (3.4) in terms of the boundary of a fundamental domain, discretization introduces some subtleties (that can safely be skipped in first instance). We end up with the familiar $2g \times 2g$ intersection numbers matrix on \diamond .

Considering $T \cup e_k$ as a rooted graph, we can prune it of all its pending branches, leaving a simple closed loop \mathfrak{N}_k^- , attached to the origin O by a simple path λ_k (see Figure 5), yielding the cycle \mathfrak{N}_k . These three cycles are deformation retract of one another, $\mathfrak{N}_k^- \subset \mathfrak{N}_k \subset T \cup e_k$ hence are equal in homology.

In the continuous case, a basis of the homology can be realized by $2g$ simple arcs, transverse to one another and meeting only at the base point. It defines an isometric

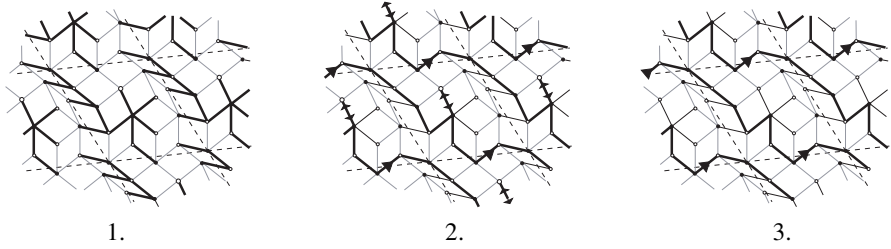


Figure 5. 1. A maximal rooted tree in a quadrilateral decomposition of the torus. 2. An additional edge defines a rooted cycle \aleph_1 , pruned of its dangling trees. 3. Its un-rooted version, the simple loop \aleph_1^- .

model of the surface as a fundamental domain homeomorphic to a disc and bordered by $4g$ arcs to identify pairwise. In the discrete case, by definition, the set $\diamondset \setminus \aleph$ of the cellular complex minus the edges taking part into the cycles basis is homeomorphic to a disc hence the surface is realized as a polygonal *fundamental domain* \mathcal{M} whose boundary edges are identified pairwise.

But it is sometimes impossible to choose a basis of the homology verifying (3.3) by simple discrete cycles which are transverse to one another. For instance, if the path λ_k is not empty, the cycle \aleph_k is not even simple. Moreover, some edges may belong to several cycles. In this case, the edges on the boundary of this fundamental polygon can not be assigned a unique element of the basis or its inverse, and therefore can not be grouped into only $4g$ continuous paths to identify pairwise but more than $4g$.

In fact, the information contained into the basis \aleph is more than simply this polygon, the set of edges composing the concatenated cycle

$$(\aleph_1, \aleph_{g+1}, \aleph_1^{-1}, \aleph_{g+1}^{-1}, \aleph_2, \dots, \aleph_g^{-1}, \aleph_{2g}^{-1}) \quad (3.5)$$

encodes a cellular complex \mathcal{M}_+ which is *not* a combinatorial surface and consists of the fundamental polygon \mathcal{M} plus some *dangling trees*, corresponding to the edges which belong to more than one cycle or participate more than once in a cycle (the paths λ_k), as exemplified in Figure 6. By construction, the edge e_k belongs to the cycle \aleph_k only, hence these trees are in fact without branches, simple paths whose only leaf is the base point O . To retrieve the surface, the edges of this structure \mathcal{M}_+ are identified group-wise, an edge participating k times in cycles will have $[k/2] + 2$ representatives to identify together, two on the fundamental polygon and the rest as edges of dangling trees.

Eliminating repetition, that is to say looking at (3.5) not as a *sequence* of edges but as a simplified *cycle* (or a simplified word in edges), thins \mathcal{M}_+ into \mathcal{M} , pruning away the dangling paths. The fundamental polygon boundary loses its structure as $4g$ arcs to be identified pairwise, in general a basis cycle will be disconnected around the fundamental domain and a given edge can not be assigned to a particular cycle.

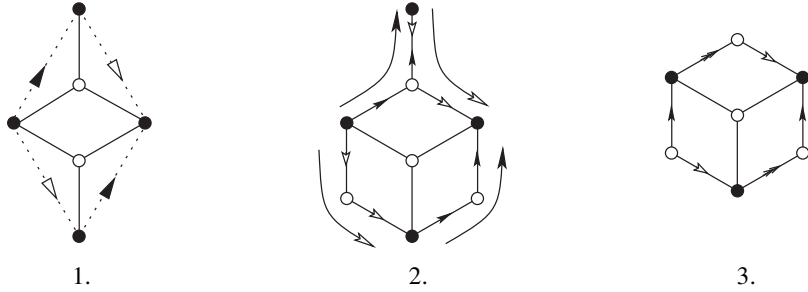


Figure 6. Three different fundamental polygons of a decomposition of the torus ($g = 1$) by three quadrilaterals: 1. The standard fundamental domain where the $4g$ paths are not adapted to \diamond . 2. \mathcal{M}_+ is composed of edges of \diamond composing $4g$ arcs (which may have portions in common) to identify pairwise, each edge corresponds to an element of the basis \aleph or its inverse, except for edges of “dangling trees” which are associated with two such elements. 3. \mathcal{M} is composed of edges of \diamond composing more than $4g$ arcs to identify pairwise, there is no correspondence with a basis of cycles.

This peculiarity gives a more complex yet well-defined meaning to the contour integral formula for a 1-form θ defined on the boundary edges of \mathcal{M}_+ ,

$$\oint_{\partial \mathcal{M}} \theta = \sum_{k=1}^{2g} \oint_{\aleph_k} \theta + \oint_{\aleph_k^{-1}} \theta. \tag{3.6}$$

This basis gives rise to cycles \aleph^Γ and \aleph^{Γ^*} whose homology classes form a basis of the group for each respective graph, that we compose into \aleph^Λ defined by

$$\begin{aligned} \aleph_k^\Lambda &= \aleph_k^\Gamma, & \aleph_{k+g}^\Lambda &= \aleph_k^{\Gamma^*}, \\ \aleph_{k+2g}^\Lambda &= \aleph_{k+g}^{\Gamma^*}, & \aleph_{k+3g}^\Lambda &= \aleph_{k+g}^\Gamma, \end{aligned} \tag{3.7}$$

for $1 \leq k \leq g$ so that while the intersection numbers matrix on \diamond is given by the $2g \times 2g$ matrix

$$(\aleph_k \cdot \aleph_\ell)_{k,\ell} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \tag{3.8}$$

the intersection numbers matrix on Λ is the $4g \times 4g$ matrix with the same structure

$$(\aleph_k^\Lambda \cdot \aleph_\ell^\Lambda)_{k,\ell} = \begin{pmatrix} \Gamma & \Gamma^* & \Gamma^* & \Gamma \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ -I & 0 & 0 & 0 \\ 0 & -I & 0 & 0 \end{pmatrix} \begin{matrix} \Gamma \\ \Gamma^* \\ \Gamma^* \\ \Gamma \end{matrix}. \tag{3.9}$$

3.3 Bilinear relations

Proposition 3.1. *Given a canonical dissection \aleph , for two closed forms $\theta, \theta' \in Z^1(\diamond)$,*

$$\iint_{\diamond} \theta \wedge \theta' = \sum_{j=1}^g \left(\oint_{\aleph_j} \theta \oint_{\aleph_{j+g}} \theta' - \oint_{\aleph_{j+g}} \theta \oint_{\aleph_j} \theta' \right); \tag{3.10}$$

for two closed forms $\theta, \theta' \in Z^1(\Lambda)$,

$$\iint_{\diamond} \theta \wedge \theta' = \sum_{j=1}^{2g} \left(\oint_{\aleph_j^\Lambda} \theta \oint_{\aleph_{j+2g}^\Lambda} \theta' - \oint_{\aleph_{j+2g}^\Lambda} \theta \oint_{\aleph_j^\Lambda} \theta' \right). \tag{3.11}$$

Proof. Each side is bilinear and depends only on the cohomology classes of the forms. Decompose the forms onto the cohomology basis (α_k) (see below). On Λ , use Equation (3.15) for the LHS and the duality property Equation (3.14) for the RHS. On \diamond , use their counterparts. \square

Notice that for a harmonic form $\theta \in \mathcal{H}^1(\Lambda)$, the form $*\theta$ is closed as well, therefore its norm is given by its boundary values on the fundamental polygon:

$$\theta \in \mathcal{H}^1(\Lambda) \implies \|\theta\|^2 = \sum_{j=1}^{2g} \left(\oint_{\aleph_j} \theta \oint_{\aleph_{j+2g}} *\bar{\theta} - \oint_{\aleph_{j+2g}} \theta \oint_{\aleph_j} *\bar{\theta} \right). \tag{3.12}$$

3.4 Basis of harmonic forms, basis of holomorphic forms

We define α^Λ , the basis of real harmonic 1-forms, dual to the homology basis \aleph^Λ , as described in Section 3.1,

$$\begin{aligned} \alpha_k^\Lambda &:= \eta_{\aleph_{k+2g}^\Lambda} & \text{and} \\ \alpha_{k+2g}^\Lambda &:= -\eta_{\aleph_k^\Lambda} & \text{for } 1 \leq k \leq 2g \end{aligned} \tag{3.13}$$

which verify

$$\begin{aligned} \oint_{\aleph_k^\Lambda} \alpha_\ell &= \delta_{k,\ell}, \\ \oint_{\aleph_{k+2g}^\Lambda} \alpha_{\ell+2g} &= \delta_{k,\ell}, \end{aligned} \tag{3.14}$$

and dually, the intersection matrix elements are given by

$$\aleph_k^\Lambda \cdot \aleph_\ell^\Lambda = \iint_{\diamond} \alpha_k^\Lambda \wedge \alpha_\ell^\Lambda = (\alpha_k^\Lambda, -*\alpha_\ell^\Lambda). \tag{3.15}$$

On \diamond , the elements $\alpha_k^\diamond := \eta_{\aleph_{k+g}}$ and $\alpha_{k+g}^\diamond := -\eta_{\aleph_k}$ for $1 \leq k \leq g$, defined up to $d\varepsilon$, verify $A(\alpha_k^\diamond) = \alpha_k^\Lambda + \alpha_{k+g}^\Lambda$, $A(\alpha_{k+g}^\diamond) = \alpha_{k+2g}^\Lambda + \alpha_{k+3g}^\Lambda$ and form a basis of the

cohomology on \diamond dual to \aleph as well,

$$\begin{aligned}\alpha_k^\diamond &:= \eta_{\aleph_{k+g}^\diamond} & \text{and} \\ \alpha_{k+g}^\diamond &:= -\eta_{\aleph_k^\diamond} & \text{for } 1 \leq k \leq g,\end{aligned}\tag{3.16}$$

they fulfill the first identity in Equation (3.15) but the second is meaningless in general since $*$ can not be defined on \diamond . We will drop the mention Λ when no confusion is possible.

Proposition 3.2. *The matrix of inner products on Λ ,*

$$(\alpha_k, \alpha_\ell)_{k,\ell} = \iint_{\diamond} \alpha_k \wedge * \bar{\alpha}_\ell = \begin{cases} + \int_{\aleph_{k+2g}} * \alpha_\ell, & 1 \leq k \leq 2g, \\ - \int_{\aleph_{k-2g}} * \alpha_\ell, & 2g < k \leq 4g. \end{cases} =: \begin{pmatrix} A & D \\ B & C \end{pmatrix}\tag{3.17}$$

is a real symmetric positive definite matrix.

Proof. It is real because the forms are real, and symmetric because the scalar product (2.13) is skew symmetric. Definition Equation (3.13) and Equation (3.1) lead to the integral formulae. Positivity follows from the bilinear relation Equation (3.11): for $\theta = \sum_{k=1}^{4g} \xi_k \alpha_k$, with $\xi_k \in \mathbb{C}$, $\sum_{k=1}^{4g} |\xi_k|^2 > 0$,

$$\begin{aligned}\|\theta\|^2 &= \sum_{j=1}^{2g} \left[\int_{\aleph_j} \theta \int_{\aleph_{2g+j}} * \bar{\theta} - \int_{\aleph_{2g+j}} \theta \int_{\aleph_{2j}} * \bar{\theta} \right] \\ &= \sum_{k,\ell=1}^{4g} \xi_k \bar{\xi}_\ell \sum_{j=1}^{2g} \left[\int_{\aleph_j} \alpha_k \int_{\aleph_{2g+j}} * \alpha_\ell - \int_{\aleph_{2g+j}} \alpha_k \int_{\aleph_{2j}} * \alpha_\ell \right] \\ &= \sum_{k,\ell=1}^{4g} \xi_k \bar{\xi}_\ell (\alpha_k, \alpha_\ell) > 0.\end{aligned}\tag{3.18}$$

□

The form α_k is supported by only one of the two graphs Γ or Γ^* , the form $*\alpha_k$ is supported by the other one, and the wedge product $\theta_\Gamma \wedge \theta_{\Gamma^*} = 0$ is null for two 1-forms supported by the same graph. Therefore the matrices A and C are $g \times g$ -block diagonal and B is anti-diagonal.

$$A = \begin{pmatrix} A_\Gamma & 0 \\ 0 & A_{\Gamma^*} \end{pmatrix}, \quad B = \begin{pmatrix} 0 & B_{\Gamma^*,\Gamma} \\ B_{\Gamma,\Gamma^*} & 0 \end{pmatrix}, \quad C = \begin{pmatrix} C_{\Gamma^*} & 0 \\ 0 & C_\Gamma \end{pmatrix}.\tag{3.19}$$

The matrices of intersection numbers (3.9) and of inner products differ only by the Hodge star $*$. Because $*$ preserves harmonic forms and the inner product, we get its matrix representation in the basis α ,

$$* = \begin{pmatrix} -D & A \\ -C & B \end{pmatrix}\tag{3.20}$$

and because $*^2 = -1$,

$$B^2 - C \cdot A + I = 0 \tag{3.21}$$

$$A \cdot B = {}^t B \cdot A \tag{3.22}$$

$$C \cdot {}^t B = B \cdot C. \tag{3.23}$$

On \diamond , while the Hodge star $*$ can not be defined, we can obviously consider the following positive scalar product on the classes of closed forms modulo $d\varepsilon$, to which the set (α_k^\diamond) belong:

$$\begin{aligned} (\alpha^\diamond, \beta^\diamond) &:= (A(\alpha^\diamond), A(\beta^\diamond)) \\ &= \sum_{\substack{(x,y,x',y') \in \diamond_2 \\ \rho = \rho(x,x'), \\ \rho^* = \rho(y,y')}} {}^t \begin{pmatrix} \int_{(x,y)} \alpha \\ \int_{(y,x')} \alpha \\ \int_{(x',y')} \alpha \\ \int_{(y',x)} \alpha \end{pmatrix} \cdot \begin{pmatrix} +\rho+\rho^* & +\rho-\rho^* & -\rho-\rho^* & -\rho+\rho^* \\ +\rho-\rho^* & +\rho+\rho^* & -\rho+\rho^* & -\rho-\rho^* \\ -\rho-\rho^* & -\rho+\rho^* & +\rho+\rho^* & +\rho-\rho^* \\ -\rho+\rho^* & -\rho-\rho^* & +\rho-\rho^* & +\rho+\rho^* \end{pmatrix} \cdot \begin{pmatrix} \int_{(x,y)} \bar{\beta} \\ \int_{(y,x')} \bar{\beta} \\ \int_{(x',y')} \bar{\beta} \\ \int_{(y',x)} \bar{\beta} \end{pmatrix}. \end{aligned}$$

and it yields

$$(\alpha_k^\diamond, \alpha_\ell^\diamond)_{k,\ell} = \begin{pmatrix} A_\Gamma + A_{\Gamma^*} & {}^t B_{\Gamma^*} + {}^t B_{\Gamma^*\Gamma} \\ B_{\Gamma^*\Gamma} + B_{\Gamma^*\Gamma} & C_\Gamma + C_{\Gamma^*} \end{pmatrix}, \tag{3.24}$$

which, in general, can not be understood as the periods of a set of forms on \diamond along the basis \aleph .

Let us decompose the space of harmonic forms into two orthogonal supplements,

$$\mathcal{H}^1(\Lambda) = \mathcal{H}_\parallel^1 \oplus^\perp \mathcal{H}_\perp^1 \tag{3.25}$$

where the first vector space are the harmonic forms whose holonomies on one graph are equal to their holonomies on the dual, that is to say

$$\mathcal{H}_\parallel^1 := \text{Vect}(\alpha_k + \alpha_{k+g}, 1 \leq k \leq g \text{ or } 2g < k \leq 3g). \tag{3.26}$$

Definition (3.13) and Equation (3.1) imply that

$$\mathcal{H}_\perp^1 = \text{Vect}(*\alpha_k - *\alpha_{k+g}, 1 \leq k \leq g \text{ or } 2g < k \leq 3g). \tag{3.27}$$

These elements in the basis $(\alpha_k + \alpha_{k+g}, \alpha_k - \alpha_{k+g})$, for $1 \leq k \leq g$ and $2g < k \leq 3g$, are represented by the following invertible matrix:

$$\begin{pmatrix} I & 0 & {}^t B_{\Gamma^*} - {}^t B_{\Gamma^*\Gamma} & A_\Gamma - A_{\Gamma^*} \\ 0 & I & C_\Gamma - C_{\Gamma^*} & B_{\Gamma^*} - B_{\Gamma^*\Gamma} \\ 0 & 0 & {}^t B_{\Gamma^*} + {}^t B_{\Gamma^*\Gamma} & A_\Gamma + A_{\Gamma^*} \\ 0 & 0 & C_\Gamma + C_{\Gamma^*} & B_{\Gamma^*} + B_{\Gamma^*\Gamma} \end{pmatrix}. \tag{3.28}$$

It implies in particular that the lower right $g \times g$ block is invertible, therefore so is Equation (3.24).

3.5 Period matrix

Proposition 3.3. *The matrix $\Pi = C^{-1} \cdot (i - B)$ is the period matrix of the basis of holomorphic forms*

$$\zeta_k := (i - *) \sum_{\ell=1}^{2g} C_{k,\ell}^{-1} \alpha_{\ell+2g} \quad (3.29)$$

in the canonical dissection \aleph , that is to say

$$\oint_{\aleph_k} \zeta_\ell = \begin{cases} \delta_{k,\ell} & \text{for } 1 \leq k \leq 2g, \\ \Pi_{k-2g,\ell} & \text{for } 2g < k \leq 4g, \end{cases} \quad (3.30)$$

and Π is symmetric, with a positive definite imaginary part.

The proof is essentially the same as in the continuous case [7] and we include it for completeness.

Proof of Proposition 3.3. Let $\omega_j := \alpha_j + i * \alpha_j$ for $1 \leq j \leq 4g$. These holomorphic forms fulfill

$$\begin{aligned} P_{k,j} &:= \frac{1}{2}(\omega_k, \omega_j) = (\alpha_k, \alpha_j) + i(\alpha_k, - * \alpha_j) \\ &= \begin{cases} -i \int_{\aleph_{j+2g}} \omega_k, & 1 \leq j \leq 2g, \\ i \int_{\aleph_{j-2g}} \omega_k, & 2g < j \leq 4g. \end{cases} \end{aligned} \quad (3.31)$$

P is the period matrix of the forms (ω) in the homology basis \aleph . The first $2g$ forms $(\omega_j)_{1 \leq j \leq 2g}$ are a basis of holomorphic forms. It has the right dimension and they are linearly independent:

$$\begin{aligned} &\sum_{j=1}^{2g} (\lambda_j + i \mu_j) (\alpha_j + i * \alpha_j) \\ &= \sum_{j=1}^{2g} \left((\lambda_j + \sum_{k=1}^{2g} \mu_k B_{j,k}) \alpha_j + \sum_{k=1}^{2g} \mu_k C_{j,k} \alpha_{2g+j} \right) \\ &\quad + i \sum_{j=1}^{2g} \left((\mu_j + \sum_{k=1}^{2g} \lambda_k B_{j,k}) \alpha_j + \sum_{k=1}^{2g} \lambda_k C_{j,k} \alpha_{2g+j} \right) \end{aligned} \quad (3.32)$$

is null, for $\lambda, \mu \in \mathbb{R}$ only when $\lambda = \mu = 0$ because C is positive definite. Similarly for the last $2g$ forms. The change of basis $i C^{-1}$ on them provides the basis of holomorphic forms (ζ) . The last $2g$ rows of P is the $2g \times 4g$ matrix $(B - i I, C)$ hence the periods of (ζ) in \aleph are given by (I, Π) . \square

The first identity in Equation (3.30) uniquely defines the basis ζ and a holomorphic 1-form is completely determined by whether its periods on the first $2g$ cycles of \aleph , or their real parts on the whole set.

Notice that because C is $g \times g$ block diagonal and B is anti-diagonal, Π is decomposed into four $g \times g$ blocks, the two diagonal matrices form $i C^{-1}$ and are pure imaginary, the other two form $-C^{-1} \cdot B$ and are real.

$$\Pi = \begin{pmatrix} \Pi_{i_*} & \Pi_r \\ \Pi_{r_*} & \Pi_i \end{pmatrix} = \begin{pmatrix} i C_{\Gamma^*}^{-1} & -C_{\Gamma^*}^{-1} \cdot B_{\Gamma^*, \Gamma} \\ -C_{\Gamma}^{-1} \cdot B_{\Gamma, \Gamma^*} & i C_{\Gamma}^{-1} \end{pmatrix}. \tag{3.33}$$

Therefore the holomorphic forms ζ_k are real on one graph and pure imaginary on its dual,

$$\begin{aligned} 1 \leq k \leq g &\implies \zeta_k \in C_{\mathbb{R}}^1(\Gamma) \oplus i C_{\mathbb{R}}^1(\Gamma^*) \\ g < k \leq 2g &\implies \zeta_k \in C_{\mathbb{R}}^1(\Gamma^*) \oplus i C_{\mathbb{R}}^1(\Gamma). \end{aligned} \tag{3.34}$$

We will call

$$\Pi_{\Gamma} = \Pi_r + \Pi_{i_*} \tag{3.35}$$

the period matrix on the graph Γ the sum of the real periods of ζ_k , $1 \leq k \leq g$, on Γ , with the associated pure imaginary periods on the dual Γ^* , and similarly for ζ_k , $g < k \leq 2g$, the period matrix on Γ^* .

It is natural to ask how close Π_{Γ} and Π_{Γ^*} are from one another, and whether their mean can be given an interpretation. Criticality [10], [11], that we will define in the next Section 4 answers partially the issue.

4 Criticality and integrable system

The theory of discrete holomorphy can be viewed as the simplest (it is linear) of a series of integrable theories [1]. We will present its quadratic counterpart, which leads to another version of discrete analytic functions, based on circle patterns. Along the way, we will see how discrete exponentials and discrete polynomials emerge due to integrable systems theory pieces of technology named the Bäcklund or Darboux transform [1].

4.1 Criticality

Criticality clarifies how purely combinatorial data can be connected to an underlying geometry and gives a meaning to approximation theorems.

Definition 4.1. A discrete conformal structure (\diamond, ρ) of a simply connected set is critical if there exists a discrete holomorphic map Z such that the quadrilateral faces \diamond_2 can be simultaneously embedded into rhombi by Z in the complex plane. We call Z a critical map. The non simply connected case is covered using an atlas of overlapping simply connected patches and critical maps.

A discrete holomorphic function Z only maps vertices to the complex plane. We extend it linearly to quadrilateral edges and harmonically inside the quadrilateral faces. Because of the Gauss–Bonnet theorem, it is not possible to globally embed a compact surface into the plane, therefore we allow for an atlas of local critical maps with a finite number of fixed local conic singularities. When a continuous limit is taken, their number, angle and position should not change, and the theorem of isolated singularities helps us wipe them out as inessential.

It is a simple calculation to check that if Z is a critical map, any discrete holomorphic function $f \in \Omega(\diamond)$ gives rise through (2.6) to a *holomorphic* 1-form $f dZ$.

For a holomorphic function f , the equality $f dZ \equiv 0$ is equivalent to $f = \lambda \varepsilon$ for some $\lambda \in \mathbb{C}$ with ε the bi-constant $\varepsilon|_\Gamma \equiv +1$, $\varepsilon|_\Gamma^* \equiv -1$.

Following Duffin [5], [6], we introduce the

Definition 4.2. For a holomorphic function f , define on a flat simply connected map U the holomorphic functions f^\dagger , the *dual* of f , and f' , the *derivative* of f , by the following formulae:

$$f^\dagger(z) := \varepsilon(z) \bar{f}(z), \quad (4.1)$$

where \bar{f} denotes the complex conjugate, $\varepsilon = \pm 1$ is the bi-constant, and

$$f'(z) := \frac{4}{\delta^2} \left(\int_O^z f^\dagger dZ \right)^\dagger + \lambda \varepsilon, \quad (4.2)$$

defined up to ε .

It is an immediate calculation [11] to check the following

Proposition 4.3. *The derivative f' fulfills*

$$df = f' dZ. \quad (4.3)$$

4.2 Period matrix

Criticality allows to state how close the two period matrices of a refining sequence of critical maps are:

Theorem 4.1. *In the genus one critical case, the period matrices Π_Γ and Π_{Γ^*} are equal to the period matrix Π_Σ of the underlying surface Σ . For higher genus, given a refining sequence (\diamond^k, ρ_k) of critical maps of Σ , the discrete period matrices Π_{Γ^k} and $\Pi_{\Gamma^{*k}}$ converge to the period matrix Π_Σ .*

Proof. The genus one case is postponed to Section 4.3. We proved [11] that, given a refining sequence of critical maps, any holomorphic function can be approximated by a sequence of discrete holomorphic functions. Taking the real parts, this implies as well that any harmonic function can be approximated by discrete harmonic functions. In particular, the discrete solutions f_k to a Dirichlet or Neumann problem on

a simply connected set converge to the continuous solution f because the latter can be approximated by discrete harmonic functions g_k and the difference $f_k - g_k$ being harmonic and small on the boundary, converges to zero. In particular, each form in the basis (α_ℓ^\diamond) , provides a solution to the Neumann problem Equation (3.16) and a similar procedure, detailed afterwards, define a converging sequence of forms ζ_ℓ^\diamond , yielding the result. \square

We can try to replicate the work done on Λ on the graph \diamond . A problem is that $A_\Gamma + A_{\Gamma^*}$ and $C_\Gamma + C_{\Gamma^*}$ need not be positive definite. Moreover, the Hodge star $*$ does not preserve the space $(A(\alpha_k^\diamond))$ of harmonic forms with equal holonomies on the graph and on its dual, so we can not define the analogue of $\alpha + i * \alpha$ on \diamond . We first investigate what happens when we can partially define these analogues:

Assume that for $2g < k \leq 3g$, the holonomies of $*\alpha_k$ on Γ are equal to the holonomies of $*\alpha_{k+g}$ on Γ^* , that is to say $C_\Gamma = C_{\Gamma^*} =: \frac{1}{2}C_\diamond$ and $D_{\Gamma\Gamma^*} = D_{\Gamma^*\Gamma} =: \frac{1}{2}D_\diamond$. It implies that the transposes fulfill $B_{\Gamma\Gamma^*} = B_{\Gamma^*\Gamma} =: \frac{1}{2}B_\diamond$ as well. We can then define $\beta_{k-g}^\diamond \in Z^1(\diamond)$ such that $A(\beta_{k-g}^\diamond) = *\alpha_{k+g}$, uniquely up to $d\varepsilon$. The last g columns ${}^t(B_\diamond, C_\diamond)$ of the matrix of scalar product Equation (3.24) are related to their periods in the homology basis \aleph^\diamond in a way similar to Equation (3.17). By the same reasoning as before, the forms

$$\zeta_k^\diamond = \sum_{\ell=1}^g C_{\diamond, k, \ell}^{-1} \left(\alpha_{\ell+g}^\diamond - i\beta_{\ell+g}^\diamond \right), \quad 1 \leq k \leq g \tag{4.4}$$

verify $A(\zeta_k^\diamond) = \frac{\zeta_k + \zeta_{k+g}}{2}$ and have periods on \aleph^\diamond given by the identity for the first g cycles and the following $g \times g$ matrix, mean of the period matrices on the graph and on its dual:

$$\Pi^\diamond = C_\diamond^{-1}(i - B_\diamond) = \frac{\Pi_\Gamma + \Pi_{\Gamma^*}}{2}. \tag{4.5}$$

The same reasoning applies when the periods of the forms $*\alpha_k$ on the graph and on its dual are not equal but close to one another. In the context of refining sequences, we said that the basis (α_ℓ^\diamond) , converges to the continuous basis of harmonic forms defined by the same Neumann problem Equation (3.16). Therefore

$$C_\Gamma - C_{\Gamma^*} = o(1), \quad B_{\Gamma\Gamma^*} - B_{\Gamma^*\Gamma} = o(1). \tag{4.6}$$

A harmonic form $\nu_{k+g} = o(1)$ on Γ^* can be added to $*\alpha_{k+g}$ such that there exists $\beta_{k-g}^\diamond \in Z^1(\diamond)$ with $A(\beta_{k-g}^\diamond) = *\alpha_{k+g} + \nu_{k+g}$, yielding forms ζ_k^\diamond , verifying $A(\zeta_k^\diamond) = \frac{1}{2}(\zeta_k + \zeta_{k+g}) + o(1)$ and whose period matrix is $\Pi^\diamond + o(1)$. Since the periods of α_k converge to the same periods as its continuous limit, this period matrix converges to the period matrix Π_Σ of the surface. Which is the claim of Theorem 4.1.

In the paper [3], R. Costa-Santos and B. McCoy define a period matrix on a special cellular decomposition Γ of a surface by squares. They do not consider the dual graph Γ^* . Their period matrix is equal to one of the two diagonal blocks of the double period matrix we construct in this case. They do not have to consider the off-diagonal

blocks because the problem is so symmetric that their period matrix is pure imaginary. The triangular case is investigated in [4].

4.3 Genus one case

Criticality solves partially the problem of having two different $g \times g$ period matrices instead of one since they converge to one another in a refining sequence. However, on a genus one critical torus, the situation is simpler: The overall curvature is null and a critical map is everywhere flat. Therefore the cellular decomposition is the quotient of a periodic cellular decomposition of the plane by two independent periods. They can be normalized to $(1, \tau)$. The continuous period matrix is the 1×1 -matrix τ . A basis of the two dimensional holomorphic 1-forms is given by the real and imaginary parts of dZ on Γ and Γ^* respectively, and the reverse. The discrete period matrix is the 2×2 matrix $\begin{pmatrix} \text{Im } \tau & \text{Re } \tau \\ \text{Re } \tau & \text{Im } \tau \end{pmatrix}$ and the period matrices on the graph and on its dual are both equal to the continuous one.

For illustration purposes, the whole construction, of a basis of harmonic forms, then projected onto a basis of holomorphic forms, yielding the period matrix, can be checked explicitly on the critical maps of the genus 1 torus decomposed by square or triangular/hexagonal lattices:

Consider the critical square (rectangular) lattice decomposition of a torus $\diamond = (\mathbb{Z}e^{i\theta} + \mathbb{Z}e^{-i\theta})/(2pe^{i\theta} + 2qe^{-i\theta})$, with horizontal parameter $\rho = \tan \theta$ and vertical parameter its inverse. Its modulus is $\tau = \frac{q}{p}e^{2i\theta}$. The two dual graphs Γ and Γ^* are isomorphic. An explicit harmonic form α_1^Γ is given by the constant $1/2p$ on horizontal and downwards edges of the graph Γ and 0 on all the other edges. Its holonomies are 1 and 0 on the p , resp. q cycles. Considering $1/2q$ and the dual graph, we construct in the same fashion $\alpha_2^\Gamma, \alpha_1^{\Gamma^*}, \alpha_2^{\Gamma^*}$. The matrix of inner products is

$$(\alpha_k, \alpha_\ell)_{k,\ell} = \frac{1}{\sin 2\theta} \begin{pmatrix} \frac{q}{p} & & & \cos 2\theta \\ & \frac{q}{p} & \cos 2\theta & \\ & \cos 2\theta & \frac{p}{q} & \\ \cos 2\theta & & & \frac{p}{q} \end{pmatrix} \quad (4.7)$$

using $\frac{\rho+1/\rho}{2} = 1/\sin 2\theta$ and $\frac{\rho-1/\rho}{2} = -1/\tan 2\theta$ so that the period matrix is

$$\Pi = \frac{q}{p} \begin{pmatrix} i \sin 2\theta & \cos 2\theta \\ \cos 2\theta & i \sin 2\theta \end{pmatrix}. \quad (4.8)$$

Therefore there exists a holomorphic form which has the same periods on the graph and on its dual, it is the average of the two half forms of Equation (3.30) and its periods are $(1, \frac{q}{p}e^{2i\theta})$ along the p , resp. q cycles, yielding the continuous modulus. This holomorphic form is simply the normalized fundamental form $\frac{dZ}{pe^{-i\theta}}$.

In the critical triangular/hexagonal lattice, we just point out to the necessary check by concentrating on a tile of the torus, composed of two triangles, pointing up and

down respectively. We show that there exists an explicit holomorphic form which has the same shift on the graph and on its dual, along this tile. Let ρ_-, ρ_\setminus and $\rho_/$ the three parameters around a given triangle. Criticality occurs when $\rho_- \rho_\setminus + \rho_\setminus \rho_/ + \rho_/ \rho_- = 1$. The form which is 1 on the rightwards and South-West edges and 0 elsewhere is harmonic on the triangular lattice. Its pure imaginary companion on the dual hexagonal lattice exhibits a shift by $i \rho_\setminus$ in the horizontal direction and $i(\rho_\setminus + \rho_-)$ in the North-East direction along the tile. Dually, on the hexagonal lattice, the form which is $\rho_\setminus \rho_-$ along the North-East and downwards edges and $1 - \rho_\setminus \rho_-$ along the South-East edges, is a harmonic form. Its shift in the horizontal direction is 1, in the North-East direction 0, and its pure imaginary companion on the triangular lattice exhibits a shift by $i \rho_\setminus$ in the horizontal direction and $i(\rho_\setminus + \rho_-)$ in the North-East direction along the tile as before. Hence their sum is a holomorphic form with equal holonomies on the triangular and hexagonal graphs and the period matrix it computes is the same as the continuous one. This simply amounts to pointing out that the fundamental form dz can be explicitly expressed in terms of the discrete conformal data.

4.4 $\bar{\partial}$ operator

For holomorphic or anti-holomorphic functions, df is, locally on each pair of dual diagonals, proportional to dZ , resp. $d\bar{Z}$, we define ∂ and $\bar{\partial}$ operators (not to be confused with the boundary operator on chains) that decompose the exterior derivative into holomorphic and anti-holomorphic parts yielding

$$df \wedge \bar{d}f = (|\partial f|^2 + |\bar{\partial} f|^2) dZ \wedge d\bar{Z}$$

where the derivatives naturally live on faces:

In the continuous theory, for any derivable function f on the complex plane, the derivatives $\partial = \frac{d}{dx} + i \frac{d}{dy}$ and $\bar{\partial} = \frac{d}{dx} - i \frac{d}{dy}$ with respect to $z = x + i y$ and $\bar{z} = x - i y$ yield

$$f(z + z_0) = f(z_0) + z(\partial f)(z_0) + \bar{z}(\bar{\partial} f)(z_0) + o(|z|).$$

These derivatives can be seen as a limit of a contour integral over a small loop γ around z_0 :

$$(\partial f)(z_0) = \lim_{\gamma \rightarrow z_0} \frac{i}{2\mathcal{A}(\gamma)} \oint_{\gamma} f d\bar{z}, \quad (\bar{\partial} f)(z_0) = - \lim_{\gamma \rightarrow z_0} \frac{i}{2\mathcal{A}(\gamma)} \oint_{\gamma} f dz,$$

which leads to the following definitions in the discrete setup:

$$\partial: C^0(\diamond) \rightarrow C^2(\diamond)$$

$$f \mapsto \partial f = [(x, y, x', y') \mapsto - \frac{i}{2\mathcal{A}(x, y, x', y')} \oint_{(x,y,x',y')} f d\bar{Z}]$$

$$= \frac{(f(x') - f(x))(\bar{y}' - \bar{y}) - (\bar{x}' - \bar{x})(f(y') - f(y))}{(x' - x)(\bar{y}' - \bar{y}) - (\bar{x}' - \bar{x})(y' - y)},$$

$$\bar{\partial}: C^0(\diamond) \rightarrow C^2(\diamond)$$

$$f \mapsto \bar{\partial}f = [(x, y, x', y') \mapsto -\frac{i}{2\mathcal{A}(x, y, x', y')} \oint_{(x, y, x', y')} f dZ]$$

$$= \frac{(f(x')-f(x))(y'-y)-(x'-x)(f(y')-f(y))}{(x'-x)(y'-y)-(x'-x)(y'-y)}.$$

A holomorphic function f verifies $\bar{\partial}f \equiv 0$ and (with $Z(u)$ noted simply u)

$$\partial f(x, y, x', y') = \frac{f(y') - f(y)}{y' - y} = \frac{f(x') - f(x)}{x' - x}.$$

Notice that the statement $f = (\int \partial f dz)$ has no meaning, ∂ is not a derivation endomorphism in the space of functions on the vertices of the double.

On the other hand, these differential operators can be extended (see [8]) into operators (the Kasteleyn operator) $\partial_{20}, \bar{\partial}_{20}: C^2(\diamond) \rightarrow C^0(\diamond)$ simply by transposition, $\partial_{20} = -{}^t \partial_{02}$, leading to endomorphisms of $C^0(\diamond) \oplus C^2(\diamond)$. They are such that their composition, restricted to the vertices \diamond_0 , gives back the Laplacian:

$$\Delta = \frac{1}{2}(\partial \circ \bar{\partial} + \bar{\partial} \circ \partial).$$

Furthermore, the double derivative $\partial_{20} \circ \partial_{02}$ is a well defined endomorphism of $C^0(\diamond)$.

4.5 Geometrical interpretation

Following an idea by Colin de Verdière, generalized by Kenyon, holomorphicity at criticality can be recognized by eye. Consider the embedding of a graph Γ given by half the diagonals of a critical map Z . We can embed likewise its dual graph Γ^* . They can actually be represented both at the same time by shrinking every face by a half. For example the triangular and hexagonal lattices are represented in Figure 7.1. A discrete holomorphic function is mapping these polygons to *similar* polygons: Let F a discrete holomorphic function whose derivative is the discrete holomorphic function f (this is well defined only up to the bi-constant ε). It maps the polygon $x^* \in \Gamma_2^*$, dual to the vertex $x \in \Gamma_0$, to a similar polygon centered at $F(x)$, $F(x) + f(x)x^*$ where the multiplication by the complex $re^{i\theta}$ encodes the similitude of ratio r and angle θ . The fact that these polygons still touch at their vertices in the same manner as in the reference case Z is precisely the holomorphicity condition $\oint_{\partial x^*} f dZ = 0$.

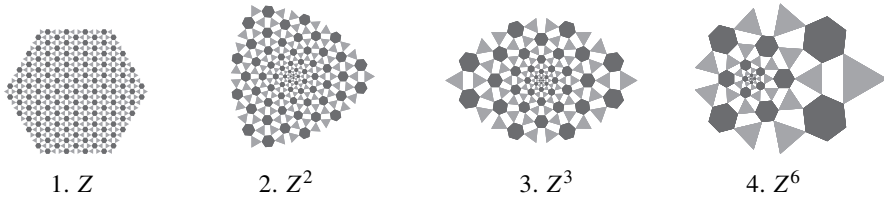


Figure 7. Discrete polynomials in the triangular/hexagonal case

4.6 Quasiconformal maps

For a general discrete function, the derivation with respect to the Z and \bar{Z} parameters allow one to define the *dilatation* coefficient $D_f: \diamond_2 \rightarrow \mathbb{R}^+$:

$$D_f := \frac{|\partial f| + |\bar{\partial} f|}{|\partial f| - |\bar{\partial} f|}.$$

For $D_f \geq 1$, that is to say $|\bar{\partial} f| \leq |\partial f|$, we will say that f is *quasiconformal*. The real dilatation can be written in term of the *complex dilatation*:

$$\mu_f = \frac{\bar{\partial} f}{\partial f} = \frac{(f(x') - f(x))(y' - y) - (x' - x)(f(y') - f(y))}{(f(x') - f(x))(\bar{y}' - \bar{y}) - (\bar{x}' - \bar{x})(f(y') - f(y))}.$$

Locally, a discrete harmonic function can be written as a sum of a holomorphic function and a anti-holomorphic function. Their derivative with respect to Z and \bar{Z} respectively give rise to a geometrical interpretation related to Figure 8: The complex dilatation

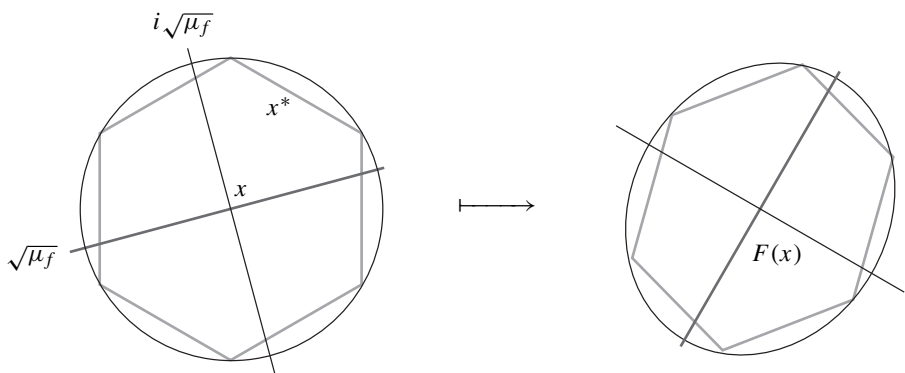


Figure 8. A quasiconformal map shears a polygon to a non similar shape.

$\mu_f(x)$ can then be defined at a vertex $x \in \Gamma$ and its square root direction is mapped to the semimajor axis of an ellipse which supports a non similar image of the polygon x^* centered at $F(x)$.

The Jacobian $J = |\partial f|^2 - |\bar{\partial} f|^2$ relates the area of the quadrilateral and its image:

$$\iint_{(x,y,x',y')} df \wedge \bar{d}f = J \iint_{(x,y,x',y')} dZ \wedge \bar{d}Z.$$

4.7 Discrete exponential

Definition 4.4. For a constant $\lambda \in \mathbb{C}$, we define the discrete exponential of parameter λ , noted $\exp(:\lambda: Z)$, as the solution of

$$\begin{aligned} \exp(:\lambda: O) &= 1 \\ d \exp(:\lambda: Z) &= \lambda \exp(:\lambda: Z) dZ. \end{aligned} \quad (4.9)$$

We define its derivatives with respect to the continuous parameter λ :

$$Z^k: \exp(:\lambda: Z) := \frac{\partial^k}{\partial \lambda^k} \exp(:\lambda: Z). \quad (4.10)$$

The discrete exponential on the square lattice was defined by Lelong-Ferrand [9], generalized in [13] and studied independently in [8]. For $|\lambda| \neq 2/\delta$, an immediate check shows that it is a rational fraction in λ at every point: For the vertex $x = \sum \delta e^{i\theta_k}$,

$$\exp(:\lambda: x) = \prod_k \frac{1 + \frac{\lambda\delta}{2} e^{i\theta_k}}{1 - \frac{\lambda\delta}{2} e^{i\theta_k}} \quad (4.11)$$

where (θ_k) are the angles defining $(\delta e^{i\theta_k})$, the set of $(Z$ -images of) \diamond -edges between x and the origin. Because the map is critical, Equation (4.11) only depends on the end points (O, x) . It is a generalization of a well known formula, in a slightly better version,

$$\exp(\lambda x) = \left(1 + \frac{\lambda x}{n}\right)^n + O\left(\frac{\lambda^2 x^2}{n}\right) = \left(\frac{1 + \frac{\lambda x}{2n}}{1 - \frac{\lambda x}{2n}}\right)^n + O\left(\frac{\lambda^3 x^3}{n^2}\right) \quad (4.12)$$

in the case when the path from the origin to the point $x = \sum_1^n \frac{x}{n} = \sum \delta e^{i\theta_k}$ is not restricted to straight equal segments but to a general path of $O(|x|/\delta)$ segments of any directions.

The integration with respect to λ gives an interesting analogue of $Z^{-k}: \exp(:\lambda: Z)$. It is defined up to a globally defined discrete holomorphic map. One way to fix it is to integrate from a given λ_0 of modulus $2/\delta$, which is not a pole of the rational fraction, along a path that does not cross the circle of radius $2/\delta$ again.

Proposition 4.5. For pointwise multiplication, at every point $x \in \diamond_0$,

$$\exp(:\lambda: x) \cdot \exp(:-\lambda: x) = 1. \quad (4.13)$$

The specialization at $\lambda = 0$ defines monomials $Z^{:k:} := Z^{:k:} \exp(:\lambda: Z)|_{\lambda=0}$ which fulfill

$$Z^{:k:} = k \int Z^{:k-1:} dZ. \tag{4.14}$$

The anti-linear duality \dagger maps exponentials to exponentials:

$$\exp(:\lambda:)^\dagger = \exp(:\frac{4}{\delta^2 \bar{\lambda}}:). \tag{4.15}$$

In particular, $\exp(:\infty:) = 1^\dagger = \varepsilon$ is the bi-constant. Because discrete exponential and discrete monomials do not fulfill the usual pointwise multiplication relations $\exp(\lambda z) \exp(\mu z) = \exp((\lambda + \mu)z)$ for $\lambda \neq -\mu$, and $Z^k Z^\ell = Z^{k+\ell}$, we denote the arguments by $: \lambda :$ and $: k :$ to stress the problem behind the similarity.

Proof of Proposition 4.5. The first assertion is immediate.

The derivation of (4.9) with respect to λ yields

$$d \frac{\partial^k}{\partial \lambda^k} \exp(:\lambda: Z) = \left(\lambda \frac{\partial^k}{\partial \lambda^k} \exp(:\lambda: Z) + k \frac{\partial^{k-1}}{\partial \lambda^{k-1}} \exp(:\lambda: Z) \right) dZ \tag{4.16}$$

which implies (4.14).

Derivation of $\exp(:\lambda:)^\dagger$ gives,

$$\begin{aligned} (\exp(:\lambda:)^\dagger)' &= \frac{4}{\delta^2} \left(\int_O^z \exp(:\lambda:) dZ \right)^\dagger + \mu \varepsilon \\ &= \frac{4}{\delta^2} \left(\frac{\exp(:\lambda:) - 1}{\lambda} \right)^\dagger + \mu \varepsilon \\ &= \frac{4}{\delta^2 \bar{\lambda}} \exp(:\lambda:)^\dagger + \nu \varepsilon \end{aligned} \tag{4.17}$$

with μ, ν some constants, so that the initial condition $\exp(:\lambda: O)^\dagger = 1$ at the origin and the difference equation $d \exp(:\lambda:)^\dagger = \frac{4}{\delta^2 \bar{\lambda}} \exp(:\lambda:)^\dagger dZ$ yields the result. \square

Note that it is natural to define $\exp(:\lambda: (x - x_0)) := \frac{\exp(:\lambda: x)}{\exp(:\lambda: x_0)}$ as a function of x with x_0 a fixed vertex. It is simply a change of origin. But apart on a lattice where addition of vertices or multiplication by an integer can be given a meaning as maps of the lattice, there is no easy way to generalize this construction to other discrete holomorphic functions such as $\exp(:\lambda: (x + n y))$ with $x, y \in \diamond_0$ and $n \in \mathbb{Z}$.

4.8 Series

The series $\sum_{k=0}^\infty \frac{\lambda^k Z^{:k:}}{k!}$, wherever it is absolutely convergent, coincides with the rational fraction (4.11): Its value at the origin is 1 and it fulfills the defining difference equation (4.9). Using Equation (4.14), a Taylor expansion of $\exp(:\lambda: x)$ at $\lambda = 0$ gives back the same result. We are now interested in the growth rate of the monomials.

Direct analysis gives an estimate of $Z^{:k}$:

Proposition 4.6. For $x \in \diamond$, at a combinatorial distance $d(x, O)$ of the origin, and any $k \in \mathbb{N}$,

$$\left| \frac{Z^{:k}(x)}{k!} \right| \leq \left(\frac{\alpha + 1}{\alpha - 1} \right)^{d(x, O)} \left(\alpha \frac{\delta}{2} \right)^k, \quad (4.18)$$

for any $\alpha > 1$ arbitrarily close to 1.

Corollary 4.7. The series $\sum_{k=0}^{\infty} \frac{\lambda^k Z^{:k}}{k!}$ is absolutely convergent for $|\lambda| < \frac{2}{\delta}$.

Proof of Proposition 4.6. It is proved by double induction, on the degree k and on the combinatorial distance to the origin.

For $k = 0$, it is valid for any x since $\frac{\alpha+1}{\alpha-1} = 1 + \frac{2}{\alpha-1} > 1$, with equality only at the origin.

Consider $x \in \diamond$ a neighbor of the origin, $Z(x) = \delta e^{i\theta}$, then an immediate induction gives for $k \geq 1$,

$$\frac{Z^{:k}(x)}{k!} = 2 \left(\frac{\delta e^{i\theta}}{2} \right)^k \quad (4.19)$$

which fulfills the condition Equation (4.18) for any $k \geq 1$ because $\frac{\alpha+1}{\alpha-1} \alpha^k > 2$. This was done merely for illustration purposes since it is sufficient to check that the condition holds at the origin, which it obviously does.

Suppose the condition is satisfied for a vertex x up to degree k , and for its neighbor y , one edge further from the origin, up to degree $k - 1$. Then,

$$\frac{Z^{:k}(y)}{k!} = \frac{Z^{:k}(x)}{k!} + \frac{Z^{:k-1}(x) + Z^{:k-1}(y)}{(k-1)!} \frac{Z(y) - Z(x)}{2} \quad (4.20)$$

in absolute value fulfills

$$\begin{aligned} \left| \frac{Z^{:k}(y)}{k!} \right| &\leq \left(\frac{\alpha + 1}{\alpha - 1} \right)^{d(x, O)} \left(\alpha \frac{\delta}{2} \right)^{k-1} \left(\left(\alpha \frac{\delta}{2} \right) + \left(1 + \frac{\alpha + 1}{\alpha - 1} \right) \frac{\delta}{2} \right) \\ &= \left(\frac{\alpha + 1}{\alpha - 1} \right)^{d(x, O)} \left(\alpha \frac{\delta}{2} \right)^k \left(1 + \frac{2}{\alpha - 1} \right) \\ &= \left(\frac{\alpha + 1}{\alpha - 1} \right)^{d(y, O)} \left(\alpha \frac{\delta}{2} \right)^k, \end{aligned} \quad (4.21)$$

thus proving the condition for y at degree k . It follows by induction that the condition holds at any point and any degree. \square

4.9 Basis

The discrete exponentials form a basis of discrete holomorphic functions on a finite critical map: given any set of pairwise different reals $\{\lambda_k\}$ of the right dimension, the

associated discrete exponentials will form a basis of the space of discrete holomorphic functions. See [1] for the formula

$$f(x) = \frac{1}{2i\pi} \int_{\gamma} g(\lambda) \exp(\lambda : x) d\lambda \tag{4.22}$$

for a certain fixed contour γ in the space of parameters λ , and the definition of $g(\lambda)$ as a fixed contour integral in \diamond involving f .

The polynomials however do not form a basis in general: the combinatorial surface has to fulfill a certain condition called “combinatorial convexity”. A quadrilateral, when traversed from one side to its opposite, defines a unique chain of quadrilaterals, that we call a “train-track”. The condition we ask is that two different train-tracks have different slopes.

On a combinatorially convex set, the discrete polynomials form a basis as well.

4.10 Continuous limit

In a critical map, where quadrilaterals are mapped to rhombi of side δ , we identify a vertex x with its image $Z(x)$.

The combinatorial distance $d_{\diamond}(x, O)$ is related to the modulus $|x|$ through

$$d_{\diamond}(x, O) \frac{\sin \theta_m}{4} \leq \frac{|x|}{\delta} \leq d_{\diamond}(x, O) \tag{4.23}$$

where θ_m is the minimum of all rhombi angles. When the rhombi do not flatten, the combinatorial distance and the modulus (over δ) are equivalent distances.

Lemma 4.8. *Let $(ABCD)$ be a four-sided polygon of the Euclidean plane such that its diagonals are orthogonal and the vertices angles are in $[\eta, 2\pi - \eta]$ with $\eta > 0$. Let (M, M') be a pair of points on the polygon. There exists a path on $(ABCD)$ from M to M' of minimal length ℓ . Then*

$$\frac{MM'}{\ell} \geq \frac{\sin \eta}{4}.$$

It is a straightforward study of a several variables function. If the two points are on the same side, $MM' = \ell$ and $\sin \eta \leq 1$. If they are on adjacent sides, the extremal position with MM' fixed is when the triangle $MM'P$, with P the vertex of $(ABCD)$ between them, is isosceles. The angle in P being less than η , $\frac{MM'}{\ell} \geq \sin \frac{\eta}{2} > \frac{\sin \eta}{2}$. If the points are on opposite sides, the extremal configuration is given by Figure 9.2., where $\frac{MM'}{\ell} = \frac{\sin \eta}{4}$. □

A function $f: \diamond_0 \rightarrow \mathbb{C}$ on the combinatorial surface can be extended to a function on the image of the critical map in the complex plane $\hat{f}: U \rightarrow \mathbb{C}$ by stating that $\hat{f}(Z(x)) = f(x)$ for the image of a vertex, and extend it linearly on the segments $[Z(x), Z(y)]$ image of an edge, and harmonically inside each rhombus.



1. M, M' on adjacent sides.

2. M, M' on opposite sides.

Figure 9. The two extremal positions.

Theorem 4.2. *Let (\diamond_k) a sequence of simply connected critical maps, U the non-empty intersection of their images in the complex plane and a holomorphic function $f: U \rightarrow \mathbb{C}$. If the sequence of minimum angles are bounded away from 0 and the sequence of rhombi side lengths (δ_k) converge to 0, then the function f can be approximated by a sequence of discrete holomorphic functions $f_n \in \Omega(\diamond_k)$ converging to f . The convergence is not only pointwise but C^∞ on the intersection of images. Conversely a converging sequence of discrete holomorphic functions converges to a continuous holomorphic function, in particular the discrete polynomials and discrete exponentials with fixed parameters.*

Corollary 4.9. *On a Riemann surface, any 1-form can be approximated by a sequence of discrete holomorphic 1-forms on a refining sequence of critical maps with fixed conic singularities.*

The proof relies on the convergence of polynomials seen as iterated primitives of the constant function.

Lemma 4.10. *Given a sequence of discrete holomorphic functions (f_k) on a refining sequence of critical maps converging to a holomorphic function f , the sequence of primitives $(\int f_k dZ)$ converges to $\int f(z) dz$. Moreover, in the compact case, if the convergence of the functions is of order $O(\delta_k^2)$, the same property holds for the primitives.*

Proof. Suppose that we are given a sequence of flat vertices $O_k \in \diamond_k$ where the face containing the fixed flat origin $O \in U$ is adjacent to O_k . For a given integer k , let \hat{F}_k the continuous piecewise harmonic extension of the discrete primitive $\int_{O_k} f_k dZ$ to U . We want to prove that for any $x \in U$, the following sequence tends to zero

$$\left(\left| (\hat{F}_k(x) - \hat{F}_k(O)) - \int_O^x f(z) dz \right| \right)_{k \in \mathbb{N}}. \tag{4.24}$$

For each integer k consider a vertex $x_k \in \diamond_0$ on the boundary of the face of \diamond_2 containing x .

We decompose the difference (4.24) into three parts, inside the face containing the origin O and its neighbor O_k , similarly for x and x_k , and purely along the edges of the graph \diamond^k itself.

$$\begin{aligned} & \left| (\hat{F}_k(x) - \hat{F}_k(O)) - \int_O^x f(z) dz \right| \tag{4.25} \\ &= \left| (\hat{F}_k(x) - \hat{F}_k(x_k)) + \int_{O_k}^{x_k} f_k dZ + (\hat{F}_k(O) - \hat{F}_k(O_k)) - \int_O^x f(z) dz \right| \leq \\ &\leq \left| \hat{F}_k(x) - \hat{F}_k(x_k) - \int_{x_k}^x f(z) dz \right| + \left| \int_{O_k}^{x_k} f_k dZ - \int_{O_k}^{x_k} f(z) dz \right| \\ &\quad + \left| \hat{F}_k(O_k) - \hat{F}_k(O) - \int_O^{O_k} f(z) dz \right|. \end{aligned}$$

On the face of \diamond containing x , the primitive $\xi \mapsto \int_{x_k}^\xi f(z) dz$ is a holomorphic, hence harmonic function as well as $\xi \mapsto \hat{F}_k(\xi)$. By the maximum principle, the harmonic function $\xi \mapsto \hat{F}_k(\xi) - \hat{F}_k(x_k) - \int_{x_k}^\xi f(z) dz$ reaches its maximum on that face, along its boundary. The difference of the discrete primitive along the boundary edge $(x_k, y) \in \diamond_1$ at the point $\xi = (1 - \lambda)x_k + \lambda y$ is equal by definition to

$$\hat{F}_k((1 - \lambda)x_k + \lambda y) - \hat{F}_k(x_k) = \lambda(y - x_k) \frac{f_k(x_k) + f_k(y)}{2}. \tag{4.26}$$

The holomorphic f is differentiable with a bounded derivative on U , so averaging the first order expansions at x_k and y , we get

$$\begin{aligned} \int_{x_k}^\xi f(z) dz &= \lambda(y - x_k) \frac{f(x_k) + f(y)}{2} \\ &\quad + (y - x_k)^2 \frac{\lambda^2 f'(x_k) + (1 - \lambda)^2 f'(y)}{4} \\ &\quad + o((\xi - x_k)^3) + o((\xi - y)^3) \\ &= \lambda(y - x_k) \frac{f(x_k) + f(y)}{2} + O(\delta_k^2) \end{aligned} \tag{4.27}$$

therefore

$$|\hat{F}_k(x) - \hat{F}_k(x_k) - \int_{x_k}^x f(z) dz| = O(\delta_k^2). \tag{4.28}$$

Similarly for the term around the origin.

By definition of \hat{f}_k , the 1-form $\hat{f}_k(z) dz$ along edges of the graph \diamond is equal to the discrete form $f_k dZ$ so that $\int_{O_k}^{x_k} f_k dZ = \int_{O_k}^{x_k} \hat{f}_k(z) dz$ on a path along \diamond edges. Therefore the difference

$$\left| \int_{O_k}^{x_k} f_k dZ - \int_{O_k}^{x_k} f(z) dz \right| \leq \int_{O_k}^{x_k} |(\hat{f}_k(z) - f(z)) dz| \tag{4.29}$$

is of the same order as the difference $|f_k(z) - f(z)|$ times the length $\ell(\gamma_k)$ of a path on \diamond_k from O_k to x_k . This length is bounded as $\ell(\gamma_k) \leq \frac{4}{\sin \theta_m} |x_k - O_k|$. Since we are interested in the compact case, this length is bounded uniformly and the difference (4.29) is of the same order as the pointwise difference. We conclude that the sequence of discrete primitives converges to the continuous primitive and if the limit for the functions was of order $O(\delta^2)$, it remains of that order. \square

The discrete polynomials of degree less than three agree pointwise with their continuous counterpart, $Z^{:2:}(x) = Z(x)^2$.

A simple induction then gives the following

Corollary 4.11. *The discrete polynomials converge to the continuous ones, the limit is of order $O(\delta_k^2)$.*

Which implies the main theorem:

Proof of Theorem 4.2. On the simply connected compact set U , a holomorphic function f can be written, in a local map z as a series,

$$f(z) = \sum_{k \in \mathbb{N}} a_k z^k. \quad (4.30)$$

Therefore, by a diagonal procedure, there exists an increasing integer sequence $(N(n))_{n \in \mathbb{N}}$ such that the sequence of discrete holomorphic polynomials converges to the continuous series and the convergence is C^∞ ;

$$\left(\sum_{k=0}^{N(n)} a_k Z^{:k:} \right)_{n \in \mathbb{N}} \rightarrow f. \quad (4.31)$$

\square

4.11 Cross-ratio preserving maps

Once the isometry Z is chosen, holomorphicity of a function f can be written on a quadrilateral $(x, y, x', y') \in \diamond_2$, writing $x = Z(x)$ for a vertex $x \in \diamond_0$, as

$$\frac{f(y') - f(y)}{f(x') - f(x)} = \frac{y' - y}{x' - x} \quad (4.32)$$

and f is understood as a *diagonal ratio preserving map*, and each value at a corner vertex can be linearly solved in terms of the three others.

A quadratic version is given by the *cross-ratio preserving maps*: A function f is said to be *quadratic holomorphic* iff

$$\frac{(f(y) - f(x))(f(y') - f(x'))}{(f(x) - f(y'))(f(x') - f(y))} = \frac{(y - x)(y' - x')}{(x - y')(x' - y)}. \quad (4.33)$$

A rhombic tiling gives rise to two sets of isoradial circle patterns: a set of circles of common radius δ , whose centers are the vertices of Γ and intersections are the vertices of Γ^* and vice-versa. Two interesting families of cross-ratio preserving maps are given by circle patterns with the same combinatorics and intersection angles as one of these two *circle patterns*.

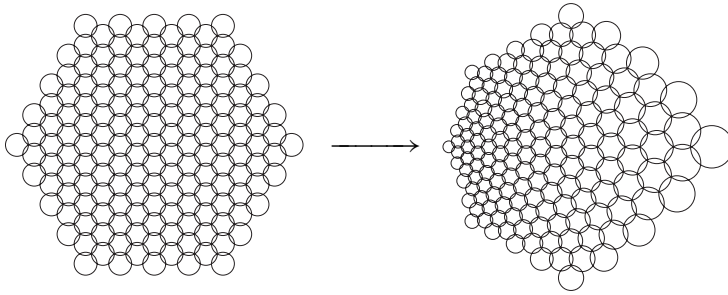


Figure 10. A circle pattern with prescribed angles as a cross ratio preserving map.

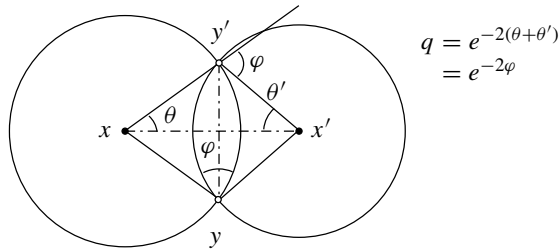


Figure 11. The cross-ratio q is given by the intersection angles.

A change of coordinates helps understanding diagonal ratio preserving maps as a linearized version of the cross-ratio preserving maps. We will say that the function $w : \diamond_2 \rightarrow \mathbb{C}$ solves the *Hirota system* if, around a face $(x, y, x', y') \in \diamond_2$,

$$(y - x)w(x)w(y) + (x - y')w(y')w(x) + (y' - x')w(x')w(y') + (x' - y)w(y)w(x') = 0. \tag{4.34}$$

This is to be understood as a quadratic version of the Morera theorem $\oint f dz = 0$ and w is a half of the derivative of a holomorphic function:

Proposition 4.12. *If w solves the Hirota system, then the function $f : \diamond_2 \rightarrow \mathbb{C}$ defined up to an additive constant by*

$$f(y) - f(x) = (y - x)w(x)w(y) \tag{4.35}$$

is quadratic holomorphic.

Proof. The function f is well defined because the associated 1-form is closed by definition of the Hirota system. The function w disappears in the cross-ratio of f , leaving the original cross-ratio. \square

Conversely, a quadratic holomorphic function defines a solution to the Hirota system, unique up to multiplication by λ on Γ , $1/\lambda$ on Γ^* . Concerning circle patterns families, w is real on the centers and unitary on the intersections, and encodes the variation of radius, resp. of direction of the image of the circle:

$$f(y) - f(x) = r(x)e^{i\theta(y)}(y - x). \quad (4.36)$$

Proposition 4.13. *The logarithmic derivative of the Hirota system associated to a family of cross-ratio preserving maps is a diagonal ratio preserving map.*

In other words, for $(1 + \varepsilon g)w$ to continue solving the Hirota system at first order, the deformation g must satisfy

$$\frac{g(y') - g(y)}{g(x') - g(x)} = \frac{f(y') - f(y)}{f(x') - f(x)}. \quad (4.37)$$

Proof of Proposition 4.13. The ε contribution of the closeness condition (4.34) for $(1 + \varepsilon g)w$ gives

$$(g(x) + g(y)) f(x) f(y) (y - x) + (g(y) + g(x')) f(y) f(x') (x' - y) \\ + (g(x') + g(y')) w(x') w(y') (y' - x') + (g(y') + g(x)) w(y') w(x) (x - y') = 0,$$

which reads, referring to f :

$$\frac{g(y') - g(y)}{g(x') - g(x)} = \frac{f(y') - f(y)}{f(x') - f(x)}. \quad \square$$

4.12 Bäcklund transformation

The Bäcklund transformation is a way to associate, to a given solution of an integrable problem, a family of deformed solutions. The two problems under consideration here are the linear and quadratic holomorphicity constraints on each face. They are given by a linear, resp. quadratic algebraic relation between the four values of a solution at the vertices of each face. These relations involve only values supported by the edges of the rhombus, which are equal on opposite sides, namely the complex label $y - x = x' - y'$.

The Bäcklund transformation is defined by imposing such constraints over new virtual faces added over each edge, with “vertical edges” labelled by a complex constant λ :

Definition 4.14. Given a linear holomorphic function $f \in \Omega(\diamond)$, complex numbers $u, \lambda \in \mathbb{C}$, its Bäcklund transformation $f_\lambda = B_\lambda^u(f)$ is defined by

$$\begin{aligned} f_\lambda(0) &= u, \\ \frac{f_\lambda(x) - f(y)}{f_\lambda(y) - f(x)} &= \frac{\lambda + x - y}{\lambda + y - x}. \end{aligned} \tag{4.38}$$

Given a quadratic holomorphic function f , complex numbers $u, \lambda \in \mathbb{C}$, its Bäcklund transformation $f_\lambda = B_\lambda^u(f)$ is defined by

$$\begin{aligned} \lambda(0) &= u, \\ \frac{f_\lambda(y) - f_\lambda(x)}{f_\lambda(x) - f(x)} \frac{f(x) - f(y)}{f(y) - f_\lambda(y)} &= \frac{(y-x)^2}{\lambda^2}. \end{aligned} \tag{4.39}$$

The right hand sides are the values respectively of the diagonal ratio and cross-ratio of a parallelogram faces of sides $(y-x)$ and λ seen as “over” the edge $(x, y) \in \diamond_1$.

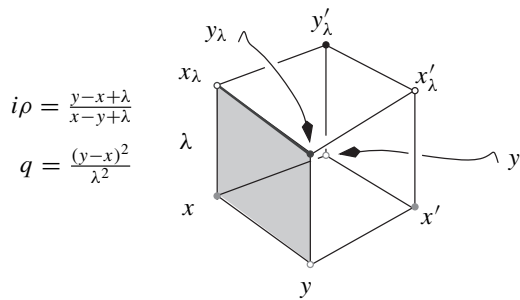


Figure 12. The face $(x_\lambda, y_\lambda, x'_\lambda, y'_\lambda)$ “over” the face $(x, y, x', y') \in \diamond_2$.

Proposition 4.15. *This transformation is well defined in the critical case.*

This condition, called *three dimensional consistency* is an over-determination constraint: if the cube “over” the face $(x, y, x', y') \in \diamond_2$ is split into two hexagons along the cycle $(y, x', y', y'_\lambda, x_\lambda, y_\lambda)$, one can see that, given values at these six vertices, the values at the centers of each hexagons, namely at x'_λ and at x are overdetermined.

Therefore only certain values at the six vertices are allowed, defined by two algebraic relations between them. The compatibility condition is that these two algebraic relations are equivalent. It is a simple computation to check it is the case for critical maps.

This transformation verifies

$$B_{\lambda^{-1}}^{f(O)}(B_\lambda^u(f)) = f \tag{4.40}$$

for any (u, λ) . It is an analytic transformation in all the parameters therefore its derivative is a linear map between the tangent spaces, that is to say between diagonal

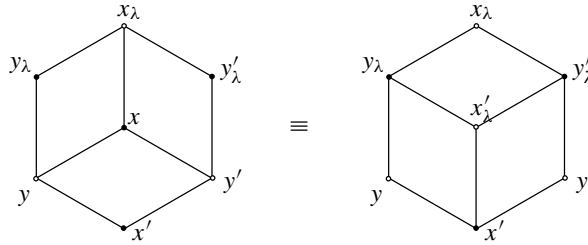


Figure 13. The cube split into two hexagons yielding equivalent compatibility constraints.

ratio preserving maps,

$$d B_\lambda^u(f): \Omega(f) \rightarrow \Omega(B_\lambda^u(f)). \tag{4.41}$$

It is not injective and I define the discrete exponential at f as being the direction of this 1-dimensional kernel. It can be characterized as a derivative with respect to the initial value at the origin:

$$\exp_u(:\lambda:f) := \frac{\partial}{\partial v} B_{\lambda-1}^v(B_\lambda^u(f))|_{v=f(O)} \in \ker(d B_\lambda^u(f)) \tag{4.42}$$

because $B_\lambda^u(B_{\lambda-1}^v(g)) = g$ for all λ, g and v .

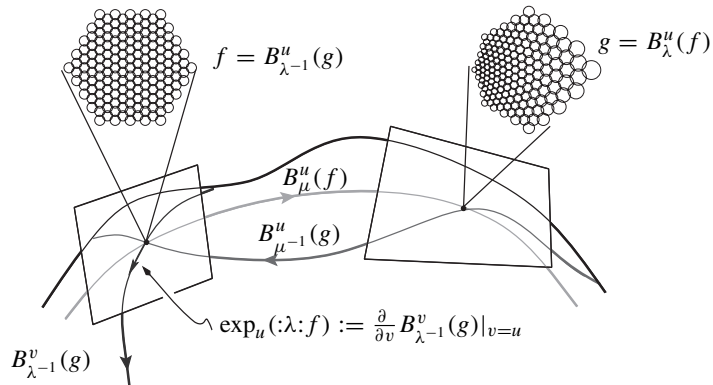


Figure 14. The discrete exponential $\exp_u(:\lambda:f)$ is the kernel of the linear transformation $d B_\lambda^u(f)$ (here $u = f(O)$).

As in the discrete exponential case, the value of the Bäcklund transformation at a given vertex is the image of values at neighbouring vertices by a homography. These homographies can be encoded as projective operators $L(e; \lambda) \in GL_2(\mathbb{C})[\lambda]$ lying on

the edges $e \in \diamond_2$, called a *zero curvature representation*:

$$L((x, y); \lambda) = \begin{pmatrix} \lambda + y - x & -2(y - x)(f(x) + f(y)) \\ 0 & \lambda + x - y \end{pmatrix} \quad \text{for the linear case,} \quad (4.43)$$

$$L((x, y); \lambda) = \begin{pmatrix} 1 & -(y - x)w(y) \\ -\lambda(y - x)/w(x) & w(y)/w(x) \end{pmatrix} \quad \text{for the Hirota system.} \quad (4.44)$$

Then we define [15] the *moving frame* $\Psi: \diamond_2 \rightarrow \text{GL}_2(\mathbb{C})(\lambda)$ by a prescribed value at the origin and recursively by $\Psi(y; \lambda) = L((x, y); \lambda)\Psi(x; \lambda)$ and its logarithmic derivative with respect to λ

$$A(e; \lambda) = \frac{d\Psi(e; \lambda)}{d\lambda} \Psi^{-1}(e; \lambda) \quad (4.45)$$

is meromorphic in λ for each edge e . We call f , resp. w , *isomonodromic* if the positions and orders of the poles do not depend on the edge e . The two points discrete Green function (the discrete logarithm) $G(O, x)$ is the inverse of the Laplacian in the sense that

$$\Delta G(O, \bullet) = \delta_{O, \bullet}. \quad (4.46)$$

It can be constructed as the unique isomonodromic solution with some prescribed data [1], which allows us to give an explicit formula for it, recovering results of Kenyon [8]: an integral over a loop in the space of discrete exponentials,

$$G(O, x) = -\frac{1}{8\pi^2 i} \oint_C \exp(:\lambda: x) \frac{\log \frac{\delta}{2} \lambda}{\lambda} d\lambda \quad (4.47)$$

where the integration contour C contains all the possible poles of the rational fraction $\exp(:\lambda: x)$ but avoids the half line through $-x$. It is real (negative) on half of the vertices and imaginary on the others. Because of the logarithm, this imaginary part is multivalued.

References

- [1] A. I. Bobenko, Ch. Mercat, and Y. B. Suris, Linear and nonlinear theories of discrete analytic functions. Integrable structure and isomonodromic Green's function. *J. Reine Angew. Math.* 583 (2005), 117–161. [542](#), [556](#), [566](#), [574](#)
- [2] A. I. Bobenko, Discrete differential geometry. Integrability as consistency. In *Discrete integrable systems*, Lecture Notes in Phys. 644, Springer-Verlag, Berlin 2004, 85–110. [542](#)
- [3] R. Costa-Santos and B. M. McCoy, Dimers and the critical Ising model on lattices of genus > 1 . *Nuclear Phys. B* 623 (3) (2002), 439–473. [558](#)

- [4] R. Costa-Santos and B. M. McCoy, Finite size corrections for the Ising model on higher genus triangular lattices. *J. Statist. Phys.* 112 (2003), 889–920. [559](#)
- [5] R. J. Duffin, Basic properties of discrete analytic functions. *Duke Math. J.* 23 (1956), 335–363. [557](#)
- [6] R. J. Duffin, Potential theory on a rhombic lattice. *J. Combin. Theory* 5 (1968), 258–272. [557](#)
- [7] H. M. Farkas and I. Kra, *Riemann surfaces*. Second edition, Grad. Texts in Math. 71, Springer-Verlag, New York 1992. [542](#), [548](#), [555](#)
- [8] R. Kenyon, The Laplacian and Dirac operators on critical planar graphs. *Invent. Math.* 150 (2) (2002), 409–439. [542](#), [561](#), [563](#), [574](#)
- [9] J. Lelong-Ferrand, *Représentation conforme et transformations à intégrale de Dirichlet bornée*. Gauthier-Villars, Paris 1955. [563](#)
- [10] Ch. Mercat, Holomorphic discrète et modèle d’Ising. PhD thesis, Université Louis Pasteur, Strasbourg, France, 1998; available at <http://www-irma.u-strasbg.fr/annexes/publications/pdf/98014.pdf> [542](#), [548](#), [556](#)
- [11] Ch. Mercat, Discrete Riemann surfaces and the Ising model. *Comm. Math. Phys.* 218 (1) (2001), 177–216. [542](#), [556](#), [557](#)
- [12] Ch. Mercat, Exponentials form a basis of discrete holomorphic functions on a compact. *Bull. Soc. Math. France* 132 (2004), 305–326. [542](#)
- [13] Ch. Mercat, Discrete period matrices and related topics. Preprint, 2002; arXiv:math-ph/0111043. [563](#)
- [14] D. Mumford, *Tata lectures on theta*. I. Progr. Math. 28, Birkhäuser, Boston, MA, 1983. [542](#)
- [15] F. W. Nijhoff, A. Ramani, B. Grammaticos, and Y. Ohta, On discrete Painlevé equations associated with the lattice KdV systems and the Painlevé VI equation. *Stud. Appl. Math.* 106 (3) (2001), 261–314. [574](#)
- [16] M. Troyanov, Les surfaces euclidiennes à singularités coniques. *Enseign. Math.* (2) 32 (1–2) (1986), 79–94. [543](#)

Part D

The quantum theory, 1

Chapter 14

On quantizing Teichmüller and Thurston theories

Leonid O. Chekhov and Robert C. Penner

Contents

1	Introduction	580
2	Classical Teichmüller spaces	584
	2.1 Graph description of Teichmüller spaces	585
	2.1.1 Global coordinates on Teichmüller space	585
	2.1.2 Weil–Petersson form	588
	2.1.3 Mapping class group description using graphs	589
	2.2 Poisson algebras of geodesic functions	591
	2.2.1 Multicurves	591
	2.2.2 Classical skein relation	591
	2.2.3 Poisson brackets for geodesic functions	592
	2.2.4 Poisson geodesic algebras for higher genera	593
3	Quantization	594
	3.1 Quantizing Teichmüller spaces	594
	3.2 Geodesic length operators	599
	3.3 Algebra of quantum geodesics	601
	3.4 Quantizing the Nelson–Regge algebras	603
	3.5 Improving the quantum ordering	603
4	Classical Thurston theory of surfaces	606
	4.1 Measured foliations and Thurston’s boundary	606
	4.2 Train tracks	611
	4.3 Laminations	615
	4.4 Dynamics on train tracks	616
	4.5 Decorated measured foliations and freeways	618
	4.6 Shear coordinates for measured foliations	620
5	On quantizing Thurston theory	622
	5.1 Proper length of geodesics	622
	5.2 Approximating laminations and the main theorem	623
	5.3 Elements of the proof	626
	5.3.1 Continued fraction expansion	626
	5.3.2 Mapping class group transformations and the unzipping procedure	631

5.3.3 Asymptotic regime	632
5.3.4 Quantum continued fraction expansion	636
6 Conclusion	638
Appendix A. Combinatorial proof of Theorem 5.1	639
Appendix B. Degeneracy of the Poisson structure	641
References	643

1 Introduction

One manifestation of Teichmüller space in contemporary mathematical physics is as the Hilbert space and the algebra of observables for three-dimensional (3D) quantum gravity since E. Verlinde and H. Verlinde [39] have argued that the classical phase space of Einstein gravity in a 3D manifold is the Teichmüller space of its boundary. (Analogously, the classical phase space for 3D Chern–Simons theory is the moduli space of flat connections on the boundary; this theory was quantized in [12], [13].) Teichmüller space possesses its canonical (Weil–Petersson) Poisson structure, whose symmetry group is the mapping class group of orientation-preserving homeomorphisms modulo isotopy. The algebra of observables is the collection of geodesic length functions of geodesic representatives of homotopy classes of essential closed curves together with its natural mapping class group action.

Given this Poisson structure, one can turn to the problem of quantizing it, thereby obtaining a variant of the quantum 3D gravity description. According to the correspondence principle: (1) the algebra of observables of the corresponding quantum theory is the noncommutative deformation of the $*$ -algebra of functions on it governed by the Poisson structure; (2) the Hilbert spaces of the theory are the representation spaces of these $*$ -algebras; and (3) the symmetry group acts on the algebra of observables by automorphisms. Under the assumption that the quantization of a Poisson manifold exists and is unique, to solve this problem it suffices to construct a family of $*$ -algebras, which depend on the quantization parameter \hbar , and an action of the mapping class group on this family by outer automorphisms, and to show that the algebra and the action thus constructed reproduces the classical algebra, the classical action, and the classical Poisson structure in the limit $\hbar \rightarrow 0$. This program has been successfully performed in [6] and [16], and we describe it in Sections 2 and 3 of this chapter. Actually, the problem that was solved in these papers differs slightly from the original formulation because the methods of [6] are suited for describing only *open* surfaces (surfaces with nonempty boundary, components of which can, however, reduce to a puncture). The corresponding Teichmüller space has a degenerate Weil–Petersson Poisson structure, while the mapping class group acts as symmetry group. We describe the deformation quantization of the corresponding Teichmüller space, the action of the mapping class group by outer automorphisms, the representations of the observable algebra, and the induced action of the mapping class group on the representation space following [6].

As in [39], the representation space of the observable algebra can also be interpreted as the space of conformal blocks of the Liouville conformal field theory. This program is under current development (see [33] and [34]). Our construction can be therefore interpreted as the construction of the conformal block spaces and the mapping class group actions for this CFT.

The key point of the quantization procedure is constructing quantum mapping class group transformations that define in a consistent way the morphisms between quantum $*$ -algebras simultaneously preserving the quantum geodesic algebra. The main mathematical ingredient of the construction is a version of the quantum dilogarithm by L. D. Faddeev [8]. We interpret the corresponding five-term relation as the only nontrivial relation in a certain groupoid that has the mapping class group as maximal subgroup. A similar construction has been made independently and simultaneously by R. M. Kashaev [16]. The key difference between these two constructions lies in the dimensions of the Poisson leaves of the two theories: given a graph with v three-valent vertices, e edges, and $f = s$ faces suitably embedded in a surface F_g^s of genus g with s ideal boundary components, the genus is given by Euler's formula $v - e + f = 2 - 2g$. In Kashaev's approach, there are $2e$ variables and $v + f$ central elements (at each vertex and at each face), so the Poisson leaf dimension is $2e - v - f = 8g - 8 + 3s$, while in our approach, there are e variables and f central elements, so the Poisson dimension $e - f = 6g - 6 + 2s$ exactly coincides with the dimension of the Teichmüller space of Riemann surfaces of genus g with s punctures. The approach of [6] is thus appropriate for describing 2D topological theories while that of [16] is suitable for describing Liouville field theory as a lattice theory (for instance, the Liouville field central charge can be calculated, see [17]).

One of the mathematical tools employed in the quantization [6] is the decorated Teichmüller theory [25], and the relevant aspects are briefly reviewed in Section 2. In effect in the classical case, to each edge of a trivalent “fatgraph” Γ (i.e., a graph plus a cyclic ordering of the half-edges about each vertex) embedded as a “spine” (i.e., a deformation retract of the surface) $F = F_g^s$ is assigned a number Z_α , where α here and below indexes the edges of Γ . The tuple (Z_α) gives global coordinates on an appropriate Teichmüller space $\mathcal{T}_H = \mathcal{T}_H(F)$ of F (as first studied by Thurston and later by Fock), where any or all of the “punctures” of F_g^s are permitted to be uniformized instead as circular boundary components (and the subscript H stands for “holes”); the details are given in Section 2.1. The Weil–Petersson Kähler two-form, which is known [27] in the (Z_α) coordinates, pulls back to a degenerate two-form, i.e., to a degenerate Poisson structure on \mathcal{T}_H . Furthermore, the action of the mapping class group $\text{MC} = \text{MC}(F)$ of F on \mathcal{T}_H , which is again known [27] in the (Z_α) coordinates as well as combinatorially [28], leaves invariant this Poisson structure, and this is the Poisson manifold \mathcal{T}_H with MC-action which has been quantized.

More explicitly still, for each edge α , we may associate a pair of Möbius transformations $R_{Z_\alpha}, L_{Z_\alpha}$ depending upon Z_α with the following property. For any homotopy class γ of geodesic in F with corresponding closed edge-path P on Γ , consider the serial product P_γ of operators $R_{Z_\alpha}, L_{Z_\alpha}$ taken in order as one traverses P in some

orientation from some starting point, where one inserts the former (or latter, respectively) if immediately after traversing edge α , then P turns right (or left) in Γ ; thus, the combinatorial geometry of serially traversing edges of Γ as dictated by P determines an ordered product P_γ of matrices depending upon (Z_α) , and the length l_γ of the geodesic representative of γ for the point of \mathcal{T}_H determined by (Z_α) is given by $G_\gamma = 2 \cosh(l_\gamma/2) = \text{tr } P_\gamma$. It is the Poisson algebra of these geodesic functions G_γ , the algebra of observables, which has been quantized.

In the quantum case (after passing to a suitable subspace on which the Poisson structure is non-degenerate), standard techniques of deformation quantization produce the appropriate Hilbert space \mathcal{H} of the quantum theory, as well as pairs of operators $R_{Z_\alpha}, L_{Z_\alpha}$ on \mathcal{H}^2 for each edge α of Γ . Again, to the homotopy class of a geodesic γ in F or its corresponding closed edge-path P on Γ , we may assign the ordered product P_γ of these operators as dictated by the combinatorial geometry of P in Γ , whose trace $G_\gamma = \text{tr } P_\gamma$ is the “quantum geodesic operator.” The main point is to prove the invariance under the action of MC, which is intimately connected with functional properties of the quantum dilogarithm as was mentioned before.

In [6] was proved the existence and uniqueness of an appropriate “proper quantum ordering” of operators that enjoy MC-invariance as well as satisfy the standard physical requirements. In Section 3.5, we observe that the natural operatorial ordering given by the combinatorial geometry of edge-paths in Γ can be used to derive this physically correct quantum ordering. The improved ordering is required in the subsequent quantization (discussed below).

In order to explain the further new results in this chapter, we must recall aspects of Thurston’s seminal work on surface geometry, topology, and dynamics from the 1970–1980s, which is surveyed in Section 4. Very briefly, Thurston introduced the space $\mathcal{P}\mathcal{F}_0 = \mathcal{P}\mathcal{F}_0(F)$ of “projective measured foliations of compact support in F ” as a boundary for the Teichmüller space $\mathcal{T} = \mathcal{T}(F)$, where “Thurston’s compactification” $\overline{\mathcal{T}} = \mathcal{T} \cup \mathcal{P}\mathcal{F}_0$ is a closed ball with boundary sphere $\mathcal{P}\mathcal{F}_0$, where the action of MC on \mathcal{T} extends continuously to the natural action on $\overline{\mathcal{T}}$. Furthermore, the sphere $\mathcal{P}\mathcal{F}_0$ contains the set of all homotopy classes of geodesics in F as a dense subset, i.e., $\mathcal{P}\mathcal{F}_0$ is an appropriate completion of this set. (Unfortunately, the action of MC on $\mathcal{P}\mathcal{F}_0$ has dense orbits, so the beautiful structure of Thurston’s compactification does not descend in any reasonable way, with the current state of understanding, to a useful structure on the level of Riemann’s moduli space.) Thurston also devised an elegant graphical formalism for understanding $\mathcal{P}\mathcal{F}_0$ using “train tracks”, which are graphs embedded in F with the further structure of a “branched one-submanifold” (cf. Section 4.2). In effect, a maximal train track gives a chart on the sphere $\mathcal{P}\mathcal{F}_0$, and furthermore, the combinatorial expression for inclusion of charts effectively captures the dynamics of the action of diffeomorphisms of F (cf. Section 4.4).

Thus, Thurston theory arises as a natural tool to understand degenerations in Teichmüller space and dynamics on F , and in light of remarks above, its quantization should provide a natural tool for studying degenerations of 3D gravity or Liouville conformal

field theory. Many aspects of the survey of classical Thurston surface theory (in Section 4) are required for our subsequent quantization.

In any case, a mathematically natural problem armed with [6] is to “quantize Thurston theory”: assign operators H_γ on \mathcal{H} to each homotopy class γ of geodesic in such a way that as γ converges in \mathcal{PF}_0 to a projectivized measured foliation $[\mathcal{F}, \bar{\mu}]$, then the corresponding operators H_γ converge in an appropriate sense to a well-defined operator $H_{[\mathcal{F}, \bar{\mu}]}$ on \mathcal{H} . Upon choosing a spine $\Gamma \subseteq F$ (i.e., an embedding, up to homotopy, of a graph whose inclusion is a homotopy equivalence), any homotopy class of curve γ may be essentially uniquely realized as an edge-path on Γ . Define the “graph length” $\text{g.l.}(\gamma)$ to be the total number of edges of Γ traversed by this edge-path counted with multiplicities.

In the special case of the torus F_1^1 , we have succeeded here (in Section 5) in showing that the ratio of operators

$$H_\gamma = \text{p.l.}(\gamma)/\text{g.l.}(\gamma)$$

converge weakly to a well-defined operator as γ converges in $\mathcal{PF}_0(F_1^1)$, where the “proper length” is defined by

$$\text{p.l.}(\gamma) = \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr} \log 2T_n \left(\frac{1}{2} P_\gamma \right),$$

with T_n the Chebyshev polynomials. In fact in the classical case, $\text{p.l.}(\gamma)$ agrees with half the length of the geodesic homotopic to γ in the Poincaré metric, and in the quantum case, there is an appropriate operatorial interpretation, both of which are described in Section 5.1. In particular, for several spines whose corresponding charts cover the circle $\mathcal{PF}_0(F_1^1)$, the analysis involves rather intricate estimates. This leads to a natural operatorial quantization (in Section 5.3.4) of the standard simple continued fractions, which are intimately connected with Thurston theory on F_1^1 (as we shall describe in Section 4.5). To complete the basic theory on the torus, one would like an intrinsic operatorial description of the circle of unbounded operators we have constructed, as well as an explication of the mapping class group action on it, viz., Thurston’s classification of surface automorphisms.

This chapter is organized as follows: Section 2 covers classical Teichmüller theory and Section 3 the Chekhov–Fock quantization; Section 4 surveys classical Thurston theory of surfaces, and Section 5 gives our quantization of Thurston’s boundary for the punctured torus. Excerpts of Sections 2 and 3 are derived from an earlier manuscript of Chekhov–Fock, and we strive to include further mathematical detail. Section 4 surveys aspects of a large literature on Thurston theory and train tracks, explicitly covering only what is required in Section 5. Section 5 should be regarded as work in physics in the sense that some of the formal calculations depend upon manipulations of asymptotic spectral expansions for which there may be remaining mathematical issues. Closing remarks in Section 6 discuss the natural extension of these results to more complicated surfaces as well as other related work. Appendix A includes a novel proof of the required convergence in the classical case for any surface (in a sense, a new proof of the existence of Thurston’s compactification), which may yet be useful

in the quantum case. Appendix B contains an analysis of the Casimir operators in the Poisson algebra in the Z -variables and the appropriate diagonalization of Poisson structure.

Acknowledgements. We are indebted to L. D. Faddeev, V. V. Fock, M. Lapidus, and F. Bonahon for useful discussions. The work was partially supported by the RFFI Grant No. 01-01-00549 (L.Ch.), by the Program Mathematical Methods in Nonlinear Dynamics (L.Ch.), and by the COBASE Project.

2 Classical Teichmüller spaces

To begin, we shall briefly recall the two related roles played by graphs in Teichmüller theory as both aspects will be required here.

A *fatgraph* or *ribbon graph* is a graph Γ together with a cyclic ordering on the half-edges incident on each vertex, and we canonically associate to Γ a surface $F(\Gamma)$ with boundary obtained by “fattening each edge of the graph into a band” in the natural way; we shall tacitly require all vertices to have valence at least three unless stated otherwise, and we shall call a fatgraph *cubic* if each vertex has valence three. To each homotopy class of homotopy equivalence $\iota: \Gamma \rightarrow F$ for some surface F , where ι respects the orientation, there is a corresponding cell in $\mathcal{T}_g^s \times \mathbb{R}_{>0}^s$ as explained in [29] in the hyperbolic setting [25] and in the conformal setting [32]. Thus, a homotopy class of $\iota: \Gamma \rightarrow F$ is the name of a cell in the canonical cell decomposition. We shall sometimes suppress the mapping $\iota: \Gamma \rightarrow F$ and refer to Γ itself as a *spine* of F , where Γ is identified with $\iota(\Gamma) \subseteq F$. The cell decomposition is invariant under the action of the mapping class group (induced by post-composition of ι with homeomorphisms), and this has been an effective tool for studying Riemann’s moduli space; for instance, we shall recall here the corresponding presentation of the mapping class groups.

The second role of fatgraphs is exclusive to the hyperbolic setting, namely, fatgraphs provide a kind of “basis” for geometrically natural global parameterizations of Teichmüller space. Specifically, fix a homotopy class $\iota: \Gamma \rightarrow F$ as above, where we now demand that Γ is cubic, and let $E = E(\Gamma)$ denote the set of edges of Γ . In several different contexts, one can naturally identify $\mathbb{R}_{>0}^E$ with a suitable modification of an appropriate Teichmüller space of F .

For instance, for punctured surfaces, recall [25] that the lambda length of a pair of horocycles is $\sqrt{2} e^\delta$, where δ is the signed hyperbolic distance between the horocycles. Lambda lengths give a global real-analytic parametrization of the decorated Teichmüller space as the trivial bundle $\tilde{\mathcal{T}}_g^s = \mathcal{T}_g^s \times \mathbb{R}_{>0}^s$ over Teichmüller space, where the fiber over a point is the space of all s -tuples of horocycles in the surface, one horocycle about each puncture (parameterized by hyperbolic length).

For another example, Thurston’s shear coordinates [36], [2] give global parameters not only on Teichmüller space (cf. Section 2.1.1) but also on the related space of measured foliations (cf. Section 4.6).

On the level of Teichmüller space, the two global coordinate systems (lambda lengths and shear coordinates) are closely related, and we choose to give the exposition here principally in shear coordinates with the parallel lambda length discussion relegated to a series of ongoing remarks. On the other hand, certain proofs of identities involving shear coordinates are easy calculations in lambda lengths.

2.1 Graph description of Teichmüller spaces

2.1.1 Global coordinates on Teichmüller space. In addition to the Teichmüller space \mathcal{T}_g^s and decorated Teichmüller space $\tilde{\mathcal{T}}_g^s$, we shall also require the following modification. Given an open Riemann surface F of finite topological type, a neighborhood of an ideal boundary component is either an annulus or a punctured disk; in the former case, the ideal boundary component will be called a “true” boundary component and in the latter will be called a “puncture.” We shall study the latter as a degeneration of the former with an elaboration

$$\mathcal{T}eich(F) = \text{Hom}'(\pi_1(F), \text{PSL}_2(\mathbb{R})) / \text{PSL}_2(\mathbb{R})$$

of the usual Teichmüller space, where Hom' denotes the space of all discrete faithful representations with no elliptic elements, i.e., $|\text{tr}(\rho(\gamma))| \geq 2$ for all $\gamma \in \pi_1(F)$ for any representation $\rho: \pi_1(F) \rightarrow \text{PSL}_2(\mathbb{R})$.

Assume that $\iota: \Gamma \rightarrow F$ is a homotopy class of homotopy equivalence and $\gamma \in \pi_1(F)$ is conjugate in $\pi_1(F)$ to the boundary of a regular neighborhood of an ideal boundary component of $F(\Gamma)$. Thus, the ideal boundary component is a puncture if and only if $\rho(\gamma)$ is a parabolic transformation, i.e., $|\text{tr}(\rho(\gamma))| = 2$. For any $\gamma \in \pi_1(F)$ with $|\text{tr}(\rho(\gamma))| > 2$, the underlying free homotopy class of unbased curves contains a unique hyperbolic geodesic whose length l_γ is given by $G_\gamma = 2 \cosh(l_\gamma) = |\text{tr} \rho(\gamma)|$, where G_γ is called the *geodesic operator* and is constant on the conjugacy class of $\gamma \in \pi_1(F)$. Furthermore, by definition,

$$\begin{aligned} \mathcal{T}_g^s = \{[\rho] \in \mathcal{T}eich(F) : \text{for all } \gamma \in \pi_1(F) \text{ freely homotopic into} \\ \text{the boundary, we have } |\text{tr}(\rho(\gamma))| = 2\} \\ \subseteq \mathcal{T}eich(F) \end{aligned}$$

Finally, define the space $\mathcal{T}_H(F)$ to be the 2^s -fold cover of $\mathcal{T}eich(F)$ branched over \mathcal{T}_g^s , where the fiber is given by the set of all orientations on the boundary components of F .

Theorem 2.1. *Fix any spine $\Gamma \subseteq F$, where Γ is a cubic fatgraph. Then there is a real-analytic homeomorphism $\mathbb{R}^{E(\Gamma)} \rightarrow \mathcal{T}_H(F)$. The hyperbolic length l_γ of a true boundary component γ is given by $l_\gamma = |\sum Z_i|$, where the sum is over the set of all edges traversed by γ counted with multiplicity. Furthermore, $\sum Z_i = 0$ if and only if the corresponding ideal boundary component is a puncture, so $\mathcal{T}_g^s \subseteq \mathcal{T}_H(F)$ is determined by s independent linear constraints.*

The theorem is due to Thurston with a systematic study by Fock. We shall not give a proof here (though there is not a complete proof in the literature so far as we know), but we shall at least give the construction that defines the homeomorphism $\mathbb{R}^{E(\Gamma)} \rightarrow \mathcal{T}_H(F)$.

The basic idea is to associate to each edge of Γ an appropriate cross ratio. To set this up, consider the topological surface $F^+ \subseteq F$ obtained by adjoining a punctured disk to each true boundary component of F . The fatgraph $\iota(\Gamma) \subseteq F \subseteq F^+$ is thus also a spine of F^+ , and its Poincaré dual in F^+ is an ideal triangulation Δ of F^+ (i.e., a decomposition into triangles with vertices among the punctures). In the universal cover of F^+ , each arc of Δ thus separates two complementary triangles which combine to give a topological quadrilateral, and the basic idea is to associate to each edge the cross-ratio of this quadrilateral.

To make this precise and describe the homeomorphism in the theorem, let $\alpha = 1, \dots, E = E(\Gamma)$ index the edges of Γ , and let (Z_α) denote a point of \mathbb{R}^E . We associate the Möbius transformation

$$X_{Z_\alpha} = \begin{pmatrix} 0 & -e^{Z_\alpha/2} \\ e^{-Z_\alpha/2} & 0 \end{pmatrix}. \quad (2.1)$$

to the edge α . To explicate this definition, consider an ideal quadrilateral in the hyperbolic plane triangulated by a diagonal into two ideal triangles T_1, T_2 . We may conjugate in $\mathrm{PSL}_2(\mathbb{R})$ to arrange that the vertices of T_1 are $0, -1, \infty$ and the vertices of T_2 are $0, -1, t$, where $0 < t < \infty$, and an appropriate cross ratio of the original quadrilateral is t . Setting $Z_\alpha = \log t$ in the formula above, X_{Z_α} is the Möbius transformation interchanging $0, \infty$ and sending -1 to t , i.e., sending T_1 to T_2 . Notice that $X_{Z_\alpha}^2$ is the identity in $\mathrm{PSL}_2(\mathbb{R})$, so X_{Z_α} also sends T_2 to T_1 .

We also introduce the “right” and “left” turn matrices

$$R = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}, \quad L = R^2 = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}, \quad (2.2)$$

and define the corresponding operators R_Z and L_Z ,

$$R_Z \equiv RX_Z = \begin{pmatrix} e^{-Z/2} & -e^{Z/2} \\ 0 & e^{Z/2} \end{pmatrix}, \quad (2.3)$$

$$L_Z \equiv LX_Z = \begin{pmatrix} e^{-Z/2} & 0 \\ -e^{-Z/2} & e^{Z/2} \end{pmatrix}. \quad (2.4)$$

Consider a closed oriented edge-path P in Γ , where we assume that P never consecutively traverses an oriented edge followed by its reverse, i.e., there is no “turning back”. Choosing also an initial base point on P , we may imagine the corresponding curve serially traversing the oriented edges of Γ with coordinates Z_1, \dots, Z_n turning left or right from Z_i to Z_{i+1} , for $i = 1, \dots, n$ (with the indices mod n so that

$Z_{n+1} = Z_1$). Assign to P the corresponding composition

$$P_{Z_1 \dots Z_n} = L_{Z_n} L_{Z_{n-1}} R_{Z_{n-2}} \dots R_{Z_2} L_{Z_1}, \tag{2.5}$$

where the matrices L_{Z_i} or R_{Z_i} are inserted depending on which turn – left or right – the path takes at the corresponding stage.

Proposition 2.2. *There is a one-to-one correspondence between the set of conjugacy classes of elements of $\pi_1(F)$ and free homotopy classes of closed oriented geodesics in F . For any spine of F , each free homotopy class is uniquely represented by a cyclically defined closed edge-path P with no turning back, and the length of γ is determined by*

$$G_\gamma \equiv 2 \cosh(l_\gamma/2) = |\operatorname{tr} P_{Z_1 \dots Z_n}|. \tag{2.6}$$

By construction if P corresponds to a boundary component γ of $F(\Gamma)$, then the associated matrix has the form $R_{Z_1} R_{Z_2} \dots R_{Z_n}$, or $L_{Z_1} L_{Z_2} \dots L_{Z_n}$ depending on the orientation. In this case, because all of the matrices R_x (L_x) are upper (lower) triangular, formula (2.6) gives

$$l_\gamma = \left| \sum_{i=1}^n Z_i \right|, \tag{2.7}$$

where the sign of this sum gives the orientation of the boundary component, and $l_\gamma = 0$ corresponds to a puncture. This proves the assertions about boundary lengths.

The Z -coordinates (i.e., log cross ratios) are called (*Thurston*) *shear coordinate* [36], [2] and can alternatively be defined by dropping perpendiculars from each of the two opposite vertices to the diagonal α of a quadrilateral, and measuring the signed hyperbolic distance Z_α along α between these two projections.

Assume that there is an enumeration of the edges of Γ and that edge α has distinct endpoints. Given a spine Γ of F , we may produce another spine Γ_α of F by contracting and expanding edge α of Γ , the edge labelled Z in Figure 1, to produce Γ_α as in the figure; the fattening and embedding of Γ_α in F is determined from that of Γ in the natural way. Furthermore, an enumeration of the edges of Γ induces an enumeration of the edges of Γ_α in the natural way, where the vertical edge labelled Z in Figure 1 corresponds to the horizontal edge labelled $-Z$. We say that Γ_α arises from Γ by a *Whitehead move* along edge α . We shall also write $\Gamma_{\alpha\beta} = (\Gamma_\alpha)_\beta$, for any two indices α, β of edges, to denote the result of first performing a move along α and then along β ; in particular, $\Gamma_{\alpha\alpha} = \Gamma$ for any index α .

Proposition 2.3 ([6]). *Setting $\phi(Z) = \log(e^Z + 1)$ and adopting the notation of Figure 1 for shear coordinates of nearby edges, the effect of a Whitehead move is illustrated in the figure, viz.,*

$$W_Z: (A, B, C, D, Z) \rightarrow (A + \phi(Z), B - \phi(-Z), C + \phi(Z), D - \phi(-Z), -Z). \tag{2.8}$$

In the various cases where the edges are not distinct and identifying an edge with its shear coordinate in the obvious notation we have: if $A = C$, then $A' = A + 2\phi(Z)$; if $B = D$, then $B' = B - 2\phi(-Z)$; if $A = B$ (or $C = D$), then $A' = A + Z$ (or $C' = C + Z$); if $A = D$ (or $B = C$), then $A' = A + Z$ (or $B' = B + Z$).

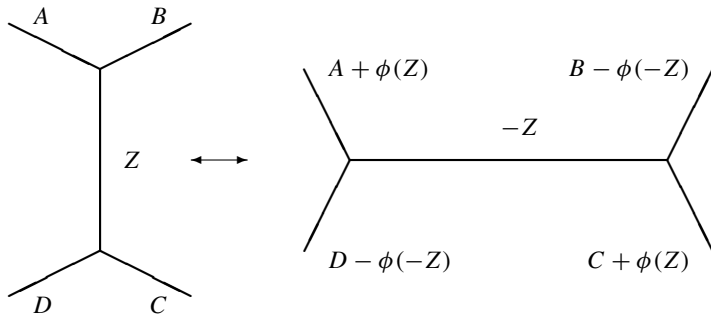


Figure 1. Whitehead move on shear coordinates.

Sketch of proof. Assume that e is the diagonal of a quadrilateral with consecutive sides a, b, c, d , where e separates a, b from c, d . Identifying an edge with its lambda length, the shear coordinate is given by $Z = \log \frac{bd}{ac}$, i.e., $\frac{bd}{ac}$ is the required cross-ratio [25]. Furthermore, if f is the lambda length of the other diagonal, then the lambda lengths satisfy Ptolemy’s relation $ef = ac + bd$ [25], and the transformation laws for shear coordinates in the proposition are readily derived from this either in the surface for (2.8) or in the universal cover of the surface in the various cases. \square

Insofar as hyperbolic lengths of geodesics are well-defined invariants of homotopy classes of curves in F , these lengths must be invariant under Whitehead moves, so we have the following

Lemma 2.4. Transformation (2.8) preserves the traces of products over paths (2.6).

2.1.2 Weil–Petersson form. $\mathcal{T}_H(F)$ supports its canonical Weil–Petersson Poisson structure, which has a very simple form in shear coordinates.

Theorem 2.5 ([10]). In the coordinates (Z_α) on any fixed spine, the Weil–Petersson bracket B_{WP} is given by

$$B_{WP} = \sum_v \sum_{i=1}^3 \frac{\partial}{\partial Z_{v_i}} \wedge \frac{\partial}{\partial Z_{v_{i+1}}}, \tag{2.9}$$

where the sum is taken over all vertices v and $v_i, i = 1, 2, 3 \text{ mod } 3$, are the labels of the cyclically ordered half-edges incident on this vertex.

The proof [10] relies on the independence of this form under Whitehead moves as in [28]. Indeed, the equivalent expression for the Weil–Petersson Kähler two-form in the punctured case was first given in lambda length coordinates in [27] starting from Wolpert’s formula [41], and the formula in shear coordinates follows from direct calculation using the expression $Z = \log \frac{bd}{ac}$. (There is more to this geometrically, however, and one must show [3] that the same expression is the Weil–Petersson form for surfaces with boundary.)

The set of Casimir functions is described by the following proposition, whose proof is given in Appendix B.

Proposition 2.6. *The center of the Poisson algebra (2.9) is generated by elements of the form $\sum Z_\alpha$, where the sum is over all edges of Γ in a boundary component of $F(\Gamma)$ and the sum is taken with multiplicity.*

2.1.3 Mapping class group description using graphs. Recall that the mapping class group $\text{MC}(F)$ of an open surface F is the group of homotopy classes of orientation-preserving homeomorphisms of F . No special constraints are imposed by the circle boundary components, i.e., a homeomorphism must fix each boundary component only setwise, and the homotopies must likewise fix each boundary component only setwise. Thus, if F has b boundary component circles, p punctures, and genus g , then $\text{MC}(F) \approx \text{MC}(F_g^{b+p})$, so we generally write $\text{MC}_g^s = \text{MC}(F)$ for any surface of genus g with s boundary components. In this section, we establish the combinatorial presentation of MC_g^s associated with the cell decomposition of decorated Teichmüller space.

Recall that a cell in the decomposition of $\tilde{\mathcal{T}}_g^s$ is described by the homotopy classes of an embedding $\iota: \Gamma \rightarrow F$ of a fatgraph Γ as a spine of F . MC_g^s acts on the set of homotopy classes of such embeddings by post-composition, and the cell decomposition of decorated Teichmüller space $\tilde{\mathcal{T}}_g^s$ descends to an orbifold cell decomposition of $\tilde{\mathcal{M}}_g^s = \tilde{\mathcal{T}}_g^s / \text{MC}(F)$.

The modular groupoid $\text{MG}_g^s = \text{MG}(F)$ is the fundamental path groupoid of $\tilde{\mathcal{M}}_g^s$, and MC_g^s arises as the subgroup of paths based at any point. Specifically, consider the dual graph $\mathcal{G}_g^s = \mathcal{G}(F)$ of the codimension-two skeleton of this decomposition of $\tilde{\mathcal{M}}_g^s$ (where there is one vertex for each top-dimensional cell, edges correspond to Whitehead moves, and two-dimensional cells correspond to pairs of homotopic paths in the one-skeleton which are homotopic to real endpoints in $\tilde{\mathcal{M}}_g^s$.) The fundamental path groupoid of \mathcal{G}_g^s is the modular groupoid, and in particular, MC_g^s is the stabilizer in MG_g^s of any vertex of \mathcal{G}_g^s .

We may think of a Whitehead move along edge α of fatgraph Γ producing another fatgraph Γ_α as an ordered pair (Γ, Γ_α) , i.e., an oriented edge of \mathcal{G}_g^s . Letting $[\Gamma_1, \Gamma_2]$ denote the MC_g^s -orbit of a pair (Γ_1, Γ_2) by the diagonal action of the mapping class group, the natural composition descends to a well-defined product

$$[\Gamma_1, \Gamma_3] = [\Gamma_1, \Gamma_2][\Gamma_2, \Gamma_3].$$

Theorem 2.7 ([28], [27]). *The modular groupoid MG_g^s is generated by Whitehead moves and relabelings by fatgraph symmetries. A complete list of relations in MG_g^s is given by relabelings under fatgraph symmetries together with the two following relations.*

If α and β are two edges with no common endpoints, then

$$[\Gamma_{\alpha\beta}, \Gamma_\alpha][\Gamma_\alpha, \Gamma] = [\Gamma_{\alpha\beta}, \Gamma_\beta][\Gamma_\beta, \Gamma] \quad (\text{Commutativity}).$$

If α and β share exactly one common endpoint, then (see Figure 2 drawn for the dual graph)

$$[\Gamma, \Gamma_\alpha][\Gamma_\alpha, \Gamma_{\beta\alpha}][\Gamma_{\beta\alpha}, \Gamma_{\alpha\beta}][\Gamma_{\alpha\beta}, \Gamma_\beta][\Gamma_\beta, \Gamma] = 1 \quad (\text{Pentagon}).$$

Furthermore, the expression in Theorem 2.5 for the Weil–Petersson form is invariant under Whitehead moves.

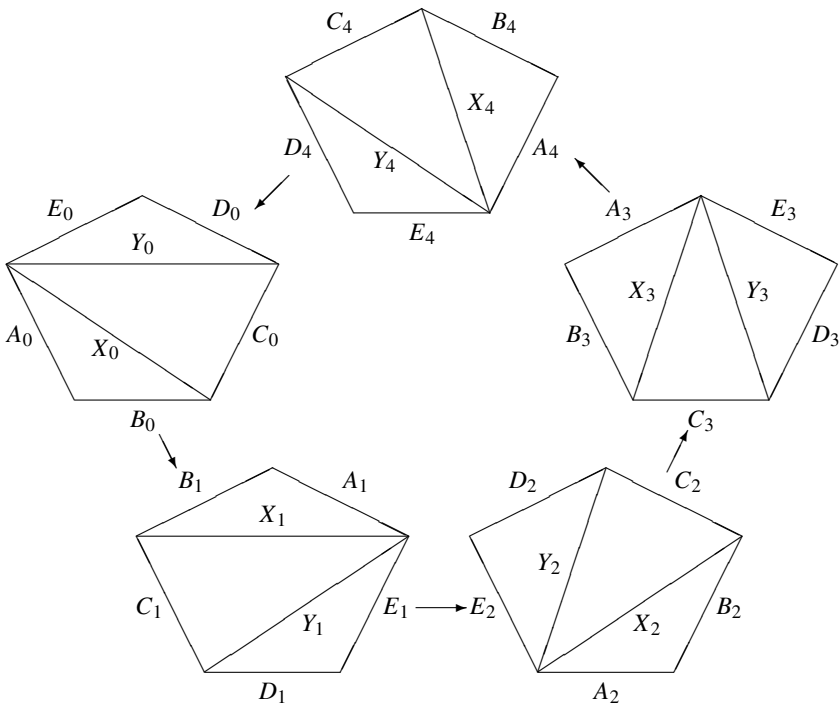


Figure 2. Pentagon identity.

Proof. The first parts are immediate consequences of the cell decomposition. Specifically, by connectivity of $\tilde{\mathcal{T}}_g^s$, any two points can be joined by a smooth path, which we

may put into general position with respect to the codimension-one faces; this proves the first part. For the second part, a homotopy between edge-paths in $\tilde{\mathcal{M}}_g^s$ can likewise be put into general position with respect to the codimension-two faces; there are two possibilities for a pair of edges depending upon whether their vertices are disjoint or not, corresponding respectively to the commutativity and pentagon relations, proving the second part. The invariance of the expression for the Weil–Petersson form under Whitehead moves is a direct calculation in lambda lengths [29] or shear coordinates using (2.8). \square

2.2 Poisson algebras of geodesic functions

The algebra generated (with multiplication and with the Weil–Petersson Poisson bracket) by the functions $\{G_\gamma\}$ (2.6) was first studied by W. Goldman [14].

2.2.1 Multicurves. In the sequel, disjointly embedded families of geodesics will play a special role as they constitute a basis for the algebra of observables in both the classical case considered here and the quantum case discussed in Section 3. The homotopy class of such a family is called a *multiple curve*. A *multicurve* is multiset based on the set of curves in a multiple curve.

Definition 2.1. Consider the homotopy class of a finite collection $C = \{\gamma_1, \dots, \gamma_n\}$ of disjointly embedded (unoriented) simple closed curves γ_i in a topological surface F , where C need *not* be a mutiple curve. A *generalized multicurve* (GMC) \hat{C} in F is a multiset based on C ; one thus imagines $s_i \geq 1$ parallel copies of components of C , or in other words, positive integral weights s_i on each component of C , where s_i is the multiplicity of γ_i in \hat{C} . Further, given a hyperbolic structure on F , we associate to \hat{C} the product $G_{\hat{C}} = G_{\gamma_1}^{s_1} \dots G_{\gamma_n}^{s_n}$ of geodesic operators (2.6) of all geodesics constituting a GMC; these operators Poisson commute in the classical case since the components of C are disjoint. In particular, a GMC containing a contractible component (of length zero) is twice the GMC with this curve removed.

An edge-path on a spine $\Gamma \subseteq F$ or its corresponding geodesic γ in F is said to be *graph simple path* with respect to Γ if it does not pass more than once through any edge of Γ . Obviously, the set of graph simple geodesics depends upon Γ and is *not* invariant under Whitehead moves. Nevertheless, this notion will be useful in what follows.

2.2.2 Classical skein relation. The trace relation $\text{tr}(AB) + \text{tr}(AB^{-1}) - \text{tr} A \cdot \text{tr} B = 0$ for arbitrary 2×2 matrices A and B with unit determinant allows one to “disentangle” any product of geodesic functions, i.e., express it uniquely as a finite linear combination of GMCs. Introducing the additional factor $\#G$ to be the total number of components

in a GMC, we can then uniformly present the classical skein relation as

$$(-1)^{\#G} \begin{array}{c} \diagup \quad \diagdown \\ \diagdown \quad \diagup \end{array} + (-1)^{\#G} \begin{array}{c} \text{)} \quad (\\ (\quad) \end{array} + (-1)^{\#G} \begin{array}{c} \text{) } \\ \text{(} \end{array} = 0. \tag{2.10}$$

2.2.3 Poisson brackets for geodesic functions. Turning attention now to the Poisson structure, two geodesic functions Poisson commute if the underlying geodesics are disjointly embedded. By the Leibnitz rule for the Poisson bracket, it suffices to consider only “simple” intersections of pairs of geodesics with respective geodesic functions G_1 and G_2 of the form

$$G_1 = \text{tr}^1 \dots X_C^1 R^1 X_Z^1 L^1 X_A^1 \dots, \tag{2.11}$$

$$G_2 = \text{tr}^2 \dots X_B^2 L^2 X_Z^2 R^2 X_D^2 \dots, \tag{2.12}$$

where the superscripts 1 and 2 pertain to operators and traces in two different matrix spaces.

The bracket between X_C^1 and X_B^2 possesses a simple r -matrix structure

$$\{X_C^1, X_B^2\} = \frac{1}{4}(-1)^{i+j} (e_{ii}^1 \otimes e_{jj}^2) X_C^1 \otimes X_B^2, \tag{2.13}$$

where the “elementary” matrix e_{ij} has entry unity in its i th row and j th column and zero otherwise. Direct calculations then give

$$\{G_1, G_2\} = \frac{1}{2}(G_H - G_I), \tag{2.14}$$

where G_I corresponds to the geodesic that is obtained by erasing the edge Z and joining together the edges “ A ” and “ D ” as well as “ B ” and “ C ” in a natural way as illustrated in the middle diagram in (2.10); G_H corresponds to the geodesic that passes over the edge Z twice, so it has the form $\text{tr} \dots X_C R_Z R_D \dots \dots X_B L_Z L_A \dots$ as illustrated in the rightmost diagram in (2.10). These relations were first obtained in [14] in the continuous parametrization (the classical Turaev–Viro algebra).

Torus example. For the torus, $\mathcal{T}_H(F_1^1)$ has three generators X, Y, Z , where

$$\{X, Y\} = \{Y, Z\} = \{Z, X\} = 2,$$

corresponding to the combinatorially unique cubic spine and the Casimir element is $X + Y + Z$. The geodesic functions for the three graph simple geodesics are

$$\begin{aligned} G_X &= \text{tr} L X_Y R X_Z = e^{-Y/2-Z/2} + e^{-Y/2+Z/2} + e^{Y/2+Z/2}, \\ G_Y &= \text{tr} R X_X L X_Z = e^{-Z/2-X/2} + e^{-Z/2+X/2} + e^{Z/2+X/2}, \\ G_Z &= \text{tr} R X_Y L X_X = e^{-X/2-Y/2} + e^{-X/2+Y/2} + e^{X/2+Y/2}. \end{aligned} \tag{2.15}$$

Introducing the geodesic function

$$\tilde{G}_Z = \text{tr } L_Y R_Z R_X L_Z = e^{-X/2-Y/2-Z} + e^{X/2-Y/2}(e^{-Z} + e^Z + 2) + e^{X/2+Y/2+Z},$$

obtained from G_Z by a Whitehead move, we find that $\{G_X, G_Y\} = \tilde{G}_Z/2 - G_Z/2$, and because relation (2.10) implies that $G_X G_Y = G_Z + \tilde{G}_Z$, we have

$$\{G_X, G_Y\} = \frac{1}{2}G_X G_Y - G_Z, \quad (2.16)$$

plus the cyclic permutations in X, Y, Z , i.e., the classical Poisson algebra closes in the algebraic span of the geodesic functions $\{G_X, G_Y, G_Z\}$.

2.2.4 Poisson geodesic algebras for higher genera. In order to generalize the torus example, we must find a graph on which graph simple geodesics constitute a convenient algebraic basis. Such a graph is illustrated in Figure 3, where m edges pairwise connect two horizontal line segments. Graph simple closed geodesics in this picture are those and only those that pass through exactly two different “vertical” edges, and they are therefore enumerated by ordered pairs of edges; we denote the corresponding geodesic functions \mathcal{G}_{ij} where $i < j$. The Poisson algebra for the functions \mathcal{G}_{ij} is described by

$$\{\mathcal{G}_{ij}, \mathcal{G}_{kl}\} = \begin{cases} 0, & j < k, \\ 0, & k < i, j < l, \\ \mathcal{G}_{ik}\mathcal{G}_{jl} - \mathcal{G}_{kj}\mathcal{G}_{il}, & i < k < j < l, \\ \frac{1}{2}\mathcal{G}_{ij}\mathcal{G}_{jl} - \mathcal{G}_{il}, & j = k, \\ \mathcal{G}_{il} - \frac{1}{2}\mathcal{G}_{ij}\mathcal{G}_{il}, & i = k, j < l \\ \mathcal{G}_{ik} - \frac{1}{2}\mathcal{G}_{ij}\mathcal{G}_{kj}, & j = l, i < k. \end{cases} \quad (2.17)$$

The graph in Figure 3 has genus $\frac{m}{2} - 1$ and two faces (holes) if m is even and genus $(m - 1)/2$ and one face (hole) if m is odd. Such geodesic bases for m even were considered in [21]. The Poisson algebras of geodesics obtained there *coincide exactly* with (2.17). These are the so-called $\mathfrak{so}_q(m)$ algebras whose representations were constructed in [15].

In the mathematical literature, this algebra has also appeared as the Poisson algebra of the monodromy data (Stokes matrices) of certain matrix differential equations [38] and on the symplectic groupoid of upper-triangular matrices \mathcal{G} [4]. These matrices have entries given by unity on the main diagonal (i.e., we set $\mathcal{G}_{ii} \equiv 1$) and the entries \mathcal{G}_{ij} above it. For $m \times m$ -matrices, there are $\left[\frac{m}{2}\right]$ central elements of this algebra generated by the polynomial invariants $f_{\mathcal{G}}(\lambda) \equiv \det(\mathcal{G} + \lambda\mathcal{G}^T) = \sum f_i(\mathcal{G})\lambda^i$. The total Poisson dimension d of algebra (2.17) is $\frac{m(m-1)}{2} - \left[\frac{m}{2}\right]$, and for $m = 3, 4, 5, 6, \dots$ we have $d = 2, 4, 8, 12, \dots$. The dimensions of the corresponding Teichmüller spaces are $D = 2, 4, 8, 10, \dots$, so we see that the Teichmüller spaces are embedded as the Poisson leaves in the algebra (2.17).

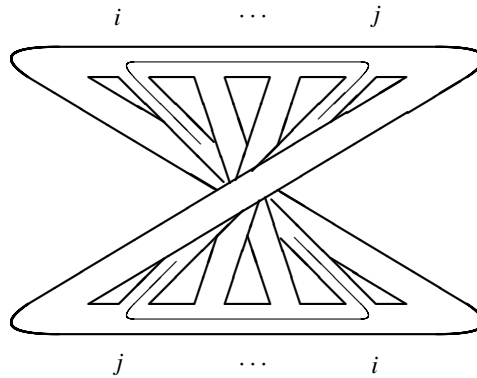


Figure 3. The special spine for higher genera.

3 Quantization

A quantization of a Poisson manifold, which is equivariant under the action of a discrete group \mathcal{D} , is a family of $*$ -algebras \mathcal{A}^{\hbar} depending on a positive real parameter \hbar with \mathcal{D} acting by outer automorphisms and having the following properties:

1. (Flatness) All algebras are isomorphic (noncanonically) as linear spaces.
2. (Correspondence) For $\hbar = 0$, the algebra is isomorphic as a \mathcal{D} -module to the $*$ -algebra of complex-valued functions on the Poisson manifold.
3. (Classical limit) The Poisson bracket on \mathcal{A}^0 given by $\{a_1, a_2\} = \lim_{\hbar \rightarrow 0} \frac{[a_1, a_2]}{\hbar}$ coincides with the Poisson bracket given by the Poisson structure of the manifold.

3.1 Quantizing Teichmüller spaces

Here we construct a quantization $\mathcal{T}^{\hbar}(F)$ of the Teichmüller space $\mathcal{T}_H(F)$ that is equivariant with respect to the action of the mapping class group $\mathcal{D} = \text{MC}(F)$.

Fix a cubic fatgraph Γ as spine of F , and let $\mathcal{T}^{\hbar} = \mathcal{T}^{\hbar}(\Gamma)$ be the algebra generated by Z_{α}^{\hbar} , one generator for each unoriented edge α of Γ , with relations

$$[Z_{\alpha}^{\hbar}, Z_{\beta}^{\hbar}] = 2\pi i \hbar \{z_{\alpha}, z_{\beta}\} \quad (3.1)$$

(cf. (2.9)) and the $*$ -structure

$$(Z_{\alpha}^{\hbar})^* = Z_{\alpha}^{\hbar}, \quad (3.2)$$

where z_{α} and $\{\cdot, \cdot\}$ denotes the respective coordinate functions and the Poisson bracket on the classical Teichmüller space. Because of (2.9), the righthand side of (3.1) is a constant taking only five values $0, \pm 2\pi i \hbar$, and $\pm 4\pi i \hbar$ depending upon the coincidences of endpoints of edges labelled α and β .

Lemma 3.1. *The center \mathcal{Z}^{\hbar} of the algebra \mathcal{T}^{\hbar} is generated by the sums $\sum_{\alpha \in I} Z_{\alpha}^{\hbar}$ over all edges $\alpha \in I$ surrounding a given boundary component, and the Poisson structure is non-degenerate on the quotient $\mathcal{T}^{\hbar}/\mathcal{Z}^{\hbar}$.*

A standard Darboux-type theorem for non-degenerate Poisson structures then gives the following result.

Corollary 3.2. *There is a basis for $\mathcal{T}^{\hbar}/\mathcal{Z}^{\hbar}$ given by operators p_i, q_i , for $i = 1, \dots, 6g - 6 + 2s$, satisfying the standard commutation relations $[p_i, q_j] = 2\pi i \hbar \delta_{ij}$, $i, j = 1, \dots, 3g - 3 + s$.*

Not only is the proof of Lemma 3.1 given in Appendix B, but also an algorithm for diagonalizing this Poisson structure is described there.

Now, define the Hilbert space \mathcal{H} to be the set of all L^2 functions in the q -variables and let each q -variable act by multiplication and each corresponding p variable act by differentiation, $p_i = 2\pi i \hbar \frac{\partial}{\partial p_i}$. For different choices of diagonalization of non-degenerate Poisson structures, these Hilbert spaces are canonically isomorphic.

Torus example. In the case of the bordered torus, we have three generators, X^{\hbar} , Y^{\hbar} , and Z^{\hbar} , the commutation relations (3.1) have the form $[X^{\hbar}, Y^{\hbar}] = [Y^{\hbar}, Z^{\hbar}] = [Z^{\hbar}, X^{\hbar}] = 4\pi i \hbar$, and the single central element is $X^{\hbar} + Y^{\hbar} + Z^{\hbar}$. In the Darboux-type representation, we can identify, e.g., $(X^{\hbar} + Y^{\hbar})/2$ with q and $(-X^{\hbar} + Y^{\hbar})/2$ with $(-2\pi i \hbar) \partial / \partial q$.

On the level of the modular groupoid as a category where all morphisms are invertible and any two objects are related by a morphism, we have constructed one $*$ -algebra per object. In order to describe the \mathcal{D} -equivariance we must associate a homomorphism of the corresponding $*$ -algebras to any morphism in the modular groupoid. For this, we associate a morphism of algebras to any Whitehead move and must verify that the relations in Theorem 2.7 are satisfied.

We now define the *quantum Whitehead move* or *flip* along an edge of Γ by Equation (2.8) using the (quantum) function

$$\phi(z) \equiv \phi^{\hbar}(z) = -\frac{\pi \hbar}{2} \int_{\Omega} \frac{e^{-ipz}}{\sinh(\pi p) \sinh(\pi \hbar p)} dp, \tag{3.3}$$

where the contour Ω goes along the real axis bypassing the origin from above. The function (3.3) is Faddeev’s generalization [8] of the quantum dilogarithm.

Proposition 3.3. *For each unbounded self-adjoint operator Z on \mathcal{H} , $\phi(Z)$ is a well-defined unbounded self-adjoint operator on \mathcal{H} .*

Proof. The function $\phi^{\hbar}(Z)$ satisfies the relations (see [6])

$$\begin{aligned} \phi^{\hbar}(Z) - \phi^{\hbar}(-Z) &= Z, \\ \phi^{\hbar}(Z + i\pi \hbar) - \phi^{\hbar}(Z - i\pi \hbar) &= \frac{2\pi i \hbar}{1 + e^{-Z}}, \end{aligned}$$

$$\phi^{\hbar}(Z + i\pi) - \phi^{\hbar}(Z - i\pi) = \frac{2\pi i}{1 + e^{-Z/\hbar}}$$

and is meromorphic in the complex plane with the poles at the points $\{\pi i(m + n\hbar), m, n \in \mathbb{Z}_+\}$ and $\{-\pi i(m + n\hbar), m, n \in \mathbb{Z}_+\}$.

Hence the function $\phi^{\hbar}(Z)$ is holomorphic in the strip $|\text{Im } Z| < \pi \min(1, \text{Re } \hbar) - \varepsilon$ for any $\varepsilon > 0$, so we need only its asymptotic behavior as $Z \in \mathbb{R}$ and $|Z| \rightarrow \infty$, for which we have (see, e.g., [18])

$$\phi^{\hbar}(Z)|_{|Z| \rightarrow \infty} = (Z + |Z|)/2 + O(1/|Z|). \tag{3.4}$$

Therefore, the function $\phi^{\hbar}(Z)$ increases as Z goes to plus infinity and represents an operator in \mathcal{H} by the functional calculus [31]. □

Theorem 3.4. *The family of algebras $\mathcal{T}^{\hbar} = \mathcal{T}^{\hbar}(\Gamma)$ is a quantization of $\mathcal{T}_H(F)$ for any cubic fatgraph spine Γ of F , that is:*

1. *in the limit $\hbar \mapsto 0$, morphism (2.8) using (3.3) coincides with the classical morphism (2.8) where $\phi(Z) = \log(1 + e^Z)$;*
2. *morphism (2.8) using (3.3) is indeed a morphism of $*$ -algebras;*
3. *a flip W_Z satisfies $W_Z^2 = I$, cf. (2.8), and flips satisfy the commutativity relation;*
4. *flips satisfy the pentagon relation.*¹

Furthermore,

5. *the morphisms $\mathcal{T}^{\hbar}(\Gamma) \rightarrow \mathcal{T}^{1/\hbar}(\Gamma)$ given by $Z_{\alpha}^{\hbar} \mapsto Z_{\alpha}^{1/\hbar}$ commute with morphisms (2.8).*

Sketch of proof. Property 1 follows since $\lim_{\hbar \rightarrow 0} \phi^{\hbar}(z) = \log(e^z + 1)$, and Property 3 is obvious.

In order to prove Property 2, we must first verify that $[A + \phi^{\hbar}(Z), B - \phi^{\hbar}(-Z)] = 0$ and $[A + \phi^{\hbar}(Z), D - \phi^{\hbar}(-Z)] = -2\pi i \hbar$ (since the other relations are obviously satisfied), which follows from the identity $\phi^{\hbar}(z) - \phi^{\hbar}(-z) = z$.

For Property 5, we must verify that the morphism $\mathcal{T}^{\hbar}(\Gamma) \rightarrow \mathcal{T}^{1/\hbar}(\Gamma)$ commutes with a flip, that is, $(A + \phi^{\hbar}(Z))/\hbar = A/\hbar + \phi^{\hbar}(Z/\hbar)$, $(B - \phi^{1/\hbar}(-Z))/\hbar = B/\hbar - \phi^{\hbar}(-Z/\hbar)$, etc., which follows from $\phi^{\hbar}(z)/\hbar = \phi^{1/\hbar}(z/\hbar)$.

Turning finally to the most nontrivial Property 4, we may reformulate it as follows. There are seven generators involved in the sequence of flips depicted in Figure 2 for the dual cell decomposition, which are denoted A, B, C, D, E, X, Y as in the figure. As a result of a flip, the piece of graph shown in Figure 1 just gets cyclically rotated. Denote by $A_i, B_i, C_i, D_i, E_i, X_i,$ and Y_i the algebra elements associated to the edges of this piece of graph after i flips are performed. From (2.8), (3.3), these elements

¹This result was independently obtained by R. M. Kashaev [16].

evolve as follows:

$$\begin{pmatrix} X_{i+1} \\ Y_{i+1} \\ A_{i+1} \\ B_{i+1} \\ C_{i+1} \\ D_{i+1} \\ E_{i+1} \end{pmatrix} \longrightarrow \begin{pmatrix} Y_i - \phi^{\hbar}(-X_i) \\ -X_i \\ D_i \\ E_i \\ A_i + \phi^{\hbar}(X_i) \\ B_i - \phi^{\hbar}(-X_i) \\ C_i + \phi^{\hbar}(X_i) \end{pmatrix}. \tag{3.5}$$

We must prove that this operator is periodic with period five.

Assume for a moment that this five-periodicity of X_i has been established. Then five-periodicity of Y_i and other variables follow from simple calculations.

It suffices to prove this five-periodicity for the X_i since the five-periodicity of the other operators such as Y_i then follows from elementary calculations. As to the five-periodicity of X_i , let us “take logarithms” and introduce four new algebra elements

$$U_i = e^{X_i}; \quad V_i = e^{Y_i}; \quad \tilde{U}_i = e^{X_i/\hbar}; \quad \tilde{V}_i = e^{Y_i/\hbar},$$

which satisfy the following commutation relations

$$\begin{aligned} U_i V_i &= q^{-2} V_i U_i, & \tilde{U}_i \tilde{V}_i &= \tilde{q}^{-2} \tilde{V}_i \tilde{U}_i, \\ U_i \tilde{V}_i &= \tilde{V}_i U_i, & V_i \tilde{U}_i &= \tilde{U}_i V_i, \end{aligned} \tag{3.6}$$

where

$$q = e^{-\pi i \hbar}, \quad \tilde{q} = e^{-\pi i / \hbar}.$$

Under the flip, these variables are transformed in an especially simple way,

$$U_{i-1} = V_i^{-1}, \tag{3.7}$$

$$V_{i-1} = U_i(1 + q V_i), \tag{3.8}$$

$$\tilde{U}_{i-1} = \tilde{V}_i^{-1}, \tag{3.9}$$

$$\tilde{V}_{i-1} = V_i(1 + \tilde{q} \tilde{V}_i). \tag{3.10}$$

As the first step of the proof, we consider the inverse transformation laws for X_i and Y_i :

$$X_{i-1} = -Y_i; \quad Y_{i-1} = X_i + \phi^{\hbar}(Y_i).$$

Equations (3.7) and (3.9) are obvious. Using the standard formula

$$e^{A+F(B)} = e^{\frac{1}{[A,B]} \int_B^{B+[A,B]} F(z) dz} e^A,$$

we obtain

$$\begin{aligned}
 V_{i-1} &= e^{Y_{i-1}} = e^{X_i + \phi^{\hbar}(Y_i)} = e^{X_i} e^{\int_Y^{Y+2\pi i\hbar} \phi^{\hbar}(z) dz} \\
 &= U_i \exp\left(-\frac{\pi\hbar}{2} \int_{\Omega} \frac{e^{-ipz}(e^{-2p\pi\hbar} - 1)}{(-ip) \sinh(\pi p) \sinh(\pi\hbar p)} dp\right) \\
 &= U_i \exp\left(\frac{\pi\hbar}{i} \int_{\Omega} \frac{e^{-ip(z-\pi i\hbar)}}{p \sinh(\pi p)} dp\right) \\
 &= U_i(1 + q^{-2}V_i).
 \end{aligned}$$

The proof of (3.10) is analogous.

Now, in order to finally prove that X_i is five-periodic, it suffices to verify that both U_i and \tilde{U}_i are five-periodic. Indeed, if only the operator U_i is five-periodic, it does not suffice because the logarithm of an operator is ambiguously defined. However, if we have two families of operators U and \tilde{U} , which depend continuously on \hbar , then, assuming the existence of an operator X (depending continuously on \hbar) such that $U = e^X$ and $\tilde{U} = e^{X/\hbar}$, then this operator is evidently unique. (It can be found as $\lim_{(m+n/\hbar)\rightarrow 0} (U^m \tilde{U}^n)/(m + n/\hbar)$ for any irrational value of \hbar .) The five-periodicity of sequence (3.8) (and (3.10)) is a direct calculation using (3.6).

The only subtlety remaining is the possibility that some of the edges A, B, C, D, E coincide. (Note, however, that X and Y must have exactly one common vertex.) If, say, edges A and C coincide, then the value of the commutator $[A, X]$ is doubled by definition, and we can then fictitiously *split* the edge $A = C$ into two half-edges with the matrices $X_{A/2}$ assigned to each half-edge using the formula

$$X_A = X_{A/2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} X_{A/2}. \tag{3.11}$$

The quantities $A/2 = C/2$ have the same commutation relations with the rest of variables as well as the same transformation laws as the quantities A and C before, so the earlier formulas remain valid if we simply replace there A' by $A'/2$ and A by $A/2$. The net effect is that commutators with $A = C$ are doubled.

If edges A and B coincide, then we must correct formulas (3.4) using a splitting as above but demanding $A'/2 = A/2 + X/2$. Obviously, $[A, X] = 0$ in this case (in which formulas for the quantum ordering below must be also corrected, see Section 3.5).

The formulas in Proposition 2.3 can thus be realized for exponentiated quantities, although in the current quantum case there will be corrections. For instance, formulas (3.8) and (3.9) will be different; indeed, letting $A = C = X$, we calculate that $e^{X/2} \mapsto e^{X/2 + \phi^{\hbar}(Z)} =: e^{X/2}(1 + e^Z) := e^{X/2} + e^{X/2+Z} = e^{X/2}(1 + e^{-i\pi\hbar}e^Z)$ since the commutator of X and Z is doubled, where the normal ordering (the Weyl ordering) $\cdot \cdot \cdot$ is explained in the next section. The transformation for e^X itself becomes more complicated when $A = C$, namely, $e^X = (e^{X/2})^2 \mapsto e^X(1 + q^3e^Z)(1 + qe^Z)$. \square

Corollary 3.5. 1. Let K be an operator acting in the Hilbert space $L^2(\mathbb{R})$ and having the integral kernel

$$K(x, z) = F^{\hbar}(z)e^{-\frac{xz}{2\pi i\hbar}}, \tag{3.12}$$

where

$$F^{\hbar}(z) = \exp\left(-\frac{1}{4} \int_{\Omega} \frac{e^{-ipz}}{p \sinh(\pi p) \sinh(\pi \hbar p)} dp\right) \tag{3.13}$$

Then the operator K is unitary up to a multiplicative constant and satisfies the identity

$$K^5 = \text{const.} \tag{3.14}$$

2. Let $\hbar = m/n$ be a rational number and assume that both m and n are odd. Introduce a linear operator $L(u)$ acting in the space \mathbb{C}^n and depending on one positive real parameter u through its matrix

$$L(u)^i_j = F^{\hbar}(j, u)q^{-4ij}, \tag{3.15}$$

where

$$F^{\hbar}(j, u) = (1 + u)^{j/n} \prod_{k=0}^{j-1} (1 + q^{-4k+2}u^{1/n})^{-1}.$$

Then the following identity holds:

$$L(u)L(v + uv)L(v + vu^{-1} + u^{-1})L(u^{-1}v^{-1} + u^{-1})L(v^{-1}) = 1. \tag{3.16}$$

Using this construction, Kashaev [18] constructed the set of eigenfunctions of the quantum Dehn twist transformation. Namely, Kashaev’s dilogarithm function $e_b(z)$ is $1/F^{\hbar}(2z)$ from (3.13). We return to this discussion when considering sets of quantum Dehn twists for the torus in Section 5.

3.2 Geodesic length operators

We next embed the algebra of geodesics (2.6) into a suitable completion of the constructed algebra \mathcal{T}^{\hbar} . For any γ , the geodesic function G_{γ} (2.6) can be expressed in terms of shear coordinates on \mathcal{T}_H :

$$G_{\gamma} \equiv \text{tr } P_{Z_1 \dots Z_n} = \sum_{j \in J} \exp\left\{\frac{1}{2} \sum_{\alpha \in E(\Gamma)} m_j(\gamma, \alpha) z_{\alpha}\right\}, \tag{3.17}$$

where $m_j(\gamma, \alpha)$ are integers and J is a finite set of indices. In order to find the quantum analogues of these functions, we denote by $\hat{\mathcal{T}}^{\hbar}$ a completion of the algebra \mathcal{T}^{\hbar} containing $e^{xZ_{\alpha}}$ for any real x .

For any closed path γ on F , define the quantum geodesic operator $G_{\gamma}^{\hbar} \in \hat{\mathcal{T}}^{\hbar}$ to be

$$G_{\gamma}^{\hbar} \equiv \times \text{tr } P_{Z_1 \dots Z_n} \times \equiv \sum_{\substack{j \in J \\ \kappa \in \{j\}}} \exp\left\{\frac{1}{2} \sum_{\alpha \in E(\Gamma)} (m_j(\gamma, \alpha) Z_{\alpha}^{\hbar} + 2\pi i \hbar c_j^{\kappa}(\gamma, \alpha))\right\}, \tag{3.18}$$

where the *quantum ordering* $\times_{\hbar} \cdot \times_{\hbar}$ implies that we vary the classical expression (3.17) by introducing additional integer coefficients $c_j^K(\gamma, \alpha)$, which must be determined from the conditions below. Notice that the operators $\{G_{\gamma}^{\frac{1}{\hbar}}\}$ themselves can be considered as belonging to the algebra $\widehat{\mathcal{T}}^{\hbar}$ insofar as

$$G_{\gamma}^{\frac{1}{\hbar}} = \sum_{j \in J} \exp \left\{ \frac{1}{2\hbar} \sum_{\alpha \in E(\Gamma)} (m_j(\gamma, \alpha) Z_{\alpha}^{\hbar} + 2\pi i c_j^K(\gamma, \alpha)) \right\}. \quad (3.19)$$

In what follows for the notational simplicity, we shall sometimes omit the superscript \hbar from G^{\hbar} and write it merely G assuming that G is either an operator or a classical geodesic function depending on the context. We shall concentrate on the case of quantum functions of the G^{\hbar} -type; the consideration of the sector $G^{1/\hbar}$ is analogous and does not lead to new effects (at least at the present stage of understanding), so we omit it. We also call a quantum geodesic function merely a quantum geodesic (since implicitly quantum objects admit only functional, not geometrical, descriptions).

We wish to associate an operatorial *quantum multicurve* QMC to a multiset \widehat{C} of quantum geodesics corresponding to disjointly embedded families of nonnegative integrally weighted geodesics. One ansatz will be that operators corresponding to disjoint underlying geodesics must commute; this implies that the ordering in which the quantum geodesics enter the product QMC is immaterial, where the product is defined as for GMCs. We next formulate the defining properties of quantum geodesics.

1. If closed paths γ and γ' do not intersect, then the operators G_{γ}^{\hbar} and $G_{\gamma'}^{\hbar}$ commute.
2. *Naturality.* The mapping class group $\text{MC}(F)$ (2.8) acts naturally, i.e., for any $\{G_{\gamma}^{\hbar}\}$, $\delta \in \text{MC}(F)$ and closed path γ in a spine Γ of F , we have $\delta(G_{\gamma}^{\hbar}) = G_{\delta\gamma}^{\hbar}$.
3. *Geodesic algebra.* The product of two quantum geodesics is a linear combination of QMC's governed by the (quantum) skein relation [37].
4. *Orientation invariance.* Quantum traces of direct and inverse geodesic operators coincide.
5. *Exponents of geodesics.* A quantum geodesic $G_{n\gamma}$ corresponding to the n -fold concatenation of γ is expressed via G_{γ} exactly as in the classical case, namely,

$$G_{n\gamma} = 2T_n(G_{\gamma}/2), \quad (3.20)$$

where $T_n(x)$ are Chebyshev's polynomials.

6. *Duality.* For any γ and γ' , the operators G_{γ}^{\hbar} and $G_{\gamma'}^{\frac{1}{\hbar}}$ commute.

We shall let the standard normal ordering symbol $: e^{a_1} e^{a_2} \dots e^{a_n} :$ denote the *Weyl ordering* $e^{a_1 + \dots + a_n}$, i.e.,

$$: e^{a_1} e^{a_2} \dots e^{a_n} : = 1 + (a_1 + \dots + a_n) + \frac{1}{2!} (a_1 + \dots + a_n)(a_1 + \dots + a_n) + \dots \quad (3.21)$$

for any set of exponents with $a_i \neq -a_j$ for $i \neq j$. In particular, the Weyl ordering implies total symmetrization in the subscripts.

Proposition 3.6. *For any graph simple geodesic G with respect to any spine Γ , the coefficients $c_j^k(\gamma, \alpha)$ in (3.18) are identically zero.*

Proof. Consider term-by-term the trace of the matrix product for the quantum graph simple geodesic and expand it in Laurent monomials in $e^{Z_i/2}$. It is easy to see that each term $e^{Z_i/2}$ comes either in power $+1$, or -1 in the corresponding monomial and there are no equivalent monomials in the sum. This means that in order to have a Hermitian operator, we must apply the Weyl ordering with no additional q -factors (by the correspondence principle, each such factor must be again a Laurent monomial in q standing by the corresponding term, which breaks the self-adjointness unless all such monomials are unity). Since quantum Whitehead moves must preserve the property of being Hermitian, if a graph-simple geodesic transforms to another graph-simple geodesic, then a Weyl-ordered expression transforms to a Weyl-ordered expression, and only these expressions are self-adjoint. \square

Torus example. For the torus with one hole, there are three graph simple quantum geodesics for any spine, which are exactly (2.15) in the Weyl-ordered form. The quantum geodesics \tilde{G}_Z obtained from G_Z by the flip transformation is

$$\begin{aligned} \tilde{G}_Z = & e^{-X/2-Y/2-Z} + e^{X/2-Y/2-Z} + e^{X/2-Y/2} \cdot 2 \cos(\pi \hbar) \\ & + e^{X/2-Y/2+Z} + e^{X/2+Y/2+Z}. \end{aligned} \quad (3.22)$$

The product of two graph simple quantum geodesics is

$$G_X G_Y = e^{i\pi \hbar/2} \tilde{G}_Z + e^{-i\pi \hbar/2} G_Z. \quad (3.23)$$

Denoting $q \equiv e^{-i\pi \hbar}$, $[A, B]_q \equiv q^{1/2} AB - q^{-1/2} BA$, and $\xi = q - q^{-1}$, we obtain from (3.23)

$$[G_X, G_Y]_q = \xi G_Z, \quad [G_Y, G_Z]_q = \xi G_X, \quad [G_Z, G_X]_q = \xi G_Y. \quad (3.24)$$

This algebra is exactly the $\mathfrak{so}_q(3)$ quantum algebra studied in [15]. There is a unique central element, the quantum Markov relation

$$\mathcal{M} = G_X G_Y G_Z - q^{1/2} (G_X^2 + q^{-2} G_Y^2 + G_Z^2). \quad (3.25)$$

3.3 Algebra of quantum geodesics

Let G_1 and G_2 correspond to the respective graph simple geodesics with respect to the same spine having one nontrivial intersection. For G_1 and G_2 , formula (3.18) implies, by virtue of Proposition 3.6, the mere Weyl ordering.

After some algebra, we obtain (cf. (2.14))

$$G_1 G_2 = e^{-i\pi \hbar/2} G_Z + e^{i\pi \hbar/2} \tilde{G}_Z, \quad (3.26)$$

where G_Z coincides with the Weyl-ordered G_1 in the classical case (cf. (2.14)) while \tilde{G}_Z contains the quantum correction term

$$\begin{aligned} \tilde{G}_Z &= \times \text{tr}^1 \text{tr}^2 \dots (e_{ij}^1 \otimes e_{ji}^2)[X_Z^1 \otimes X_Z^2] \dots \times \\ &= : \text{tr}^1 \text{tr}^2 \dots (e_{ij}^1 \otimes e_{ji}^2)[X_Z^1 \otimes X_Z^2 + 2(1 - \cos \pi \hbar)e_{11}^1 \otimes e_{22}^2] \dots : . \end{aligned}$$

Here $e_{ij}^1 \otimes e_{ji}^2$ is the standard r -matrix that permutes the spaces “1” and “2,” and as a result, the “skein” relation of form (2.14) appears. Locally, this relation has exactly the form proposed by Turaev [37], i.e., for two graph simple geodesics intersecting at a single point, we have the defining relation

$$G_1 \times G_2 = e^{-i\pi\hbar/2} \left(\text{arc}_1 \right) \left(\text{arc}_2 \right) + e^{i\pi\hbar/2} \left(\text{arc}_1 \right) \left(\text{arc}_2 \right) . \tag{3.27}$$

(The order of crossing lines corresponding to G_1 and G_2 depends on which quantum geodesic occupies the first place in the product; the rest of the graph remains unchanged for all items in (3.27)). Note, however, that if the quantum geodesics G_1 and G_2 correspond to graph simple geodesics, we may turn the geodesic \tilde{G}_Z again into the graph simple geodesic \tilde{G}'_Z by performing the quantum flip with respect to the edge Z .

If we now compare two unambiguously determined expressions: \tilde{G}'_Z , which must be Weyl ordered, and \tilde{G}_Z obtained from the geodesic algebra, we find that $\tilde{G}_Z = \tilde{G}'_Z$. This enables us to formulate the main assertion.

Lemma 3.7 ([6]). *There exists a unique quantum ordering $\times \dots \times$ (3.18), which is generated by the quantum geodesic algebra (3.27) and consistent with the quantum mapping class groupoid transformations (2.8), i.e., so that the quantum geodesic algebra is invariant under the action of the quantum mapping class groupoid.*

We can now relax the constraints of graph simplicity of curves: as the quantum geodesic algebra is quantum mapping class group invariant, having two arbitrary embedded geodesics with a single intersection, we can transform them using quantum morphisms to a canonical form of graph simple geodesics and employ the Weyl order. Relation (3.27) remains valid in both cases.

Let us now address the problem of multiple intersections. Here, we have the following lemma.

Lemma 3.8 ([6]). *If more than one intersection of two QMCs occurs, then the quantum skein relations (3.27) must be applied simultaneously at all intersection points.*

This lemma implies the standard Reidemeister moves for curves on a graph, where the empty loop gives rise to a factor $-e^{-i\pi\hbar} - e^{i\pi\hbar}$; that is, for geodesics intersecting generically, apply (3.27) simultaneously at all intersection points to obtain the Reidemeister moves.

Remark 3.1. The quantum algebra \mathcal{M}_1^2 was studied in [6], where the exact correspondence with Kauffman bracket skein quantization of the corresponding Poisson algebra of geodesics (see [5]) was observed. In [4], this algebra arises as the quantum deformation algebra of the classical groupoid Poisson relations for the group $SL(4)$. This algebra is however only one among many quantum Nelson–Regge algebras corresponding to Riemann surfaces of higher genera which are described in the next section.

3.4 Quantizing the Nelson–Regge algebras

The algebra (2.17) was quantized by the deformation quantization method in [21], [20]. We now explicitly implement the quantization conditions (3.1). It is convenient to represent the (classical or quantum) elements \mathcal{G}_{ij} as chords connecting the points of the cyclically ordered set of indices i, j . There is then a trichotomy: if two chords do not intersect, then the corresponding geodesics do not intersect either, and the quantum geodesics commute (Figure 4 (a)); if two chords have a common endpoint, then the corresponding geodesics intersect at one point, and the three quantum geodesics \mathcal{G}_{ij}^{\hbar} , \mathcal{G}_{jk}^{\hbar} , \mathcal{G}_{ki}^{\hbar} (as depicted in Figure 4 (b)) constitute the quantum subalgebra $so_q(3)$; if two chords intersect at an interior point (as depicted in Figure 4 (c)), then the corresponding geodesics intersect in two points, and the corresponding quantum geodesics \mathcal{G}_{ij}^{\hbar} and \mathcal{G}_{kl}^{\hbar} , $i < k < j < l$, satisfy the commutation relation

$$[\mathcal{G}_{ij}^{\hbar}, \mathcal{G}_{kl}^{\hbar}] = \xi (\mathcal{G}_{ik}^{\hbar} \mathcal{G}_{jl}^{\hbar} - \mathcal{G}_{il}^{\hbar} \mathcal{G}_{jk}^{\hbar}) \tag{3.28}$$

with the usual commutator (not the q -commutator) and where again $\xi = q - q^{-1}$.

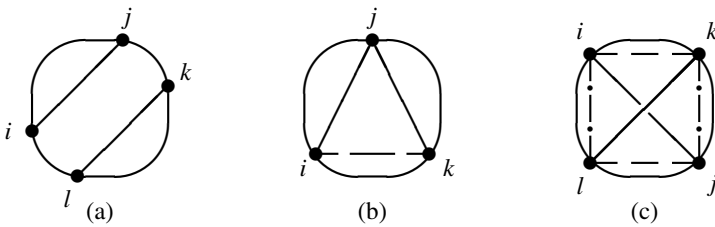


Figure 4. Picture of Nelson–Regge quantum relations.

3.5 Improving the quantum ordering

We now extend the construction of QMCs by taking the products of operatorial matrices X_Z , L , and R along oriented geodesics as before, but we do not apply the trace operation as in the definition of the geodesic operators.

For the three cases of oriented curves depicted in Figure 5 below, we may apply the indicated quantum Whitehead move and calculate as follows:

For curve 1, we obtain

$$\begin{aligned}
 & X_{B'} L X_{-Z} L X_{A'} \\
 &= \begin{pmatrix} e^{-B'/2} e^{-Z/2} e^{A'/2} + e^{-B'/2} e^{Z/2} e^{A'/2} & -e^{-B'/2} e^{Z/2} e^{-A'/2} \\ e^{B'/2} e^{-Z/2} e^{A'/2} & 0 \end{pmatrix} \\
 &= \begin{pmatrix} e^{A/2-B/2} & -e^{-A/2-B/2-\frac{1}{4}[A,B]} \\ e^{A/2+B/2-\frac{1}{4}[A,B]} & 0 \end{pmatrix} \text{ for } B' \neq A' \quad (3.29) \\
 &= e^{-\frac{1}{8}[A,B]} \begin{pmatrix} e^{-B/2} e^{A/2} & -e^{-B/2} e^{-A/2} \\ e^{B/2} e^{A/2} & 0 \end{pmatrix} \text{ for } B' \neq A' \\
 &= \begin{cases} e^{-\frac{1}{8}[A,B]} X_B L X_A & \text{for } A \neq B, \\ X_B L X_A & \text{for } A = B. \end{cases}
 \end{aligned}$$

For curve 2, we obtain

$$\begin{aligned}
 & X_{C'} R X_{-Z} L X_{A'} \\
 &= \begin{pmatrix} e^{-C'/2} e^{Z/2} e^{A'/2} & -e^{-C'/2} e^{Z/2} e^{-A'/2} \\ e^{C'/2} e^{Z/2} e^{A'/2} + e^{C'/2} e^{-Z/2} e^{A'/2} & -e^{C'/2} e^{Z/2} e^{-A'/2} \end{pmatrix} \\
 &= \begin{pmatrix} e^{A/2-C/2+Z/2+\frac{1}{4}[A,B]} & -e^{-A/2-C/2+Z/2} - e^{-A/2-C/2-Z/2} \\ e^{A/2+C/2+Z/2} & -e^{-A/2+C/2+Z/2-\frac{1}{4}[A,B]} \end{pmatrix} \quad (3.30) \\
 &= \begin{pmatrix} e^{-C/2} e^{Z/2} e^{A/2} & -e^{-C/2} e^{Z/2} e^{-A/2} - e^{-C/2} e^{-Z/2} e^{-A/2} \\ e^{C/2} e^{Z/2} e^{A/2} & -e^{C/2} e^{Z/2} e^{-A/2} \end{pmatrix} \\
 &= X_C L X_Z R X_A,
 \end{aligned}$$

and if edges A and C coincide, we merely use the same splitting as in (3.11); no additional factors arise.

For curve 3, we obtain

$$\begin{aligned}
 & X_{C'} R X_{-Z} R X_{D'} \\
 &= \begin{pmatrix} 0 & -e^{-C'/2} e^{Z/2} e^{-D'/2} \\ e^{C'/2} e^{-Z/2} e^{D'/2} & -e^{C'/2} e^{Z/2} e^{-D'/2} - e^{-C'/2} e^{-Z/2} e^{-D'/2} \end{pmatrix} \\
 &= \begin{pmatrix} 0 & -e^{-C/2-D/2+\frac{1}{4}[A,B]} \\ e^{C/2+D/2+\frac{1}{4}[A,B]} & -e^{C/2-D/2} \end{pmatrix} \text{ for } C' \neq D' \quad (3.31) \\
 &= e^{\frac{1}{8}[A,B]} \begin{pmatrix} 0 & -e^{-C/2} e^{-D/2} \\ e^{C/2} e^{D/2} & -e^{C/2} e^{-D/2} \end{pmatrix} \text{ for } C' \neq D'
 \end{aligned}$$

$$= \begin{cases} e^{\frac{1}{8}[A,B]} X_C R X_D & \text{for } C \neq D, \\ X_C R X_D & \text{for } C = D. \end{cases}$$

In these formulas, we have used the identities

$$e^{A'/2+C'/2+Z/2} + e^{A'/2+C'/2-Z/2} = : e^{A/2+C/2} \frac{1}{1+e^{-Z}} (e^{Z/2} + e^{-Z/2}) : \tag{3.32}$$

$$= e^{A/2+C/2+Z/2}$$

and

$$e^{-A'/2-C'/2+Z/2} = : e^{-A/2-C/2} (1+e^{-Z}) e^{Z/2} : \tag{3.33}$$

$$= e^{-A/2-C/2+Z/2} + e^{-A/2-C/2-Z/2},$$

where vertical dots denote the Weyl ordering (3.21) as before implying total symmetrization with respect to all the variables $\{A, B, C, D, Z\}$.

As we have just observed and amazingly enough, the only thing that changes under a quantum Whitehead move is the overall factor standing by the product of matrices, and even this factor can easily be taken into account if included with each left-turn matrix L is an overall factor $e^{-\frac{1}{8}[A,B]}$, and included with each right-turn matrix R is an overall factor $e^{\frac{1}{8}[A,B]}$.

The subtlety of potentially multiple intersections between curves corresponds to the possibility that edges of the graph may coincide. As in the previously useful convenient fiction of splitting such an edge into two half-edges, the formulas (3.29), (3.30), and (3.31) remain valid if we replace there $A'/2$ by $A'/4$ and $A/2$ by $A/4$ in case $A = C$ for instance.

We thus have the quantum analogue of Lemma 2.4.

Lemma 3.9. *For any oriented non boundary-parallel geodesic γ , taking the finite (periodic) sequence of matrices with quantum entries in X_{Z_i} as before but making the replacement*

$$\tilde{L} = q^{-1/4} L, \quad \tilde{R} = q^{1/4} R, \tag{3.34}$$

where $q = e^{-i\pi\hbar}$, the resulting product of matrices

$$P_\gamma = (X_{Z_n} \tilde{L} X_{Z_{n-1}} \tilde{R} \dots X_{Z_3} \tilde{L} X_{Z_2} \tilde{R} X_{Z_1} \tilde{R}) \tag{3.35}$$

is invariant under quantum Whitehead moves.

Remark 3.2. In the case of a boundary-parallel curve, where a sub-word of the form $X_{D'} \tilde{R} X_{-Z} \tilde{R}$ transforms to a sub-word of the form $X_D \tilde{\tilde{R}}$, we must set the resulting turn matrices to be

$$\tilde{\tilde{L}} = q^{-1/2} L, \quad \tilde{\tilde{R}} = q^{1/2} R,$$

i.e., they have quantum factors doubled in comparison with \tilde{L} and \tilde{R} .

We now address the question of the proper quantum ordering. Using mapping class group transformations, we can reduce any simple curve either to the form of

a graph simple curve with exactly one left and one right turn (c.f. formulas (2.15) and Figure 5) if this curve is not boundary-parallel, or to the form $X_D \tilde{R}$ or $X_D \tilde{L}$ for a boundary-parallel curve. In both cases, each term in the corresponding Laurent polynomial must be self-adjoint, which immediately results in the Weyl ordering by Proposition 3.6.

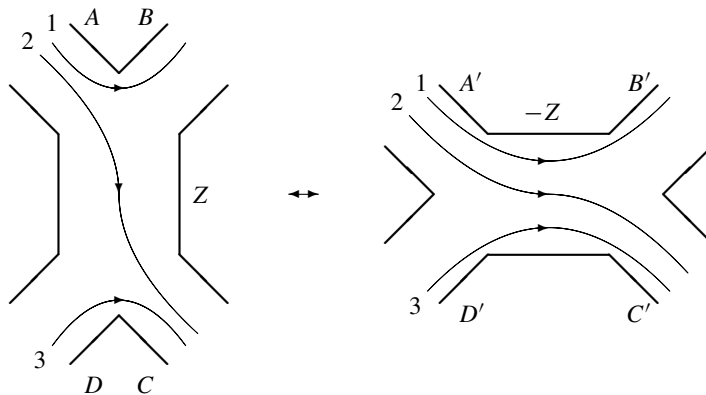


Figure 5. Three cases of flips for geodesics.

This observation does not suffice to derive the proper quantum ordering in our quantization of Thurston theory since we cannot consider in a consistent way an infinite product of matrices corresponding to infinite leaves of a measured foliation. Our tools for analyzing the Thurston theory in this case will devolve to naturality of the mapping class group action on the QMC algebra and an operatorial version of infinite continued fractions in Section 5.3.4. On the other hand, the improved quantum ordering is used in our analysis of closed geodesics on the torus, i.e., of operatorial finite continued fraction expansions in Section 5.3.1.

4 Classical Thurston theory of surfaces

Let F_g^s denote an oriented smooth surface with $s \geq 0$ punctures (so F_g^s may be closed without boundary in this section), with genus $g \geq 0$, and with negative Euler characteristic $2 - 2g - s < 0$.

4.1 Measured foliations and Thurston’s boundary

Define a *measured foliation* on $F = F_g^s$ to be a one-dimensional topological foliation \mathcal{F} of F , where in a neighborhood of any $p \in F$, \mathcal{F} must restrict (in an appropriate chart) to the horizontal foliation as in Figure 6 (a) or it must restrict to a foliation with one n -pronged singularity at p , for $n \geq 3$, as illustrated in Figure 6 (b).

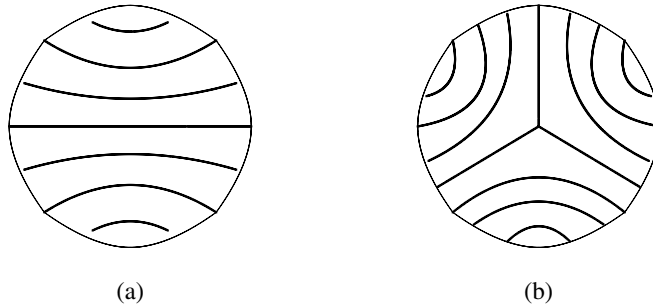


Figure 6. Pictures of foliations.

Furthermore, \mathcal{F} comes equipped with a *transverse measure* μ , which assigns to any arc a in F that is transverse to \mathcal{F} a real number $\mu(a) \in \mathbb{R}_{\geq 0}$, where μ is required to satisfy:

No holonomy. If a_0, a_1 are homotopic through arcs a_t transverse to \mathcal{F} , for $0 \leq t \leq 1$, keeping the endpoints of a_t on the same leaf for all t , then $\mu(a_0) = \mu(a_1)$.

σ -Additivity. If a is the serial concatenation of transverse arcs a_1, a_2, \dots , then $\mu(a) = \sum_{i \geq 1} \mu(a_i)$.

In other words, in the neighborhood of a non-singular point, there is a local chart $\phi: U \rightarrow \mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$ so that $\phi^{-1}(\{y = \text{constant}\})$ are the leaves of the foliation $\mathcal{F} \cap U$. If two charts U_i and U_j intersect, then the transition function ϕ_{ij} are of the form $\phi_{ij}(x, y) = (h_{ij}(x, y), c_{ij} \pm y)$, where c_{ij} is constant. In these coordinates, the transverse measure is $|dy|$. In case the transition functions can be chosen with constant sign $\phi_{ij}(x, y) = (h_{ij}(x, y), c + y)$, i.e., if the foliation is “transversely orientable”, then (away from the singular points) y is the primitive of a closed one-form on F .

Another canonical construction of a measured foliation on a Riemann surface is given by taking the leaves of the foliation to be the level sets of a harmonic function, where the transverse measure is given by integrating the conjugate differential along transverse arcs. Still another example is given by the homotopy class of a finite collection of disjointly embedded (weighted) curves, as we shall see.

There is an equivalence relation on measured foliations generated by isotopy and *Whitehead collapse* as illustrated in Figure 7, and the set of all equivalence classes (including the empty measured foliation \emptyset) is denoted $\mathcal{MF} = \mathcal{MF}(F)$, where the class of (\mathcal{F}, μ) is denoted $[\mathcal{F}, \mu]$. To naturally topologize \mathcal{MF} , we introduce the discrete set $\mathcal{S} = \mathcal{S}(F)$ consisting of all free homotopy classes $[c]$ of simple closed curves c in F which are neither null homotopic nor puncture-parallel. We shall also require the set $\mathcal{S}' = \mathcal{S}'(F)$ consisting of homotopy classes of all disjointly embedded families of curves in F , where each component of the family lies in \mathcal{S} and no two components are homotopic.

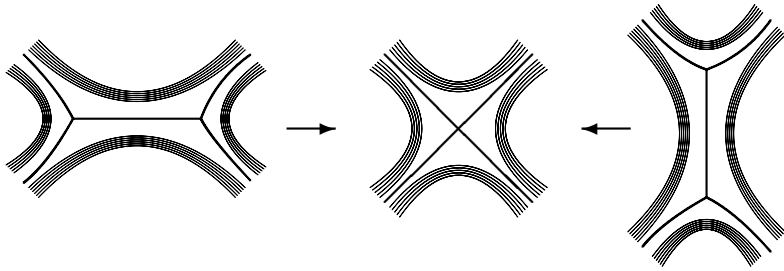


Figure 7. Whitehead collapses on foliations.

One can show [9] that for any $[c] \in \mathcal{S}$, there is a representative $c_{\mathcal{F}} \in [c]$ which minimizes the μ -transverse measure in its homotopy class; furthermore, given equivalent measured foliations (\mathcal{F}_i, μ_i) , for $i = 1, 2$, we have $\mu_1(c_{\mathcal{F}_1}) = \mu_2(c_{\mathcal{F}_2})$ [9], and hence there is a well-defined mapping

$$J : \mathcal{MF} \rightarrow \mathbb{R}_{\geq 0}^{\mathcal{S}},$$

$$(\mathcal{F}, \mu) \mapsto (i_{(\mathcal{F}, \mu)} : [c] \mapsto \mu(c_{\mathcal{F}})),$$

where the empty measured foliation \emptyset is identified with $\vec{0} = \{0\}^{\mathcal{S}}$. This mapping J is an injection [9] and induces a topology on \mathcal{MF} (where a neighborhood of \emptyset is homeomorphic to a cone from \emptyset over $\mathcal{MF}/\mathbb{R}_{>0}$, with the natural action by homothety of $\mathbb{R}_{>0}$ on measures). The function $i_{(\mathcal{F}, \mu)}$ is called the (geometric) intersection function of the measured foliation (\mathcal{F}, μ) .

Given $[c] \in \mathcal{S}$, there is a corresponding measured foliation defined as follows. Choose a representative c of $[c]$ and collapse $F - c$ onto a spine (making further choices) to build a foliation \mathcal{F}_c of F , whose leaves either lie in the spine or are homotopic to c , and choose a transverse measure μ_c on \mathcal{F}_c that pulls back under the collapsing map the counting measure δ_c on c . It is a classical fact due to Whitehead [9] that the resulting Whitehead equivalence class $[\mathcal{F}_c, \mu_c]$ is well-defined independent of any choices. Taking the projective class of this foliation, we may thus regard

$$\mathcal{S} \subseteq \mathcal{PF}_0.$$

Furthermore by construction, $i_{[\mathcal{F}_c, \mu_c]}([d])$, for $[d] \in \mathcal{S}$, is just the geometric intersection number of $[c]$ and $[d]$, i.e., the total number of intersections of representatives c with d , where c and d intersect minimally. More generally, given a family of curves c_1, \dots, c_n representing a point of \mathcal{S}' , together with a collection $w_1, \dots, w_n \in \mathbb{R}_{>0}$ of “weights”, we may again collapse to a spine of $F - \bigcup\{c_i\}_1^n$ to produce a foliation \mathcal{F} and choose a measure μ on \mathcal{F} that pulls back the weighted sum $\sum_1^n w_i \delta_{c_i}$ to get a well-defined Whitehead equivalence class $[\mathcal{F}, \mu]$. As mentioned before, one may thus associate a measured foliation to a weighted curve family.

Of special interest is the subspace $\mathcal{MF}_0 = \mathcal{MF}_0(F)$ consisting of all measured foliations (\mathcal{F}, μ) of compact support, i.e., any leaf of \mathcal{F} with a transverse arc a so that $\mu(a) > 0$ must be disjoint from a neighborhood of the punctures and no such leaf is puncture parallel. There is a natural $\mathbb{R}_{>0}$ -action on $\mathcal{MF} - \{\emptyset\}$ and $\mathcal{MF}_0 - \{\emptyset\}$ given by scaling the transverse measure, and the corresponding quotients

$$\begin{aligned} \mathcal{PF} &= \mathcal{PF}(F) = (\mathcal{MF}(F) - \{\emptyset\})/\mathbb{R}_{>0}, \\ \mathcal{PF}_0 &= \mathcal{PF}_0(F) = (\mathcal{MF}_0(F) - \{\emptyset\})/\mathbb{R}_{>0} \end{aligned}$$

are the spaces of central interest in the sequel. The projective class of (\mathcal{F}, μ) and $[\mathcal{F}, \mu]$, respectively, will be denoted $(\mathcal{F}, \bar{\mu})$ and $[\mathcal{F}, \bar{\mu}]$.

A point of $\mathcal{T} = \mathcal{T}(F_g^s) = \mathcal{T}_g^s$ may be regarded as the class of a hyperbolic metric on F , i.e., a complete finite-area Riemannian metric on F of constant Gauss curvature -1 . We shall also require the ‘‘Yamabe space’’ $\mathcal{Y} = \mathcal{Y}(F)$ of all complete finite-area Riemannian metrics on F of constant Gauss curvature $-x^2$, for some $x \in \mathbb{R}_{>0}$. \mathcal{Y} is canonically homeomorphic to $\mathcal{T}_g^s \times \mathbb{R}_{>0}$, where (ρ, x) corresponds to the class of the metric $x\rho$, and we let $\pi: \mathcal{Y} \rightarrow \mathcal{T}$ denote the projection onto the first factor. Define the map

$$\begin{aligned} I: \mathcal{Y} &\rightarrow \mathbb{R}_{\geq 0}^s, \\ \rho &\mapsto \ell_\rho(\cdot), \end{aligned}$$

where $\ell_\rho([c])$ is the ρ -length of the unique ρ -geodesic in the homotopy class $[c]$. Thus, if $\rho \in \mathcal{Y}$ corresponds to $(\pi(\rho), x)$, then $I(\rho) = I(\pi(\rho), x) = x I(\pi(\rho), 1)$.

The basic facts [9] are that $I: \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}^s$ and $J: \mathcal{MF}_0(F) \rightarrow \mathbb{R}_{\geq 0}^s$ are embeddings with disjoint images, and we may define a completion $\bar{\mathcal{Y}}$ of \mathcal{Y} in $\mathbb{R}_{\geq 0}^s$ by setting

$$\bar{\mathcal{Y}} = I(\mathcal{Y}) \cup J(\mathcal{MF}_0)$$

and identifying \mathcal{Y} with $I(\mathcal{Y})$. Passing to quotients under the homothetic actions of $\mathbb{R}_{>0}$ on $\mathbb{R}_{\geq 0}^s - \{0\}$, on $\mathcal{Y} - \{0\}$, and on $\mathcal{MF}_0 - \{\emptyset\}$, we obtain Thurston’s compactification

$$\bar{\mathcal{T}}_g^s = (\mathcal{T}_g^s \cup \mathcal{PF}_0) \approx (\bar{\mathcal{Y}} - \{\vec{0}\})/\mathbb{R}_{>0}$$

of $\mathcal{T} \approx (\mathcal{Y} - \{0\})/\mathbb{R}_{>0}$ by $\mathcal{PF}_0 \approx (\mathcal{MF}_0 - \{\emptyset\})/\mathbb{R}_{>0}$.

Theorem 4.1. 1. ([35], [9], [24]) $\mathcal{PF}_0(F_g^s)$ is naturally a piecewise linear sphere of dimension $6g - 7 + 2s$ which compactifies \mathcal{T}_g^s to produce a closed ball $\bar{\mathcal{T}}_g^s$.

2. ([35], [26]) The action of the mapping class group MC_g^s on \mathcal{T}_g^s extends continuously to an action on $\bar{\mathcal{T}}_g^s$, where the action on $\mathcal{PF}_0(F_g^s)$ is the natural one, and there are explicit piecewise linear formulas for the action of Dehn twist generators.

3. ([35], [9]) Suppose that a sequence of hyperbolic metrics ρ_i on F tends to a point $(\mathcal{F}, \bar{\mu}) \in \mathcal{PF}_0$. In the projectivization of $\mathbb{R}_{\geq 0}^s - \{0\}$, the projectivized length functions $\bar{\ell}_{\rho_i}$ of ρ_i converge to the projectivized intersection function $\bar{i}_{(\mathcal{F}, \bar{\mu})}$ of $(\mathcal{F}, \bar{\mu})$.

4. ([35], [1]) The function $i_{(\mathcal{F}, \mu)}: \mathcal{S} \rightarrow \mathbb{R}_{\geq 0}$ extends continuously to the “geometric intersection pairing” $\mathcal{MF}_0 \times \mathcal{MF}_0 \rightarrow \mathbb{R}_{\geq 0}$, that vanishes on the diagonal and which is also invariant under MC_g^s .

5. ([35], [23], [3]) The Weil–Petersson Kähler two-form on \mathcal{T}_g^s continuously extends (in the appropriate sense on Yamabe space) to a non-degenerate symplectic form, called “Thurston’s symplectic form”, on $\mathcal{MF}_0(F_g^s) (\approx \overline{\mathcal{Y}} - \mathcal{Y})$, which is invariant under MC_g^s .

Though there is all this beautiful natural structure on Thurston’s boundary, the quotient $\mathcal{PF}_0(F_g^s)/\text{MC}_g^s$ is maximally non-Hausdorff (i.e., its largest Hausdorff quotient is a singleton) as we shall see, so there is no correspondingly nice Thurston compactification on the level of Riemann’s moduli space.

Torus example. Recall that for the once-punctured torus $F = F_1^1$, the Teichmüller space is $\mathcal{T}_1^1 \approx \{z \in \mathbb{C} : |z| < 1\}$, and the mapping class group is $\text{MC}_g^s \approx \text{PSL}_2(\mathbb{Z})$. Indeed, the right Dehn twists M and L on the meridian and longitude, respectively, generate MC_1^1 , and a complete list of relations between them is given by $\iota = MLM = LML$ and $\iota^2 = 1$.

A point of $\mathcal{PF}_0(F)$ is uniquely determined by its “slope”, defined as follows. Fix two disjointly embedded ideal arcs x, y asymptotic to the puncture p which decompose F into an ideal quadrilateral, where $x \cup \{p\}$ is homotopic to the meridian, and $y \cup \{p\}$ is homotopic to the longitude. Given $[\mathcal{F}, \mu] \in \mathcal{MF}_0(F_1^1) - \{\emptyset\}$, compact support guarantees that $i_{[\mathcal{F}, \mu]}(x)$ and $i_{[\mathcal{F}, \mu]}(y)$ are well-defined and finite, and the ratio $|\theta| = i_{[\mathcal{F}, \mu]}(y)/i_{[\mathcal{F}, \mu]}(x) \in [0, \infty]$ is therefore projectively well-defined; we further imbue θ with a sign (when it is finite and non-zero) in the natural way, where the sign is positive if one (in fact, any) leaf of \mathcal{F} immediately after meeting x then meets the copy of y in the frontier of the ideal quadrilateral which lies to the right (where the orientation of the ideal quadrilateral is inherited from that of F_1^1). It is easy to see that the slope θ is a well-defined and complete invariant of $[\mathcal{F}, \bar{\mu}] \in \mathcal{PF}_0(F_1^1)$, where we regard $\theta \in S^1$ in the natural way. Thus, $\mathcal{PF}_0(F_1^1) \approx S^1$, and $\overline{\mathcal{T}}_1^1 = \mathcal{T}_1^1 \cup \mathcal{PF}_0(F_1^1)$ is a closed ball (which you should *not* identify with the Poincaré disk together with its circle at infinity). The slope $\theta = p/q$ is rational if and only if the measured foliation corresponds to the simple closed curve wrapping p times around the meridian and q times around the longitude, where p and q are relatively prime integers. Geometrically, deforming hyperbolic structure to pinch this curve, it is clear from elementary considerations of hyperbolic geometry that the corresponding geodesic length functions converge projectively to the geometric intersection number with this curve. Furthermore, the geometric intersection number of curves with slopes $p/q, r/s$ written in least terms is given by $|ps - qr|$.

4.2 Train tracks

A *train track* $\tau \subseteq F$ is a graph (where vertices are called “switches” and edges are called “branches”) together with the following extra structure:

Smoothness. τ is C^1 away from its switches. Furthermore, for each switch v of τ , there is a tangent line ℓ to τ at v in the tangent plane to F at v so that for each half-branch whose closure contains v , the one-sided tangent at v lies in ℓ .

Non-degeneracy. Vertices of τ are at least trivalent, and for any switch v of τ , there is an embedding $(0, 1) \rightarrow F$ with $f(\frac{1}{2}) = v$ which is C^1 as a map into F .

Geometry. Suppose that C is a component of $F - \tau$, and let $D(C)$ denote the double of C along the C^1 frontier edges of C so the non-smooth points in the frontier of C give rise to punctures of $D(C)$. We demand that the Euler characteristic of $D(C)$ be negative.

The smoothness condition is synonymously called the structure of a “branched one-submanifold” and leads to the fundamental notion of a graph smoothly supporting a curve or another train track as we shall see. According to the non-degeneracy condition, the half-branches incident on a fixed switch decompose canonically into two non-empty sets of “incoming” and “outgoing” branches. The geometric condition rules out the following complementary regions: smooth disks (i.e., nullgons), monogons, bigons, smooth annuli, and once-punctured nullgons, and will be further explained below.

Let $B(\tau)$ denote the set of branches of τ . A function $\mu: B(\tau) \rightarrow \mathbb{R}_{\geq 0}$ induces $\mu: \{\text{half-branches of } \tau\} \rightarrow \mathbb{R}_{\geq 0}$ in the natural way (where $\mu(b_{\frac{1}{2}}) = \mu(b)$ if $b_{\frac{1}{2}} \subseteq b$) and satisfies the *switch conditions* provided that for each switch v of τ , we have

$$\sum_{\substack{\text{outgoing} \\ \text{half-branches } b}} \mu(b) = \sum_{\substack{\text{incoming} \\ \text{half-branches } b}} \mu(b).$$

Such a function satisfying the switch conditions is called a *transverse measure* on τ , and τ itself is said to be *recurrent* if it supports a *positive* measure μ with $\mu(b) > 0$ for each branch of τ . In the sequel, train tracks will tacitly be assumed to be recurrent.

Torus example. There is a unique combinatorial type of recurrent trivalent train track τ in the surface $F = F_1^1$, and two embeddings of it as spine are illustrated in Figure 8. $F - \tau$ consists of a single once-punctured bigon. There are two branches of τ so that μ is uniquely determined by its values on these branches, and the weight on the remaining branch of τ is given by their sum according to the switch condition. Notice that this train track is “orientable” in the sense that the graph underlying τ admits an orientation where incoming points toward outgoing at each vertex. Thus, fixing an orientation on τ , a measure μ on τ uniquely determines a homology class in $H_1(F_1^1, \mathbb{R})$. Equivalently, every measured foliation of compact support on F_1^1 is transversely orientable.

Construction 4.1. Given a train track τ and a positive measure μ on it, we may construct a measured foliation of a neighborhood of τ in the following way. For each branch b of τ , take a rectangle of width $\mu(b)$ and length unity foliated by horizontal leaves. For each switch, place the rectangles of the incoming branches next to one another and likewise for the outgoing branches, and then finally glue the vertical edges of all the incoming to all the outgoing rectangles at each switch in the natural way preserving the transverse measure along the widths of the rectangles by the switch conditions. This produces from μ a measured foliation of a *tie neighborhood* of τ , where a vertical leaf in any rectangle is called a *tie*, and the *singular ties* arise from the vertical sides of the rectangles. As before by Whitehead's result, the Whitehead equivalence class of the resulting measured foliation is well-defined.

If a measured foliation arises in this way from a measure on a train track, then we say that the train track *carries* the measured foliation.

Let $U(\tau)$ denote the cone of all measures on τ , i.e., the subspace of $\mathbb{R}_{\geq 0}^{B(\tau)}$ determined by the switch conditions. There is again the natural $\mathbb{R}_{>0}$ -action on $U(\tau) - \{\vec{0}\}$ by homothety, and the quotient $V(\tau) = (U(\tau) - \{\vec{0}\})/\mathbb{R}_{>0}$ is the *polygon of projective measures* on τ . Construction 4.1 thus gives well-defined maps $U(\tau) \rightarrow \mathcal{MF}_0(F)$ and $V(\tau) \rightarrow \mathcal{PF}_0(F)$.

A recurrent train track is *maximal* if it is not a proper sub track of any recurrent train track. For general $F = F_g^s$, complementary regions to a maximal train track are either trigons or once-punctured monogons, but in the special case of the once-punctured torus, a maximal train track has a single complementary once-punctured bigon.

Theorem 4.2 ([35], [24]). *For any maximal recurrent train track τ in F , Construction 4.1 determines continuous embeddings*

$$U(\tau) \rightarrow \mathcal{MF}_0(F) \quad \text{and} \quad V(\tau) \rightarrow \mathcal{PF}_0(F)$$

onto open sets.

In fact, the geometric condition in the definition of train track precisely guarantees the injectivity in this theorem. In light of this result, one may regard a maximal train track in F as indexing a chart on the manifold \mathcal{PF}_0 , and we next study the transition functions of this putative manifold structure.

Torus example. For $F = F_1^1$, two embeddings of train tracks as spine are illustrated in Figure 8, and in fact, every foliation is carried by one of these two train tracks. The corresponding charts on the circle are also illustrated as well as the two points of intersection in the closures of these charts.

It is most convenient now to restrict to the "generic" case, where all switches of τ are trivalent. For each switch of τ , the decomposition of incident half-branches into incoming/outgoing thus consists of one singleton and one doubleton, and we say a branch of τ is *large* if it is a singleton at both its endpoints (which are then necessarily distinct) as illustrated with the branch labeled e in the left-hand side of Figure 9.

Likewise, if a branch is a doubleton at both its endpoints, then it is called *small*, while a branch which is neither small nor large is called *half-large*.

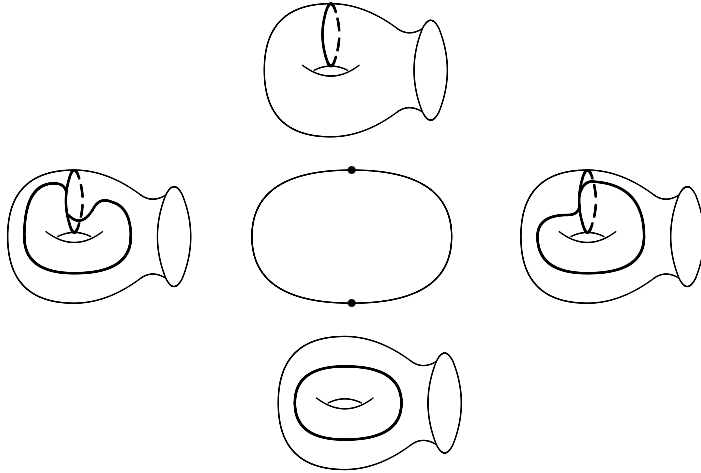


Figure 8. Charts for torus.

Define the combinatorial *splitting* of a measured train track (τ, μ) along a large branch e as illustrated in Figure 9, where we identify an edge with its μ transverse measure $\mu(e)$ for convenience. One imagines separating bands of horizontal leaves in the rectangle associated to e by excavating along the two “singular leaves” beginning at the endpoints of the large branch, i.e., beginning at the singular ties. If the measure μ is so that (either of) the singular leaves starting at an endpoint of e turn left or right, then the respective split is called a *left* (case 1) or *right* (case 3) split, while if the two singular leaves coincide for e , then the split is called a *collision* (case 2).

Splitting and smooth isotopy of measured train tracks generates an equivalence relation on the set of all measured train tracks in F , and we shall let $[\tau, \mu]$ denote the equivalence class of the measured train track (τ, μ) and $[\tau, \bar{\mu}]$ denote the equivalence class of the projectively measured train track $(\tau, \bar{\mu})$.

Theorem 4.3 ([24]). *If (τ_i, μ_i) are positively measured train tracks giving rise to corresponding measured foliations (\mathcal{F}_i, μ_i) via Construction 4.1, for $i = 1, 2$, then*

$$[\mathcal{F}_1, \mu_1] = [\mathcal{F}_2, \mu_2] \text{ if and only if } [\tau_1, \mu_1] = [\tau_2, \mu_2].$$

Thus, the space of all Whitehead equivalence classes of (projectivized) measured foliations is identified with the space of all splitting equivalence classes of (projectivized) measured train tracks up to isotopy.

There is another aspect to the splitting equivalence relation on the set of all measured train tracks. In addition to splitting, one considers also *shifting* along a half-large

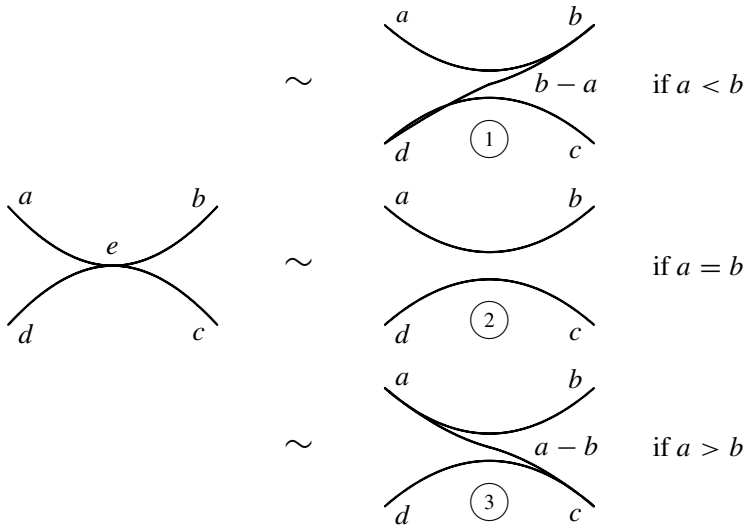


Figure 9. Splitting.

branch by pushing two confluent branches of a train track past one another as illustrated in Figure 10. Shifting plays a role in the later discussion, and a basic result in train track theory [24] is that if two train tracks are related by shifting, splitting, and smooth isotopy, then they are also related by splitting and smooth isotopy alone.

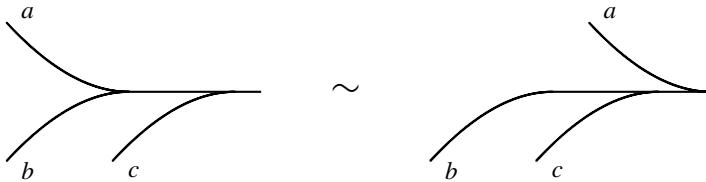


Figure 10. Shifting.

Let us finally give the idea of the proof that \mathcal{S} is dense in \mathcal{MF}_0 , as was mentioned before, by explaining density of \mathcal{S} in each chart $V(\mathcal{T}) \subseteq \mathcal{PF}_0$, for some maximal train track $\tau \subseteq F$. We may approximate any $\mu \in V(\tau) \subseteq (\mathbb{R}_{\geq 0}^{B(\tau)} - \{\vec{0}\})/\mathbb{R}_{>0}$ by a rational measure $\mu' \in (\mathbb{Q}_{\geq 0}^{B(\tau)} - \{\vec{0}\})/\mathbb{R}_{>0}$ (satisfying the switch conditions). Furthermore clearing denominators in μ' , there are $N \in \mathbb{Z}_{>0}$ and $\nu \in (\mathbb{Z}_{\geq 0}^{B(\tau)} - \{\vec{0}\})/\mathbb{R}_{>0}$ so that $N\mu' = \nu$. We may construct an embedded family of curves in F from ν by arranging $\nu(b) \geq 0$ tie-transverse strands parallel to b in a tie neighborhood of τ . By the switch conditions, there are at each vertex exactly as many incoming strands

as outgoing, and there is a unique way to combine strands near vertices to produce a disjointly embedded family of curves. Let us give each component curve a weight $1/N$ and combine any parallel curves while adding their weights to produce our desired weighted family of disjointly embedded curves. Let $[\mathcal{F}, \mu_1]$ denote the corresponding measured foliation (discussed before); tracing through the constructions, one finds that $[\tau, \mu'] = 1/N[\tau, \nu]$ gives rise to $[\mathcal{F}, \mu_1]$. Letting $\mu' \rightarrow \mu$ and projectivizing, it follows easily that families of disjointly embedded curves are dense in $V(\tau)$. With a little more work [24], one can approximate (in the topology of $V(\tau)$) such disjointly embedded families with a single curve, and this gives the asserted density of \mathcal{S} itself.

Thus, Thurston's boundary $\mathcal{PF}_0(F)$ is a completion of the set \mathcal{S} . In fact, one can approximate with a single non-separating curve (provided $g \neq 0$); since any two such curves are equivalent under the action of MC_g^s , it follows that the action of MC_g^s on $\mathcal{PF}_0(F)$ has a dense orbit, and the maximal non-Hausdorffness of the quotient, which was mentioned before, is thereby established.

4.3 Laminations

Each basic formulation of the objects presented so far, namely, measured foliations and measured train tracks, requires passage to the quotient under an appropriate equivalence relation. Thurston has given a more ethereal, elemental, and elegant description of these objects as “measured geodesic laminations”, where no passage to equivalence classes is necessary. Here we simply give the definition and a few basic properties referring the reader to [24] for instance for further details. A “lamination” \mathcal{L} in F is a foliation of a closed subset of F , and a “(transverse) measure” to \mathcal{L} is defined much as before as a σ -additive measure on arcs transverse to \mathcal{L} with the analogous condition of no-holonomy (where the homotopy is through arcs transverse to \mathcal{L} with endpoints disjoint from \mathcal{L}). \mathcal{L} is a “geodesic lamination” if its leaves are geodesic for some specified hyperbolic metric. The simplest case of a geodesic lamination is the geodesic representative of an element of $\mathcal{S}'(F)$. (In fact, for different choices of metric, the spaces of measured geodesic laminations are naturally identified via the circle at infinity in their universal covers, so we may speak of a geodesic lamination without the *a priori* specification of a metric.) A measured geodesic lamination has zero measure in F , and the intersection of a measured geodesic lamination \mathcal{L} with a transverse arc a in F is a Cantor set together with isolated points corresponding to intersections with simple geodesic curve components or arc components of \mathcal{L} , if any. There is a natural topology on the set of all measured geodesic laminations in F , which is induced by the weak topology on the set of all $\pi_1(F)$ -invariant measures supported on the Möbius band past infinity. $\mathcal{ML}(F)$ (and $\mathcal{ML}_0(F)$) is the corresponding space of measured geodesic laminations (and with compact support) and corresponding projectivization $\mathcal{PL}(F)$ (and $\mathcal{PL}_0(F)$). A basic result in Thurston theory is $\mathcal{ML}(F) \approx \mathcal{MF}(F)$, $\mathcal{ML}_0(F) \approx \mathcal{MF}_0(F)$, $\mathcal{PL}(F) \approx \mathcal{PF}(F)$, $\mathcal{PL}_0(F) \approx \mathcal{PF}_0(F)$, where train tracks give suitable charts on any of these piecewise linear manifolds. Furthermore,

the deformation theory due to Thurston, called “earthquaking” cf. [7], which we do not further discuss here, is most conveniently expressed in the context of laminations.

4.4 Dynamics on train tracks

In this section, we simply recall Thurston’s classification of surface automorphisms as well as recall several basic facts about “pseudo-Anosov” mappings. In the process, we develop further basic techniques which will be required in quantization. Since Thurston’s compactification produces a closed ball upon which the mapping class group acts continuously, one immediately is led to consider fixed points of this action.

Theorem 4.4 (Thurston’s classification ([35], [9])). *Any orientation-preserving homeomorphism $f: F \rightarrow F$ is homotopic to a diffeomorphism $f': F \rightarrow F$ which satisfies one of the following conditions (and the only overlap is between 1. and 2.).*

1. f' fixes a unique point of \mathcal{T} and is of finite order.
2. f' is “reducible” in the sense that f' fixes an element of \mathcal{S}' .
3. f' is “pseudo-Anosov” in the sense that there is some $\lambda > 1$ together with two measured foliations $(\mathcal{F}_\pm, \mu_\pm)$, which share singular points and are otherwise transverse, so that $f'(\mathcal{F}_\pm, \mu_\pm) = \lambda^{\pm 1}(\mathcal{F}_\pm, \mu_\pm)$. The projective classes $[\mathcal{F}_\pm, \bar{\mu}_\pm]$ are the unique fixed points of f on $\bar{\mathcal{T}}$. The invariant λ is called the dilatation of f or f' .

Notice the similarity with the trichotomy elliptic/parabolic/hyperbolic for fractional linear transformations. In the reducible case, one simplifies the dynamics by cutting F along a representative of the invariant element of \mathcal{S}' . A pseudo-Anosov mapping is the analogue of an Anosov map of the torus in the current context of surfaces with negative Euler characteristic.

In fact, train tracks provide a powerful tool for analyzing the dynamics of surface automorphisms owing to the fact that since a train track has a well-defined tangent line at each point, there is a coherent notion of a train track smoothly “carrying” a curve, another train track, or a lamination.

Suppose that κ is a smooth curve, a train track, or a measured geodesic lamination. We say that the train track τ carries κ and write $\kappa < \tau$, if there is a C^1 map $\phi: F \rightarrow F$ homotopic to the identity, called the supporting map, so that $\phi(\kappa) \subseteq \tau$, where the restriction of the differential $d\phi_p$ to the tangent line to κ at p is non-zero for every $p \in \kappa$.

We think of ϕ as squashing together nearly parallel strands of κ . For instance, any curve arising as before from an integral measure on τ is carried by τ , and more generally, any curve, train track or lamination κ which lies in a tie-neighborhood of τ and is transverse to the ties satisfies $\kappa < \tau$, where the supporting map collapses ties.

For example, if a train track σ arises from τ by splitting and shifting (but no collapsing), then $\sigma < \tau$, and we say that σ arises from τ by *unzipping*.

Suppose that the train track σ is transverse to a tie-neighborhood of τ , say with supporting map $\phi: F \rightarrow F$. Let us enumerate the branches b_j of τ , for $j = 1, \dots, n$, and a_i of σ , for $i = 1, \dots, m$, and choose $x_j \in b_j$ for each j . There is then an $m \times n$ -matrix $A = (A_{ij})$ called the *incidence matrix*, where A_{ij} is the cardinality of $\phi^{-1}(x_j) \cap a_i$. It is clear that the incidence matrix $A: \mathbb{R}_{\geq 0}^m \rightarrow \mathbb{R}_{\geq 0}^n$ describes the inclusion $U(\sigma) \rightarrow U(\tau)$ in the train track coordinates.

We close this section with several basic results about pseudo-Anosov mappings.

Theorem 4.5 ([35], [22]). *A homeomorphism $f: F \rightarrow F$ is a pseudo-Anosov map if and only if there is a train track τ in F , with each component of $F - \tau$ an at most once punctured polygon, so that τ unzips to $f(\tau)$ with no collisions. Furthermore, the incidence matrix A of the carrying $f(\tau) < \tau$ is Perron–Frobenius, the eigenvector of A corresponding to the spectral radius λ gives the projective measure $\bar{\mu}_+$, and likewise the extreme eigenvector of the transpose of A gives $\bar{\mu}_-$.*

Given a measured train track (τ, μ) , consider the foliated neighborhood of τ determined by μ via Construction 4.1. Choose some enumeration of the switches of τ and serially follow the singular leaves from the switches until the first splitting (ignoring shifting), for the first switch, second switch, \dots , last switch, and then begin anew from the first switch. Suppose there are no collisions, and record the resulting sequence of right or left splits, so as to produce a semi-infinite word of rights and lefts.

Theorem 4.6 ([22]). *The right-left sequence is eventually periodic if and only if the corresponding measured foliation is fixed by some pseudo-Anosov mapping.*

An explicit and simple construction of pseudo-Anosov maps is given by the following result.

Theorem 4.7 ([29]). *Suppose that $\mathcal{C}, \mathcal{D} \in \mathcal{S}'$ admit representative arc families C, D intersecting minimally which satisfy the condition that each component of $F - \cup(C \cup D)$ is an at most once-punctured polygon. Take any composition w of Dehn twists to the right along elements of \mathcal{C} and to the left along elements of \mathcal{D} so that the Dehn twist along each element of \mathcal{C} or \mathcal{D} occurs at least once in w . Then w is pseudo-Anosov.*

Torus example. Consider a generic train track τ in F_1^1 , so τ has one large branch e and two small branches. The two small branches are canonically linearly ordered by first taking the branch a to the right and then the branch b to the left at either endpoint of e , and furthermore $U(\tau) \approx \mathbb{R}_{\geq 0}^{\{a,b\}}$, i.e., the measures of the small branches a, b are unconstrained and uniquely determine the measure on e as well. Given a measure $\mu \in U(\tau) - \{0\}$, start unzipping (τ, μ) along either singular leaf, i.e., split along e , to produce another measured train track (τ_1, μ_1) ; of course, τ_1 is combinatorially equivalent to τ . For definiteness, suppose that $B = \mu(b) > \mu(a) = A$, so the split is a left split. The edge corresponding to b is the large edge of τ_1 , and the two small

edges, in right/left order, have measures $(A, B - A)$. Continue unzipping, i.e., next split (τ_1, μ_1) along its large edge to produce (τ_2, μ_2) . Again suppose that $B > 2A$ for definiteness, so the second split is a left split as well, and the small branches of τ_2 have measures $(A, B - 2A)$. Continue unzipping (under the assumption that there are no collisions) until the first right split, say there are $a_1 = \lfloor \frac{B}{A} \rfloor > 1$ left splits before the first right split. Perform the right split along the large branch of (τ_{a_1}, μ_{a_1}) , where the measures on the small branches are $(A, B_1) = (A, B - m_1 A)$, to produce a train track whose small edges have measures $(A - B_1, B_1)$. Continue unzipping and suppose there are no collisions, i.e., suppose A and B are not rationally related, to produce a semi-infinite sequence of symbols L (for left splits) and R (for right splits). Let a_1 denote the number of L 's that begin this sequence, a_2 denote the length of the next consecutive sequence of R 's, a_3 the length of the next consecutive sequence of L 's, and so on. It follows from the discussion above that the continued fraction expansion of B/A is given by

$$B/A = a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}$$

Continued fractions occur in another related guise as well. The isomorphism $\text{MC}_1^1 \approx \text{PSL}_2(\mathbb{Z})$ is induced by $M \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $L \mapsto \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$, where M and L are the right Dehn twists on the meridian and longitude respectively. A mapping class in MC_1^1 is pseudo-Anosov, reducible, periodic if and only if the corresponding fractional linear transformation is hyperbolic, parabolic, elliptic respectively. Every hyperbolic element of $\text{PSL}_2(\mathbb{Z})$ is conjugate to a product

$$\begin{pmatrix} 1 & m_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n_1 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & m_k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n_k & 1 \end{pmatrix},$$

where $m_i, n_i > 0$ are unique up to cyclic permutation. Furthermore, $m_1, n_1, \dots, m_k, n_k$ are the partial quotients of the periodic continued fraction expansion of the dilatation of the corresponding pseudo-Anosov map. It follows from this discussion that in MC_1^1 , all pseudo-Anosov mappings arise from the previous theorem. Indeed, the theorem gives the construction of two semi-groups corresponding to right/left or left/right twisting on meridian/longitude. For each semi-group, it is easy to construct a train track τ in F_1^1 so that the matrix representation above precisely describes the action of the corresponding semi-group on the measures of the linearly ordered small branches of τ ; indeed, these two train tracks are illustrated in Figure 8.

4.5 Decorated measured foliations and freeways

In this section, we recall material from [23] which is required for quantization. If $\Gamma \subseteq F$ is a cubic fatgraph spine of F , then we may blow-up each vertex of Γ into a

little trigon as illustrated in Figure 11. The resulting object $\tau = \tau_\Gamma$ has both a natural branched one-submanifold structure and a fattening, and furthermore, components of $F - \tau$ are either little trigons or once-punctured nullgons. Thus, τ is not a train track, but it is almost a train track, and is called the *freeway* associated to Γ . Notice that each edge of Γ gives rise to a corresponding large branch of τ , and each vertex gives rise to three small branches. It is easy to see that every measured lamination of compact support in F is carried by the freeway τ . The frontier of a once-punctured nullgon component of $F - \tau$ is a puncture-parallel curve called a *collar curve* of F . A small branch is contained in exactly one collar curve, while a large branch may be contained in either one or two collar curves.

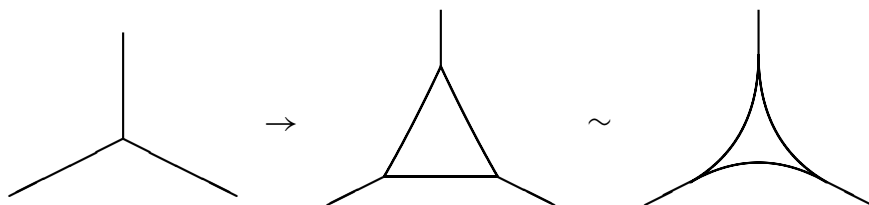


Figure 11. Freeway from fatgraph.

A *measure* on a freeway τ is a function $\mu \in \mathbb{R}^{B(\tau)}$ satisfying the switch conditions, where we wish to emphasize that the measure is not necessarily nonnegative (as it is for train tracks). Let $U(\tau)$ denote the vector space of all measures on τ . Notice that $\mu \in U(\tau)$ is uniquely determined by its values on the small branches alone, and the switch conditions are equivalent to the following “coupling equations”

$$\mu(a_1) + \mu(b_1) = \mu(e) = \mu(a_2) + \mu(b_2),$$

for any large branch e whose closure contains the switches $v_1 \neq v_2$, where a_i, b_i are the small branches incident on v_i for $i = 1, 2$. On the other hand, the values on the large branches alone also uniquely determine μ , and in fact, these values are unconstrained by the switch conditions. Indeed, letting a_i denote the large branches incident on a little trigon with opposite small branches α_i , for $i = 1, 2, 3$, we may uniquely solve for a measure μ on τ , where

$$\mu(\alpha_i) = \frac{1}{2}\{\mu(a_1) + \mu(a_2) + \mu(a_3) - 2\mu(a_i)\},$$

and so we identify $U(\tau) \approx \mathbb{R}^{\text{LB}(\tau)}$, where $\text{LB}(\tau)$ denotes the set of large branches of τ .

In particular, if μ is a nonnegative measure on τ , then the analogue of Construction 4.1 in the current context produces a well defined equivalence class of measured foliations in F , where this measured foliation will typically contain a collection of puncture-parallel annuli foliated by curves homotopic to collar curves. Deleting these

foliated annuli produces a well-defined (but possibly empty) class in $\mathcal{MF}_0(F)$. Thus, a nonnegative measure on τ canonically determines a point of $\mathcal{MF}_0(F)$ together with a nonnegative “collar weight”, i.e., the transverse measure of a transverse arc connecting the boundary components of the corresponding foliated annulus.

In the general case that μ is not necessarily nonnegative, suppose that C is a collar curve of τ . The switches of τ decompose C into a collection of arcs, each of which inherits a corresponding real-valued weight from μ . Let $\{\gamma_i\}_1^n$ denote the collection of real numbers associated to the small branches of τ that occur in C . Define the *collar weight* of C for μ to be $\mu_C = \min\{\gamma_i\}_1^n$. Define a *collar weight* on F itself to be the assignment of such a weight to each puncture.

We may modify the original measure $\mu \in U(\tau)$ by defining $\mu'(b) = \mu(b) - \mu_C$ if b is contained in the collar curve C for any small branch b of τ . Thus, μ' is a nonnegative measure on μ with identically vanishing collar weights that determines a corresponding element of $\mathcal{MF}_0(F)$.

We are led to define the space $\widetilde{\mathcal{MF}}_0 = \mathcal{MF}_0(F) \times \mathbb{R}^s$ of *decorated measured foliation* and summarize the previous discussion:

Theorem 4.8 ([23]). *The space $U(\tau) \approx \mathbb{R}^{\text{LB}(\tau)}$ gives global coordinates on $\widetilde{\mathcal{MF}}_0$, and there is a canonical fiber bundle $\Pi: \widetilde{\mathcal{MF}}_0 \rightarrow \mathcal{MF}_0$, where the fiber over a point is the set \mathbb{R}^s of all collar weights on F .*

Remark 4.1. The natural action of MC_g^s is by bundle isomorphisms of Π . Furthermore, Π admits a natural MC_g^s -invariant section $\sigma: \mathcal{MF}_0 \rightarrow \widetilde{\mathcal{MF}}_0$ which is determined by the condition of identically vanishing collar weights. The restriction of σ to $\mathcal{MF}_0 \subseteq \widetilde{\mathcal{MF}}_0$ gives a piecewise-linear embedding of the piecewise-linear manifold \mathcal{MF}_0 into the linear manifold (vector space) $\widetilde{\mathcal{MF}}_0 \approx U(\tau) \approx \mathbb{R}^{\text{LB}(\tau)}$.

4.6 Shear coordinates for measured foliations

We now give an equivalent parametrization of measured foliations in terms of “Thurston’s shear coordinates” that are close analogues of Thurston’s shear coordinates Z_α on $\mathcal{T}_H(F)$. In fact, we have already encountered these quantities when describing the splitting procedure train tracks (see Figure 9). There, excavating along two different singular leaves, we have obtained the “new” edge, which can turn either left or right (for splittings) or be absent (for collisions).

We assign a corresponding signed quantity (positive for right, negative for left) as follows. Given a measure μ on the long branches of the freeway τ associated to the fatgraph spine $\Gamma \subseteq F$, define the (*Thurston’s foliation-*)*shear coordinate* of the edge indexed by α to be

$$\zeta_\alpha = \frac{1}{2}(\mu(A) - \mu(B) + \mu(C) - \mu(D)),$$

in the notation of Figure 1 for nearby branches. From the very definition, ζ_α is independent of collar weights. Again, Thurston's foliation-shear coordinates are alternatively defined in terms of the signed transverse length of the arc between the singular leaves along Z_α , in analogy to the geometric interpretation given before of the shear coordinates on \mathcal{T}_H .

Note that the shear coordinates ζ_α are not independent. They are subject to the restrictions that

$$\sum_{\alpha \in I} \zeta_\alpha = 0 \quad (4.1)$$

for the sum over edges $\alpha \in I$ surrounding any given boundary component, and we shall refer to these conditions as the *face conditions* for shear coordinates. Thus, the space of foliation-shear coordinates is of dimension $\text{LB}(\tau) - n$, where we let n denote the number of boundary components. One sees directly that for any assignment of shear coordinates, there is a well-defined point of \mathcal{MF}_0 realizing them, thereby establishing a homeomorphism between \mathcal{MF}_0 and this sub-vector space $\mathbb{R}^{\text{LB}(\tau)-n} \subseteq \mathbb{R}^{\text{LB}(\tau)}$ of shear coordinates on the long branches of τ .

To describe the action of the mapping class group on foliation-shear coordinates, we shall give the transformation under Whitehead moves, i.e., derive the analogue of formula (2.8) for measured foliations, which is an elementary calculation using the formulas for splitting as follows.

Lemma 4.9. *Under the Whitehead move in Figure 1, the corresponding foliation-shear coordinates of the edges A, B, C, D, and Z situated as in Figure 1 are transformed according to formula (2.8)*

$$\begin{aligned} M_Z: (\zeta_A, \zeta_B, \zeta_C, \zeta_D, \zeta_Z) \\ \mapsto (\zeta_A + \phi_H(\zeta_Z), \zeta_B - \phi_H(-\zeta_Z), \zeta_C + \phi_H(\zeta_Z), \zeta_D - \phi_H(-\zeta_Z), -\zeta_Z) \end{aligned}$$

with

$$\phi_H(\zeta_Z) = (\zeta_Z + |\zeta_Z|)/2, \quad (4.2)$$

i.e., $\phi_H(x) = x$, for $x > 0$, and zero otherwise. All other shear coordinates on the graph remain unchanged.

Remark 4.2. Comparing expressions for the classical function $\phi(x) = \log(1 + e^x)$ and (4.2), one finds that the latter is a *projective limit* of the former:

$$\phi_H(x) = \lim_{\lambda \rightarrow +\infty} \frac{1}{\lambda} \phi(\lambda x) = \lim_{\lambda \rightarrow +\infty} \frac{1}{\lambda} \phi^{\hbar}(\lambda x), \quad (4.3)$$

that is, all three transformations coincide asymptotically in the domain of large absolute values (or large eigenvalues for the corresponding operators) of Teichmüller space coordinates $\{Z_\alpha\}$. We shall actively use this property in Section 5.3.2 when proving the existence of the quantization of Thurston's boundary for the punctured torus.

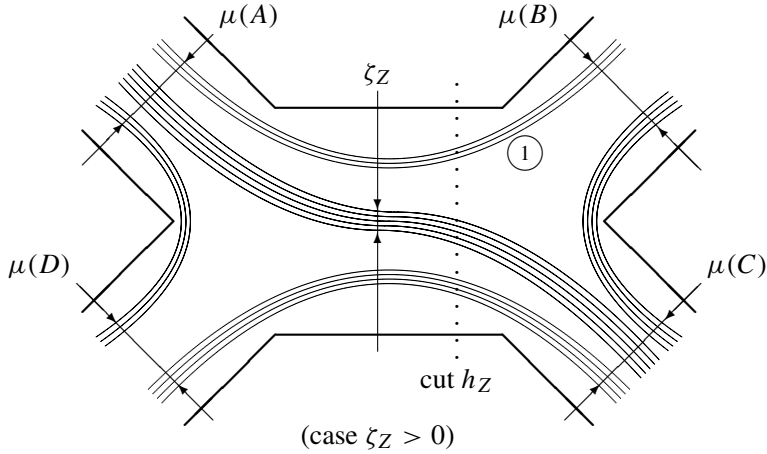


Figure 12. Foliation-shear coordinates.

5 On quantizing Thurston theory

5.1 Proper length of geodesics

Definition 5.1. The *proper length* $p.l.(\gamma)$ of a closed curve γ in the classical or quantum case is constructed from the quantum ordered operator P_γ associated to a closed oriented edge-path with basepoint (to begin the linearly ordered word P_γ) as

$$p.l.(\gamma) = \lim_{n \rightarrow \infty} \frac{1}{n} \operatorname{tr} \log 2T_n(P_\gamma/2), \tag{5.1}$$

where we take the principal branch of the logarithm and T_n are Chebyshev’s polynomials (cf. (3.20)). Since $T_n(\cosh \frac{t}{2}) = \cosh \frac{nt}{2}$, it follows that $p.l.(\gamma)$ agrees with half the hyperbolic length of γ in the Poincaré metric in the classical case.

More explicitly in the operatorial case, we can determine $p.l.(\gamma)$ explicitly in terms of the spectral expansion of the operator G_γ , which is known exactly. Namely, the basis of eigenfunctions of G_γ is “doubly reduced” in the sense that each eigenvalue (except 2, which is singular) with corresponding eigenfunction α_S , has the form $e^{S/2} + e^{-S/2}$, where S ranges over the entire real axis, and α_S has the same eigenvalue as α_{-S} . In fact, these functions coincide, so there is actually a representation on the positive real axis, which is nevertheless complete, and is singular at infinity and zero. We may define the proper length operator to be the one with the same eigenfunctions α_S for S positive (which constitute a basis in the function space) and with eigenvalues to be $|S/2|$. This operator $p.l.(\gamma)$ is then a well-defined operator on any compactum in function space.

The *proper length* of a QMC or GMC \hat{C} , again denoted $\text{p.l.}(\hat{C})$, is the sum of the proper lengths of the constituent geodesic length operators (or the sum of half geodesic lengths calculated in the Poincaré metric in the classical case) weighted by the number of appearances in the multiset.

5.2 Approximating laminations and the main theorem

Fix once and for all a spine Γ of F with corresponding freeway τ . A measure μ on τ gives rise to a (possibly empty) measured foliation in F together with a collar weight on the boundary components of F . Erasing collars yields an underlying measure $\mu_1 \geq 0$ on τ whose support is a sub-train track $\tau_1 \subseteq \tau$, and the measured train track (τ_1, μ_1) determines a (possibly empty) measured foliation. Via the canonical embedding of \mathcal{MF}_0 into $\widetilde{\mathcal{MF}}_0$ with vanishing collars, we may thus uniquely determine a point of \mathcal{MF}_0 by specifying foliation-shear coordinates on the long branches of τ satisfying the face conditions 4.1, i.e., \mathcal{MF}_0 is naturally identified with a codimension n subspace of $\mathbb{R}^{\text{LB}(\tau)}$. Passing to projective foliations, a point of \mathcal{PF}_0 is given by the projectivization $P\vec{\zeta}$ of a vector of foliation-shear coordinates $\vec{\zeta} = (\zeta_i)$, where $\zeta \in \mathbb{R}^{\text{LB}(\tau)} - \{0\}$, and i indexes the long branches of τ , i.e., the edges of Γ .

Definition 5.2. A sequence $\vec{n}^\beta = (n_i^\beta)$, for $\beta \geq 1$, of integer-valued n_i , for $i = 1, \dots, \text{LB}(\tau)$, on τ is an *approximating sequence* for the projectivized measured foliation $P\vec{\zeta}$ if the face conditions 4.1 hold on \vec{n}^β and if $\lim_{\beta \rightarrow \infty} n_i^\beta / n_j^\beta = \zeta_i / \zeta_j$ for all i, j with $\zeta_j \neq 0$.

Constructed from \vec{n} as an integral measure on τ is a GMC \hat{C} with integral collar weights. Just as with decorated measured foliations, components of \hat{C} which are puncture- or boundary-parallel can be erased to produce a corresponding multicurve to be denoted $\hat{C}_{\vec{n}}$. $\hat{C}_{\vec{n}}$ is carried by a sub-train track of τ , and it traverses the long branch of τ indexed by i some number, say, $m_i \geq 0$ of times, so m_i is the standard train track coordinate of integral transverse measure. In the usual notation as in Figure 12, one sees directly that $n_Z = \frac{1}{2}(m_A - m_B + m_C - m_D)$. We may also sometimes write $\hat{C}_{\vec{m}}$ for $\hat{C}_{\vec{n}}$

Definition 5.3. A *graph length* function with respect to the spine Γ is any linear function

$$\text{g.l.}_{\vec{a}}(\hat{C}_{\vec{n}}) = \text{g.l.}_{\vec{a}}(\hat{C}_{\vec{m}}) = \sum_i a_i m_i. \tag{5.2}$$

In particular, when all a_i are unity, the graph length is just the combinatorial length of $\hat{C}_{\vec{m}}$, i.e., the total number of edges of Γ traversed (with multiplicities) by all the component curves of $\hat{C}_{\vec{m}}$. When the spine Γ and \vec{a} are fixed or unimportant, then we shall write simply $\text{g.l.}(\hat{C}_{\vec{n}})$ or $\text{g.l.}(\hat{C}_{\vec{m}})$ for the graph length.

Any graph length function is evidently additive over disjoint unions of multicurves, and more generally, is a linear function of \vec{m} .

We next describe the bordered or punctured torus case in detail. Each multicurve on the torus is uniquely determined by three nonnegative integers (m_X, m_Y, m_Z) that satisfy one of the three triangle equalities $m_i = m_j + m_k$, where $\{i, j, k\} = \{X, Y, Z\}$. Projectivization allows us to re-scale so that m_j and m_k are relatively prime using this degree of freedom, so that the corresponding multicurve has just one component.

As in Figure 8, the space $\mathcal{PF}_0(F_1^1)$ of projectivized measured foliations with compact support is a piecewise-linear circle for $r + s = 1$, and an alternative family of charts on this circle is given in Figure 13. The relation \sim in Figure 13 denotes the equivalence between different boundary cases between two different charts, and arrows represent one-simplices in $\mathcal{PF}(F_1^1)$.

In Figure 13, we use the previous notation G_X , etc. (see (2.15)) but in a slightly different sense. Now, these quantities are (2×2) -matrices, not just geodesic functions, i.e., we do not evaluate traces in the corresponding formulas. There is thus an ambiguity in choosing the place in the graph where the matrix products begin. We indicate this place by drawing the reference cut (the dotted line). Changing the reference cut when moving along the circle corresponds to passing from one chart to another in the chart covering of the circle. Of course, choosing the reference cut does not affect the quantum trace operation.

In order to have a good transition in the boundary cases, for instance, in the upper case in Figure 13 (and the other cases are similar and omitted), we must ensure that the corresponding functions for the quantities $G_X^m G_Z^{(1)}$ and $G_X^m \tilde{G}_Z^{(1)}$ must coincide in the limit $m \rightarrow \infty$ with each other and with the corresponding quantity calculated merely for the “short” geodesic function G_X . To prove this, given two (2×2) -matrices G_X and G_Z corresponding to geodesic curves, we can conjugate them by respective unitary transformations U_X and U_Z to diagonal form with real eigenvalues $e^{\pm l_X/2}$ and $e^{\pm l_Z/2}$ since G_X, G_Z are hyperbolic. We then have

$$\text{tr } G_X^m G_Z = \text{tr} \begin{pmatrix} e^{ml_X/2} & 0 \\ 0 & e^{-ml_X/2} \end{pmatrix} V \begin{pmatrix} e^{l_Z/2} & 0 \\ 0 & e^{-l_Z/2} \end{pmatrix} V^{-1}, \quad V = U_X^{-1} U_Z \tag{5.3}$$

and the proper length (5.1) is $ml_X/2 + O(1)$.² In order to have a well-defined projective limit, we shall “kill” the factor m in a consistent way, and this can be achieved by dividing the result by any graph length function of the curve since $\text{g.l.}(G_X^m G_Z) = m \cdot \text{g.l.}(G_X) + \text{g.l.}(G_Z)$.

In the quantum case, however, the situation is much more involved. Indeed, let us consider an example of the product, which is of form $U^m V$, as in (5.3), where $U =$

²Unless the matrix $U_X^{-1} G_Z U_X$ has the form $\begin{pmatrix} 0 & -r^{-1} \\ r & \phi \end{pmatrix}$. For this matrix to determine a hyperbolic element, the quantity ϕ must be real greater than two. Multiplying this matrix by the diagonal matrix above, we obtain $\begin{pmatrix} 0 & -r^{-1} e^{ml_X/2} \\ r e^{-ml_X/2} & \phi e^{-ml_X/2} \end{pmatrix}$, so for sufficiently large m , the resulting product ceases to be hyperbolic, which is absurd.

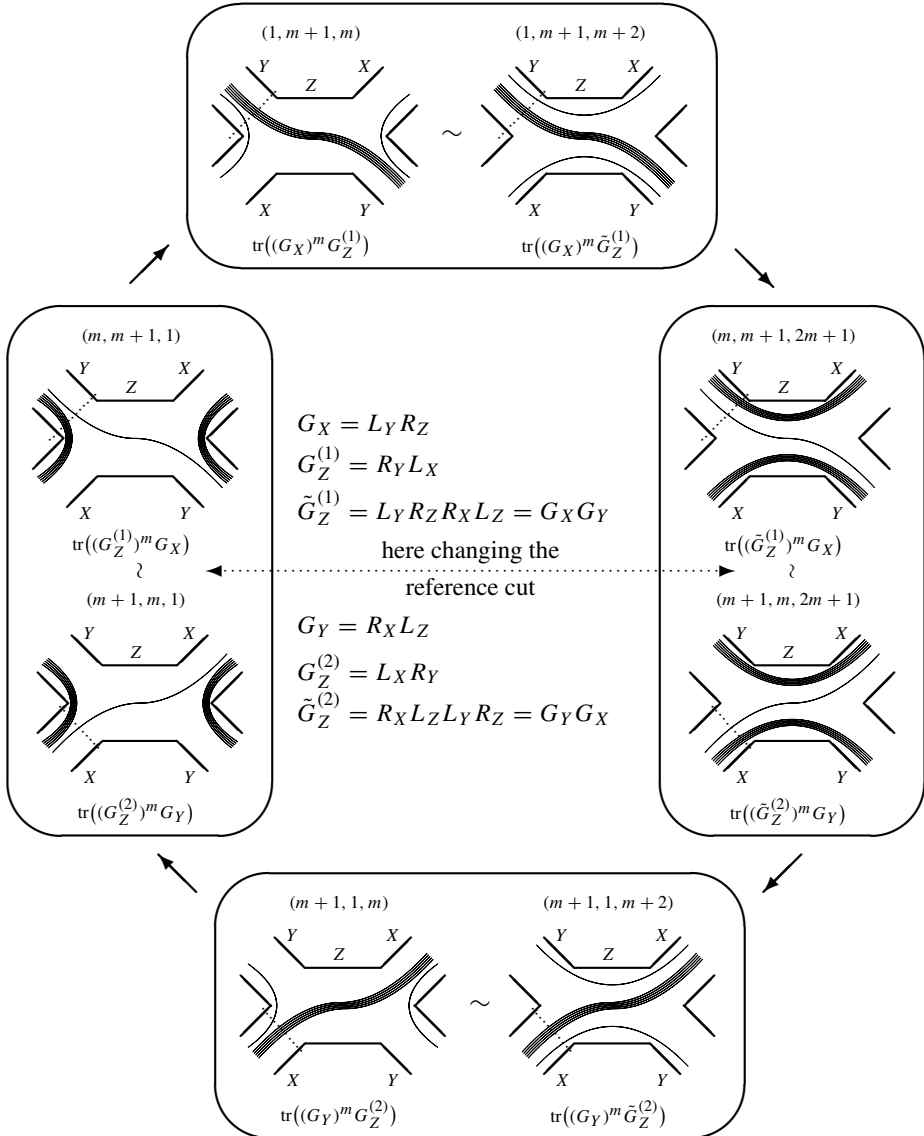


Figure 13. The circle $\mathcal{P} \mathcal{L}_0(F_1^1)$.

$e^{X/2}$, $V = e^{Y/2}$, and $[X, Y] = 4\pi i\hbar$. Thus, $(U^m)V = e^{mX/2+Y/2+m\pi i\hbar/2}$ while $V(U^m) = e^{mX/2+Y/2-m\pi i\hbar/2}$, and the corresponding logarithms do not coincide as $m \rightarrow \infty$. Moreover, even the Hermiticity condition does not often suffice to determine the proper length. For instance, given operators $U = e^{\alpha X^2}$ and $V = e^{i\beta\partial_x}$, we may calculate that $V(U^m)V = e^{m\alpha X^2+2i\beta\partial_x+m\alpha\beta/3}$, i.e., the proper length in this case is $\alpha X^2 + \alpha\beta/3$ and depends on the parameter β (of course, this correction is purely quantum). This illustrates that proving continuity for the boundary transitions in the quantum case requires more subtle estimates, which we perform in the next section after deriving recurrence relations for the operators of quantum approximating multicurves.

We may now formulate our main result on quantizing Thurston theory:

Theorem 5.1. *Fix a spine Γ of F_1^1 with corresponding freeway τ . Fix any projectivized vector $P\vec{\zeta}$ of foliation-shear coordinates on τ and any graph length function $g.l.$ For any approximating sequence \vec{n}^β to $P\vec{\zeta}$, the limit*

$$\lim_{\beta \rightarrow \infty} \frac{\text{p.l.}(\hat{C}_{\vec{n}^\beta})}{g.l.(\hat{C}_{\vec{n}^\beta})} \quad (5.4)$$

exists both in the classical case as a real number and in the quantum case as a weak operatorial limit.

Because both the numerator and denominator in the limit are additive, this limit is projectively invariant and defines a continuous function (in the classical case) or a weakly continuous family of operators (in the quantum case) on the circle $\mathcal{P}\mathcal{L}_0(F_{1,r}^s)$, for $r + s = 1$.

The proof of the previous theorem occupies the remainder of this section. The continued fraction structure intrinsic to the torus case is used extensively, and various analogous operatorial recursions are derived and studied. There is a second essentially combinatorial proof of this result, however, only in the classical case since we have no means to control the quantum ordering of the procedure. Nevertheless, the structures discovered are interesting, and we present this second proof in Appendix A, which depends upon the recursion (5.7) derived later in Lemma 5.2. Indeed, this basic recursion arises from ‘‘Rauzy–Veech–Zorich induction’’ [30] in the special case of the torus, which is derived from first principles in the next section.

5.3 Elements of the proof

5.3.1 Continued fraction expansion. In each one-simplex in $\mathcal{P}\mathcal{L}_0(F)$, illustrated as arrows in Figure 13, the approximating multicurve is determined by two nonnegative integers, m_1 and m_2 , where we assume that $m_1 > m_2$ with m_1 and m_2 relatively prime.

Define $\tilde{L}_0 \equiv G_X$ and introduce \tilde{L}_i , for $i \geq 1$, which is the matrix L_i (composed from the elementary matrices G_X and G_Y) in which the first two symbols of elementary matrices must be interchanged: if the first symbol of L_i is G_Y and the second is G_X , then the first symbol of \tilde{L}_i is G_X and the second is G_Y ; all other elementary matrices retain their forms. We illustrate this procedure in Figure 14.³

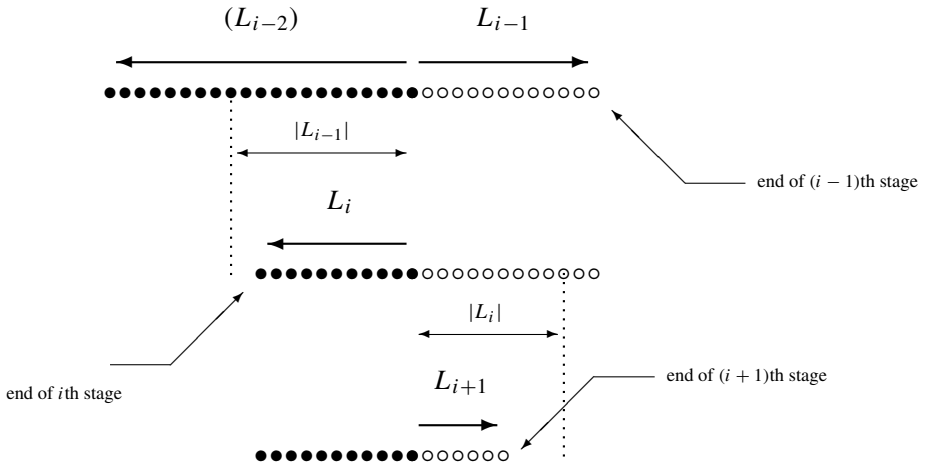


Figure 14. Threads a curve.

One thus sees directly that

$$\begin{aligned}
 L_2 &= (L_1)^{a_2-1} \tilde{L}_0 L_1 \\
 L_3 &= (\tilde{L}_2)^{a_3-1} L_1 L_2 \\
 L_4 &= (L_3)^{a_4-1} \tilde{L}_2 L_3 \\
 L_5 &= (\tilde{L}_4)^{a_5-1} L_3 L_4 \\
 &\vdots
 \end{aligned}$$

and this leads to the following recurrence relation.

Lemma 5.2. *Given the simple continued fraction expansion (5.5) of m_2/m_1 , the sequence of matrices L_n associated to the corresponding geodesic is given by the*

³Note that we terminate a stage whenever we come closer to the starting thread; this occurs at each stage on the opposite side from the previous stage, just as for continued fractions. The appearance of tilded quantities is explained as follow: each time during the recursion when we start a string “parallel” to some L_k from the right to the starting thread (in the circled domain), we must follow the same string of branches because, by definition, there are no threads in the string L_k except the very first thread that appears at a distance closer than $|L_k|$ to the starting thread. On the other hand, if we start from the left of the starting thread, then we must interchange exactly the first two appearances of matrices in the resulting string.

following recursion

$$\begin{aligned} L_{2i} &= (L_{2i-1})^{a_{2i}-1} \tilde{L}_{2i-2} L_{2i-1} \\ L_{2i+1} &= (\tilde{L}_{2i})^{a_{2i+1}-1} L_{2i-1} L_{2i} \end{aligned} \quad \text{for } i \geq 1, \quad \tilde{L}_0 = G_X, \quad L_1 = (G_X)^{a_1} G_Y. \quad (5.7)$$

Turning to the quantum case, we first show that the proper length operator (5.1) $\text{p.l.}(\times L_{2i} \times) / \text{g.l.}(L_{2i})$ must agree with $\text{p.l.}(\times L_{2i-1} \times) / \text{g.l.}(L_{2i-1})$ for $a_{2i} \rightarrow \infty$, that is, the operators corresponding to continued fractions of form (5.5) with large coefficient a_n must converge to the operator corresponding to the continued fraction terminated at the $(n - 1)$ st step. To this end, we must analyze the structure of matrix products and corresponding operators. Indeed, the operators L_i and \tilde{L}_i enjoy elegant commutation relations as we shall next see.

Notice that for every stage i we have a geodesic corresponding to the matrix L_i (because we can close the corresponding geodesic line without self-intersections), and we can therefore define the corresponding QMC

$$\mathcal{L}_i \equiv \times \text{tr } L_i \times.$$

The first observation pertains to L_i and \tilde{L}_i : One of the corresponding curves can be obtained from the other by a parallel shift along the cut h_Z illustrated in Figure 12. Thus, the curves are disjoint and hence are homotopic on the torus, and so $\text{tr } L_i = \text{tr } \tilde{L}_i$, i.e.,

$$\mathcal{L}_i = \tilde{\mathcal{L}}_i \quad \text{for } i \geq 1. \quad (5.8)$$

Furthermore, the curves L_i and L_{i+1} can be perturbed to have exactly one intersection as one sees by considering how the corresponding geodesic curves pass through the cut h_Z .⁴ In the case where $i = 2k$, we obtain (in the notation of (3.24))

$$\begin{aligned} [\mathcal{L}_{2k}, \mathcal{L}_{2k+1}]_q &= \xi \times \text{tr } L_{2k} L_{2k+1} \times, \quad \text{and} \\ [\mathcal{L}_{2k+1}, \mathcal{L}_{2k}]_q &= \xi \times \text{tr } L_{2k-1} (\tilde{L}_{2k})^{a_{2k+1}-1} \times, \end{aligned} \quad (5.9)$$

where the proper quantum ordering is assumed for the terms in the right-hand sides.

Formulas (5.9) are crucial when proving the continuity. Letting

$$\begin{aligned} I_m &\equiv \times \text{tr} (L_{2i-1})^{m-1} \tilde{L}_{2i-2} L_{2i-1} \times, \\ \mathcal{L}_{2i-1} &\equiv e^{\ell_X/2} + e^{-\ell_X/2}, \end{aligned}$$

we find

$$I_{m-1} \mathcal{L}_{2i-1} = q^{1/2} I_m + q^{-1/2} I_{m-2}, \quad (5.10)$$

$$\mathcal{L}_{2i-1} I_{m-1} = q^{-1/2} I_m + q^{1/2} I_{m-2}, \quad (5.11)$$

⁴One of these curves necessarily has odd subscript $2k + 1$, and we can make a small shift of all threads of this curve to the right from the threads of the second curve L_{2k} (or L_{2k+2} , depending on the situation). We find that there are no intersections of threads outside the region between terminating points of L_{2k+1} and L_{2k} (or L_{2k+2}), and in this domain, when closing the curves, there is produced exactly one intersection point.

where the basis of the recursion is given by $I_0 \equiv e^{l_Y/2} + e^{-l_Y/2}$ and $I_{-1} \equiv e^{l_Z/2} + e^{-l_Z/2}$. From (5.10), we have the exact equalities

$$I_m = q^{-m/2} I_0 (e^{m\ell_X/2} + e^{-m\ell_X/2}) - q^{-m/2-1/2} I_{-1} (e^{(m-1)\ell_X/2} + e^{-(m-1)\ell_X/2}), \quad (5.12)$$

or, equivalently,

$$I_m = q^{m/2} (e^{m\ell_X/2} + e^{-m\ell_X/2}) I_0 - q^{m/2+1/2} (e^{(m-1)\ell_X/2} + e^{-(m-1)\ell_X/2}) I_{-1}. \quad (5.13)$$

We wish to present the expression (5.12) or (5.13) in the form $e^{mH_1+H_0}$, where H_1, H_0 are Hermitian operators independent of m , so the proper length is then just H_1 (while H_0 introduces quantum corrections that do not affect the proper length limit but are important for ensuring the proper commutation relations).

Let us choose a compact domain \mathcal{F} in the function space $L^2(\mathbb{R})$ such that norms of all the operators in play are bounded for functions from this domain. Notice that neither term on the right-hand side in (5.12) is self-adjoint, but if one of these terms prevails (and it can be only the first term since we have an expression with coefficients that are all positive in the classical limit of the right-hand side, and hence we must have a positive left-hand side as well), then we can replace the total sum by this prevailing term and obtain an approximate equality

$$q^{-m/2} I_0 (e^{m\ell_X/2} + e^{-m\ell_X/2}) \sim q^{m/2} (e^{m\ell_X/2} + e^{-m\ell_X/2}) I_0$$

in this limit. By considering the spectral expansion with respect to the eigenfunctions of the operator ℓ_X , we immediately conclude that $I_0 \sim e^{-2\pi i \hbar \partial / \partial |\ell_X|}$ in this limit, and then

$$H_1 = |\ell_X|/2, \quad (5.14)$$

where the modulus has to be understood in terms of the spectral expansion: having a QMC operator \mathcal{L}_X which admits a spectral decomposition (see formulas (5.23)–(5.25) below) in functions $|\alpha_S\rangle$, we define the operator $|\ell_X|$ by its action on these functions: $|\ell_X| |\alpha_S\rangle = |S| |\alpha_S\rangle$.

A potentially problematic situation is when neither of the terms prevails and their difference remains finite as $m \rightarrow \infty$. This would correspond, as in the classical case discussed before, to a situation where the corresponding element fails to be hyperbolic. As we next show, we must obtain “long” curves when m goes to infinity, so this is also impossible in the quantum case; that is, for G_X, G_Y, G_Z , and \tilde{G}_Z from (2.15), (3.22) and representing $G_X = e^{\ell_X/2} + e^{-\ell_X/2}$, we must prove the operatorial inequality

$$G_Y e^{|\ell_X|/2} > q^{1/2} \tilde{G}_Z. \quad (5.15)$$

To prove this, first express $e^{\ell_X/2}$ through G_X . Taking the positive branch of the square root, we find from the left-hand side of (5.15) the expression

$$G_Y \left(\frac{G_X}{2} + \sqrt{\frac{G_X^2}{4} - 1} \right),$$

and since $G_Y G_X = q^{1/2} \tilde{G}_Z + q^{-1/2} G_Z$, we must compare two expressions $G_Y \sqrt{\frac{G_X^2}{4} - 1}$ and $(q^{1/2} \tilde{G}_Z - q^{-1/2} G_Z)/2$. Multiplying by the Hermitian conjugate on the right in both expressions, we eliminate the square root and arrive at Laurent polynomial expressions. After some simple algebra, we come to the inequality to be proved:

$$\frac{1}{2}(q \tilde{G}_Z G_Z + q^{-1} G_Z \tilde{G}_Z) > G_Y^2.$$

To see this, we expand the left-hand side

$$\begin{aligned} & (e^{-X-Z} + e^{X-Z} + e^{X+Z} + (q + q^{-1}) e^X) \\ & + \frac{q + q^{-1}}{2} (e^{-X-Y-Z} + e^{X+Y+Z} + e^Z + 2e^{-Z} + (q + q^{-1})) \\ & + \frac{q^2 + q^{-2}}{2} (e^{Y+Z} + e^{-Y-Z} + e^{-Y+Z} + (q + q^{-1}) e^{-Y}) + \frac{q^3 + q^{-3}}{2} e^Z, \end{aligned}$$

while the right-hand side is expressed as

$$G_Y^2 = e^{-X-Z} + e^{X-Z} + e^{X+Z} + 2 + (q + q^{-1})(e^X + e^{-Z}).$$

Subtracting this expression from the previous one, we obtain that this difference is

$$\frac{q^2 + q^{-2}}{2} G_X^2 + \frac{q + q^{-1}}{2} [e^{X+Y+Z} + e^{-X-Y-Z} - q - q^{-1}], \tag{5.16}$$

and both these terms are positive definite for $|q| = 1$.

The operatorial inequality (5.15) has therefore been established. This proves that the limit (5.4) exists and is well defined at rational points of the continued fraction expansion, and we next turn to the case of infinite continued fraction expansions, i.e., infinite sequences of elementary operators.

5.3.2 Mapping class group transformations and the unzipping procedure. We consider now an infinite continued fraction expansion $a_1, a_2, \dots, a_n, a_{n+1}, \dots$ extending the notation of (5.5). As we shall see, there is a corresponding sequence of unzippings of the freeway τ associated to a spine Γ of F as in Section 4.5 as well as an associated sequence of mapping class group elements, expressed as Dehn twists, which reduce an approximating multicurve to one of two possible graph simple curves.

Given the recursive representation (5.7) for the operator of a geodesic curve determined by a continued fraction expansion (5.5) and applying two (unitary) operators D_X and D_Y of the modular transformations of the form (3.12) that correspond to the respective Dehn twists along the corresponding closed curves γ_X and γ_Y (with the respective geodesic functions G_X and G_Y), we shall construct the sequence of zipping or unzipping transformations.

Definition 5.4. An approximating multicurve is determined by two nonnegative integers m_1 and m_2 . As in Figure 13, enumerate such a pair as a triple $(m_1, m_2, m_1 + m_2)$.

If $m_1 > m_2$, the action of the Dehn twist D_Y^{-1} along γ_Y (an *unzipping* transformation – the Dehn twist in the opposite direction) is

$$D_Y^{-1}: (m_1, m_2, m_1 + m_2) \mapsto (m_1 - m_2, m_2, m_1)$$

while if $m_1 < m_2$, we apply the unzipping transformation along the curve γ_X , which gives

$$D_X^{-1}: (m_1, m_2, m_1 + m_2) \mapsto (m_1, m_2 - m_1, m_2).^5$$

Given a continued fraction expansion (5.5), we construct the sequence of unzipping transformations

$$D_{(X \text{ or } Y)}^{-a_n} D_{(Y \text{ or } X)}^{-a_{n-1}} \cdots D_Y^{-a_3} D_X^{-a_2} D_Y^{-a_1}, \quad (5.17)$$

which, when applied to the approximating multicurve $(m_1, m_2, m_1 + m_2)$, reduces it either to $(1, 0, 1) \equiv \gamma_Y$ for n even or to $(0, 1, 1) \equiv \gamma_X$ for n odd.

Definition 5.5. Equivalently, we can consider the *zipping* procedure, that is, given a sequence of Dehn twists $D_Y^{a_1} D_X^{a_2} D_Y^{a_3} \cdots D_Y^{a_{n-1}} D_X^{a_n}$ applied to the curve γ_Y , we obtain the curve $(m_1, m_2, m_1 + m_2)$.

Considering the sequence (5.17) of quantum Dehn twist operators (3.12) and exploiting the quantum invariance from Lemma 3.9, we come to the main observation that having an involved expression for the proper limit (5.1) of a QMC operator constructed by the rules described in Lemmas 3.9 and 5.2 in terms of the elementary operators X, Y, Z , we may perform the sequence (5.17) of unzipping quantum modular transformations, which reduces this operator to a standard form of the quantum operator G_Y or G_X expressed through the new operators $X^{(n)}, Y^{(n)}, Z^{(n)}$ related to the initial operators by this sequence of quantum modular transformations. This is the operatorial statement of naturality of lengths under the mapping class group action.

It is intuitively natural to imagine that as the geodesic lengths must diverge as m_1, m_2 tend to infinity, we must eventually come to an asymptotic regime where all quantities $X^{(n)}, Y^{(n)}$ are large in the literal or operatorial sense for all sufficiently large n . The inexorability of the approach to this asymptotic regime is not obvious and is described in the next section.

5.3.3 Asymptotic regime. Let us recall the modular transformations for X, Y , and Z variables:

$$D_X^{-1}: (X, Y, Z) \mapsto (X + 2\phi^{\hbar}(Z), -Z, Y - 2\phi^{\hbar}(-Z)) \quad (5.18)$$

and

$$D_Y^{-1}: (X, Y, Z) \mapsto (-Z, Y - 2\phi^{\hbar}(-Z), X + 2\phi^{\hbar}(Z)). \quad (5.19)$$

⁵In terms of the symbolic dynamics of elementary operators G_X and G_Y , we can present this action as follows: $D_X^{-1}(G_X) = G_X$, $D_X^{-1}(G_X G_Y) = G_Y$ and $D_Y^{-1}(G_X G_Y) = G_X$, $D_Y^{-1}(G_Y) = G_Y$.

In terms of the quantities $U \equiv e^{X/2}$ and $V \equiv e^{-Y/2}$ in the case where $X+Y+Z=0$, we have

$$D_X^{-1} \begin{pmatrix} U \\ V \end{pmatrix} D_X = \begin{pmatrix} e^{X/2} + e^{-X/2-Y} \\ e^{-Y/2+X/2} \end{pmatrix} \equiv \begin{pmatrix} U + VU^{-1}V \\ q^{1/2}U^{-1}V \end{pmatrix}, \quad (5.20)$$

$$D_Y^{-1} \begin{pmatrix} U \\ V \end{pmatrix} D_Y = \begin{pmatrix} e^{X/2+Y/2} \\ e^{-Y/2} + e^{Y/2+X} \end{pmatrix} \equiv \begin{pmatrix} q^{1/2}UV^{-1} \\ V + UV^{-1}U \end{pmatrix}. \quad (5.21)$$

Worth mentioning is that since each operator Z_α is Hermitian, each exponential is positive definite, so we can always write, for instance, that $U + VU^{-1}V > U$ in the sense of spectral expansion: $\langle f|U + VU^{-1}V|f \rangle > \langle f|U|f \rangle$ for any function $f \in L^2(\mathbb{R})$.

Using now an alternating sequence of transformations (5.20), (5.21), we shall subsequently show that we attain the asymptotic regime of large positive X (large U) and large in absolute value negative Y (large V) starting from every pair of X and Y lying in a compact domain of the (X, Y) -plane in the classical case or acting within a compactum of test functions with bounded derivatives in the function space in the quantum operatorial case.

In this section, we verify that the asymptotic regime is attained for the distinguished sequence of modular transformations corresponding to the Fibonacci number sequence (golden mean), namely, for alternating D_X^{-1} and D_Y^{-1} . The proof in the general case is analogous although the structure is more involved, as described in the next section.

Given the sequence of transformations $D_X^{-1}D_Y^{-1} \dots D_X^{-1}D_Y^{-1} = (D_X^{-1}D_Y^{-1})^n$, we obtain

$$D_X^{-1}D_Y^{-1} \begin{pmatrix} U \\ V \end{pmatrix} \quad (5.22)$$

$$= \begin{pmatrix} UV^{-1}U + V \\ V^{1/2}(V^{-1}Uq^{-1/2} + U^{-1}Vq^{1/2})U(V^{-1}Uq^{-1/2} + U^{-1}Vq^{1/2})V^{1/2} + V^{1/2}U^{-1}V^{1/2} \end{pmatrix}.$$

In the classical case, the asymptotics is already clear from this formula; for the first entry in (5.22), we have $UV^{-1}U + V = U(V^{-1}U + U^{-1}V) > 2U$ as the expression in the parentheses has the form $e^S + e^{-S} \geq 2$ for any real S . The same logic applies to the second entry in (5.22), and we deduce that the classical part $V(U + U^{-1})$ has the same property.

The proof given below in the quantum case is more subtle as it needs a thorough operatorial analysis. Nevertheless, the estimates turn out to be close to those in the classical case, which we briefly discuss here: we must prove that a lower bound on the operatorial spectrum on a compactum in the function space diverges with n . This is a routine procedure, which uses that the action of operators X and Y in the basis of,

say, normalized Hermitian functions h_n has the form

$$\begin{aligned} X|h_n\rangle &= \sqrt{4\pi\hbar}(\sqrt{n+1}|h_{n+1}\rangle + \sqrt{n}|h_{n-1}\rangle) \\ Y|h_n\rangle &= \sqrt{4\pi\hbar}\frac{i}{2}(\sqrt{n+1}|h_{n+1}\rangle - \sqrt{n}|h_{n-1}\rangle). \end{aligned}$$

These operators “almost” commute in the domain of large n , which allows the combinatorics to be analyzed semiclassically.

In the quantum case, we recall the construction of quantum Dehn twists and their eigenfunctions from [18]. The generator of the Dehn twist D_X has the form

$$D_X = e^{q_1^2/2\pi i\hbar} F^\hbar(q_1 + p_1), \quad q_1 = X/2, \quad p_1 = 2\pi i\hbar\partial_X \quad (5.23)$$

and because it commutes with the geodesic length operator G_X , they share the common set of eigenfunctions

$$|\alpha_S\rangle = e^{-X^2/16\pi i\hbar} F^\hbar(S+X) F^\hbar(-S+X) \quad (5.24)$$

with the eigenvalues

$$G_X|\alpha_S\rangle = 2 \cosh(S/2)|\alpha_S\rangle; \quad D_X|\alpha_S\rangle = e^{S^2/2\pi i\hbar}|\alpha_S\rangle. \quad (5.25)$$

The functions $|\alpha_S\rangle$ constitute a complete set of functions in the sense that

$$\langle\alpha_T|\alpha_S\rangle = \delta(S-T)v^{-1}(S), \quad v(S) = 4 \sinh(\pi S) \sinh(\pi\hbar S)$$

and

$$\int_0^\infty v(S)dS |\alpha_S\rangle\langle\alpha_S| = \text{Id}$$

We now split the plane of the variables $(X, Y) = (X^{(0)}, Y^{(0)})$ into four sub-domains and consider the action of the Dehn twists D_X^{-1} and D_Y^{-1} in each sub-domain.

Domain I. $\{X^{(0)} > 0, Y^{(0)} > 0\} \cup \{X^{(0)} > 0, Y^{(0)} < 0 \text{ and } |X^{(0)}| > |Y^{(0)}|\}$:

$$\begin{aligned} D_X^{-1}(X^{(0)}, Y^{(0)}) &= (X^{(1)}, Y^{(1)}) = (X^{(0)}, X^{(0)} + Y^{(0)}) \in \text{Domain I}, \\ D_Y^{-1}(X^{(0)}, Y^{(0)}) &= (X^{(1)}, Y^{(1)}) = (X^{(0)} + Y^{(0)}, -2X^{(0)} - Y^{(0)}) \in \text{Domain II}; \end{aligned}$$

Domain II. $\{X^{(0)} > 0, Y^{(0)} < 0 \text{ and } |X^{(0)}| < |Y^{(0)}|\} \cup \{X^{(0)} < 0, Y^{(0)} < 0\}$:

$$\begin{aligned} D_X^{-1}(X^{(0)}, Y^{(0)}) &= (X^{(1)}, Y^{(1)}) = (-X^{(0)} - 2Y^{(0)}, X^{(0)} + Y^{(0)}) \in \text{Domain I}, \\ D_Y^{-1}(X^{(0)}, Y^{(0)}) &= (X^{(1)}, Y^{(1)}) = (X^{(0)} + Y^{(0)}, Y^{(0)}) \in \text{Domain II}; \end{aligned}$$

Domain IIIa. $\{X^{(0)} < 0, Y^{(0)} > 0 \text{ and } |X^{(0)}| < |Y^{(0)}|\}$:

$$\begin{aligned} D_X^{-1}(X^{(0)}, Y^{(0)}) &= (X^{(1)}, Y^{(1)}) = (X^{(0)}, X^{(0)} + Y^{(0)}) \in \text{Domain IIIa or IIIb}, \\ D_Y^{-1}(X^{(0)}, Y^{(0)}) &= (X^{(1)}, Y^{(1)}) = (X^{(0)} + Y^{(0)}, -2X^{(0)} - Y^{(0)}) \in \text{Domain I or II}; \end{aligned}$$

Domain IIIb. $\{X^{(0)} < 0, Y^{(0)} > 0 \text{ and } |X^{(0)}| > |Y^{(0)}|\}$:

$$D_X^{-1}(X^{(0)}, Y^{(0)}) = (X^{(1)}, Y^{(1)}) = (-X^{(0)} - 2Y^{(0)}, X^{(0)} + Y^{(0)}) \in \text{Domain I or II,}$$

$$D_Y^{-1}(X^{(0)}, Y^{(0)}) = (X^{(1)}, Y^{(1)}) = (X^{(0)} + Y^{(0)}, Y^{(0)}) \in \text{Domain IIIa or IIIb.}$$

We see that only Domain III is potentially problematic. This regime is however unstable: absolute values of X and Y variables decrease in this regime and they eventually leave the asymptotic regime as soon as we *remain* in Domain III; immediately upon leaving this domain, we come to Domains I and II and will never leave this three quarters of the (X, Y) -plane. The above considerations of U, V just demonstrate that even if we were initially in Domain III, we come to the nonasymptotic domain of bounded X and Y and then will leave this compactum moving toward asymptotic expansions in Domains I and II.

The asymptotic dynamics always takes place in the first three quarters of the (X, Y) -plane. Nevertheless, even this dynamics is rather involved. The stable regime corresponds to the case where we are in Domain II before applying one or several operators D_X^{-1} . The application of the first of these operators brings us to Domain I, and upon subsequent applications of the operators D_X^{-1} we remain in Domain I. Next, if we were in Domain I, then the very first application of the operator D_Y^{-1} brings us to Domain II, and we then remain in Domain II upon subsequent applications of D_Y^{-1} .

We turn now to actual geodesic lengths of curves or proper lengths of operators. If a sequence of unzipping transformations terminates, this means that we have a graph simple geodesic, which is either G_X if the last transformation was D_Y^{-1} or G_Y if the last transformation was D_X^{-1} . Considering the corresponding geodesic or proper lengths, we find that up to exponentially small corrections, the leading contributions in the above domains are

$$\text{p.l.}(\gamma_Y) = X + Y/2 \quad \text{in Domain I,} \tag{5.26}$$

$$\text{p.l.}(\gamma_X) = -Y - X/2 \quad \text{in Domain II} \tag{5.27}$$

(see expressions (2.15)).

Thus, although the transformation laws for the variables X, Y themselves do not possess the property of linearity with respect to the parameters a_i, a_j , when applying sequences of transformations $(D_X^{-1})_i^{a_i} \equiv D_X^{-a_i}$ or $(D_Y^{-1})^{a_j} \equiv D_Y^{-a_j}$, the proper lengths do possess this property! Namely, starting with variables $(X^{(0)}, Y^{(0)})$ lying in the corresponding domains and applying the sequences of transformations $D_X^{-a_i}$ or $D_Y^{-a_j}$, we obtain for the resulting proper lengths the following expressions:

$$\text{p.l.}(\gamma_{Y^{(i)}}) = -\frac{Y^{(0)}}{2} + a_i \left(-Y^{(0)} - \frac{X^{(0)}}{2} \right) \quad \text{for } X^{(0)}, Y^{(0)} \in \text{Domain II} \tag{5.28}$$

$$\text{p.l.}(\gamma_{X^{(j)}}) = \frac{X^{(0)}}{2} + a_j \left(X^{(0)} + \frac{Y^{(0)}}{2} \right) \quad \text{for } X^{(0)}, Y^{(0)} \in \text{Domain I} \tag{5.29}$$

Let us now explore the asymptotic formulas (5.26), (5.27) and (5.28), (5.29) first in the classical case to close this section, relegating the discussion of the quantum case to the next section.

Assume that we start from the variables $(X^{(0)}, Y^{(0)})$ in Domain I and have the corresponding initial length p.l. $(\gamma_{Y^{(0)}})$ from (5.26). Applying the transformation $D_Y^{-a_j}$, we obtain new variables $(X^{(j)}, Y^{(j)})$ and the new proper length p.l. $(\gamma_{X^{(j)}})$ (5.29) having form (5.27) in these new variables, which must now lie in Domain II. Note that explicitly

$$(X^{(j)}, Y^{(j)}) = (X^{(0)} + Y^{(0)} + (a_j - 1)(-2X^{(0)} - Y^{(0)}), -2X^{(0)} - Y^{(0)}).$$

We then apply the transformation $D_X^{-a_i}$ to obtain variables $(X^{(j,i)}, Y^{(j,i)})$, and the proper length p.l. $(\gamma_{Y^{(j,i)}})$ is expressed as in (5.28), where the term multiplied by a_i is none other than p.l. $(\gamma_{X^{(j)}})$ and the term $-\frac{Y^{(j)}}{2}$ is exactly p.l. $(\gamma_{Y^{(0)}})$. We thus find in the asymptotic regime that the corresponding lengths are related by *exactly the same* recurrence relation as for a graph length (the latter of which follows immediately from (5.7)):

$$\text{p.l.}(\gamma_{Y^{(j,i)}}) = a_i \text{p.l.}(\gamma_{X^{(j)}}) + \text{p.l.}(\gamma_{Y^{(0)}}), \quad (5.30)$$

$$\text{g.l.}(\gamma_{Y^{(j,i)}}) = a_i \text{g.l.}(\gamma_{X^{(j)}}) + \text{g.l.}(\gamma_{Y^{(0)}}). \quad (5.31)$$

It is then easy to conclude that the ratio of these two quantities has a definite limit as $i \rightarrow \infty$ for any sequence of numbers a_i . It is a standard estimate: given two numerical sequences (5.30) and (5.31) and denoting the relative error of their ratio as ε_i , i.e., at the i th step, the ratio is $S(1 + \varepsilon_i)$, where S is constant, for $a_{i+1} > 1$ at the $(i + 1)$ th step, we obtain $\varepsilon_{i+1} < \varepsilon_i / (a_{i+1} - 1/2)$, or if we have two coefficients $a_{i+1} = a_i = 1$, then $\varepsilon_{i+1} < \varepsilon_{i-1} / 1.5$. In general, for ε_i small enough, we always have $\varepsilon_{i+1} < \varepsilon_i$. This shows that the relative error decreases exponentially with the index i .

5.3.4 Quantum continued fraction expansion. Let us turn again to the sequence (5.17) of unzipping transformations. In order to obtain operatorial expressions, we consider the unitary operators D_X , D_Y and explicitly indicate the variables in which these operators are expressed, i.e., we write $D_{X^{(j)}} \equiv D_X(X^{(j)}, Y^{(j)})$ for the Dehn twist along γ_X at the $(j + 1)$ th step. Thus,

$$\begin{aligned} (X^{(n)}, Y^{(n)}) & \quad (5.32) \\ &= D_{X^{(n-1)}}^{-a_n} D_{Y^{(n-2)}}^{-a_{n-1}} \dots D_{X^{(1)}}^{-a_2} D_{Y^{(0)}}^{-a_1} (X^{(0)}, Y^{(0)}) D_{Y^{(0)}}^{a_1} D_{X^{(1)}}^{a_2} \dots D_{Y^{(n-2)}}^{a_{n-1}} D_{X^{(n-1)}}^{a_n}. \end{aligned}$$

In order to represent such long strings of operators in terms of the original operators $(X^{(0)}, Y^{(0)})$, we invert the dependence, i.e., we remember that, for instance,

$$D_{X^{(1)}}^{-a_2} = D_{Y^{(0)}}^{-a_1} D_{X^{(0)}}^{-a_2} D_{Y^{(0)}}^{a_1},$$

etc., which gives

$$(X^{(n)}, Y^{(n)}) \quad (5.33)$$

$$= D_{Y^{(0)}}^{-a_1} D_{X^{(0)}}^{-a_2} \dots D_{Y^{(0)}}^{-a_{n-1}} D_{X^{(0)}}^{-a_n} (X^{(0)}, Y^{(0)}) D_{X^{(0)}}^{a_n} D_{Y^{(0)}}^{a_{n-1}} \dots D_{X^{(0)}}^{a_2} D_{Y^{(0)}}^{a_1}.$$

We shall compute with bases of functions that are convenient in the asymptotic regime. Let

$$|f_{\mu,s}\rangle \equiv e^{i\mu(x-s)^2/2} \quad \mu, s \in \mathbb{R}. \quad (5.34)$$

These functions constitute a basis at each μ :

$$\langle f_{\mu,t} | f_{\mu,s} \rangle = \frac{2\pi}{\mu} \delta(s-t), \quad \int_{-\infty}^{\infty} ds |f_{\mu,s}\rangle \langle f_{\mu,s}| = \frac{2\pi}{\mu} \text{Id}. \quad (5.35)$$

For two arbitrary real numbers w and γ , we have

$$e^{iwx^2/2} |f_{\mu,s}\rangle = e^{i\mu ws^2/2} |f_{\mu+w, \frac{s}{1+w/\mu}}\rangle, \quad (5.36)$$

$$e^{i\gamma\partial_x^2/2} |f_{\mu,s}\rangle = \frac{1}{\sqrt{1+\gamma\mu}} |f_{\frac{1}{\gamma+1/\mu}, s}\rangle, \quad (5.37)$$

and

$$\langle f_{\mu,s} | x | f_{\mu,t} \rangle = \frac{2\pi}{i\mu^2} \delta'(s-t) + \frac{2\pi}{\mu} s \delta(s-t), \quad (5.38)$$

$$\langle f_{\mu,s} | \frac{1}{i} \partial_x | f_{\mu,t} \rangle = \frac{2\pi}{i\mu} \delta'(s-t). \quad (5.39)$$

We now define the dimensionless variable x and set

$$X |f_{\mu,s}\rangle = \sqrt{4\pi\hbar} x \cdot |f_{\mu,s}\rangle, \quad Y |f_{\mu,s}\rangle = \sqrt{4\pi\hbar} \frac{1}{i} \frac{\partial}{\partial x} \cdot |f_{\mu,s}\rangle. \quad (5.40)$$

The explicit formulas for the operators D_X and D_Y acting on $|f_{\mu,s}\rangle$ in the asymptotic regime are

$$D_X^{a_i} |f_{\mu,s}\rangle = e^{i(a_i-1)x^2/2} e^{-i\partial_x^2} e^{ix^2/2} |f_{\mu,s}\rangle, \quad (5.41)$$

$$D_Y^{a_j} |f_{\mu,s}\rangle = e^{i(a_j-1)\partial_x^2/2} e^{-ix^2} e^{i\partial_x^2/2} |f_{\mu,s}\rangle. \quad (5.42)$$

In order to establish the required recurrence relation, we must compare matrix elements of the three consecutive length operators in the corresponding operatorial decompositions:

$$A_{st}^{(0)} = \langle f_{\mu,s} | X + \frac{Y}{2} | f_{\mu,t} \rangle,$$

$$A_{st}^{(j)} = \langle f_{\mu,s} | D_Y^{-a_j} | -Y - \frac{X}{2} | D_Y^{a_j} f_{\mu,t} \rangle,$$

$$A_{st}^{(j,i)} = \langle f_{\mu,s} | D_Y^{-a_j} D_X^{-a_i} | X + \frac{Y}{2} | D_X^{a_i} D_Y^{a_j} f_{\mu,t} \rangle.$$

Now, using formulas (5.36)–(5.42), it is straightforward to show that

$$A_{st}^{(j,i)} = A_{st}^{(0)} + a_i A_{st}^{(j)}, \quad (5.43)$$

for all s, t , i.e., we again attain the recurrence relation (5.30) in the asymptotic regime, but now for the matrix elements of the operators of the quantum proper lengths. Estimates show that the corrections due to both the (operatorial) deviations from the asymptotic regime and the error parameters ε_i (as for (5.30), (5.31)) decrease exponentially with the index i , so the limit (5.4) exists in a weak operatorial sense. We conclude that ratios (5.4) define a weakly continuous family of operators parameterized by projective transverse measures on the freeway associated to a spine of the once-punctured torus. This completes the proof of Theorem 5.1.

6 Conclusion

We hope to have added to the mathematical foundation and general understanding of the quantization of Teichmüller space and its geometric underpinnings in the first several sections of this exposition. We also hope that the survey given here of train tracks and their extensions might be useful.

The quantization of Thurston’s boundary in general seems to be a substantial project, which we have only just begun here with the quantization of continued fractions. First of all, one would like a better understanding of the operators we have constructed, for instance, an intrinsic characterization or an explicit calculational framework for them. At the same time, our current constructions depend upon a choice of spine, and there would seem to be a more invariant version of the theory, where the choice of spine is dictated by the combinatorics of the cell decomposition of Teichmüller space; the calculations in this chapter apply to each such spine (since there is a combinatorially unique cubic one) for the once-punctured torus.

Second of all, the quantization of Thurston’s boundary for higher-genus or multiply-punctured surfaces may be approachable using the improved quantum ordering. Namely, in any fixed spine of the surface, there is a fixed finite family of “edge-simple” closed edge-paths which by definition never twice traverse the same oriented edge. It is elementary to see that any closed edge-path on Γ may be written non-uniquely as a concatenation of edge-simple paths, where the particular concatenation depends upon a starting point. (Edge-simple paths were studied as “canonical curves” on train tracks in [26]; they contain the extreme points of the polyhedron of projective measures on the track.) It follows that an arbitrary leaf of a measured foliation carried by a freeway can be written as a concatenation of paths from this finite collection of edge-simple paths. The corresponding quantum operatorial statement results from the improved quantum ordering described here. Thus, whereas the quantization of the once-punctured torus devolved, in effect, to an analysis of two-letter words, the quantization of Thurston’s

boundary sphere in general may depend upon an analysis of words comprised of letters which are edge-simple paths.

One appealing long-term goal would be to discover the Thurston classification already on the operatorial level, for instance, with the dilatation in the pseudo-Anosov case explicitly computable from the mapping class group operator or from the invariant projective foliation operator.

Another intriguing aspect involves generalizations of graph length functions insofar as the proof of Theorem 5.1 holds taking as graph length any continuous positive definite function which is homogeneous of degree one. A natural choice of such a function is induced by the geodesic length of the corresponding geodesic curve taken for a fixed basepoint in Teichmüller space on its fixed spine, for instance, vanishing shear coordinates on the usual spine in the once-punctured torus. What sort of regularity (e.g., piecewise smoothness) is achieved in the operators corresponding to points of Thurston's boundary under such "gauge fixing"?

Also worth mentioning are very recent advances in the description of quantum $sl(n, \mathbb{R})$ connections [11], where one finds an improved quantum ordering in a more complicated higher-dimensional setting.

Appendix A. Combinatorial proof of Theorem 5.1

In this appendix, we give a complementary, combinatorial proof of the classical Theorem 5.1 using the recurrence relation (5.7). At the present state of understanding, the proof applies only to the classical case as we cannot control the quantum ordering.

Let us recall the structure of the matrix product (2.5). It is a sequence of matrices L_Z, R_Z with different Z . It can be always segregated into clusters of matrices

$$L_{\bar{z}} \equiv L_{Z_{i+s}} L_{Z_{i+s-1}} \dots L_{Z_i}$$

and

$$R_{\bar{z}} \equiv R_{Z_{j+k}} R_{Z_{j+k-1}} \dots R_{Z_j}.$$

The periodic extension of expression (5.7) is always an alternating sequence of matrices $L_{\bar{z}}$ and $R_{\bar{z}}$:

$$P_{Z_1, \dots, Z_n} = \dots (L_{\bar{z}_s} R_{\bar{z}_{s-1}}) \dots (L_{\bar{z}_1} R_{\bar{z}_1}) \dots$$

First note that it is impossible to have arbitrarily long sequences of only left or right matrices for a given graph: the maximum length is restricted to be less or equal the maximum graph length of geodesics around holes. This means that the length of a single cluster for a given graph is always bounded once the topology is fixed.

One can directly calculate the product $(L_{\bar{Z}_1} R_{\bar{Z}_2})$ for $L_{\bar{Z}_1} = L_{Z_1} \dots L_{Z_m}$ and $R_{\bar{Z}_2} = R_{Z_{m+1}} \dots R_{Z_{m+k}}$:

$$(L_{\bar{Z}_1} R_{\bar{Z}_2}) = A s_1^+ s_2^+ + B(s_1^- s_2^- + S_1 s_2^- + S_1 S_2) + D(s_1^+ s_2^- + s_1^+ S_2) + P S_1 s_2^+, \tag{A.1}$$

where S_j, s_j^\pm , for $j = 1, 2$, are the following coefficient functions:

$$s_1^\pm = e^{\pm \sum_{i=1}^m Z_i/2}, \quad s_2^\pm = e^{\pm \sum_{j=m+1}^{m+k} Z_j/2}$$

$$S_1 = \sum_{q=2}^m e^{+\sum_{r=1}^{q-1} Z_r/2 - \sum_{r=q}^m Z_r/2}, \quad S_2 = \sum_{q=m+2}^{m+k} e^{-\sum_{r=m+1}^{q-1} Z_r/2 + \sum_{r=q}^{m+k} Z_r/2},$$

and A, B, D, P are the special (2×2) -matrices (“letters”):

$$A = \begin{pmatrix} +1 & 0 \\ -1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 0 & +1 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & -1 \\ 0 & +1 \end{pmatrix}, \quad P = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}. \tag{A.2}$$

These letters possess interesting multiplication properties which are summarized in the next lemma, whose proof is a routine calculation.

Lemma A.1 (The alphabet lemma). *The multiplication table of letters (A.2) reads:*

	A	B	D	P	
A	A	0	D	0	
B	P	B	B	P	,
D	A	D	D	A	
P	P	0	B	0	

(A.3)

so the trace of any product of these matrices is either unity or zero. In the product of t matrices of form (A.1), the only monomials that survive are

$$[(A + D)^{i_\alpha} D] B^{j_\beta} (B + P) \dots [(A + D)^{i_\rho} D] B^{j_\omega} (B + P) \text{ and } (A + D)^t, B^t. \tag{A.4}$$

The main point is that almost all cancellations of letters in long words are due to the local multiplication rules (A.3). This means that, having a long sequence of letters, say, L_{I+N} from Lemma 5.2, we can split it into pieces depending on sequences of letters $L_I, L_{I-1}, \tilde{L}_I$, and \tilde{L}_{I-1} , where the index I is also assumed to be big enough. That is, let $L_{I+N} = L_I L_{I-1} \tilde{L}_I L_I \dots L_I$ comprise p_N entries L_I and \tilde{L}_I and q_N entries L_{I-1} and \tilde{L}_{I-1} . We have then the following estimate:⁶

$$|\log \text{tr } L_{I+N} - p_N \log \text{tr } L_I - q_N \log \text{tr } L_{I-1}| < C (q_N + p_N), \tag{A.5}$$

⁶This estimate also follows from the properties of long geodesic lines in hyperbolic geometry: for two lines of large lengths L_1 and L_2 intersecting at angle α , the length L_3 of the third side of the resulting triangle is $L_1 + L_2 + \log((1 - \cos \alpha)/2) + O(1/L)$. This also shows that our estimate is very rough.

where the constant C depends only on the Teichmüller space coordinates Z_α and on the genus and the number of holes of the Riemann surface, and we have also used (5.8). The ratio of the coefficients is given by the continued fraction

$$q_N/p_N = \frac{1}{a_{I+1} + \frac{1}{a_{I+2} + \dots + \frac{1}{a_{I+N-1} + \frac{1}{a_{I+N}}}}}, \tag{A.6}$$

and also has a definite limit as $N \rightarrow \infty$. Now the estimate follows: up to exponential corrections, $\text{p.l.}(L)$ coincides with the $\log \text{tr } L$, so for any $\varepsilon > 0$, let us choose the index I such that $\varepsilon \text{p.l.}(L_{I-1})/2 > C$ and $\varepsilon \text{p.l.}(L_I)/2 > C$. Thus,

$$\frac{\text{p.l.}(L)(L_{I+N})}{\text{g.l.}(L_{I+N})} = (1 + O(\varepsilon/2)) \frac{p_N \text{p.l.}(L)(L_I) + q_N \text{p.l.}(L)(L_{I-1})}{p_N \text{g.l.}(L_I) + q_N \text{g.l.}(L_{I-1})}, \tag{A.7}$$

and because the ratio q_N/p_N has a definite limit as $N \rightarrow \infty$, there exists N_0 such that the relative error of this ratio times the sum of ratios of proper and graph lengths of L_I and L_{I-1} will not exceed $\varepsilon/2$. Thus, the collective relative error for such fixed I and for all $N > N_0$ is less than ε , proving the theorem.

Appendix B. Degeneracy of the Poisson structure

We shall explicitly calculate the degeneracy of the Poisson brackets (2.9) for a special graph and choose the graph whose “building blocks” are depicted in Figure 15. Namely, we have a line tree subgraph comprising edges X_i with attached subgraphs as in Figure 15 (a) and 15 (b). Attaching a subgraph of type a corresponds to adding a handle (increasing g by unity) while a subgraph of type b corresponds to adding a hole (increasing s by unity). We shall assume that $2g + 2s > 5$ to avoid the once-punctured torus, which is already handled separately in Section 3.

For the variables A, B, C, D, E in Figure 15 (a) we have the Poisson bracket (sub)matrix

	A_i	B_i	C_i	D_i	E_i
A_i	0	1	-1	0	0
B_i	-1	0	1	1	-1
C_i	1	-1	0	1	-1
D_i	0	-1	-1	0	2
E_i	0	1	1	-2	0

,

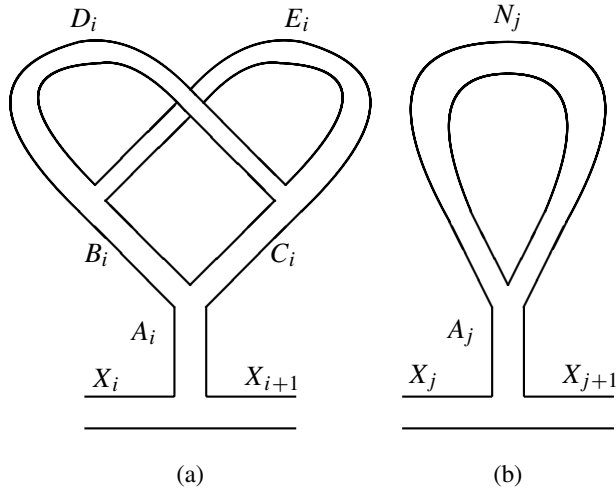


Figure 15. Building blocks.

where the entries are the Poisson brackets between the corresponding variables. Adding the last row to the next-to-the-last row as well as adding the last column to the next-to-the-last column, then adding the third row to the second row as well as the third column to the second column, we obtain the matrix

$$\begin{pmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -2 \\ 1 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 2 & 1 & -2 & 0 \end{pmatrix},$$

which obviously has rank four and can be further reduced (without adding the first column or row to any other) to the form

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & +1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & +2 \\ 0 & 0 & 0 & -2 & 0 \end{pmatrix}.$$

Thus, erasing all columns and rows corresponding to the variables B_i , C_i , D_i , and E_i leaves invariant the rank of the Poisson bracket matrix.

Adjoining the subgraph in Figure 15 (b) creates exactly one degeneracy as the variable N_j Poisson commutes with everything (as it must be when adding a hole).

It remains only to calculate the rank of the matrix corresponding to a tree graph with edges X_i and A_i remaining after erasing all B -, C -, D -, E -, and N -variable rows and columns. The corresponding Poisson bracket matrix has dimension $2g + 2s - 5 > 0$ and the simple block-diagonal form

0	1	-1	0	0	.	.	.
-1	0	1	0	0			
1	-1	0	1	-1			
	0	-1	0	1	0	0	
	0	1	-1	0	1	-1	.
			0	-1	0	1	
			0	1	-1	0	⋱
						⋱	⋱

Adding each even-index row to its predecessor as well as adding each even-index column to its predecessor, this reduces to the matrix whose only nonzero elements are $+1$ on the main super-diagonal and -1 on the main sub-diagonal. Since this matrix has corank one, the discussion is complete.

References

- [1] F. Bonahon, Bouts des variétés hyperboliques de dimension 3. *Ann. of Math.* 124 (1986), 441–479. [610](#)
- [2] F. Bonahon, Shearing hyperbolic surfaces, bending pleated surfaces and Thurston’s symplectic form. *Ann. Fac. Sci. Toulouse Math.* (6) 5 (1996), 233–297. [584](#), [587](#)
- [3] F. Bonahon and Y. Sözen, The Weil-Petersson and Thurston symplectic forms. *Duke Math J.* 108 (2001), 581–597. [589](#), [610](#)
- [4] A. Bondal, A symplectic groupoid of triangular bilinear forms and the braid groups. *Izv. Ross. Akad. Nauk Ser. Mat.* 68 (4) (2004), 19–74; English transl. *Izv. Math.* 68 (4) (2004), 659–708. [593](#), [603](#)
- [5] D. Bullock and J. H. Przytycki, Multiplicative structure of Kauffman bracket skein module quantizations. *Proc. Am. Math. Soc.* 128 (3) (2000), 923–931. [603](#)
- [6] L. Chekhov and V. Fock, Talk given at *St. Petersburg Meeting on Selected Topics in Mathematical Physics*, LOMI, 26–29 May, 1997; A quantum Teichmüller space. *Theor. Math. Phys.* 120 (1999), 1245–1259; Quantum mapping class group, pentagon relation, and geodesics. *Proc. Steklov Math. Inst.* 226 (1999), 149–163. [580](#), [581](#), [582](#), [583](#), [587](#), [595](#), [602](#), [603](#)

- [7] D. B. A. 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. [616](#)
- [8] L. D. Faddeev, Discrete Heisenberg–Weyl group and modular group. *Lett. Math. Phys.* 34 (1995), 249–254. [581](#), [595](#)
- [9] A. Fathi, F. Laudenbach, and V. Poénaru (eds.), Travaux de Thurston sur les surfaces. *Astérisque* 66–67 (1979). [608](#), [609](#), [616](#)
- [10] V. V. Fock, Description of moduli space of projective structures via fat graphs. Preprint, 1994; arXiv:hep-th/9312193. [588](#), [589](#)
- [11] V. V. Fock and A. B. Goncharov, Moduli spaces of local systems and higher Teichmüller theory. *Publ. Math. Inst. Hautes Études Sci.* 103 (2006), 1–211; Cluster ensembles, quantization and the dilogarithm. Preprint, 2004; arXiv:math.AG/0311245. [639](#)
- [12] V. V. Fock and A. A. Rosly, Poisson structures on moduli of flat connections on Riemann surfaces and the r -matrix. In *Moscow Seminar in Mathematical Physics*, Amer. Math. Soc. Transl. Ser. 2 191, Amer. Math. Soc., Providence, RI, 1999, 67–86. [580](#)
- [13] V. V. Fock and A. A. Rosly, Flat connections and polytopes. *Theoret. and Math. Phys.* 95 (1993), 526–535. [580](#)
- [14] W. M. Goldman, Invariant functions on Lie groups and Hamiltonian flows of surface group representations. *Invent. Math.* 85 (1986), 263–302. [591](#), [592](#)
- [15] M. Havlíček, A. V. Klimyk, and S. Pošta, Representations of the cyclically symmetric q -deformed algebra $so_q(3)$. *J. Math. Phys.* 40 (4) (1999), 2135–2161. [593](#), [601](#)
- [16] R. M. Kashaev, Quantization of Teichmüller spaces and the quantum dilogarithm. *Lett. Math. Phys.* 43 (2) (1998), 105–115. [580](#), [581](#), [596](#)
- [17] R. M. Kashaev, Liouville central charge in quantum Teichmüller theory. *Proc. Steklov Math. Inst.* 226 (1999), 62–70. [581](#)
- [18] R. M. Kashaev, On the spectrum of Dehn twists in quantum Teichmüller theory. In *Physics and Combinatorics* (Nagoya 2000), World Sci. Publ., River Edge, N.J., 2001, 63–81. [596](#), [599](#), [634](#)
- [19] M. Kontsevich, Intersection theory on the moduli space of curves and the matrix Airy function. *Comm. Math. Phys.* 147 (1992), 1–23.
- [20] J. E. Nelson and T. Regge, $2 + 1$ quantum gravity. *Phys. Lett. B* 272 (1991), 213–216; Invariants of $2 + 1$ gravity. *Comm. Math. Phys.* 155 (1993), 561–568. [603](#)
- [21] J. E. Nelson and T. Regge, Homotopy groups and $(2+1)$ -dimensional quantum gravity. *Nuclear Phys. B* 328 (1989), 190–199; J. E. Nelson, T. Regge, and F. Zertuche, Homotopy groups and $(2 + 1)$ -dimensional quantum de Sitter gravity. *Nuclear Phys. B* 339 (1990), 516–532. [593](#), [603](#)
- [22] A. Papadopoulos and R. C. Penner, Enumerating pseudo-Anosov conjugacy classes. *Pacific J. Math* 142 (1990), 159–173. [617](#)
- [23] A. Papadopoulos and R. C. Penner, The Weil-Petersson symplectic structure at Thurston’s boundary. *Trans. Amer. Math. Soc.* 335 (1993), 891–904. [610](#), [618](#), [620](#)
- [24] R. C. Penner and J. L. Harer, *Combinatorics of train tracks*. Ann. of Math. Stud. 125, Princeton University Press, Princeton, N.J., 1992. [609](#), [612](#), [613](#), [614](#), [615](#)

- [25] R. C. Penner, The decorated Teichmüller space of Riemann surfaces. *Comm. Math. Phys.* 113 (1988), 299–339. [581](#), [584](#), [588](#)
- [26] R. C. Penner, The action of the mapping class group on isotopy classes of curves and arcs in surfaces. Thesis, Massachusetts Institute of Technology, 1982. [609](#), [638](#)
- [27] R. C. Penner, Weil-Petersson volumes. *J. Differential Geom.* 35 (1992), 559–608. [581](#), [589](#), [590](#)
- [28] R. C. Penner, Universal constructions in Teichmüller theory. *Adv. Math.* 98 (1993), 143–215. [581](#), [589](#), [590](#)
- [29] R. C. Penner, A construction of pseudo-Anosov homeomorphisms. *Proc. Amer. Math. Soc.* 104 (1988), 1–19. [584](#), [591](#), [617](#)
- [30] G. Rauzy, Echanges d’intervalles et transformations induites. *Acta Arith.* 34 (1979), 325–328. [626](#)
- [31] M. Reed and B. Simon, *Methods of modern mathematical physics*. I. Functional analysis, second edition, Academic Press, Inc., New York 1980. [596](#)
- [32] K. Strebel, *Quadratic differentials*. *Ergeb. Math. Grenzgeb.* (3) 5, Springer-Verlag, Berlin 1984. [584](#)
- [33] J. Teschner, From Liouville theory to the quantum geometry of Riemann surfaces. Preprint, 2003; arXiv:hep-th/0308031. [581](#)
- [34] J. Teschner, An analog of a modular functor from quantized Teichmüller theory. *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 685–760. [581](#)
- [35] W. P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc.* 19(1988), 417–431. [609](#), [610](#), [612](#), [616](#), [617](#)
- [36] W. P. Thurston, Minimal stretch maps between hyperbolic surfaces. Preprint, 1984; arXiv:math.GT/9801039. [584](#), [587](#)
- [37] V. G. Turaev, Skein quantization of Poisson algebras of loops on surfaces. *Ann. Sci. École Norm. Sup.* (4) 24 (1991), 635–704. [600](#), [602](#)
- [38] M. Ugaglia, On a Poisson structure on the space of Stokes matrices. *Internat. Math. Res. Notices* 1999 (9) (1999), 473–493. [593](#)
- [39] E. Verlinde and H. Verlinde, Conformal field theory and geometric quantization. In *Superstrings ’89* (Trieste, 1989), World Scientific, River Edge, N.J., 1990, 422–449. [580](#), [581](#)
- [40] O. Ya. Viro, Lectures on combinatorial presentations of manifolds. In *Differential Geometry and Topology* (Alghero, 1992), World Scientific, River Edge, N.J., 1993, 244–264.
- [41] S. Wolpert, On the symplectic geometry of deformations of a hyperbolic surface. *Ann. of Math.* 117 (1983), 207–234. [589](#)

Dual Teichmüller and lamination spaces

Vladimir V. Fock and Alexander B. Goncharov

Contents

1	Introduction	647
2	Surfaces and triangulations	649
3	Laminations	652
	3.1 Unbounded measured laminations	652
	3.2 Bounded measured laminations	656
4	Teichmüller spaces	659
	4.1 Teichmüller space of surfaces with holes $\mathcal{T}^x(S)$	660
	4.2 Teichmüller space of decorated surfaces $\mathcal{T}^a(S)$	664
5	Tropicalisation	668
6	Poisson and degenerate symplectic structures	669
7	Canonical pairing	671
	7.1 Additive canonical pairing	672
	7.2 Multiplicative canonical pairing	673
	7.3 Proofs of main properties	675
	Appendix A. Combinatorial description of $\mathcal{D}(S)$	676
	Appendix B. Markov numbers	679
	References	682

1 Introduction

The Teichmüller space of a surface S is the space of complex structures on S modulo diffeomorphisms isotopic to the identity. For open surfaces this definition should be made more precise since we need to specify the behaviour of the complex structure at the boundary of the surface thus giving different versions of these spaces.

The space of measured laminations for a surface S is defined by William Thurston [25]. A dense subset of the space of laminations is formed by the ones with finite number of leaves – just collections of weighted closed curves without intersections and self intersections. For open surfaces there are different versions of such spaces.

All versions of both Teichmüller and lamination spaces share the following properties:

- (1) They are topologically trivial.
- (2) Isotopy classes of embeddings of surfaces induce canonical maps between these spaces (forward or backward).
- (3) There exist canonical compactifications of Teichmüller spaces by the quotients of lamination spaces by multiplication by positive real numbers.
- (4) There exists a canonical Poisson or degenerate symplectic structure (depending on a version) on these spaces.
- (5) A lamination with integral weights provides a function on Teichmüller space. Such functions form a basis in the algebra of regular functions on Teichmüller space with respect to a certain algebraic structure on it.

The original approach to Teichmüller spaces used extensively functional analysis and was highly nonconstructive. The aim of this chapter is to give a description of all these spaces for open surfaces using elementary algebra and geometry and make all the properties obvious.

We also consider generalisations of Riemann surfaces – the so-called *ciliated surfaces* – surfaces with a distinguished set of points on the boundary. Though at the beginning it requires a little more elaborate definitions, it provides us with more simple examples of Teichmüller spaces containing most features of the general cases. Namely, the configurations of points on the boundary of a disk or a cylinder fit perfectly into the picture.

We included a somehow more technical Section 7 about the canonical pairing between lamination and Teichmüller spaces in the case of ciliated surfaces.

The present exposition borrows a lot from [4] and Section 11 of [6]. We tried to restrict ourselves to the purely geometric part of the picture and make it as elementary as possible. It can be used as an elementary introduction to more algebraic subjects, such as classical and quantum cluster varieties. To give the reader a further perspective, let us briefly mention some other sources and related topics.

Explicit coordinate description of the Teichmüller spaces for an open surface goes back to William Thurston [25] and Robert Penner [21]. This subject was developed further in [4]. Quantum Teichmüller spaces were constructed in [5] and independently by Rinat Kashaev in [17]. Jörg Teschner [24] proved that the quantum Teichmüller space acts on the space of conformal blocks of the Liouville conformal field theory, as was conjectured in [4], [5]. In [6] the higher Teichmüller spaces were defined, and it was shown that they parametrise certain discrete, faithful representations of the fundamental group of the surface in a split real simple Lie group G of higher rank. These spaces are closely related to the ones studied by Nigel Hitchin [16]. In particular, for $G = SL_3(\mathbb{R})$ the corresponding higher Teichmüller space turns out [8] to coincide with the space of real projective structures on S studied by William Goldman and Suhyong Choi [14], [3]. In [6], [7] it was shown that the \mathcal{A} and \mathcal{X} versions of the Teichmüller and lamination spaces can be obtained as the positive real and tropical points of certain *cluster \mathcal{A} - and \mathcal{X} -varieties*. The closely related objects, cluster algebras, were introduced by Sergey Fomin and Andrei Zelevinsky [11], and

studied by them and Arkady Berenstein, Mikhail Gekhtman, Mikhail Shapiro, Alek Vainshtein ([12] and references therein). As a result of these developments, the cluster theory was enriched by new examples as well as new features (such as duality, Poisson structure and quantisation, relations to algebraic K -theory and to the dilogarithm).

In particular the canonical pairings from Section 6, in the special case of a ciliated disc, can be viewed as the canonical pairings for cluster \mathcal{A} and \mathcal{X} -varieties of finite type A_n , predicted by the general duality conjectures [7].

We are very grateful to Athanase Papadopoulos and Vladimir Turaev, who suggested to write this chapter for the Handbook on Teichmüller spaces. The first author is grateful to the Fields Institute and Brown University, where this exposition was mainly written, for hospitality. The authors were supported by the NSF grants DMS-0400449 (the second author) and CRDF 2622;2660 (the first author). We are very indebted also to Andrei Levin, Yuri Neretin, Stepan Orevkov, Athanase Papadopoulos, Robert Penner and especially to Aleksei Rosly for very valuable discussions and for Guillaume Théret for carefully reading the text and correcting many typos and mistakes.

2 Surfaces and triangulations

In this section we shall briefly recall basic facts about triangulations of surfaces.

A *ciliated surface* is a compact oriented surface with boundary and with a finite set of marked points on the boundary called *cilia*.

A boundary component without cilia is called a *hole*.

A *triangulation* Γ of a ciliated surface is a decomposition of the surface with contracted holes into triangles such that every vertex of a triangle is either a cilium or a shrunk hole.

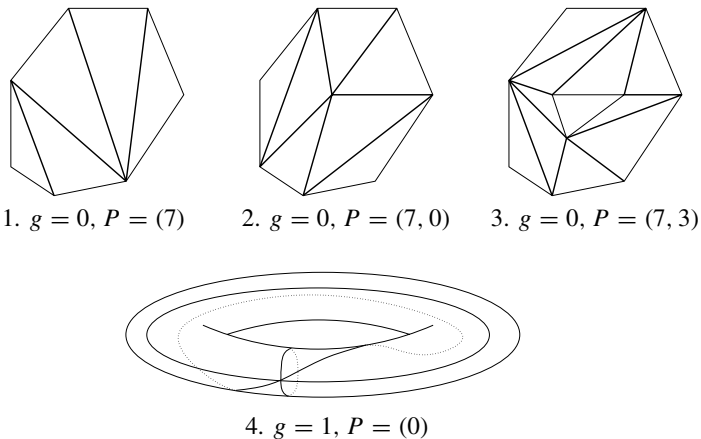


Figure 1. Examples of triangulations of ciliated surfaces. Internal edges are thick.

The edges of the triangulation belonging to the boundary of the surface are called *external*. Other edges are called *internal*. Denote by $F(\Gamma)$, $E(\Gamma)$, $E_0(\Gamma)$, $V(\Gamma)$ the set of triangles, edges, external edges and vertices of the triangulation, respectively.

Topologically a ciliated surface is defined by its genus g and a finite collection of positive integers $P = (p_1, \dots, p_s)$, where s is the number of boundary components and p_i is the number of cilia on the i -th component. There is no canonical order on the set of boundary components. Denote the number of holes by h , the total number of cilia by $c = \sum_i p_i$, and the number of internal edges by n .

Observe, that the numbers of faces, vertices and edges, both external and internal, are determined by the topology:

- (1) $\sharp V(\Gamma) = h + c$;
- (2) $\sharp E_0(\Gamma) = c$;
- (3) $\sharp E(\Gamma) = 6g - 6 + 3s + 2c$;
- (4) $n = 6g - 6 + 3s + c$;
- (5) $\sharp F(\Gamma) = 4g - 4 + 2s + c$.

The first two statements are obvious. The others follow from the expression for the Euler characteristic of the surface,

$$\sharp F(\Gamma) - \sharp E(\Gamma) + \sharp V(\Gamma) = 2 - 2g + h - s,$$

and the expression

$$3\sharp F(\Gamma) = 2\sharp E(\Gamma) - \sharp E_0(\Gamma)$$

of the fact that each triangle has three sides.

The topology of the triangulation can be encoded into a skew-symmetric matrix $\varepsilon^{\alpha\beta}$, where $\alpha, \beta \in E(\Gamma)$, defined as

$$\varepsilon^{\alpha\beta} = \sum_{i \in F(\Gamma)} \langle \alpha, i, \beta \rangle,$$

where $\langle \alpha, i, \beta \rangle$ is equal to $+1$ (resp. -1) if both α and β are sides of the triangle i and α is in the counterclockwise (resp. clockwise) direction from β with respect to their common vertex. Otherwise $\langle \alpha, i, \beta \rangle$ is equal to zero. The entries of the matrix $\varepsilon^{\alpha\beta}$ might have five possible values: $0, \pm 1, \pm 2$.

The number of triangulations of a given ciliated surface is infinite except for $g = 0$ and for either $P = (k)$ or $P = (k, 0)$. However one triangulation can be obtained from another one by a sequence of elementary moves called *flips* or *Whitehead moves*. One triangulation is a flip of another if it is obtained by removing one internal edge and replacing it by another diagonal of the arising quadrilateral (Figure 2). A flip can be done with any internal edge unless it belongs to a unique triangle of the triangulation.

Observe that there is a canonical correspondence between the edges of a triangulation and the edges of a flipped one. We shall use the notation α' for the new edge corresponding to the old edge α . However one should be aware that compositions

of a sequence of flips may restore the original triangulations but a nontrivial correspondence between the edges. An example of this phenomenon can be illustrated by Figure 16 where the composition of five flips does not change the triangulation of the pentagon, but the canonical correspondence interchanges the diagonals.

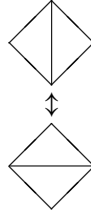


Figure 2. A flip.

It is a simple exercise to check that the matrix $\varepsilon^{\alpha'\beta'}$ encoding the combinatorics of a triangulation after a flip in the edge γ is given by the formula (unless one of the triangle of the triangulation is glued to itself):

$$\varepsilon^{\alpha'\beta'} = \begin{cases} -\varepsilon^{\alpha\beta} & \text{if } \alpha = \gamma \text{ or } \beta = \gamma, \\ \varepsilon^{\alpha\beta} + \frac{1}{2}(\varepsilon^{\alpha\gamma}|\varepsilon^{\gamma\beta}| + |\varepsilon^{\alpha\gamma}|\varepsilon^{\gamma\beta}) & \text{otherwise.} \end{cases} \quad (2.1)$$

To give some definitions we shall need to fix the orientation of the holes of the surface. By that we mean fixing orientation of the boundary components in a way not necessarily induced by the orientation of the surface. We say that the orientation of a hole is *positive* (resp. *negative*) if it agrees (resp. disagrees) with the orientation of S .

For a given boundary component ρ with *even* number of cilia introduce a function $\zeta_\rho: E_0(\Gamma) \rightarrow \mathbb{Z}/2$. First assign numbers to edges of Γ by the following rule. If ρ is a hole assign 1 to the corresponding vertex. If ρ is a boundary component with positive even number of cilia assign $1/2$ to vertices corresponding to every second cilium of ρ and $-1/2$ to the remained vertices corresponding to ρ . Assign then 0 to the vertices corresponding to other boundary components. The value $\zeta_\rho(\alpha)$ is just the sum of numbers assigned to the ends of the edge α .

The *mapping class group* $\mathcal{D}(S)$ of a ciliated surface S is the group of connected components of the diffeomorphisms of S preserving the set of cilia and the orientation of S . In our examples of Figure 1 the mappings class groups are, respectively $\mathbb{Z}/7\mathbb{Z}$, $\mathbb{Z}/7\mathbb{Z}$, $\mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/7\mathbb{Z}$ and $\text{PSL}(2, \mathbb{Z})$. Observe that the mapping class group is finite if and only if the number of triangulations is finite.

In 7.3 we give a combinatorial description of the mapping class group in terms of the surface triangulations.

3 Laminations

Taking into account that the reader may be unfamiliar with Thurston's notion of a measured lamination [25], we are going to give all definitions here in the form which is almost equivalent to the original one (the only difference is in the treatment of the holes, punctures and cilia), but more convenient for us. The construction of coordinates on the space of laminations we are going to describe is a slight modification of Thurston's "train tracks" ([25], Section 9).

It seems worth mentioning here that the definitions of measured laminations are very similar to the definitions of singular homology groups, and is in a sense an unoriented version of the latter ones.

There are two different ways to define the notion of measured lamination for surfaces with boundary, which are analogous to the definition of homology group with compact and closed support, respectively.

Measured laminations are certain collection of weighted curves on the surface. By a *curve* here we mean a curve without self-intersections either closed or connecting two points on the boundary, disjoint from cilia and considered up to homotopy within the class of such curves. We call a curve *special* if it is retractable to a hole or to an interval in the boundary containing exactly one cilium. We call a curve *contractible* if it can be retracted to a point within this class of curves. In particular a nonclosed curve is contractible if and only if it is retractable to a segment of the boundary containing no cilia.

3.1 Unbounded measured laminations

Definition. A *rational unbounded measured lamination* or an \mathcal{X} -*lamination* on a 2-dimensional ciliated surface is given by orientations of some boundary components and a homotopy class of a finite collection of non-selfintersecting and pairwise nonintersecting curves with positive rational weights either closed or ending at the boundary and disjoint from cilia, subject to the following equivalence relations:

- (1) A lamination containing a contractible or a special curve is equivalent to the lamination with this curve removed.
- (2) A lamination containing two homotopy equivalent curves of weights u and v , respectively, is equivalent to the lamination with one of these curves removed and with the weight $u + v$ on the other.
- (3) Oriented boundary components are holes unless they are disjoint from the curves.

The set of all rational unbounded laminations on a given surface S is denoted by $T^x(S, \mathbb{Q})$. This space has a natural subset, given by collections of curves with integral weights. This space is denoted by $T^x(S, \mathbb{Z})$. We shall also omit the arguments if they are clear from the context.

There is an action of the multiplicative group of positive rational numbers on unbounded laminations $\mathbb{Q}_{>0} \times T^x(\mathbb{Q}) \rightarrow T^x(\mathbb{Q})$ given by multiplication of the weights of all curves by a fixed positive number.

If the surface S is open the set of laminations without closed curves form an open dense subset of $T^x(\mathbb{Q})$.

To each hole ρ we associate a number r^ρ equal to the total weight (resp. the negative of total weight) of curves entering the hole if the orientation of the hole is positive (resp. negative).

To each boundary component ρ with *even* number of cilia we associate a number also denoted by r^ρ equal to the sum of total weights of curves entering the segments between cilia taken with alternating signs.

Construction of coordinates. Let us first rotate each boundary component without cilia infinitely many times in the direction prescribed by the orientation as shown in Figure 3. Then deform the curves in such a way that they do not cross any edge of the

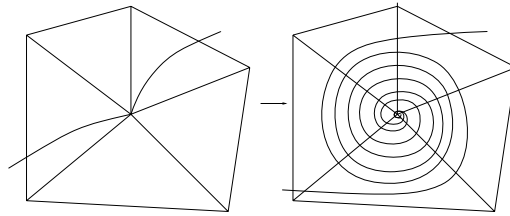


Figure 3. Twisting curves incident to a boundary component without cilia.

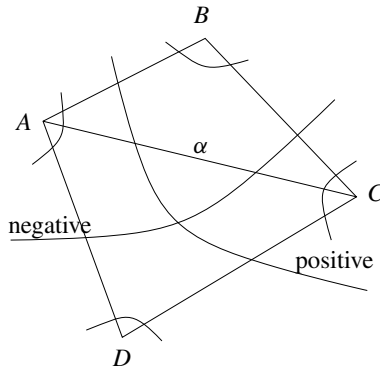


Figure 4. Positive and negative intersection points.

triangulation consecutively in opposite directions. Consider a quadrilateral formed by two triangles sharing an edge α . Denote its vertices by $ABCD$ in the clockwise direction with respect to the orientation, starting from a vertex of α . We call an

intersection point of the edge α and a curve of the lamination *positive* if it belongs to a segment of a curve connecting AB and CD and *negative* if it belongs to a segment of a curve connecting BC and AD . Since the curves of the lamination do not intersect, having both negative and positive intersection points on the same edge is impossible. Notice that we do not assign a sign, say, to a curve intersecting BC and CD .

Now assign to each edge α the sum over the positive intersection points of weights of the respective curves minus the sum over the negative intersection points of weights of the respective curves. The collection of these numbers, one for each internal edge of Γ , is the desired set of coordinates.

Observe that although the curves spiralling around the holes without cilia intersect some edges infinitely many times, only a finite number of these intersection points are positive or negative, and therefore all numbers on edges obtained in this way are finite.

Now we need to prove that these numbers are global coordinates indeed, i.e., that the correspondence between rational unbounded measured laminations and rational numbers on edges is one-to-one. We shall do it by describing an inverse construction.

Note that if we are able to construct a lamination corresponding to the set of numbers $\{x^\alpha\}$, we can also do it for the set $\{ux^\alpha\}$ for any rational $u \geq 0$. Therefore we can reduce our task to the case where all numbers on edges are integral.

Take a triangle, parametrise each side of it by \mathbb{R} respecting the orientation induced by the orientation of the triangle, and connect each point with parameter $i \in \frac{1}{2} + \mathbb{Z}_{\geq 0}$ on one side to the point with parameter $-i$ on the next side in the clockwise direction with respect to the orientation of the surface. This can be done by curves without mutual intersections. Now replace every triangle of our triangulation with such triangles with curves in such a way that the point with parameter i on one side of an edge α is glued to the point with parameter $x^\alpha - i$ on the other side (Figure 5).

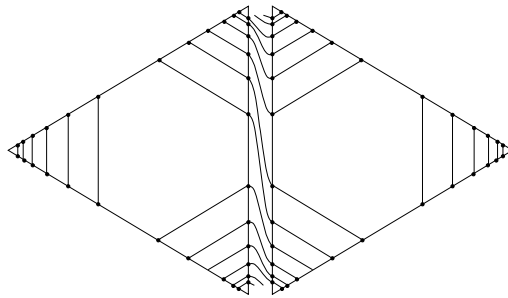


Figure 5. Gluing triangles for $x^\alpha = 2$.

Observe that although we have started with an infinite bunch of curves, the resulting lamination is finite: all these curves glue together into a finite number of connected components and a possibly infinite number of special curves. Indeed, any connected component either intersects positively or negatively at least one edge or is closed. Since the total number of positive or negative intersection points $\sum_{\alpha \in E(\Gamma) - E_0(\Gamma)} |x^\alpha|$ is finite, the resulting lamination contains no more than this number of connected

components. In particular the number of connected components equals this number provided all numbers x^α are all nonpositive or all nonnegative.

Properties of the coordinates. The constructed coordinates on the space of laminations correspond to a particular choice of triangulation. If we change the triangulation, the corresponding coordinates change. The rule for changing the coordinates under a flip of the triangulation is given by an explicit formula:

$$x'^{\beta'} = \begin{cases} -x^\alpha & \text{if } \beta = \alpha, \\ x^\beta + \varepsilon^{\alpha\beta} \max(0, x^\alpha) & \text{if } \varepsilon^{\alpha\beta} \geq 0, \\ x^\beta + \varepsilon^{\alpha\beta} \max(0, -x^\alpha) & \text{if } \varepsilon^{\alpha\beta} \leq 0. \end{cases} \quad (3.1)$$

If all edges of the quadrilateral taking part in the flip are different, the change of coordinates can be shown by the graphical rule Figure 6.

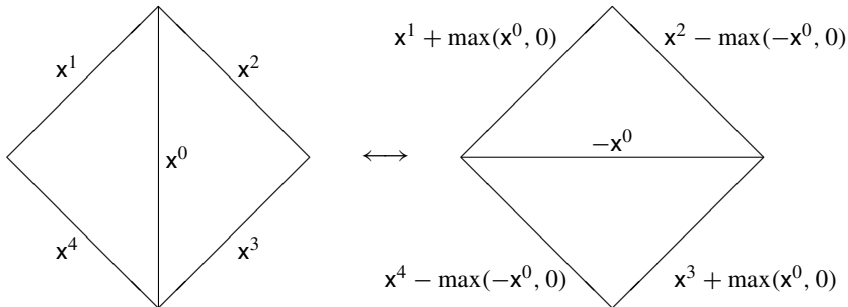


Figure 6. Changing coordinates under a flip of the triangulation. Only the changing coordinates are shown here. The numbers on the other edges remain unchanged.

An \mathcal{X} -lamination is *integral* if all its coordinates are integral.

If ρ is a boundary component with even number of cilia, the canonical map $r^\rho: \mathbb{T}^x \rightarrow \mathbb{Q}$ is given by $r^\rho = \sum_{\alpha} \zeta_{\rho}(\alpha)x$.

Given an edge α a lamination with coordinates $x^\alpha = 1, x^\beta = 0$ for $\beta \neq \alpha$ is just a single curve homotopy equivalent to the edge α with the ends shifted slightly from cilia in the counterclockwise direction. A lamination having nonnegative coordinates is just a union of such curves. It implies that for any lamination without closed curves there exists a triangulation such that all coordinates of the lamination are nonnegative. Conversely if a lamination have positive coordinates with respect to a triangulation then such triangulation is unique.

3.2 Bounded measured laminations

Definition. A rational bounded measured lamination or a rational \mathcal{A} -lamination on a 2-dimensional ciliated surface is a homotopy class of a collection of finite number of self- and mutually nonintersecting unoriented curves either closed or connecting two points of the boundary disjoint from cilia with rational weights and subject to the following conditions and equivalence relations.

- (1) Weights of all curves are positive, unless a curve is special.
- (2) A lamination containing a curve of weight zero is considered to be equivalent to the lamination with this curve removed.
- (3) A lamination containing a contractible curve is considered to be equivalent to the lamination with this curve removed.
- (4) A lamination containing two homotopy equivalent curves with weights u and v is equivalent to the lamination with one of these curves removed and with the weight $u + v$ on the other.

Recall that a contractible curve in this context is either closed or retractable to a segment of the boundary without cilia.

The set of all rational bounded laminations on a given surface S is denoted by $T^a(S, \mathbb{Q})$. This space has a natural subset called the set of *integral* bounded laminations, consisting of laminations with integral weights. It is denoted by $T^a(S, \mathbb{Z})$. We shall omit the arguments of T^a if they are clear from the context or if the statement is valid for any value of the arguments.

The space $T^a(S)$ has a subspace $T_0^a(S)$ consisting of the laminations such that for any segment of the boundary between two cilia the total weight of curves ending at it vanishes.

There is a canonical action of the multiplicative group of positive rational numbers on bounded lamination $\mathbb{Q}_{>0} \times T^a(\mathbb{Q}) \rightarrow T^a(\mathbb{Q})$ given by multiplication of weights of all curves by a fixed positive number.

For every hole ρ and for every connected component with *even* number of cilia there is an action r_ρ of the additive group of rational numbers $r_\rho: \mathbb{Q} \times T^a \rightarrow T^a$. For a hole this group acts by adding a closed loop with a certain weight around the hole. Similarly, for every boundary component with even number of cilia ρ the group acts by adding a special curve about every second cilium with a certain weight w and adding a special curve with weight $-w$ about the remaining cilia of ρ .

For every hole ρ there is a map $A_\rho: T^a \rightarrow \mathbb{Q}$ (called a *collar map* [20], where it was defined and studied) given by the total weight of the loops surrounding the hole.

There is a canonical mapp: $T^a(S) \rightarrow T^x(S)$ since every bounded lamination can be considered as an unbounded one. This map obviously commutes with multiplication by positive numbers and sends integral laminations to integral ones.

Construction of coordinates. Suppose we are given a triangulation Γ of a ciliated surface S . We are going to assign, for a given lamination, rational numbers on

edges of Γ and show that these numbers are global coordinates on the space of laminations.

Deform the curves of the lamination in such a way that every curve intersects every edge at the minimal possible number of points. Assign to each edge α half of the sum over the intersection points of weights of the respective curves. Collection of numbers, one for each edge of Γ , is the desired set of coordinates. For the space $T_0^g(S)$, the numbers assigned in this way to any external edge vanishes. Thus only numbers assigned to internal edges serve as coordinates for this space.

Reconstruction. Now we need to prove that these numbers are coordinates indeed, i.e., that they provide a bijection between the space of bounded measured laminations and collections of rational numbers on edges. For this purpose we just describe an inverse construction, which gives a lamination, starting from numbers on edges.

Let us first fix a pair of rationals u and v and introduce the numbers $\{\tilde{a}_\alpha \mid \alpha \in E(\Gamma)\}$ by

$$\tilde{a}_\alpha = u a_\alpha + v.$$

By choosing u and v one can make the numbers $\{\tilde{a}_\alpha\}$ to be positive integers and to satisfy the triangle inequality for every triangle of the triangulation with edges α, β, γ :

$$|\tilde{a}_\alpha - \tilde{a}_\beta| \leq \tilde{a}_\gamma \leq \tilde{a}_\alpha + \tilde{a}_\beta. \quad (3.2)$$

Now mark $2\tilde{a}_\alpha$ points on each edge α and join the points on the sides of each triangle pairwise by nonintersecting and non-selfintersecting lines connecting points on different sides and passing inside the triangle. One can easily see that once the triangle inequality is satisfied this can be done in a unique way up to homotopy (Figure 7). Now take each hole and each cilium and add a closed curve surrounding them with

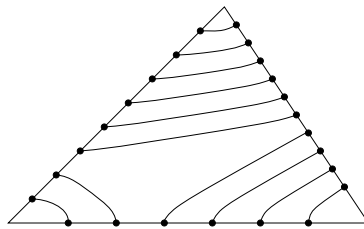


Figure 7. $(\tilde{a}_\alpha, \tilde{a}_\beta, \tilde{a}_\gamma) = (3, 4, 5)$.

weight $-v$. Then divide the weights of all curves by u . The construction is finished. The proof that the construction is indeed inverse to the previous one is self-evident.

Properties of the coordinates. Just as for the case of \mathcal{X} -laminations, the constructed coordinates on the space of laminations correspond to a particular choice of the triangulation. The rule stating how the coordinates change under a flip of triangulation is

given by the formula:

$$a'_{\beta'} = \begin{cases} \max(\sum_{\delta|\varepsilon^{\beta\delta} > 0} \varepsilon^{\beta\delta} a_\delta, -\sum_{\delta|\varepsilon^{\beta\delta} < 0} \varepsilon^{\beta\delta} a_\delta) - a_\alpha & \text{if } \beta = \alpha, \\ a_\beta & \text{if } \beta \neq \alpha. \end{cases} \quad (3.3)$$

Since any triangulation change is a composition of flips, these rules allow to express the coordinate change for any triangulation change.

If all edges of the quadrilateral taking part in the flip are different, the coordinates change can be shown by the graphical rule on Figure 8.

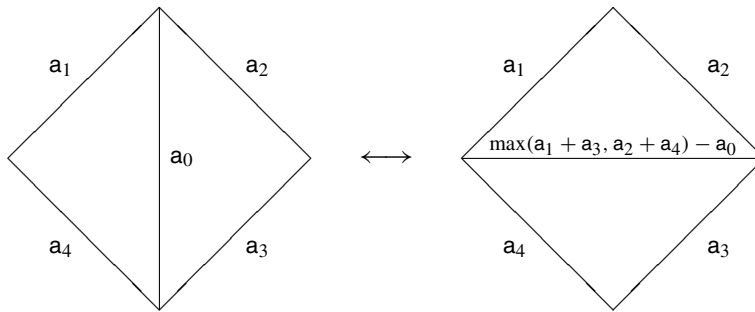


Figure 8. Change of coordinates under a flip of the triangulation. Only the changing coordinates are shown here, the numbers on the other edges remain unchanged.

A bounded lamination is integral if and only if all coordinates are integral or half integral and for every triangle with sides α, β, γ the sum $a_\alpha + a_\beta + a_\gamma$ is integral.

If we multiply a lamination by a positive number, the coordinates get multiplied by the same number.

A bounded lamination is called *even* if all coordinates are integers. One can easily check that this condition does not depend on the coordinate system. This set is denoted by T_{ev}^a . Observe that an even lamination divided by 2 is not necessarily integral.

The canonical map $p: T^a \rightarrow T^x$ is given by $x^\alpha = -\sum_\beta \varepsilon^{\alpha\beta} a_\beta$.

The action r_ρ of the additive group \mathbb{Q} corresponding to a boundary component ρ with even number of cilia is given by

$$r_\rho(w): a_\alpha \mapsto w\zeta_\rho(\alpha) + a_\alpha.$$

If ρ is a hole, then the collar map A_ρ is given by

$$A_\rho = \max_t (a_{\alpha_t} + a_{\beta_t} - a_{\delta_t}),$$

where t runs over all triangle having ρ as a vertex, α_t and β_t are the edges of t incident to ρ and δ_t is the third edge.

Real lamination spaces. Since the transformation rules for coordinates (8) and (6) are continuous w.r.t. the standard topology of \mathbb{Q}^n , the coordinates define a natural topology on the lamination spaces. One can now define the spaces of *real measured laminations* (resp. bounded and unbounded) as a completion of the corresponding spaces of rational laminations. These spaces are denoted as $T^a(\mathbb{R})$ and $T^x(\mathbb{R})$, respectively. Of course, we automatically have coordinate systems on these spaces.

Observe that to define real measured laminations it is not enough to replace rational numbers by real numbers in the definition of the space of laminations unless the surface S is a disc or a cylinder with one hole. Such definition would not be equivalent to the one above since a sequence of more and more complicated curves with smaller and smaller weights may converge to a real measured lamination which is not presentable by a finite collection of curves.

4 Teichmüller spaces

The *Teichmüller space* $\mathcal{T}(S)$ (resp. *moduli space* $\mathcal{M}(S)$) of a closed surface S is the space of complex structures on S modulo diffeomorphisms isotopic to the identity (resp. modulo all diffeomorphisms). We are going to give two different extensions of this notion to the case of open surfaces with cilia. But before we recall some basic facts about relations between complex structures, constant negative curvature metrics and discrete subgroups of the group $\mathrm{PSL}(2, \mathbb{R})$. For more details we recommend the reviews [1] and [19].

Consider first a vicinity of a boundary component. Topologically it is a cylinder, but as a complex surface it can be isomorphic either to a cylinder $\{z \in \mathbb{C} \mid 0 < r < |z| < 1\}$ or to a punctured disk $\{z \in \mathbb{C} \mid 0 < |z| < 1\}$. The boundary components of the second kind are called *punctures*.

According to the Poincaré uniformisation theorem any complex surface S (except a sphere with less than three holes and a torus without holes) can be represented as a quotient of the hyperbolic plane H by a discrete subgroup Δ of its automorphism group (sometimes called the Möbius group) $\mathrm{PSL}(2, \mathbb{R})$ of real 2×2 matrices with unit determinant considered up to the factor -1 . The group Δ is defined up to conjugation, is discrete as a subgroup of $\mathrm{PSL}(2, \mathbb{R})$, is canonically isomorphic to the fundamental group of the surface $\pi_1(S)$ and acts freely on H . The subgroups with these properties are called *Fuchsian groups*. In particular, this implies that the Teichmüller space $\mathcal{T}(S)$ for a surface S without cilia is isomorphic to a connected component in the space of Fuchsian groups considered up to conjugation. The connected component is singled out by the topology of the quotient of the upper half plane. The group Δ corresponding to a given complex structure is called the *monodromy group*.

Recall that an element g of $\mathrm{PSL}(2, \mathbb{R})$ is called *hyperbolic* (resp. *parabolic*, *elliptic*) if it has two distinct real eigenspaces (resp. one one-dimensional real eigenspace, two conjugate complex eigenspaces). Since eigenspaces are stable points of the action

of g on $\mathbb{C}P^1$, only parabolic and hyperbolic elements act on H freely and thus all elements of Δ must be of these two types.

The quotient of H by the action of the subgroup generated by a parabolic element is a punctured disk. Thus the only elements of $\pi_1(S)$ which can be mapped to parabolic elements are loops surrounding punctures.

On H there exists a unique $\mathrm{PSL}(2, \mathbb{R})$ -invariant curvature -1 Riemannian metric given in the standard coordinates by $(\mathrm{Im} z)^{-2} dz d\bar{z}$. It induces a metric on S . Since this metric is of negative curvature, any homotopy class of closed curves contains a unique geodesic unless a curve surrounds a puncture. In the latter case the length of the curve can be made arbitrarily small and. In this case we say that the geodesic length is zero.

Any nontrivial element of Δ has one or two fixed points on $\mathbb{R}P^1$, which is the boundary (absolute) of H . The *singular set* of a ciliated surface S is the subset of $\mathbb{R}P^1$ consisting of all fixed points of all elements of the monodromy group Δ and of preimages of all cilia. This set is Δ -invariant by construction and thus its convex hull in H is also Δ -invariant. The *convex core* S_0 of the surface S is the quotient of this convex hull by Δ . To describe it geometrically in terms of the original surface S consider the closed geodesics surrounding holes as well as geodesics connecting adjacent cilia on the same boundary components. Cut out of the surface the pieces facing the respective boundary components. These pieces as Riemannian manifolds are isomorphic either to the positive quadrant of H or its quotients by a cyclic group. The remaining part of the surface S is just its convex core S_0 .

For example, if our surface is a ciliated disk, its convex core is an ideal polygon.

Homotopy classes of closed curves on a surface are in one-to-one correspondence with the conjugacy classes of its fundamental group. Denote by γ an element of $\pi_1(S)$ and by $l(\gamma)$ the length of the corresponding geodesic. Then a simple computation shows that

$$l(\gamma) = \left| \log \frac{\lambda_1}{\lambda_2} \right|, \quad (4.1)$$

where λ_1 and λ_2 are the eigenvalues of the monodromy along the loop γ . This number is obviously well defined, i.e., it does not depend on the choices of particular representation of $\pi_1(S)$, of a particular element of $\pi_1(S)$ representing a given loop, and of a particular 2×2 matrix representing the element of $\mathrm{PSL}(2, \mathbb{R})$. This formula agrees with the convention that the length of the geodesic surrounding a puncture is zero.

4.1 Teichmüller space of surfaces with holes $\mathcal{T}^x(S)$

Definition. *The Teichmüller space of a ciliated surface with holes or the Teichmüller \mathcal{X} -space, denoted by $\mathcal{T}^x(S)$, is the space of complex structures on S together with an orientation of all holes except the punctures up to the diffeomorphisms homotopy equivalent to the identity and preserving cilia pointwise.*

It is not *a priori* obvious that this space possesses a natural topology in which it is connected, but this will be clarified later.

For every hole ρ there exists a map $r^\rho : \mathcal{T}^x(S) \rightarrow \mathbb{R}_{>0}$ given by the exponential of the length of the geodesic surrounding this hole (resp. exponential of minus the length of the geodesics surrounding the hole) if the orientation of the hole agrees (resp. disagrees) with the orientation of the surface.

For every boundary component with *even* number of cilia there exists a map denoted by the same letter $r^\rho : \mathcal{T}^x(S) \rightarrow \mathbb{R}_{>0}$ defined as follows. Take an horocycle about one cilium. Then take another one about the next cilium tangent to the first one. Repeat this procedure until we come back to the original cilium. The value of r^ρ is $e^{\pm l}$, where l is the distance between the two horocycles about the first cilium and the sign depends on whether the final horocycle is inside or outside the original one. The value of r^ρ does not depend on the choice of the original horocycle and may change to its inverse if we have chosen another starting point.

Construction of coordinates. Let Γ be a triangulation of S . For any point of $\mathcal{T}^x(S)$ we are going to describe a rule for assigning a positive real number to each internal edge of Γ . The collection of these numbers will give us a global parameterisation of $\mathcal{T}^x(S)$.

Every edge of Γ connecting two cilia, a cilium and a puncture, or two punctures can be made geodesic in a canonical way. In order to make geodesic the edges connecting a cilium or a puncture with a hole first draw a geodesic around each hole and choose an arbitrary point on it. Then deform the edge to a geodesic ray connecting the cilium and the chosen point on the geodesic. Observe that the orientation of the hole induces an orientation of the geodesic. Now move the chosen point along the geodesic in the direction given by this orientation. The limiting position of the ray is just the desired one. By construction it is spiralling around the hole. Analogous procedure can be done for the edges connecting two holes giving a geodesic spiralling around one closed geodesic at one end and around another one at another end.

We end up by a triangulation of a part (actually of the convex core S_0) of the surface S into ideal triangles.

Now lift the triangulation to the upper half plane and consider an edge together with two adjacent triangles forming an ideal quadrilateral. The quadruple of points of this quadrilateral have one invariant under the action of $\text{PSL}(2, \mathbb{R})$. For our purpose it is convenient to choose the cross-ratio x by choosing a coordinate on H (sending it to the upper half plane) equal to 0 and ∞ at the ends of the edge and to -1 at the third vertex of the quadrilateral (Figure 9). Then the value of the coordinate at the fourth vertex is the desired parameter x associated to the chosen edge. It is an easy exercise to show that there are two ways to choose the coordinate and both give the same value of the cross-ratio.

Reconstruction. Our goal now is to construct a ciliated surface with hyperbolic metric starting from a triangulation of a ciliated surface with real positive numbers $\{x^\alpha\}$

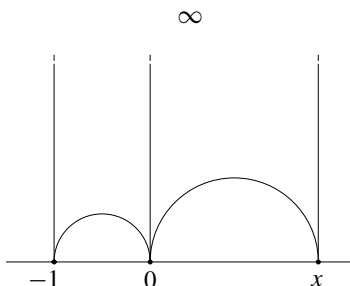


Figure 9. Definition of the cross-ratio.

assigned to internal edges. First of all we give a simple prescription for how to restore orientations of the boundary components from these data: The orientation of a boundary component corresponding to a hole γ is just induced from the orientation of the surface (resp. opposite to the induced one) if the sum $\sum x^\alpha$ over all α incident to γ is positive (resp. negative). If the sum is zero, it means that it is not a hole, but a puncture.

The construction of the surface itself can be achieved in two equivalent ways. We shall describe both since one is more transparent from the geometric point of view and the other is useful for practical computations.

Construction by gluing. We are going to glue the convex core S_0 of our surface S out of ideal hyperbolic triangles. The lengths of the sides of ideal triangles are infinite and therefore we can glue two triangles in many ways which differ by shifting one triangle with respect to another along the side. The ways of gluing triangles can be parameterised by the cross-ratios of four vertices of the obtained quadrilateral (considered as points of $\mathbb{R}P^1$) just as defined by Figure 9.

Now consider the triangulated surface S . Replace every triangle of the triangulation by an ideal one and glue them together along the edges just as they are glued in S using numbers assigned to the edges as gluing parameters.

Note, although this is not quite obvious that the resulting surface is not necessarily complete even if the surface has no cilia. In fact it is the complex core S_0 with boundary components corresponding to holes removed.

Construction of the Fuchsian group. We are now going to construct a discrete monodromy subgroup Δ of $\mathrm{PSL}(2, \mathbb{R})$ starting from a triangulation of S with positive real numbers on edges. Construct first a graph out of the triangulation in the following way (Figure 10). Draw a small edge of the graph transversal to every edge of the triangulation and for every triangle connect the three ends of edges in it pairwise. Orient the edges of the arising triangle in the clockwise direction. Now assign to each of these edges the matrix $U = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$. Assign to the remaining edges the matrix $B(x^\alpha) = \begin{pmatrix} 0 & (x^\alpha)^{1/2} \\ -(x^\alpha)^{-1/2} & 0 \end{pmatrix}$, where α is the edge of the triangulation it intersects.

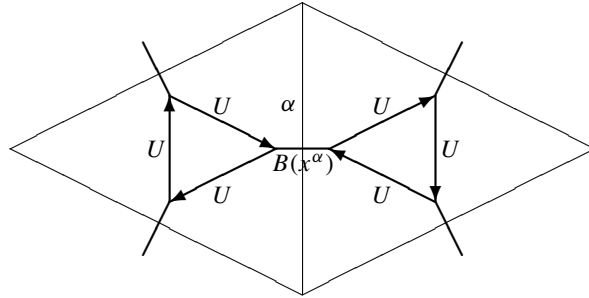


Figure 10. Graph for computing monodromy for $\mathcal{T}^x(S)$. The original triangulation is shown by the thin lines.

Now to any oriented path on this graph we can associate a matrix by multiplying consecutively all matrices we meet along it, taking U^{-1} instead of U each time the orientation of the path disagrees with the orientation of the edge. (The orientations of the edges transversal to the edges of the triangulation are not to be taken into account, since $B(x)^2 = -1$ and therefore $B(x)$ coincides with its inverse in the group $\text{PSL}(2, \mathbb{R})$.) In particular, if we take closed paths starting from a fixed vertex of the graph, we get a homomorphism of the fundamental group of Γ to the group $\text{PSL}(2, \mathbb{R})$. The image of this homomorphism is the desired group Δ .

The fact that these two constructions are inverse to the above construction of coordinates is almost obvious, especially for the first one. The only note we would like to make here is to show where the matrices I and $B(x)$ came from. Consider two ideal triangles in the upper half plane with vertices at the points $-1, 0, \infty$ and $x, \infty, 0$, respectively. Then the Möbius transformation which permutes the vertices of the first triangle is given by the matrix I , and the one which maps one triangle to another (respecting the order of vertices given two lines above) is given by $B(x)$.

Properties of the coordinates. The constructed coordinates on Teichmüller space correspond to a particular choice of triangulation. The rule of coordinate change for a flip is given by the formula

$$x'^{\beta'} = \begin{cases} (x^\alpha)^{-1} & \text{if } \beta = \alpha, \\ x^\beta(1 + x^\alpha)^{\varepsilon^{\alpha\beta}} & \text{if } \varepsilon^{\alpha\beta} \geq 0, \\ x^\beta(1 + (x^\alpha)^{-1})^{\varepsilon^{\alpha\beta}} & \text{if } \varepsilon^{\alpha\beta} \leq 0. \end{cases} \quad (4.2)$$

If all edges of the quadrilateral taking part in the flip are different, the coordinate change can be shown by the graphical rule Figure 11.

The canonical map $r^\rho: \mathcal{T}^x \rightarrow \mathbb{R}$ is given by

$$r^\rho = \prod_{\alpha} (x^\alpha)^{\zeta(\alpha)}, \quad (4.3)$$

where the product is taken over all edges of the triangulation Γ .

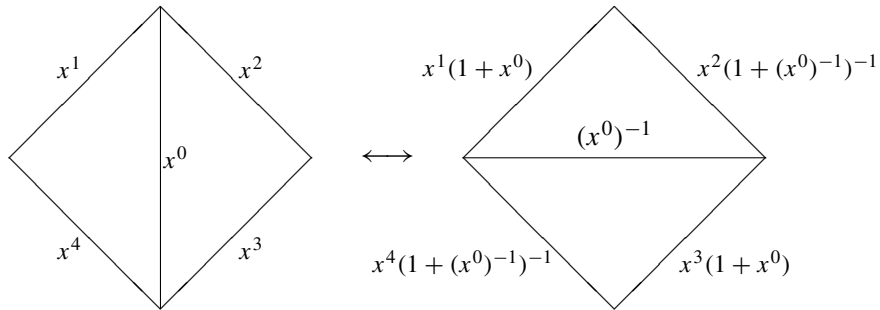


Figure 11. Changing of coordinates under a flip of the triangulation. Only the changing coordinates are shown here, the numbers on the other edges remain unchanged.

Integral monodromy. It is obvious from the reconstruction formulae that if all coordinates are equal to 1 then the monodromy group Δ is a subgroup of $\mathrm{PSL}(2, \mathbb{Z})$. Of course the subgroup Δ obtained in this way is a finite index subgroup and is free (and thus torsion free). It turns out that a kind of converse statement is also true. Namely for any torsion free finite index subgroup $\Delta \subset \mathrm{PSL}(2, \mathbb{Z})$ there exists a surface S with a triangulation $|\Gamma|$ which is unique up to the mapping class group action such that Δ is the monodromy group of a hyperbolic structure on S corresponding to all coordinates equal to 1. Thus we have an explicit bijection between conjugacy classes of torsion free finite index subgroups of $\mathrm{PSL}(2, \mathbb{Z})$ and triangulated surfaces.

We shall give an explicit construction of the triangulated surface out of a torsion free discrete subgroup Δ of $\mathrm{PSL}(2, \mathbb{Z})$ borrowed from Kontsevich [18].

Observe that the group $\mathrm{PSL}(2, \mathbb{Z})$ is the free product of the subgroups $\mathbb{Z}/2\mathbb{Z}$ generated by the matrix $\sigma = B(1) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\mathbb{Z}/3\mathbb{Z}$ generated by the matrix $\nu = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$. Every nontrivial torsion element of $\mathrm{PSL}(2, \mathbb{Z})$ is conjugated to either σ or ν . Consider the finite set $F = \Delta \setminus \mathrm{PSL}(2, \mathbb{Z})$. Since Δ is torsion free, the action of the subgroups $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z}$ on F is free. For every $\mathbb{Z}/3\mathbb{Z}$ orbit $f, \nu f, \nu^2 f$ take a triangle with vertices at the points of this orbit. Then for every $\mathbb{Z}/3\mathbb{Z}$ orbit $f, \sigma f$ glue together the triangles $f, \nu f, \nu^2 f$ to the triangle $\sigma f, \nu \sigma f, \nu^2 \sigma f$ along one side sending f to $\nu \sigma f$ and σf to τf . We get the desired triangulated surface without boundary and such that the triangles are numerated by the orbits of τ , edges are the orbits of σ and vertices are orbits of $\nu \sigma$.

4.2 Teichmüller space of decorated surfaces $\mathcal{T}^a(S)$

In this section we are going to reproduce some results of Penner [21].

Before giving a definition of the Teichmüller space $\mathcal{T}^a(S)$, recall what is a *horocycle*. A horocycle in the upper half plane H at $A \in \mathbb{R}P^1$ is a limit of a circle passing through a fixed point on H when its hyperbolic centre tends to A . Equivalently a

horocycle at A is an orbit of a subgroup of $\mathrm{PSL}(2, \mathbb{R})$ stabilising A . A horocycle at the point $A \in \mathbb{R} \subset \mathbb{R}P^1$ is a Euclidean circle belonging to the upper half plane tangent to the real axis at A . A horocycle at $\infty \in \mathbb{R}P^1$ is a horizontal line. Points of one horocycle are equidistant from another horocycle based at the same point. The set of all horocycles is isomorphic to the space $(\mathbb{R}^2 - \{(0, 0)\})/\pm 1$. This isomorphism is equivariant with respect to the $\mathrm{PSL}(2)$ action and sends the vector $(x, y) \in \mathbb{R}^2$ into a horocycle about the point x/y and of euclidean diameter y^{-2} .

For a punctured Riemann surface $S = H/\Delta$, where Δ is a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$, a horocycle at a cilium or a puncture A is a Δ -invariant set of horocycles at each preimage of A on $\mathbb{R}P^1$. By definition the area of a horocycle at the puncture is the area inside the horocycle of the quotient of H by the subgroup of Δ stabilising the basepoint. If the horocycle is small enough, its image on S is a small circle surrounding a puncture or tangent to the boundary at the cilium and orthogonal to any geodesics coming out of this puncture or cilium. In this case its area coincides with the actual area inside the image. However the projection of a general horocycle to the surface may have a relatively complicated topology and its area inside the image differs from the area of the horocycle.

We say that two horocycles on the ciliated surface S are *tangent* along the path γ connecting their base points if the corresponding horocycles at the ends of the lift of the path to the universal covering H are tangent.

Definition. A *decorated ciliated Riemann surface* is a ciliated hyperbolic surface such that every hole is a puncture and with a horocycle chosen at each puncture and cilium. The Teichmüller space of decorated ciliated surfaces is called the *Teichmüller \mathcal{A} -space* and is denoted by $\mathcal{T}^{\mathcal{A}}(S)$.

The space $\mathcal{T}^{\mathcal{A}}(S)$ has a canonical subspace $\mathcal{T}_0^{\mathcal{A}}(S)$ consisting of decorated surfaces such that for every path connecting two cilia and retractable to a segment of the boundary without cilia, the corresponding horocycles are tangent.

Forgetting the horocycles we get a canonical map $p: \mathcal{T}^{\mathcal{A}} \rightarrow \mathcal{T}^x$.

For every cilium and for every hole ρ there is an action of the multiplicative group of positive numbers $r_\rho: \mathbb{R}_{>0} \times \mathcal{T}^{\mathcal{A}} \rightarrow \mathcal{T}^{\mathcal{A}}$ given by changing the size of the corresponding horocycle.

For every hole ρ there is an area map $A_\rho: \mathcal{T}^{\mathcal{A}} \rightarrow \mathbb{R}_{>0}$ given by the area inside the corresponding horocycle.

There is a canonical map $p: \mathcal{T}^{\mathcal{A}} \rightarrow \mathcal{T}^x$ given by forgetting horocycles.

Construction of coordinates. Let Γ be a triangulation of S . For any point of $\mathcal{T}^{\mathcal{A}}(S)$, we are going to assign a positive real number to each edge of Γ . The collection of these numbers will give us a global parameterisation of $\mathcal{T}^{\mathcal{A}}(S)$.

Consider an edge α of the triangulation, deform it into a geodesic and lift it to the universal covering H . There are the horocycles attached to the ends of the edges. The number a_α which we assign to the edge α is the exponential of half of the length (resp. of minus the length) of the segment of the lifted edge between the intersection

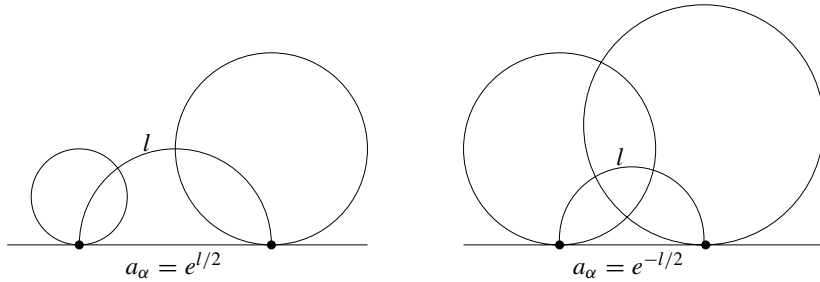


Figure 12. Penner coordinates.

points with the horocycles if the horocycles do not intersect (resp. if the horocycles do intersect), as shown on Figure 12.

This number can be easily computed algebraically using the correspondence between horocycles and vectors in \mathbb{R}^2 . If (x_1, y_1) and (x_2, y_2) are vectors, corresponding to the horocycles, then $a_\alpha = |\det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}|$.

Reconstruction. This is quite similar to that of holed surfaces. There is a canonical mapping $p: \mathcal{T}^a(S) \rightarrow \mathcal{T}^x(S)$ which just forgets the horocycles and is given explicitly in coordinates by (4.5). Therefore to reconstruct the surface itself we can apply the reconstruction procedure for $\mathcal{T}^x(S)$. To reconstruct the horocycles, consider an ideal triangle which we have used to glue the surface. On each edge, the a length of the corresponding geodesics between the horocycles allows us to restore unambiguously intersection points of the horocycles with the edges.

Observe that given two points A and B on $\mathbb{R}P^1$ and a horocycle about A one can associate a canonical coordinate z on the upper half plane H , such that $z(A) = \infty$, $z(B) = 0$ and the horocycle is given by the line $\text{Im } z = 1$. Therefore given a lift of an oriented edge of the triangulation to the upper half plane we get a canonical coordinate on the latter taking for A the beginning of this edge and for B its end. Changing the orientation of the edge results in a Möbius transformation of the coordinate with the matrix $D(a_\alpha) = \begin{pmatrix} 0 & a_\alpha \\ -a_\alpha^{-1} & 0 \end{pmatrix}$ and passing from the side α of a triangle with the edges α, β, γ to the edge β with orientations towards the same vertex results in the Möbius transformation with the matrix $F\left(\frac{a_\gamma}{a_\alpha a_\beta}\right) = \begin{pmatrix} 1 & 0 \\ \frac{a_\gamma}{a_\alpha a_\beta} & 1 \end{pmatrix}$.

Similarly to the case of the space $\mathcal{T}^x(S)$ one can give a rule to find the corresponding Fuchsian group as a monodromy group of a connection on a graph. Construct first a graph out of the triangulation in the following way (Figure 13). Draw a small edge of the graph along every edge of the triangulation and for every triangle connect the six ends of edges to form a hexagon. Orient the edges surrounding every vertex of the triangulation in the clockwise direction. Assign to each edge of the graph parallel to the edge α the matrix $D(a_\alpha)$ and assign to the edge connecting sides α and β of the triangle $\alpha\beta\gamma$ the matrix $F\left(\frac{a_\gamma}{a_\alpha a_\beta}\right)$. Now for any closed path on the surface one can

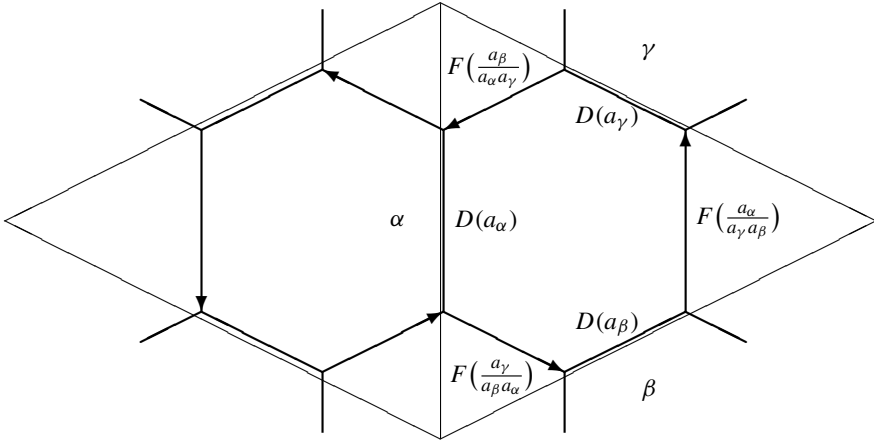


Figure 13. Graph for reconstructing the monodromy for $\mathcal{T}^\alpha(S)$.

retract this path to the graph and the monodromy operator can be computed by the product of the corresponding matrices on the edges.

Moreover this graph with matrices allows to compute the signed distance between any two horocycles. Indeed, consider the path connecting two corresponding vertices of the triangulation. Move each endpoint along an edge of the triangulation to the nearest vertex of the graph and then deform it to pass along the edges of the graph. Associate to it the product of the matrices along this path. Then the absolute value of the upper right element of this matrix is equal to the exponential of signed length between the horocycles. The result does not depend on how we have deformed the path since the monodromy along the edges surrounding the vertex is strictly lower triangular.

Properties of the coordinates. Just as for the three cases considered above, the constructed coordinates correspond to a particular choice of triangulation. The rule describing how the coordinate changes under a flip of an edge α is given by

$$a'_{\beta'} = \begin{cases} \frac{\prod_{\delta|\varepsilon\beta\delta > 0} a_\delta^{\varepsilon\beta\delta} + \prod_{\delta|\varepsilon\beta\delta < 0} a_\delta^{-\varepsilon\beta\delta}}{a_\alpha} & \text{if } \beta = \alpha, \\ a_\beta & \text{if } \beta \neq \alpha. \end{cases} \quad (4.4)$$

If all edges of the quadrilateral taking part in the flip are different, the coordinate change can be shown by the graphical rule shown on Figure 14.

The canonical map $p: \mathcal{T}^\alpha \rightarrow \mathcal{T}^x$ is given by

$$x^\alpha = \prod_{\beta} (a_\beta)^{-\varepsilon^{\alpha\beta}}. \quad (4.5)$$

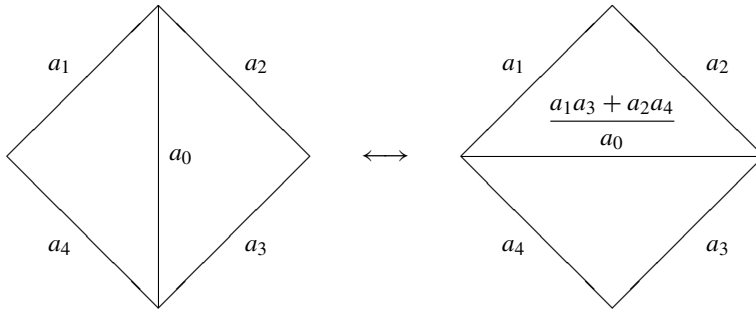


Figure 14. Changing of coordinates under a flip of the triangulation. Only the changing coordinates are shown here, the numbers on the other edges remain unchanged.

The action r_ρ of the multiplicative group $\mathbb{R}_{>0}$ corresponding to a hole with even number of cilia is given by

$$r_\rho(w): a_\alpha \mapsto w^{\zeta_\rho(\alpha)} a_\alpha.$$

If ρ is a hole, then the area map A_ρ is given by

$$A_\rho = \sum_t \frac{a_{\delta_t}}{a_{\alpha_t} a_{\beta_t}},$$

where t runs over all triangle having γ as a vertex, α_t and β_t are the edges of t incident to γ and δ_t is the third edge.

The subspace \mathcal{T}_0^a is the space where all numbers assigned to external edges are equal to one.

5 Tropicalisation

In this short section we shall state explicitly in which sense the spaces of laminations are “tropicalisations” of the corresponding Teichmüller spaces.

Recall that a set \mathbb{F} with two binary operations: addition $+$ and multiplication \cdot is called a *semifield* if it satisfies the following axioms:

- (1) \mathbb{F} is an Abelian semigroup with respect to addition.
- (2) \mathbb{F} is an Abelian group with respect to multiplication.
- (3) Addition and multiplication are related by the distributivity law: $a \cdot (b + c) = a \cdot b + a \cdot c$.

The simplest examples of a semifield are just positive real or rational numbers with respect to the usual addition and multiplication. Another class of examples are given

by the sets of real, rational and integral numbers with the usual addition used for multiplication and usual maximum used for addition. The corresponding semifields are called *real, rational and integral tropical semifields* and are denoted by \mathbb{R}^t , \mathbb{Q}^t and \mathbb{Z}^t respectively. Every algebraic expression which does not contain subtraction (such expressions are called *subtraction free*) makes sense in a semifield. Consider for example the polynomial $a^3 + 3ab^2 + 2$. In tropical semifields this expression can be rewritten using usual operation as $\max(3a, 2a + b, 0)$. We say that the second expression is the *tropical analogue* of the first one.

Observe that tropical semifields can be considered as certain limits of nontropical ones. Namely fix a positive real constant ϵ and consider the set of real numbers provided with the addition: $(a, b) \mapsto \epsilon \log(e^{a/\epsilon} + e^{b/\epsilon})$ and the multiplication given by the ordinary addition. This semifield is obviously isomorphic to the semifield of positive numbers by the isomorphism $a \mapsto e^{a/\epsilon}$. In the limit as $\epsilon \rightarrow 0$ the isomorphism is no longer defined, but the limit of the operations does exist and coincides with the operations in the tropical semifield.

In particular the tropical analogue of a subtraction free rational function of n variables $f(a_1, \dots, a_n)$ is given by

$$f(\mathbf{a}_1, \dots, \mathbf{a}_n) = \lim_{\epsilon \rightarrow 0} \epsilon \log f(e^{a_1/\epsilon}, \dots, e^{a_n/\epsilon}).$$

One can easily see by comparison of the Sections 3 and 4 that the formulae for the spaces of laminations are tropical analogues of the respective formulae for Teichmüller spaces. For example the space of laminations T^x has the same set of coordinate systems as the Teichmüller space \mathcal{T}^x , while the transition functions between the coordinate system (3.1) for the former space are given by tropical analogues of the respective functions (4.2) for the latter one. Similarly the lamination spaces T^a and T_0^a are tropical analogues of the spaces \mathcal{T}^a and \mathcal{T}_0^a respectively.

In particular one can see that Teichmüller spaces can be canonically compactified by adding the quotients of the respective real lamination spaces by action of the multiplicative group of positive real numbers. One can show that it is just a version of the *Thurston compactification* [25] for ciliated surfaces.

6 Poisson and degenerate symplectic structures

We are going to give a definition (by explicit formulae) of a Poisson bracket P_{WP} on the spaces $\mathcal{T}^x(S)$ and of a degenerate symplectic structure ω_{WP} on $\mathcal{T}^a(S)$. For surfaces S without cilia such structures are well known under the name of Weil–Petersson ones and can be defined as follows.

The spaces of homomorphisms of the fundamental group of a surface S into any reductive group G considered up to conjugation is known to have a canonical Poisson structure (see [9], [10] and references therein). Its symplectic leaves are numerated by conjugacy classes of elements corresponding to loops surrounding holes. The

Teichmüller space $\mathcal{T}^x(S)$ can be mapped to this space (for $G = \mathrm{PSL}(2, \mathbb{R})$) and if the surface S has no cilia the corresponding map is a local diffeomorphism in a vicinity of a generic point. This map induces the Poisson structure P_{WP} .

Another way of defining the same Poisson structure was given by William Goldman in [15]. For any closed loop one can associate the absolute value of the trace of the element of $\mathrm{PSL}(2, \mathbb{R})$ corresponding to it. Goldman gave an explicit expression for the Poisson bracket between function of this type as a linear combination of functions of the same type.

The image of $\mathcal{T}^a(S)$ in $\mathcal{T}^x(S)$ under the map p forgetting the horocycles corresponds to the representations of $\pi_1(S)$ in $\mathrm{PSL}(2, \mathbb{R})$ sending all loops surrounding holes to parabolic elements. It implies that the Poisson structure restricted to the image is non-degenerate. The inverse image under the map p of the corresponding symplectic structure defines a degenerate symplectic structure ω_{WP} on $\mathcal{T}^a(S)$.

The Poisson structure on $\mathcal{T}^x(S)$ for a ciliated surface S is given by the formula:

$$P_{\mathrm{WP}} = \sum_{\alpha\beta \in E(\Gamma)} \varepsilon^{\alpha\beta} x^\alpha \frac{\partial}{\partial x^\alpha} \wedge x^\beta \frac{\partial}{\partial x^\beta}. \quad (6.1)$$

The degenerate symplectic structure ω_{WP} on $\mathcal{T}^a(S)$ for a ciliated surface S is given by the formula:

$$\omega_{\mathrm{WP}} = \sum_{\alpha\beta \in E(\Gamma)} \varepsilon^{\alpha\beta} \frac{da_\alpha}{a_\alpha} \wedge \frac{da_\beta}{a_\beta}. \quad (6.2)$$

The main property of these Poisson and symplectic structures is that they are independent of the particular triangulation. This property can be checked explicitly by substitution of the transformation laws (4.2) (resp. (4.4)) into the expression for P_{WP} (resp. ω_{WP}), taking into account the transformation law (2.1) for the matrix $\varepsilon^{\alpha\beta}$.

Another useful property is the compatibility of the bracket with unramified covering of surfaces. If $\tilde{S} \rightarrow S$ is an unramified covering of degree N then there are obvious maps $i^x: \mathcal{T}^x(S) \rightarrow \mathcal{T}^x(\tilde{S})$ and $i^a: \mathcal{T}^a(S) \rightarrow \mathcal{T}^a(\tilde{S})$. Then one can easily see that $(i^a)^* \tilde{\omega}_{\mathrm{WP}} = N \omega_{\mathrm{WP}}$, where $\tilde{\omega}_{\mathrm{WP}}$ is the canonical degenerate symplectic form on $\mathcal{T}^x(\tilde{S})$. The map i^x is not a Poisson map, however it maps symplectic leaves to symplectic leaves. If $i^x: L \rightarrow \tilde{L}$, where L and \tilde{L} are the symplectic leaves of $\mathcal{T}^x(S)$ and $\mathcal{T}^x(\tilde{S})$ respectively, then $(i^x)^*(\tilde{P}_{\mathrm{WP}})^{-1}|_{\tilde{L}} = N(P_{\mathrm{WP}})^{-1}|_L$.

One can check that if the surface S has no cilia the formula gives the known Poisson and symplectic structures. The simplest way to do it is to compute explicitly the bracket between traces of matrices corresponding to closed loops using both our definition and Goldman's one. Due to the compatibility with coverings it is sufficient to check the coincidence of the brackets for two loops having one intersection point, where Goldman's expression is especially simple. We do not give the computation here since it is straightforward and technical.

The kernel of the symplectic form on the space \mathcal{T}_0^a is spanned by the vector fields corresponding to the action r_ρ , where ρ runs through the boundary components with even number of cilia. The kernel of the Poisson structure on the space \mathcal{T}^x is spanned by

the differentials of the functions r^ρ . In particular if the number of cilia is odd on every boundary component the symplectic and the Poisson structures are nondegenerate.

The kernel of the symplectic form on the space \mathcal{T}^a is generated by the orbits of resizing of all horocycle at a given boundary component by the same amount.

The lamination spaces \mathbb{T}^x and \mathbb{T}^a also have a canonical Poisson structure P_{WP} and the degenerate symplectic structure w_{WP} , respectively. They are given by

$$P_{WP} = \sum_{\alpha\beta \in E(\Gamma)} \varepsilon^{\alpha\beta} \frac{\partial}{\partial \mathbf{x}^\alpha} \wedge \frac{\partial}{\partial \mathbf{x}^\beta},$$

$$w_{WP} = \sum_{\alpha\beta \in E(\Gamma)} \varepsilon^{\alpha\beta} d\mathbf{a}_\alpha \wedge d\mathbf{a}_\beta.$$

All the properties of these structures are similar to the respective properties of the Poisson and symplectic structures on their Teichmüller counterparts.

7 Canonical pairing

Certain curves or collections of curves on a surface define functions on Teichmüller and lamination spaces. Namely:

1. A closed curve γ on S defines a function $\ell(\gamma)$ on $\mathcal{T}^a(S)$ as well as on $\mathcal{T}^x(S)$ by the length of the geodesic w.r.t. the curvature -1 metric in the homotopy class defined by the curve.
2. If a curve γ on S connects two points on the boundary it defines a function $\ell(\gamma)$ on $\mathcal{T}^a(S)$ in the following way. Move the ends of the curve which do not belong to holes, in the counterclockwise direction until they hit the cilia. Then take a geodesic in the same homotopy class and take the exponent of the signed half of the length of the segment between the corresponding horocycles like in Figure 12.
3. If a collection of curves $m = \{\gamma_i\}$ with rational weights $\{w_i\}$ has the property that for every segment of the boundary between cilia and for every hole the total weight of the curves hitting it vanishes, then it defines a function on $\mathcal{T}^x(S)$ in the following way. Move first the ends of curves in the counterclockwise direction to hit the cilia. Lift then the collection of curves to H and choose horocycles at their ends. Then take the weighted sum of signed lengths of geodesic segments between the horocycles $\ell(m) = \sum w_i \ell(\gamma_i)$. It is easy to see that this sum does not depend on the choice of the horocycles due to the vanishing property.

These functions can be considered as functions of two arguments: the first is a curve or a weighted collection of curves and the second is a point of the appropriate Teichmüller or lamination space. Our aim now is to extend the first argument to be not just a single curve, but a point of an appropriate space of laminations. The resulting functions of two arguments are called *pairings*.

There are two versions or pairings between Teichmüller and lamination spaces – additive and multiplicative. The main feature of the additive one is that it can be extended to real laminations as a continuous function. The multiplicative one is defined for integral laminations only. For every lamination this pairing defines a function on the Teichmüller space which is a Laurent polynomial in the coordinates with positive integral coefficients.

7.1 Additive canonical pairing

The additive canonical pairing is a function of two arguments. One argument is a rational (or, later on, real) lamination, which can be either from \mathbb{T}_0^a or \mathbb{T}^x , and the other is a point of the opposite type Teichmüller space, \mathcal{T}^x or \mathcal{T}_0^a , respectively.

$$\mathcal{I} : \mathbb{T}^x \times \mathcal{T}_0^a \rightarrow \mathbb{R},$$

$$\mathcal{I} : \mathcal{T}^x \times \mathbb{T}_0^a \rightarrow \mathbb{R}.$$

Abusing notation, we shall denote all of them by a single letter \mathcal{I} .

We will also define the *intersection pairing*:

$$l : \mathbb{T}^x \times \mathbb{T}_0^a \rightarrow \mathbb{R}.$$

Sometimes it is useful to consider the functions \mathcal{I} and l as functions of one variable taking another one as a parameter. In this case we shall put this parameter as a lower index. For example $\mathcal{I}_{m^x}(m^a) := \mathcal{I}(m^x, m^a)$.

The additive and the intersection pairings are defined as follows.

Definition.

- (1) Let $\gamma^x \in \mathbb{T}^x$ be a single closed curve and $m^a \in \mathcal{T}_0^a$. Then $\mathcal{I}(\gamma^x, m^a)$ is its length $\ell(\gamma^x)$ (Rule 1).
- (2) Let $m^x \in \mathcal{T}^x$ and let $\gamma^a \in \mathbb{T}^a$ be a single closed curve. Then $\mathcal{I}(m^x, \gamma^a)$ is equal to \pm a half of its length $\ell(\gamma^a)$ (Rule 1). The sign is $+$ unless γ^a surrounds a negatively oriented hole.
- (3) Let $\gamma^x \in \mathbb{T}^x$ be a curve connecting two points of the boundary of S and $m^a \in \mathcal{T}^a$. Then $\mathcal{I}(\gamma^x, m^a)$ is the signed length $\ell(\gamma^x)$ given by Rule 2.
- (4) Let $\gamma^x \in \mathbb{T}^x$ and $\gamma^a \in \mathbb{T}^a$ be two curves. Move the ends of γ^a (if any and but those belonging to holes) counterclockwise to hit cilia. Then $l(\gamma^x, \gamma^a)$ is a half of the minimal number of intersection points between γ^x and γ^a .
- (5) Let $m^x \in \mathcal{T}^x$ and let $m^a \in \mathbb{T}_0^a$ be a collection of curves. Then the function $\ell(l^a)$ defined by this Rule 3 is equal to $\mathcal{I}(m^x, m^a)$. (The vanishing property follows from the definition of \mathbb{T}_0^a .)
- (6) Let u, v be positive rational numbers. Let $m_1^x, m_2^x \in \mathbb{T}^x$ be two collections of nonintersecting curves such that no curve from m_1^x intersects m_2^x , and $m^a \in \mathcal{T}^a$.

Then $\mathcal{I}(um_1^x + vm_2^x, m^a) = u\mathcal{I}(m_1^x, m^a) + v\mathcal{I}(m_2^x, m^a)$. If $m^x \in \mathcal{T}^x$ and $m_1^a, m_2^a \in \mathcal{T}_0^a$ be nonintersecting collections of curves such that no curve from m_1^a intersects m_2^a then $\mathcal{I}(m^x, um_1^a + vm_2^a) = u\mathcal{I}(m^x, m_1^a) + v\mathcal{I}(m^x, m_2^a)$.

Properties of the additive canonical pairing. 1. The additive canonical pairings \mathcal{I} as well as the intersection pairing l are continuous functions.

2. $\mathcal{I}(p(m^a), m^a) = \mathcal{I}(p(m^a), m^a)$, where $m^a \in \mathcal{T}_0^a$ and $m^a \in \mathcal{T}_0^a$.

3. If a point m^x of the space \mathcal{T}^x has positive coordinates x^1, \dots, x^n and a point m^a of the Teichmüller space \mathcal{T}_0^a has coordinates a_1, \dots, a_n , then $\mathcal{I}(m^x, m^a) = \sum_{\alpha} x^{\alpha} \log a_{\alpha}$.

4. Let $m^x \in \mathcal{T}^x$ be a point of a Teichmüller space \mathcal{T}^x with coordinates x_1, \dots, x_n . Let C be a positive real number. Denote by $(m^x)^C \in \mathcal{T}^x$ a point with coordinates $(x_1)^C, \dots, (x_n)^C$ and by $m^x \in \mathcal{T}^x$ a lamination with coordinates $x^1 = \log x^1, \dots, x^n = \log x^n$. Then for any lamination $m^a \in \mathcal{T}^a$ we have

$$\lim_{C \rightarrow \infty} \mathcal{I}((m^x)^C, m^a) / C = l(m^x, m^a).$$

5. Analogously let $m^a \in \mathcal{T}_0^a$ be a point of Teichmüller space with coordinates a_1, \dots, a_n . Let C be a positive real number. Denote by $(m^a)^C \in \mathcal{T}_0^a$ the point in Teichmüller space with coordinates $(a_1)^C, \dots, (a_n)^C$ and by $m^a \in \mathcal{T}_0^a$ the lamination with coordinates $a_1 = \log a_1, \dots, a_n = \log a_n$. Then for any lamination $m^x \in \mathcal{T}^x$ we have

$$\lim_{C \rightarrow \infty} \mathcal{I}(m^x, (m^a)^C) / C = \mathcal{I}(m^x, m^a).$$

For closed surfaces there exists a canonical pairing $\mathcal{I} : \mathcal{T}^x \times \mathcal{T}_0^a \rightarrow \mathbb{R}$ defined by Francis Bonahon [2], such that the pairings defined here are its limits. Unfortunately we do not know any explicit construction of this pairing.

7.2 Multiplicative canonical pairing

The multiplicative canonical pairing is a function of two arguments. One argument is an integral lamination which can be either in $\mathcal{T}_0^a(\mathbb{Z})$ or $\mathcal{T}^x(\mathbb{Z})$, and the other is a point of the opposite type Teichmüller space, \mathcal{T}^x or \mathcal{T}_0^a , respectively.

$$\mathbb{I} : \mathcal{T}^x(\mathbb{Z}) \times \mathcal{T}_0^a \rightarrow \mathbb{R}_{>0},$$

$$\mathbb{I} : \mathcal{T}^x \times \mathcal{T}_0^a(\mathbb{Z}) \rightarrow \mathbb{R}_{>0}.$$

Abusing notations, we shall denote all of them by a single letter \mathbb{I} . Sometimes it is convenient to consider the function \mathbb{I} as a function of one variable taking another one as a parameter. In this case we shall put this parameter as the lower index. For example $\mathbb{I}_{m^x}(m^a) := \mathbb{I}(m^x, m^a)$.

The multiplicative pairing is defined by the following properties:

Definition.

- (1) Let $\gamma^x \in \mathbb{T}^x$ be a single closed curve with weight k and $m^a \in \mathcal{T}^a$. Then $\mathbb{I}(\gamma^x, m^a)$ is the absolute value of the trace of the monodromy of $(\gamma^x)^k$.
- (2) Let $m^x \in \mathbb{T}^x$ and let $\gamma^a \in \mathbb{T}^a$ be a single closed curve which is not retractable to a hole. Then $\mathbb{I}(\gamma^x, m^a)$ is the absolute value of the trace of the monodromy of $(\gamma^x)^k$.
- (3) Let $m^x \in \mathcal{T}^x$ and let $\gamma^a \in \mathbb{T}^a$ be a single closed curve which is retractable to a hole and the hole is oriented positively (resp. negatively). Then $\mathcal{I}(\gamma^x, m^a)$ is equal to the absolute value of the largest (resp. smallest) eigenvalue of the monodromy of $(\gamma^x)^k$.
- (4) Let $\gamma^x \in \mathbb{T}^x$ be a curve with weight k connecting two points of the boundary of S and $m^a \in \mathcal{T}^a$. Then $\mathbb{I}(\gamma^x, m^a) = \exp \mathcal{I}(\gamma^x, m^a)$.
- (5) Let $m^x \in \mathcal{T}^x$ and let $m^a \in \mathbb{T}_0^a$ be a collection of open curves with integral weights. Then $\mathbb{I}(m^x, m^a) = \exp \mathcal{I}(m^x, m^a)$.
- (6) Let u, v be positive integral numbers. Let $m_1^x, m_2^x \in \mathbb{T}^x$ be two collections of non-intersecting curves such that no curve from m_1^x intersects or coincides with a curve from m_2^x , and $m^a \in \mathcal{T}^a$. Then $\mathbb{I}(um_1^x + vm_2^x, m^a) = (\mathbb{I}(m_1^x, m^a))^u (\mathbb{I}(m_2^x, m^a))^v$.
If $m^x \in \mathcal{T}^x$ and $m_1^a, m_2^a \in \mathbb{T}_0^a$ are nonintersecting collections of curves such that no curve from m_1^a intersects or coincides with a curve from m_2^a , then $\mathbb{I}(m^x, um_1^a + vm_2^a) = (\mathbb{I}(m^x, m_1^a))^u (\mathbb{I}(m^x, m_2^a))^v$.

Properties of the multiplicative canonical pairing. 1. Let a_1, \dots, a_n be coordinates of a lamination $m^a \in \mathbb{T}_0^a(S, \mathbb{Z})$. Then the function \mathbb{I}_{m^a} on \mathcal{T}^x is a Laurent polynomial with positive integral coefficients of $(x^1)^{1/2}, \dots, (x^n)^{1/2}$, where x^1, \dots, x^n are the coordinates on \mathcal{T}^x . Its highest term is equal to $\prod_{\alpha} (x^{\alpha})^{a_{\alpha}}$ and the lowest one is $\prod_{\alpha} (x^{\alpha})^{-a_{\alpha}}$. Moreover \mathbb{I}_{m^a} is a Laurent polynomial of x^1, \dots, x^n if and only if the lamination m^a is even.

2. Let x^1, \dots, x^n be coordinates of a lamination $m^x \in \mathbb{T}^x(S, \mathbb{Z})$. Then the function \mathbb{I}_{m^a} on \mathcal{T}^x is a Laurent polynomial with positive integral coefficients of the coordinates a^1, \dots, a^n on \mathcal{T}_0^a .

3. Let $m^x \in \mathbb{T}^x(S, \mathbb{Z})$ be an integral \mathcal{X} -lamination and let $m^a \in \mathcal{T}_0^a$ be a point of the Teichmüller \mathcal{A} -space. Then $\lim_{C \rightarrow \infty} \log(\mathbb{I}(Cm^x, m^a))/C = \mathcal{I}(m^x, m^a)$. Analogously let $m^x \in \mathcal{T}^x$ be a point of the Teichmüller \mathcal{X} -space and let $m^a \in \mathbb{T}_0^a$ be an \mathcal{A} -lamination. Then $\lim_{C \rightarrow \infty} \log(\mathbb{I}(Cm^x, m^a))/C = \mathcal{I}(m^x, m^a)$.

4. Let $m^a \in \mathbb{T}_0^a$ be an integral bounded lamination. Then the function \mathbb{I}_{m^a} on \mathbb{T}^x is the tropical analogue of the function \mathbb{I}_{m^a} on \mathcal{T}^x in any coordinate system. Analogously let $m^x \in \mathbb{T}^x$ be an integral unbounded lamination. Then the function \mathbb{I}_{m^x} on \mathbb{T}_0^a is the tropical analogue of the function \mathbb{I}_{m^x} on \mathcal{T}^a in any coordinate system.

5. The coordinate functions on the space \mathcal{T}^a as well as \mathcal{T}_0^a are particular cases of the pairing, namely the value of the coordinate function a_{α} at the point m^a of the

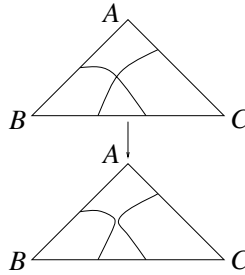


Figure 15. Rearranging the intersection points.

Teichmüller space \mathcal{T}^a is $\mathbb{I}(m^x, m^a)$, where m^x is the edge α of the triangulation with weight 1.

Conjecturally the converses to Properties 1 and 2 are also true. Namely, the only functions on \mathcal{T}^x (resp. \mathcal{T}_0^a) given by a Laurent polynomial with positive integral coefficients in every coordinate chart and undecomposable into a sum of two such polynomials are given by the multiplicative canonical pairing.

7.3 Proofs of main properties

Here we are going to give main ideas of the proofs of the most important properties of pairings: the continuity of the additive pairing (Property 1) and Properties 1 and 2 of the multiplicative pairing. Other properties are either obvious or follow from the proof of these three.

Continuity. Let us first prove the continuity of the pairing between \mathbb{T}^a and \mathcal{T}^x . To do this it suffices to prove it for laminations without curves with negative weights. Indeed, if we add such a curve to a lamination the length obviously changes continuously. We are going to show that the length of integral lamination is a convex function of its coordinates, i.e. that

$$\mathcal{I}(m^x, m_1^a) + \mathcal{I}(m^x, m_2^a) \leq \mathcal{I}(m^x, m_{1+2}^a) \quad (7.1)$$

where by m_{1+2}^a we have denoted a lamination with coordinates being sums of the respective coordinates of m_1^a and m_2^a . Taking into account the homogeneity of \mathcal{I} , one sees that Inequality (7.1) holds for all rational laminations and therefore can be extended by continuity to all real laminations.

Let us prove Inequality (7.1). Draw both laminations m_1^a and m_2^a on the surface and deform them to be geodesic. These laminations in general intersect each other in a finite number of points. Construct now a new collection of curves out of the union of m_1^a and m_2^a in the following way. For every intersection point consider the

triangle ABC of the triangulation it belongs to. Without loss of generality we may assume that the segments of the curves containing the intersection points connect the sides AB with BC and AC with BC , respectively. Then remove the neighbourhood of the intersection point and replace it by a union of two nonintersecting arcs in such a way that the segments of the resulting curves still connect AB with BC and AC with BC (Figure 15). The resulting collection of curves considered as a lamination is just m_{1+2}^a . The total length of the constructed curves can be made as close to $\mathcal{I}(m^x, m_1^a) + \mathcal{I}(m^x, m_2^a)$ as one desires. However it is not geodesic any longer. When we deform this collection to the geodesic one its length only decreases, thus giving the desired inequality (refined).

Laurent property and positivity. Now we are going to show that the functions \mathbb{I}_{m^a} and \mathbb{I}_{m^x} are given by Laurent polynomials with positive integral coefficients in any coordinate charts. We shall show it for laminations representable by a single curve and then extend the result to arbitrary laminations by multiplicativity.

Consider first a single closed curve with weight k as an element of the space \mathbb{T}^a . If M is the monodromy operator around this curve, then by definition $\mathbb{I}_{m^a} = \text{tr}(M^k)$. According to the construction of the Fuchsian group from Section 3.1 the operator M is given by the expression

$$M = B(x^{i_1})I^{\pm 1}B(x^{i_2}) \dots I^{\pm 1}B(x^{i_p})I^{\pm 1},$$

where i_1, \dots, i_p is the sequence of edges of the triangulation the path intersects and the signs for $I^{\pm 1}$ are chosen according to whether the path turns left or right passing through a triangle. Multiplying every $I^{\pm 1}$ by $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ from the left and every B by J from the right we do not change the product since $J^2 = 1$ in $\text{PSL}(2, \mathbb{R})$ and transform the expression for M into $M = H(x^{i_1})E^{\pm}H(x^{i_2}) \dots H(x^{i_p})E^{\pm}$, where $E^+ = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $E^- = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $H(x) = \begin{pmatrix} x^{1/2} & 0 \\ 0 & x^{-1/2} \end{pmatrix}$. Since all coefficients of these matrices are Laurent monomials with positive coefficients, the trace of any power of M is a Laurent polynomial with positive coefficients.

Now consider the case of the space $\mathcal{T}^a(S)$. The reconstruction rule implies that the traces of the monodromy matrices are Laurent polynomials of the coordinates in integral powers of the coordinates $\{a_\alpha\}$. Positivity follows from the existence of the monomial map $p: \mathcal{T}^a(S) \rightarrow \mathcal{T}^x(S)$.

We leave the proof of the Laurent property and positivity for non closed curves as an exercise.

Appendix A. Combinatorial description of $\mathcal{D}(S)$

Denote by $|\Gamma|(S)$ the set of combinatorial types of triangulations of S . For each element of $|\Gamma|(S)$ fix a *marking*, i.e., a numbering of its edges. Denote by $\Gamma(S)$ the set of isotopy classes of marked triangulations of S . The presence of the marking

changes the set of triangulations since some of them may have nontrivial symmetry group. Introducing the marking is a tool to remove this symmetry. (Here and below the vertical lines $|\cdot|$ indicate the diffeomorphism class.)

The mapping class group $\mathcal{D}(S)$ obviously acts freely on the space of marked triangulations with the space of combinatorial types of triangulations as a quotient,

$$\Gamma(S)/\mathcal{D}(S) = |\Gamma|(S).$$

Recall that a group can be thought of as a category with one object and invertible morphisms. Similarly, a *groupoid* is a category where all morphisms are isomorphisms, and any two objects are isomorphic. Since the automorphism groups of different objects of a groupoid are isomorphic, we can associate a group to a groupoid, well defined up to an isomorphism. We are going to construct a groupoid providing the mapping class group and admitting a simpler description by generators and relations than the mapping class group itself.

Definition. Let $|\Gamma|(S)$ be the set of objects. For any two triangulations $|\Gamma|, |\Gamma_1| \in |\Gamma(S)|$ let a morphism from $|\Gamma|$ to $|\Gamma_1|$ be a pair of triangulations of S of types $|\Gamma|$ and $|\Gamma_1|$ modulo the diagonal mapping class group action; we denote this morphism by $|\Gamma, \Gamma_1|$. For any three triangulations $\Gamma, \Gamma_1, \Gamma_2$, the composition of $|\Gamma, \Gamma_1|$ and $|\Gamma_1, \Gamma_2|$ is $|\Gamma, \Gamma_2|$. The described category is called the *modular groupoid*.

One can easily verify that (1) composition of morphisms is well defined; (2) the class of the pair of identical triangulations $|\Gamma, \Gamma|$ is the identity morphism and the inverse of the morphism $|\Gamma, \Gamma_1|$ is $|\Gamma_1, \Gamma|$; (3) the group of automorphisms of an object is the mapping class group $\mathcal{D}(S)$.

To give a description of the modular groupoid by generators and relations, we need to introduce distinguished sets of morphisms called *flips* and *symmetries*. Recall that a morphism $|\Gamma, \Gamma_\alpha|$ is a flip if the triangulation Γ_α is obtained from Γ by removing an edge α and replacing it by another diagonal of the resulting quadrilateral. We use the notation Γ_α to emphasize the relation of this triangulation to the triangulation Γ . Note that for a given triangulation Γ , several marked embedded graphs may be denoted by Γ_α because no marking of Γ_α is indicated.

To each symmetry σ of a triangulation Γ we assign an automorphism $|\Gamma, \sigma\Gamma|$.

There is no canonical bijection between the edges of different triangulations even if a morphism between them is given. However, if two triangulations are related by a flip, we can introduce such a bijection since all edges of one triangulation coincide with the edges of the other. We exploit this identification and denote the corresponding edges of different graphs by the same letter if it is clear which sequence of flips relating these graphs is considered. To avoid confusion we note that this identification has nothing to do with the marking.

In this notation, the triangulation $\Gamma_{\alpha_1 \dots \alpha_n}$ is the graph obtained as a result of consecutive flips $\alpha_n, \dots, \alpha_1$ of edges of a given graph Γ .

There are three kinds of relations between flips, which are satisfied for any choice of marking of the triangulations entering the relations.

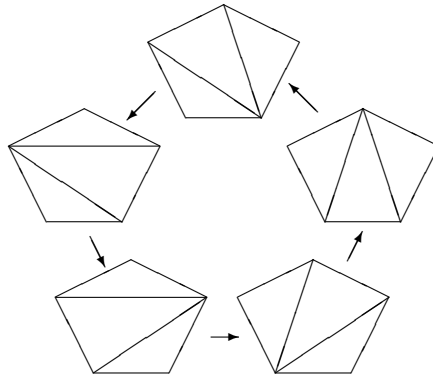


Figure 16. Pentagon relation.

Proposition. *A square of a flip is a symmetry: if $|\Gamma_\alpha, \Gamma|$ is a flip in an edge α , then $|\Gamma, \Gamma_\alpha|$ is also a flip and¹*

$$|\Gamma, \Gamma_\alpha||\Gamma_\alpha, \Gamma| = 1. \tag{R.2}$$

Flips in disjoint edges commute: if α and β are two edges, then

$$|\Gamma_{\alpha\beta}, \Gamma_\alpha||\Gamma_\alpha, \Gamma| = |\Gamma_{\alpha\beta}, \Gamma_\beta||\Gamma_\beta, \Gamma|. \tag{R.4}$$

Five consecutive flips in edges α and β having one common vertex is the identity: for such α and β , the triangulations $\Gamma_{\alpha\beta}$ and $\Gamma_{\beta\alpha}$ are related by a flip and

$$|\Gamma, \Gamma_\alpha||\Gamma_\alpha, \Gamma_{\beta\alpha}||\Gamma_{\beta\alpha}, \Gamma_{\alpha\beta}||\Gamma_{\alpha\beta}, \Gamma_\beta||\Gamma_\beta, \Gamma| = 1. \tag{R.5}$$

The proofs of Relations R.2 and R.4 are obvious. Relation R.5 is obvious from the triangulation of a pentagon shown in Figure 16.

Replacing the mapping class group by the modular groupoid, we can express the latter through generators and relations in a simpler way.

Note that a symmetry can be represented as a ratio of two flips in a given edge and the modular groupoid is therefore generated by the flips only. We do not describe relations between flips and graph symmetries in detail because they are quite obvious. In fact, the symmetry groups of Γ and Γ_α act transitively on the set of flips $|\Gamma, \Gamma_\alpha|$, and this action can be considered as relations between flips and symmetries.

The proposition can be proved using direct combinatorial methods of the simplicial topology. However, we give the main idea of another proof, which is more specific for dimension 2.

Proof of the proposition. To any connected simplicial complex \mathcal{S} we can associate a groupoid by taking a point in each top-dimensional simplex objects and the ho-

¹The notation $R.n$ indicates the number n of graphs entering this relation.

motopy classes of oriented paths connecting the chosen points as morphisms. The corresponding group is the fundamental group of the topological space given by the complex.

To any codimension one simplex we can associate two classes of paths (differing by orientation and having the identity morphism as their product) connecting adjacent top-dimensional simplices. It is natural to call them *flips*. We can associate a relation between flips to any codimension two simplex. It is obvious that this set of flips generates the groupoid and that the only relations between the flips are given by codimension two simplices.

The same is true for an orbifold simplicial complex, where we replace simplices by quotients of simplices by finite groups. In this case, we must choose one generic point per each top-dimensional simplex as an object and orbifold homotopy classes of paths as morphisms. The corresponding group is the orbifold fundamental group of the orbifold given by the complex. The groupoid is now generated by flips and groups of top dimensional simplices and still the only nontrivial relations are those given by codimension two simplices.

Consider now the orbifold simplicial decomposition of moduli space of complex structures on S by Strebel, Hubbard, Mazur and Penner [21]. The orbifold fundamental group of the moduli space \mathcal{M} is just the mapping class group $\mathcal{D}(S)$. Recall that Strebel orbisimplices are enumerated by ribbon graphs corresponding to S and the dimension of a simplex is equal to the number of its edges. One can easily see that the groupoid of the Strebel complex coincides with the modular groupoid. Moreover, the flips of the former correspond to the flips of the latter. The relations between flips are given by codimension two cells, which correspond either to graphs with two four-valent vertices (which produces Relation R.4) or to graphs with one five-valent vertex (which produces Relation R.5). Relation R.2 holds true for any simplicial complex. \square

Appendix B. Markov numbers

The set of Markov numbers is a remarkable collection of integers that appeared first in number theory but is related to lengths of closed geodesics on a holed torus with hyperbolic structures. In this appendix we shall give a simple description of these numbers as well as a proof of a few of their properties as an illustration of the technique developed here.

Consider a torus with one hole T . The space of homotopy classes of simple (i.e., without self intersections) unoriented closed paths on it can be parameterised by points of $\mathbb{Q}P^1$. Indeed, once we have chosen an orientation of the path, we can consider it as an element of the first homology group of T with compact support. It is also obvious that any simple (indivisible) class is represented by a unique simple oriented closed path. Since the first homology group is \mathbb{Z}^2 , it just gives the desired parameterisation.

Introduce the equiharmonic complex structure on T , i.e. the structure which has maximal symmetry group $\mathbb{Z}/3\mathbb{Z}$. For any closed path γ on T without self-intersections the numbers $X_\gamma = \frac{2}{3} \cosh l(\gamma)$, where $l(\gamma)$ are the lengths of the corresponding geodesic, are called *Markov numbers*.

The main properties of Markov numbers are the following:

1. Markov numbers are positive integers.
2. Markov numbers include Fibonacci numbers with even indices, 2, 5, 13, 34, 89, 233, ...

Call a *Markov triple* a triple of Markov numbers (X, Y, Z) corresponding to three geodesics having pairwise one intersection point.

3. Elements of a Markov triple satisfy the Markov equation:

$$X^2 + Y^2 + Z^2 = 3XYZ. \quad (1)$$

4. Any integer solution of this equation is a Markov triple.

5. For any Markov triple (X, Y, Z) the triples (Y, Z, X) and $(Z, Y - 3XZ, X)$ are also Markov triples. Any Markov triple can be obtained from the triple $(1, 1, 1)$ by a sequence of such transformations.

Since homotopy classes of closed non-selfintersecting curves can be parameterised by $\mathbb{Q}P^1$, one can choose affine coordinates on $\mathbb{Q}P^1$ in such a way that the curves with coordinates 0, 1 and ∞ have Markov numbers 1. Denote by $M(u)$ the Markov number corresponding to the curve with the coordinate $u \in \mathbb{Q}$.

6. The function $\psi\left(\frac{p}{q}\right) = \frac{1}{q} \operatorname{arcosh}\left(\frac{3}{2}M\left(\frac{p}{q}\right)\right)$, where $\gcd(p, q) = 1$, is extensible to a continuous convex function on \mathbb{R} .

$$7. M(x) = M(1-x) = M\left(\frac{1}{x}\right) = M\left(\frac{1}{1-x}\right) = M\left(\frac{x}{x-1}\right) = M\left(\frac{x-1}{x}\right)$$

8. For any closed geodesics γ on S there exists a unique geodesics γ' going from the puncture to the puncture which doesn't intersect γ . Let $l(\gamma')$ be the length of the piece of γ' between the intersection points with the horocycle surrounding the region of area 3. Then $e^{l(\gamma')} = M(\gamma)$.

9. (Markov conjecture). The famous unproven Markov conjecture says that two Markov numbers $M(x)$ and $M(y)$ are different unless x and y are related by transformations from Property 7.

Taking into account that the segment $[0, 1]$ is the fundamental domain of the action of transformations from Property 7, one can reformulate the Markov conjecture as follows: if $M(x) = M(y)$ and $x, y \in [0, 1]$ then $x = y$.

Proofs of the properties (unfortunately, without the last one and Property 4). There is only one combinatorial triangulation corresponding to the holed torus. It has one vertex, three edges and two triangles. This triangulation has obvious $\mathbb{Z}/3\mathbb{Z}$ symmetry group cyclically permuting the edges. Let x, y, z be the corresponding coordinates on the Teichmüller space $\mathcal{T}^x(S)$.

A closed curve on S can be considered as a bounded lamination if we assign to it the weight 1. The standard coordinates of such laminations are given by three nonnegative integers n_1, n_2, n_3 . These three numbers have no common factor, because

otherwise the weight of the curve would be greater than 1. On the other hand one of the numbers should be a sum of the two others since otherwise there would be a component surrounding the hole. The relation between this parameterisation by n_1, n_2, n_3 and the parameterisation by $\mathbb{Q}P^1$ described above is given by

$$x = \begin{cases} \frac{-n_2}{n_1} & \text{if } n_3 = n_1 + n_2, \\ \frac{n_2}{n_1} & \text{if } n_1 = n_2 + n_3 \text{ or } n_2 = n_3 + n_1. \end{cases} \tag{2}$$

Denote by Z, X and Y one thirds of the traces of the elements of the Fuchsian group corresponding to the curves with coordinates $(1, 1, 0), (0, 1, 1)$ and $(1, 0, 1)$ respectively. They can be easily computed using the explicit formulae for the Fuchsian group:

$$\begin{aligned} Z &= \frac{1}{3}(x^{1/2}y^{1/2} + x^{1/2}y^{-1/2} + x^{-1/2}y^{-1/2}), \\ X &= \frac{1}{3}(y^{1/2}z^{1/2} + y^{1/2}z^{-1/2} + y^{-1/2}z^{-1/2}), \\ Y &= \frac{1}{3}(z^{1/2}x^{1/2} + z^{1/2}x^{-1/2} + z^{-1/2}x^{-1/2}). \end{aligned} \tag{3}$$

Using these expressions we can verify the equality

$$X^2 + Y^2 + Z^2 - 3XYZ = -\frac{1}{9}(xyz - 2 + (xyz)^{-1}). \tag{4}$$

The symmetry of the graph obviously cyclically permutes the coordinates and therefore the numbers Z, X, Y . A flip of an edge acts by Rule (3.3) and it results in the mapping

$$(Z, X, Y) \mapsto (X, 3YZ - X, Z). \tag{5}$$

If all three coordinates x, y, z are ones, the corresponding complex surface is just the equiharmonic punctured torus.

Properties 1, 3, 5, 6 immediately follows from this picture. One can easily check that $M(n)$ for $n \in \mathbb{N}$ are just the Fibonacci numbers which gives Property 2. The property 7 is an immediate consequence of the convexity property of the lamination length function. The property 4 was proved by Markov himself.

Property 8 stands a little apart from the others since it is related to the spaces $\mathcal{T}^a(S)$ and $\mathcal{T}^x(S)$ rather than $\mathcal{T}^a(S)$ and $\mathcal{T}^x(S)$ respectively. Consider a coordinate system U, V, W on $\mathcal{T}^a(S)$. Let $A = A_\rho$ be the area of the region inside the only horocycle ρ . It easily follows from the expression for the area

$$(U^2 + V^2 + W^2) = UVWA. \tag{6}$$

The cyclic symmetry of the triangulation acts by cyclic permutation of U, V, W . A flip of an edge acts by

$$(U, V, W) \mapsto \left(W, \frac{U^2 + W^2}{V}, U \right). \tag{7}$$

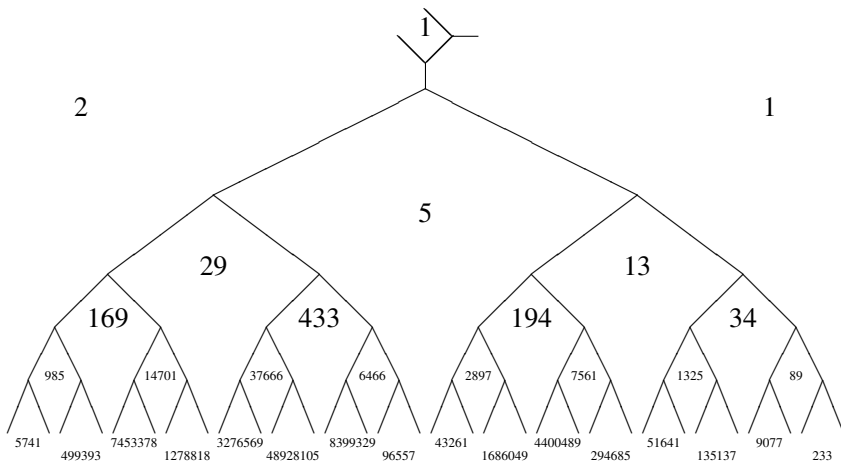
On the other hand this transformation law can be rewritten taking into account Equation (6):

$$(U, V, W) \mapsto (W, UWA - V, U). \quad (8)$$

This rule coincides with (5) for $A = 3$.

Now consider the decorated surface with $U = V = W = 1$. This is the surface with the area inside the horocycle $A = 3$. Applying modular transformations we get obviously the Markov triples, which proves the property 8. \square

There exists a canonical decomposition (called *Farey tessellation*) of the upper half plane H into ideal triangles with vertices in all rational points of its ideal boundary. The dual graph to this tessellation is the universal three-valent tree. The faces of this tree are therefore in one-to-one correspondence with rational numbers. On the pictures below we have drawn a fragment of this tree with corresponding Markov numbers written on the faces.



As a concluding remark of this section note that, as it was observed by A. Bondal, Markov triples are dimensions of elements of distinguished sets of sheaves on $\mathbb{C}P^2$. The relations between these two ways of obtaining Markov numbers are completely unclear and very exciting.

References

- [1] W. Abikoff, *The real analytic theory of Teichmüller space*. Lecture Notes in Math. 820, Springer-Verlag, Berlin 1980. 659
- [2] F. Bonahon, Teichmüller space and geodesic currents. *Invent. Math.* 92 (1988), 139–162. 673

- [3] S. Choi and W. Goldman, The deformation spaces of convex \mathbb{RP}^2 structures on 2-orbifolds. The deformation spaces of convex \mathbb{RP}^2 -structures on 2-orbifolds. *Amer. J. Math.* 127 (5) (2005), 1019–1102. [648](#)
- [4] V. V. Fock, Dual Teichmüller spaces. Preprint, 1998; arXiv:dg-ga/9702018. [648](#)
- [5] V. V. Fock, L. O. Chekhov, Quantum Teichmüller spaces. *Teoret. Mat. Fiz.* 120 (3) (1999), 511–528; English transl. *Theoret. and Math. Phys.* 120 (3) (1999), 1245–1259. [648](#)
- [6] V. V. Fock and A. B. Goncharov, Moduli spaces of local systems and higher Teichmüller theory. *Publ. Math. Inst. Hautes Études Sci.* 103 (2006), 1–211. [648](#)
- [7] V. V. Fock, A. B. Goncharov, Cluster ensembles, quantization and the dilogarithm. Preprint, 2004; arXiv:math/0311245. [648](#), [649](#)
- [8] V. V. Fock, A. B. Goncharov, Moduli spaces of convex real projective structures on surfaces. Preprint, 2006; arXiv:math.DG/0405348. [648](#)
- [9] V. V. Fock and A. A. Rosly, Poisson structures on moduli of flat connections on Riemann surfaces and the r -matrix. In *Moscow Seminar in Mathematical Physics*, Amer. Math. Soc. Transl. Ser. 2 191, Amer. Math. Soc., Providence, RI, 1999, 67–86. [669](#)
- [10] V. V. Fock and A. A. Rosly, Flat connections and polyubles. *Theoret. and Math. Phys.* 95 (1993), 526–535. [669](#)
- [11] S. Fomin and A. Zelevinsky, Cluster algebras. *J. Amer. Math. Soc.* 15 (2) (2002), 497–529. [648](#)
- [12] S. Fomin and A. Zelevinsky, Cluster algebras: Notes for the CDM-03 conference. In *Current developments in mathematics, 2003*, International Press, Somerville, MA, 2003, 1–34. [649](#)
- [13] M. Gekhtman, M. Shapiro, A. Vainshtein, Cluster algebras and Weil-Petersson forms. *Duke Math. J.* 127 (2) (2005), 291–311.
- [14] W. M. Goldman, Convex real projective structures on surfaces. *J. Differential Geom.* 31 (1990), 795–845. [648](#)
- [15] W. M. Goldman, The symplectic nature of fundamental groups of surfaces. *Adv. Math.* 54 (2) (1984), 200–225. [670](#)
- [16] N. J. Hitchin, Lie groups and Teichmüller space. *Topology* 31 (3) (1992), 449–473. [648](#)
- [17] R. M. Kashaev, Quantum dilogarithm and Quantisation of Teichmüller space. *Lett. Math. Phys.* 43 (2) (1998), 105–115. [648](#)
- [18] M. L. Kontsevich, Intersection theory on the moduli space of curves. *Funct. Anal. Appl.* 25 (2) (1991), 123–129. [664](#)
- [19] S. Krushkal, B. Apanasov and N. Gusevskii, *Kleinian groups and uniformization in examples and problems*. Transl. Math. Monogr. 62, Amer. Math. Soc., Providence, R.I., 1986. [659](#)
- [20] A. Papadopoulos and R. C. Penner, La forme symplectique de Weil-Petersson et le bord de Thurston de l'espace de Teichmüller. *C. R. Acad. Sci. Paris Sér. I Math.* 312 (11) (1991), 871–874. [656](#)
- [21] R. C. Penner, The decorated Teichmüller space of Riemann surfaces. *Comm. Math. Phys.* 113 (1988), 299–339. [648](#), [664](#), [679](#)

- [22] R. C. Penner, Weil-Peterson volumes in moduli spaces. Preprint, Inst. Mittag-Leffler, 1989.
- [23] K. Strebel, *Quadratic differentials*. Ergeb. Math. Grenzgeb. (3) 5, Springer-Verlag, Berlin 1984.
- [24] J. Teschner, On the relation between quantum Liouville theory and the quantized Teichmüller spaces. *Internat. J. Modern Phys. A* 19 (2004), 459–477. [648](#)
- [25] W. Thurston, *The geometry and topology of three-manifolds*. Princeton University Notes, <http://www.msri.org/publications/books/gt3m>. [647](#), [648](#), [652](#), [669](#)

Chapter 16

An analog of a modular functor from quantized Teichmüller theory

Jörg Teschner

Contents

1	Introduction	687
1.1	Motivation	687
1.2	Aims	689
1.3	Overview	690
1.4	Outlook	691
Part I. Coordinates for the Teichmüller spaces		693
2	The Penner coordinates	693
2.1	Triangulations and fat graphs	693
2.2	Penner coordinates	695
2.3	Fock coordinates	695
3	The Ptolemy groupoids	697
3.1	Groupoids vs. complexes	698
3.2	Change of the triangulation	698
3.3	The representation of the mapping class group	700
4	Teichmüller space as the phase space of a constrained system	700
4.1	Kashaev's coordinates	700
4.1.1	The constraints	701
4.1.2	Change of fat graph	701
4.2	The structure of the Kashaev space W_φ	702
4.2.1	Fock variables vs. Kashaev's variables	702
4.2.2	Splitting of W_φ	702
5	The Fenchel–Nielsen coordinates	703
5.1	Definition of the Fenchel–Nielsen coordinates	704
5.2	Symplectic structure	705
5.3	Geodesic lengths from the Penner coordinates	705
6	Coordinates for surfaces with holes of finite size	707
6.1	Useful fat graphs on surfaces with holes of finite size	707
6.2	Kashaev type coordinates	708
Part II. Quantization of the Teichmüller spaces		709
7	Quantization of the Teichmüller spaces	709

7.1	Canonical quantization	709
7.2	Quantization of the Kashaev space W_φ	711
7.3	Reduction to the quantized Teichmüller spaces	712
8	Representations of the mapping class groups	714
8.1	Projective unitary representations of groupoids	714
8.2	Representations of symmetries of a groupoid	715
8.3	The projective representation of the Ptolemy groupoid on $\mathcal{K}(\varphi)$	717
8.4	Reduction to the quantized Teichmüller spaces	717
Part III. A stable modular functor from the quantum Teichmüller spaces		720
9	The notion of a stable unitary modular functor	721
9.1	Rigged Riemann surfaces	721
9.2	Stable unitary modular functors	722
9.3	Representations of the modular groupoid versus modular functors	724
10	The modular groupoid	726
10.1	Notations and conventions	726
10.2	Generators	726
10.3	Relations	728
11	From markings to fat graphs	731
11.1	The complex $\mathcal{M}^{\text{ad}}(\Sigma)$	731
11.2	Separated variables	733
11.3	Quantized Teichmüller spaces for surfaces with holes	735
12	Geodesic length operators	736
12.1	Overview	736
12.2	Construction of the length operators	737
12.3	Spectrum	739
12.4	Relation with the Dehn twist generator	740
13	Passage to the length representation	741
13.1	The length representation	741
13.2	Construction of the operator $D(\sigma)$	742
14	Realization of $M(\Sigma)$	743
14.1	Two constructions for the generators	743
14.2	The main result	744
14.3	Concluding remarks	746
Appendix A. The special functions $e_b(x)$ and $s_b(x)$		747
Appendix B. Operator-theoretical background		748
Appendix C. On the proof of Theorem 2		749
Appendix D. Proof of Proposition 10		751
Appendix E. On the proof of Theorem 6		753
E.1	A technical preliminary	753
E.2	Invariance of length operators	755
References		757

1 Introduction

The program of the quantization of the Teichmüller spaces $\mathcal{T}(\Sigma)$ of Riemann surfaces Σ which was started in [15], [17] and independently in [30]¹ is motivated by certain problems and conjectures from mathematical physics. One of the main aims of this program is to construct a one-parameter family of maps

$$\Sigma \longrightarrow (\mathcal{H}_b^\mathcal{T}(\Sigma), \mathbf{O}_b^\mathcal{T}(\Sigma), \mathbf{M}_b^\mathcal{T}(\Sigma)), \tag{1.1}$$

where

- (i) b is a deformation parameter, related to the traditional \hbar via $b = \sqrt{\hbar}$,
- (ii) Σ is a two-dimensional topological surface possibly with boundary,
- (iii) $\mathcal{H}_b^\mathcal{T}(\Sigma)$ is a Hilbert-space (possibly infinite-dimensional),
- (iv) $\mathbf{O}_b^\mathcal{T}(\Sigma)$ is an algebra of bounded operators on $\mathcal{H}_b^\mathcal{T}(\Sigma)$ and,
- (v) $\mathbf{M}_b^\mathcal{T}(\Sigma)$ is a unitary projective representation of the mapping class group of Σ on $\mathcal{H}_b^\mathcal{T}(\Sigma)$.

The data $(\mathbf{O}_b^\mathcal{T}(\Sigma), \mathbf{M}_b^\mathcal{T}(\Sigma))$ are restricted by the requirement that a suitably defined limit of $\mathbf{O}_b^\mathcal{T}(\Sigma)$ for $b \rightarrow 0$ should reproduce the commutative algebra of functions on the Teichmüller space $\mathcal{T}(\Sigma)$, whereas a natural limit $b \rightarrow 0$ of the automorphisms of $\mathbf{O}_b^\mathcal{T}(\Sigma)$ which are induced by the representation $\mathbf{M}_b^\mathcal{T}(\Sigma)$ should correspond to the natural action of the mapping class group $\text{MC}(\Sigma)$ on $\mathcal{T}(\Sigma)$.

1.1 Motivation

Motivation for studying this problem comes from mathematical physics. A conjecture of H. Verlinde [53] can be formulated very schematically as the statement that

$$(\mathcal{H}_b^\mathcal{T}(\Sigma), \mathbf{M}_b^\mathcal{T}(\Sigma)) \simeq (\mathcal{H}_c^L(\Sigma), \mathbf{M}_c^L(\Sigma)), \tag{1.2}$$

where

- (i) $\mathcal{H}_c^L(\Sigma)$ has a definition in terms of the representation theory of the Virasoro algebra with central charge c as the so-called space of conformal blocks associated to Σ ,

and

- (ii) $\mathbf{M}_c^L(\Sigma)$ is an action of the mapping class group $\text{MC}(\Sigma)$ on $\mathcal{H}_c^L(\Sigma)$ which is canonically associated to the representation-theoretic definition of $\mathcal{H}_c^L(\Sigma)$.

Part of the interest in the space $\mathcal{H}_c^L(\Sigma)$ from the side of mathematical physics is due to the fact that the elements of $\mathcal{H}_c^L(\Sigma)$ represent the basic building blocks in the so-called Liouville conformal field theory [49]. Deep connections between the perturbative approach to quantum Liouville theory on the one hand and Teichmüller

¹See [18], [10] for recent generalizations.

theory on the other hand have been exhibited by Takhtajan, Zograf and Teo see e.g. [48] and references therein.

This conjecture may be seen as a non-compact analog of similar relations between the quantization of moduli spaces of flat connections on Riemann surfaces on the one hand, and rational conformal field theories on the other hand. For K being a compact group, the geometric quantization of the moduli space $\mathcal{M}_K(\Sigma)$ of flat K -connections on a Riemann surface was performed in [27], [1]. Alternative approaches were based on more explicit descriptions of the symplectic structure on $\mathcal{M}_K(\Sigma)$ [19], [2], [3], [8], [4].² In either case one may schematically describe one of the main results of these constructions as an assignment

$$\Sigma \longrightarrow (\mathcal{H}_k^{\mathcal{M}}(\Sigma), M_k^{\mathcal{M}}(\Sigma)), \tag{1.3}$$

where

- (i) $\mathcal{H}_k^{\mathcal{M}}(\Sigma)$ is a finite-dimensional vector space,
 - (ii) $M_k^{\mathcal{M}}(\Sigma)$ is a projective representation of the mapping class group of Σ on $\mathcal{H}_k^{\mathcal{M}}(\Sigma)$.
- Part of the interest in these results was due to the close relations between the representation $M_k^{\mathcal{M}}(\Sigma)$ and the Reshetikhin–Turaev invariants of three manifolds [43]. Another source of interest were the relations to rational conformal field theory, which were predicted in [56], see [47] for a review of mathematical approaches to the problem and further references. These relations may, again schematically, be summarized as the existence of canonical isomorphisms

$$\begin{array}{ccc}
 & (\mathcal{H}_k^{\mathcal{M}}(\Sigma), M_k^{\mathcal{M}}(\Sigma)) & \\
 \swarrow \cong & & \nwarrow \cong \\
 (\mathcal{H}_k^{\mathcal{W}}(\Sigma), M_k^{\mathcal{W}}(\Sigma)) & \xleftarrow{\cong} & (\mathcal{H}_k^{\text{RT}}(\Sigma), M_k^{\text{RT}}(\Sigma))
 \end{array} \tag{1.4}$$

where

- (i) $\mathcal{H}_k^{\mathcal{W}}(\Sigma)$ is the space of conformal blocks in the WZNW-model associated to the compact group K , which can be defined in terms of the representation theory of the affine Lie algebra $\hat{\mathfrak{g}}_k$ with level k associated to the Lie algebra \mathfrak{g} of K ,
- (ii) $M_k^{\mathcal{W}}(\Sigma)$ is the natural action of the mapping class group on $\mathcal{H}_k^{\mathcal{W}}(\Sigma)$, which can be defined by means of the monodromy representation of the Knizhnik–Zamolodchikov connection,
- (iii) $\mathcal{H}_k^{\text{RT}}(\Sigma)$ is the space of invariants in certain tensor products of representations of the quantum group $\mathcal{U}_q(\mathfrak{g})$,
- (iv) $M_k^{\text{RT}}(\Sigma)$ is the mapping class group representation on $\mathcal{H}_k^{\text{RT}}(\Sigma)$ defined by the construction of Reshetikhin–Turaev.

²The equivalence between the different quantization schemes has not been discussed in detail so far. It boils down to the verification that the monodromy representation of the KZ-connection constructed within the geometric quantization framework of [27], [1] is equivalent to the mapping class group representation defined in [2], [3], [8]. It seems that e.g. combining the results of [36] and [6] does the job.

The quantization program (1.1) can be seen as a non-compact analog of (1.3) in the following sense. In (1.1) the role of the moduli space of flat connections $\mathcal{M}_K(\Sigma)$ is taken by the Teichmüller space $\mathcal{T}(\Sigma)$, which can be identified with the component in the moduli space $\mathcal{M}_G(\Sigma)$ of flat $G = \mathrm{SL}(2, \mathbb{R})$ -connections that has maximal Euler class [26], [23]. Moreover, the natural symplectic structure on the moduli space of flat $\mathrm{SL}(2, \mathbb{R})$ -connections restricts to the Weil–Petersson symplectic form on $\mathcal{T}(\Sigma)$ [22]. A quantization of the Teichmüller space $\mathcal{T}(\Sigma)$ may therefore be regarded as providing a quantization of a topological component in the moduli space $\mathcal{M}_G(\Sigma)$.

We expect that any non-compact counterpart of the developments mentioned above will be mathematically at least as rich as the already known results associated to compact groups K . In particular we expect that certain analogs of the constructions of Reshetikhin–Turaev and/or Turaev–Viro will capture information on the geometry of hyperbolic three manifolds, similar and probably related to the appearance of hyperbolic volumes in the asymptotic behavior of certain link invariants [29].

1.2 Aims

A major step towards establishing H. Verlinde’s conjecture (1.2) is to show that the quantization of the Teichmüller spaces (1.1) as initiated in [15], [17], [30] produces an analog of a *modular functor*. The basic data of a modular functor are assignments such as (1.3), which are required to satisfy a natural set of axioms as discussed in Section 9. One of the most important implications of the axioms of a modular functor are simple relations between the representations of the mapping class groups associated to Σ and $\Sigma^{\dagger c}$ respectively, where $\Sigma^{\dagger c}$ is the surface that is obtained from Σ by cutting along a simple closed curve c . These relations imply that the representation $M_k^{\mathcal{M}}(\Sigma)$ restricts to – and is generated by – the representations $M_k^{\mathcal{M}}(\Sigma')$ which are associated to those subsurfaces Σ' that can be obtained from Σ by cutting along a set of non-intersecting simple closed curves. This crucial *locality* property can be seen as the hard core of the notion of a modular functor.

Within the formalisms introduced in [15], [17], [30]–[32] it is far from obvious that the quantization of Teichmüller spaces constructed there has such properties. To show that this is indeed the case is the main problem solved in this chapter. The representation $M_b^{\mathcal{P}}(\Sigma)$ constructed and investigated in [30]–[32] is obtained by exploiting the fact that the mapping class group can be embedded into the so-called Ptolemy groupoid associated to the transformations between different triangulations of a Riemann surface Σ . A representation of the Ptolemy groupoid is constructed in [30]–[32], which then canonically induces a projective unitary representation of the mapping class group $\mathrm{MC}(\Sigma)$. The simplicity of the Ptolemy groupoid, which underlies the elegance of the constructions in [30]–[32] now turns out to cause a major problem from the point of view of our aims, since the above-mentioned locality properties implied in the notion of a modular functor are not transparently realized by the Ptolemy groupoid.

Essentially our task is therefore to go from triangulations to pants decompositions, which is the type of decomposition of a Riemann surface Σ that is naturally associated

to the concept of a modular functor. This requires to construct a change of representation for $\mathcal{H}_b^\sigma(\Sigma)$ from the one naturally associated to triangulations of Σ [30]–[32] to another one which is associated to pants decompositions. The main tool for doing this are the geodesic length operators introduced and studied in [15], [16] and [32], [33], which are the observables on the quantized Teichmüller spaces that are associated with the geodesic length functions³ on the classical Teichmüller spaces. The length operators associated to a maximal set of non-intersecting simple closed curves turn out to furnish a set of commuting self-adjoint operators, and the simultaneous diagonalization of these operators defines the sought-for change of representation.

There is a natural groupoid associated with the transformations between different pants decompositions. Of particular importance for us will be a certain refinement of this groupoid which will be called the modular groupoid $M(\Sigma)$. The modular groupoid $M(\Sigma)$ has been introduced for the study of rational conformal field theories by Moore and Seiberg in [38], and it was further studied in particular in [5]. Constructing a modular functor is essentially equivalent to constructing a tower of representations of the modular groupoid. Our main aim in the present exposition will be to show that the quantization of Teichmüller spaces allows one to construct a tower of representations of the modular groupoid by unitary operators in a natural way.

1.3 Overview

This chapter has three main parts. The first of these parts collects the necessary results from the “classical” theory of Riemann surfaces. This includes a review of two types of coordinate systems for the Teichmüller spaces $\mathcal{T}(\Sigma)$, one of which is associated to triangulations of Σ , the other to pants decompositions. The coordinates associated to triangulations were first introduced by Penner in [39]. We will also need to discuss variants of these coordinates due to Fock [15] and Kashaev [30] respectively. The changes of the underlying triangulation of Σ generate a groupoid, the Ptolemy groupoid $\text{Pt}(\Sigma)$, which has a useful representation in terms of generators and relations (Theorem 2).

The coordinates associated to pants decompositions are the classical Fenchel–Nielsen coordinates, which we review briefly in §5.1. We furthermore explain how the coordinates of Penner [39] and Kashaev [30], which were originally introduced to parameterize the Teichmüller spaces of surfaces Σ with punctures only, can be used to provide coordinates also for the case where the surface Σ has holes represented by geodesics of finite length.

The material in this part is mostly known, but it is scattered over many places in the literature, and some basic results were stated in the original references without a proof. We have therefore tried to give a reasonably self-contained and complete

³See [57], [58], [59] for some classical work on the symplectic nature of the Fenchel–Nielsen coordinates which represents important background for our results.

presentation of the relevant material, providing proofs where these are not available elsewhere.

The second part gives a largely self-contained presentation of the foundations of the quantization of Teichmüller spaces. Our presentation is heavily inspired by [30]–[32], but we deviate from these references in some important points. The treatment presented in this chapter seems to be the first complete and mathematically rigorous formulation of the quantum theory of the Teichmüller spaces.

The main aims of this chapter are finally achieved in the third part. We begin in §9 by introducing the notion of a stable unitary modular functor, and by explaining why having a stable unitary modular functor is equivalent to having a tower of unitary projective representations of the modular groupoid.

In §10 we will reformulate the main result of [38], [5] concerning the description of $M(\Sigma)$ in terms of generators and relations in a way that is convenient for us.

Of particular importance for us will be §11, where important first relations between certain subgroupoids of $M(\Sigma)$ and $Pt(\Sigma)$ are observed.

In §12 we define the geodesic length operators and establish their main properties. These results are of independent interest since some important properties of the geodesic length operators had not been proven in full generality before.

A key step in our constructions is taken in §13 by constructing a change of representation from the original one to a representation in which the length operators associated to a pants decomposition are simultaneously diagonalized. An important feature of this construction is the fact that the unitary operator which describes the change of representation factorizes into operators associated to the individual three holed spheres (trinions) which appear in a pants decomposition.

In §14 we construct the corresponding representation of the modular groupoid $M(\Sigma)$. The operators which represent $M(\Sigma)$ are constructed out of compositions of the representatives for the transformations in $Pt(\Sigma)$. This makes it relatively easy to verify the relations of $M(\Sigma)$, but the price to pay is that some crucial locality properties are more difficult to prove.

1.4 Outlook

In a sequel [51] to this exposition we will calculate the matrix coefficients of the operators which generate the representation of the modular groupoid explicitly. A close relation to the modular double $\mathcal{D}\mathcal{U}_q(\mathfrak{sl}(2, \mathbb{R}))$ of $\mathcal{U}_q(\mathfrak{sl}(2, \mathbb{R}))$ as defined and studied in [12], [42], [9] will be found.

It should be noted that $(\mathcal{H}_b^{\mathcal{J}}(\Sigma), M_b^{\mathcal{J}}(\Sigma))$ will not satisfy all the usual axioms of a modular functor, which require, in particular, that the vector space $\mathcal{V}(\Sigma)$ assigned to each Riemann surface should be finite-dimensional. Most importantly, however, the assignment $\Sigma \rightarrow (\mathcal{H}_b^{\mathcal{J}}(\Sigma), M_b^{\mathcal{J}}(\Sigma))$ was up to now only constructed for surfaces Σ which have at least one boundary component.

What will allow us to overcome this unsatisfactory feature are the remarkable analytic properties that the matrix coefficients of the operators which represent the

modular groupoid will be shown to have. It turns out that the mapping class group representation $M_b^\sigma(\Sigma)$ assigned to a Riemann surface with a boundary represented by geodesics of a fixed length depends *analytically* on the values of these lengths. The analytic properties of the matrix coefficients will furthermore allow us to “close a hole” by taking a limit where the length parameter assigned to this boundary component approaches a certain *imaginary* value. It will be shown in [51] that the resulting mapping class group representation is equivalent to the one on the surface which is obtained by gluing a disc into the relevant boundary component.

Concerning the representation theoretic side of H. Verlinde’s conjecture (1.2) it should be mentioned that a complete mathematical construction of $(\mathcal{H}_c^L(\Sigma), M_c^L(\Sigma))$ is not available so far, but nontrivial steps in the direction of constructing and describing $(\mathcal{H}_c^L(\Sigma), M_c^L(\Sigma))$ precisely have been taken in [50] in the case of surfaces Σ of genus zero. This includes in particular the derivation of explicit formulae for a set of basic data which characterize the resulting representation of the braid group uniquely.

The explicit computation of the matrix coefficients of the operators which generate the representation of the modular groupoid carried out in [51] will therefore allow us to verify H. Verlinde’s conjecture (1.2) in the case of Riemann surfaces of genus zero.

We furthermore expect that it should be possible to construct non-compact analogs of $(\mathcal{H}_k^{\text{RT}}(\Sigma), M_k^{\text{RT}}(\Sigma))$ based the non-compact quantum group $\mathcal{D}\mathcal{U}_q(\mathfrak{sl}(2, \mathbb{R}))$, and thereby complete a non-compact analog of the triangle (1.4).

Acknowledgements. The author would like to thank L. Chekhov, V. Fock and especially R. Kashaev for useful discussions.

We furthermore gratefully acknowledge financial support from the DFG by a Heisenberg fellowship, as well as the kind hospitality of the Humboldt University Berlin, the School of Natural Sciences of the IAS Princeton, the Physics Department of Chicago University and the Caltech, Pasadena.

Part I. Coordinates for the Teichmüller spaces

We will consider two-dimensional surfaces Σ with genus $g \geq 0$ and $s \geq 1$ boundary components such that

$$M \equiv 2g - 2 + s > 0.$$

On Σ we will consider metrics of constant negative curvature -1 . Our main interest will be the case where the boundary components can be represented by geodesics of finite length. Such boundary components will also be called holes in the following. However, to begin with we will focus on the case where the boundary components are *punctures*, i.e. holes of vanishing geodesic circumference.

The space of deformations of the metrics of constant negative curvature is called the Teichmüller space $\mathcal{T}(\Sigma)$. It will be of basic importance for us to have useful systems of coordinates for $\mathcal{T}(\Sigma)$.

We will consider two classes of coordinate systems which are associated to two types of graphs drawn on the Riemann surfaces respectively. The first class of coordinates goes back to Penner and is associated to triangulations of the Riemann surface or the corresponding dual graphs, the so-called fat graphs. We will also describe two useful variants of the Penner coordinates due to Kashaev and Fock respectively.

The second class of coordinates are the classical Fenchel–Nielsen length-twist coordinates. One may view them as being associated to a second type of graph on a surface Σ called *marking* that determines in particular a decomposition of the surface into three-holed spheres (trinions).

In the following sections of the first part of this chapter we shall describe these coordinate systems in some detail, discuss the graphs on Σ that these coordinates are associated to, as well as the groupoids generated by the transformations between different choices of these graphs.

2 The Penner coordinates

2.1 Triangulations and fat graphs

Consider a fixed oriented topological surface Σ of genus g with $s \geq 1$ punctures. An ideal triangulation τ of Σ is the isotopy class of a collection of disjointly embedded arcs in Σ running between the punctures such that τ decomposes Σ into triangles. There are $2M$ triangles and $3M$ edges for any ideal triangulation. As an example we have drawn a triangulation of the once-punctured torus in Figure 1.

The graph dual to a triangulation is a trivalent fat graph, i.e. a trivalent graph embedded in the surface with fixed cyclic order of the edges incident to each vertex. An example for a fat graph on the once-punctured torus is depicted in Figure 2. The sets of vertices and edges of a fat graph φ will be denoted φ_\circ and $\varphi_\mathbf{1}$ respectively.

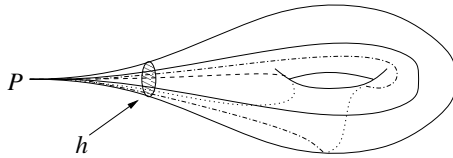


Figure 1. Triangulation of the once-punctured torus.

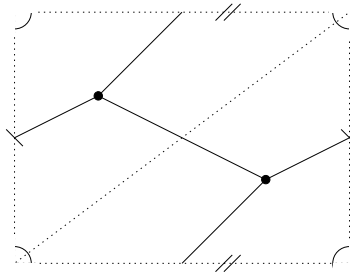


Figure 2. Another representation of the triangulation from Figure 1 and the dual fat graph.

An ideal triangulation is called *decorated* if the triangles are numbered and if a corner is marked for each triangle [30]. The decoration of the triangle t_v dual to a vertex $v \in \varphi_o$ can be used to fix a numbering convention for the edges $e_i^v, i = 1, 2, 3$ which emanate from v as defined in Figure 3.

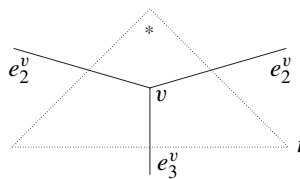


Figure 3. Graphical representation of the vertex v dual to a triangle t . The marked corner defines a corresponding numbering of the edges that emanate at v .

Remark 1. Decorated ideal triangulations are dual to decorated fat graphs, which means that the vertices are numbered, and for each vertex $v \in \varphi_o$ one has chosen a distinguished edge $e_v \in \varphi_1$. As a convention we will assume that fat graphs always carry such a decoration unless otherwise stated.

2.2 Penner coordinates

It turns out to be useful to consider a somewhat enlarged object which keeps track of the choices of horocycles around each of the punctures, the so-called decorated Teichmüller space $\tilde{\mathcal{T}}(\Sigma)$. $\tilde{\mathcal{T}}(\Sigma)$ is defined as a principal \mathbb{R}_+^s -bundle over $\mathcal{T}(\Sigma)$ by taking the s -tuple of horocycles around each of the punctures as the fiber over a point of $\mathcal{T}(\Sigma)$. The ordered s -tuple of hyperbolic lengths of the horocycles gives coordinates for the fibers.

Given any point P in the decorated Teichmüller space $\tilde{\mathcal{T}}(\Sigma)$ and an ideal triangulation τ of Σ , Penner assigns a coordinate value to each of the edges in τ by means of the following construction. By means of Fuchsian uniformization one may equip the surface Σ with a unique hyperbolic metric g associated to our chosen point $P \in \tilde{\mathcal{T}}(\Sigma)$. Let τ_1 be the set of edges of a triangulation τ . Each edge e in τ_1 may be straightened to a geodesic for the hyperbolic metric g . The coordinate $l_e(P)$ is defined as the hyperbolic length of the segment of e that lies between the two horocycles surrounding the punctures that e connects, taken with positive sign if the two horocycles are disjoint, with negative sign otherwise. We are going to consider the tuple $(l_e)_{e \in \tau_1}$ as a vector in the vector space \mathbb{R}^{τ_1} of dimension $3M$.

Theorem 1 (Penner [39], [40]). (a) *For any fixed ideal triangulation τ of Σ , the function*

$$l: \tilde{\mathcal{T}}(\Sigma) \rightarrow \mathbb{R}^{\tau_1}, \quad P \rightarrow (l_e(P))_{e \in \tau_1}$$

is a homeomorphism.

(b) *The pull-back of the Weil–Petersson two-form ω on $\mathcal{T}(\Sigma)$ is given by the expression*

$$\omega = - \sum_{t \in \tau_2} (dl_{e_1(t)} \wedge dl_{e_2(t)} + dl_{e_2(t)} \wedge dl_{e_3(t)} + dl_{e_3(t)} \wedge dl_{e_1(t)}),$$

where the summation is extended over the set τ_2 of triangles of τ , and $e_i(t)$, $i = 1, 2, 3$ are the edges bounding the triangle t , labelled in the counter-clockwise sense.

The Teichmüller space $\mathcal{T}(\Sigma)$ itself can finally be described as the space of orbits in $\tilde{\mathcal{T}}(\Sigma)$ under the following symmetry. Choose a number $d(\mathfrak{p})$ for each puncture \mathfrak{p} . Let the action of the symmetry be defined by

$$l'_e \equiv l_e + d(\mathfrak{p}) + d(\mathfrak{p}') \tag{2.1}$$

if the edge e connects the punctures \mathfrak{p} and \mathfrak{p}' .

2.3 Fock coordinates

There is a useful variant of the Penner coordinates which was introduced by V. Fock in [15]. In terms of the Penner coordinates one may define the Fock coordinates in

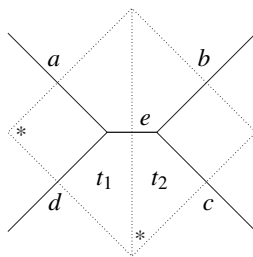


Figure 4. Two adjacent triangles and the dual fat graph.

terms of certain cross-ratios. Given a quadrilateral formed by two adjacent triangles we are going to keep the labelling of the edges introduced in Figure 4. Let

$$z_e = l_a + l_c - l_b - l_d. \quad (2.2)$$

The dependence of the Penner coordinates on the choice of horocycles drops out in the Fock coordinates. However, the variables z_e assigned to the $3M$ edges in a triangulation are not all independent. To describe the relations that they satisfy it is convenient to think of the Fock coordinates as being assigned to the edges of the fat graph dual to the given triangulation. Each closed curve c on Σ is homotopic to a unique path g_c on the fat graph which has minimal length w.r.t. the metric defined by assigning each edge of φ the length one. Such paths will also be called graph geodesics in the following. The path g_c may be described by a sequence of edges $e_1^c, \dots, e_{n_c}^c \in \varphi_1$. To a closed curve c let us associate

$$f_{\varphi,c} \equiv \sum_{i=1}^{n_c} z_e e_i^c. \quad (2.3)$$

The definition (2.2) then implies the relations

$$f_{\varphi,c} = 0 \quad (2.4)$$

for any closed curve c that encircles one puncture only. If one uses the equations (2.4) to express s of the variables z_e in terms of the others one obtains a set of coordinates for $\mathcal{T}(\Sigma)$.

On the $3M$ -dimensional space F_φ that is spanned by the coordinate functions $z_e(P)$ one may define a Poisson bracket Ω_{WP} which is such that

- (i) the elements f_a , $a = 1, \dots, s$ span the set C_φ of all $\mathfrak{c} \in F_\varphi$ such that

$$\Omega_{\text{WP}}(\mathfrak{c}, \mathfrak{w}) = 0 \quad \text{for all } \mathfrak{w} \in F_\varphi, \quad (2.5)$$

- (ii) the Poisson bracket which is induced by Ω_{WP} on the quotient F_φ/C_φ coincides with the Poisson bracket which corresponds to the Weil–Peterson symplectic form on $\mathcal{T}(\Sigma)$.

There exists a rather simple description for this Poisson-bracket [15]:

$$\Omega_{WP}(z_e, z_{e'}) = n_{e,e'}, \quad \text{where } n_{e,e'} \in \{-2, -1, 0, 1, 2\}. \quad (2.6)$$

The value of $n_{e,e'}$ depends on how edges e and e' are embedded into a given fat graph. If e and e' don't have a common vertex at their ends, or if one of e, e' starts and ends at the same vertex then $n_{e,e'} = 0$. In the case that e and e' meet at two vertices one has $n_{e,e'} = 2$ (resp. $n_{e,e'} = -2$) if e' is the first edge to the right⁴ (resp. left) of e at both vertices, and $n_{e,e'} = 0$ otherwise. In all the remaining cases $n_{e,e'} = 1$ (resp. $n_{e,e'} = -1$) if e' is the first edge to the right (resp. left) of e at the common vertex.

The coordinates z_e also have a nice geometrical meaning [15]. In the Fuchsian uniformization the two triangles that share the common edge e will be mapped into ideal hyperbolic triangles in the upper half plane. The edges are then represented by half-circles, and the corners will be at points x_1, \dots, x_4 on the real line, see Figure 5.

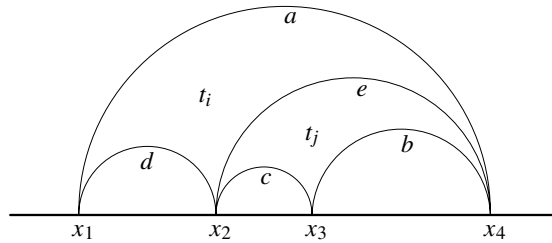


Figure 5. Representation of the triangles t_i and t_j in the upper half plane.

We then have

$$\exp(z_e) = \frac{(x_4 - x_1)(x_3 - x_2)}{(x_4 - x_3)(x_2 - x_1)}. \quad (2.7)$$

By means of Möbius transformations $x_i \rightarrow \frac{ax_i+b}{cx_i+d}$ one may map the corners of one of the two triangles to $-1, 0$ and ∞ respectively. The variable z_e , being expressed in terms of the Möbius-invariant cross-ratio therefore parameterizes the different ways of gluing two ideal hyperbolic triangles along a common edge modulo Möbius transformations. Given the variables z_e one may reconstruct the Riemann surface as represented in the Fuchsian uniformization by successively mapping ideal hyperbolic triangles into the upper half-plane, glued along the edges e in the way prescribed by the given value z_e [15].

3 The Ptolemy groupoids

In the previous section we had associated coordinate systems to fat graphs on a surface Σ : A change of graph will of course induce a change of coordinates. The groupoid

⁴W.r.t. to the orientation induced by the embedding of the fat-graph into the surface.

generated by the moves between different fat graphs will be the subject of the present section.

3.1 Groupoids vs. complexes

The groupoids that we will be interested in can be conveniently described as 2-dimensional connected CW complexes \mathcal{G} . The set of vertices \mathcal{G}_0 of \mathcal{G} will be represented by certain sets of graphs, within this section called fat graphs. The (directed) edges $E \in \mathcal{G}_1$ that connect these vertices correspond to the generators (“elementary moves”) of the groupoid, while the faces $F \in \mathcal{G}_2$ of \mathcal{G} yield the relations.

The groupoid G associated to the 2-dimensional connected CW complexes \mathcal{G} will then simply be the path groupoid of \mathcal{G} , which has the vertices in \mathcal{G}_0 as objects and the homotopy classes of edge paths between two vertices as morphisms. The homotopy class of paths leading from vertex $V_1 \in \mathcal{G}_0$ to vertex $V_2 \in \mathcal{G}_0$ will be denoted by $[V_2, V_1]$. Similarly we will sometimes use the notation $[W_E, V_E]$ for the element of G which corresponds to an edge $E \in \mathcal{G}_1$.

A path π which represents an element in the homotopy class $[W, V]$ may be represented by a *chain* of edges $E \in \mathcal{G}_1$, i.e. an ordered sequence $(E_{\pi, n(\pi)}, \dots, E_{\pi, 1})$, $E_{\pi, j} \in \mathcal{G}_1$ for $j = 1, \dots, n(\pi)$ such that $E_{\pi, j} \in [V_{\pi, j+1}, V_{\pi, j}]$ for $j = 1, \dots, n(\pi) - 1$, and $V_{\pi, 1} = V$, $V_{\pi, n(\pi)} = W$. We will also use the suggestive notation $E_n \circ E_{n-1} \circ \dots \circ E_1$ to denote a chain.

3.2 Change of the triangulation

In the case of the Ptolemy groupoids $Pt(\Sigma)$ we will consider a complex $\mathcal{P}t(\Sigma)$, where the set $\mathcal{G}_0 = \mathcal{P}t_0(\Sigma)$ is defined to be the set of fat graphs on Σ . Let us furthermore define $\mathcal{P}t_1(\Sigma)$ to consist of the following elementary moves.

- (i) *Permutation* (vw) . Exchanges the labels of the vertices v and w .
- (ii) *Rotation* ρ_v . See Figure 6.

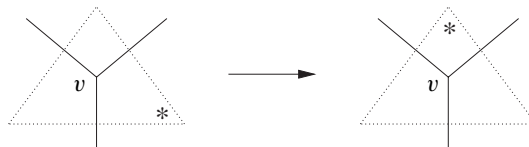


Figure 6. Transformation ρ_v changes the marked corner of the triangle dual to a vertex $v \in \varphi_0$.

- (iii) *Flip* ω_{vw} . See Figure 7.

Proposition 1 ([39, Proposition 7.1], [32]). *The complex $\mathcal{P}t(\Sigma)$ is connected, i.e. for a given surface Σ , any two fat graphs φ and φ' can be connected by a chain of elementary transformations.*

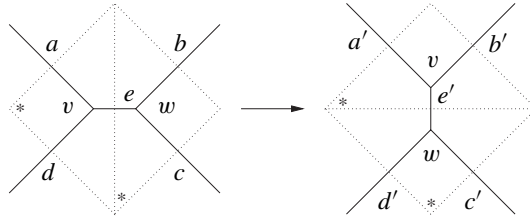


Figure 7. The flip transformation ω_{vw} changes the diagonal in the quadrilateral formed by the two adjacent triangles t_v and t_w .

Validity of the following relations in $\text{Pt}(\Sigma)$ can easily be verified pictorially [32].

$$\begin{aligned}
 \rho_v \circ \rho_v \circ \rho_v &= \text{id}, \\
 \omega_{u_1 u_2} \circ \omega_{u_3 u_4} &= \omega_{u_3 u_4} \circ \omega_{u_1 u_2}, \quad u_r \neq u_s \text{ for } r \neq s, \\
 \omega_{vw} \circ \omega_{uw} \circ \omega_{uv} &= \omega_{uv} \circ \omega_{vw}, \\
 (\rho_v^{-1} \times \rho_w) \circ \omega_{vw} &= \omega_{vw} \circ (\rho_v^{-1} \times \rho_w), \\
 \omega_{wv} \circ \rho_v \circ \omega_{vw} &= (vw) \circ (\rho_v \times \rho_w).
 \end{aligned} \tag{3.1}$$

Theorem 2. *The complex $\mathcal{Pt}(\Sigma)$ is simply connected, i.e. any relation between the generators (vw) , ρ_v and ω_{vw} of the Ptolemy groupoid is a consequence of the relations (3.1) together with the relations of the permutation group.*

The proof of this theorem is explained in Appendix C.

One of the main virtues of the Penner coordinates is that the corresponding change of coordinates can be described rather simply.

Lemma 1 (Lemma A.1a of [40]). *Let τ' be the triangulation obtained by applying the flip of Figure 7 to a pair of adjacent triangles in a given initial triangulation τ , and denote e and e' the diagonal edge before and after the flip. The coordinates associated to τ and τ' will then agree for each edge that the two triangulations have in common, and*

$$\lambda_{e'} = \frac{1}{\lambda_e} (\lambda_a \lambda_c + \lambda_b \lambda_d), \quad \lambda_f \equiv \sqrt{2} \exp\left(\frac{1}{2} l_f\right) \text{ for all } f \in \varphi_1, \tag{3.2}$$

where we have labelled the edges according to Figure 7.

The corresponding transformation of the Fock variables is also easy to describe:

$$\begin{aligned}
 e^{-z_{a'}} &= e^{-z_a} (1 + e^{-z_e}) & e^{-z_{b'}} &= e^{+z_b} (1 + e^{+z_e}) \\
 e^{+z_{d'}} &= e^{+z_d} (1 + e^{+z_e}) & z_{e'} &= -z_e & e^{+z_{c'}} &= e^{-z_c} (1 + e^{-z_e})
 \end{aligned} \tag{3.3}$$

3.3 The representation of the mapping class group

The mapping class group $\text{MC}(\Sigma)$ of the topological surface Σ is the group of isotopy classes of orientation-preserving diffeomorphisms of Σ . An element $\mu \in \text{MC}(\Sigma)$ of the mapping class group will map a given (decorated) triangulation τ into another one, $\mu.\tau$. The fact that any two triangulations can be mapped into each other by a composition of the elementary transformations introduced in Section 3.2 therefore leads to an embedding of the mapping class group into the Ptolemy groupoid:

$$\Phi_\tau : \text{MC}(\Sigma) \rightarrow \text{Pt}(\Sigma), \quad \Phi_\tau(f) \equiv [\mu.\tau, \tau]. \quad (3.4)$$

More precisely, Φ_τ induces a homomorphism $\text{MC}(\Sigma) \rightarrow \text{Pt}(\Sigma)$ in the sense that

$$\Phi_\tau(\mu_2 \circ \mu_1) = \Phi_{\mu_1.\tau}(\mu_2) \circ \Phi_\tau(\mu_1) \quad \text{for any } \mu_2, \mu_1 \in \text{MC}(\Sigma),$$

which embeds $\text{MC}(\Sigma)$ *injectively* into $\text{Pt}(\Sigma)$ [41, Theorem 1.3].

4 Teichmüller space as the phase space of a constrained system

As a preparation for the description of the quantum Teichmüller spaces it will be useful to parameterize the Teichmüller spaces by means of variables assigned to the *triangles* instead of the *edges* of a triangulation [30]. In the following section we shall elaborate upon the results and constructions in [30], strengthening them somewhat.

4.1 Kashaev's coordinates

Assume given a fat graph φ with set of vertices φ_\circ . For each vertex $v \in \varphi_\circ$ one may introduce a pair of variables (q_v, p_v) according to the following rule. Let us label the edges that emanate from the vertex v by e_i^v , $i = 1, 2, 3$ according to Figure 3. We will denote the Penner coordinates associated to the edges e_i^v by l_i^v , $i = 1, 2, 3$. Let us then define the pair of variables (q_v, p_v) as

$$(q_v, p_v) = (l_3^v - l_2^v, l_1^v - l_2^v). \quad (4.1)$$

Following Kashaev [30] we will consider the vector space $V_\varphi \simeq \mathbb{R}^{4M}$ obtained by regarding the variables q_v, p_v as the components $q_v(\mathfrak{v}), p_v(\mathfrak{v})$ of vectors $\mathfrak{v} \in V_\varphi$. The space of linear coordinate functions on V_φ will be called the Kashaev space W_φ . On W_φ we will consider the Poisson bracket Ω_φ defined by

$$\Omega_\varphi(p_v, q_w) = \delta_{vw}, \quad \Omega_\varphi(q_v, q_w) = 0, \quad \Omega_\varphi(p_v, p_w) = 0. \quad (4.2)$$

The assignment (4.1) associates a vector $\mathfrak{v}(P)$ in a subspace $T_\varphi \subset V_\varphi$ to each point $P \in \tilde{\mathcal{T}}(\Sigma)$. Kashaev has observed that the subspace T_φ can be characterized by a suitable set of linear forms $h_c \in W_\varphi$ (“constraints”).

4.1.1 The constraints. To define the linear forms h_c let us introduce an embedding of the first homology $H_1(\Sigma, \mathbb{R})$ into W_φ as follows. Each graph geodesic g_c which represents an element $[c] \in H_1(\Sigma, \mathbb{R})$ may be described by an ordered sequence of vertices $v_i \in \varphi_\circ$, and edges $e_i \in \varphi_\bullet$, $i = 0, \dots, n$, where $v_0 = v_n$, $e_0 = e_n$, and we assume that v_{i-1} , v_i are connected by the single edge e_i . We will define $\omega_i = 1$ if the arcs connecting e_i and e_{i+1} turn around the vertex v_i in the counterclockwise sense, $\omega_i = -1$ otherwise. The edges emanating from v_i will be numbered e_j^i , $j = 1, 2, 3$ according to the convention introduced in Figure 3. To each $[c] \in H_1(\Sigma, \mathbb{R})$ we will assign

$$h_c \equiv \sum_{i=1}^n u_i, \quad u_i := \omega_i \begin{cases} -q_{v_i} & \text{if } \{e_i, e_{i+1}\} = \{e_3^i, e_1^i\}, \\ p_{v_i} & \text{if } \{e_i, e_{i+1}\} = \{e_2^i, e_3^i\}, \\ q_{v_i} - p_{v_i} & \text{if } \{e_i, e_{i+1}\} = \{e_1^i, e_2^i\}. \end{cases} \quad (4.3)$$

h_c is independent of the choice of representative c within the class $[c]$. Let C_φ be the subspace in W_φ that is spanned by the h_c , $[c] \in H_1(\Sigma, \mathbb{R})$.

Lemma 2 ([30]). (i) *The mapping $H_1(\Sigma, \mathbb{R}) \ni [c] \mapsto h_c \in C_\varphi$ is an isomorphism of vector spaces.*

(ii) *The restriction of Ω_φ to C_φ coincides with the intersection form I on $H_1(\Sigma, \mathbb{R})$,*

$$\Omega_\varphi(h_{c_1}, h_{c_2}) = I(c_1, c_2).$$

(iii) *The linear forms h_c , $[c] \in H_1(\Sigma, \mathbb{R})$ vanish identically on the subspace T_φ .*

The equations $h_c(\mathbf{v}) = 0$, $[c] \in H_1(\Sigma, \mathbb{R})$ characterize the image of $\tilde{\mathcal{T}}(\Sigma)$ within V_φ . It is useful to recall that $H_1(\Sigma, \mathbb{R})$ splits as $H_1(\Sigma, \mathbb{R}) = H_1(\Sigma_{\text{cl}}, \mathbb{R}) \oplus B(\Sigma)$, where $B(\Sigma)$ is the $s - 1$ -dimensional subspace spanned by the homology classes associated to the punctures of Σ , and Σ_{cl} is the compact Riemann surface which is obtained by “filling” the punctures of Σ . The corresponding splitting of C_φ will be written as $C_\varphi = H_\varphi \oplus B_\varphi$.

4.1.2 Change of fat graph. In order to describe the change of Kashaev variables induced by a change of fat graph let us, following [30], define the following two transformations associated to the elementary moves ω_{vw} and ρ_v respectively.

$$A_v: (q_v, p_v) \mapsto (p_v - q_v, -q_v), \quad (4.4)$$

$$T_{vw}: \begin{cases} (U_v, V_v) \mapsto (U_v U_w, U_v V_w + V_v), \\ (U_w, V_w) \mapsto (U_w V_v (U_v V_w + V_v)^{-1}, V_w (U_v V_w + V_v)^{-1}), \end{cases} \quad (4.5)$$

where we have set $U_v \equiv e^{q_v}$ and $V_v \equiv e^{p_v}$ for all $v \in \varphi_\circ$.

Lemma 3. *The maps $A_v: W_\varphi \rightarrow W_{\rho_v \circ \varphi}$ and $T_{vw}: W_\varphi \rightarrow W_{\omega_{vw} \circ \varphi}$ defined in (4.4) and (4.5) respectively are canonical, i.e. they preserve the Poisson structure Ω_φ .*

The proof is again straightforward. Lemma 3 implies in particular that the mapping class group acts on W_φ by *canonical* transformations.

4.2 The structure of the Kashaev space W_φ

4.2.1 Fock variables vs. Kashaev’s variables. There is a canonical way to reconstruct the Fock-variables in terms of Kashaev’s variables which is found by combining equations (2.2) and (4.1). The result may be formulated as follows. Let $v, w \in \varphi_\circ$ be the vertices that are connected by the edge $e \in \varphi_1$, and let $e_i^v, i = 1, 2, 3$ be the edges introduced in Figure 3.

$$\hat{z}_e = \hat{z}_{e,v} + \hat{z}_{e,w}, \quad \hat{z}_{e,v} = \begin{cases} p_v & \text{if } e = e_1^v, \\ -q_v & \text{if } e = e_2^v, \\ q_v - p_v & \text{if } e = e_3^v. \end{cases} \quad (4.6)$$

The definition (4.6) defines a linear map $I_\varphi: F_\varphi \rightarrow \hat{F}_\varphi \subset W_\varphi$. It will be useful to describe the properties of this map a bit more precisely.

- Lemma 4.** (i) $\hat{z}_e(v(P)) = z_e(P)$ for all $e \in \varphi_1, P \in \tilde{\mathcal{T}}(\Sigma)$.
 (ii) $\Omega_\varphi(\hat{z}_e, \hat{z}_f) = \Omega_{\text{WP}}(z_e, z_f)$ for all $e, f \in \varphi_1$.
 (iii) $\Omega_\varphi(\hat{z}_e, h_c) = 0$ for all $e \in \varphi_1, c \in H_1(\Sigma, \mathbb{R})$.
 (iv) $\hat{f}_c \equiv I_\varphi(f_c) = h_c$ for all $[c] \in B(\Sigma)$.

Proof. Straightforward verifications. □

It is also useful to remark that the transformation of the Fock variables $\hat{z}_e, e \in \varphi_1$ that is induced by (4.4), (4.5) coincides with (3.3).

4.2.2 Splitting of W_φ . The linear forms $h_c \in B_\varphi$ turn out to be the Hamiltonian generators for the symmetry (2.1) [30]. It is therefore natural to consider the subspace $M_\varphi \subset W_\varphi$ which is spanned by the Hamiltonian vector fields that are generated by the linear forms $h_c \in B_\varphi$, as well as $N_\varphi \equiv M_\varphi \oplus B_\varphi$.

Proposition 2. *There exists a canonical transformation establishing the isomorphism of Poisson vector spaces $W_\varphi \simeq T_\varphi \oplus N_\varphi \oplus H_\varphi$, such that*

- (i) $T_\varphi \simeq \mathcal{T}'(\Sigma)$ is the space of linear functions on the Teichmüller space $\mathcal{T}(\Sigma)$;
- (ii) the restriction of Ω_φ to T_φ coincides with the Poisson bracket induced by the Weil–Peterson symplectic form.

Proof. As a warmup it may be instructive to count dimensions: We have $\dim(W_\varphi) = 8g - 8 + 4s$ and $\dim(C_\varphi) = \dim(H_1(\Sigma, \mathbb{R})) = 2g + s - 1$. In order to determine $\dim(N_\varphi)$ let us choose a canonical basis for $H_1(\Sigma, \mathbb{R})$, represented by curves

$\alpha_1, \dots, \alpha_g, \beta_1, \dots, \beta_g, \gamma_1, \dots, \gamma_{s-1}$ such that the only nontrivial intersection pairings are $I(\alpha_i, \beta_j) = \delta_{ij}$. N_φ is spanned by the images of the classes $[\gamma_1], \dots, [\gamma_{s-1}]$, together with the Hamiltonian vector fields that they generate. It follows that $\dim(N_\varphi) = 2s - 2$.

The main point that remains to be demonstrated is the existence of a decomposition of \hat{F}_φ as the direct sum

$$\hat{F}_\varphi = T_\varphi \oplus B_\varphi \quad \text{such that } \Omega_\varphi(t, h) = 0 \text{ for all } t \in T_\varphi, h \in N_\varphi. \quad (4.7)$$

Thanks to Lemma 3 we may choose a convenient fat graph to this aim. Let us pick a basis $\mathfrak{B}(\Sigma)$ for $B(\Sigma)$ represented by the curves which encircle $s - 1$ of the s punctures. It is easy to see that we can always construct a fat graph φ' such that the elements of $\mathfrak{B}(\Sigma)$ are represented by single edges in φ'_1 . These edges end in a unique vertex $v(c)$ for each $c \in \mathfrak{B}(\Sigma)$. It is clear that the expression for h_c only involves the variables $(q_{v(c)}, p_{v(c)})$ for all $c \in \mathfrak{B}(\Sigma)$. It follows from (4.2) that the Hamiltonian vector field generated by h_c can likewise be expressed in terms of $(q_{v(c)}, p_{v(c)})$ only. The existence of the sought-for decomposition (4.7) is obvious in this case.

The result is carried over to the general case with the help of Lemma 3. It is clear that the subspace $T_\varphi \subset \hat{F}_\varphi$ is defined uniquely by the condition (4.7). T_φ may then also be described as the quotient of \hat{F}_φ by the conditions $h_c = 0$ for all $[c] \in B(\Sigma)$. It therefore follows from our discussion in §2.3 that T_φ is canonically isomorphic to $\mathcal{T}'(\Sigma)$, the vector space of linear functions on $\mathcal{T}(\Sigma)$.

To complete the proof it remains to observe that we have

- (i) $\Omega_\varphi(t, h) = 0$ for all $t \in T_\varphi, h \in H_\varphi$,
- (ii) $\Omega_\varphi(h, n) = 0$ for all $h \in H_\varphi, n \in N_\varphi$.

(i) follows directly from Lemma 4, whereas part (ii) can easily be verified by considering the fat graph φ' above. □

5 The Fenchel–Nielsen coordinates

We will now be interested in the case of Riemann surfaces with a boundary that is represented by a collection of $s > 0$ geodesics. Another standard set of coordinates for the Teichmüller spaces is associated to the decomposition of a Riemann surface into trinions (three-holed spheres). We are now going to review the definition of these coordinates. Different sets of Fenchel–Nielsen coordinates will be associated to different *markings* of the Riemann surface in a way which is analogous to the relation between the Penner coordinates and fat graphs.

Let us denote by S_3 the sphere with three holes (trinion). As a concrete model we may e.g. choose

$$S_3 \equiv \{z \in \mathbb{C}; |z| \geq \varepsilon, |1 - z| \geq \varepsilon, |z| < 1/\varepsilon\}. \quad (5.1)$$

Any trinion with a smooth boundary is diffeomorphic to S_3 .

Definition 1. A marking σ of a surface Σ consists of the following data.

- (i) A *cut system* \mathcal{C}_σ , which is a set $\mathcal{C}_\sigma = \{c_1, \dots, c_{3g-3+s}\}$ of simple non-intersecting oriented closed curves c_i on Σ . Cutting Σ along the curves in \mathcal{C} decomposes the surface into a collection \mathcal{P}_σ of trinions.
- (ii) A choice of a trivalent graph Γ_T with a single vertex v_T in each trinion $T \in \mathcal{P}_\sigma$ such that the graphs on the different trinions glue to a connected graph Γ_σ on Σ .
- (iii) A choice of a distinguished boundary component for each trinion $T \in \mathcal{P}_\sigma$.

These data will be considered up to isotopy.

An example for the graphs Γ_T is depicted in Figure 10.

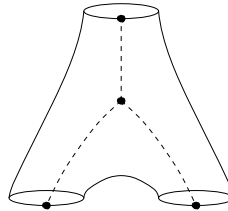


Figure 8. A trinion equipped with a marking graph.

5.1 Definition of the Fenchel–Nielsen coordinates

The basic observation underlying the definition of the Fenchel–Nielsen twist coordinates is the fact that for each triple (l_1, l_2, l_3) of positive real numbers there is a *unique* metric of constant curvature -1 on the three-holed sphere (trinion) such that the boundary components are geodesics with lengths l_i , $i = 1, 2, 3$. A trinion with its metric of constant curvature -1 will be called *hyperbolic trinion*. There furthermore exist three distinguished geodesics on each hyperbolic trinion that connect the boundary components pairwise.

Let us assume that the geodesic c separates two trinions T_a and T_b . Pick boundary components c_a and c_b of T_a and T_b respectively by starting at c , following the marking graphs, and turning left at the vertices. As mentioned above, there exist distinguished geodesics on T_a and T_b that connect c with c_a and c_b respectively. Let δ_c be the signed geodesic distance between the end-points of these geodesics on c , and let

$$\theta_c = 2\pi \frac{\delta_c}{l_c} \quad (5.2)$$

be the corresponding twist-angle. In a similar way one may define θ in the case that cutting along c opens a handle.

Given a cut system $\mathcal{C} = \{c_1, \dots, c_\kappa\}$, $\kappa = 3g - 3 + s$, we thereby associate to each Riemann surface Σ a tuple $(l_1, \dots, l_\kappa; \theta_1, \dots, \theta_\kappa)$ of real numbers. It can be shown (see e.g. [28]) that the Riemann surface Σ is characterized uniquely by the tuple $(l_1, \dots, l_\kappa; e^{i\theta_1}, \dots, e^{i\theta_\kappa}) \in (\mathbb{R}^+)^{\kappa} \times (S^1)^{\kappa}$. In order to describe the Teichmüller space $\mathcal{T}(\Sigma)$ of *deformations* of Σ it suffices to allow for arbitrary *real* values of the twist angles θ_i . Points in $\mathcal{T}(\Sigma)$ are then parametrized by tuples $(l_1, \dots, l_\kappa; \theta_1, \dots, \theta_\kappa) \in (\mathbb{R}^+)^{\kappa} \times \mathbb{R}^{\kappa}$.

Remark 2. The marking graph Γ_σ allows one to distinguish systems of Fenchel–Nielsen coordinates which are related to each other by Dehn twists, $\theta'_c = \theta_c + 2\pi k_c$, $k \in \mathbb{Z}$, $c \in \mathcal{C}$. To use the markings for the parametrization of different systems of Fenchel–Nielsen coordinates is then closely analogous to using fat graphs for the specification of systems of Penner coordinates.

The definition of the Fenchel–Nielsen coordinates does not use the choice of a distinguished boundary component for each trinion. The latter has been included into the definition 1 for later convenience only.

5.2 Symplectic structure

Let us furthermore notice that the Weil–Petersson symplectic form becomes particularly simple in terms of the Fenchel–Nielsen coordinates:

Theorem 3 ([58]).

$$\omega_{\text{WP}} = \sum_{i=1}^{\kappa} d\tau_i \wedge dl_i, \quad \tau_i = \frac{1}{2\pi} l_i \theta_i. \tag{5.3}$$

The content of the theorem may also be paraphrased as follows:

- (i) The geodesic length functions associated to non-intersecting closed curves Poisson-commute.
- (ii) The Hamiltonian flows generated by the geodesic length functions coincide with the Fenchel–Nielsen twist flows.

5.3 Geodesic lengths from the Penner coordinates

A nice feature of the Fock coordinates is that they lead to a particularly simple way to reconstruct the Fuchsian group corresponding to the point P in Teichmüller space that is parametrized by the variables $z_e(P)$, $e \in \varphi_1$. Assume given a graph geodesic g_c on the fat graph homotopic to a simple closed curve c on Σ . Let the edges be labelled e_i , $i = 1, \dots, r$ according to the order in which they appear on g_c , and define σ_i to be 1 if the path turns left⁵ at the vertex that connects edges e_i and e_{i+1} , and to be equal to

⁵W.r.t. to the orientation induced by the embedding of the fat-graph into the surface.

−1 otherwise. The generator $X(c)$ of the Fuchsian group that corresponds to c is then constructed as follows [15].

$$X_c = V^{\sigma_r} E(z_{e_r}) \dots V^{\sigma_1} E(z_{e_1}), \quad (5.4)$$

where the matrices $E(z)$ and V are defined respectively by

$$E(z) = \begin{pmatrix} 0 & +e^{+\frac{z}{2}} \\ -e^{-\frac{z}{2}} & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}. \quad (5.5)$$

Given the generator X_c of the Fuchsian group one may then calculate the hyperbolic length of the closed geodesic isotopic to c via

$$2 \cosh\left(\frac{1}{2}l_c\right) = |\text{tr}(X_c)|. \quad (5.6)$$

The proof of (5.4) was omitted in [15]. We are therefore now going to explain how to verify the validity of equation (5.4).

It was remarked in the Section 2.3 that for given values of the coordinates z_e one may construct a uniformized representation of the corresponding Riemann surface by successively mapping ideal hyperbolic triangles into the upper half plane which are glued according to the values z_e . Iterating this procedure ad infinitum one generates a tessellation of the upper half plane by ideal hyperbolic triangles. Let us now consider a generator $X(c)$ of the Fuchsian group, represented on the upper half plane by a Möbius transformation $M_{X(c)}$, where $M_X(u) \equiv \frac{au+b}{cu+d}$ if $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The element $c \in \pi_1(\Sigma)$ may then be represented by an open path on the upper half plane which leads from a chosen base point u to its image under M_X .

But one may equivalently represent the motion along the path by standing still at the base point and moving the tessellation around by means of Möbius transformations. More precisely, let us assume that our base point u is located within the ideal hyperbolic triangle t_0 with corners at $-1, 0, \infty$, and that the path \mathcal{P}_c representing our chosen element $c \in \pi_1(\Sigma)$ crosses the edges e_i $i = 1, \dots, r$ in the order of the labelling. We may assume that the edge e_1 connects the points 0 and $e^{z_{e_1}}$. After having crossed edge e_1 one would have left the triangle t_0 into the triangle t_1 with corners at $0, e^{z_{e_1}}, \infty$. The Möbius transformation M_1 corresponding to $E(z_{e_1})$ brings one back into t_0 : It can be checked that it leaves the set of corners $\{-1, 0, e^{z_{e_1}}, \infty\}$ on the two triangles glued along e_1 invariant, but exchanges the two triangles. To continue along the path \mathcal{P}_c in this fashion we now need to map the next edge e_2 to be crossed to the edge going from 0 to ∞ before we can apply the Möbius transformation corresponding to $E(z_{e_2})$ in the same manner as before. This is precisely what the Möbius transformation $M_{V\sigma_1}$ does: It simply rotates the edges of our fundamental triangle t_0 . Moreover, the triangle t_2 that would be reached when leaving t_0 through e_2 will be mapped by $M_{V\sigma_1}$ into the ideal hyperbolic triangle with corners $0, e^{z_{e_2}}, \infty$. The fact that $e^{z_{e_2}}$ is indeed the position that the corner of t_2 is mapped into by $M_{V\sigma_1}$ follows from the prescription for gluing t_1 and t_2 along e_2 in terms of z_{e_2} and the fact that $M_{V\sigma_1}$ preserves cross-ratios.

By continuing in this fashion one generates the Möbius transformation $M_{X(c)}$ that evidently maps the original tessellation representing the chosen point P in $\mathcal{T}(\Sigma)$ into

another one that is equally good as a representation for P . By assumption, $M_{X(c)}$ represents a closed path on the considered fat graph. This means that the points that are mapped into each other by $M_{X(c)}$ are to be identified as different representatives for the same points on the surface corresponding to our point $P \in \mathcal{T}(\Sigma)$.

6 Coordinates for surfaces with holes of finite size

Throughout we are mainly interested in the case of Riemann surfaces which have a boundary $\partial\Sigma$ represented by s geodesics of finite length. We therefore need to discuss how to introduce analogs of the previously described coordinate systems for $\mathcal{T}(\Sigma)$ for the cases of interest here. When considering surfaces with holes of finite size one has to choose if one wants to keep the geodesic lengths of the boundary components variable, or if one wants to consider surfaces Σ_Λ which have fixed boundary length given by the tuple $\Lambda = (l_1, \dots, l_s) \in \mathbb{R}_+^s$. We shall find the first option often more convenient to work with. Passing to a representation in which the boundary lengths are fixed will then be almost trivial.

6.1 Useful fat graphs on surfaces with holes of finite size

Riemann surfaces Σ with s holes can always be represented by considering a Riemann surface Σ^c with s pairs of punctures, from which Σ is obtained by cutting Σ^c along the geodesics b_1, \dots, b_s that encircle the pairs of punctures. This simple observation allows us to use the coordinates discussed previously in order to define coordinates for the Teichmüller spaces of surfaces with s holes. In order to spell out more precisely how to do this, let us first introduce a convenient class of fat graphs.

Let us consider a pair of punctures (P_1, P_2) . Let c be a geodesic such that cutting Σ along c produces two connected components one of which is a two-punctured disc D with punctures P_1 and P_2 . A given fat graph φ^c will be said to have standard form near D if there exists a neighborhood of the disc D in which φ^c is homotopic to the fat graph depicted on the left half of Figure 9.

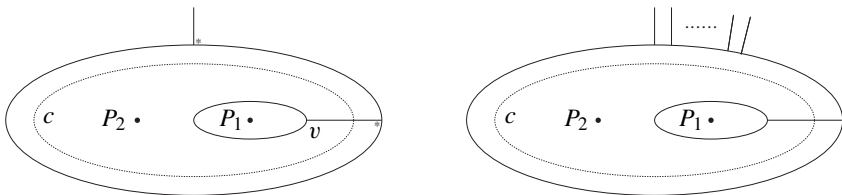


Figure 9. Simple fat graphs in a neighborhood of the disc defined by the geodesic c which encircles two punctures P_1 and P_2 .

For surfaces Σ^e with s pairs of punctures there exist fat graphs φ^e which are of standard form in the neighborhood of $s - 1$ discs D_i . The simplest possible form of a fat graph around the remaining two punctures is indicated on the right of Figure 9. A fat graph φ^e on a surface with $2s$ punctures will be said to have standard form if it has standard form near $s - 1$ discs D_i , and if it has the form depicted on the right of Figure 9 in a neighborhood of the remaining disc.

We will finally say that a fat graph φ on a Riemann surface Σ with s geodesic boundaries has standard form if (Σ, φ) can be obtained from a pair (Σ^e, φ^e) consisting of a $2s$ -punctured surface Σ and a fat graph φ^e of standard form by cutting Σ^e along s geodesics b_1, \dots, b_s , each of which encircles a pair of punctures. The embedding $\Sigma \hookrightarrow \Sigma^e$ furthermore induces an embedding $\text{MC}(\Sigma) \hookrightarrow \text{MC}(\Sigma^e)$ of the respective mapping class groups. The subgroup of $\text{MC}(\Sigma^e)$ which is generated by the diffeomorphisms that are supported on $\Sigma \subset \Sigma^e$ preserves the set of fat graphs which have standard form.

6.2 Kashaev type coordinates

If we only use fat graphs of standard form, it becomes easy to adapt the previously discussed systems of coordinates to the case of interest in the rest of this chapter. We may in particular consider the space W_φ of Kashaev variables associated to the fat graph φ . A subspace T_φ of W_φ can again be defined by means of the decomposition (4.7). It is furthermore convenient to introduce the set φ'_1 which only contains the edges of φ that do not end in boundary components of Σ .

Lemma 5. (i) *We have $T_\varphi \simeq T_{\varphi^e}$.*

(ii) *The Fock coordinates $\{z_e; e \in \varphi'_1\}$ form a set of coordinates for T_φ .*

Proof. Let us first note that the linear form h_1 which is via (4.3) associated to puncture P_1 can be expressed in terms of the variables (p_v, q_v) associated to the vertex v in Figure 9 only. This means that both p_v and q_v are contained in N_{φ^e} . Instead of the linear form h_2 associated to puncture P_2 we may consider $h_c = h_1 + h_2$, which can be expressed exclusively in terms of the variables associated to the vertices contained in Σ . Part (i) of the lemma follows easily from these observations.

In order to verify part (ii) one may again consider φ^e . When writing the relations $f_c = 0$, $c \in B(\Sigma)$ in terms of the Fock variables z_e , $e \in \varphi_1^e$ one will always find contributions containing the z_e , $e \in \varphi_1^e \setminus \varphi'_1$. It is then easy to convince oneself that the relations $\hat{f}_c = 0$ may be used to express the z_e , $e \in \varphi_1^e \setminus \varphi'_1$ in terms of the z_e , $e \in \varphi'_1$. After this is done, all relations $f_c = 0$, $c \in B(\Sigma)$ are satisfied, the variables z_e , $e \in \varphi'_1$ are therefore unconstrained. \square

However, in this case the relation between the Teichmüller space $\mathcal{T}(\Sigma)$ and T_φ is slightly more complicated. In order to describe this relation let us consider the spaces $\text{Fun}(\mathcal{T}(\Sigma))$ and $\text{Fun}(T_\varphi)$ of smooth functions on $\mathcal{T}(\Sigma)$ and T_φ respectively. These

spaces carry canonical Poisson brackets $\{ \cdot, \cdot \}_{WP}$ and $\{ \cdot, \cdot \}_\varphi$ uniquely defined by the bilinear forms Ω_{WP} and Ω_φ respectively.

Proposition 3. *We may represent $\text{Fun}(\mathcal{T}(\Sigma))$ as the subspace of $\text{Fun}(T_\varphi)$ which is defined by the conditions*

$$\{F, l_i\}_\varphi = 0, \quad i = 1, \dots, s, \quad F \in \text{Fun}(T_\varphi), \quad (6.1)$$

where l_i is the length function which is associated to the i -th boundary component via equations (5.4) and (5.6).

It should be noted that the length functions associated to the boundary components are contained in $\text{Fun}(\mathcal{T}(\Sigma))$.

Part II. Quantization of the Teichmüller spaces

Our aim is to construct certain classes of infinite-dimensional representations of the mapping class groups $\text{MC}(\Sigma)$. One possible approach to this problem is to “quantize” Poisson manifolds like the Teichmüller spaces on which $\text{MC}(\Sigma)$ acts as a group of symmetries.

Our construction will proceed in two main steps. Quantization of the Kashaev spaces W_φ leads to a rather elegant construction of projective unitary representations of the mapping class groups [30]. However, these representations turn out to be reducible. The second step will therefore be to identify distinguished subrepresentations within the representations coming from the quantization of the Kashaev spaces as the mapping class group representations which are naturally associated to the quantization of the Teichmüller spaces. A direct construction of the latter is not known, which is why this somewhat indirect construction seems to be most efficient at the moment.

7 Quantization of the Teichmüller spaces

7.1 Canonical quantization

Quantization of a Poisson manifold \mathcal{P} means “deforming” the space of functions on \mathcal{P} into a one-parameter (\hbar) family of noncommutative algebras \mathcal{P}_\hbar in such a way that the deformed product $f *_\hbar g$ satisfies

$$f *_\hbar g = fg + \hbar\Omega(f, g) + \mathcal{O}(\hbar^2), \quad (7.1)$$

where fg is the ordinary commutative product of functions on \mathcal{P} and $\Omega(f, g)$ is the Poisson bracket on \mathcal{P} . If the Poisson manifold \mathcal{P} has a group G of symmetries it is

natural to demand that these symmetries are preserved by quantization in the sense that any $g \in G$ is realized as an automorphism $f \rightarrow \mathfrak{a}_g(f)$ of \mathcal{P}_\hbar .

Representations of the group G can be constructed by studying representations of the algebra \mathcal{P}_\hbar by operators $\mathbf{O}(f)$ on a Hilbert space \mathcal{H} ,

$$\mathbf{O}(f) \cdot \mathbf{O}(g) = \mathbf{O}(f *_{\hbar} g).$$

The Hilbert space \mathcal{H} will then typically come equipped with a unitary projective representation of the group G of symmetries by operators \mathbf{U}_g such that the automorphisms $\mathbf{A}_g(\mathbf{O}(f)) \equiv \mathbf{O}(\mathfrak{a}_g(f))$ are realized as

$$\mathbf{A}_g(\mathbf{O}) = \mathbf{U}_g \cdot \mathbf{O} \cdot \mathbf{U}_g^{-1}.$$

Quantization is particularly simple if there exist coordinate functions q_1, \dots, q_N and p_1, \dots, p_N defined globally on \mathcal{P} such that the Poisson bracket takes the form

$$\Omega(p_v, q_w) = \delta_{vw}, \quad \Omega(q_v, q_w) = 0, \quad \Omega(p_v, p_w) = 0, \quad (7.2)$$

for $v, w \in \{1, \dots, N\}$. One may then define \mathcal{P}_\hbar in such a way that the relations

$$i[p_v, q_w] = \hbar \delta_{vw}, \quad [q_v, q_w] = 0, \quad [p_v, p_w] = 0, \quad (7.3)$$

hold for $[f, g] \equiv f *_{\hbar} g - g *_{\hbar} f$. There exists a standard representation \mathbf{O} of these commutation relations on (dense subspaces of) the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^N)$ of square-integrable functions $\Psi(\mathfrak{q})$, $\mathfrak{q} = (q_1, \dots, q_N)$, which is generated by pairs of operators

$$\mathfrak{p}_v \equiv \frac{1}{2\pi\hbar} \mathbf{O}(p_v), \quad \mathfrak{q}_v \equiv \mathbf{O}(q_v), \quad v = 1, \dots, N,$$

that are defined respectively by

$$\mathfrak{q}_v \Psi(\mathfrak{q}) \equiv q_v \Psi(\mathfrak{q}), \quad \mathfrak{p}_v \Psi(\mathfrak{q}) \equiv \frac{1}{2\pi i} \frac{\partial}{\partial q_v} \Psi(\mathfrak{q}). \quad (7.4)$$

This simple example for the quantization of Poisson manifolds is often referred to as “canonical quantization”.

Remark 3. The representation that is constructed in this way is irreducible in the following sense. If \mathbf{O} is a bounded operator on $L^2(\mathbb{R}^N)$ which commutes⁶ with the operators \mathfrak{p}_v and \mathfrak{q}_v for all $v = 1, \dots, N$ then $\mathbf{O} = \chi$, the operator of multiplication with the complex number χ .

Following the discussion in Sections 2, it seems natural to define the quantized Teichmüller spaces as the noncommutative algebra $\mathcal{T}_\hbar(\Sigma)$ with generators z_e and

⁶Commutativity $[\mathbf{O}, \mathbf{A}] = 0$ with a self-adjoint unbounded operator \mathbf{A} is, by convention, understood in the sense of commutativity with the spectral projections of \mathbf{A} . For the reader's convenience, we have collected the relevant operator-theoretical results in Appendix B.

relations

$$i[z_{e_2}, z_{e_1}] = \hbar \Omega_{\text{WP}}(z_{e_2}, z_{e_1}) \tag{7.5}$$

$$f_{\varphi, c} = 0, \quad \text{for all } c \in B(\Sigma). \tag{7.6}$$

The space $\mathcal{H}(\Sigma)$ will be defined as an irreducible representation of the commutation relations (7.5) which satisfies the additional conditions (7.6).

7.2 Quantization of the Kashaev space W_φ

For each given fat graph φ let us define the Hilbert space $\mathcal{K}(\varphi)$ as the space of square integrable functions $\Psi(\mathfrak{q})$ of the Kashaev-variables $\mathfrak{q} = (q_1, \dots, q_{2M})$. On $\mathcal{K}(\varphi)$ we shall consider the basic operators $\mathfrak{p}_v, \mathfrak{q}_v$ defined in (7.4) for $v = 1, \dots, 2M$. The noncommutative algebras of operators which are generated by the operators $\mathfrak{q}_v, \mathfrak{p}_w$ with the commutation relations

$$[\mathfrak{p}_v, \mathfrak{q}_w] = (2\pi i)^{-1} \delta_{v,w} \tag{7.7}$$

may be considered as representing quantized algebras of functions on the Kashaev space W_φ .

Remark 4. It is worth noting that the decoration of the triangles is used to define the concrete realization of the space $\mathcal{K}(\varphi)$ as a space of square integrable functions.

When quantizing the Kashaev space W_φ we get more than we ultimately want. In order to see this, let us introduce quantum analogs of the coordinate functions h_c and \hat{z}_e respectively, i.e. self-adjoint operators \hat{h}_c and \hat{z}_e on $\mathcal{H}(\varphi)$ which are defined by formulae very similar to (4.3) and (4.6) respectively, normalized in such a way that the following commutation relation hold:

$$i[\hat{z}_{e_1}, \hat{z}_{e_2}] = \hbar \Omega_{\text{WP}}(z_{e_1}, z_{e_2}), \tag{7.8}$$

$$i[\hat{h}_{c_1}, \hat{h}_{c_2}] = \hbar I(c_1, c_2). \tag{7.9}$$

$$[\hat{z}_e, \hat{h}_c] = 0 \quad \text{for all } c \in H_1(\Sigma, \mathbb{R}). \tag{7.10}$$

We observe that we *do* have a representation, henceforth denoted Z_φ , of the algebra (7.5), but this representation is neither irreducible, nor does it fulfill the additional relations (7.6). The latter point becomes most clear if one introduces the operators \hat{f}_c , $c \in B(\Sigma)$ associated to the relations (7.6) which are defined by replacing $z_e \rightarrow \hat{z}_e$ in (2.3). We may then observe that

$$\hat{f}_c = h_c \quad \text{for } c \in B(\Sigma), \tag{7.11}$$

which is verified in the same way as statement (iv) in Lemma 4.

7.3 Reduction to the quantized Teichmüller spaces

In order to see that the representations Z_φ “contain” irreducible representations of the quantized Teichmüller spaces $\mathcal{T}_\hbar(\Sigma)$, let us consider the noncommutative algebra $Z_\hbar(\Sigma)$ with generators z_e and the only relations (7.5). One should observe that the f_c generate the center of the algebra Z_\hbar . Irreducible unitary representations Z_f of this algebra are parametrized by linear functions $f: B(\Sigma) \rightarrow \mathbb{R}$. Such representations are such that the operators $f_c = Z_f(f_c)$ are realized as the operators of multiplication with the real numbers $f(c)$, $c \in B(\Sigma)$.

The representations can be constructed concretely by forming linear combinations t_k and $t_{k'}^\vee$, $k = 1, \dots, 3g - 3 + s$ of the z_e which

- (i) are mutually linearly independent, and linearly independent of the f_c , and
- (ii) which satisfy the commutation relations

$$2\pi [t_k, t_{k'}^\vee] = i \delta_{kk'}. \tag{7.12}$$

Canonical quantization realizes Z_f on the Hilbert space $\mathcal{H}_f(\varphi) \simeq L^2(\mathbb{R}^{3g-3+s})$ which consists of square-integrable functions $\Phi(\mathbf{t})$, $\mathbf{t} = (t_1, \dots, t_{3g-3+s})$. The operators $Z_{f,e} \equiv Z_f(z_e)$ are then realized as linear combination of the operators

$$t_k \Phi_f(\mathbf{t}) = t_k \Phi_f(\mathbf{t}), \quad t_k^\vee \Phi_f(\mathbf{t}) = \frac{1}{2\pi i} \frac{\partial}{\partial t_k} \Phi_f(\mathbf{t}), \quad f_c \Phi_f(\mathbf{t}) = f(c) \Phi_f(\mathbf{t}),$$

where $k = 1, \dots, 3g - 3 + s$ and $c \in B(\Sigma)$.

Our aim is to describe how the representation Z_φ decomposes into the representations Z_f . In order to do this, let us introduce the representation

$$Z'_\varphi \equiv \int_{B'(\Sigma)}^\oplus d\mathbf{f} Z_f \quad \text{on the space} \quad \mathcal{H}_z(\varphi) \equiv \int_{B'(\Sigma)}^\oplus d\mathbf{f} \mathcal{H}_f(\varphi).$$

The space $\mathcal{H}_z(\varphi)$ is spanned by square-integrable families $\Phi \equiv (\Phi_f)_{f \in B'(\Sigma)}$ of functions $\Phi_f \in \mathcal{H}_f(\varphi)$ which are associated to the linear functions $f: B(\Sigma) \rightarrow \mathbb{R}$ in the dual $B'(\Sigma) \simeq \mathbb{R}^{s-1}$ up to a set of measure zero. The representatives $z_e \equiv Z_\varphi(z_e)$ are defined as follows

$$z_e \Phi \equiv (Z_{f,e} \Phi_f)_{f \in B'(\Sigma)}. \tag{7.13}$$

Proposition 4. *The decomposition of the representation Z_φ into irreducible representations of Z_\hbar may be written as follows:*

$$Z_\varphi \simeq \left(\int_{B'(\Sigma)}^\oplus d\mathbf{f} Z_f \right) \otimes 1_{\mathcal{H}_h(\varphi)}, \tag{7.14}$$

where the space $\mathcal{H}_h(\varphi)$ is isomorphic to $L^2(\mathbb{R}^g)$. There exists a unitary operator l_φ , $l_\varphi: \mathcal{K}(\varphi) \rightarrow \mathcal{H}_z(\varphi) \otimes \mathcal{H}_h(\varphi)$ such that

$$l_\varphi \cdot \hat{z}_e \cdot l_\varphi^{-1} = z_e \otimes 1 \quad \text{and} \quad l_\varphi \cdot \hat{h}_c \cdot l_\varphi^{-1} = 1 \otimes h_c, \tag{7.15}$$

for any $e \in \varphi_1$ and $c \in H_1(\Sigma_{cl}, \mathbb{R})$, respectively.

Proof. Let us recall the direct sum decomposition

$$W_\varphi \simeq T_\varphi \oplus N_\varphi \oplus H_\varphi. \quad (7.16)$$

To each of the three spaces T_φ , N_φ , H_φ one may choose coordinates which bring the Poisson bracket to the canonical form (7.2). The corresponding operators

$$\begin{aligned} (\hat{\mathfrak{t}}_k, \hat{\mathfrak{t}}_k^\vee), & \quad k = 1, \dots, 3g - 3 + s, \\ (\hat{\mathfrak{f}}_l, \hat{\mathfrak{f}}_l^\vee), & \quad l = 1, \dots, s - 1, \\ (\hat{\mathfrak{h}}_m, \hat{\mathfrak{h}}_m^\vee), & \quad m = 1, \dots, g, \end{aligned}$$

can be constructed as linear combinations of the $(\mathfrak{p}_v, \mathfrak{q}_v)$, $v \in \varphi_1$, in such a way that the only nontrivial commutation relations are

$$\begin{aligned} 2\pi [\hat{\mathfrak{t}}_k, \hat{\mathfrak{t}}_{k'}^\vee] &= ib^2 \delta_{kk'}, \quad k, k' = 1, \dots, 3g - 3 + s, \\ 2\pi [\hat{\mathfrak{f}}_l, \hat{\mathfrak{f}}_{l'}^\vee] &= ib^2 \delta_{ll'}, \quad l, l' = 1, \dots, s - 1, \\ 2\pi [\hat{\mathfrak{h}}_m, \hat{\mathfrak{h}}_{m'}^\vee] &= ib^2 \delta_{mm'}, \quad m, m' = 1, \dots, g. \end{aligned} \quad (7.17)$$

It will be convenient to form the following vectors with $4M$ operator-valued components:

$$\begin{aligned} \mathbf{v} &= (\dots, \mathfrak{q}_v, \dots, \mathfrak{p}_w, \dots), \\ \hat{\mathbf{v}} &= (\dots, \hat{\mathfrak{t}}_k, \dots, \hat{\mathfrak{f}}_l, \dots, \hat{\mathfrak{h}}_m, \dots, \hat{\mathfrak{t}}_k^\vee, \dots, \hat{\mathfrak{f}}_l^\vee, \dots, \hat{\mathfrak{h}}_m^\vee, \dots). \end{aligned}$$

The linear change of variables $\hat{\mathbf{v}} = \hat{\mathbf{v}}(\mathbf{v})$ can then be represented by a symplectic $(4M \times 4M)$ -matrix J_φ ,

$$\hat{\mathbf{v}} = J_\varphi \mathbf{v}, \quad J_\varphi \in \text{Sp}(2M, \mathbb{R}). \quad (7.18)$$

On the other hand let us note that $\mathcal{H}_z(\varphi) \otimes \mathcal{H}_h(\varphi)$ is canonically isomorphic to $L^2(\mathbb{R}^{2M})$ via

$$\mathbb{K}: \Phi \otimes \psi \rightarrow \Psi, \quad \Psi(\mathfrak{t}, \mathfrak{f}, \mathfrak{h}) \equiv \Phi_{\mathfrak{f}}(\mathfrak{t})\psi(\mathfrak{h}). \quad (7.19)$$

The corresponding representation of the commutation relations (7.17) on the Hilbert space $L^2(\mathbb{R}^{2M})$ is obtained by renaming the operators q_v, p_v as follows

$$(\dots, \mathfrak{q}_v, \dots, \mathfrak{p}_w, \dots) \equiv (\dots, \mathfrak{t}_k, \dots, \mathfrak{f}_l, \dots, \mathfrak{h}_m, \dots, \mathfrak{t}_k^\vee, \dots, \mathfrak{f}_l^\vee, \dots, \mathfrak{h}_m^\vee, \dots).$$

It follows from the Stone–von Neumann uniqueness theorem for the representation of the commutation relations (7.17) that these two representation must be related by a unitary transformation. This transformation may be characterized more precisely as follows.

Lemma 6. a) *To each $\gamma \in \text{Sp}(2M, \mathbb{R})$ there exists a unitary operator J_γ on $L^2(\mathbb{R}^{2M})$ such that*

$$J_\gamma \cdot \mathbf{v} \cdot J_\gamma^{-1} = \gamma \mathbf{v}. \quad (7.20)$$

The operators J_γ generate a projective unitary representation of $\text{Sp}(2M, \mathbb{R})$.

b) The operators J_γ can be represented in the form

$$J_\gamma = \exp(i J_\gamma(\mathbf{v})), \quad (7.21)$$

where $J_\gamma(\mathbf{v})$ is a quadratic expression in the operators \mathbf{v} .

Proof. Part a) is a classical result of I. E. Segal, [45]. Part b) follows easily from the observation that the quadratic functions of the operators \mathbf{v} generate a representation of the Lie algebra of $\mathrm{Sp}(2M, \mathbb{R})$ which satisfies the infinitesimal version of (7.20), see e.g. [24] for more details. \square

One may therefore find an operator J_φ on $L^2(\mathbb{R}^{2M})$ which represents the transformation (7.18) in the sense that

$$J_\varphi \cdot \mathbf{v} \cdot J_\varphi^{-1} = \hat{\mathbf{v}}(\mathbf{v}) = J_\varphi \mathbf{v}. \quad (7.22)$$

The sought-for isomorphism \mathfrak{l}_φ can finally be constructed as $\mathfrak{l}_\varphi = \mathbf{K}^{-1} \cdot J_\varphi^{-1}$. \square

Remark 5. It is worth noting that the definition of $\mathcal{H}_\mathfrak{f}(\varphi)$ depends only on the combinatorial structure of the fat graph φ , not on the way it is embedded into the Riemann surface Σ . It follows that the isomorphism $\mathcal{H}_\mathfrak{f}(\mu.\varphi) \simeq \mathcal{H}_\mathfrak{f}(\varphi)$, $\mu \in \mathrm{MC}(\Sigma)$ is *canonical*.

8 Representations of the mapping class groups

The representations of the mapping class group associated to the quantized Teichmüller spaces will be obtained by means of a very general construction which produces representations of the group G of symmetries of a two-dimensional CW complex \mathfrak{g} out of representations of the edge path groupoid of \mathfrak{g} . We shall first describe this construction, before we discuss how to construct representations of the Ptolemy groupoid on the quantized Teichmüller spaces. The latter will then induce the sought-for representation of $\mathrm{MC}(\Sigma)$.

8.1 Projective unitary representations of groupoids

Let us recall that a groupoid G is a category such that all morphisms are invertible. The objects of G will here be denoted by letters U, V, W, \dots . Anticipating that the groupoids G we will be interested in are path groupoids of some topological space we will use the notation $[W, V] \equiv \mathrm{Hom}_G(V, W)$. The elements of $[W, V]$ will also be called “paths”.

Definition 2. A unitary projective representation of the groupoid G consists of the following data:

- (i) A Hilbert space $\mathcal{H}(V)$ associated to each object $V \in \text{Ob}(G)$,
 (ii) a map u which associates to each path $\pi \in [W, V]$ in G a unitary operator

$$u(\pi): \mathcal{H}(V) \rightarrow \mathcal{H}(W),$$

- (iii) a family of maps ζ_V , $V \in \text{Ob}(G)$ which associate to each closed path $\pi \in [V, V]$ a number $\zeta_V(\pi) \in \mathbb{C}$ with $|\zeta(\pi)| = 1$.

These data are required to satisfy the relations

$$\begin{aligned} \text{a) } u(\pi_2 \circ \pi_1) &= u(\pi_2) \cdot u(\pi_1), & \text{b) } u(\pi^{-1}) &= u^\dagger(\pi), \\ \text{c) } \zeta_V(\pi_2 \circ \pi_1) &= \zeta_V(\pi_2)\zeta_V(\pi_1), & \text{d) } \zeta_V(\pi^{-1}) &= (\zeta_V(\pi))^*, \\ \text{e) } u(\pi) &= \zeta_V(\pi) \text{ if } \pi \in [V, V], & \text{f) } u(\text{id}) &= 1, \end{aligned} \quad (8.1)$$

where we use the notation $\zeta_V(\pi)$ also to denote the operator which multiplies each vector of \mathcal{H}_V by the number $\zeta_V(\pi)$.

The groupoids of interest will be the path groupoids G of two-dimensional CW complexes \mathcal{G} . The set of objects is given by the set of vertices \mathcal{G}_0 , whereas the set of morphisms coincides with the set of paths in the complex \mathcal{G} . Since each path π may be represented as a chain $E_{\pi, n(\pi)} \circ \cdots \circ E_{\pi, 2} \circ E_{\pi, 1}$ of edges in \mathcal{G}_1 it is clear that a projective unitary representation of the path groupoid G of a two-dimensional CW complex \mathcal{G} is characterized completely by specifying the images $u(E)$ for $E \in \mathcal{G}_1$. Existence of the family of maps ζ_V such that relation e) is fulfilled represents a rather nontrivial constraint that the operators $u(E)$, $E \in \mathcal{G}_1$ have to satisfy. Of course it suffices to satisfy these constraints for the 2-cells $\pi \in \mathcal{G}_2$.

8.2 Representations of symmetries of a groupoid

The group of symmetries G of a two-dimensional CW complex \mathcal{G} is the group of all invertible mappings

$$\mu: \begin{cases} \mathcal{G}_0 \ni V \longrightarrow \mu.V \in \mathcal{G}_0, \\ \mathcal{G}_1 \ni E \longrightarrow \mu.E \in \mathcal{G}_1. \end{cases}$$

There is an associated action on the paths in the complex \mathcal{G} ,

$$[W, V] \ni \pi \rightarrow \mu.\pi \in [\mu.W, \mu.V],$$

which is such that

$$\mu(\pi_2 \circ \pi_1) = \mu(\pi_2) \circ \mu(\pi_1), \quad \mu(\pi^{-1}) = (\mu(\pi))^{-1}. \quad (8.2)$$

We will assume that we are given a unitary projective representation of G which is compatible with the symmetry G in the sense that $\mathcal{H}(V)$ is canonically isomorphic with $\mathcal{H}(\mu.V)$, $\mathcal{H}(V) \simeq \mathcal{H}(\mu.V)$. We are going to show that the given representation of the groupoid G canonically induces a representation of its group G of symmetries.

Let us fix a base point $V \in \text{Ob}(G)$ and assume having chosen a path $\pi_V(\mu) \in [\mu.V, V]$ for each $\mu \in G$. Let then $R(\mu): \mathcal{H}(V) \rightarrow \mathcal{H}(V) \simeq \mathcal{H}(\mu.V)$ be defined by

$$R_V(\mu) = u(\pi_V(\mu)). \tag{8.3}$$

We are going to assume that the paths $\pi_{\mu_1.V}(\mu_2)$ are the translates of $\pi_V(\mu_2)$ under μ_1 , i.e. that $\pi_{\mu_1.V}(\mu_2) = \mu_1.\pi_V(\mu_2)$. It follows that

$$R_{\mu_1.V}(\mu_2) = R_V(\mu_2).$$

The operators $R_V(\mu)$ satisfy the relations

$$R_V(\mu_2) \cdot R_V(\mu_1) = \vartheta_V(\mu_2, \mu_1) R_V(\mu_2 \circ \mu_1), \tag{8.4}$$

$$\vartheta_V(\mu_2, \mu_1) = \zeta_V(\pi_V^{-1}(\mu_2 \circ \mu_1) \circ \pi_{\mu_1.V}(\mu_2) \circ \pi_V(\mu_1)). \tag{8.5}$$

We may next observe that the apparent dependence on the base point $V \in \mathcal{G}_o$ is inessential. Let $V, W \in \mathcal{G}_o$, and let us pick a path $\pi_{W,V} \in [W, V]$. For an operator $O_V: \mathcal{H}(V) \rightarrow \mathcal{H}(V)$ we define

$$A_{[W,V]}(O_V) = u(\pi_{W,V}) \cdot O_V(\mu) \cdot u^\dagger(\pi_{W,V}). \tag{8.6}$$

It is easy to convince oneself that $A_{W,V}(O_V(\mu))$ does not depend on the choice of a path $\pi_{W,V} \in [W, V]$. We furthermore have

$$A_{W,V}(R_V(\mu)) = R_W(\mu). \tag{8.7}$$

It easily follows that $\vartheta_V(\mu_2, \mu_1)$ does not depend on V , i.e. $\vartheta_V(\mu_2, \mu_1) \equiv \vartheta(\mu_2, \mu_1)$.

To summarize: The operators $R_V(\mu)$ generate a projective unitary representation $R_V(G)$ of G on \mathcal{H} ,

$$R_V(\mu_2) \cdot R_V(\mu_1) = \vartheta(\mu_2, \mu_1) R_V(\mu_2 \circ \mu_1). \tag{8.8}$$

The operators $A_{[W,V]}$ express the unitary equivalence of the representations R_V associated to the different $V \in \mathcal{G}_o$, which allows us to regard

$$R \equiv [(R_V)_{V \in \mathcal{G}_o}, (A_E)_{E \in \mathcal{G}_1}]$$

as the representation of G canonically associated to the given representation of the groupoid G .

Remark 6. There is of course some ambiguity in the construction, coming from the choice of a representative $\pi_V(\mu) \in [\mu.V, V]$. However, it is clearly natural to consider two representations r, r' as equivalent if the generators $r_V(\mu)$ and $r'_V(\mu)$ differ from each other just by multiplication with a (possibly μ -dependent) central element. The cocycle ϑ of the representation r will differ from the cocycle ϑ' of r' by a coboundary.

8.3 The projective representation of the Ptolemy groupoid on $\mathcal{K}(\varphi)$

Following [32] closely we shall define a projective representation of the Ptolemy groupoid in terms of the following set of unitary operators on $\mathcal{K}(\varphi)$:

$$\begin{aligned} \mathbf{A}_v &\equiv e^{\frac{\pi i}{3}} e^{-\pi i(\mathfrak{p}_v + \mathfrak{q}_v)^2} e^{-3\pi i \mathfrak{q}_v^2}, \\ \mathbf{T}_{vw} &\equiv e_b(\mathfrak{q}_v + \mathfrak{p}_w - \mathfrak{q}_w) e^{-2\pi i \mathfrak{p}_v \mathfrak{q}_w}, \end{aligned} \quad \text{where } v, w \in \varphi_\circ. \quad (8.9)$$

The special function $e_b(U)$ can be defined in the strip $|\Im z| < |\Im c_b|$, $c_b \equiv i(b + b^{-1})/2$ by means of the integral representation

$$\log e_b(z) \equiv \frac{1}{4} \int_{i0-\infty}^{i0+\infty} \frac{dw}{w} \frac{e^{-2izw}}{\sinh(bw) \sinh(b^{-1}w)}. \quad (8.10)$$

We refer to Appendix A for more details on this remarkable special function. These operators are unitary for $(1 - |b|)\Im b = 0$. They satisfy the following relations [32]

$$\mathbf{T}_{vw} \mathbf{T}_{uw} \mathbf{T}_{uv} = \mathbf{T}_{uv} \mathbf{T}_{vw}, \quad (8.11)$$

$$\mathbf{A}_v \mathbf{T}_{uv} \mathbf{A}_u = \mathbf{A}_u \mathbf{T}_{vu} \mathbf{A}_v, \quad (8.12)$$

$$\mathbf{T}_{vu} \mathbf{A}_u \mathbf{T}_{uv} = \zeta \mathbf{A}_u \mathbf{A}_v \mathbf{P}_{uv}, \quad (8.13)$$

$$\mathbf{A}_u^3 = \text{id}, \quad (8.14)$$

where $\zeta = e^{\pi i c_b^2/3}$, $c_b \equiv \frac{i}{2}(b + b^{-1})$. The relations (8.11) to (8.14) allow us to define a projective representation of the Ptolemy groupoid as follows.

- Assume that $\omega_{uv} \in [\varphi', \varphi]$. To ω_{uv} let us associate the operator

$$\mathfrak{u}(\omega_{uv}) \equiv \mathbf{T}_{uv} : \mathcal{K}(\varphi) \ni \mathfrak{v} \rightarrow \mathbf{T}_{uv} \mathfrak{v} \in \mathcal{K}(\varphi').$$

- For each fat graph φ and vertices $u, v \in \varphi_\circ$ let us define the following operators

$$\mathbf{A}_u^\varphi : \mathcal{K}(\varphi) \ni \mathfrak{v} \rightarrow \mathbf{A}_u \mathfrak{v} \in \mathcal{K}(\rho_u \circ \varphi),$$

$$\mathbf{P}_{uv}^\varphi : \mathcal{K}(\varphi) \ni \mathfrak{v} \rightarrow \mathbf{P}_{uv} \mathfrak{v} \in \mathcal{K}((uv) \circ \varphi).$$

It follows immediately from (8.11)–(8.14) that the operators \mathbf{T}_{uv} , \mathbf{A}_u and \mathbf{P}_{uv} can be used to generate a unitary projective representation of the Ptolemy groupoid in $\mathcal{K}(\varphi) \simeq L^2(\mathbb{R}^{2M})$.

8.4 Reduction to the quantized Teichmüller spaces

Theorem 4. *The isomorphism \mathfrak{l}_φ maps the operators $\mathfrak{u}(\pi)$ which represent the Ptolemy groupoid on $\mathcal{K}(\varphi)$ to operators of the form $\mathfrak{u}'(\pi) = \mathbf{V}_z(\pi) \otimes \mathbf{V}_h(\pi)$, where $\mathbf{V}_z(\pi) \equiv (\mathbf{V}_\mathfrak{f}(\pi))_{\mathfrak{f} \in B'(\Sigma)}$ is a family of unitary operators $\mathbf{V}_\mathfrak{f}(\pi)$ on $\mathcal{H}_\mathfrak{f}(\varphi)$.*

For each fixed $\mathfrak{f} \in B'(\Sigma)$ one may use the operators $\mathbf{V}_\mathfrak{f}(\pi)$ to generate a unitary projective representation of the Ptolemy groupoid.

Proof. To begin with, let us note that each path $\pi \in \rho \equiv [\varphi', \varphi]$ canonically defines a map $c \mapsto c' \equiv \mathfrak{A}(c)$ for each $c \in H_1(\Sigma, \mathbb{R})$. This map is defined in an obvious way for the elementary moves depicted in Figures 6, 7 if we require that c' coincides with c outside the triangles depicted in these figures.

Lemma 7. a) *The map $c \mapsto \mathfrak{A}(c)$ preserves the symplectic (intersection) form on $H_1(\Sigma, \mathbb{R})$.*

b) *The operator $u(\pi)$ maps $u(\pi) \cdot h_{\varphi,c} \cdot (u(\pi))^{-1} = h_{\varphi',\mathfrak{A}(c)}$.*

Proof. Direct verifications. □

Proposition 5. *For each path $\pi \in [\varphi', \varphi]$ there exists an operator $H(\pi)$ on $\mathcal{K}(\varphi)$ such that the operators $V(\pi)$ on $L^2(\mathbb{R}^{2M})$ defined by*

$$V(\pi) = H(\pi) \cdot U(\pi), \quad U(\pi) \equiv J_{\varphi'}^{-1} \cdot u(\pi) \cdot J_{\varphi},$$

- (i) *commute with all operators $h_m, h_m^\vee, m = 1, \dots, g$ and $f_l, l = 1, \dots, s - 1$,*
- (ii) *generate a unitary projective representation of $\text{Pt}(\Sigma)$ on $L^2(\mathbb{R}^{2M})$.*

The operators $H(\pi)$ can be represented in the form

$$H(\pi) \equiv \exp(iH_\pi(h_{\varphi'})), \tag{8.15}$$

where $H_\pi(h_{\varphi'})$ is a quadratic function of the $2g$ operators $h'_1, \dots, h'_g, h_1^\vee, \dots, h_g^\vee$ on $\mathcal{K}(\varphi')$.

Proof. The existence of operators $H(\pi)$ of the form (8.15) which are such that statement (i) of the proposition is verified follows directly from Lemma 6 if one takes into account that the transformation $h_{\varphi',c} \rightarrow h_{\varphi',c'}$ is represented by an element of $\text{Sp}(g, \mathbb{R})$ according to part a) of Lemma 7.

In order to prove statement (ii) of the proposition, we mainly need to check that the operators $V(\pi)$ satisfy the relations of the Ptolemy groupoid. Let us consider a closed path $\pi \in [\varphi, \varphi]$ which decomposes into a chain of edges as $\pi = \pi_n \circ \dots \circ \pi_1$.

$$V(\pi) \equiv V(\pi_n) \dots V(\pi_1)$$

On the one hand one may observe that $V(\pi)$ can be factorized as

$$V(\pi) = H(\pi) \cdot U(\pi_n) \dots U(\pi_1) = H(\pi)\zeta_\varphi(\pi). \tag{8.16}$$

The operator $H(\pi)$ in (8.16) can be represented as follows:

$$H(\pi) = H(\pi_n) \cdot [U(\pi'_n) \cdot H(\pi_{n-1}) \cdot U(\pi'_n)^\dagger] \dots [U(\pi'_1) \cdot H(\pi_1) \cdot U(\pi'_1)^\dagger],$$

where $\pi'_{n-k} \equiv \pi_n \circ \dots \circ \pi_{n-k}$. It follows from Lemma 7 together with (8.15) that

$$U(\pi'_{j+1}) \cdot H(\pi_j) \cdot U(\pi'_{j+1})^\dagger = \exp(iH_{\pi_j}(h_\varphi)).$$

Taking into account equation (7.15) we conclude that

$$K^{-1} \cdot H(\pi) \cdot K = 1 \otimes H_h(\pi). \tag{8.17}$$

On the other hand let us note that $V(\pi)$ commutes with all operators h_m, h_m^\vee , $m = 1, \dots, g$. Equation (8.16) implies that the same is true for $H(\pi)$. However, the representation of the operators h_m, h_m^\vee , $m = 1, \dots, g$ on $\mathcal{H}_h(\varphi)$ is irreducible (see Remark 3). This allows us to conclude that $H(\pi) = \eta_\varphi(\pi) \in \mathbb{C}$, $|\eta_\varphi(\pi)| = 1$. Inserting this into (8.16) proves our claim. \square

It follows from statement (i) in the proposition that

$$K^{-1} \cdot V(\pi) \cdot K = V_z(\pi) \otimes 1. \tag{8.18}$$

The task remains to describe the operators $V_z(\pi)$ more precisely.

Proposition 6. *There exists a family of unitary operators $V_f(\pi) : \mathcal{H}_f(\varphi) \rightarrow \mathcal{H}_f(\varphi')$, $f \in B'(\Sigma)$ which represent the action of the operator $V_z(\pi) : \mathcal{H}_z(\varphi) \rightarrow \mathcal{H}_z(\varphi')$.*

Proof. To begin with, let us observe that it follows from (8.9) that the operators $u(E)$ associated to the edges $E = \rho_v$, $E = \omega_{vw}$ in $\text{Pt}_1(\Sigma)$ can all be factorized as $u(E) = Q(E) \cdot G(z_e)$, where $Q(E)$ is of the form $Q(E) = \exp(iQ_E(v))$ for a quadratic function Q_E , and $G \equiv 1$ if $E = \rho_v$ and $G(z) = e_b(z)$ if $E = \omega_{vw}$. It follows that the corresponding operator $V(E)$ defined in Proposition 5 can be factorized as

$$V(E) = Q'(E) \cdot G(z_e), \quad \text{where } Q'(E) \equiv H(E) \cdot J_{\varphi'}^{-1} \cdot Q(E) \cdot J_\varphi. \tag{8.19}$$

The operator $Q'(E)$ is a product of four operators $J_{\gamma_k} \in \text{Sp}(2M, \mathbb{R})$ for $k = 1, 2, 3, 4$. It follows from Lemma 6 that it can be represented in the form

$$Q'(E) = \exp(iJ_E(v)),$$

for some expression $J_E(v)$ which is quadratic in v . Note that the operators $G(z_e)$ and $V(E)$ commute with all operators h_m, h_m^\vee , $m = 1, \dots, g$ and f_l , $l = 1, \dots, s - 1$. It follows that the same is true for $Q'(E)$, which implies that $J_E(v) \equiv J_E(z)$ depends only on the vector $z \equiv (z_e)_{e \in \varphi_1}$. $V(E)$ is therefore of the form $V(E) = \exp(iJ_E(z)) \cdot G(z_e)$. Our claim follows easily, $W_f(E) = \exp(iJ_E(z_{f,e})) \cdot G(z_{f,e})$ does the job. \square

Theorem 4 follows by combining Propositions 5 and 6. \square

We are finally in the position to define more precisely what we will regard as the quantized Teichmüller spaces. To this aim let us note that the Hilbert spaces $\mathcal{H}_o(\varphi)$ associated to the origin o in $B'(\Sigma)$ form irreducible representations of the relations (7.5), (7.6). Functions of the operators $z_{o,e}$ generate the algebras $\mathbf{B}(\mathcal{H}_o(\varphi))$ of all bounded operators on $\mathcal{H}_o(\varphi)$, which suggests to interpret $\mathbf{B}(\mathcal{H}_o(\varphi))$ as particular representations of the quantized algebras $\mathcal{T}_\hbar(\Sigma)$ of functions on the Teichmüller spaces.

The operators $V_o(\pi)$ generate a unitary projective representation of the Ptolemy groupoid which allows us to regard two operators $O_{\varphi_2} \in \mathbf{B}(\mathcal{H}_o(\varphi_2))$ and $O_{\varphi_1} \in \mathbf{B}(\mathcal{H}_o(\varphi_1))$ as equivalent, $O_{\varphi_2} \sim O_{\varphi_1}$, iff

$$O_2 = V_o(\pi) \cdot O_1 \cdot (V_o(\pi))^{-1}, \quad \pi \in [\varphi_2, \varphi_1]. \tag{8.20}$$

Let furthermore MC_φ be the unitary projective representation of the mapping class group $\text{MC}(\Sigma)$ which is generated from the operators $V_\circ(\pi)$ by means of the construction in Section 8.2.

Definition 3. (i) We define the algebra $\mathcal{T}_\hbar(\Sigma)$ as the algebra generated by the families $\mathcal{O} \equiv (\mathcal{O}_\varphi)_{\varphi \in \mathcal{P}_{t_\circ}(\Sigma)}$ of bounded operators \mathcal{O}_φ on $\mathcal{H}_\circ(\varphi)$ such that $\mathcal{O}_{\varphi_2} \sim \mathcal{O}_{\varphi_1}$ for all $\varphi_2, \varphi_1 \in \mathcal{P}_{t_\circ}(\Sigma)$. The algebra $\mathcal{T}_\hbar(\Sigma)$ will be called the quantized algebra of functions on the Teichmüller spaces.

(ii) Let $\mathcal{MC}_\hbar(\Sigma)$ be the subalgebra of $\mathcal{T}_\hbar(\Sigma)$ generated by the families $\mathcal{MC}(\mu) \equiv (\text{MC}_\varphi(\mu))_{\varphi \in \mathcal{P}_{t_\circ}(\Sigma)}$ for all $\mu \in \text{MC}(\Sigma)$.

Part III. A stable modular functor from the quantum Teichmüller spaces

Let us recall that systems of Fenchel–Nielsen coordinates are naturally associated to markings of the surfaces Σ . The transformations between the different markings of a Riemann surface Σ generate yet another groupoid, called the stable modular groupoid. Given that the quantum version of the changes between the Penner coordinates associated to different fat graphs was represented by a unitary projective representation of the Ptolemy groupoid it is natural to expect that the quantization of the Fenchel–Nielsen coordinates should similarly come with a unitary projective representation of the modular groupoid.

Pants decompositions have one big advantage over ideal triangulations: The gluing operation allows us to build “larger” surfaces from simple pieces of the same type, namely hyperbolic surfaces with holes. It is natural to require that the unitary projective representations of the modular groupoid assigned to surfaces Σ should correspondingly be organized in a “tower-like” fashion: They should allow restriction to, and should be generated by the representations assigned to the surfaces Σ' which are obtained from Σ by cutting along simple closed curves on Σ .

Our aim in the rest of this chapter will be to show that such a structure can be constructed from the quantized Teichmüller spaces considered in the previous part. This is of great importance since

having a tower of projective unitary representations of the stable modular groupoid is equivalent to having a stable unitary modular functor.

The notion of a stable unitary modular functor will be introduced in the next section. One main difference to the more conventional (two-dimensional) modular functors as defined e.g. in [52], [6] is that one restricts attention to Riemann surfaces Σ of genus g and with n parametrized boundary components which are stable in the sense that $2g - 2 + n > 0$.

We will then explain why having a stable unitary modular functor is equivalent to having a tower of projective unitary representations of the modular groupoid before we take up the task to actually construct the latter from the quantization of the Teichmüller spaces as described previously.

9 The notion of a stable unitary modular functor

Given that the usual definitions of a modular functor take several pages to fully write them down [52], [6], we shall only briefly explain the most important features. The missing details will not differ much from the definitions discussed in [52], [6].

9.1 Rigged Riemann surfaces

We will consider compact oriented surfaces Σ with boundary $\partial\Sigma = \coprod_{\beta \in A(\Sigma)} b_\beta$, where $A(\Sigma) \equiv \pi_0(\partial\Sigma)$ is the set of connected components of $\partial\Sigma$. A surface Σ is called an extended surface if one has chosen orientation-preserving homeomorphisms $p_\beta: b_\beta \rightarrow S^1$ for each connected component b_β of the boundary. To be concrete, let $S^1 = \{z \in \mathbb{C}; |z| = 1\}$. An e-surface Σ of genus g and with n boundary circles is called stable if $2g - 2 + n > 0$.

We will use the terminology rigged Riemann surface, or r-surface for short, for triples $\hat{\Sigma} = (\Sigma, y, c)$, where

- ▷ Σ is a stable extended surface,
- ▷ y is a Lagrangian subspace of $H_1(\Sigma, \mathbb{R})$, and
- ▷ $c: A(\Sigma) \rightarrow \mathcal{L}$ is a coloring of the boundary of Σ by elements of a set \mathcal{L} .

Given an r-surface $\hat{\Sigma}$ and given $\beta, \beta' \in A(\Sigma)$ such that $c(\beta) = c(\beta')$ we can define a new r-surface $\hat{\Sigma}' \equiv \sqcup_{\beta\beta'} \hat{\Sigma} = (\Sigma', y', c')$ which is obtained from $\hat{\Sigma}$ by gluing the boundary components b_β and $b_{\beta'}$.

- ▷ The surface $\Sigma' \equiv \sqcup_{\beta\beta'} \Sigma$ is defined by identifying all points $p \in b_\beta$ with $(p_{\beta'}^{-1} \circ p_\beta^o)(p) \in b_{\beta'}$, where $p_\beta^o: b_\beta \rightarrow S^1$ is defined by $p_\beta^o(p) = \overline{p_\beta(p)}$. There is a corresponding projection $P_{\beta\beta'}: \Sigma \rightarrow \Sigma' \equiv \sqcup_{\beta\beta'} \Sigma$ which maps $b_\beta, b_{\beta'}$ to the same simple closed curve on Σ' .
- ▷ The Lagrangian subspace y' is given by the image of $H_1(\Sigma, \mathbb{R})$ under the projection $P_{\beta\beta'}$.
- ▷ The coloring c' is obtained from c by putting $c'(\alpha') = c(\alpha)$ if $P_{\beta\beta'}(b_\alpha) = b_{\alpha'} \in \partial\Sigma'$.

9.2 Stable unitary modular functors

Let \mathcal{L} now be a space with a measure $d\nu$. A stable modular functor with central charge is the following collection of data.

▷ Assignment $\hat{\Sigma} = (\Sigma, y, \mathfrak{c}) \rightarrow \mathcal{H}(\Sigma, y, \mathfrak{c})$, where

- $\hat{\Sigma}$ is an r -surface,
- $\mathcal{H}(\Sigma, y, \mathfrak{c})$ is a separable Hilbert space.

▷ *Mapping class group*: Assignment

$$[f] \longrightarrow (\mathbf{U}_{[f]}: \mathcal{H}(\Sigma, y, \mathfrak{c}) \rightarrow \mathcal{H}(\Sigma', y'_f, \mathfrak{c}'_f)),$$

- $[f]$ is the isotopy class of a homeomorphism $f: \Sigma \rightarrow \Sigma'$,
- $\mathbf{U}_{[f]}$ is a unitary operator,
- y'_f is the Lagrangian subspace of $H_1(\Sigma', \mathbb{R})$ determined from y via f ,
- $\mathfrak{c}'_f: A(\Sigma') \rightarrow \mathcal{L}$: the coloring of boundary components of Σ' induced from \mathfrak{c} via f .

▷ *Disjoint union*: There exist unitary operators

$$\mathbf{G}_{2,1}: \mathcal{H}(\Sigma_2 \sqcup \Sigma_1, y_2 \oplus y_1, \mathfrak{c}_2 \sqcup \mathfrak{c}_1) \xrightarrow{\sim} \mathcal{H}(\Sigma_2, y_2, \mathfrak{c}_2) \otimes \mathcal{H}(\Sigma_1, y_1, \mathfrak{c}_1).$$

▷ *Gluing*: Let (Σ', y') be obtained from (Σ, y) by gluing of two boundary components α, β . There then exists a unitary operator

$$\mathbf{G}_{\alpha\beta}: \int_{\mathcal{L}} d\nu(s) \mathcal{H}(\Sigma, y, \mathfrak{c}_{s|\alpha\beta}) \xrightarrow{\sim} \mathcal{H}(\Sigma', y', \mathfrak{c}'),$$

where the coloring $\mathfrak{c}_{s|\alpha\beta}: A(\Sigma) \rightarrow \mathcal{L}$ is defined from \mathfrak{c}' via

$$\begin{aligned} \mathfrak{c}_{s|\alpha\beta}(c) &= \mathfrak{c}'(P_{\alpha\beta}(c)) & \text{if } c \in A(\Sigma) \setminus \{\alpha, \beta\}, \\ \mathfrak{c}_{s|\alpha\beta}(c) &= s & \text{if } c \in \{\alpha, \beta\}. \end{aligned}$$

These data are required to satisfy the following “obvious” consistency and compatibility conditions.

Multiplicativity. For all homeomorphisms $f: \Sigma_1 \rightarrow \Sigma_2, g: \Sigma_2 \rightarrow \Sigma_3$ there exists $\zeta(f, g) \in S^1$ such that we have

$$\mathbf{U}_{[f]}\mathbf{U}_{[g]} = \zeta(f, g)\mathbf{U}_{[f \circ g]}. \quad (9.1)$$

$\zeta(f, g)$ has to satisfy the condition $\zeta(f, g)\zeta(f \circ g, h) = \zeta(g, h)\zeta(f, g \circ h)$.

Functoriality. The gluing isomorphisms and the disjoint union isomorphisms are functorial in Σ .⁷

⁷One is considering the category with objects r -surfaces, and morphisms isotopy classes of homeomorphisms of r -surfaces, equipped additionally with the gluing and disjoint union operations.

Compatibility. The gluing isomorphisms and the disjoint union isomorphisms are mutually compatible.

Symmetry of gluing. $G_{\alpha\beta} = G_{\beta\alpha}$.

It would take us several pages to write out all conditions in full detail, we therefore refer to [52], [6] for more details. However, it seems that the following two “naturality” requirements represent a key to the understanding of the notion of the modular functor:

Naturality. a) Let $f_1 : \Sigma_1 \rightarrow \Sigma'_1$, $f_2 : \Sigma_2 \rightarrow \Sigma'_2$ be r -homeomorphisms. We then have

$$G_{21} \cdot U_{[f_2 \sqcup f_1]} \cdot G_{21}^\dagger \equiv U_{[f_2]} \otimes U_{[f_1]}.$$

b) If the r -homeomorphism $f : \Sigma_1 \rightarrow \Sigma_2$ induces an r -homeomorphism $f' : \Sigma'_1 \rightarrow \Sigma'_2$ of the surfaces Σ'_1, Σ'_2 obtained from Σ_1, Σ_2 by the gluing construction, we have

$$G_{\alpha\beta} \cdot \left(\int_{\mathcal{L}} dv(s) U_{[f]}(\Sigma, y, c_{s|\alpha\beta}) \right) \cdot G_{\alpha\beta}^\dagger = U_{[f']}(\Sigma', y', c').$$

These requirements make clear how the mapping class group representations on the spaces $\mathcal{H}(\Sigma', y', c')$ restrict to and are generated by the representations assigned to the surfaces Σ which are obtained from Σ' by cutting along simple closed curves on Σ' .

Remark 7. The standard definitions of modular functors assume that the Hilbert spaces $\mathcal{H}(\Sigma, y, c)$ are *finite-dimensional*. They are therefore not suitable for *nonrational* conformal field theories. Our definition should be seen as a first step towards the definition of analogs of modular functors which are associated to nonrational conformal field theories in a way similar to the connections between rational conformal field theories and modular functors mentioned in the introduction.

However, there is one important ingredient of the usual definition that does not have an obvious counterpart in our framework. In the more standard definitions of modular functors it is required that there exists a distinguished element s_\circ in \mathcal{L} which has the property that coloring a boundary component β with s_\circ is equivalent to “closing” this boundary component. More precisely, let $c_{s|\beta}$ be a coloring of the boundary components of an extended surface Σ which assigns $s \in \mathcal{L}$ to the component with label β , and let $\hat{\Sigma}_\beta = (\Sigma_\beta, y_\beta, c_\beta)$ be the r -surface obtained from $\hat{\Sigma} = (\Sigma, y, c)$ by gluing a disc to b_β . The more standard definitions of modular functors assume or imply existence of an element s_\circ in \mathcal{L} such that

$$\mathcal{H}(\Sigma, y, c_{s_\circ|\beta}) \simeq \mathcal{H}(\Sigma_\beta, y_\beta, c_\beta), \quad U_{[f]}(\Sigma, y, c_{s_\circ|\beta}) \simeq U_{[f]}(\Sigma_\beta, y_\beta, c_\beta).$$

This yields additional relations between the mapping class group representations assigned to surfaces with different numbers of boundary components.

In the case of the quantized Teichmüller spaces it ultimately turns out that an analog of the element s_\circ in \mathcal{L} exists only if one considers the analytic continuation of $\mathcal{H}(\Sigma, y, c)$ and $U_{[f]}(\Sigma, y, c)$ with respect to the boundary labels $c(\beta) \in \mathcal{L}$. This fact, and the corresponding improvement of our definition of a stable modular functor will be elaborated upon elsewhere.

9.3 Representations of the modular groupoid versus modular functors

It turns out to be very useful to reformulate the notion of a stable unitary modular functor in terms of generators and relations as follows. Any surface Σ' can be glued from a surface $\Sigma_\circ = \coprod_{p \in \sigma_\circ} T_p$ which is a disjoint union of trinions. The different ways of doing this can be parametrized⁸ by markings σ . The gluing construction determines a canonical Lagrangian subspace y'_σ of $H_1(\Sigma', \mathbb{R})$ from the tautological Lagrangian subspace $y_\circ \equiv H_1(\Sigma_\circ, \mathbb{R})$. By iterating the gluing and disjoint union isomorphisms one defines unitary operators $\mathbb{G}(\sigma, \sigma'): \mathcal{H}(\Sigma', y'_\sigma, \mathbf{c}') \rightarrow \mathcal{H}(\sigma, \mathbf{c})$, where

$$\mathcal{H}(\sigma, \mathbf{c}) \equiv \int_{\mathcal{L}} d\nu_\sigma(S) \bigotimes_{p \in \sigma_\circ} \mathcal{H}(T_p, \mathbf{c}_{\sigma,p}^S). \quad (9.2)$$

We have used the following notation:

- The integration is extended over the set \mathcal{L} of all colorings $S: \mathcal{C}_\sigma \ni c \rightarrow s_c \in \mathcal{L}$ equipped with the canonical product measure $d\nu_\sigma(S)$ obtained from $d\nu$ by choosing any numbering of the elements of \mathcal{C}_σ .
- $\mathbf{c}_{\sigma,p}^S$ is the coloring of the boundary components of T_p which is defined by assigning

$$\mathbf{c}_{\sigma,p}^S(\beta) = \begin{cases} \mathbf{c}'(\beta') & \text{if } P_{\sigma,p}(\beta) = \beta' \in A(\Sigma'), \\ s_c & \text{if } P_{\sigma,p}(\beta) = c \in \mathcal{C}_\sigma. \end{cases}$$

$P_{\sigma,p}(\beta)$ is the embedding $T_p \hookrightarrow \Sigma'$ defined from σ by the gluing construction.

Unitary mappings between the different spaces $\mathcal{H}(\sigma, \mathbf{c})$ arise in two ways: First, one may have different markings σ_2, σ_1 such that the Lagrangian subspaces defined by the gluing construction coincide, $y_{\sigma_1} = y_{\sigma_2}$. In this case one gets unitary operators $\mathbb{F}_{\sigma_2\sigma_1}(\mathbf{c}): \mathcal{H}(\sigma_1, \mathbf{c}) \rightarrow \mathcal{H}(\sigma_2, \mathbf{c})$ from the composition

$$\mathbb{F}_{\sigma_2\sigma_1}(\mathbf{c}) \equiv \mathbb{G}(\sigma_2, \mathbf{c}) \cdot (\mathbb{G}(\sigma_1, \mathbf{c}))^\dagger. \quad (9.3)$$

Secondly, one has the mappings $\mathbb{U}_{[f]}$ which may map between spaces $\mathcal{H}(\Sigma, y_\sigma, \mathbf{c})$ and $\mathcal{H}(\Sigma', (y_\sigma)'_f, \mathbf{c}'_f)$. In the case that $\mu: \Sigma \rightarrow \Sigma'$ represents an element of the mapping class group it is natural to define operators $\mathbb{V}_{\sigma\mu}(\mathbf{c}): \mathcal{H}(\sigma, \mathbf{c}) \rightarrow \mathcal{H}(\sigma'_\mu, \mathbf{c}'_\mu)$ by

$$\mathbb{V}_{\sigma\mu}(\mathbf{c}) = \mathbb{G}(\sigma'_\mu, \mathbf{c}'_\mu) \cdot \mathbb{U}_{[\mu]} \cdot (\mathbb{G}(\sigma, \mathbf{c}))^\dagger, \quad (9.4)$$

where σ'_μ is the image of the marking σ under μ .

It turns out – as will be reviewed in the next section – that there exists a set $\mathcal{M}_1(\Sigma)$ of elementary moves between the different markings σ such that any two markings σ_2, σ_1 can be connected by paths π which are composed out of the elementary moves. There furthermore exists a set $\mathcal{M}_2(\Sigma)$ of relations which makes the resulting two-dimensional CW complex $\mathcal{M}(\Sigma)$ simply connected. The corresponding path groupoid is called the modular groupoid and denoted by $\mathbb{M}(\Sigma)$.

⁸Note that markings with the same cut system will yield equivalent representations for Σ' . This redundancy will give a useful book-keeping device when the representation of the mapping class group is considered.

Our construction of the operators $F_{\sigma_2\sigma_1}(\mathfrak{c})$, $V_{\sigma\mu}(\mathfrak{c})$ assigns unitary operators to each of the elementary moves . This yields unitary operators $U(\pi, \mathfrak{c}): \mathcal{H}(\sigma_1, \mathfrak{c}) \rightarrow \mathcal{H}(\sigma_2, \mathfrak{c}_\pi)$ for each path $\pi \in [\sigma_2, \sigma_1]$, where \mathfrak{c}_π is the coloring of boundary components which is obtained from \mathfrak{c} by tracking the relabeling of boundary components defined by π . Due to the fact that multiplicativity holds only projectively, cf. eqn. (9.1), one will find that there exists $\zeta(\pi) \in S^1$ such that $U(\pi) = \zeta(\pi)$ for all *closed* paths, starting and ending at the same marking σ . In other words, a stable unitary modular functor canonically defines projective unitary representations of the stable modular groupoid.

Conversely, the definition of a *tower of projective unitary representations of the modular groupoids* $M(\Sigma)$ involves the following data.

- ▷ Assignment $(\sigma, \mathfrak{c}) \rightarrow \mathcal{H}(\sigma, \mathfrak{c})$, where $\mathcal{H}(\sigma, \mathfrak{c})$ is constructed out of spaces $\mathcal{H}(S_3, \mathfrak{c}_3)$ assigned to trinions as in (9.2).
- ▷ Assignment

$$\pi \in [\sigma_2, \sigma_1] \longrightarrow (U(\pi, \mathfrak{c}): \mathcal{H}(\sigma_1, \mathfrak{c}) \rightarrow \mathcal{H}(\sigma_2, \mathfrak{c}_\pi)).$$

It is required that the operators $U(\pi, \mathfrak{c})$ generate unitary projective representations of the modular groupoids associated to the surfaces Σ . In order to formulate the additional requirements which turn this collection of representations of $M(\Sigma)$ into a *tower* let us note that markings can be glued in a natural way. We will use the notation $\sqcup_{\alpha\beta}\sigma$ for the marking obtained by gluing boundary components α and β . Gluing and disjoint union now become realized in a trivial manner,

$$\mathcal{H}(\sigma_2 \sqcup \sigma_1, \mathfrak{c}_2 \sqcup \mathfrak{c}_1) = \mathcal{H}(\sigma_2, \mathfrak{c}_2) \otimes \mathcal{H}(\sigma_1, \mathfrak{c}_1), \tag{9.5}$$

$$\mathcal{H}(\sqcup_{\alpha\beta}\sigma, \mathfrak{c}') = \int_{\mathcal{L}} d\nu(s) \mathcal{H}(\sigma, \mathfrak{c}_{s|\alpha\beta}). \tag{9.6}$$

The data specified above are then required to satisfy the following naturality conditions.

Naturality. a) Let $\pi = \pi_2 \sqcup \pi_1 \in [\sigma'_2, \sigma_2] \sqcup [\sigma'_1, \sigma_1]$. We then have

$$U(\pi_2 \sqcup \pi_1, \mathfrak{c}_2 \sqcup \mathfrak{c}_1) \equiv U(\pi_2, \mathfrak{c}_2) \otimes U(\pi_1, \mathfrak{c}_1). \tag{9.7}$$

b) Given a path $\pi \in [\sigma_2, \sigma_1]$ in $\mathcal{M}(\Sigma)$ let $\sqcup_{\alpha\beta}\pi \in [\sqcup_{\alpha\beta}\sigma_2, \sqcup_{\alpha\beta}\sigma_1]$ be the corresponding path in $\mathcal{M}(\sqcup_{\alpha\beta}\Sigma)$ defined by the gluing construction. We then have

$$U(\sqcup_{\alpha\beta}\pi, \mathfrak{c}') = \int_{\mathcal{L}} d\nu(s) U(\pi, \mathfrak{c}_{s|\alpha\beta}). \tag{9.8}$$

From a tower of projective unitary representations of the modular groupoids one can reconstruct a stable unitary modular functor as follows. The system of isomorphisms $U(\pi, \mathfrak{c}): \mathcal{H}(\sigma_1, \mathfrak{c}) \rightarrow \mathcal{H}(\sigma_2, \mathfrak{c})$ allows us to identify the Hilbert spaces associated to different markings σ , thereby defining $\mathcal{H}(\Sigma, \mathfrak{c})$. The representation of the modular groupoid defines a representation of the mapping class group on $\mathcal{H}(\Sigma, \mathfrak{c})$ via the construction in Section 8.1. We refer to [6] for more details.

It is important to note (see next section for a detailed discussion) that the elements of the set $\mathcal{M}_1(\Sigma)$ of elementary moves only change the markings within subsurfaces of genus zero with three or four holes, or within subsurfaces of genus one with one hole. The corresponding operators will be called Moore–Seiberg data. They characterize a tower of representations of the modular groupoid completely. Let us furthermore note that the faces/relations that one needs to define a two-dimensional CW complex $\mathcal{M}(\Sigma)$ which has as set of edges $\mathcal{M}_1(\Sigma)$ turn out to involve only subsurfaces of genus zero with three to five holes, and subsurfaces of genus one with one or two holes. This means that one only needs to verify a finite number of relations to show that a given set of Moore–Seiberg data defines a stable unitary modular functor.

10 The modular groupoid

The modular groupoid is the groupoid generated from the natural transformations relating the different markings of a Riemann surface Σ [38], [5]. A complete set of generators and relations has been determined in [38], [5], [20].

Remark 8. The CW complex $\mathcal{M}(\Sigma)$ defined below will be a subcomplex of the complex denoted $\mathcal{M}^{\max}(\Sigma)$ in [5], since that reference allowed for cut systems that yield connected components with less than three boundary components, whereas we will exclusively consider cut systems that yield connected components with exactly three holes. All other deviations from [5] are due to slightly different conventions in the definition of the generators.

10.1 Notations and conventions

We have a unique curve $c(\sigma, e) \in \mathcal{C}'$ associated to each edge $e \in \sigma_1$ of the marking graph Γ_σ . The trinions $T_p \in \mathcal{P}_\sigma$ are in one-to-one correspondence with the vertices $p \in \sigma_o$ of Γ_σ .

The choice of a distinguished boundary component c_p for each trinion T_p , $p \in \sigma_o$ will be called the *decoration* of the marking graph Γ_σ . The distinguished boundary component c_p will be called *outgoing*, the other two boundary components of T_p *incoming*. Two useful graphical representation for the decoration are depicted in Figure 10.

10.2 Generators

The set of edges $\mathcal{M}_1(\Sigma)$ will be given by elementary moves denoted as (pq) , Z_p , B_p , F_{pq} and S_p . The indices $p, q \in \sigma_o$ specify the relevant trinions within the pants decomposition of Σ that is determined by σ . The move (pq) will simply be the operation in which the labels p and q get exchanged. Graphical representations for the elementary moves Z_p , B_p , F_{pq} and S_p are given in Figures 11–14.

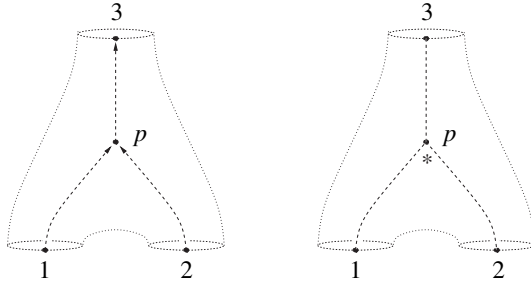


Figure 10. Two representations for the decoration on a marking graph.

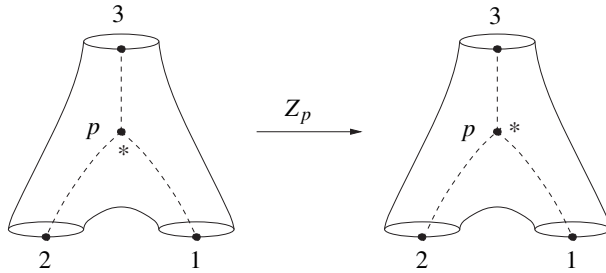


Figure 11. The Z -move.

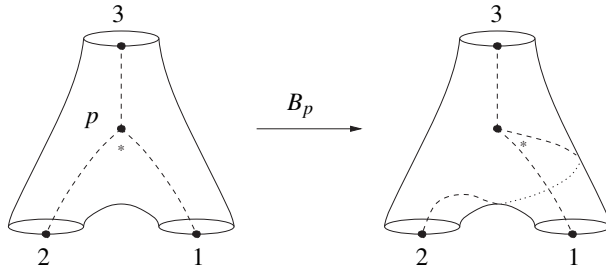


Figure 12. The B -move.

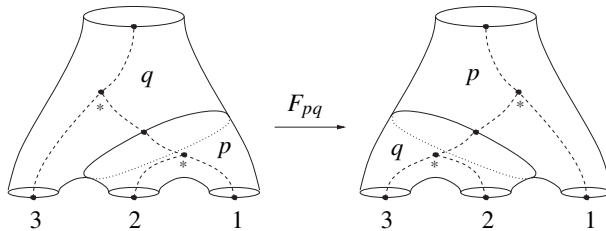


Figure 13. The F -move.

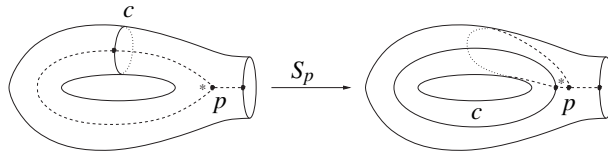


Figure 14. The S-move.

10.3 Relations

The relations of $M(\Sigma)$ correspond to the faces $\varpi \in \mathcal{M}_2(\Sigma)$. In the following we will define a set \mathcal{R} of faces which is large enough to make the complex $\mathcal{M}(\Sigma)$ simply connected. A face ϖ in $\mathcal{M}_2(\Sigma)$ may be characterized by choosing a chain $\Gamma_\varpi = E_{\varpi, n(\varpi)} \circ \dots \circ E_{\varpi, 1}$, where $E_{\varpi, i} \in \mathcal{M}_1(\Sigma)$ for $i = 1, \dots, n(\varpi)$. In order to simplify notation we will generically factorize Γ_ϖ as $\Gamma_\varpi = \Gamma_\varpi^2 \circ \Gamma_\varpi^1$ and write $(\Gamma_\varpi^2)^{-1} = \Gamma_\varpi^1$ instead of $\Gamma_\varpi = \text{id}$.

Locality. Let us introduce the notation $\text{supp}(m)$ by $\text{supp}(m) = \{p\}$ if $m = Z_p, B_p, S_p$, $\text{supp}(m) = \{p, q\}$ if $m = (pq), F_{pq}$ and $\text{supp}(m_2 \circ m_1) = \text{supp}(m_2) \cup \text{supp}(m_1)$. We then have

$$m_2 \circ m_1 = m_1 \circ m_2 \quad \text{whenever } \text{supp}(m_1) \cap \text{supp}(m_2) = \emptyset. \tag{10.1}$$

We will list the remaining relations ordered by the topological type of the surfaces on which the relevant graphs can be drawn.

Relations supported on surfaces of genus zero.

$$g = 0, s = 3: \quad Z_p \circ Z_p \circ Z_p = \text{id}. \tag{10.2}$$

$$g = 0, s = 4: \quad \begin{aligned} \text{a) } & F_{qp} \circ B_p \circ F_{pq} = (pq) \circ B_q \circ F_{pq} \circ B_p, \\ \text{b) } & F_{qp} \circ B_p^{-1} \circ F_{pq} = (pq) \circ B_q^{-1} \circ F_{pq} \circ B_p^{-1}, \\ \text{c) } & A_{pq} \circ A_{qp} = (pq). \end{aligned} \tag{10.3}$$

$$g = 0, s = 5: \quad F_{qr} \circ F_{pr} \circ F_{pq} = F_{pq} \circ F_{qr}. \tag{10.4}$$

We have used the abbreviation

$$A_{pq} \equiv Z_q^{-1} \circ F_{pq} \circ Z_q^{-1} \circ Z_p. \tag{10.5}$$

In Figures 15 and 16 we have given diagrammatic representations for relations (10.3), b) and (10.4) respectively.

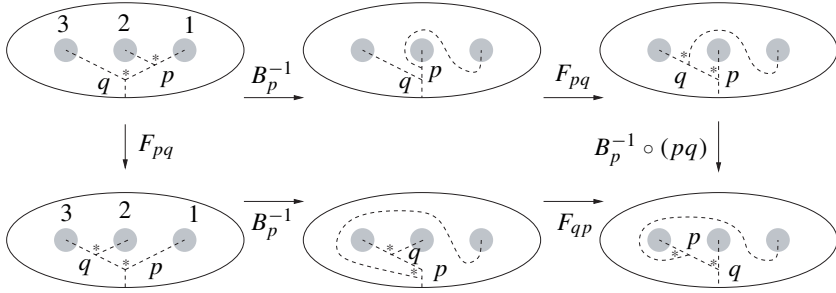


Figure 15. The hexagon relation.

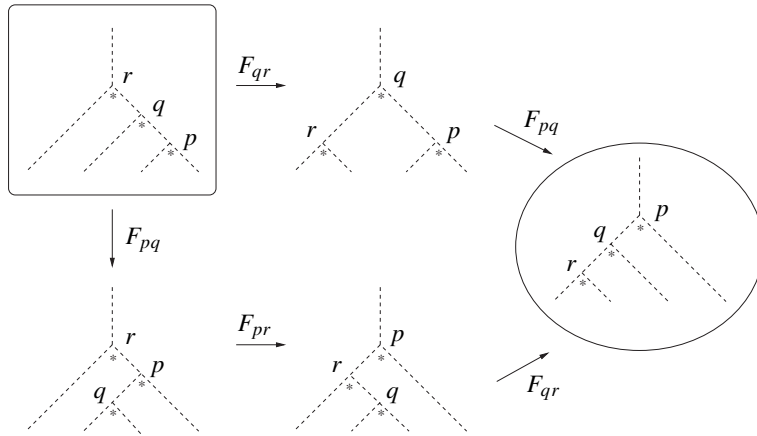


Figure 16. The pentagon identity.

Relations supported on surfaces of genus one. In order to write the relations transparently let us introduce the following composites of the elementary moves.

$$\begin{aligned}
 g = 0, s = 3: \quad & \text{a) } B'_p \equiv Z_p^{-1} \circ B_p \circ Z_p^{-1}, \\
 & \text{b) } T_p \equiv Z_p^{-1} \circ B_p \circ Z_p \circ B_p,
 \end{aligned}
 \tag{10.6}$$

$$g = 0, s = 4: \quad B_{qp} \equiv Z_q^{-1} \circ F_{qp}^{-1} \circ B'_q \circ F_{pq}^{-1} \circ Z_q^{-1} \circ (pq),
 \tag{10.7}$$

$$g = 1, s = 2: \quad S_{qp} \equiv (F_{qp} \circ Z_q)^{-1} \circ S_p \circ (F_{qp} \circ Z_q).
 \tag{10.8}$$

It is useful to observe that the move T_p represents the Dehn twist around the boundary component of the trinion \mathfrak{t}_p numbered by $i = 1$ in Figure 10. With the help of these

definitions we may write the relations supported on surfaces of genus one as follows:

$$\begin{aligned}
 g = 1, s = 1: \quad & \text{a) } S_p^2 = B'_p, \\
 & \text{b) } S_p \circ T_p \circ S_p = T_p^{-1} \circ S_p \circ T_p^{-1}. \tag{10.9}
 \end{aligned}$$

$$g = 1, s = 2: \quad B_{qp} = S_{qp}^{-1} \circ T_q^{-1} T_p \circ S_{pq}. \tag{10.10}$$

Relation (10.10) is represented diagrammatically in Figure 17.

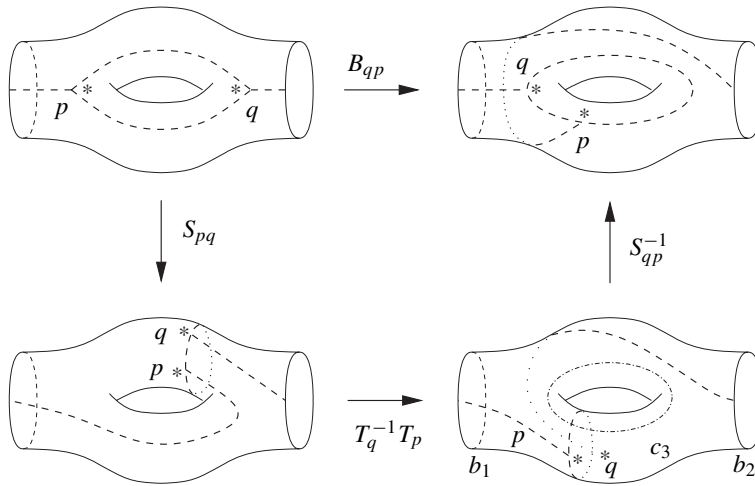


Figure 17. Relation for the two-punctured torus.

Theorem 5. *The complex $\mathcal{M}(\Sigma)$ is connected and simply connected for any e-surface Σ .*

Proof. The theorem follows easily from [5], Theorem 5.1. We noted previously that our complex $\mathcal{M}(\Sigma)$ differs from the complex $\mathcal{M}^{\max}(\Sigma)$ of [5] in having a set of vertices which corresponds to decompositions of Σ into connected components with *exactly* three holes. The edges of $\mathcal{M}_1^{\max}(\Sigma)$ which correspond to the F-move of [5] simply can't appear in $\mathcal{M}_1(\Sigma)$. Otherwise the set of edges of $\mathcal{M}(\Sigma)$ coincides with the relevant subset of $\mathcal{M}_1^{\max}(\Sigma)$, with the exception that our move B_p is a composition of the B- and the Z-move of [5]. To complete the proof it remains to check that our set of faces is equivalent to the subset of $\mathcal{M}_2^{\max}(\Sigma)$ which involves only the vertices $\mathcal{M}_0(\Sigma)$ of our smaller complex $\mathcal{M}(\Sigma)$. This is a useful exercise. \square

Definition 4. Let $\mathcal{M}'(\Sigma)$ be the complex which has the same set of vertices as $\mathcal{M}(\Sigma)$, a set of edges given by the moves

$$(pq), Z_p, B_p, B'_p, S_p, T_p, F_{pq}, A_{pq}, B_{pq}, S_{pq} \tag{10.11}$$

defined above, as well as faces given by equations (10.1)–(10.10).

11 From markings to fat graphs

A key step in our construction of a stable modular functor from the quantized Teichmüller spaces will be the definition of a distinguished class of fat graphs φ_σ which are associated to the elements σ of a certain subset of the set $\mathcal{M}_o(\Sigma)$ of all markings of Σ .

Definition 5. Let $\mathcal{A}_\sigma \subset \mathcal{C}_\sigma$ be the set of all curves c which are incoming for both adjacent trinions. We will say that a marking σ is admissible iff there is no curve $c \in \mathcal{C}_\sigma$ which is outgoing for both adjacent trinions, and if cutting Σ along all curves $c \in \mathcal{A}_\sigma$ yields connected components all of which have genus zero. The set of all admissible markings will be denoted by $\mathcal{M}_o^{\text{ad}}(\Sigma)$.

To each admissible marking σ we may naturally associate a fat graph φ_σ on Σ by the following construction. In order to construct the fat graph φ_σ , it will be useful to consider a certain refinement of the pants decomposition associated to σ which is defined as follows. For each curve $c \in \mathcal{A}_\sigma$ let A_c be a small annular neighborhood of c which contains c in its interior, and which is bounded by a pair of curves (c^+, c^-) that are isotopic to c . If $c \in A(\Sigma)$ represents a boundary component of Σ we may similarly consider an annular subset B_c of Σ bounded by $c^+ \equiv c$ and another curve c^- isotopic to c . If $c \in \mathcal{C}_\sigma$ belongs to neither of these two classes we will simply set $c^+ \equiv c \equiv c^-$. By cutting along all such curves c^\pm we obtain a decomposition of Σ into trinions T_p , $p \in \sigma_o$ and annuli A_c , $c \in \mathcal{A}_\sigma$, each equipped with a marking graph. We may then replace the markings on each of these connected components by fat graphs according to Figures 18–20. The re-gluing of trinions T_p , $p \in \sigma_o$ and annuli A_c , B_c to recover the surface Σ may then be performed in such a way that the fat graphs on the connected components glue to a fat graph φ_σ on Σ .

11.1 The complex $\mathcal{M}^{\text{ad}}(\Sigma)$

It is natural to consider the complex $\mathcal{M}^{\text{ad}}(\Sigma)$ for which the set of vertices $\mathcal{M}_o^{\text{ad}}(\Sigma)$ is the subset of $\mathcal{M}_o(\Sigma)$ which consists of the admissible markings, and which has a set of edges $\mathcal{M}_1^{\text{ad}}(\Sigma)$ given by the subset of $\mathcal{M}'_1(\Sigma)$ that contains those edges which connect two admissible markings.

Proposition 7. *The complex $\mathcal{M}^{\text{ad}}(\Sigma)$ is connected and simply connected.*

Proof. Let us consider two admissible markings $\sigma, \sigma' \in \mathcal{M}_o(\Sigma)$. There exists a path ϖ in $\mathcal{M}(\Sigma)$ which connects σ and σ' . This path may be represented as a chain $C_\varpi = E_{\varpi, n(\varpi)} \circ \cdots \circ E_{\varpi, 1}$, $E_{\varpi, i} \in \mathcal{M}_i(\Sigma)$ composed out of the moves Z_p, F_{pq}, B_p and S_p .

For a given marking $\sigma \in \mathcal{M}_o(\Sigma)$ let $[\sigma]$ be the set of all markings σ' which differ from σ only in the choice of decoration. The moves Z_p act transitively on $[\sigma]$. By

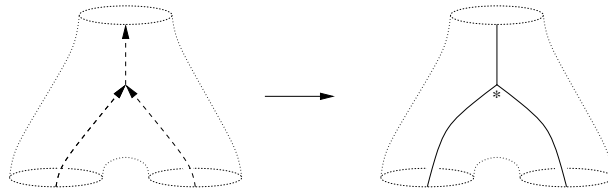


Figure 18. Substitution for a trinion T .

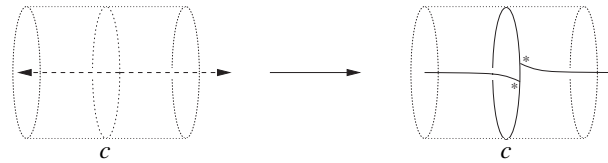


Figure 19. Substitution for an annulus A .

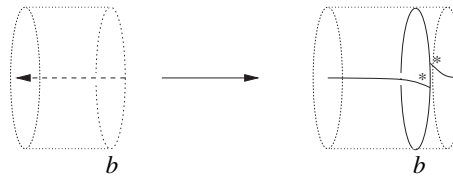


Figure 20. Substitution for a boundary component B .

inserting Z_p moves if necessary we may therefore modify C_{ϖ} to a chain D_{ϖ} which takes the form

$$D_{\varpi} = Z_{\varpi, n(\varpi)} \circ F_{\varpi, n(\varpi)} \circ Z_{\varpi, n(\varpi)-1} \circ \cdots \circ F_{\varpi, 1} \circ Z_{\varpi, 0},$$

where $Z_{\varpi, i}$, $i = 0, \dots, n(\varpi)$ are chains composed out of Z_p -moves only, and the moves $F_{\varpi, n(\varpi)} \in \mathcal{M}_1(\Sigma)$ connect markings $\tilde{\sigma}_{\varpi, i}$ and $\sigma_{\varpi, i}$, $i = 1, \dots, n(\varpi)$ which are *admissible*. We clearly must have $[\sigma_{\varpi, i+1}] = [\tilde{\sigma}_{\varpi, i}]$, $i = 1, \dots, n(\varpi) - 1$ and $[\sigma_{\varpi, 1}] = [\sigma]$, $[\sigma_{\varpi, n(\varpi)}] = [\sigma']$. Connectedness of $\mathcal{M}^{\text{ad}}(\Sigma)$ would follow if the chains $Z_{\varpi, i}$, $i = 0, \dots, n(\varpi)$ are homotopic to chains $Y_{\varpi, i}$ which represent paths in $\mathcal{M}^{\text{ad}}(\Sigma)$. That this indeed the case follows from the following lemma.

Lemma 8. *Assume that $\sigma_a, \sigma_b \in \mathcal{M}_0^{\text{ad}}(\Sigma)$ satisfy $\sigma_b \in [\sigma_a]$. There then exists a path ϖ_{ab} in $\mathcal{M}^{\text{ad}}(\Sigma)$ which connects σ_a and σ_b .*

Sketch of proof. Let us recall that a marking σ is *irreducible* if it is admissible and if there are no edges e in \mathcal{A}_{σ} such that cutting Σ along $c(e, \sigma)$ yields two disconnected components. A marking σ which is irreducible has only one outgoing external edge.

With the help of Lemma 17 in Appendix D it is easy to show that for an admissible graph σ there always exists a sequence of Z_p -moves in $\mathcal{M}_1^{\text{ad}}(\Sigma)$ which transforms σ to a graph σ' that is irreducible.

We may and will therefore assume that σ_a and σ_b are both irreducible. Using Lemma 17 again allows us to transform σ_b to a marking σ_c which is such that the outgoing external edges of σ_a and σ_c correspond to the same outgoing boundary component of Σ .

The graphs σ_a and σ_c can finally be connected by a chain which is composed out of moves B_p and F_{qp} it only. This follows from the connectedness of $\mathcal{M}(\Sigma_\circ)$ in the case of a surface Σ_\circ of genus zero [38]. In this way one constructs a sequence of moves that connects σ_a to σ_b □

It remains to prove that $\mathcal{M}^{\text{ad}}(\Sigma)$ is simply connected as well. Let us consider any closed path ϖ in $\mathcal{M}^{\text{ad}}(\Sigma)$. The path ϖ is contractible in $\mathcal{M}'(\Sigma)$. The deformation of ϖ to a trivial path may be performed recursively, face by face. The crucial observation to be made is the following one.

Lemma 9. *The paths ϖ which represent the boundaries of the faces of $\mathcal{M}'_2(\Sigma)$ are paths in $\mathcal{M}^{\text{ad}}(\Sigma)$.*

Proof. By direct inspection of the relations (10.2)–(10.10). □

Lemma 9 implies that deforming a path $\varpi \in \mathcal{M}^{\text{ad}}(\Sigma)$ by contracting a face in $\mathcal{M}'_2(\Sigma)$ will produce a path ϖ' which still represents a path in $\mathcal{M}^{\text{ad}}(\Sigma)$. It follows that ϖ is contractible in $\mathcal{M}^{\text{ad}}(\Sigma)$. □

11.2 Separated variables

We had observed in Section 6 that the Fock variables $z_e, e \in \varphi'_1$ form a set of coordinates for the Teichmüller spaces of surfaces with holes. When considering fat graphs φ_σ associated to a marking σ it will be useful to replace the Fock variables $z_e, e \in \varphi'_1$ by an alternative set of coordinates $(q_c, p_c), c \in \mathcal{C}_\sigma$ for T_φ which satisfy

$$\begin{aligned} \Omega_\varphi(p_{c_1}, p_{c_2}) &= 0 = \Omega_\varphi(q_{c_1}, q_{c_2}), & c_1, c_2 \in \mathcal{C}_\sigma. \\ \Omega_\varphi(p_{c_1}, q_{c_2}) &= \delta_{c_1, c_2}, \end{aligned} \tag{11.1}$$

These coordinates are constructed as follows.

For $c \in \mathcal{A}_\sigma$ let A_c be an annular neighborhood of c such that the part of φ_σ which is contained in A_c is isotopic to the model depicted in Figure 21.

Let e_1 and e_2 be the two edges that are entirely contained in A_c with labelling defined by Figure 21. Out of z_{e_1} and z_{e_2} we may then define

$$q_c \equiv \frac{1}{2}(z_{e_1} - z_{e_2}), \quad p_c \equiv -\frac{1}{2}(z_{e_1} + z_{e_2}). \tag{11.2}$$

In the case $c \in \mathcal{C}_\sigma \setminus \mathcal{A}_\sigma$ let us note that there is a unique trinion T_c for which the curve c is the outgoing boundary component. The part of φ_σ contained in T_c is depicted in Figure 22.

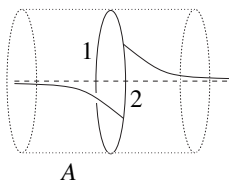


Figure 21. Annulus A_c and fat graph φ on A .

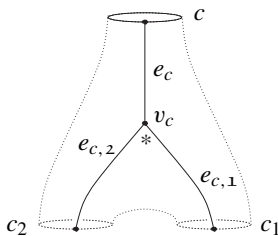


Figure 22. Part of φ_σ contained in T_c .

Let $e_{c,2}$ and $e_{c,1}$ be the edges of φ_σ as indicated in Figure 22, and let c_ε , $\varepsilon = 1, 2$ be the curves which represent the corresponding boundary components of T_p . We shall then define

$$q_c \equiv y_{c_2}, \quad p_c \equiv -y_{c_1}, \tag{11.3}$$

where

$$y_{c_\varepsilon} \equiv z_{e_{c,\varepsilon}} + \frac{1}{2} f_{c_\varepsilon}, \quad \varepsilon = 1, 2. \tag{11.4}$$

The same construction yields linear combinations (\hat{p}_c, \hat{q}_c) , $c \in \mathcal{C}_\sigma$ of the Kashaev variables which represent the coordinates (p_c, q_c) within W_{φ_σ} . Note that our construction of the fat graph φ_σ implies that T_c contains a unique vertex $v_c \in \varphi_{\sigma,0}$. It turns out that the variables \hat{q}_c, \hat{p}_c have a simple relation to the Kashaev variables q_{v_c}, p_{v_c} .

Lemma 10. (i) $\hat{q}_c = q_{v_c} + h_{c_2}, \hat{p}_c = p_{v_c} - h_{c_1}$.

(ii) $\Omega_\varphi(p_c, q_{c'}) = \delta_{c,c'}, c, c' \in \mathcal{C}_\sigma$.

Proof. In order to prove part (i) let us first look at the vertices $v \in \varphi_{\sigma,0}$ which appear on the graph geodesic homotopic to c . The contribution to y_{c_ε} of the edges that are incident to v is $\hat{z}_{e'_1} + \hat{z}_{e'_2} + \hat{z}_{e'_3}$ if e'_i , $i = 1, 2, 3$ are the three edges incident to v . It then follows from (4.6) that y_c does not depend on both p_v and q_v . What remains are contributions from the vertex v_c , as well as contributions from the vertices contained in annuli $A_{c'}$ which are determined as follows. Note that the homology class $[c]$ can be decomposed as a linear combination of classes $[c']$ with $c' \in \mathcal{A}_\sigma$. It is straightforward to check that each $c' \in \mathcal{A}_\sigma$ which appears in this decomposition yields a contribution $-h_{c'}$ to \hat{f}_c . These contributions sum up to give $-h_c$. What remains

is the contribution from the vertex v_c . Part (i) of the lemma now follows easily by recalling the definition (4.6).

Part (ii) of the lemma is a trivial consequence of part (i). □

11.3 Quantized Teichmüller spaces for surfaces with holes

The coordinates introduced in the previous section make it straightforward to modify the discussion of the quantization of Teichmüller spaces from the case of punctured Riemann surfaces (Section 7) to the case of Riemann surfaces with holes.

Bearing in mind that the Fock variables $z_e, e \in \varphi'_1$ are unconstrained in the present case (see Lemma 5) leads us to identify the algebra of functions on the Teichmüller spaces with the algebras of function of the variables $z_e, e \in \varphi'_1$. A convenient set of coordinates is given by the coordinate functions $(p_c, q_c), c \in \mathcal{C}_\sigma$. Canonical quantization of the Teichmüller spaces is therefore straightforward, and leads to an algebra of operators with generators $(p_c, q_c), c \in \mathcal{C}_\sigma$, which is irreducibly represented (in the sense of Remark 3) on the Hilbert space $\mathcal{H}_z(\sigma) \simeq L^2(\mathbb{R}^{3g-3+2s})$. We will normalize the operators $(p_c, q_c), c \in \mathcal{C}_\sigma$, such that

$$\begin{aligned} \text{(i)} \quad & [p_c, q_{c'}] = (2\pi i)^{-1} \delta_{cc'}, \quad c, c' \in \mathcal{C}_\sigma, \\ \text{(ii)} \quad & [p_c, p_{c'}] = 0 = [q_c, q_{c'}]. \end{aligned} \tag{11.5}$$

Quantization of the Kashaev space W_{φ_σ} produces a *reducible* representation of the algebra (11.5) which is generated by operators (\hat{q}_c, \hat{p}_c) associated to the pairs of variables (\hat{q}_c, \hat{p}_c) . It also yields operators $h_c, c \in H_1(\Sigma, \mathbb{R})$ which represent the quantization of the Poisson vector space H_{φ_σ} with basis $h_c, c \in H_1(\Sigma, \mathbb{R})$. The algebra generated by the $h_c, c \in H_1(\Sigma, \mathbb{R})$ has a center generated by the $h_c, c \in B(\Sigma)$. Following the discussion at the beginning of Section 7.3 one constructs a representation of this algebra on the space

$$\mathcal{H}_h(\sigma) \equiv \int_{B'(\Sigma)} d\mathfrak{f} \mathcal{H}_{h, \mathfrak{f}}(\sigma),$$

where $\mathcal{H}_{h, \mathfrak{f}}(\sigma) \simeq L^2(\mathbb{R}^g)$ is an irreducible representation of the algebra $i[h_{c_2}, h_{c_1}] = b^2 I(c_2, c_1)$ for $c_2, c_1 \in H_1(\Sigma_{cl}, \mathbb{R})$.

The following Proposition 8 describes how the quantized Teichmüller spaces are related to the quantized Kashaev space.

Proposition 8. *There exists a unitary operator l_σ ,*

$$l_\sigma : \mathcal{K}(\varphi_\sigma) \rightarrow \mathcal{H}_z(\sigma) \otimes \mathcal{H}_h(\sigma)$$

such that

$$l_\sigma \cdot \hat{p}_c \cdot l_\sigma^{-1} = p_c \otimes 1, \quad l_\sigma \cdot \hat{q}_c \cdot l_\sigma^{-1} = q_c \otimes 1 \quad \text{and} \quad l_\sigma \cdot h_c \cdot l_\sigma^{-1} = 1 \otimes h'_c,$$

for any $e \in \sigma_1$ and $c \in H_1(\Sigma, \mathbb{R})$, respectively.

Let us recall that a move $m \in [\tau_m, \sigma_m]$, $m \in \mathcal{M}_1(\Sigma)$ is admissible if both τ_m and σ_m are admissible. Given an element $m \in \mathcal{M}_1^{\text{ad}}(\Sigma)$ it is natural to consider the corresponding fat graphs $\theta_m \equiv \varphi_{\tau_m}$, $\varphi_m \equiv \varphi_{\sigma_m}$ on Σ , to pick a path $\pi_m \in [\theta_m, \varphi_m]$ and consider the operator $u(m) \equiv u(\pi_m)$. The reduction to the quantized Teichmüller spaces proceeds as in Section 8.4. By multiplying the operators $l_\sigma \cdot u(m) \cdot l_\sigma^{-1}$ with suitably chosen operators $H(m)$ one gets operators $v(m)$ which factorize as $v(m) = w(m) \otimes 1$. The resulting operators $w(m): \mathcal{H}_z(\varphi_m) \rightarrow \mathcal{H}_z(\theta_m)$ will then generate a unitary projective representation of the path groupoid of $\mathcal{M}^{\text{ad}}(\Sigma)$.

12 Geodesic length operators

Of fundamental importance for us will be to define and study quantum analogs of the geodesic length functions on the Teichmüller spaces, the geodesic length operators.

12.1 Overview

When trying to define operators which represent the geodesic length functions one has to face the following difficulty: The classical expression for $L_{\varphi,c} \equiv 2 \cosh \frac{1}{2} l_c$ as given by formula 5.6 is a linear combination of monomials in the variables $e^{\pm \frac{z_e}{2}}$ of a very particular form,

$$L_{\varphi,c} = \sum_{\tau \in \mathcal{F}} C_{\varphi,c}(\tau) e^{x(\tau)}, \quad x(\tau) \equiv \sum_{e \in \varphi_1} \tau(e) z_e, \tag{12.1}$$

where the summation is taken over the space \mathcal{F} of all maps $\varphi_1 \ni e \rightarrow \tau(e) \in \frac{1}{2}\mathbb{Z}$. The coefficients $C_{\varphi,c}(\tau)$ are positive integers, and non-vanishing for a finite number of $\tau \in \mathcal{F}$ only.

In the quantum case one is interested in the definition of length operators $L_{\varphi,c}$ which should be representable by expressions similar to (12.1),

$$L_{\varphi,c} = \sum_{\tau \in \mathcal{F}} C_{\varphi,c}^b(\tau) e^{x(\tau)}, \quad x(\tau) \equiv \sum_{e \in \varphi_1} \tau(e) z_e. \tag{12.2}$$

The following properties seem to be indispensable if one wants to interpret an operator of the general form (12.2) as the quantum counterpart of the functions $L_{\varphi,c} = 2 \cosh \frac{1}{2} l_c$:

- (a) *Spectrum.* $L_{\varphi,c}$ is self-adjoint. The spectrum of $L_{\varphi,c}$ is simple and equal to $[2, \infty)$. This is necessary and sufficient for the existence of an operator $l_{\varphi,c}$ – the *geodesic length operator* – such that $L_{\varphi,c} = 2 \cosh \frac{1}{2} l_c$.
- (b) *Commutativity.*

$$[L_{\varphi,c}, L_{\varphi,c'}] = 0 \quad \text{if } c \cap c' = \emptyset.$$

(c) *Mapping class group invariance.*

$$\mathbf{a}_\mu(L_{\varphi,c}) = L_{\mu,\varphi,c}, \quad \mathbf{a}_\mu \equiv \mathbf{a}_{[\mu,\varphi,\varphi]} \quad \text{for all } \mu \in \text{MC}(\Sigma).$$

(d) *Classical limit.* The coefficients $C_{\varphi,c}^b(\tau)$ which appear in (12.2) should be deformations of the classical coefficients $C_{\varphi,c}(\tau)$ in the sense that

$$\lim_{b \rightarrow 0} C_{\varphi,c}^b(\tau) = C_{\varphi,c}(\tau).$$

Length operators were first defined and studied in the pioneering work [16]. It was observed in [16] that the necessary deformation of the coefficients $C_{\varphi,c}^b(\tau)$ is indeed nontrivial in general. However, a full proof that the length operators introduced in [16] fulfil the requirements (a) and (c) does not seem to be available yet. We will therefore present an alternative approach to this problem, which will allow us to define length operators that satisfy (a)-(d) in full generality.

12.2 Construction of the length operators

Our construction of the length operators will proceed in two steps. First, we will construct length operators $L_{\sigma,c}$ in the case that the fat graph φ under consideration equals φ_σ . This will facilitate the verification of the properties (a)-(d) formulated above. In order to define the length operators $L_{\varphi,c}$ in the general case we shall then simply pick any marking σ such that the given curve c is contained in the cut system \mathcal{C}_σ , and define

$$L_{\varphi,c} \equiv \mathbf{a}_{[\varphi,\varphi_\sigma]}(L_{\sigma,c}). \tag{12.3}$$

Independence of this construction from the choice of σ will follow from Theorem 6 below. Definition (12.3) reduces the proof of properties (a), (b) to the proof of the corresponding statements for the length operators $L_{\sigma,c}$ which will be given below. Property (c) follows from $\mathbf{a}_{[\mu,\varphi,\varphi]} \circ \mathbf{a}_{[\varphi,\varphi_\sigma]} = \mathbf{a}_{[\mu,\varphi,\varphi_\sigma]}$.

In order to prepare for our construction of length operators it is useful us recall the construction of the fat graph φ_σ in Section 11.

Definition 6. *Case $c \in \mathcal{A}_\sigma$.* Let A be an annular neighborhood of c such that the part of φ_σ which is contained in A is isotopic to the model depicted in Figure 21. We will then define

$$L_{\varphi,c} \equiv e^{-2\pi b q_c} + 2 \cosh 2\pi b p_c. \tag{12.4}$$

Case $c \notin \mathcal{A}_\sigma$. If a curve c is not contained in \mathcal{A}_σ , it is necessarily the outgoing boundary component of a trinion T_p (cf. Figure 18). Let $c_\varepsilon, \varepsilon = 1, 2$ be the curves which represent the incoming boundary components of T_p as indicated in Figure 22. Given that $L_{\sigma;c_i}, i = 1, 2$ are already defined we will define $L_{\sigma;c}$ by

$$L_{\sigma,c} = 2 \cosh(y_{c_2} + y_{c_1}) + e^{-y_{c_2}} L_{\sigma,c_1} + e^{y_{c_1}} L_{\sigma,c_2} + e^{y_{c_1} - y_{c_2}}, \tag{12.5}$$

where $y_{c_\varepsilon}, \varepsilon = 1, 2$ are defined as $y_{c_2} = 2\pi b q_c, y_{c_1} = -2\pi b p_c$.

It is easy to see that this recursively defines length operators for all remaining $c \in \mathcal{C}_\sigma$.

Remark 9. Let us note that (11.5) implies that $[y_{c_\varepsilon}, L_{\sigma, c_{\varepsilon'}}] = 0$ for $\varepsilon, \varepsilon' \in \{1, 2\}$. We therefore do not have an issue of operator ordering in (12.5).

Proposition 9. *The length operators $L_{\sigma, c}, L_{\sigma, c'}$ associated to different curves $c, c' \in \mathcal{C}_\sigma$ commute with each other.*

Proof. Let us recall that cutting the surface Σ along all of the curves $c \in \mathcal{A}_\sigma$, yields a set of connected components which all have genus zero. For a given curve $c \in \mathcal{C}_\sigma \setminus \mathcal{A}_\sigma$ let Σ_c be the connected component which contains c . Cutting Σ_c along c produces two connected components. The component which has c as its outgoing boundary component will be denoted Σ'_c . It follows from Definition 6 that $L_{\sigma, c}$ is an operator function of the operators p_d and q_d , where $d \in \mathcal{C}_\sigma$ is contained in Σ'_c . The claim therefore follows immediately from (11.5) if Σ'_c and $\Sigma'_{c'}$ are disjoint.

Otherwise we have the situation that one of $\Sigma'_c, \Sigma'_{c'}$, say $\Sigma'_{c'}$ is a subsurface of the other. The crucial point to observe is that the resulting expression for $L_{\sigma, c}$ depends on the variables p_d and q_d associated to the subsurface $\Sigma'_{c'}$ exclusively via $L_{\sigma, c'}$. The claim therefore again follows from (11.5). \square

The following theorem expresses the consistency of our definition with the automorphisms induced by a change of the marking σ .

Theorem 6. *For a given curve c let $\sigma_i, i = 1, 2$ be markings such that c is contained in both cut systems \mathcal{C}_{σ_1} and \mathcal{C}_{σ_2} . We then have*

$$a_{[\varphi_{\sigma_2}, \varphi_{\sigma_1}]}(L_{\sigma_1, c}) = L_{\sigma_2, c}. \tag{12.6}$$

On the proof of Theorem 6: The description of the modular groupoid in terms of generators and relations given in Section 10 reduces the proof of Theorem 6 to the case that σ_2 and σ_1 are connected by one of the elementary moves m defined in Section 10.2. In order to reduce the proof of Theorem 6 to a finite number of verifications one would need to have simple standard choices for the paths $\pi_m \in [\varphi_{\sigma_2}, \varphi_{\sigma_1}]$ for all elementary moves m . Existence of such standard paths π_m turns out to be nontrivial, though. The task to find suitable paths π_m is particularly simple for a subclass of moves m which is defined as follows.

Definition 7. (i) Let $\pi \in [\varphi', \varphi]$ be a path in the complex $\mathcal{P}t(\Sigma)$ which is described by a sequence $S_\pi \equiv (\varphi' \equiv \varphi_n, \dots, \varphi_1 \equiv \varphi)$ of fat graphs such that consecutive elements of S_π are connected by edges in $\mathcal{P}t_1(\Sigma)$. We will say that $\pi \in [\varphi', \varphi]$ is realized locally in a subsurface $\Sigma' \hookrightarrow \Sigma$ if the restrictions of φ_i to $\Sigma \setminus \Sigma'$ coincide for all $i = 1, \dots, n$.

(ii) We will say that the move $m = [\tau_m, \sigma_m] \in \mathcal{M}'_1(\Sigma)$ can be realized locally if there exists a path $\pi \in [\theta_m, \varphi_m]$ in the complex $\mathcal{P}t(\Sigma)$ that is realized locally in Σ_m in the sense of (i).

For moves $m = [\tau_m, \sigma_m] \in \mathcal{M}_1(\Sigma)'$ that can be realized locally we may choose essentially the *same* path $\pi_m \in [\theta_m, \varphi_m]$ for all surfaces Σ into which Σ_m can be embedded. It is then crucial to observe the following fact

Proposition 10. *If $m \in \mathcal{M}'_1(\Sigma)$ is admissible, but can not be realized locally, there always exists a path ϖ_m which is (i) homotopic to m within $\mathcal{M}^{\text{ad}}(\Sigma)$, and (ii) takes the form*

$$\varpi_m = Y_m \circ m \circ Y_m^{-1}, \tag{12.7}$$

where Y_m is a chain composed out of Z_p -moves and F_{pq} -moves which can all be realized locally.

The proof of Proposition 10 is given in Appendix D. It therefore suffices to prove Theorem 6 in the case that σ_2 and σ_1 are connected by any elementary move m that can be realized locally. This amounts to a finite number of verifications which can be carried out by straightforward, but tedious calculations. Some details are given in the Appendix E. □

12.3 Spectrum

Theorem 7. *The spectrum of $L_{\sigma,c}$ is simple and equal to $[2, \infty)$.*

Proof. To begin with, let us consider the following simple model for the length operators:

$$L \equiv 2 \cosh 2\pi b p + e^{-2\pi b q}, \tag{12.8}$$

where p, q are operators on $L^2(\mathbb{R})$ which satisfy the commutation relations $[p, q] = (2\pi i)^{-1}$.

A basic fact is that L is self-adjoint. Indeed, being a sum of two *positive* self-adjoint operators L is self-adjoint on the intersection of the domains of the summands. The main spectral properties of this operator are summarized in the following proposition.

Proposition 11 ([33]). (i) *We have $\text{Spec}(L) = (2, \infty)$.*

(ii) *The spectrum of L in $L^2(\mathbb{R})$ is simple.*

Validity of Theorem 7 in the case $c \in \mathcal{A}_\sigma$ is a direct consequence of Proposition 11. It remains to treat the case $c \notin \mathcal{A}_\sigma$. We will keep the notations introduced in Definition 6. The main ingredient will be an operator $C_{\sigma,c}: \mathcal{H}(\sigma) \rightarrow \mathcal{H}(\sigma)$ which maps all length operators $L_{\sigma,c}$ to the simple standard form $L_{\sigma,c}^{\text{st}}$,

$$L_{\sigma,c}^{\text{st}} = 2 \cosh 2\pi b p_c + e^{-2\pi b q_c} \tag{12.9}$$

in the sense that the following commutation relations are satisfied:

$$C_{\sigma,c} \cdot L_{\sigma,c} = L_{\sigma,c}^{\text{st}} \cdot C_{\sigma,c}. \tag{12.10}$$

Definition 8. Let the unitary operator $C_{\sigma,c}$ be defined as

$$C_{\sigma,c}^{-1} = e_b(q_c - s_2) \frac{s_b(s_1 - p_c)}{s_b(s_1 + p_c)} e^{2\pi i s_2 q_c}, \tag{12.11}$$

where

$$s_\varepsilon = (2\pi b)^{-1} \operatorname{arcosh} \frac{1}{2} L_{\sigma,c_\varepsilon} \quad \varepsilon = 1, 2,$$

and the special function $s_b(x)$ is a close relative of $e_b(x)$ defined in the Appendix A.

Lemma 11. *The unitary operator $C_{\sigma,p}$ satisfies (12.10).*

Proof. Proposition 11 is proven by means of a direct calculation using the explicit form of $L_{\sigma,p}$ given in (12.5) and the functional equations (A.3) and (A.9). \square

The proof of Theorem 7 is thereby reduced to Proposition 11. \square

12.4 Relation with the Dehn twist generator

To round off the picture we shall now discuss, following [32], the relation between the length operators $L_{\varphi,c}$ and the operator $D_{\varphi,c}$ which represents the Dehn twist D_c along c .

A closed curve c will be called a curve of simple type if the connected components of $\Sigma \setminus c$ all have more than one boundary component. It is not hard to see (using the construction in Section 11, for example) that for curves of simple type there always exists a fat graph φ and an annular neighborhood A_c of c in which φ takes the form depicted in Figure 23.

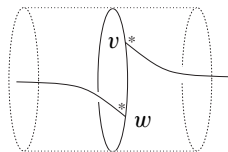


Figure 23. Annulus A_c and fat graph φ on A .

Lemma 12. *Let Σ be a surface with genus g and s holes. The pure mapping class group $MC(\Sigma)$ is generated by the Dehn twists along geodesics of simple type.*

Proof. For $g \geq 2$ it is known that the Dehn twists along non-separating closed curves suffice to generate the mapping class group $MC(\Sigma)$ [21]. Closed curves which are non-separating are always of simple type.

For $g = 0$ any closed curve is of simple type. In the remaining case $g = 1$ one may note that the only closed curve c which is not of simple type is the one which separates a one-holed torus from the rest of Σ . It is then well-known that the Dehn

twists along a - and b -cycles of the one-holed torus generate the Dehn-twist along c . If we supplement these generators by the Dehn twists along the remaining closed curves (which are all of simple type), we get a complete system of generators for $\text{MC}(\Sigma)$ [21]. \square

We may therefore use a fat graph which in an annular neighborhood of c takes the form depicted in Figure 23. It is easy to see that the action of the Dehn twist D_c on the fat graph φ can be undone by a single flip ω_{vw} . As the representative $F_{\varphi,c}$ of the Dehn twist D_c on $\mathcal{K}(\varphi)$ we may therefore choose $F_{\varphi,c} = T_{vw}$.

The operator $h_{\varphi,c}$ associated to the homology cycle c is $h_{\varphi,c} \equiv \frac{1}{2}(p_v + q_w)$. It is not very difficult to verify that the operator

$$D_{\varphi,c} = \zeta^{-6} \exp(2\pi i h_{\varphi,c}^2) F_{\varphi,c}. \tag{12.12}$$

commutes with all $h_c, c \in H_1(\Sigma, \mathbb{R})$. The prefactor $\zeta^{-6}, \zeta = e^{\pi i b^2/3}$ was inserted to define a convenient normalization. We then have the following result.

Proposition 12 ([33]). D_c coincides with the following function of the length operator l_c :

$$D_c = \zeta^{-6} \exp\left(i \frac{l_c^2}{8\pi b^2}\right).$$

This should be compared with the classical result that the geodesic length functions are the Hamiltonian generators of the Fenchel–Nielsen twist flow, which reproduces the Dehn twist for a twist angle of 2π .

13 Passage to the length representation

13.1 The length representation

Our aim is to define a representation for $\mathcal{H}(\Sigma)$ which is such that the length operators associated to a cut system are all realized as multiplication operators. For a surface Σ and a marking σ on Σ let

$$\mathcal{H}_L(\sigma) \equiv \mathcal{K}_{\text{sp}}^{\otimes \sigma_1}, \tag{13.1}$$

where $\mathcal{K}_{\text{sp}} \simeq L^2(\mathbb{R}_+, \eta)$ and η is the spectral measure of the operator $l_A = 2 \operatorname{arcosh} \frac{1}{2} L_A$. The numbering of the edges associated to σ defines canonical isomorphisms

$$\mathcal{H}_L(\sigma) \simeq \mathcal{H}_L(\Sigma) \equiv L^2(\mathcal{L}, d\eta_{\mathcal{L}}), \tag{13.2}$$

where $\mathcal{L} \simeq \mathbb{R}_+^{3g-3+2s}$ and $d\eta_{\mathcal{L}} = d\eta(l_1) \wedge \cdots \wedge d\eta(l_{3g-3+2s})$ is the corresponding product measure. For $e \in \sigma_1$ and $f: \mathbb{R}^+ \rightarrow \mathbb{C}$ let us define the multiplication operator $m_e[f]$ by

$$(m_e[f]\Psi)(\lambda_{\sigma}) \equiv f(l_e)\Psi(\lambda_{\sigma}). \tag{13.3}$$

The operator $m_e[f]$ will represent a bounded operator $m_e[f]: \mathcal{H}_L(\sigma) \rightarrow \mathcal{H}_L(\sigma)$ iff $f \in L^\infty(\mathbb{R}^+, d\eta)$.

For the rest of this section let us consider an admissible marking σ and the associated fat graph φ_σ . Definition 6 yields a family of mutually commuting length operators $L_{\sigma,c}$ associated to the curves c which constitute the cut system \mathcal{C}_σ . It follows from the spectral theorem⁹ for the family $\{L_e; e \in \sigma_1\}$ of self-adjoint operators that there exists a unitary operator

$$D(\sigma): \mathcal{H}_z(\sigma) \longrightarrow \mathcal{H}_L(\sigma)$$

that diagonalizes the set of length operators $L_{\sigma,e}$, $e \in \sigma_1$ in the sense that

$$D(\sigma) \cdot L_{\sigma,e} = m_e[2 \cosh \frac{1}{2}l] \cdot D(\sigma). \tag{13.4}$$

Our aim for the rest of this section will be to give a recursive construction for such an operator $D(\sigma)$ in terms of operators C_p which are associated to the vertices $p \in \sigma_0$ of σ .

13.2 Construction of the operator $D(\sigma)$

The main ingredient will be an operator $C_\sigma: \mathcal{H}(\sigma) \rightarrow \mathcal{H}(\sigma)$ which maps all length operators $L_{\sigma,c}$ to the simple standard form L_c^{st} ,

$$L_{\sigma,c}^{\text{st}} = e^{-2\pi b q_c} + 2 \cosh 2\pi b p_c. \tag{13.5}$$

The operators $C_{\sigma,c}$ that were defined in Definition 8 solve this task locally for every curve $c \in \mathcal{C}_\sigma \setminus \mathcal{A}_\sigma$. If the trinion T_p has c_p as the outgoing boundary component we will define

$$C_{\sigma,p} \equiv C_{\sigma,c_p}. \tag{13.6}$$

We are now ready to define

$$C_\sigma \equiv \prod_{p \in \sigma_0} C_{\sigma,p}. \tag{13.7}$$

Let us note that we do not have to specify the order in which the operators $C_{\sigma,p}$ appear thanks to the following lemma.

Lemma 13. *We have $C_{\sigma,p} C_{\sigma,q} = C_{\sigma,q} C_{\sigma,p}$.*

Proof. This follows from Proposition 9 and Definition 6, keeping in mind (11.5). \square

In order to construct the sought-for operator $D(\sigma)$ it now remains to map the length operators $L_{\sigma,e}^{\text{st}}$ to multiplication operators. Proposition 11 ensures existence of an operator

$$d_{\sigma,e}: L^2(\mathbb{R}) \rightarrow \mathcal{K}_{\text{sp}} \quad \text{such that} \quad d_{\sigma,e} \cdot L_e^{\text{st}} = m_e[2 \cosh \frac{1}{2}l] \cdot d_{\sigma,e}.$$

⁹See Appendix B for the precise statement.

Out of E_σ we may finally define the operator $D(\sigma)$ as

$$D(\sigma) \equiv d_\sigma \cdot C_\sigma, \quad d_\sigma \equiv \bigotimes_{e \in \sigma_1} d_{\sigma,e}, \tag{13.8}$$

where l_σ is the operator introduced in Proposition 4. It is straightforward to verify that the operator $D(\sigma)$ indeed satisfies the desired property (13.4).

14 Realization of $M(\Sigma)$

14.1 Two constructions for the generators

Our aim is to define operators $U(m)$ associated to the edges $m \in \mathcal{M}_1(\Sigma)$. We will give two constructions for these operators, each of which makes certain properties manifest. The proof of the equivalence of these two constructions will be the main difficulty that we will have to deal with.

The first construction. Let us recall that a move $m \in [\tau_m, \sigma_m]$, $m \in \mathcal{M}_1(\Sigma)$ is admissible if both τ_m and σ_m are admissible. Given an element $m \in \mathcal{M}_1^{\text{ad}}(\Sigma)$ it is natural to consider the corresponding fat graphs $\theta_m \equiv \varphi_{\tau_m}$, $\varphi_m \equiv \varphi_{\sigma_m}$ on Σ , to pick a path $\pi_m \in [\theta_m, \varphi_m]$ and define operators $\tilde{U}(m): \mathcal{H}_L(\sigma_m) \rightarrow \mathcal{H}_L(\tau_m)$ as

$$\tilde{U}(m) \equiv D(\tau_m) \cdot u(\pi_m) \cdot D(\sigma_m)^\dagger. \tag{14.1}$$

The second construction. We note that for all $m \in \mathcal{M}_1(\Sigma)$ the markings τ_m and σ_m will coincide outside of a subsurface $\Sigma_m \hookrightarrow \Sigma$. Let us therefore consider the restrictions τ'_m and σ'_m of τ_m and σ_m to Σ_m respectively. Admissibility of τ'_m and σ'_m is obvious for $m \in \mathcal{M}_1(\Sigma)$, allowing us to use the first construction in order to define an operator

$$U'(m): \mathcal{H}_L(\sigma'_m) \rightarrow \mathcal{H}_L(\tau'_m)$$

Out of $U'(m)$ we may then construct the sought-for operator $U(m)$ by acting with $U'(m)$ non-trivially only on those tensor factors of $\mathcal{H}_L(\sigma) = \mathcal{K}_{\text{sp}}^{\otimes \sigma_1}$ which correspond to the subsurface $\Sigma_m \hookrightarrow \Sigma$. More precisely, let $\mathbb{E}_m \subset \sigma_1$ be the set of edges in σ_1 which have nontrivial intersection with Σ_m . Out of $U'(m)$ let us then define the operator $U(m): \mathcal{K}(\sigma_m) \rightarrow \mathcal{K}(\tau_m)$ by applying definition (B.5) to the case $\mathbb{O} \equiv U'(m)$ and $\mathbb{J} \equiv \mathbb{E}_m$.

Comparison. The crucial difference between $\tilde{U}(m)$ and $U(m)$ is that the latter is manifestly acting *locally* in $\mathcal{H}_L(\sigma)$, in the sense that it acts only on the tensor factor of $\mathcal{H}_L(\sigma)$ which corresponds to the subsurface Σ_m . This is not obvious in the case of $\tilde{U}(m)$.

The length operators $L_{\sigma,c}$ associated to the boundary components $c \in A(\Sigma)$ form a commutative family of operators. The joint spectral decomposition for this family of operators leads us to represent $\mathcal{H}_L(\Sigma)$ as

$$\mathcal{H}_L(\Sigma) \simeq \int_{\mathcal{L}} d\eta_{\Sigma}(c) \mathcal{H}_L(\Sigma, c), \quad (14.2)$$

where the integration is extended over the set $\mathcal{L} \simeq \mathbb{R}_+^s$ of all colorings c of the boundary by elements of \mathbb{R}_+ . It follows from Theorem 6 that the operators $U(m)$ and $\tilde{U}(m)$ commute (up to permutations of the boundary components) with the length operators $L_{\sigma,c}$, $c \in A(\Sigma)$. Within the representation (14.2) one may therefore¹⁰ represent the operators $U(m)$ and $\tilde{U}(m)$ by families of operators $(U(m, c))_{c \in \mathcal{L}}$ and $(\tilde{U}(m, c))_{c \in \mathcal{L}}$.

14.2 The main result

The following theorem is the main result of this chapter.

Theorem 8. *The operators $U(m, c)$, $c \in \mathcal{L}$, generate a tower of projective unitary representations of the modular groupoids $M(\Sigma)$.*

The proof of Theorem 8 will take up the rest of this section.

To begin with, let us note that the necessary structure (9.2) of the Hilbert spaces follows trivially from our definition of the length representation in Section 13, where in the present case we simply have $\mathcal{H}(S_3, c_3) \simeq \mathbb{C}$. It is furthermore clear that the operators $\tilde{U}(m, c)$, $c \in \mathcal{L}$ generate a unitary projective representation of the modular groupoid $M(\Sigma)$ for each surface Σ within the considered class. Let us finally note that the naturality properties formulated in Section 9.3 clearly hold for the operators $U(m, c)$, $c \in \mathcal{L}$. Our main task is therefore to show that $\tilde{U}(m, c) = U(m, c)$, as will be established in Proposition 13 below.

Proposition 13. *For all $m \in \mathcal{M}_1^{\text{ad}}(\Sigma)$ there exists a path $\pi_m \in [\theta_m, \varphi_m]$ such that we have*

$$\tilde{U}(m, c) = U(m, c).$$

Proof. When we compare the respective definitions of $U(m)$ and $\tilde{U}(m)$, we observe that there are two main discrepancies that we need to deal with. First, it is not always true that the path π_m can be realized locally in the sense of Definition 7. It may therefore not be clear a priori why there should exist a simple relation between $U(m)$ and $\tilde{U}(m)$.

Second, we may observe that the definition of the operators $U(m)$ and $\tilde{U}(m)$ involves products of operators $C_{\sigma,p}$, where the set of vertices p that the product is extended over is generically much smaller in the case of $U(m)$. This means that most of the

¹⁰According to Proposition 14 in Appendix B.

factors $C_{\sigma,p}$ must ultimately cancel each other in the expression for $\tilde{U}(m)$. The first step will be to prove Proposition 13 in the case that m can be realized locally.

Lemma 14. *We have*

$$\tilde{U}(m) = U(m) \tag{14.3}$$

whenever $m = [\tau_m, \sigma_m] \in \mathcal{M}'_1(\Sigma)$ can be realized locally.

Proof. It follows from (12.11) and (11.5) that $C_{\sigma,p}$ can be represented as a function of the following operators

$$C_{\sigma,p} = C_{\sigma,p}(p_{v_p}, q_{v_p}; L_{c_2}, L_{c_1}),$$

where v_p is the vertex of φ_σ contained in the trinion T_p and $c_\varepsilon \equiv c_\varepsilon(p)$, $\varepsilon = 1, 2$ are the curves which represent the outgoing and incoming boundary components of T_p respectively. Let us recall that the length operators $L_{\tau_m, c_\varepsilon}$ and $L_{\sigma_m, c_\varepsilon}$ $\varepsilon = 1, 2$ satisfy (12.6). These observations imply that

$$u(\pi_m) \cdot C_{\sigma_m,p} = C_{\tau_m,p} \cdot u(\pi_m), \tag{14.4}$$

whenever the operators $p_{\sigma,p}$ and $q_{\sigma,p}$ commute with $u(\pi_m)$. Our task is therefore to determine the set of all $p \in \sigma_\circ$ for which this is the case. The condition that m can be realized locally implies that (14.4) will hold unless p is located within Σ_m . The claim now follows straightforwardly from these observations. \square

In order to treat the general case let us recall that Proposition 10 implies that π_m may be chosen as $\pi_m = y_m \circ \hat{\pi}_m \circ y_m^{-1}$, where y_m is uniquely defined by the factorization of Y_m into elementary moves that can be realized locally, and $\hat{\pi}_m$ is the (fixed) path which was chosen to represent m in the case that m can be realized locally. This leads to the following representation for $\tilde{U}(m)$:

$$\begin{aligned} \tilde{U}(m) &= D(\tau_m) \cdot u(\pi_m) \cdot D(\sigma_m)^\dagger \\ &= D(\tau_m) \cdot u(y_m) \cdot u(\hat{\pi}_m) \cdot u(y_m^{-1}) \cdot D(\sigma_m)^\dagger. \end{aligned} \tag{14.5}$$

By using Lemma 14 one may deduce from (14.5) that the following holds:

$$\tilde{U}(m) = \tilde{U}(Y_m) \cdot U(m) \cdot \tilde{U}(Y_m^{-1}) = U(Y_m) \cdot U(m) \cdot U(Y_m)^\dagger. \tag{14.6}$$

It remains to observe that

Lemma 15. *We have*

$$U(m_1) \cdot U(m_2) \cdot U(m_1)^\dagger = U(m_2)$$

whenever $\text{supp}(m_1) \cap \text{supp}(m_2) = \emptyset$.

Recall that we had defined the notation $\text{supp}(m)$ by $\text{supp}(m) = \{p\}$ if $m = Z_p, B_p, S_p$ and $\text{supp}(m) = \{p, q\}$ if $m = (pq), F_{pq}$.

Proof. Let us factorize $\mathcal{H}_L(\Sigma)$ as $\mathcal{H}_L(\Sigma) = \mathcal{K}_{\text{sp}}^{\otimes \beta_1} \otimes \mathcal{K}_{\text{sp}}^{\otimes \sigma_1 \setminus \beta_1}$, where $\beta_1 \subset \sigma_1$ is the set of all edges which end in boundary components of Σ . The representation (14.2) may be rewritten as

$$\mathcal{H}_L(\Sigma) \simeq \int_{\mathbb{R}_+^s}^{\oplus} d\eta_s(\mathbf{c}) \mathcal{H}_L(\Sigma, \mathbf{c}), \tag{14.7}$$

where $\mathcal{H}_L(\Sigma, \mathbf{c}) \simeq \mathcal{K}_{\text{sp}}^{\otimes \sigma_1 \setminus \beta_1}$ for all $\mathbf{c} \in \mathbb{R}_+^s$. Let $U(m, \mathbf{c})$, $\mathbf{c} \in \mathbb{R}_+^{s(m)}$ be the unitary operators on $\mathcal{H}_L(\Sigma_m, \mathbf{c})$ which represent the operators $U(m)$ in the representation (14.7). Within this representation it becomes almost trivial to complete the proof of Proposition 13. Let $\beta_{2,1} = \beta_2 \cup \beta_1$, where $\beta_j \subset \sigma_1$, $j = 1, 2$ are the sets of edges which correspond to boundary components of Σ_{m_j} . Let $\delta_j \subset \sigma_1$, $j = 1, 2$ be the sets of edges that are fully contained in the interior of Σ_{m_j} respectively. Let finally $\sigma'_1 = \sigma_1 \setminus (\beta_{2,1} \cup \delta_2 \cup \delta_1)$. We may then factorize $\mathcal{H}_L(\Sigma)$ in the following way:

$$\mathcal{H}_L(\Sigma) = \mathcal{K}_{\text{sp}}^{\otimes \beta_{2,1}} \otimes \mathcal{K}_{\text{sp}}^{\otimes \delta_2} \otimes \mathcal{K}_{\text{sp}}^{\otimes \delta_1} \otimes \mathcal{K}_{\text{sp}}^{\otimes \sigma'_1}, \tag{14.8}$$

where we define $\mathcal{K}_{\text{sp}}^{\otimes \delta} = \mathbb{C}$ if $\delta = \emptyset$. This may be rewritten as

$$\mathcal{H}_L(\Sigma) \simeq \int_{\mathbb{R}_+^{s_{2,1}}} d\mu(\mathbf{c}_{2,1}) \mathcal{H}_L(\Sigma, \mathbf{c}_{2,1}), \tag{14.9}$$

where $\mathcal{H}_L(\Sigma, \mathbf{c}_{2,1}) \simeq \mathcal{K}_{\text{sp}}^{\otimes \delta_2} \otimes \mathcal{K}_{\text{sp}}^{\otimes \delta_1} \otimes \mathcal{K}_{\text{sp}}^{\otimes \sigma'_1}$ for all $\mathbf{c}_{2,1} \in \mathbb{R}_+^{s_{2,1}}$, $s_{2,1} = \text{card}(\beta_{2,1})$. In the representation (14.9) we may represent $U(m_j)$, $j = 1, 2$ by families of operators $U(m_j, \mathbf{c}_{2,1})$ which take the form $U(m_2, \mathbf{c}_{2,1}) \simeq U_2(m_2, \mathbf{c}_{2,1}) \otimes \text{id} \otimes \text{id}$ and $U(m_1, \mathbf{c}_{2,1}) \simeq \text{id} \otimes U_1(m_1, \mathbf{c}_{2,1}) \otimes \text{id}$ respectively. The lemma follows easily from these observations. □

It follows from the lemma that $U(Y_m) \cdot U(m) \cdot U(Y_m)^\dagger = U(m)$, which completes the proof of Proposition 13. □

Taken together our results show that the quantization of Teichmüller spaces yields a tower of unitary projective representations of the modular groupoid in the sense of §9.3.

14.3 Concluding remarks

Let us recall that the mapping class group is a group of symmetries for both complexes $\mathcal{P}t(\Sigma)$ and $\mathcal{M}(\Sigma)$. Having projective unitary representations of the associated groupoids $\text{Pt}(\Sigma)$ and $\text{MC}(\Sigma)$ will therefore induce corresponding representations of the mapping class group by means of the construction in §8.2. It follows quite easily from Proposition 7 that these two representations are equivalent.

Finally it is clearly of interest to calculate the phases in the relations $u(\pi_{\varpi}) = \zeta^{\nu(\varpi)}$.

Conjecture 1. There exists a definition for the operators $U(m)$, $m \in \mathcal{M}_1(\Sigma)$ such that the phases $\zeta^{\hat{\nu}(\varpi)}$ which appear in the relations $u(\pi_\varpi) = \zeta^{\nu(\varpi)}$ are trivial for all but one $\varpi \in \mathcal{M}_2(\Sigma)$, which can be chosen as the relation (10.9),a). The phase which appears in the relation (10.9),a) is given as

$$\zeta^{\nu(\varpi)} = e^{\frac{\pi i}{2} c_L}, \quad c_L \equiv 1 + 6(b + b^{-1})^2.$$

We now believe to have a proof of this conjecture. Details will appear elsewhere. Our conjecture is also strongly supported by the calculation in [31] which establishes a similar result for the realization of the *pure* mapping class group on the quantized Teichmüller spaces.

Appendix A. The special functions $e_b(x)$ and $s_b(x)$

The function $s_b(x)$ may be defined with the help of the following integral representation.

$$\log s_b(x) = \frac{1}{i} \int_0^\infty \frac{dt}{t} \left(\frac{\sin 2xt}{2 \sinh bt \sinh b^{-1}t} - \frac{x}{t} \right). \tag{A.1}$$

This function, or close relatives of it like

$$e_b(x) = e^{\frac{\pi i}{2} x^2} e^{-\frac{\pi i}{24} (2-Q^2)} s_b(x), \tag{A.2}$$

have appeared in the literature under various names like “Quantum Dilogarithm” [13], “Hyperbolic G-function” [44], “Quantum Exponential Function” [60] and “Double Sine Function”, we refer to the appendix of [34] for a useful collection of properties of $s_b(x)$ and further references. The most important properties for our purposes are

(i) Functional equation: $s_b(x - i\frac{b}{2}) = 2 \cosh \pi b x s_b(x + i\frac{b}{2})$. (A.3)

(ii) Analyticity: $s_b(x)$ is meromorphic,
 poles: $x = c_b + i(nb - mb^{-1})$, $n, m \in \mathbb{Z}^{\geq 0}$, (A.4)
 zeros: $x = -c_b - i(nb - mb^{-1})$, $n, m \in \mathbb{Z}^{\geq 0}$.

(iii) Self-duality: $s_b(x) = s_{1/b}(x)$. (A.5)

(iv) Inversion relation: $s_b(x)s_b(-x) = 1$. (A.6)

(v) Unitarity: $\overline{s_b(x)} = 1/s_b(\bar{x})$. (A.7)

(vi) Residue: $\text{res}_{x=c_b} s_b(x) = e^{-\frac{\pi i}{12} (1-4c_b^2)} (2\pi i)^{-1}$. (A.8)

The function $e_b(x)$ clearly has very similar properties as $s_b(x)$. We list the properties that get modified compared to $s_b(x)$ below.

$$(i)' \text{ Functional equation: } e_b(x - i\frac{b}{2}) = (1 + e^{2\pi bx}) e_b(x + i\frac{b}{2}). \tag{A.9}$$

$$(iv)' \text{ Inversion relation: } e_b(x)e_b(-x) = e^{\pi ix^2} e^{-\frac{\pi i}{6}(1+2c_b^2)}. \tag{A.10}$$

$$(vi) \text{ Residue: } \text{res}_{x=c_b} s_b(x) = (2\pi i)^{-1}. \tag{A.11}$$

Among the most remarkable properties satisfied by the function $e_b(x)$ is the so-called pentagonal relation (A.12) which underlies the validity of the pentagonal relation (8.11),

$$e_b(p) \cdot e_b(q) = e_b(q) \cdot e_b(p + q) \cdot e_b(p), \tag{A.12}$$

Relation (A.12) is valid if q and p represent $[p, q] = (2\pi i)^{-1}$ on $L^2(\mathbb{R})$. Equation (A.12) in turn is equivalent to the following property of the function $E_b(x) \equiv e_b(-\frac{x}{2\pi b})$:

$$E_b(U) \cdot E_b(V) = E_b(U + V), \tag{A.13}$$

where $U = e^{2\pi b q}$, $V = e^{2\pi b p}$. Proofs of (A.12) and (A.13) can be found in [60], [14], [9], [54].

Appendix B. Operator-theoretical background

Let \mathcal{H} be a separable Hilbert space. The algebra of bounded operators in \mathcal{H} will be denoted by $\mathbf{B}(\mathcal{H})$. We will only need some of the most basic notions and results from functional analysis and the theory of operator algebras as summarized e.g. in [55, Chapter 14]. For the reader's convenience and to fix some notations we shall formulate the results that we need below.

Theorem 9 ([55, Theorem 14.8.14]). *Let \mathbf{C} be a commutative von Neumann subalgebra of $\mathbf{B}(\mathcal{H})$, where \mathcal{H} is a separable Hilbert space. Then there exists a compact, separable Hausdorff space X , a Radon measure μ on X , a measurable family of Hilbert spaces $\{\mathcal{H}_x\}_{x \in X}$, and a unitary bijection*

$$U: \mathcal{H} \rightarrow \mathcal{H}_{\mathbf{C}} \equiv \int_X d\mu(s) \mathcal{H}_s \tag{B.1}$$

such that

$$U \cdot \mathbf{C} \cdot U^\dagger = \{m_f; f \in L^\infty(x, d\mu)\}, \tag{B.2}$$

where m_f is the multiplication operator defined by

$$m_f: \mathcal{H}_{\mathbf{C}} \ni \{\Psi_x\}_{x \in X} \rightarrow \{f(x)\Psi_x\}_{x \in X} \in \mathcal{H}_{\mathbf{C}}. \tag{B.3}$$

If \mathbf{C} is a commutative von Neumann subalgebra of $\mathbf{B}(\mathcal{H})$ we will call an operator O on \mathcal{H} \mathbf{C} -decomposable if there exists a family of bounded operators O_x , defined on

\mathcal{H}_x for μ -almost every $x \in X$, such that

$$U \cdot O \cdot U^\dagger = \int_X d\mu(s) O_x. \tag{B.4}$$

Let \mathbf{D}_C be the algebra of all C -decomposable operators.

Proposition 14 ([55, Proposition 14.8.8]). *\mathbf{D}_C is the commutant of C within $\mathbf{B}(\mathcal{H})$.*

The unbounded operators that we will have to deal with will all be self-adjoint. We will freely use standard functional calculus for self-adjoint operators. When we say that two unbounded self-adjoint operators commute,

$$[A, B] = 0,$$

we will always mean commutativity of the spectral projections. Let $\mathcal{F} = \{A_\iota ; \iota \in \mathcal{I}\}$ be a family of commuting self-adjoint operators defined on dense domains in a separable Hilbert space \mathcal{H} . Standard functional calculus for commuting self-adjoint operators associates to \mathcal{F} a commutative von Neumann subalgebra $\mathbf{C}_{\mathcal{F}}$ of $\mathbf{B}(\mathcal{H})$. Theorem 9 applied to $\mathbf{C}_{\mathcal{F}}$ yields the existence of a common spectral decomposition for the family \mathcal{F} , where X represents the one-point compactification of the spectrum of \mathcal{F} .

Tensor product notation. For a given finite set I we will often use the notation $\mathcal{H}^{\otimes \mathbb{I}}$ instead of $\mathcal{H}^{\otimes \text{card}(\mathbb{I})}$, where $\text{card}(\mathbb{I})$ is the number of elements in \mathbb{I} . In order to avoid fixing a numbering of the elements of \mathbb{I} we shall find it useful to employ the following “leg-numbering” notation. To a given a subset $\mathbb{J} \subset \mathbb{I}$ we may associate the canonical permutation operators

$$P_{\mathbb{J}}: \mathcal{H}^{\otimes \mathbb{I}} \rightarrow \mathcal{H}^{\otimes \mathbb{J}} \otimes \mathcal{H}^{\otimes \mathbb{I} \setminus \mathbb{J}}.$$

To an operator $O \in \mathbf{B}(\mathcal{H}^{\otimes \mathbb{J}})$ we may then associate an operator $O_{\mathbb{J}} \in \mathbf{B}(\mathcal{H}^{\otimes \mathbb{I}})$ via

$$O_{\mathbb{J}} \equiv P_{\mathbb{J}}^{-1} \cdot (O \otimes \text{id}) \cdot P_{\mathbb{J}}. \tag{B.5}$$

If $\mathbb{J} = \{i_1, i_2, \dots\}$ we will sometimes write $O_{i_1 i_2 \dots}$ instead of $O_{\mathbb{J}}$. We will also abbreviate $O_{\mathbb{J}_1 \cup \mathbb{J}_2 \cup \dots}$ to $O_{\mathbb{J}_1 \mathbb{J}_2 \dots}$.

Appendix C. On the proof of Theorem 2

A similar statement is known [17] for the closely related groupoid $\tilde{\text{Pt}}(\Sigma)$ whose elements are the moves $[\tilde{\varphi}_2, \tilde{\varphi}_1]$ between fat graphs $\tilde{\varphi}_2, \tilde{\varphi}_1$ which do not have the decoration introduced in Section 3.1, but which have a numbering of the edges instead.

Theorem 2'. *The groupoid $\tilde{\text{Pt}}(\Sigma)$ is the path groupoid of the complex $\tilde{\mathcal{P}}t(\Sigma)$ which has vertices $\tilde{\varphi}$ and the following generators and relations:*

Generators:

- (i) *Flips F_e along the edges $e \in \tilde{\varphi}_1$ (see Figure 7).*
- (ii) *The exchanges (ef) of the numbers assigned to edges e and f .*

Relations:

- (a) *There is no vertex that both edges e and f are incident to: $F_f \circ F_e = F_e \circ F_f$.*
- (b) *The edges e and f are incident to the same vertex: $F_e \circ F_f \circ F_e \circ F_f \circ F_e = (ef)$.*
- (c) *$(ef) \circ F_e = F_f \circ (ef)$.*
- (d) *$F_e \circ F_e = \text{id}$.*

It was observed in [41], [17], [10] that Theorem 2' follows quite easily from the fact [39], [35] that the fat graphs $\tilde{\varphi}$ can be used to label cells in certain cell decompositions [25], [39] of $\mathcal{T}(\Sigma) \times \mathbb{R}_+^s$. This allows one to associate a path $\hat{\pi}$ in $\mathcal{T}(\Sigma) \times \mathbb{R}_+^s$ to each path $\tilde{\pi} \in [\tilde{\varphi}_2, \tilde{\varphi}_1]$ in $\tilde{\text{Pt}}(\Sigma)$. Each closed path $\hat{\pi}$ in $\mathcal{T}(\Sigma) \times \mathbb{R}_+^s$ can be deformed into small circles around the codimension two faces of the cell decompositions from [25], [39]. The latter are easily identified with the relations listed in Theorem 2'. This implies that no relations other than those listed in Theorem 2' are needed to contract a closed path $\tilde{\pi} \in [\tilde{\varphi}, \tilde{\varphi}]$ to the identity.

It remains to show that Theorem 2 follows from Theorem 2'. To this aim it is important to observe that $\tilde{\text{Pt}}(\Sigma)$ imbeds into $\text{Pt}(\Sigma)$ by means of the following construction: Given the numbered fat graph $\tilde{\varphi}$ let us define a decorated fat graph φ according to the following rule: Pick any numbering of the vertices of $\tilde{\varphi}$. For each vertex v let us distinguish among the edges which emanate from v the one with the smallest number assigned to it. For each move F_e let us choose a lift ϕ_e of the form $\phi_e = \delta'_e \circ \omega_{u_e v_e} \circ \delta_e$, where the vertices u_e and v_e represent are the ends of e and δ'_e, δ_e change the decoration only. It is then crucial to check that the relations of $\text{Pt}(\Sigma)$ ensure that the image of $\tilde{\text{Pt}}(\Sigma)$ within $\text{Pt}(\Sigma)$ is simply connected. This follows if the images of the relations listed in Theorem 2' are contractible within $\text{Pt}(\Sigma)$. We have drawn a particular example for such an image in Figure 24 below. It is easy to verify that the image of the corresponding path in $\text{Pt}(\Sigma)$ is closed. More generally one needs to consider all relations obtained by changes of the numbering of the edges. These relations are obtained by inserting $\eta^{-1} \circ \eta$ between the moves which occur in the image of the relation under consideration, where η represents the change of decoration induced by a change of numbering. In a similar way one may convince oneself that all the relations listed in Theorem 2' are mapped to closed paths in $\text{Pt}(\Sigma)$, which completes the proof that $\tilde{\text{Pt}}(\Sigma)$ embeds into $\text{Pt}(\Sigma)$.

It remains to show that each closed path in $\text{Pt}(\Sigma)$ is homotopic to a closed path in the image of $\tilde{\text{Pt}}(\Sigma)$. Let us consider a closed path $\pi \in [\varphi, \varphi] \in \text{Pt}(\Sigma)$. Write $\pi = \omega_n \circ \dots \circ \omega_1$, where each ω_k is of the form $\omega_k = \delta'_k \circ \omega_{u_k v_k} \circ \delta_k$, with δ_k, δ'_k being composed out of moves ρ_w and permutations only. Choose any numbering for the edges of φ and denote the resulting numbered fat graph by $\tilde{\varphi}$. The path π then defines a path $\tilde{\pi}$ in $\tilde{\text{Pt}}(\Sigma)$ by means of the following construction: Substitute the moves ω_r for $k = 1, \dots, n$ by the corresponding flips along the edges which connect u_k and v_k , and then multiply the result by the necessary permutations of the numbers assigned to

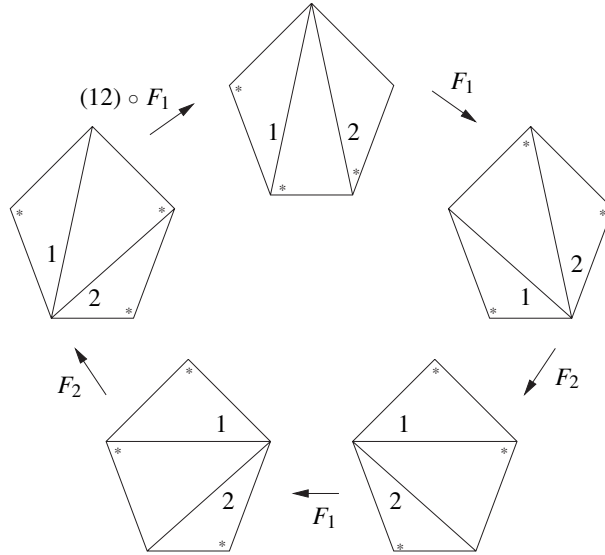


Figure 24. Image of the pentagon in $\text{Pt}(\Sigma)$.

the edges of $\tilde{\varphi}$. $\tilde{\pi}$ has the form $\tilde{\pi} = \Phi_n \circ \dots \circ \Phi_1$, where $\Phi_k \in [\tilde{\varphi}_{k+1}, \tilde{\varphi}_k]$ factorize into a single flip F_k times a change of numbering P_k , $\Phi_k = P_k \circ F_k$.

The path $\tilde{\pi}$ is mapped to a path π' in $\text{Pt}(\Sigma)$ by means of the construction above. π' has the form $\pi' = \phi'_n \circ \dots \circ \phi'_1$ with $\phi'_k \in [\varphi_{k+1}, \varphi_k]$. On the other hand let us note that the path $\pi'' = \phi''_n \circ \dots \circ \phi''_1$ with $\phi''_k \equiv \eta_{k+1}^{-1} \circ \omega_k \circ \eta_k$ is clearly homotopic to π for all changes of decorations η_k which satisfy $\eta_{n+1} = \eta_1$. For suitable choice of the η_k one gets $\phi''_k \in [\varphi_{k+1}, \varphi_k]$. It is then easy to see that ϕ'_k and ϕ''_k are homotopic. \square

Appendix D. Proof of Proposition 10

One may easily convince oneself that the moves Z_p and B_p can generically *not* be realized locally. There are simple sufficient criteria for a move m to be realizable locally.

Lemma 16. (i) *A move $m = [\tau_m, \sigma_m] \in \mathcal{M}'_1(\Sigma)$ can always be realized locally if all but one of the boundary components of Σ_m are contained in $\{c_e; e \in \mathcal{A}_\sigma\}$.*

(ii) *The moves B_p, B'_p, T_p can be realized locally if the curve $c_{p, \perp}$ which represents the boundary component of \mathfrak{t}_p assigned number \perp in Figure 12 is contained in \mathcal{A}_σ .*

(iii) *The moves $F_{pq}, A_{pq}, S_p, S_{pq}$ and B_{pq} can always be realized locally.*

Proof. Straightforward verifications. \square

Proposition 15 (= Proposition 10). *If $m \in \mathcal{M}'_1(\Sigma)$ is admissible, but can not be realized locally, there always exists a path ϖ_m which is (i) homotopic to m within $\mathcal{M}^{\text{ad}}_1(\Sigma)$, and (ii) takes the form*

$$\varpi_m = Y_m \circ m \circ Y_m^{-1}, \quad (\text{D.1})$$

where Y_m is a chain composed out of Z_p -moves and F_{pq} -moves which can all be realized locally.

Proof. The claim will follow easily from another auxiliary result that we will formulate as a separate lemma. As a preparation, let us recall that cutting a surface Σ along the curves $c(\sigma, e)$, $e \in \mathcal{A}_\sigma$ produces a surface Σ^\dagger , the connected components of which all have genus zero. Let us call a marking σ *irreducible* if the set of connected components of Σ^\dagger has only one element. It is easy to see that a marking σ is irreducible iff it has precisely one outgoing external edge.

Lemma 17. *Let σ be a marking on a surface Σ that is irreducible. For any chosen boundary component b of Σ there exists a chain of Z_p -moves and F_{pq} -moves that (i) preserves admissibility in each step, (ii) consists only of moves that can be realized locally, and (iii) transforms σ to a irreducible marking σ' whose outgoing external edge ends in the chosen component b .*

Sketch of proof. In order to check the following arguments it may be useful to think of an irreducible marking as being obtained from the corresponding marking σ^\dagger on the surface Σ^\dagger of genus zero by identifying the appropriate boundary circles. To each incoming boundary component of Σ^\dagger there corresponds a unique element of \mathcal{A}_σ .

The following claim is easy to verify. By means of F_{pq} -moves one may transform σ to a marking $\tilde{\sigma}$ which has the following two properties. First, for each vertex $p \in \tilde{\sigma}_\circ$ at least one of the edges that are incident at p is contained in \mathcal{A}_σ . Second, there is no edge $e \in \sigma_\perp$ which connects a vertex p to itself. These two properties insure that for each $p \in \tilde{\sigma}_\circ$ either Z_p or Z_p^{-1} preserves the admissibility of the marking. It is then not very hard to construct a sequence of Z_p -moves that (i) can all be realized locally and (ii) which transform $\tilde{\sigma}$ to a marking $\tilde{\sigma}'$ whose outgoing external edge ends in the chosen boundary component b .

By means of a chain of F_{pq} -moves one may finally transform $\tilde{\sigma}'$ back to a marking σ' that differs from the original marking only by the desired change of decoration. \square

End of proof of Lemma 15. Lemma 16 allows us to restrict attention to the cases $m = Z_p, B_p, B'_p, T_p$. We need to transform the original marking to another one which has the property that both incoming boundary components of the trinion \mathfrak{t}_p are contained in \mathcal{A}_σ . Given a curve $c \in \mathcal{C}_\sigma \setminus \mathcal{A}_\sigma$ there is a unique subsurface $\Sigma_c \hookrightarrow \Sigma$ with marking σ_c such that (i) c is the unique outgoing curve in the boundary of Σ_c , and (ii) the incoming boundary curves of Σ_c are contained in \mathcal{A}_σ . In order to infer the existence of the chain Y_p it clearly suffices to apply Lemma 17 to the subsurfaces

Σ_{c_1} and Σ_{c_2} respectively, where c_1 and c_2 are the incoming boundary components of the trinion with label p . \square

Lemma 15 implies that π_m may be chosen as $\pi_m = y_m \circ \hat{\pi}_m \circ y_m^{-1}$, where y_m is uniquely defined by the factorization of Y_m into elementary moves that can be realized locally, and $\hat{\pi}_m$ is the (fixed) path which was chosen to represent m in the case that m can be realized locally.

Appendix E. On the proof of Theorem 6

E.1 A technical preliminary

To begin with, let us present a technical result that facilitates the explicit computations. The operator functions we will be interested in are of the form

$$\mathbb{O} \equiv \mathbb{O}(\{z_i; i \in \varphi_1\}) = \sum_{\tau \in \mathcal{F}} C_\tau e^{x(\tau)}, \quad x(\tau) \equiv \sum_{e \in \varphi_1} \tau(e) z_e, \quad (\text{E.1})$$

where the summation is taken over the space \mathcal{F} of all maps $\varphi_1 \ni e \rightarrow \tau(e) \in \frac{1}{2}\mathbb{Z}$, and the coefficients C_τ are assumed to be non-vanishing for a finite number of $\tau \in \mathcal{F}$ only, in which case we assume $C_\tau \in \mathbb{R}_+$. These operators are densely defined and positive self-adjoint due to the self-adjointness of $x(\tau)$. The cone generated by operator functions of the form (E.1) will be denoted $\mathbb{C}_+(\varphi)$.

Proposition 16. *Let φ, φ' be two fat graphs that are related by $\rho \equiv \omega_{vw} \in \mathcal{Pt}(\Sigma)$, and let us adopt the labelling of the relevant edges given by Figure 25. For each $\mathbb{O} \in \mathbb{C}_+(\varphi)$ one has*

$$\mathbf{a}_\rho(\mathbb{O}) \equiv \mathbf{u}(\pi) \cdot \mathbb{O} \cdot \mathbf{u}(\pi)^{-1} = \mathbf{E}_b(z_{e'}) \cdot \mathbb{O}' \cdot (\mathbf{E}_b(z_{e'}))^{-1}, \quad (\text{E.2})$$

where $\mathbf{E}_b(z) \equiv e_b(-\frac{z}{2\pi b})$, and \mathbb{O}' is related to \mathbb{O} via

$$\begin{aligned} \mathbb{O}' &\equiv \mathbb{O}(z_{a'}, z_{b'} + z_{e'}, z_{c'}, z_{d'} + z_{e'}, -z_{e'}, \{z_i; i \in \varphi_1 \setminus \{a', b', c', d', e'\}\}) \text{ if} \\ \mathbb{O} &= \mathbb{O}(z_a, z_b, z_c, z_d, z_e, \{z_i; i \in \varphi_1 \setminus \{a, b, c, d, e\}\}). \end{aligned} \quad (\text{E.3})$$

Proof. Given the explicit expression for T_{vw} we only have to verify that

$$e^{-2\pi i p_v q_w} \cdot \mathbb{O} \cdot e^{-2\pi i p_v q_w} = \mathbb{O}',$$

with \mathbb{O} and \mathbb{O}' being related by (E.3). Keeping in mind (4.6) we notice that the operators z_f may be represented in the form $z_f = z_{f,v} + z_{f,w}$, where $v, w \in \varphi_\circ$ are the vertices connected by the edge $f \in \varphi_\circ$. Let us label the vertices in φ_\circ and φ'_\circ such that the edges a and d are incident to the vertices v_a and v_d besides to the vertex v respectively, and similarly for the vertices v_b, v_c and w . This leads to the following

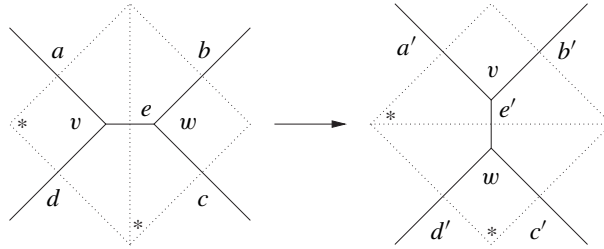


Figure 25. Labelling of the edges which are relevant for the description of a flip.

representations for the relevant Fock variables:

$$\begin{aligned}
 z_e &= 2\pi b(q_v - p_v + p_w), & z_{e'} &= 2\pi b(q'_w - q'_v - p'_w), \\
 z_a &= 2\pi b p_v + z_{a,v_a}, & z_{a'} &= 2\pi b p'_v + z_{a',v_a}, \\
 z_b &= 2\pi b(q_w - p_w) + z_{b,v_b}, & z_{b'} &= 2\pi b(q'_v - p'_v) + z_{b',v_b}, \\
 z_c &= 2\pi b(-q_w) + z_{c,v_c}, & z_{c'} &= 2\pi b(-q'_w) + z_{c',v_c}, \\
 z_d &= 2\pi b(-q_v) + z_{d,v_d}, & z_{d'} &= 2\pi b p'_w + z_{d',v_d}.
 \end{aligned}
 \tag{E.4}$$

To complete the proof of Proposition 16 is now the matter of a straightforward calculation. □

Remark 10. This yields in particular the formulae

$$\begin{aligned}
 e^{-z_{a'}} &= e^{-\frac{1}{2}z_a} (1 + e^{-z_e}) e^{-\frac{1}{2}z_a}, & e^{-z_{b'}} &= e^{+\frac{1}{2}z_b} (1 + e^{+z_e}) e^{+\frac{1}{2}z_b}, \\
 e^{+z_{d'}} &= e^{+\frac{1}{2}z_d} (1 + e^{+z_e}) e^{+\frac{1}{2}z_a}, & z_{e'} &= -z_e, & e^{+z_{c'}} &= e^{-\frac{1}{2}z_c} (1 + e^{-z_e}) e^{-\frac{1}{2}z_c}.
 \end{aligned}$$

It is quite obvious that these transformations reduce to their classical counterparts (3.3) in the limit $b \rightarrow 0$.

As an example let us consider the monomials $M_{n_a n_c}^{n_b n_d} \equiv e^{\frac{1}{2}(n_a z_a + n_b z_b + n_c z_c + n_d z_d)}$, where $n_b, b \in \{a, b, c, d\}$ are restricted by the requirement that $2N \equiv n_a + n_c - n_b - n_d$ must be even. One then finds certain simplifications on the right hand side of (E.2):

$$\begin{aligned}
 E_b(z'_e) \cdot M_{n_a n_c}^{n_b n_d} \cdot (E_b(z'_e))^{-1} &= \sqrt{M_{n_a n_c}^{n_b n_d}} \frac{E_b(z'_e + \pi i b^2 N)}{E_b(z'_e - \pi i b^2 N)} \sqrt{M_{n_a n_c}^{n_b n_d}} \\
 &= \sqrt{M_{n_a n_c}^{n_b n_d}} \left[\prod_{m=-\frac{N-1}{2}}^{\frac{N-1}{2}} (1 + q^{2m} e^{-z'_e}) \right] \sqrt{M_{n_a n_c}^{n_b n_d}}.
 \end{aligned}
 \tag{E.5}$$

E.2 Invariance of length operators

In the main text we have explained how to reduce the proof of Theorem 6 to the following lemma:

Lemma 18. *Assume that $m \in \mathcal{M}'_1(\Sigma)$ can be realized locally. We then have the relation*

$$u(\pi_m) \cdot L_{\sigma_m, c} = L_{\tau_m, c} \cdot u(\pi_m) \tag{E.6}$$

for all curves c such that $c \in \mathcal{C}_{\sigma_m}$ and $c \in \mathcal{C}_{\tau_m}$.

On the proof of Lemma 18. It is easy to see that our recursive Definition 6 of the length operators $L_{\sigma, c}$ reduces the task of proving Lemma 18 to the curves that represent a boundary component of Σ_m . Let us write $L_{\sigma, e} \equiv L_{\sigma, c(\sigma, e)}$.

In the cases $m = F_{pq}, S_p, T_p, Z_p$ we will need to verify Lemma 18 by direct calculations. Let us begin with the case $m = F_{pq}$. The claim is again trivial for the length operators associated to the incoming boundary components of Σ_m . Let us therefore focus on the length operators $L_{\sigma_m, f}, L_{\tau_m, f}$ assigned to the outgoing boundary component of Σ_m . Definition 6 yields the following expressions for the length operators $L_{\sigma_m, f}$ and $L_{\tau_m, f}$ respectively.

$$\begin{aligned} L_{\sigma_m, f} &= e^{z_{21}+y_3+y_2+y_1} + e^{z_{21}-y_3+y_2+y_1} + e^{-z_{21}-y_3-y_2-y_1} \\ &\quad + e^{y_1-y_2-y_3} + e^{y_1+y_2-y_3} + e^{-y_1-y_2-y_3} \\ &\quad + e^{z_{21}+y_2+y_1}L_3 + e^{y_1-y_3}L_2 + e^{-y_2-y_3}L_1, \\ L_{\tau_m, f} &= e^{+z'_{32}+y'_3+y'_2+y'_1} + e^{-z'_{32}-y'_3-y'_2-y'_1} + e^{-z'_{32}-y'_3-y'_2+y'_1} \\ &\quad + e^{y'_1+y'_2+y'_3} + e^{y'_1-y'_2-y'_3} + e^{y'_1+y'_2-y'_3} \\ &\quad + e^{y'_2+y'_1}L'_3 + e^{y'_1-y'_3}L'_2 + e^{-z'_{32}-y'_2-y'_3}L'_1. \end{aligned} \tag{E.7}$$

We have used the abbreviations $y_l \equiv y_{\varphi_\sigma, e_l}, y'_l \equiv y_{\varphi_\tau, e_l}, L_l \equiv L_{\sigma, e_l}$ and $L'_l \equiv L_{\tau, e_l}$ for $l \in \{1, 2, 3\}$, as well as $z_l \equiv z_{\varphi_\sigma, e_l}, z'_l \equiv z_{\varphi_\tau, e_l}$ for $l \in \{32, 21\}$. The labelling of the edges is the one introduced in Figure 26. According to Proposition 16 we need to

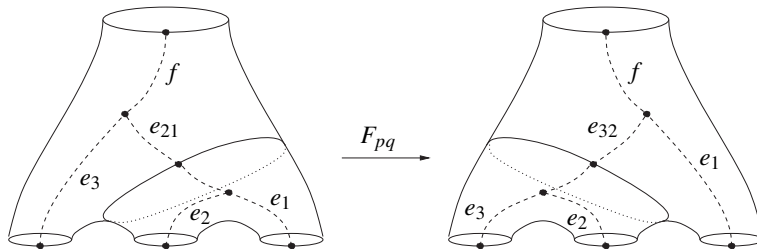


Figure 26. Labelling of the edges of σ_m for $m = F_{pq}$.

calculate

$$\begin{aligned}
 E_b(\mathbf{z}'_{3,2}) \cdot & (e^{y'_3+y'_2+y'_1+z'_{3,2}} + e^{y'_1-y'_2-y'_3} \\
 & + e^{y'_2+y'_1}L'_3 + e^{y'_1-y'_3}L'_2 + e^{-z'_{3,2}-y'_2-y'_3}L'_3 \\
 & + e^{\frac{1}{2}(y'_1+y'_2-y'_3)}(1 + e^{-z'_{3,2}})e^{\frac{1}{2}(y'_1+y'_2-y'_3)} \\
 & + e^{-\frac{1}{2}(y'_1+y'_2+y'_3+z'_{3,2})}(1 + e^{-z'_{3,2}})e^{-\frac{1}{2}(y'_1+y'_2+y'_3+z'_{3,2})}) \cdot E_b(\mathbf{z}'_{3,2})^\dagger.
 \end{aligned}
 \tag{E.8}$$

We finally need to apply equation (E.5). The terms in the first line of (E.8) have $N = 1$, those in the second line $N = 0$, and all other terms have $N = -1$. Straightforward application of equation (E.5) shows that the expression given in (E.8) equals $L_{\tau_m, f}$, as claimed.

The next case we will consider is $m = S_p$. It clearly suffices to restrict attention to the case that $\Sigma_m = \Sigma_{1,1}$, the one-holed torus. On Σ_m let us consider the fat graphs $\varphi_i, i = 1, 2, 3$ depicted in Figure 27.

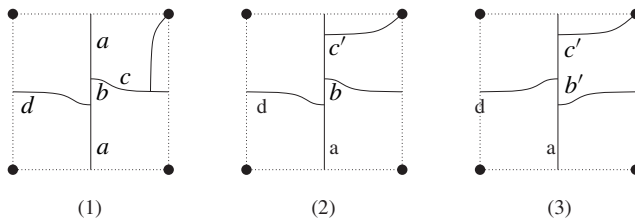


Figure 27. Fat graphs φ_i on the one-holed torus.

We have $\varphi_1 = \tilde{\varphi}'_m$ and $\varphi_3 = \theta'_m$ respectively. The sequence of fat graph $(\varphi_3, \varphi_2, \varphi_1)$ defines the path $\pi_m \in [\theta'_m, \tilde{\varphi}'_m]$ which we will use. The only relevant curve is the curve β which represents the boundary of Σ_m . Definition 6 yields the following expressions for the length operators $L_{\varphi_1, \beta}$ and $L_{\varphi_3, \beta}$ respectively:

$$\begin{aligned}
 L_{\varphi_1, \beta} &= e^{z_d-z_c} + 2 \cosh(z_c + z_d + z_a + z_b) \\
 &+ e^{z_d}(1 + e^{z_b} + e^{z_b+z_a}) + e^{-z_c}(1 + e^{-z_a} + e^{-z_a-z_b}), \\
 L_{\varphi_3, \beta} &= e^{z_{c'}-z_a} + 2 \cosh(z_a + z_d + z_{c'} + z_{b'}) \\
 &+ e^{z_{c'}}(1 + e^{z_{b'}} + e^{z_{b'}+z_d}) + e^{-z_a}(1 + e^{-z_d} + e^{-z_d-z_{b'}}).
 \end{aligned}
 \tag{E.9}$$

The labelling of the relevant edges is the one introduced in Figure 27. With the help of Proposition 16 one may calculate

$$\begin{aligned}
 L_{\varphi_2, \beta} \equiv \mathbf{a}_{[\varphi_2, \varphi_1]}(L_{\varphi_1, \beta}) &= e^{z_{c'}-z_a} + 2 \cosh(z_a + z_d + z_{c'} + z_b) \\
 &+ e^{z_{c'}}(1 + e^{z_d} + e^{z_d+z_b}) + e^{-z_a}(1 + e^{-z_b} + e^{-z_b-z_b}).
 \end{aligned}
 \tag{E.10}$$

One may then apply Proposition 16 once more to calculate $\mathbf{a}_{[\varphi_3, \varphi_2]}(L_{\varphi_2, \beta})$. The result is $L_{\varphi_3, \beta}$ as claimed.

We have furthermore verified by direct calculations that Lemma 18 holds in the cases $m = T_p, Z_p, B_p, B'_p$ respectively. These calculations proceed along very similar lines as in the previous two cases, which motivates us to omit the details.

In the remaining cases $m = A_{pq}, m = B_{pq}$ and $m = S_{pq}$ one may use the factorization of the chains π_m which follows from the definitions (10.5), (10.7) and (10.8) respectively in order to reduce the proof to the cases where the result is already established. This completes our discussion of the proof of Lemma 18. \square

References

- [1] S. Axelrod, S. Della Pietra, and E. Witten, Geometric quantization of Chern-Simons gauge theory. *J. Differential Geom.* 33 (1991), 787–902. [688](#)
- [2] A. Yu. Alekseev, H. Grosse, and V. Schomerus, Combinatorial quantization of the Hamiltonian Chern-Simons theory. I; II. *Comm. Math. Phys.* 172 (1995), 317–358; *ibid.* 174 (1996), 561–604. [688](#)
- [3] A. Yu. Alekseev and V. Schomerus, Representation theory of Chern-Simons observables. *Duke Math. J.* 85 (1996), 447–510. [688](#)
- [4] J. E. Andersen, J. Mattes, and N. Reshetikhin, Quantization of the algebra of chord diagrams. *Math. Proc. Cambridge Philos. Soc.* 124 (1998), 451–467. [688](#)
- [5] B. Bakalov and A. Kirillov, Jr., On the Lego-Teichmüller game. *Transform. Groups* 5 (2000), 207–244. [690](#), [691](#), [726](#), [730](#)
- [6] B. Bakalov and A. Kirillov, Jr., *Lectures on tensor categories and modular functors*. Univ. Lecture Ser. 21, Amer. Math. Soc., Providence, RI, 2001. [688](#), [720](#), [721](#), [723](#), [725](#)
- [7] E. W. Barnes, The theory of the double gamma function. *Philos. Trans. Roy. Soc. London Ser. A* 196 (1901), 265–387.
- [8] E. Buffenoir and P. Roche, Link invariants and combinatorial quantization of Hamiltonian Chern-Simons theory. *Comm. Math. Phys.* 181 (1996), 331–365. [688](#)
- [9] A. G. Bytsko and J. Teschner, R-operator, co-product and Haar-measure for the modular double of $U_q(\mathfrak{sl}(2, \mathbb{R}))$. *Comm. Math. Phys.* 240 (2003), 171–196. [691](#), [748](#)
- [10] L. O. Chekhov and R. C. Penner, On quantizing Teichmüller and Thurston theories. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 579–645. [687](#), [750](#)
- [11] M. Dehn, Die Gruppe der Abbildungsklassen. *Acta Math.* 69 (1938), 135–206.
- [12] L. D. Faddeev, Modular double of a quantum group. In *Conférence Moshé Flato 1999*, Vol. I (Dijon), Math. Phys. Stud. 21, Kluwer Acad. Publ., Dordrecht 2000, 149–156. [691](#)
- [13] L. D. Faddeev and R. M. Kashaev, Quantum dilogarithm. *Modern Phys. Lett. A* 9 (1994), 427–434. [747](#)
- [14] L. D. Faddeev, R. M. Kashaev, and A. Yu. Volkov, Strongly coupled quantum discrete Liouville theory. I. Algebraic approach and duality. *Comm. Math. Phys.* 219 (2001), 199–219. [748](#)

- [15] V. V. Fock, Dual Teichmüller spaces. Preprint, 1998; arXiv:hep-th/9702018. [687](#), [689](#), [690](#), [695](#), [697](#), [706](#)
- [16] V. V. Fock and L. O. Chekhov, Quantum mapping class group, pentagon relation, and geodesics. *Tr. Mat. Inst. Steklova* 226 (1999), 163–179; English transl. *Proc. Steklov Inst. Math.* 226 (1999), 149–163. [690](#), [737](#)
- [17] V. V. Fock and L. O. Chekhov, Quantum Teichmüller spaces. *Teoret. Mat. Fiz.* 120 (1999), 511–528; English transl. *Theoret. and Math. Phys.* 120 (1999), 1245–1259. [687](#), [689](#), [749](#), [750](#)
- [18] V. V. Fock and A. B. Goncharov, Cluster ensembles, quantization and the dilogarithm. Preprint, 2007; arXiv:math.AG/0311245. [687](#)
- [19] V. V. Fock and A. A. Rosly, Poisson structure on moduli of flat connections on Riemann surfaces and the r -matrix. In *Moscow Seminar in Mathematical Physics*, Amer. Math. Soc. Transl. (2) 191, Amer. Math. Soc., Providence, RI, 1999, 67–86. [688](#)
- [20] L. Funar and R. Gelca, On the groupoid of transformations of rigid structures on surfaces. *J. Math. Sci. Univ. Tokyo* 6 (1999), 599–646. [726](#)
- [21] S. Gervais, Presentation and central extensions of mapping class groups. *Trans. Amer. Math. Soc.* 348 (1996), 3097–3132. [740](#), [741](#)
- [22] W. M. Goldman, The symplectic nature of fundamental groups of surfaces. *Adv. Math.* 54 (1984), 200–225. [689](#)
- [23] W. M. Goldman, Topological components of spaces of representations. *Invent. Math.* 93 (1988), 557–607. [689](#)
- [24] V. Guillemin and S. Sternberg, *Symplectic techniques in physics*. 2nd ed., Cambridge University Press, Cambridge 1990. [714](#)
- [25] J. L. Harer, The virtual cohomological dimension of the mapping class group of an orientable surface. *Invent. Math.* 84 (1986), 157–176. [750](#)
- [26] N. J. Hitchin, The self-duality equations on a Riemann surface. *Proc. London Math. Soc.* (3) 55 (1987), 59–126. [689](#)
- [27] N. J. Hitchin, Flat connections and geometric quantization. *Comm. Math. Phys.* 131 (1990), 347–380. [688](#)
- [28] Y. Imayoshi and M. a. Taniguchi, *An introduction to Teichmüller spaces*. Springer-Verlag, Tokyo 1992. [705](#)
- [29] R. M. Kashaev, The hyperbolic volume of knots from the quantum dilogarithm. *Lett. Math. Phys.* 39 (1997), 269–275. [689](#)
- [30] R. M. Kashaev, Quantization of Teichmüller spaces and the quantum dilogarithm. *Lett. Math. Phys.* 43 (1998), 105–115. [687](#), [689](#), [690](#), [691](#), [694](#), [700](#), [701](#), [702](#), [709](#)
- [31] R. M. Kashaev, Liouville central charge in quantum Teichmüller theory. *Tr. Mat. Inst. Steklova* 226 (1999), 72–81; English transl. *Proc. Steklov Inst. Math.* 226 (1999), 63–71. [747](#)
- [32] R. M. Kashaev, On the spectrum of Dehn twists in quantum Teichmüller theory. In *Physics and combinatorics* (Nagoya, 2000), World Scientific, Singapore 2001, 63–81. [689](#), [690](#), [691](#), [698](#), [699](#), [717](#), [740](#)
- [33] R. Kashaev, The quantum dilogarithm and Dehn twists in quantum Teichmüller theory. In *Integrable structures of exactly solvable two-dimensional models of quantum field theory* (Kiev, 2000), NATO Sci. Ser. II Math. Phys. Chem. 35, Kluwer Acad. Publ., Dordrecht 2001, 211–221. [690](#), [739](#), [741](#)

- [34] S. Kharchev, D. Lebedev, and M. Semenov-Tian-Shansky, Unitary representations of $U_q(\mathfrak{sl}(2, \mathbb{R}))$, the modular double and the multiparticle q -deformed Toda chains. *Comm. Math. Phys.* 225 (2002), 573–609. [747](#)
- [35] M. Kontsevich, Intersection theory on the moduli space of curves and the matrix Airy function. *Comm. Math. Phys.* 147 (1992), 1–23. [750](#)
- [36] Y. Laszlo, Hitchin's and WZW connections are the same. *J. Differential Geom.* 49 (1998), 547–576. [688](#)
- [37] F. Luo, A presentation of the mapping class groups. *Math. Res. Lett.* 4 (1997), 735–739.
- [38] G. Moore and N. Seiberg, Classical and quantum conformal field theory. *Comm. Math. Phys.* 123 (1989), 177–254. [690](#), [691](#), [726](#), [733](#)
- [39] R. C. Penner, The decorated Teichmüller space of punctured surfaces. *Comm. Math. Phys.* 113 (1987), 299–339. [690](#), [695](#), [698](#), [750](#)
- [40] R. C. Penner, Weil-Petersson volumes. *J. Differential Geom.* 35 (1992), 559–608. [695](#), [699](#)
- [41] R. C. Penner, Universal constructions in Teichmüller theory. *Adv. Math.* 98 (1993), 143–215. [700](#), [750](#)
- [42] B. Ponsot and J. Teschner, Clebsch-Gordan and Racah-Wigner coefficients for a continuous series of representations of $\mathcal{U}_q(\mathfrak{sl}(2, \mathbb{R}))$. *Comm. Math. Phys.* 224 (2001), 613–655. [691](#)
- [43] N. Reshetikhin and V. G. Turaev, Invariants of 3-manifolds via link polynomials and quantum groups. *Invent. Math.* 103 (1991), 547–597. [688](#)
- [44] S. N. M. Ruijsenaars, First order analytic difference equations and integrable quantum systems. *J. Math. Phys.* 38 (1997), 1069–1146. [747](#)
- [45] I. E. Segal, Transforms for operators and symplectic automorphisms over a locally compact abelian group. *Math. Scand.* 13 (1963), 31–43. [714](#)
- [46] T. Shintani, On a Kronecker limit formula for real quadratic fields. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* 24 (1977), 167–199.
- [47] C. Sorger, La formule de Verlinde. *Astérisque* 237 (1996), Sémin. Bourbaki, Vol. 1994/95, Exp. No. 794, 87–114. [688](#)
- [48] Takhtajan, L. A., Teo, L.-P., Quantum Liouville theory in the background field formalism. I. Compact Riemann surfaces. *Comm. Math. Phys.* 268 (2006), 135–197. [688](#)
- [49] J. Teschner, Liouville theory revisited. *Classical Quantum Gravity* 18 (2001), R153–R222. [687](#)
- [50] J. Teschner, A lecture on the Liouville vertex operators. *Internat. J. Modern Phys. A* 19 (2004), Suppl. May, 436–458. [692](#)
- [51] J. Teschner, An analog of a modular functor from quantized Teichmüller theory, II. In preparation. [691](#), [692](#)
- [52] V. G. Turaev, *Quantum invariants of knots and 3-manifolds*. De Gruyter Stud. Math. 18, Walter de Gruyter, Berlin 1994. [720](#), [721](#), [723](#)
- [53] H. Verlinde, Conformal field theory, two-dimensional quantum gravity and quantization of Teichmüller space. *Nuclear Phys. B* 337 (1990), 652–680. [687](#)
- [54] A. Yu. Volkov, Noncommutative hypergeometry. *Comm. Math. Phys.* 258 (2005), 257–273. [748](#)

- [55] N. R. Wallach, *Real reductive groups*. II. Pure Appl. Math. 132, Academic Press, Boston 1992. [748](#), [749](#)
- [56] E. Witten, Quantum field theory and the Jones polynomial. *Comm. Math. Phys.* 121 (1989), 351–399. [688](#)
- [57] S. A. Wolpert, On the symplectic geometry of deformations of a hyperbolic surface. *Ann. of Math.* (2) 117 (1983), 207–234. [690](#)
- [58] S. A. Wolpert, On the Weil-Petersson geometry of the moduli space of curves. *Amer. J. Math.* 107 (1985), 969–997. [690](#), [705](#)
- [59] S. A. Wolpert, Geodesic length functions and the Nielsen problem. *J. Differential Geom.* 25 (1987), 275–296. [690](#)
- [60] S. L. Woronowicz, Quantum exponential function. *Rev. Math. Phys.* 12 (2000), 873–920. [747](#), [748](#)

On quantum moduli space of flat $\mathrm{PSL}_2(\mathbb{R})$ -connections on a punctured surface

Rinat M. Kashaev

Contents

1	Introduction	761
2	The decorated moduli space of flat $\mathrm{PSL}_2(\mathbb{R})$ -connections	763
	2.1 Subgroups of $\mathrm{PSL}_2(\mathbb{R})$	763
	2.2 The moduli space of flat connections	763
	2.3 A first definition of the decorated moduli space	764
	2.4 A second definition of the decorated moduli space	765
	2.5 Equivalence of the two definitions	767
	2.6 Coordinates for the decorated moduli space	767
	2.7 Reconstruction of a flat connection from the coordinates	770
	2.8 The transition functions	771
	2.9 Goldman's symplectic structure	772
3	Towards the quantum theory	772
	3.1 Quantization of an ideal triangle	773
	3.2 Quantization of the flip transformation	775
	3.3 A solution by Woronowicz and Zakrzewski	777
	3.4 Spectrum of the length operator	780
	References	781

1 Introduction

Quantum theory of the Teichmüller space of a punctured surface developed recently in the works [1], [2], [4], [8] is an example of a topological quantum field theory with infinite dimensional quantum state space. This theory, for example, is connected with the quantum theory of the Liouville equation [16]. From the purely mathematical viewpoint it leads to an interesting infinite dimensional projective unitary representation of the mapping class group of the underlying surface.

Teichmüller space is a connected component of a bigger space, namely the moduli space of $\mathrm{PSL}_2(\mathbb{R})$ flat connections on the surface, or equivalently the space of conjugacy classes of representations of the fundamental group of the surface in the group

$\mathrm{PSL}_2(\mathbb{R})$. This space is symplectic with Goldman's symplectic form which restricts to the Weil–Petersson symplectic form on the Teichmüller component. It is thus natural to ask if one can extend the quantum Teichmüller theory to a quantum theory of the moduli space of flat connections. Physically this is the question of quantizing the Chern–Simons gauge theory with a non-compact gauge group which in our case is the group $\mathrm{PSL}_2(\mathbb{R})$.

Unlike Teichmüller space, the other connected components of the moduli space are not topologically trivial [6]. This is why quantization of the whole moduli space is not expected to be a straightforward generalization of the quantum Teichmüller space.

Since the quantization procedure is not yet a functor (probably with the exception of the deformation quantization), the canonical quantization in physics is sensitive to a choice of coordinates. In the case of the Teichmüller space a convenient system of coordinates is given by real analytic coordinates, for example, the Penner coordinates coming from the decorated Teichmüller space, in which both the Weil–Petersson symplectic structure and the action of the mapping class group are described by explicit rational expressions. This is why the first step in the quantization program of the moduli space of flat connections could be an appropriate choice of coordinates.

This chapter is essentially a review of some of the results related to the quantization problem of the moduli space of irreducible flat $\mathrm{PSL}_2(\mathbb{R})$ -connections on a punctured surface with parabolicity conditions around the punctures.

Section 2 is devoted to the decorated moduli space of flat connections which generalizes Penner's decorated Teichmüller space. In addition to the constructions of the work [11], here we also give another geometrical definition for the decorated moduli space which adapts to our context Penner's original definition of the decorated Teichmüller space. It is worth noticing that the definition of [11], being more algebraic in nature, remains valid for the group $\mathrm{PSL}_2(\mathbb{F})$, where \mathbb{F} is an arbitrary field.

In Section 3, we first show the relevance of the works of Woronowicz and Zakrzewski [17], [18] to the quantization problem of the moduli space. Then, we consider a simple example of an annulus triangulated into two ideal tetrahedra, and derive a formula for the trace of the holonomy associated to the only non-contractible simple loop in the annulus. Quantization of this trace function leads to a certain difference operator in the Hilbert space of square integrable functions on the real line. The spectrum of this operator appears to be very much sensitive to the sign variable which distinguishes between different components of the moduli space.

Acknowledgement. I would like to thank V. Fock for pointing out a few inaccurate statements in the initial version of this exposition and for making a number of valuable comments. This work is supported in part by the Swiss National Science Foundation.

2 The decorated moduli space of flat $\mathrm{PSL}_2(\mathbb{R})$ -connections

2.1 Subgroups of $\mathrm{PSL}_2(\mathbb{R})$

Fix a one-parameter subgroup $U \subset \mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm 1\}$ of parabolic elements corresponding to upper triangular unipotent matrices, i.e. matrices with trace equal to ± 2 . To avoid confusion with the term “parabolic subgroup” in the algebraic group context, in this chapter subgroups conjugate to U will be called *unipotent subgroups*.

The normalizer of U , $B = N(U)$, is a Borel subgroup corresponding to upper triangular matrices. In the exact sequence of group homomorphisms

$$1 \rightarrow U \rightarrow B \rightarrow T \rightarrow 1$$

the group $T = B/U$ is identified with the Cartan subgroup represented by diagonal matrices. The Bruhat decomposition of $\mathrm{PSL}_2(\mathbb{R})$ with respect to B consists of only two cells

$$\mathrm{PSL}_2(\mathbb{R}) = B\theta B \sqcup B$$

where $\theta \in N(T) \subset G$ is a fixed representative of the only nontrivial element of the Weyl group $N(T)/T$. We choose it in a unique way by the condition $\theta^2 = 1$ so that it is represented by the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

In the standard action of $\mathrm{PSL}_2(\mathbb{R})$ by linear fractional transformations in the upper half of the complex plane, the Cartan subgroup is characterized by the condition that it fixes 0 and ∞ , while the Borel subgroup is the stabilizer of ∞ . The element θ is given by inversion with respect to the unit circle followed by complex conjugation and sign change. Elements in the two Bruhat cells are distinguished depending on whether ∞ is moved or not.

2.2 The moduli space of flat connections

Consider a punctured surface of finite type $\Sigma = \Sigma_{g,s}$ which is the complement of a finite set of points $V = \{P_1, \dots, P_s\} \subset \bar{\Sigma}$, called *punctures*, in a closed oriented surface $\bar{\Sigma}$ of genus g :

$$\Sigma = \bar{\Sigma} - V.$$

We assume the following restrictions on g and s :

$$s > 0, \quad \kappa := -\chi(\Sigma) = 2g - 2 + s > 0.$$

Denote by G the group $\mathrm{PSL}_2(\mathbb{R})$. Let \mathcal{M} be the space of gauge classes of irreducible flat G -connections on Σ with parabolic holonomies around the punctures. Connections representing points of \mathcal{M} will be called *flat connections* and the space \mathcal{M} itself will be called the *moduli space* of flat connections. There is a one-to-one correspondence between elements of \mathcal{M} and conjugacy classes of irreducible representations in G of the fundamental group of the surface with the parabolicity conditions around the

punctures. The Teichmüller space \mathcal{T} is a connected component of \mathcal{M} given by faithful representations corresponding to complete hyperbolic structures on Σ .

The decorated Teichmüller space of Penner [14] is a (trivial) principal $\mathbb{R}_{>0}^s$ -bundle $\pi_{\mathcal{T}}: \tilde{\mathcal{T}} \rightarrow \mathcal{T}$, where each fiber is given by collections of s horocycles, one about each puncture. To generalize this construction to the case of the moduli space \mathcal{M} we first define a “connection horocycle” at a puncture associated to a flat connection and which corresponds to a usual geometric horocycle at the level of gauge equivalence classes.

2.3 A first definition of the decorated moduli space

With any path a in Σ , starting at some $x \in \Sigma$ and ending at some puncture P , we associate the homotopy class $\ell(a)$ of loops based at x which go along a towards P , then go around P along a small circle in the counterclockwise direction with respect to the orientation of the surface, and then return back to x along a^{-1} .

Denote by \mathbb{H} the two-dimensional hyperbolic plane for which the group G is the group of orientation preserving isometries. Consider a flat connection f on Σ and a pair (a, h) where a is a path with initial point $x \in \Sigma$ and terminal point a puncture P , and h is a horocycle in \mathbb{H} centered at the only fixed point of the f -holonomy along the loop $\ell(a)$. We say that two such pairs (a_1, h_1) and (a_2, h_2) are *equivalent* if a_1 and a_2 have the same terminal point and $h_1 = f_{12}h_2$, where f_{12} is the f -parallel transport along any path in Σ homotopic in $\Sigma \cup \{P\}$ to the path $a_1 \cdot a_2^{-1}$. This is an equivalence relation which does not depend on the choice of the path defining the element f_{12} , since different choices give parallel transports differing only by elements of the stabilizer subgroups of the horocycles. Namely, let f_i be the f -holonomies along the loops $\ell(a_i)$ for $i = 1, 2$. One has $f_i h_i = h_i$, $i = 1, 2$. If f_{12} is the f -parallel transport along a fixed path homotopic to $a_1 \cdot a_2^{-1}$ in $\Sigma \cup \{P\}$, then any other choice for this path gives an f -parallel transport of the form $f'_{12} = f_1^m f_{12} = f_{12} f_2^m$ for some integer m . Thus, one has $f'_{12} h_2 = f_{12} h_2$.

A *connection horocycle* at the puncture P associated to the flat connection f (also called simply an f -horocycle) is an equivalence class of such pairs.

Gauge transformations of connections naturally act on connection horocycles. Let $g: \Sigma \rightarrow G$ correspond to a gauge transformation $f \mapsto f'$ given by $f'(p) = g(x)f(p)g(y)^{-1}$ for any path p in Σ connecting points x and y with $f(p)$, $f'(p)$ being the f - and f' -parallel transports along p respectively. Then, this gauge transformation acts on an f -horocycle represented by a pair (a, h) to give an f' -horocycle represented by the pair $(a, g(x)h)$ where $x \in \Sigma$ is the initial point of a .

We define the *decorated moduli space* $\tilde{\mathcal{M}}$ as the collection of gauge classes of multiples (f, H_1, \dots, H_s) , where f is a flat connection and H_i is an f -horocycle at the puncture P_i for all $1 \leq i \leq s$.

Proposition 1. *The natural projection $\pi: \tilde{\mathcal{M}} \rightarrow \mathcal{M}$ given by $\pi([(f, H_1, \dots, H_s)]) = [f]$ is a principal $\mathbb{R}_{>0}^s$ -fibration.*

Proof. A free and transitive action of the group $\mathbb{R}_{>0}$ on f -horocycles at a given puncture can be seen as follows.

Let a be a path in Σ with initial point $x \in \Sigma$ and terminal point a puncture P . Let $U_a \subset G$ be the stabilizer unipotent subgroup of the f -holonomy along the loop $\ell(a)$ with respect to the adjoint action. Any horocycle h representing an f -horocycle at P is invariant with respect to the action of U_a which, in fact, coincides with the stabilizer group of h . The normalizer subgroup $N(U_a) \subset G$ of U_a acts transitively on the space of all horocycles representing f -horocycles at P . In this way we obtain a free and transitive action of the quotient group $N(U_a)/U_a$ on the set of all f -horocycles at P . On the other hand, the group $N(U_a)/U_a$ is identified with the Cartan subgroup conjugate to the subgroup of diagonal matrices. The latter is isomorphic to $\mathbb{R}_{>0}$. \square

2.4 A second definition of the decorated moduli space

Here, following the paper [11], we give a definition of the decorated moduli space in terms of flat graph connections.

Given an immersed graph $\Gamma \subset \Sigma$, a flat *graph G -connection* on Γ is an isomorphism class of flat G -connections on Σ with fixed parallel transports along the edges of Γ . A *graph gauge transformation* of a flat graph connection is a usual gauge transformation modulo gauge transformations relating different representatives of flat graph connections.

Equivalently, one can define a flat graph G -connection as a representation in G of the edge-path groupoid of Γ , graph gauge transformations being equivalence transformations of representations.

Suppose that an immersed graph $\Gamma \subset \Sigma$ is the one-dimensional skeleton of a cell complex homotopically equivalent to Σ . Then, there is a one-to-one correspondence between equivalence classes of flat graph G -connections on Γ (with respect to graph gauge transformations) and moduli of flat connections on Σ . A typical example of a flat graph G -connection is given by a representation of the fundamental group $\pi_1(\Sigma, x)$ in G , where the immersed graph Γ is given by a single vertex x and the homotopy classes of simple loops in Σ based at x as edges. Graph gauge transformations in this case correspond to overall conjugations by elements of G .

We shall confine ourselves to special immersed graphs on Σ associated with arbitrary collections of ideal arcs. Recall that an *ideal arc* on Σ is a nontrivial isotopy class of a simple path running between punctures. Let \mathcal{A}_Σ be the set of ideal arcs. Define a pairing

$$I: \mathcal{A}_\Sigma \times \mathcal{A}_\Sigma \rightarrow \mathbb{Z}_{\geq 0}, \quad I(e, f) = \min(|a \cap b| \mid a \in e, b \in f)$$

where $|a \cap b|$ denotes the number of intersection points of a and b .

Let $\hat{\Sigma} \subset \Sigma$ be the complement of a disjoint union of small open disks centered at the punctures. The subsurface $\hat{\Sigma}$ is a compact surface homotopically equivalent to Σ with the boundary given by a disjoint union of s circles $\{L(P)\}_{P \in V}$ associated with

the punctures

$$\partial \hat{\Sigma} = \coprod_{P \in V} L(P).$$

To each collection of ideal arcs $\alpha \subset \mathcal{A}_\Sigma$ let us choose a collection of representatives $\tilde{\alpha} = \{a_e \mid e \in \alpha\}$ so that for any ideal arcs e and f in α the arcs $a_e \cap \hat{\Sigma}$ and $a_f \cap \hat{\Sigma}$ are simple paths with $|a_e \cap a_f| = I(e, f)$. Such a choice is always possible. It can be given, for example, by a collection of geodesics with respect to a complete hyperbolic structure on Σ . The set $\Gamma(\alpha) = (\hat{\Sigma} \cap \tilde{\alpha}) \cup \partial \hat{\Sigma}$ is an immersed graph in Σ with vertices $\partial \hat{\Sigma} \cap \tilde{\alpha}$ and two types of edges: “long” edges corresponding to ideal arcs, and “short” edges corresponding to segments of the boundary components of $\hat{\Sigma}$ between vertices.

If $\alpha \subset \beta$, we identify $\Gamma(\alpha)$ with the corresponding subgraph of $\Gamma(\beta)$. Thus, for any $\alpha \subset \mathcal{A}_\Sigma$, $\Gamma(\alpha)$ is a subgraph of $\Gamma(\mathcal{A}_\Sigma)$.

Let a and b be two paths having one and the same initial point in Σ and with the terminal points at some punctures. Let $m \in \mathcal{M}$ be represented by a flat connection f . The homotopy class of the path $a^{-1} \cdot b$ is called m -admissible if the f -holonomies along the loops $\ell(a)$ and $\ell(b)$ (recall that they are parabolic) do not commute.

It is clear that this notion depends only on the homotopy class $[a^{-1} \cdot b]$ and the moduli $m = [f]$, but not on the choices of their representatives. It is straightforward to see that the homotopy class $[a^{-1} \cdot b]$ is m -admissible if and only if the f -holonomies along the loops $\ell(a)$ and $\ell(b)$ have distinct fixed points. As a particular case of this definition, we have a notion of an m -admissible ideal arc.

For $\alpha \subset \mathcal{A}_\Sigma$ define an open subset $\mathcal{M}(\alpha) \subset \mathcal{M}$ by the condition that for any $m \in \mathcal{M}(\alpha)$ the set α is the maximal family of m -admissible ideal arcs. To avoid trivial cases with $\mathcal{M}(\alpha) = \emptyset$, we also define

$$\mathcal{W} := \{\alpha \subset \mathcal{A}_\Sigma \mid \mathcal{M}(\alpha) \neq \emptyset\}.$$

It is clear that

$$\mathcal{M} = \coprod_{\alpha \in \mathcal{W}} \mathcal{M}(\alpha).$$

We construct a fiber bundle $\pi : \tilde{\mathcal{M}} \rightarrow \mathcal{M}$ as the disjoint union of fiber bundles $\pi : \tilde{\mathcal{M}}(\alpha) \rightarrow \mathcal{M}(\alpha)$ for all $\alpha \in \mathcal{W}$. The fiber $\pi^{-1}(m)$ over a point $m \in \mathcal{M}(\alpha)$ consists of flat graph connections on $\Gamma(\alpha)$ which represent m and whose parallel transport operators belong to a fixed unipotent subgroup U for all short edges, and to the subset θT , for all long edges, where T is the group $N(U)/U$ identified with a subgroup of $N(U)$ through a section of the canonical projection $N(U) \rightarrow N(U)/U$. Let us show that $\tilde{\mathcal{M}}(\alpha)$ is a principal T^s -bundle over $\mathcal{M}(\alpha)$.

Our strategy is to show that any flat connection f , representing a given point $m \in \mathcal{M}(\alpha)$, is equivalent to any other flat graph connection in $\pi^{-1}(m)$.

As every vertex of $\Gamma(\alpha)$ belongs to a boundary component of $\hat{\Sigma}$, the f -holonomies around boundary components, based at vertices of $\Gamma(\alpha)$, are parabolic. Thus, one can always replace f by such an equivalent connection that all associated holonomies around the boundary components, based at vertices, belong to one and the same

unipotent subgroup $U \subset G$. This automatically makes the parallel transport operators along short edges belong to $B = N(U)$. The parallel transport operators along long edges belong to the big Bruhat cell $B\theta B$ due to the condition of admissibility. Thus, making an appropriate graph gauge transformation with values in B , we can make the parallel transport operators along all long edges belong to the subset θT . The remaining freedom in the graph gauge transformations consists of arbitrary T valued functions on vertices of $\Gamma(\alpha)$, which can be used to eliminate the T -parts of the parallel transport operators along all short edges, thus obtaining elements of $\pi^{-1}(m)$. This still leaves unfixed the graph gauge transformations, given by T valued functions taking one and the same value at all vertices associated with one boundary component. This residual gauge group is isomorphic to T^s , which acts freely and transitively on the space $\pi^{-1}(m)$. Thus, the union $\tilde{\mathcal{M}}(\alpha) = \coprod_{m \in \mathcal{M}(\alpha)} \pi^{-1}(m)$ is a principal T^s -bundle over $\mathcal{M}(\alpha)$.

2.5 Equivalence of the two definitions

To establish equivalence of the two definitions of the decorated moduli space it is enough to show that for a given point $m \in \mathcal{M}(\alpha)$ of the moduli space there exists a one-to-one correspondence between the collection of gauge classes of connection horocycles around the punctures on one hand and flat graph connections on $\Gamma(\alpha)$, on the other.

Let f be a flat graph connection on $\Gamma(\alpha)$ representing a point $\tilde{m} \in \tilde{\mathcal{M}}(\alpha)$ of the decorated moduli space in the sense of the second definition. Then we define an f -horocycle H_i at the puncture P_i to be represented by a pair (a, h_0) , where a is a path in the complement $\Sigma - \hat{\Sigma}$, connecting a vertex of the graph $\Gamma(\alpha)$ on the boundary component $L(P_i)$ to the puncture P_i , and h_0 is the horocycle in the upper half complex plane given by the equation $\Im z = 1$. This definition is consistent in the sense that it defines a unique connection horocycle for each puncture. Thus, the gauge class of the collection (f, H_1, \dots, H_s) gives a point in the decorated moduli space in the sense of the first definition.

Conversely, for each point \tilde{m} of the decorated moduli space given by the first definition with $\pi(\tilde{m}) \in \mathcal{M}(\alpha)$ there exists a unique representative (f, H_1, \dots, H_s) such that the f -holonomies along the edges of $\Gamma(\alpha)$ correspond to a graph connection in $\tilde{\mathcal{M}}(\alpha)$ in the sense of the second definition and all f -horocycles H_i are represented by the horocycle $\Im z = 1$ in the upper half complex plane.

2.6 Coordinates for the decorated moduli space

Recall that an *ideal triangulation* of Σ is a maximal family of pairwise non-intersecting ideal arcs. The set of ideal triangulations is denoted by Δ_Σ . It is a countable infinite set.

With each ideal triangulation τ we can associate an open dense subset \mathcal{M}_τ of \mathcal{M} defined as those moduli of flat connections for which τ is an admissible ideal triangulation, i.e. with all ideal arcs admissible. This set is given as the following union

$$\mathcal{M}_\tau = \coprod_{\alpha \in \mathcal{W}(\tau)} \mathcal{M}(\alpha), \quad \mathcal{W}(\tau) = \{\alpha \in \mathcal{W} \mid \tau \subset \alpha\}.$$

Following the paper [11] we describe some of the properties of the subsets \mathcal{M}_τ .

Theorem 1 ([11]). *Let $\Sigma, \mathcal{M}, \{\mathcal{M}_\tau\}_{\tau \in \Delta_\Sigma}$ be as above. Then*

- (i) *the collection $\{\mathcal{M}_\tau\}_{\tau \in \Delta_\Sigma}$ is a covering for \mathcal{M} ;*
- (ii) *there exists a finite sub-covering $\{\mathcal{M}_\tau\}_{\tau \in \Pi}, \Pi \subset \Delta_\Sigma, |\Pi| < \infty$.*

Part (i) of this theorem is equivalent to existence of an m -admissible ideal triangulation for any $m \in \mathcal{M}$. Part (ii) implies that there exists a positive integer valued topological invariant given by the minimal size of finite sub-coverings. One can estimate from above this number by

$$\frac{2^{3\kappa}}{2\kappa + 1} \binom{4\kappa}{2\kappa}.$$

This is a very rough estimation. For example, in the case of a sphere with four punctures $\Sigma = \Sigma_{0,4}$ this formula gives $2^7 7 = 896$ while one can show that there exists a covering of size 7.

Theorem 2 ([11]). *Let $\Sigma, \pi: \tilde{\mathcal{M}} \rightarrow \mathcal{M}, \{\mathcal{M}_\tau\}_{\tau \in \Delta_\Sigma}$ be as above. Then,*

- (i) *for each $\tau \in \Delta_\Sigma$ one has $\pi^{-1}(\mathcal{M}_\tau) = \coprod_{\varepsilon \in \{\pm 1\}^{2\kappa}} \tilde{\mathcal{M}}_\varepsilon(\tau)$ with each $\tilde{\mathcal{M}}_\varepsilon(\tau)$ being a principal $\mathbb{R}_{>0}^s$ -bundle homeomorphic to the complement of $\psi_{\tau,\varepsilon}^{-1}(0)$ for certain rational mapping $\psi_{\tau,\varepsilon}: \mathbb{R}_{>0}^{3\kappa} \rightarrow \mathbb{R}$ (see the end of Section 2.7 for details);*
- (ii) *if $\sigma: \{\pm 1\}^{2\kappa} \rightarrow \mathbb{Z}$ is defined by $\sigma(\varepsilon) = \frac{1}{2} \sum_{i=1}^{2\kappa} \varepsilon_i$, then for $-\kappa \leq k \leq \kappa$, the sets $\tilde{\mathcal{M}}_k = \bigcup_{\tau \in \Delta_\Sigma} \coprod_{\varepsilon \in \sigma^{-1}(k)} \tilde{\mathcal{M}}_\varepsilon(\tau)$ are principal $\mathbb{R}_{>0}^s$ -bundles which are disjoint for different k ;*
- (iii) *there exist principal bundle isomorphisms between $\tilde{\mathcal{M}}_{\pm k}$ and the decorated Teichmüller space $\tilde{\mathcal{T}}$ of Penner.*

Theorems 1 (i) and 2 (ii) imply that

$$\mathcal{M} = \coprod_{-\kappa \leq k \leq \kappa} \mathcal{M}_k, \quad \mathcal{M}_k = \tilde{\mathcal{M}}_k / \mathbb{R}_{>0}^s.$$

Part (iii) of Theorem 2 implies that $\mathcal{M}_{\pm k}$, being homeomorphic to Teichmüller space, are open cells of dimension $3\kappa - s$. The total number of components is $2\kappa + 1$ which formally at $s = 0$ coincides with the number of connected components of the moduli space of flat G connections on a closed surface [6]. In our case, however, the

components \mathcal{M}_k can be empty or not connected. For example, in the case $\kappa = 1$, $\mathcal{M}_0 = \emptyset$. The non-connectedness of \mathcal{M}_k originates from the fact that there are two distinct conjugacy classes \mathcal{C}_\pm of parabolic elements in $\mathrm{PSL}_2(\mathbb{R})$ distinguished by the sign of the off diagonal matrix element when represented by an upper triangular matrix with trace equal to 2. For example, consider a sphere with four punctures. One can show that there are two points $m, m' \in \mathcal{M}_0$ such that the holonomies around two fixed punctures are in one and the same conjugacy class for m and in distinct conjugacy classes for m' . Thus, it is clear that m and m' cannot be in one and the same connected component of the moduli space \mathcal{M} . However, taking into account the formal agreement of the invariant k with Goldman's relative Euler class, it is natural to expect that for fixed conjugacy types of the holonomies around the punctures nonempty \mathcal{M}_k 's are connected.

Part (i) of Theorem 2 extends to \mathcal{M} Penner's coordinate charts for Teichmüller space. Coordinatization is given by assigning positive real numbers to edges and signs to faces of an ideal triangulation. The geometrical meaning of the positive real numbers at edges is identical to the meaning of the Penner coordinates in the decorated Teichmüller space but adapted to our definitions of the decorated moduli space.

Given $\tilde{m} \in \tilde{\mathcal{M}}$, $m = \pi(\tilde{m})$, let a, b be two paths starting at $x \in \Sigma$ and ending at two punctures P and Q (possibly coinciding) such that the path $a^{-1} \cdot b$ represents an m -admissible ideal arc e . Choose a flat connection f representing m and two horocycles p, q in hyperbolic plane so that the pairs (a, p) and (b, q) represent the two f -horocycles at P and Q respectively. The m -admissibility of e means that the horocycles p and q are centered at distinct points u and v of the circle at infinity of hyperbolic plane. Thus, following Penner, we calculate the signed hyperbolic distance δ between p and q along the geodesic from u to v taken with positive sign if $p \cap q = \emptyset$ and with negative sign if $p \cap q \neq \emptyset$, and we assign to the ideal arc e the real positive number $\exp(\delta/2)$. It is clear that this number is independent of the choices we made: the flat connection f which represents $m \in \mathcal{M}$ or the pairs (a, p) and (b, q) representing the two f -horocycles.

To interpret the sign variables associated with ideal triangles, let three paths a, b, c starting from some $x \in \Sigma$ and ending at three punctures P, Q, R be such that the paths $a^{-1} \cdot b, b^{-1} \cdot c, c^{-1} \cdot a$ represent three m -admissible ideal arcs which thus specify an oriented ideal triangle t in Σ with the orientation induced from that of Σ . On the other hand, the fixed points of the f -holonomies along the loops $\ell(a), \ell(b), \ell(c)$ are distinct and specify an oriented ideal triangle t' in hyperbolic plane with the orientation induced from that of the hyperbolic plane. Taking into account the natural correspondence between the vertices of t and t' we compare their orientations and assign to t a positive sign if the two orientations agree and a negative sign if they disagree. In particular, the part (iii) of Theorem 2 implies that the Teichmüller component of the moduli space corresponds to putting all signs to one and the same value, and that corresponds to the fact that the orientations of all ideal triangles in Σ and the orientations of the corresponding ideal triangles in \mathbb{H} simultaneously agree or disagree.

2.7 Reconstruction of a flat connection from the coordinates

Now, let us show how one can reconstruct a flat connection together with an s -tuple of connection horocycles starting from a collection of positive real numbers on edges and signs on faces of an ideal triangulation of Σ .

A point $m \in \mathcal{M}$ is represented by infinitely many different flat connections on Σ , and first we reconstruct a flat graph connection on a two dimensional cell complex associated to an ideal triangulation.

As any ideal triangulation τ of Σ is a subset of ideal arcs, we can associate to τ the immersed graph $\Gamma(\tau) \subset \Sigma$ as in Section 2.4. Due to the definition of an ideal triangulation, any pair of edges of $\Gamma(\tau)$ intersect at most at their endpoints, and thus the graph $\Gamma(\tau)$ as a subset of Σ is a one-dimensional cell complex. The connected components of its complement in $\hat{\Sigma}$ are given by hexagons which are in one-to-one correspondence with ideal triangles of τ . Thus, we obtain a cell complex decomposition of $\hat{\Sigma}$ associated with τ .

Each short edge of $\Gamma(\tau)$ belongs to the boundary of a unique hexagon, and the boundary of each hexagon consists of three short and three long edges in alternating order. In addition, each short edge has a distinguished orientation induced from that of the surface Σ . In other words, the short edges can be identified with (oriented) corners and long edges with sides of ideal triangles of τ .

Given an ideal triangulation τ of Σ , a set of positive real numbers associated with edges of τ , and a set of signs associated with triangles of τ , we construct a flat graph G -connection on $\Gamma(\tau)$ as follows.

To each long edge with corresponding positive real number a we associate a parallel transport given by the matrix

$$\begin{bmatrix} 0 & a^{-1} \\ -a & 0 \end{bmatrix}.$$

Notice, that the square of this matrix is the negative unit matrix which corresponds to the unit element in the group G and thus the orientation of the edge is irrelevant.

To each short edge with distinguished orientation we associate a parallel transport given by the matrix

$$\begin{bmatrix} 1 & -\varepsilon a/(bc) \\ 0 & 1 \end{bmatrix}$$

where ε is the sign associated to the unique ideal triangle which has this short edge as its corner, a being the positive real number associated with this triangle's side opposite to the corner under consideration and b, c being the numbers associated with the other two sides of the triangle.

Let f be a flat connection on Σ representing the flat graph connection we have just constructed. Each vertex x of $\Gamma(\tau)$ belongs to the closure of a unique small disk in the complement $\Sigma - \hat{\Sigma}$ and thus there is a path a_x in this disk having this vertex as its initial point and the puncture P in its center as its terminal point. This path is in fact the intersection of the disk with a path representing the ideal arc associated to

x . We define an f -horocycle at P as the equivalence class of the pair (a_x, h_0) where $h_0 \subset \mathbb{H}$ is the horocycle represented in the upper half complex plane by the horizontal line $\Im z = 1$.

Finally, it remains to note that the coordinates should satisfy non-triviality conditions for holonomies around the punctures. Each such holonomy, being a product of parallel transports along short edges around a puncture, is automatically parabolic provided the upper off-diagonal element is non-zero. Thus, all the s non-triviality conditions can be given by a single inequality by taking the product of upper off-diagonal elements of all s holonomies. This is how the rational mapping $\psi_{\tau, \varepsilon}$ in Theorem 2(i) is constructed.

2.8 The transition functions

Recall that a *flip transformation* is an elementary transformation from one ideal triangulation of Σ to another obtained by taking an ideal quadrilateral composed of two distinct ideal triangles, and replacing its diagonal by the opposite diagonal. It is well known fact that any two ideal triangulations of Σ are related by a finite sequence of flip transformations [7], [13]. Thus, to describe the transition functions between coordinate charts associated with two ideal triangulations it is enough to describe the transition function in the case of a single flip.

Let $\tilde{m} \in \tilde{\mathcal{M}}$, $m = \pi(\tilde{m})$, and τ, τ' be two m -admissible ideal triangulations related by a single flip transformation. Then, the geometrical interpretation of the coordinates given in Section 2.6 implies that the positive real numbers and the signs associated with ideal arcs common to τ and τ' and ideal triangles are identical. Thus, we only need to describe the formulas relating the coordinates within the quadrilateral associated with the flip transformation. A way to obtain these formulas is to compare the parallel transports in this quadrilateral with respect to the two coordinate systems. The result is as follows.

Let a, b, c, d be the sides of the quadrilateral and e, f its diagonals, with e in τ and f in τ' and the triangles $t_1 = (a, b, e)$ and $t_2 = (c, d, e)$ in τ and $t_3 = (d, a, f)$ and $t_4 = (b, c, f)$ in τ' , see Figure 1. Then, denoting the positive real numbers associated

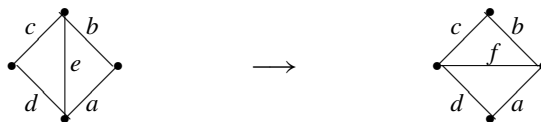


Figure 1. The flip transformation.

with the edges by the same symbols a, b, \dots and the signs of triangles t_1, t_2, \dots by $\varepsilon_1, \varepsilon_2, \dots$ we have the following equations

$$\varepsilon_3 e f = \varepsilon_1 b d + \varepsilon_2 a c, \quad \varepsilon_4 e f = \varepsilon_2 b d + \varepsilon_1 a c \tag{2.1}$$

which imply that

$$\varepsilon_1 \varepsilon_2 = \varepsilon_3 \varepsilon_4.$$

In the case of coinciding signs these equations reduce to the Ptolemy formula of [14] generating the transition functions for the decorated Teichmüller space.

2.9 Goldman's symplectic structure

In this parameterization one has an explicit description of the Goldman's symplectic structure [5]. The pull-back of the symplectic form to $\tilde{\mathcal{M}}$ is given by the same formula as that of Penner in [15] for the pull-back of the Weil–Petersson form. Namely, if ω denotes the Goldman's symplectic form on \mathcal{M} , and τ is a given ideal triangulation then, up to an overall multiplicative factor, one has

$$\pi^* \omega = \sum \frac{da \wedge db}{ab} + \frac{db \wedge dc}{bc} + \frac{dc \wedge da}{ca}$$

with the summation over all ideal triangles of τ with side coordinate functions a, b, c with the cyclic order induced from the orientation of Σ . The drawback of this form is that it is not a symplectic form, and this prevents us to immediately proceed to quantization. There are at least two possibilities to overcome this obstacle.

One possibility is to consider only gauge invariant combinations of the coordinate functions which correspond to coordinates in the moduli space \mathcal{M} . For example, for each quadrilateral with sides a, b, c, d with the cyclic order induced from the orientation of Σ , the ratio $ac/(bd)$ is gauge invariant. In this way we come to the shearing coordinates of Bonahon–Thurston and Chekhov–Fock.

Another possibility is to consider the projectivization of $\tilde{\mathcal{M}}$, $P\tilde{\mathcal{M}} = \tilde{\mathcal{M}}/\mathbb{R}_{>0}$, where the group $\mathbb{R}_{>0}$ is the diagonal subgroup of the structure group $\mathbb{R}_{>0}^s$, and then to realize it as a sub-manifold in a larger space with a symplectic structure in such a way that the moduli space \mathcal{M} is obtained by a Hamiltonian reduction. In the case of Teichmüller space this approach has been materialized in the work [8], where the larger space is given by the product space $P\tilde{\mathcal{T}} \times H^1(\Sigma; \mathbb{R})$ of even dimension

$$\dim P\tilde{\mathcal{T}} + \dim H^1(\Sigma; \mathbb{R}) = (3\kappa - 1) + (\kappa + 1) = 4\kappa$$

and which has a mapping class group invariant symplectic structure. A work on an adaptation of this approach to the case of the decorated moduli space is in progress. In the next section we consider only one aspect of this approach, namely, the flip transformation and its quantization.

3 Towards the quantum theory

The principal ingredient of the quantum Teichmüller theory is the quantum operator which realizes the flip transformation, and which among other things satisfies the

pentagonal equation. In the case of the moduli space the flip transformation acts also on sign variables, so a part of the problem of quantization is the problem of finding the associated solution of the pentagonal equation. Here we show that the solution of Woronowicz and Zakrzewski associated with the quantum ‘ $ax + b$ ’ group [18] gives rise to a viable quantum theory of the moduli space.

3.1 Quantization of an ideal triangle

The starting point is the following observation. The contribution to the pull-back of the Goldman’s symplectic form coming from an ideal triangle with side variables a , b , c can be written in the form

$$\frac{da \wedge db}{ab} + \frac{db \wedge dc}{bc} + \frac{dc \wedge da}{ca} = \frac{du \wedge dv}{uv} \quad (3.1)$$

where

$$u = a/c, \quad v = b/c. \quad (3.2)$$

The definition of the variables u , v is not unique as it does not respect the cyclic symmetry of the side variables. One can fix this ambiguity by choosing a distinguished corner in the triangle, see Figure 2. In this way, taking also into account the sign

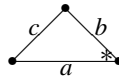


Figure 2. An ideal triangle with a distinguished corner.

variables, we associate to each ideal triangle with a distinguished corner a triple of variables (u, v, ε) taking values in $\mathbb{R}_{>0}^2 \times \{\pm 1\}$.

In quantum theory, motivated by equation (3.1), we associate to each ideal triangle with a distinguished corner a unitary \mathbb{C} -algebra \mathcal{A} , called the *quantum triangle algebra*, which contains a triple of elements (u, v, e) where a pair of invertible elements u, v (quantizations of the variables u, v) satisfies the “quantum plane” commutation relation

$$uv = q^2vu$$

with $q \in \mathbb{C} - 0$ being a quantization parameter, while element e (a quantization of the sign variable ε) satisfies the relations

$$e^2 = 1, \quad ue = eu, \quad ve = ev.$$

The minimal algebra satisfying these conditions is the algebra generated by these three elements, but for the reasons which will become clear later on, we shall choose for actual quantum triangle algebra a bigger algebra. In quantum Teichmüller theory the

sign operator is trivial and one takes for \mathcal{A} the algebra $\text{End}(L^2(\mathbb{R}))$ of linear operators in the Hilbert space $L^2(\mathbb{R})$ of square integrable functions $f(x)$ on the real axis with respect to the standard Lebesgue measure with the realizations

$$u = e^{2\pi b q}, \quad v = e^{2\pi b p}.$$

Here $b \in \mathbb{C}$, p and q are operators satisfying the Heisenberg commutation relation

$$pq - qp = (2\pi i)^{-1}, \tag{3.3}$$

so the initial quantization parameter takes the form $q = e^{i\pi b^2}$. Using the fact that the classical variables u, v are real and positive it is assumed that $b \in \mathbb{R}_{>0}$ and p, q are self-adjoint operators realized as differentiation and multiplication operators:

$$(qf)(x) = xf(x), \quad (pf)(x) = \frac{1}{2\pi i} \frac{df(x)}{dx}. \tag{3.4}$$

To take into account the sign degrees of freedom, we take for the quantum triangle algebra \mathcal{A} the algebra $\text{End}(L^2(\mathbb{R}) \otimes \mathbb{C}^2)$ with the realizations

$$u = e^{2\pi b q} \otimes \text{id}_{\mathbb{C}^2}, \quad v = e^{2\pi b p} \otimes \text{id}_{\mathbb{C}^2}, \quad e = \text{id}_{L^2(\mathbb{R})} \otimes \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{3.5}$$

This algebra has the property that if an element $x \in \mathcal{A}$ commutes both with $p \otimes \text{id}_{\mathbb{C}^2}$ and $q \otimes \text{id}_{\mathbb{C}^2}$ then there exists unique $x \in \text{End}(\mathbb{C}^2)$ such that $x = \text{id}_{L^2(\mathbb{R})} \otimes x$.

In order to simplify the notation, in what follows we shall identify elements of the algebras $\text{End}(L^2(\mathbb{R}))$ and $\text{End}(\mathbb{C}^2)$ with their images in \mathcal{A} by the canonical inclusions.

To complete the quantization of a triangle we have to restore the broken cyclic symmetry of the side variables (a, b, c) . This means that we have to construct an inner automorphism of the quantum triangle algebra corresponding to a change of the distinguished corner. Namely, consider the transformation drawn in Figure 3. Classically, the new variables $(\hat{u}, \hat{v}, \hat{\varepsilon})$ corresponding to the right hand side of Figure 3

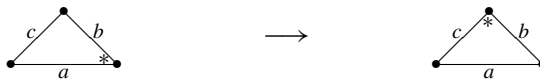


Figure 3. A change of the distinguished corner of an ideal triangle.

are given by the formula

$$(\hat{u}, \hat{v}, \hat{\varepsilon}) = (b/a, c/a, \varepsilon) = (v/u, 1/u, \varepsilon)$$

where we have used the definitions (3.2). The quantum version of this transformation has the form

$$(\hat{u}, \hat{v}, \hat{\varepsilon}) = (qu^{-1}v, u^{-1}, \varepsilon), \tag{3.6}$$

which corresponds to

$$(\hat{q}, \hat{p}) = (p - q, -q).$$

Now, there exists a unique up to multiplicative factor unitary operator $A \in \mathrm{End}(L^2(\mathbb{R}))$ which realizes the latter transformation by conjugation:

$$A(q, p)A^{-1} = (p - q, -q).$$

Explicitly we have

$$A = e^{-i\pi/3} e^{i3\pi q^2} e^{i\pi(p+q)^2} \tag{3.7}$$

where the normalization constant is chosen so that

$$A^3 = 1. \tag{3.8}$$

Thus, the operator A considered as an element of \mathcal{A} realizes by conjugation the transformation (3.6).

3.2 Quantization of the flip transformation

Recall that if a triangulated quadrilateral has side variables a, b, c, d and a diagonal variable e so that it is decomposed into two triangles $t_1 = (a, b, e)$ and $t_2 = (c, d, e)$, then the flip transformation replaces the diagonal variable e by another diagonal variable f and the triangles t_1, t_2 by triangles $t_3 = (d, a, f)$ and $t_4 = (b, c, f)$, see Figure 1, with the side and the sign variables transforming according to equations (2.1). The choice of the distinguished corners for the four triangles as in Figure 4 corresponds to the following identifications:

$$\begin{aligned} (u_1, v_1) &= (a/e, b/e), & (u_2, v_2) &= (e/d, c/d), \\ (u_3, v_3) &= (a/d, f/d), & (u_4, v_4) &= (b/f, c/f), \end{aligned}$$

so that the triple $(u_i, v_i, \varepsilon_i)$ is associated with the triangle t_i for $1 \leq i \leq 4$. The flip

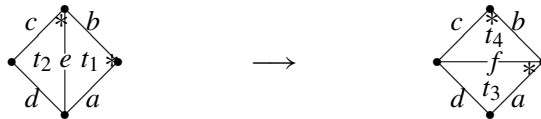


Figure 4. A choice of distinguished corners for the triangles t_1, \dots, t_4 .

transformation implies the following relations:

$$u_3 = u_1 u_2, \quad v_2 = v_3 v_4, \quad v_3 u_4 = v_1 u_2, \quad \varepsilon_3 v_3 = \varepsilon_1 v_1 + \varepsilon_2 v_2 u_1, \quad \varepsilon_1 \varepsilon_2 = \varepsilon_3 \varepsilon_4 \tag{3.9}$$

which are consistent with the equation

$$\sum_{i=1,2} \frac{du_i \wedge dv_i}{u_i v_i} = \sum_{i=3,4} \frac{du_i \wedge dv_i}{u_i v_i}.$$

In quantum theory, to each ideal triangle of a given ideal triangulation we associate a copy of the quantum triangle algebra \mathcal{A} in such a way that copies corresponding to

different ideal triangles mutually commute. Thus, in the case of the quadrilateral under consideration, the quantum algebras associated to the two triangulations are given by the tensor product algebra $\mathcal{A} \otimes \mathcal{A}$ with tensor components corresponding to quantum triangle algebras associated with the triangles t_1 and t_2 for one triangulation, and with the triangles t_3 and t_4 for another triangulation. Here, in the case of L^2 -spaces the tensor product is defined by

$$\text{End}(L^2(X)) \otimes \text{End}(L^2(Y)) := \text{End}(L^2(X \times Y)).$$

The *quantum flip transformation* by definition is an isomorphism between these two quadrilateral algebras. We assume that the same formulae (3.9) are valid on the quantum level. The *quantum flip transformation operator* $T \in \mathcal{A} \otimes \mathcal{A}$ is defined as a unitary operator realizing the quantum flip transformation by conjugation,

$$T(u_i, v_i, e_i)T^{-1} = (u_{i+2}, v_{i+2}, e_{i+2}), \quad i \in \{1, 2\},$$

where

$$(u_1, v_1, e_1) = (u \otimes 1, v \otimes 1, e \otimes 1), \quad (u_2, v_2, e_2) = (1 \otimes u, 1 \otimes v, 1 \otimes e).$$

Equivalently, we have the following system of linear equations:

$$Tq \otimes 1 = (q \otimes 1 + 1 \otimes q)T \tag{3.10}$$

$$T(p \otimes 1 + 1 \otimes p) = 1 \otimes pT \tag{3.11}$$

$$T(p \otimes 1 + 1 \otimes q) = (p \otimes 1 + 1 \otimes q)T \tag{3.12}$$

$$Te^{2\pi bp}e \otimes 1 = (e^{2\pi bp}e \otimes 1 + e^{2\pi bq} \otimes e^{2\pi bp}e)T \tag{3.13}$$

$$Te \otimes e = e \otimes eT. \tag{3.14}$$

Proposition 2. *Any solution of the system of equations (3.10)–(3.14) has the form*

$$T = e^{2\pi ip \otimes q} F \tag{3.15}$$

where

$$F = \Psi_b(e \otimes e, r \otimes r, q \otimes 1 + 1 \otimes (p - q)) + e \otimes 1 \Psi'_b(e \otimes e, r \otimes r, q \otimes 1 + 1 \otimes (p - q)) \tag{3.16}$$

with $\Psi_b, \Psi'_b: \{\pm 1\}^2 \times \mathbb{C}_{0 \leq \Im \leq b} \rightarrow \mathbb{C}$ being functions which are analytic in $\{\pm 1\}^2 \times \mathbb{C}_{0 < \Im < b}$ and for any $x \in \mathbb{R}$ satisfying one and the same functional equation

$$\Psi_b(\varepsilon, -\rho, x + ib) = (1 + \varepsilon q e^{2\pi bx}) \Psi_b(\varepsilon, \rho, x) \tag{3.17}$$

and $r \in \mathcal{A}$ is any element satisfying the relations

$$r^2 = 1, \quad rq - qr = rp - pr = re + er = 0. \tag{3.18}$$

Proof. If we define $F = e^{-2\pi i p \otimes q} T$, then the system (3.10)–(3.14) simplifies to

$$\begin{aligned} Fq \otimes 1 &= q \otimes 1F \\ F(p \otimes 1 + 1 \otimes p) &= (p \otimes 1 + 1 \otimes p)F \\ F(p \otimes 1 + 1 \otimes q) &= (p \otimes 1 + 1 \otimes q)F \\ Fe^{2\pi bp} e \otimes 1 &= e^{2\pi bp} e \otimes 1(1 + qe^{2\pi b(q \otimes 1 + 1 \otimes (p-q))} e \otimes e)F \\ Fe \otimes e &= e \otimes eF \end{aligned}$$

which implies that F must be of the form given by formula (3.16) with Ψ_b, Ψ'_b satisfying equation (3.17). The analyticity properties of Ψ_b, Ψ'_b follow from the fact that a proper definition of conjugation by a positive operator a in a Hilbert space (in our case such positive operator is of the form $e^{2\pi bp} \otimes 1$) is the analytic continuation to $t = -i$ of conjugation by the associated family of unitary operators $\{a^{|t|} \mid t \in \mathbb{R}\}$. \square

3.3 A solution by Woronowicz and Zakrzewski

The functional equation (3.17) can be solved by using Faddeev's non-compact quantum dilogarithm function $\varphi_b(z)$ [3] defined by

$$\varphi_b(z) := \exp\left(\int_{i0+\mathbb{R}} \frac{e^{-i2zw} dw}{4 \sinh(wb) \sinh(wb^{-1})w}\right) \quad (3.19)$$

in the strip $|\Im z| < (b + b^{-1})/2$ and analytically continued to the whole complex plane through the following pair of functional equations:

$$\varphi_b(z - ib^{\pm 1}/2) = \varphi_b(z + ib^{\pm 1}/2)(1 + e^{2\pi b^{\pm 1}z}). \quad (3.20)$$

It is a meromorphic function with poles given by the set

$$\mathcal{P} := \{ib(m + 1/2) + ib^{-1}(n + 1/2) \mid m, n \in \mathbb{Z}_{\geq 0}\} \quad (3.21)$$

and zeros given by the set $-\mathcal{P}$. It satisfies the “inversion relation”

$$\varphi_b(z)\varphi_b(-z) = \varphi_b(0)^2 e^{i\pi z^2}, \quad \varphi_b(0) = e^{i\pi(b^2 + b^{-2})/24} \quad (3.22)$$

and has the “unitarity property” with respect to complex conjugation

$$\overline{\varphi_b(z)} = 1/\varphi_b(\bar{z}). \quad (3.23)$$

Proposition 3. For any $\alpha \in \{\pm 1\}$ the function

$$\Psi_b(\varepsilon, \rho, z) = \frac{1 + i\alpha\rho e^{\pi z/b}(1 - \varepsilon)/2}{\varphi_b(z - ib^{-1}(1 - \varepsilon)/4)} \quad (3.24)$$

has the following properties:

- (i) it is analytic in the domain $-b/2 < \Im(z)$;
- (ii) it solves the functional equation (3.17);

(iii) it satisfies the “inversion relation”

$$\Psi_b(\varepsilon, \rho, z)\Psi_b(\varepsilon, \rho, -z) = \frac{e^{-i\pi z^2}}{2\varphi_b(0)^2}(1 + \varepsilon + i\alpha e^{i\pi b^{-2}/4}(1 - \varepsilon)\rho); \quad (3.25)$$

(iv) it has the “unitarity property” with respect to complex conjugation

$$\overline{\Psi_b(\varepsilon, \rho, z)} = 1/\Psi_b(\varepsilon, \rho, \bar{z}). \quad (3.26)$$

Proof. (i) The function $1/\varphi_b(z)$ is analytic in the domain $-(b + b^{-1})/2 < \Im(z)$ while the function $1/\varphi_b(z - ib^{-1}/2)$ is so in the domain $-b/2 < \Im(z)$. The intersection of these two domains is the domain of analyticity of the function $\Psi_b(\varepsilon, \rho, z)$.

(ii) Using the functional equation (3.20) with a plus sign, we have

$$\begin{aligned} \frac{\Psi_b(\varepsilon, -\rho, x + ib)}{\Psi_b(\varepsilon, \rho, x)} &= \frac{\varphi_b(x - ib^{-1}(1 - \varepsilon)/4)}{\varphi_b(x - ib^{-1}(1 - \varepsilon)/4 + ib)} = 1 + e^{2\pi b(x + ib/2 - ib^{-1}(1 - \varepsilon)/4)} \\ &= 1 + qe^{2\pi bx} e^{-i\pi(1 - \varepsilon)/2} = 1 + \varepsilon q e^{2\pi bx}. \end{aligned}$$

(iii) The case where $\varepsilon = 1$ is equivalent to the inversion relation (3.22). For $\varepsilon = -1$, by using the functional equation (3.20) with a minus sign and the inversion relation (3.22), we have

$$\begin{aligned} \Psi_b(-1, \rho, z)\Psi_b(-1, \rho, -z) &= \frac{(1 + i\alpha\rho e^{\pi z/b})(1 + i\alpha\rho e^{-\pi z/b})}{\varphi_b(z - ib^{-1}/2)\varphi_b(-z - ib^{-1}/2)} \\ &= i\alpha\rho e^{\pi z/b} \frac{(1 - i\alpha\rho e^{-\pi z/b})(1 + i\alpha\rho e^{-\pi z/b})}{\varphi_b(z - ib^{-1}/2)\varphi_b(-z - ib^{-1}/2)} \\ &= \frac{i\alpha\rho e^{\pi z/b}(1 + e^{-2\pi z/b})}{\varphi_b(z - ib^{-1}/2)\varphi_b(-z - ib^{-1}/2)} \\ &= \frac{i\alpha\rho e^{\pi z/b}}{\varphi_b(z - ib^{-1}/2)\varphi_b(-z + ib^{-1}/2)} \\ &= \frac{i\alpha\rho e^{\pi z/b}}{\varphi_b(0)^2 e^{i\pi(z - ib^{-1}/2)^2}} = \frac{i\alpha\rho e^{i\pi b^{-2}/4}}{\varphi_b(0)^2 e^{i\pi z^2}}. \end{aligned}$$

(iv) The case where $\varepsilon = 1$ is equivalent to equation (3.23). For $\varepsilon = -1$, by using (3.23) and (3.20) with a minus sign, we have

$$\begin{aligned} \overline{\Psi_b(-1, \rho, \bar{z})}\Psi_b(-1, \rho, z) &= (1 - i\alpha\rho e^{\pi z/b})(1 + i\alpha\rho e^{\pi z/b}) \frac{\varphi_b(z + ib^{-1}/2)}{\varphi_b(z - ib^{-1}/2)} \\ &= \frac{1 + e^{2\pi z/b}}{1 + e^{2\pi z/b}} = 1. \end{aligned} \quad \square$$

The solution (3.24) is found by Woronowicz [17]. Namely, if $F_{\hbar}(r, \rho)$ is Woronowicz’s quantum exponential function (see formulae (1.18), (1.19) of [17]), then

$$\Psi_b(\varepsilon, \rho, z) = F_{2\pi b^2}(\varepsilon e^{2\pi bz}, \alpha\rho(1 - \varepsilon)/2).$$

This solution, together with $\Psi'_b = 0$, leads to a unitary operator (3.15) when r is self-adjoint. A deep result of Woronowicz and Zakrzewski (Theorem 2.1 of [18]) implies the following

Theorem 3 ([18]). *Let b be such that*

$$ie^{i\pi b^{-2}/4} = \alpha = \pm 1 \quad (3.27)$$

and let $\Psi_b(\varepsilon, \rho, z)$ be defined by equation (3.24). Then, the unitary operator

$$\mathbb{T} = e^{2\pi i \rho \otimes \mathbf{q}} \Psi_b(\mathbf{e} \otimes \mathbf{e}, r \otimes r, \mathbf{q} \otimes 1 + 1 \otimes (\rho - \mathbf{q})) \quad (3.28)$$

satisfies the following pentagonal equation in $\mathcal{A}^{\otimes 3}$:

$$\mathbb{T}_{12}\mathbb{T}_{13}\mathbb{T}_{23} = \mathbb{T}_{23}\mathbb{T}_{12}. \quad (3.29)$$

Condition (3.27) implies that $b = 1/\sqrt{4k+2}$ for some $k \in \mathbb{Z}_{\geq 0}$, and thus $q = e^{i\pi/(4k+2)}$ is a root of unity.

The solution (3.28) of the pentagonal equation has an additional property corresponding to the symmetry with respect to changing the distinguished corners of ideal triangles.

Proposition 4. *Let b and \mathbb{T} be as in Theorem 3 and let $\mathbf{A} \in \mathcal{A}$ be defined by equation (3.7). Then, the following equation holds in $\mathcal{A}^{\otimes 2}$:*

$$\mathbb{T}\mathbf{A} \otimes 1\mathbb{P}\mathbb{T} = \zeta \mathbf{A} \otimes \mathbf{A} \quad (3.30)$$

where $\mathbb{P} \in \mathcal{A}^{\otimes 2}$ is the permutation operator and $\zeta = e^{-i\pi(b+b^{-1})^2/12}$.

Proof. Equation (3.30) is equivalent to the identity

$$e^{2\pi i \rho \otimes \mathbf{q}} e^{2\pi i (\rho - \mathbf{q}) \otimes \rho} \Psi_b(\mathbf{e} \otimes \mathbf{e}, r \otimes r, \rho \otimes 1 - 1 \otimes \mathbf{q}) \Psi_b(\mathbf{e} \otimes \mathbf{e}, r \otimes r, 1 \otimes \mathbf{q} - \rho \otimes 1) = \zeta \mathbf{A} \otimes 1\mathbb{P}$$

which, by using equations (3.25), (3.27), simplifies to

$$e^{2\pi i \rho \otimes \mathbf{q}} e^{2\pi i (\rho - \mathbf{q}) \otimes \rho} e^{-i\pi(\rho \otimes 1 - 1 \otimes \mathbf{q})^2} \varphi_b(0)^{-2} \frac{1}{2} \sum_{m,n \in \{0,1\}} (-1)^{mn} (\mathbf{e} \otimes \mathbf{e})^m (r \otimes r)^n = \zeta \mathbf{A} \otimes 1\mathbb{P}.$$

By definition of the algebra \mathcal{A} , the permutation operator \mathbb{P} has the form

$$\mathbb{P} = \sigma_{L^2(\mathbb{R}^2)} \sigma_{\mathbb{C}^4}$$

where for any vector space V the operator $\sigma_{V^{\otimes 2}} \in \mathrm{End}(V^{\otimes 2})$ is the corresponding permutation operator defined by $\sigma_{V^{\otimes 2}}(x \otimes y) = y \otimes x$ for any $x, y \in V$. One can verify by an explicit and straightforward calculation that

$$\sigma_{\mathbb{C}^4} = \frac{1}{2} \sum_{m,n \in \{0,1\}} (-1)^{mn} (\mathbf{e} \otimes \mathbf{e})^m (r \otimes r)^n$$

and

$$\sigma_{L^2(\mathbb{R}^2)} = e^{i\pi/6} \mathbf{A}^{-1} \otimes 1 e^{2\pi i \rho \otimes \mathbf{q}} e^{2\pi i (\rho - \mathbf{q}) \otimes \rho} e^{-i\pi(\rho \otimes 1 - 1 \otimes \mathbf{q})^2}.$$

Thus, equation (3.30) is true provided $\zeta = \varphi_b(0)^{-2} e^{-i\pi/6}$. \square

Equations (3.8), (3.29), (3.30) permit to construct a projective representation of the groupoid of decorated ideal triangulations of the surface Σ in the space $\mathcal{A}^{\otimes 2\kappa}$ [9]. In particular, we obtain a unitary projective representation of the mapping class group of Σ .

3.4 Spectrum of the length operator

The purpose of this section is to illustrate how different the quantum theory of the whole moduli space could be compared to that of the Teichmüller space.

Let us consider an annulus given by two ideal triangles with side coordinate functions a, b, c and d, b, c with boundary edges given by a and d , and sign variables ε_1 and ε_2 . The contribution to the pull-back of the Goldman's symplectic form coming from these two triangles can be written in the form

$$\frac{dx \wedge dy}{xy}$$

where $x = b/c$ and $y = ad/(bc)$ are two gauge invariant combinations of the coordinate functions. The same gauge invariant combinations of the coordinate functions enter the trace of the holonomy corresponding to the only generator of the fundamental group of this annulus:

$$L_\varepsilon = x + x^{-1} + \varepsilon y$$

where $\varepsilon = \varepsilon_1 \varepsilon_2$. As $\log(x)$ and $\log(y)$ are Poisson conjugate variables, the quantum version of this trace function leads to the following two operators

$$L_\pm = 2 \cosh(2\pi b\mathfrak{q}) \pm e^{2\pi b\mathfrak{p}}$$

corresponding to two possible values $\varepsilon = \pm 1$. Here again operators \mathfrak{p} and \mathfrak{q} are self-adjoint operators in a Hilbert space satisfying the Heisenberg commutation relation (3.3). In particular, in the representation (3.4) the exponential operator $e^{2\pi b\mathfrak{p}}$ becomes a finite shift operator of the form

$$(e^{2\pi b\mathfrak{p}} f)(x) = f(x - ib).$$

To make sense of such pure imaginary shifts, one has to assume that the domain $D(e^{2\pi b\mathfrak{p}})$ of this operator consists of functions $f(z)$ defined in the strip $-b \leq \Im z \leq 0$, analytic in its interior, and with square integrable restrictions to the boundary components $\Im z = 0$ and $\Im z = -b$.

In quantum Teichmüller theory all sign variables take one and the same value and thus one deals with only the operator L_+ , while L_- appears in other connected components of \mathcal{M} . The principal difference between these two non-bounded operators is that L_+ is self-adjoint while L_- is only symmetric but not self-adjoint.

Theorem 4 ([10]). *The spectrum $\mathrm{Sp}(\mathbb{L}_+)$ of the operator \mathbb{L}_+ consists of only a continuous part which is given by a subset of positive real numbers:*

$$\mathrm{Sp}(\mathbb{L}_+) = \mathrm{Sp}_c(\mathbb{L}_+) = \mathbb{R}_{\geq 2}.$$

In the Teichmüller component the absolute value of the trace of the holonomy h_γ along a closed contour γ is related to the length l_γ of the geodesic isotopic to γ through the formula

$$|\mathrm{Tr}(h_\gamma)| = 2 \cosh(l_\gamma/2)$$

Thus, the above result means that the quantum spectrum $\mathrm{Sp}(l_\gamma)$ of the length operator coincides with the set $\mathbb{R}_{\geq 0}$ of its possible classical values.

To solve the spectral problem of \mathbb{L}_- one has to study its self-adjoint extensions.

Theorem 5 ([12]). *For any $\theta \in [0, 1[$ such that $\theta + b^{-2}/4 = \Delta \notin \mathbb{Z}$, the operator \mathbb{L}_- admits a self-adjoint extension \mathbb{L}_θ , $\mathbb{L}_- \subset \mathbb{L}_\theta \subset \mathbb{L}_-^*$, with the spectrum $\mathrm{Sp}(\mathbb{L}_\theta)$ consisting of a continuous part*

$$\mathrm{Sp}_c(\mathbb{L}_\theta) = \mathbb{R}_{\geq 2}$$

and a point part

$$\mathrm{Sp}_p(\mathbb{L}_\theta) = -2 \cosh(2\pi b\sqrt{\theta - \mathbb{Z}_{<\Delta}}).$$

Appearance of a point (discrete) part in the spectrum is an interesting property which presumably reflects the non-trivial topology of the moduli space.

References

- [1] L. O. Chekhov, V. V. Fock, A quantum Teichmüller space. *Theoret. and Math. Phys.* 120 (1999), 1245–1259. [761](#)
- [2] L. O. Chekhov and R. C. Penner, On quantizing Teichmüller and Thurston theories. In *Handbook of Teichmüller theory* (A. Papadopoulos, ed.), Volume I, EMS Publishing House, Zurich 2007, 579–645. [761](#)
- [3] L. D. Faddeev, Discrete Heisenberg–Weyl group and modular group. *Lett. Math. Phys.* 34 (1995), 249–254. [777](#)
- [4] V. V. Fock, 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. [761](#)
- [5] W. M. Goldman, The symplectic nature of fundamental groups of surfaces. *Adv. Math.* 54 (1984), 200–225. [772](#)
- [6] W. M. Goldman, Topological components of spaces of representations. *Invent. Math.* 93 (1988), 557–607. [762](#), [768](#)
- [7] J. Harer, The virtual cohomological dimension of the mapping class group of an oriented surface. *Invent. Math.* 84 (1986), 157–176. [771](#)

- [8] R. M. Kashaev, Quantization of Teichmüller spaces and the quantum dilogarithm. *Lett. Math. Phys.* 43 (1998), 105–115. [761](#), [772](#)
- [9] R. M. Kashaev, Liouville central charge in quantum Teichmüller theory. *Proc. Steklov Inst. Math.* 226 (1999), 63–71. [780](#)
- [10] R. M. Kashaev, The quantum dilogarithm and Dehn twists in quantum Teichmüller theory. In *Integrable structures of exactly solvable two-dimensional models of quantum field theory* (Kiev, 2000), NATO Sci. Ser. II Math. Phys. Chem. 35, Kluwer Academic Publishers, Dordrecht 2001, 211–221. [781](#)
- [11] R. M. Kashaev, Coordinates for the moduli space of flat $\mathrm{PSL}(2, \mathbb{R})$ -connections. *Math. Res. Lett.* 12 (1) (2005), 23–36. [762](#), [765](#), [768](#)
- [12] R. M. Kashaev, On selfadjoint extensions of some difference operator. *St. Petersburg Math. J.* 17 (2006), 157–167. [781](#)
- [13] L. Mosher, Mapping class groups are automatic. *Ann. of Math.* 142 (1995), 303–384. [771](#)
- [14] R. C. Penner, The decorated Teichmüller space of punctured surfaces. *Comm. Math. Phys.* 113 (1987), 299–339. [764](#), [772](#)
- [15] R. C. Penner, Weil–Petersson volumes. *J. Differential Geom.* 35 (1992), 559–608. [772](#)
- [16] J. Teschner, Liouville theory revisited. *Classical Quantum Gravity* 18 (23) (2001), R153–R222. [761](#)
- [17] S. L. Woronowicz, Quantum exponential function. *Rev. Math. Phys.* 12 (6) (2000), 873–920. [762](#), [778](#)
- [18] S. L. Woronowicz, S. Zakrzewski, Quantum ‘ $ax + b$ ’ group. *Rev. Math. Phys.* 14 (7–8) (2002), 797–828. [762](#), [773](#), [779](#)

List of Contributors

Charalampos Charitos, Laboratory of Mathematics, Agricultural University of Athens, 75 Iera Odos, Athens, Greece; email: bakis@aua.gr

Leonid O. Chekhov, Steklov Mathematical Institute, Gubkina 8, 117966, GSP-1, Moscow, Russia; email: chekhov@mi.ras.ru.

Georgios D. Daskalopoulos, Department of Mathematics, Brown University, Providence, RI 02912, U.S.A.; email: daskal@math.brown.edu

Vladimir V. Fock, Institute for Theoretical and Experimental Physics, B.Cheremushkinskaya 25, 117259 Moscow, Russia; email: fock@math.brown.edu

Alexander B. Goncharov, Brown University, Box 1917, Providence, RI 02906, U.S.A.; email: sasha@math.brown.edu

Ursula Hamenstädt, Mathematisches Institut, Universität Bonn, Beringstraße 1, 53115 Bonn, Germany; email: ursula@math.uni-bonn.de

William J. Harvey, Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom; email: bill.harvey@kcl.ac.uk

Frank Herrlich, Mathematisches Institut II, Universität Karlsruhe, 76128 Karlsruhe, Germany; email: herrlich@math.uni-karlsruhe.de

Rinat M. Kashaev, Université de Genève, Section de mathématiques, 2-4, rue du Lièvre, CP 64, 1211 Genève 4, Suisse; email: Rinat.Kashaev@math.unige.ch

Albert Marden, School of Mathematics, University of Minnesota, Minneapolis, MN 55455, U.S.A.; email: am@umn.edu

Christian Mercat, Institut de Mathématiques et de Modélisation de Montpellier (I3M UMR CNRS 5149), Université Montpellier II – CC 51, 34095 Montpellier cedex 5, France; email: mercat@math.univ-montp2.fr

Shigeyuki Morita, Graduate School of Mathematical Sciences, University of Tokyo, Komaba, Tokyo 153-8914, Japan; email: morita@ms.u-tokyo.ac.jp

Lee Mosher, Department of Mathematics and Computer Science, Rutgers University, Newark, NJ 07102, U.S.A.; email: mosher@andromeda.rutgers.edu

Jean-Pierre Otal, Laboratoire Painlevé, Université des Sciences et Techniques de Lille 1, 59655 Villeneuve d'Ascq cedex, France; email: otal@picard.ups-tlse.fr

Ioannis Papadoperakis, Laboratory of Mathematics, Agricultural University of Athens, 75 Iera Odos, Athens, Greece; email: papadoperakis@aua.gr

Athanase Papadopoulos, Institut de Recherche Mathématique Avancée, CNRS et Université Louis Pasteur, 7 rue René Descartes, 67084 Strasbourg cedex, France; email: papadopoulos@math.u-strasbg.fr

Robert C. Penner, Departments of Mathematics and Physics/Astronomy, University of Southern California, Los Angeles, CA 90089, U.S.A.; email: rpenner@math.usc.edu

Gabriela Schmithüsen, Mathematisches Institut II, Universität Karlsruhe, 76128 Karlsruhe, Germany; email: schmithuesen@math.uni-karlsruhe.de

Jörg Teschner, Institut für theoretische Physik, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany, and Max Planck Institut für Gravitationsphysik (Albert Einstein Institut), Am Mühlenberg 1, 14476 Potsdam, Germany; teschner@mail.desy.de

Guillaume Théret, Institut de Recherche Mathématique Avancée, CNRS et Université Louis Pasteur, 7 rue René Descartes, 67084 Strasbourg cedex, France, email: theret@math.u-strasbg.fr

Marc Troyanov, École Polytechnique Fédérale de Lausanne (EPFL), Institut de Géométrie, Algèbre et Topologie, 1015 Lausanne, Switzerland; email: marc.troyanov@epfl.ch

Richard A. Wentworth Department of Mathematics, Johns Hopkins University, Baltimore, MD 21218, U.S.A.; email: wentworth@jhu.edu

Index

- absolute Galois group, 285
- accidental parabolic element, 320
- acylindrical manifold, 413
- admissible complex, 96
- admissible homotopy, 511
- admissible path, 511
- affine deformation, 302, 305
- affine diffeomorphism, 311
- affine group, 312
- Ahlfors conjecture, 412
- Ahlfors Finiteness Theorem, 413
- Ahlfors–Bers map, 228
- algebraic convergence of groups, 426
- anti-stretch line, 164
- anti-stretch ray, 164
- approximating sequence, 623
- Arakelov–Paršin theorem, 94
- arational measured foliation, 140, 391
- asymmetric metric, 165
- asymptotic rays, 141
- augmented Teichmüller space, 298, 314, 318
- automorphism groups of free groups, 361

- Bäcklund transformation, 571
- balancing condition, 97
- Beltrami coefficient, 39, 224, 258
- Beltrami differential, 40, 226, 309
- Beltrami equation, 40, 224, 259
- Beltrami harmonic differential, 45
- Belyi function, 255
 - clean, 255
- Bers boundary, 431
- Bers fiber-space theorem, 265
- Bers slice, 430, 431
 - extended, 431
- Bers universal Teichmüller space, 213
- bilinear relations, 552

- Bochner formula, 50
- boundary, 477
 - Bers, 431
 - Teichmüller, 138
 - Thurston, 193
 - visual, 141
- boundary incompressible, 413
- BP-map, 368
- Braungardt’s construction, 298, 315
- BSCC-map, 368

- canonical dissection, 548
- cataclysm coordinates, 155
- cellular decomposition, 543
- character, 528
- character variety, 38, 72
- cilium, 649
- circle pattern, 570
- coarse Hausdorff topology, 464
- coboundary, 524
- cocycle, 524
- collar lemma, 79
- collar map, 656, 658
- collar weight, 620
- collision, 613
- combination theorems, 407
- compact core of a manifold, 415
 - relative, 415
- comparison map, 477
- comparison triangle, 477
- complete geodesic lamination, 148, 449
- complete train track, 453
- completion of a geodesic lamination, 148
- complex earthquake, 438
- complex of curves, 401, 447
- complex scaling, 435
- conformal blocks, 687, 688
- conformal energy, 547
- conformal field theory, 688, 690, 723

- conformal structure, 120, 123
 - discrete, 543
- conformal structure with boundary, 124
- conformal structures
 - equivalent, 123
- continuous limit, 566
- convergence in the pullback sense, 74
- convergent ray, 139
- convex cocompact, 88
- convex cocompact subgroup, 400
- convex core, 660
- convex core of a manifold, 419
 - Sullivan's Theorem on the, 441
- Corlette–Donaldson theorem, 64, 71
- covering with cusps, 316
 - universal, 316
- critical graph, 119, 131
- critical trajectory, 57, 266
- criticality, 556
- cross ratio, 342, 696, 697
- cross ratio preserving map (discrete), 569
- curvature less than or equal to, 477
- curve graph, 447
- cusps, 142, 265
 - on $\partial\mathfrak{T}(G)$, 429
 - on Bers boundary, 431
 - rank one, 414
 - double horodisk, 414
 - solid cusp cylinder, 413
 - solid pairing tube, 414
 - rank two, 414
 - solid cusp torus, 413
- cusp points, 519
- cusped annulus, 473
- cut system, 341
- cylinder
 - core curve of, 335
 - horizontal, 323
 - maximal, 331
- decent neighbourhood, 317
- decorated measured foliation, 620
- decorated moduli space, 762
- decorated Teichmüller space, 210
- decorated Teichmüller space of the solenoid, 217
- decorated universal Teichmüller space, 212
- deformation map, 298, 319
- deformation of Kleinian groups, 425
- deformation of surface groups, 429
- deformation space boundary, 427
- degenerate hyperbolic pair of pants, 144
- degenerate right-angled hexagon, 145
- degree of a complex bundle, 70
- Dehn twist, 363, 705, 740
- Dehn–Nielsen theorem, 358
- Deligne–Mumford compactification, 297
- δ -hyperbolic, 448
- $\bar{\partial}$ operator, 560
- dessin, 255
 - balanced, 256
 - clean, 256
 - Galois, 256
- developing map, 513
- diffeotopy group, 355
- dilatation, 39, 126, 301
- Dirichlet energy, 547
- discrete area, 547
- discrete Cauchy–Riemann equation, 543
- discrete conformal structure, 543
- discrete derivative, 557
- discrete exponential, 563
- discrete Green function, 574
- discrete Hodge star, 544
- discrete holomorphic function, 543
- discrete Laplacian, 544
- discrete polynomial, 564
- discrete quasiconformal map, 562
- discrete Riemann surface, 542
- discrete series, 564
- discrete wedge product, 546
- Disk Theorem, 441
- distance parameters, 488

- distinguished oriented edge of a tessellation, 211
- distinguished oriented edge of Farey tessellation, 209
- distinguished parameter of a quadratic differential, 130
- domain of discontinuity, 399
- double annulus, 326
- double cusped disc, 473
- doubly cusped region, 317, 320
- duality, 543

- earthquake, 432
- earthquake disk, 441
- Earthquake Theorem, 433
- Eells–Sampson theorem, 49
- Eilenberg–MacLane space, 358
- energy, 47
 - first variation of, 48
 - second variation of, 48
- energy and holomorphic maps, 48, 93
- energy density, 47, 60
- energy extremal problem, 47
- energy minimizer, 47
- equivalent conformal structures, 123
- equivalent hyperbolic structures, 142
- essential curve, 118, 389
- essentially simple, 498
- euclidean triangulation, 510
- extending boundary maps, 416
- extremal length, 128
- extremal length of a homotopy class of curves, 128
- extremal length of a measured foliation class, 129
- extremal map, 39, 267

- family of Riemann surfaces, 337
- Farb–Kaimanovich–Masur theorem, 95
- Farey tessellation, 208
- fat graph, 584
- fat graphs, 693, 698, 701, 707, 731
- Fenchel–Nielsen coordinates, 146, 475, 690, 693, 703, 720
- Fenchel–Nielsen flow, 195
- Fermat curves, 272
- filling curves, 88
- filling foliations, 89
- fine moduli space for Riemann surfaces with Schottky structure, 338
- flat connection, 763
- flat graph connection, 765
- flat structure, 301
- flip transformation, 650, 771
- flux homomorphism, 379
- Fock coordinates, 695, 699, 702
- foliation
 - horizontal, 40, 130, 136
 - measured, 116
 - partial, 118
 - vertical, 40, 130, 136
- freeway, 619
- Fricke space, 38
- \mathcal{F} -structure, 266
- Fuchsian deformation, 225
- Fuchsian group, 38, 224, 659
 - deformation of, 40
- Fuchsian uniformization, 695, 697, 705

- Gardiner’s formula, 69
- generalized hyperbolic pair of pants, 145
- geodesic, 136, 476, 495
 - essentially simple, 498
- geodesic current, 230
- geodesic Hölder distribution, 232
- geodesic lamination, 147, 449
 - complete, 148
 - completion of, 148
 - maximally stretched, 166
 - measured, 150
- geodesic length, 660, 704, 705
- geodesic length operators, 690, 691, 736
- geodesic line, 136, 476
- geodesic metric space, 136
- geodesic operator, 585
- geodesic ray, 136, 303, 309, 322, 476
- geodesic segment, 476

- geometric convergence of groups, 428
- geometric intersection, 133
- geometrically finite group/manifold, 414
- geometrically infinite group/manifold, 414
- Gerstenhaber–Rauch conjecture, 56
- gluing maps, 494
- Goldman’s symplectic structure, 772
- graph length, 623
- graph simple path, 591
- Gromov boundary, 464
- Gromov hyperbolic, 391, 448
- Gromov–Hausdorff convergence, 74
- group
 - handlebody, 344
 - Hecke, 270
 - Möbius, 209
 - mapping class, 81, 123, 253, 260, 294, 312, 339
 - modular, 209
- group cohomology, 524
- groupoid, 677

- h.s.c.s., 478
- handlebody group, 344
- harmonic map, 47, 61
 - combinatorial, 98
 - equivariant, 63
 - existence of, 49, 64
 - order of, 62
 - uniqueness of, 51
- Hausdorff distance, 449
- Hausdorff topology, 449
- Hecke group, 270
- height of a measured foliation with one cylinder, 119
- Higgs bundle, 69
 - moduli space of, 71
 - polystable, 70
 - stability of, 70
- Higgs bundle and Teichmüller’s theorem, 73
- Higgs bundle and the Milnor–Wood inequality, 73
- hinge, 512
- Hirota system, 570
- Hitchin map, 72
- Hitchin–Simpson theorem, 70
- holomorphic form
 - discrete, 544
- holomorphic function
 - discrete, 543
- holomorphic motion, 239, 442
- holomorphic quadratic differential, 450
 - admissible, 302
 - critical point of, 300, 302
 - regular point of, 300, 302
- holonomy, 514
- homeotopy group, 359
- Hopf differential, 51
- horizontal foliation, 130, 136
- horocycle, 664
- horocycle topology, 315
- horocyclic area map, 665, 668
- horocyclic foliation, 497
- horocyclic measured foliation, 152
- horocyclic measured geodesic lamination, 153
- horogeodesic path, 158
 - minimal, 159
- Hubbard–Masur theorem, 58
- hyperbolic metric space, 392
- hyperbolic pair of pants, 144, 473
 - degenerate, 144
 - generalized, 145
- hyperbolic structure, 142
 - ideal, 494
- hyperbolic surface with conical singularities, 478
- Hyperbolization Theorem, 421

- IA automorphism group, 362
- ideal boundary, 64

- ideal hyperbolic metric, 494
- ideal hyperbolic structure, 494
- ideal triangulation, 146, 210
- intersection form, 450
- intersection graph, 297
- intersection number, 548
- irreducible subgroup, 397
- Ishihara's theorem, 49
- isothermal coordinates, 38
- isotrivial, 93

- Jenkins–Strebel differential, 58, 278, 322
- Jenkins–Strebel geodesic, 137
- Jenkins–Strebel ray, 298, 323
 - end point of, 299, 328
- Jenkins–Strebel rays
 - modularly equivalent, 138
 - similar, 137
- Jenkins–Strebel structure, 131
- Johnson homomorphism, 371
- Johnson subgroup, 369

- Kashaev coordinates, 700–702, 708
- Kleinian group, 412
- Knizhnik–Zamolodchikov connection, 688
- Korevaar–Schoen limits, 75

- lambda length, 206
- λ -Lemma, 442
- lamination, 652
 - bounded, 656
 - even, 658
 - integral, 655, 656
 - real, 659
 - unbounded, 652
- lattice in $\mathrm{PSL}_2(\mathbb{R})$, 296
- Laurent property, 674
- leaf, 147
- length functional, 143
- length space, 56
- limit set, 224, 398
- line
 - anti-stretch, 164
 - stretch, 164
- line of minima, 200
- Liouville map, 231
- Liouville measure, 231
- Liouville theory, 687
- local quasiconformal dilatation, 126
- locally ringed space, 315
- lower central series, 371

- Malcev completion, 372
- map
 - Hitchin, 72
 - quasiconformal, 39, 126, 562
 - quasisymmetric, 213, 259
 - stretch, 164
 - Teichmüller, 134
- mapping class
 - as outer automorphism, 81
 - as Weil–Petersson isometry, 83
 - finite order, 393
 - pseudo-Anosov, 83, 392, 394
 - pseudoperiodic, 84
 - reducible, 83, 392, 394
 - semisimple, 84
 - Thurston classification, 83, 392
- mapping class group, 81, 123, 253, 260, 294, 312, 339, 357, 387, 448, 521, 589, 651, 676, 687, 700, 702
 - classification of subgroups, 91
 - superrigidity of, 95
- mapping class group representation, 687
- marked Riemann surface, 294
- marking, 38, 703, 720, 731
- Markov numbers, 679
- maximal dilatation, 301
- maximal geodesic lamination, 449
- measure-equivalence, 119
- measured foliation, 57, 116, 390, 606
 - arational, 140, 391
 - extremal length of a, 68
 - horocyclic, 152
 - intersection number, 59
 - projective equivalence, 58

- totally transverse, 154
- uniquely ergodic, 140
- measured foliation class, 117
 - extremal length of a , 129
- measured foliation with one cylinder, 119
 - height of a , 119
- measured geodesic lamination, 150, 450
 - horocyclic, 153
 - length of a , 151
 - minimal, 194
 - projective, 151
 - topological, 191
 - totally transverse, 194
 - uniquely ergodic, 194
- metric
 - asymmetric, 165
 - Teichmüller, 43, 133, 261, 267, 294, 301
 - Weil–Petersson, 78
- metric graph, 377
- minimal action, 67
- minimal geodesic lamination, 449
- minimal measured geodesic lamination, 194
- mirror Veech group, 314
- Möbius group, 209
- modular curve \mathcal{C}_g , 253
- modular double, 691
- modular family of Riemann surfaces $\mathcal{C}_{g,n}$, 264
- modular functor, 689–691, 720
- modular group, 209
 - commensurability, 274
- modular group
 - Teichmüller (*see also* mapping class group), 261
- modular groupoid, 589, 690, 720, 726, 743
- modularly equivalent Jenkins–Strebel rays, 138
- moduli space, 253, 262, 294, 531, 689
 - boundary of, 297
 - coarse, 295
 - compactification of, 297, 318
 - fine, 295
- moduli space of flat connections, 763
- moduli space of metric graphs, 378
- moduli space of Riemann surfaces, 377
- moduli spaces, 688
- modulus, 124, 463
- modulus of a cylinder, 128
- modulus of a homotopy class of curves, 128
- monodromy group, 659
- monotonicity formula for energy minimizers, 61
- Morgan–Shalen compactification, 75
- Mostow rigidity, 421
- moving frame, 574
- multiple curve, 591
- Mumford–Mahler compactness, 82
- nearest point retraction, 420
- Nielsen realization problem, 85, 87
- nilpotent quotient, 371
- nonpositively curved space, 54, 56
- nonrefraction, 80
- norm of a quadratic differential, 130
- origami, 296
- outer automorphism groups of free groups, 361
- outer automorphisms, 521
- Outer Space, 376
- pairing
 - additive, 672
 - intersection, 672
 - multiplicative, 673
- pants decomposition, 447, 689, 690, 720
- parabolic element, 413
- partial foliation, 118
- patchwork of double annuli, 326
- patchwork of rectangles, 323
- peakless function, 189

- Penner coordinates, 690, 693, 695, 699, 705
- pentagon relation, 678
- period matrix, 555
- piecewise flat surface, 510
- pinched, 213
- ping pong
 - hyperbolic, 403
 - topological, 395
- pleated surface, 434
- Poincaré metric, 38
- polygon, 478
 - realizable, 478
- polygonal pattern, 561
- polystable, 70
- precisely invariant, 316
- pro-finite completion, 215
- product foliation structure, 129
- projective foliation class, 117
- projective measured geodesic
 - lamination, 151
- proper action, 65
- proper length, 622
- proper metric space, 136
- pseudo-Anosov homeomorphism, 257, 392
- Ptolemy groupoid, 689, 697, 717
- Ptolemy transformation, 208
- punctured solenoid, 214
- punctured surface, 508, 518
- pure mapping class group, 521
- quadratic differential, 40, 130, 389
 - initial, 42
 - Jenkins–Strebel, 58
 - Jenkins–Strebel type, 278
 - terminal, 42
- quadrilateral, 124
 - modulus of a, 124
- quantization, 594
- quantization of Teichmüller spaces, 687, 689, 709
- quantum dilogarithm function, 777
- quantum flip transformation, 776
- quantum geodesic, 599
- quantum multicurve, 600
- quantum ordering, 600
- quantum triangle algebra, 773
- quantum Whitehead move, 595
- quasi-Fuchsian deformation, 225
- quasi-Fuchsian group, 418
- quasi-Fuchsian representation, 88
- quasi-geodesic, 454
- quasiconformal deformation space, 423
- quasiconformal dilatation, 126
 - local, 126
- quasiconformal homeomorphism, 39, 126, 224, 259
 - Beltrami coefficient of, 39
 - dilatation of, 39
 - extremal, 39
- quasiconformal map
 - discrete, 562
- quasisymmetric homeomorphism, 213, 259
- R-space, 316
- rational conformal field theory, 688, 690, 723
- ray, 64
 - anti-stretch, 164
 - convergent, 139
 - stretch, 164
- rays
 - asymptotic, 141
- recurrent train track, 453
- reducible mapping class, 394
- reducible subgroup, 397
- reductive, 71
- regular B-group, 320
- regular point, 62
- representation variety, 425, 524
 - discreteness locus of, 426
- representations of groupoids, 714
- representations of mapping class groups, 687

- Reshetikhin–Turaev invariants, 688
- ribbon graph, 584
- Riemann moduli space, 82
 - Deligne–Mumford compactification of, 82
- Riemann surface, 120
 - discrete, 542
- Riemann surfaces
 - universal family of, 294
- Riemann’s extension theorem, 316
- rigged Riemann surface, 721
- right-angled hexagon
 - degenerate, 145
- rigidity of hyperbolic manifolds, 421
- saddle connection, 331
- Schottky basis, 336
- Schottky covering, 299, 336
 - generalized, 300
- Schottky group, 299, 336, 418
- Schottky space, 299, 337, 339
 - extended, 343
 - partial compactification of, 341
- Schottky structure, 338, 342
- Schottky subgroup, 402
- Schottky uniformization, 299, 336
- self-duality equations, 70
- semifield, 668
- semisimple mapping class, 84
- semisimple representation, 65
- shear coordinate, 587, 620
- shift parameter, 146, 495
- shifting, 613
- Siegel modular group, 367
- similar Jenkins–Strebel rays, 137
- similar representations, 529
- simultaneous uniformization, 430
- singular edge, 494
- singular point, 62
- singular set, 92, 96, 660
- singularity order, 517
- Skora’s theorem, 67
- slope, 70
- small action, 67
- spine, 584
- splitting sequence, 454
- square-tiled surface, 296
- stability, 70
- stable Riemann surface, 297, 317, 341
 - Schottky covering of, 341
- stable Riemann surfaces
 - family of, 321
- standard generators, 339
- stretch line, 164
- stretch map, 164
- stretch ray, 164
- structure sheaf, 315
- stump, 151
- sublamination, 450
- sufficiently large, 91
- surface
 - ciliated, 649
 - decorated, 665
- switch conditions, 611
- symplectic basis, 366
- symplectic class, 369
- symplectic homomorphism, 343
- symplectic topology, 378
- symplectomorphism group, 356
- systole function, 200
- Tameness Theorem, 415
- Teichmüller boundary, 138, 267
- Teichmüller compactification, 139
- Teichmüller curve, 296, 311, 314, 346
 - boundary points of, 334
 - cusps of, 297, 299
- Teichmüller deformation, 295, 300, 301
- Teichmüller deformation of
 - punctured Riemann surface, 302
- Teichmüller disc, 265, 295, 303, 304, 404
 - boundary points of, 299, 330, 334

- global stabilizer of, 312
- image in Schottky space of, 347
- stabilizer of, 296
- Teichmüller embedding, 295, 303, 346
 - extension of, 299, 332
- Teichmüller existence theorem, 41, 57
- Teichmüller extremal problem, 39
- Teichmüller geodesic, 295, 391
- Teichmüller line, 135
 - horizontal foliation of a, 136
 - vertical foliation of a, 136
- Teichmüller map, 41, 134
- Teichmüller metric, 43, 133, 261, 267, 294, 301, 448
- Teichmüller modular group, 261, 294
- Teichmüller ray, 135
- Teichmüller space, 42, 44, 123, 143, 209, 294, 357, 376, 389, 448, 531, 659
 - \mathcal{A} -space, 664
 - \mathcal{X} -space, 660
 - augmented, 314
 - cotangent space of, 45
 - partial compactification of, 314, 318
 - tangent space of, 45
- Teichmüller space is Stein, 90
- Teichmüller space of decorated surfaces, 665
- Teichmüller structure, 295, 320
 - universal, 339
- Teichmüller theorem, 43, 53, 73, 87
- Teichmüller uniqueness theorem, 41, 55
- tension field, 47
- tesselation, 211
- three dimensional consistency, 573
- Thurston boundary, 76, 193, 267
- Thurston compactification, 76, 669
- Tits alternative, 397
- topological cylinder, 128
 - modulus of, 128
- topological measured geodesic
 - lamination, 191
- topological triangle, 124
- Torelli group, 367
- totally transverse measured foliation, 154
- totally transverse measured geodesic lamination, 194
- train track, 452, 611
- trajectory, 322
 - critical, 322
 - horizontal, 322
 - regular, 322
 - vertical, 322
- translation length
 - Weil–Petersson, 84
- translation length of a representation, 63
- translation surface, 269
- transverse measure, 450, 453, 607, 611
- transversely recurrent train track, 453
- tree, 57
 - folding of, 67
 - minimal action on a, 67
 - morphism of, 67
 - small action on a, 67
- tree dual to a foliation, 58
- tree of projective lines, 342
- triangle, 477
 - imaginary, 485
- triangulation, 649
- tropical analogue, 669
- tropicalization, 668
- twist parameter, 474
- 2-dimensional simplicial complex, 493
- uniformization theorem, 38
- unipotent subgroup, 763
- uniquely ergodic measured foliation, 140
- uniquely ergodic measured geodesic lamination, 194
- universal branched cover, 511
- universal family, 294, 319
- universal family over $\bar{\mathcal{T}}_{g,n}$, 320
- universal Teichmüller space, 212
- unparametrized quasi-geodesic, 462
- Veech curve, 272, 273
- Veech group, 296, 311, 314

- parabolic elements of, 335
- Veech surface, 311
- vertex, 494
- vertex cycle, 454
- vertical foliation, 130, 136
- visual boundary, 141
- volume preserving diffeomorphism group, 356
- vortex equations, 73

- weakly hyperbolic, 401
- Weil–Petersson form, 588, 669, 689, 705, 762

- Weil–Petersson metric, 78
 - completion of, 79
 - curvature of, 78
 - geodesic convexity of, 88
 - geodesics of, 83
 - incompleteness of, 78
 - isometry group of, 83
- Weyl ordering, 600
- Whitehead collapse, 607
- Whitehead equivalence, 116
- Whitehead move, 116, 587

- Zygmund spaces, 243