

6

Unique ergodicity

This chapter is dedicated to a distinguished class of dynamical systems, characterized by the fact that they admit exactly one invariant probability measure. Initially, in Section 6.1, we give alternative formulations of this property and we analyze the properties of the unique invariant measure.

The relation between unique ergodicity and minimality is another important theme. A dynamical system is said to be *minimal* if every orbit is dense in the ambient space. As we observe in Section 6.2, every uniquely ergodic system is minimal, restricted to the support of the invariant measure, but the converse is not true, in general.

The main construction of uniquely ergodic transformations is algebraic in nature. In Section 6.3 we introduce the notion of the Haar measure of a topological group and we show that every transitive translation on a compact metrizable topological group is minimal and even uniquely ergodic: the Haar measure is the unique invariant probability measure.

In Section 6.4 we present a remarkable application of the idea of unique ergodicity in the realm of arithmetics: the theorem of Hermann Weyl on the equidistribution of polynomial sequences.

Throughout this chapter, unless stated otherwise, it is understood that M is a compact metric space and $f : M \rightarrow M$ is a continuous transformation.

6.1 Unique ergodicity

We say that a transformation $f : M \rightarrow M$ is *uniquely ergodic* if it admits exactly one invariant probability measure. The corresponding notion for flows is defined in precisely the same way. This denomination is justified by the observation that the invariant probability measure μ is necessarily ergodic. Indeed, suppose there existed some invariant set $A \subset M$ with $0 < \mu(A) < 1$. Then the normalized restriction of μ to A , defined by

$$\mu_A(E) = \frac{\mu(E \cap A)}{\mu(A)} \quad \text{for every measurable set } E \subset M,$$

would be an invariant probability measure, different from μ , which would contradict the assumption that f is uniquely ergodic.

Proposition 6.1.1. *The following conditions are equivalent:*

- (i) f admits a unique invariant probability measure;
- (ii) f admits a unique ergodic probability measure;
- (iii) for every continuous function $\varphi : M \rightarrow \mathbb{R}$, the sequence of time averages $n^{-1} \sum_{j=0}^{n-1} \varphi(f^j(x))$ converges at every point to a constant;
- (iv) for every continuous function $\varphi : M \rightarrow \mathbb{R}$, the sequence of time averages $n^{-1} \sum_{j=0}^{n-1} \varphi \circ f^j$ converges uniformly to a constant.

Proof. It is easy to see that (ii) implies (i). Indeed, since invariant measure is a convex combination of ergodic measures (Theorem 5.1.3), if there is a unique ergodic probability measure then the invariant probability measure is also unique. It is clear that (iv) implies (iii), since uniform convergence implies pointwise convergence. To see that (iii) implies (ii), suppose that μ and ν are ergodic probability measures of f . Then, given any continuous function $\varphi : M \rightarrow \mathbb{R}$,

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) = \begin{cases} \int \varphi d\mu & \text{at } \mu\text{-almost every point} \\ \int \varphi d\nu & \text{at } \nu\text{-almost every point.} \end{cases}$$

Since, by assumption, the limit does not depend on the point x , it follows that

$$\int \varphi d\mu = \int \varphi d\nu$$

for every continuous function φ . Using Proposition A.3.3 we find that $\mu = \nu$.

We are left to prove that (i) implies (iv). Start by recalling that f admits some invariant probability measure μ (by Theorem 2.1). The idea is to show that if (iv) does not hold then there exists some probability measure $\nu \neq \mu$ and, hence, (i) does not hold either. Suppose then that (iv) does not hold, that is, that there exists some continuous function $\varphi : M \rightarrow \mathbb{R}$ such that $n^{-1} \sum_{j=0}^{n-1} \varphi \circ f^j$ does not converge uniformly to any constant; in particular, it does not converge uniformly to $\int \varphi d\mu$. By definition, this means that there exists $\varepsilon > 0$ such that for every $k \geq 1$ there exist $n_k \geq k$ and $x_k \in M$ such that

$$\left| \frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi(f^j(x_k)) - \int \varphi d\mu \right| \geq \varepsilon. \tag{6.1.1}$$

Let us consider the sequence of probability measures

$$\nu_k = \frac{1}{n_k} \sum_{j=0}^{n_k-1} \delta_{f^j(x_k)}.$$

Since the space $\mathcal{M}_1(M)$ of probability measures on M is compact for the weak* topology (Theorem 2.1.5), up to replacing this sequence by a subsequence,

we may suppose that it converges to some probability measure ν on M . By Lemma 2.2.4 applied to the Dirac measure δ_x , the probability measure ν is invariant under f . On the other hand, the fact that $(\nu_k)_k$ converges to ν in the weak* topology implies that

$$\int \varphi d\nu = \lim_k \int \varphi d\nu_k = \lim_k \frac{1}{n_k} \sum_{j=0}^{n_k-1} \varphi(f^j(x_k)).$$

Then, recalling (6.1.1), we have that

$$\left| \int \varphi d\nu - \int \varphi d\mu \right| \geq \varepsilon.$$

In particular, $\nu \neq \mu$. This concludes the argument.

6.1.1 Exercises

- 6.1.1. Give an example of a transformation $f : M \rightarrow M$ in a compact metric space such that $(1/n) \sum_{j=0}^{n-1} \varphi \circ f^j$ converges uniformly, for every continuous function $\varphi : M \rightarrow \mathbb{R}$, but f is not uniquely ergodic.
- 6.1.2. Let $f : M \rightarrow M$ be a transitive continuous transformation in a compact metric space. Show that if $(1/n) \sum_{j=0}^{n-1} \varphi \circ f^j$ converges uniformly, for every continuous function $\varphi : M \rightarrow \mathbb{R}$, then f is uniquely ergodic.
- 6.1.3. Let $f : M \rightarrow M$ be an isometric homeomorphism in a compact metric space M . Show that if μ is an ergodic measure for f then, for every $n \in \mathbb{N}$, the function $\varphi(x) = d(x, f^n(x))$ is constant on the support of μ .

6.2 Minimality

Let $\Lambda \subset M$ be a closed invariant set of $f : M \rightarrow M$. We say that Λ is *minimal* if it coincides with the closure of the orbit $\{f^n(x) : n \geq 0\}$ of every point $x \in \Lambda$. We say that the transformation f is minimal if the ambient M is a minimal set.

Recall that the support of a measure μ is the set of all points $x \in M$ such that $\mu(V) > 0$ for every neighborhood V of x . It follows immediately from the definition that the complement of the support is an open set: if $x \notin \text{supp } \mu$ then there exists an open neighborhood V such that $\mu(V) = 0$; then V is contained in the complement of the support. Therefore, $\text{supp } \mu$ is a closed set.

It is also easy to see that the support of any invariant measure is an invariant set, in the following sense: $f(\text{supp } \mu) \subset \text{supp } \mu$. Indeed, let $x \in \text{supp } \mu$ and let V be any neighborhood of $y = f(x)$. Since f is continuous, $f^{-1}(V)$ is a neighborhood of x . Then $\mu(f^{-1}(V)) > 0$, because $x \in \text{supp } \mu$. Hence, using that μ is invariant, $\mu(V) > 0$. This proves that $y \in \text{supp } \mu$.

Proposition 6.2.1. *If $f : M \rightarrow M$ is uniquely ergodic then the support of the unique invariant probability measure μ is a minimal set.*

Proof. Suppose that there exists $x \in \text{supp } \mu$ whose orbit $\{f^j(x) : j \geq 0\}$ is not dense in the support of μ . This means that there exists some open subset U of M such that $U \cap \text{supp } \mu$ is non-empty and

$$f^j(x) \notin U \cap \text{supp } \mu \quad \text{for every } j \geq 0. \quad (6.2.1)$$

Let ν be any accumulation point of the sequence of probability measures

$$\nu_n = n^{-1} \sum_{j=0}^{n-1} \delta_{f^j(x)}, \quad n \geq 1$$

with respect to the weak* topology. Accumulation points do exist, by Theorem 2.1.5, and ν is an invariant probability measure, by Lemma 2.2.4. The condition (6.2.1) means that $\nu_n(U) = 0$ for every $n \geq 1$. Hence, using Theorem 2.1.2 (see also part 3 of Exercise 2.1.1) we have that $\nu(U) = 0$. This implies that no point of U is in the support of μ , which contradicts the fact that $U \cap \text{supp } \mu$ is non-empty.

The converse to Proposition 6.2.1 is false in general:

Theorem 6.2.2 (Furstenberg). *There exists some real-analytic diffeomorphism $f : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ that is minimal, preserves the Lebesgue measure m on the torus, but is not ergodic for m . In particular, f is not uniquely ergodic.*

In the remainder of this section we give a brief sketch of the proof of this result. A detailed presentation may be found in the original paper of Furstenberg [Fur61], as well as in Mañé [Mañ87]. In Section 7.3.1 we mention other examples of minimal transformations that are not uniquely ergodic.

To prove Theorem 6.2.2, we look for a transformation $f : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ of the form $f(x, y) = (x + \alpha, y + \phi(x))$, where α is an irrational number and $\phi : S^1 \rightarrow \mathbb{R}$ is a real-analytic function with $\int \phi(x) dx = 0$. Note that f preserves the Lebesgue measure on \mathbb{T}^2 . Let us also consider the map $f_0 : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ given by $f_0(x, y) = (x + \alpha, y)$. Note that no orbit of f_0 is dense in \mathbb{T}^2 and that the system (f_0, m) is not ergodic.

Let us consider the *cohomological equation*

$$u(x + \alpha) - u(x) \equiv \phi(x). \quad (6.2.2)$$

If ϕ and α are such that (6.2.2) admits some measurable solution $u : S^1 \rightarrow \mathbb{R}$ then (f_0, m) and (f, m) are ergodically equivalent (see Exercise 6.2.1) and, consequently, (f, m) is not ergodic. On the other hand, one can show that if (6.2.2) admits no continuous solution then f is minimal (the converse to this fact is Exercise 6.2.2). Therefore, it suffices to find ϕ and α such that the cohomological equation admits a measurable solution but not a continuous solution.

It is convenient to express these conditions in terms of the Fourier expansion $\phi(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}$. To ensure that ϕ is real-analytic it is enough to

require that:

$$\text{there exists } \rho < 1 \text{ such that } |a_n| \leq \rho^n \text{ for every } n \text{ sufficiently large.} \quad (6.2.3)$$

Indeed, in that case the series $\sum_{n \in \mathbb{Z}} a_n z^n$ converges uniformly on every corona $\{z \in \mathbb{C} : r \leq |z| \leq r^{-1}\}$ with $r > \rho$. In particular, its sum in the unit circle, which coincides with ϕ , is a real-analytic function. Since we want ϕ to take values in the real line and to have zero average, we must also require:

$$a_0 = 0 \quad \text{and} \quad a_{-n} = \bar{a}_n \text{ for every } n \geq 1. \quad (6.2.4)$$

According to Exercise 6.2.3, the cohomological equation admits a solution in the space $L^2(m)$ if and only if

$$\sum_{n=1}^{\infty} \left| \frac{a_n}{e^{2\pi n i \alpha} - 1} \right|^2 < \infty. \quad (6.2.5)$$

Moreover, the solution is uniquely determined: $u = \sum_{n \in \mathbb{Z}} b_n e^{2\pi i n x}$ with

$$b_n = \frac{a_n}{e^{2\pi i n \alpha} - 1} \quad \text{for every } n \in \mathbb{Z}. \quad (6.2.6)$$

Fejér’s theorem (see [Zyg68]) states that if u is a continuous function then the sequence of partial sums of its Fourier expansion converges Cesàro uniformly to u :

$$\frac{1}{n} \sum_{k=1}^n \left(\sum_{j=-k}^k b_j e^{2\pi i j x} \right) \text{ converges uniformly to } u(x). \quad (6.2.7)$$

Hence, to ensure that u is not continuous it suffices to require:

$$\left(\sum_{j=-k}^k b_j \right)_k \text{ is not Cesàro convergent.} \quad (6.2.8)$$

In this way, the problem is reduced to finding α and $(a_n)_n$ that satisfy (6.2.3), (6.2.4), (6.2.5) and (6.2.8). Exercise 6.2.4 hints at the issues involved in the choice of such objects.

6.2.1 Exercises

- 6.2.1. Show that if u is a measurable solution of the cohomological equation (6.2.2) then $h : \mathbb{T}^2 \rightarrow \mathbb{T}^2, h(x, y) = (x, y + u(x))$ is an ergodic equivalence between (f_0, m) and (f, m) , that is, h is an invertible measurable transformation that preserves the measure m and conjugates the two maps f and f_0 . Deduce that (f, m) cannot be ergodic.
- 6.2.2. Show that if u is a continuous solution of the cohomological equation (6.2.2) then $h : \mathbb{T}^2 \rightarrow \mathbb{T}^2, h(x, y) = (x, y + u(x))$ is a topological conjugacy between f_0 and f . In particular, f cannot be transitive.

6.2.3. Check that if $u(x) = \sum_{n \in \mathbb{Z}} b_n e^{2\pi i n x}$ is a solution of (6.2.2) then

$$b_n = \frac{a_n}{e^{2\pi i n \alpha} - 1} \quad \text{for every } n \in \mathbb{Z}. \quad (6.2.9)$$

Moreover, $u \in L^2(m)$ if and only if $\sum_{n=1}^{\infty} |b_n|^2 < \infty$.

6.2.4. We say that an irrational number α is *Diophantine* if there exist $c > 0$ and $\tau > 0$ such that $|q\alpha - p| \geq c|q|^{-\tau}$ for any $p, q \in \mathbb{Z}$ with $q \neq 0$. Show that the condition (6.2.5) is satisfied whenever α is Diophantine and ϕ satisfies (6.2.3).

6.2.5. (Theorem of Gottschalk) Let $f : M \rightarrow M$ be a continuous map in a compact metric space M . Show that the closure of the orbit of a point $x \in M$ is a minimal set if and only if $R_\varepsilon = \{n \in \mathbb{Z} : d(x, f^n(x)) < \varepsilon\}$ is a syndetic set for every $\varepsilon > 0$.

6.2.6. Let $f : M \rightarrow M$ be a continuous map in a compact metric space M . We say that $x, y \in M$ are *close* if $\inf_n d(f^n(x), f^n(y)) = 0$. Show that if $x \in M$ is such that the closure of its orbit is a minimal set then, for every neighborhood U of x and every point y close to x , there exists an increasing sequence $(n_i)_i$ such that $f^{n_i + \dots + n_{i_k}}(x)$ and $f^{n_i + \dots + n_{i_k}}(y)$ are in U for any $i_1 < \dots < i_k$ and $k \geq 1$.

6.2.7. (Theorem of Hindman) A theorem of Auslander and Ellis (see [Fur81, Theorem 8.7]) states that in the conditions of Exercise 6.2.6 the closure of the orbit of every $y \in M$ contains some point x that is close to y and such that the closure of its orbit is a minimal set. Deduce the following refinement of the theorem of van der Waerden: given any decomposition $\mathbb{N} = S_1 \cup \dots \cup S_q$ of the set of natural numbers into pairwise disjoint sets, there exists j such that S_j contains a sequence $n_1 < \dots < n_i < \dots$ such that $n_{i_1} + \dots + n_{i_k} \in S_j$ for every $k \geq 1$ and any $i_1 < \dots < i_k$.

6.3 Haar measure

We are going to see that every compact topological group carries a remarkable probability measure, called the *Haar measure*, that is invariant under every translation and every surjective group endomorphism. Assuming that the group is metrizable, every transitive translation is uniquely ergodic, with the Haar measure as the unique invariant probability measure.

6.3.1 Rotations on tori

Fix $d \geq 1$ and a rationally independent vector $\theta = (\theta_1, \dots, \theta_d)$. As we have seen in Section 4.2.1, the rotation $R_\theta : \mathbb{T}^d \rightarrow \mathbb{T}^d$ is ergodic with respect to the Lebesgue measure m on the torus. Our goal now is to show that, in fact, R_θ is uniquely ergodic.

According to Proposition 6.1.1, we only have to show that, given any continuous function $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$, there exists $c_\varphi \in \mathbb{R}$ such that

$$\varphi_n = \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ R_\theta^j \quad \text{converges to } c_\varphi \text{ at every point.} \quad (6.3.1)$$

Take $c_\varphi = \int \varphi dm$. By ergodicity, the sequence $(\varphi_n)_n$ of time averages converges to c_φ at m -almost every point. In particular, $\varphi_n(x) \rightarrow c_\varphi$ for a dense subset of values of $x \in \mathbb{T}^d$.

Let d be the distance induced in the torus $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ by the usual norm in \mathbb{R}^d : the distance between any two points in the torus is the minimum of the distances between all their representatives in \mathbb{R}^d . It is clear that the rotation R_θ preserves that distance:

$$d(R_\theta(x), R_\theta(y)) = d(x, y) \quad \text{for every } x, y \in \mathbb{T}^d.$$

Then, using that φ is continuous, given any $\varepsilon > 0$ we may find $\delta > 0$ such that

$$d(x, y) < \delta \Rightarrow d(R_\theta^j(x), R_\theta^j(y)) < \delta \Rightarrow |\varphi(R_\theta^j(x)) - \varphi(R_\theta^j(y))| < \varepsilon$$

for every $j \geq 0$. Then,

$$d(x, y) < \delta \Rightarrow |\varphi_n(x) - \varphi_n(y)| < \varepsilon \quad \text{for every } n \geq 1.$$

Since ε does not depend on n , this proves that the sequence $(\varphi_n)_n$ is equicontinuous.

This allows us to use the theorem of Ascoli to prove the claim (6.3.1), as follows. Suppose that there exists $\bar{x} \in \mathbb{T}^d$ such that $(\varphi_n(\bar{x}))_n$ does not converge to c_φ . Then there exists $c \neq c_\varphi$ and some subsequence $(n_k)_k$ such that $\varphi_{n_k}(\bar{x})$ converges to c when $k \rightarrow \infty$. By the theorem of Ascoli, up to restricting to a subsequence we may suppose that $(\varphi_{n_k})_k$ is uniformly convergent. Let ψ be its limit. Then ψ is a continuous function such that $\psi(x) = c_\varphi$ for a dense subset of values of $x \in \mathbb{T}^d$ but $\psi(\bar{x}) = c$ is different from c_φ . It is clear that such a function does not exist. This contradiction proves our claim that R_θ is uniquely ergodic.

6.3.2 Topological groups and Lie groups

Recall that a *topological group* is a group (G, \cdot) endowed with a topology with respect to which the two operations

$$G \times G \rightarrow G, (g, h) \mapsto gh \quad \text{and} \quad G \rightarrow G, g \mapsto g^{-1} \tag{6.3.2}$$

are continuous. In all that follows it is assumed that the topology of G is such that every set consisting of a single point is closed. When G is a manifold and the operations in (6.3.2) are differentiable, we say that (G, \cdot) is a *Lie group*. See Exercise 6.3.1.

The Euclidean space \mathbb{R}^d is a topological group, and even a Lie group, relative to addition $+$, and the same holds for the torus \mathbb{T}^d . Recall that \mathbb{T}^d is the quotient of \mathbb{R}^d by its subgroup \mathbb{Z}^d . This construction may be generalized as follows:

Example 6.3.1. Given any closed normal subgroup H of a topological group G , let G/H be the set of equivalence classes for the equivalence relation defined

in G by $x \sim y \Leftrightarrow x^{-1}y \in H$. Denote by xH the equivalence class that contains each $x \in G$. Consider the following group operation in G/H :

$$xH \cdot yH = (x \cdot y)H.$$

The hypothesis that H is a normal subgroup ensures that this operation is well defined. Let $\pi : G \mapsto G/H$ be the canonical projection, given by $\pi(x) = xH$. Consider in G/H the quotient topology, defined in the following way: a function $\psi : G/H \rightarrow X$ is continuous if and only if $\psi \circ \pi : G \rightarrow X$ is continuous. The hypothesis that H is closed ensures that the points are closed subsets of G/H . It follows easily from the definitions that G/H is a topological group. Recall also that if the group G is abelian then all subgroups are normal.

Example 6.3.2 (Linear group). The set $G = \text{GL}(d, \mathbb{R})$ of invertible real matrices of dimension d is a Lie group for the multiplication of matrices, called *real linear group* of dimension d . Indeed, G may be identified with an open subset of the Euclidean space $\mathbb{R}^{(d^2)}$ and, thus, has a natural structure of a differentiable manifold. Moreover, it follows directly from the definitions that the multiplication of matrices and the inversion map $A \mapsto A^{-1}$ are differentiable with respect to this manifold structure. G has many important Lie subgroups, such as the *special linear group* $\text{SL}(d, \mathbb{R})$, consisting of the matrices with determinant 1, and the *orthogonal group* $\text{O}(d, \mathbb{R})$, formed by the orthogonal matrices.

We call *left-translation* and *right-translation* associated with an element g of the group G , respectively, the maps

$$L_g : G \rightarrow G, L_g(h) = gh \quad \text{and} \quad R_g : G \rightarrow G, R_g(h) = hg.$$

An *endomorphism* of G is a continuous map $\phi : G \rightarrow G$ that preserves the group operation, that is, such that $\phi(gh) = \phi(g)\phi(h)$ for every $g, h \in G$. When ϕ is an invertible endomorphism, that is, a bijection whose inverse is also an endomorphism, we call it an *automorphism*.

Example 6.3.3. Let $A \in \text{GL}(d, \mathbb{Z})$; in other words, A is an invertible matrix of dimension d with integer coefficients. Then, as we have seen in Section 4.2.5, A induces an endomorphism $f_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$. It can be shown that every endomorphism of the torus \mathbb{T}^d is of this form.

A topological group is *locally compact* if every $g \in G$ has some compact neighborhood. For example, every Lie group is locally compact. On the other hand, the additive group of rational numbers, with the topology inherited from the real line, is not locally compact.

The following theorem is the starting point of the ergodic theory of locally compact groups:

Theorem 6.3.4 (Haar). *Let G be a locally compact topological group. Then:*

- (i) *There exists some Borel measure μ_G on G that is invariant under all left-translations, finite on compact sets and positive on open sets;*
- (ii) *If η is a measure invariant under all left-translations and finite on compact sets then $\eta = c\mu_G$ for some $c > 0$.*
- (iii) *$\mu_G(G) < \infty$ if and only if G is compact.*

We are going to sketch the proof of parts (i) and (ii) in the special case when G is a Lie group. It will be apparent that in this case μ_G is a volume measure on G . The proof of part (iii), for any topological group, is proposed in Exercise 6.3.4.

Starting with part (i), let e be the unit element and $d \geq 1$ be the dimension of the Lie group. Consider any inner product \cdot in the tangent space T_eG . For each $g \in G$, represent by $\mathcal{L}_g : T_eG \rightarrow T_gG$ the derivative of the left-translation L_g at the point e . Next, consider the inner product defined in T_gG in the following way:

$$u \cdot v = \mathcal{L}_g^{-1}(u) \cdot \mathcal{L}_g^{-1}(v) \quad \text{for every } u, v \in T_gG.$$

It is clear that this inner product depends differentiably on g . Therefore, it defines a Riemannian metric in G . It is also clear from the construction that this metric is invariant under left-translations: noting that $\mathcal{L}_{hg} = DL_h(g)\mathcal{L}_g$, we see that

$$\begin{aligned} DL_h(g)(u) \cdot DL_h(g)(v) &= \mathcal{L}_{hg}^{-1}DL_h(g)(u) \cdot \mathcal{L}_{hg}^{-1}DL_h(g)(v) \\ &= \mathcal{L}_g^{-1}(u) \cdot \mathcal{L}_g^{-1}(v) = u \cdot v \end{aligned}$$

for any $g, h \in G$ and $u, v \in T_gG$. Let μ_G be the volume measure induced by this Riemannian metric. This measure may be characterized in the following way. Given any $x = (x_1, \dots, x_d)$ in G , consider

$$\rho(x) = \det \begin{pmatrix} g_{1,1}(x) & \cdots & g_{1,d}(x) \\ \vdots & \ddots & \vdots \\ g_{d,1}(x) & \cdots & g_{d,d}(x) \end{pmatrix} \quad \text{where } g_{ij} = \frac{\partial}{\partial x_i} \cdot \frac{\partial}{\partial x_j}.$$

Then $\mu_G(B) = \int_B |\rho(x)| dx_1 \cdots dx_d$, for any measurable set B contained in the domain of the local coordinates. Noting that the function ρ is continuous and non-zero, for every local chart, it follows that μ_G is positive on open sets and finite on compact sets. Moreover, since the Riemannian metric is invariant under left-translations, the measure μ_G is also invariant under left-translations.

Now we move on to discussing part (ii) of Theorem 6.3.4. Let ν any measure as in the statement. Denote by $B(g, r)$ the open ball of center g and radius r , relative to the distance associated with the Riemannian metric. In other words, $B(g, r)$ is the set of all points in G that may be connected to g by some curve of length less than r . Fix $\rho > 0$ such that $\nu(B(e, \rho))$ is finite (such a ρ does exist

because G is locally compact and ν is finite on compact sets). We claim that

$$\limsup_{r \rightarrow 0} \frac{\nu(B(g, r))}{\mu_G(B(g, r))} \leq \frac{\nu(B(e, \rho))}{\mu_G(B(e, \rho))} \tag{6.3.3}$$

for every $g \in G$. This may be seen as follows.

First, the limit on the left-hand side of the inequality does not depend on g , because both measures are assumed to be invariant under left-translations. Therefore, it is enough to consider the case $g = e$. Let $(r_n)_n$ be any sequence converging to zero and such that:

$$\lim_n \frac{\nu(B(e, r_n))}{\mu_G(B(e, r_n))} = \limsup_{r \rightarrow 0} \frac{\nu(B(e, r))}{\mu_G(B(e, r))}. \tag{6.3.4}$$

By the Vitali lemma (Theorem A.2.16), we may find $(g_j)_j$ in $B(e, \rho)$ and $(n_j)_j$ in \mathbb{N} such that

1. the balls $B(g_j, r_{n_j})$ are contained in $B(e, \rho)$ and they are pairwise disjoint;
2. the union of these balls has full μ_G -measure in $B(e, \rho)$.

Moreover, given any $a \in \mathbb{R}$ smaller than the limit in (6.3.4), we may suppose that the integers n_j are sufficiently large that $\nu(B(g_j, r_{n_j})) \geq a\mu_G(B(g_j, r_{n_j}))$ for every j . It follows that

$$\nu(B(e, \rho)) \geq \sum_j \nu(B(g_j, r_{n_j})) \geq \sum_j a\mu_G(B(g_j, r_{n_j})) = a\mu_G(B(e, \rho)).$$

Since a may be taken arbitrarily close to (6.3.4), this proves the claim (6.3.3).

Next, we claim that ν is absolutely continuous with respect to μ_G . Indeed, let b be any number larger than the quotient on the right-hand side of (6.3.3). Given any measurable set $B \subset G$ with $\mu_G(B) = 0$, and given any $\varepsilon > 0$, let $\{B(g_j, r_j) : j\}$ be a cover of B by balls of small radii, such that $\nu(B(g_j, r_j)) \leq b\mu_G(B(g_j, r_j))$ and $\sum_j \mu_G(B(g_j, r_j)) \leq \varepsilon$. Then,

$$\nu(B) \leq \sum_j \nu(B(g_j, r_j)) \leq b \sum_j \mu_G(B(g_j, r_j)) \leq b\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\nu(B) = 0$. Therefore, $\nu \ll \mu_G$, as claimed. Now, by the Lebesgue derivation theorem (Theorem A.2.15),

$$\frac{d\nu}{d\mu_G}(g) = \lim_{r \rightarrow 0} \frac{1}{\mu_G(B(g, r))} \int_{B(g, r)} \frac{d\nu}{d\mu_G} d\mu_G = \lim_{r \rightarrow 0} \frac{\nu(B(g, r))}{\mu_G(B(g, r))}$$

for μ -almost every $g \in G$. The limit on the left-hand side does not depend on g and, by (6.3.3), it is finite. Let $c \in \mathbb{R}$ be that limit. Then $\nu = c\mu_G$, as stated in part (ii) of Theorem 6.3.4.

In the case when the group G is compact, it follows from Theorem 6.3.4 that there exists a unique probability measure that is invariant under left-translations, positive on open sets and finite on compact sets. This probability measure μ_G is called the *Haar measure* of the group. For example,

the normalized Lebesgue measure is the Haar measure on the torus \mathbb{T}^d . See also Exercises 6.3.5 and 6.3.6. The Haar measure features some additional properties:

Corollary 6.3.5. *Assume that G is compact. Then the Haar measure μ_G is invariant under right-translations and under every surjective endomorphism of G .*

Proof. Given any $g \in G$, consider the probability measure $(R_g)_*\mu_G$. Observe that $L_h \circ R_g = R_g \circ L_h$ for every $h \in G$. Hence,

$$(L_h)_*(R_g)_*\mu_G = (R_g)_*(L_h)_*\mu_G = (R_g)_*\mu_G.$$

In other words, $(R_g)_*\mu_G$ is invariant under every left-translation. By uniqueness, it follows that $(R_g)_*\mu_G = \mu_G$ for every $g \in G$, as claimed.

Given any surjective endomorphism $\phi : G \rightarrow G$, consider the probability $\phi_*\mu_G$. Given any $h \in G$, choose some $g \in \phi^{-1}(h)$. Observe that $L_h \circ \phi = \phi \circ L_g$. Hence,

$$(L_h)_*\phi_*\mu_G = \phi_*(L_g)_*\mu_G = \phi_*\mu_G.$$

In other words, $\phi_*\mu_G$ is invariant under every left-translation. By uniqueness, it follows that $\phi_*\mu_G = \mu_G$, as claimed.

More generally, when we do not assume G to be compact, the argument in Corollary 6.3.5 shows that for every $g \in G$ there exists $\lambda(g) > 0$ such that

$$(L_g)_*\mu_G = \lambda(g)\mu_G.$$

The map $G \rightarrow (0, \infty)$, $g \mapsto \lambda(g)$ is a group homomorphism.

6.3.3 Translations on compact metrizable groups

We call a distance d in a topological group G *left-invariant* if it is invariant under every left-translation: $d(L_h(g_1), L_h(g_2)) = d(g_1, g_2)$ for every $g_1, g_2, h \in G$. Analogously, we call a distance *right-invariant* if it is invariant under every right-translation. In this section we always take the group G to be compact and metrizable. We start by observing that it is always possible to choose the distance in G in such a way that it is invariant under all the translations:

Lemma 6.3.6. *If G is a compact metrizable topological group then there exists some distance compatible with the topology of G that is both left-invariant and right-invariant.*

Proof. Let $(U_n)_n$ be a countable basis of neighborhoods of the unit element e of G . By Lemma A.3.4, for every n there exists a continuous function $\varphi_n : G \rightarrow [0, 1]$ such that $\varphi_n(e) = 0$ and $\varphi_n(z) = 1$ for every $z \in G \setminus U_n$. Define

$$\varphi : G \rightarrow [0, 1], \quad \varphi(z) = \sum_{n=1}^{\infty} 2^{-n} \varphi_n(z).$$

Then, φ is continuous and $\varphi(e) = 0 < \varphi(z)$ for every $z \neq e$. Now define

$$d(x,y) = \sup\{|\varphi(gxh) - \varphi(gyh)| : (g,h) \in G^2\} \tag{6.3.5}$$

for every $x,y \in G$. The supremum is finite, since we take G to be compact. It is easy to see that d is a distance in G . Indeed, note that $d(x,y) = 0$ means that $\varphi(gxh) = \varphi(gyh)$ for every $g,h \in G$. In particular, taking $g = e$ and $h = y^{-1}$, we get that $\varphi(xy^{-1}) = \varphi(e)$. By the construction of φ , this implies that $x = y$. All the other axioms of the notion of distance follow directly from the definition of d . It is also clear from the definition that d is invariant under both left-translations and right-translations.

We are left to prove that the distance d is compatible with the topology of the group G . It is easy to check that, given any neighborhood V of a point $x \in G$, there exists $\delta > 0$ such that $B(x,\delta) \subset V$. Indeed, since $U = x^{-1}V$ is a neighborhood of $e \in G$, the properties of φ ensure that there exists $\delta > 0$ such that $\varphi(z) \leq 1 - \delta$ for every $z \notin U$. Then, $y \notin V$ implies that $\varphi(x^{-1}y) \leq 1 - \delta$ or, in other words, that $|\varphi(e) - \varphi(x^{-1}y)| \geq \delta$. Taking $g = x^{-1}$ and $h = e$ in the definition (6.3.5), we see that this last inequality implies that $d(x,y) \geq \delta$, that is, $y \notin B(x,\delta)$. Now let us check the converse: given $x \in G$ and $\delta > 0$, there exists some neighborhood V of x contained in $B(x,\delta)$. By continuity, for every pair $(g,h) \in G^2$ there exists an open neighborhood $U \times V \times W$ of (g,x,h) in G^3 such that

$$|\varphi(gxh) - \varphi(g'x'h')| \leq \delta/2 \quad \text{for every } (g',x',h') \in U \times V \times W. \tag{6.3.6}$$

The sets $U \times W$ obtained in this way, with x fixed and g,h variable, form an open cover of G^2 . Let $U_i \times W_i, i = 1, \dots, k$ be a finite subcover and $V_i, i = 1, \dots, k$ be the corresponding neighborhoods of x . Take $V = \bigcap_{i=1}^k V_i$ and consider any $y \in V$. Given any $(g,h) \in G^2$, the condition (6.3.6) implies that $|\varphi(gxh) - \varphi(gyh)| \leq \delta/2$. It follows that $d(x,y) \leq \delta/2$ and, consequently, $y \in B(x,\delta)$.

Example 6.3.7. Given a matrix $A \in \text{GL}(d, \mathbb{R})$, denote by $\|A\|$ its operator norm, that is, $\|A\| = \sup\{\|Av\| : \|v\| = 1\}$. Observe that $\|OA\| = \|A\| = \|AO\|$ for every O in the orthogonal group $O(d, \mathbb{R})$. Define

$$d(A,B) = \log(1 + \|A^{-1}B - \text{id}\| + \|B^{-1}A - \text{id}\|).$$

Then d is a distance in $\text{GL}(d, \mathbb{R})$, invariant under left-translations:

$$d(CA, CB) = \log(1 + \|A^{-1}C^{-1}CB - \text{id}\| + \|B^{-1}C^{-1}CA - \text{id}\|) = d(A,B)$$

for every $C \in \text{GL}(d, \mathbb{R})$. This distance is not invariant under right-translations in $\text{GL}(d, \mathbb{R})$ (Exercise 6.3.3). However, it is right-invariant (and left-invariant)

restricted to the orthogonal group $O(d, \mathbb{R})$: for every $O \in O(d, \mathbb{R})$,

$$\begin{aligned} d(AO, CO) &= \log(1 + \|O^{-1}A^{-1}BO - \text{id}\| + \|O^{-1}B^{-1}AO - \text{id}\|) \\ &= \log(1 + \|O^{-1}(A^{-1}B - \text{id})O\| + \|O^{-1}(B^{-1}A - \text{id})O\|) \\ &= d(A, B). \end{aligned}$$

Theorem 6.3.8. *Let G be a compact metrizable topological group and let $g \in G$. The following conditions are equivalent:*

- (i) L_g is uniquely ergodic;
- (ii) L_g is ergodic with respect to μ_G ;
- (iii) the subgroup $\{g^n : n \in \mathbb{Z}\}$ generated by g is dense in G .

Proof. It is clear that (i) implies (ii). To prove that (ii) implies (iii), consider the invariant distance d given by Lemma 6.3.6. Let H be the closure of $\{g^n : n \in \mathbb{Z}\}$ and consider the continuous function

$$\varphi(x) = \min\{d(x, y); y \in H\}.$$

Observe that this function is invariant under L_g : using that $gH = H$, we get that

$$\begin{aligned} \varphi(x) &= \min\{d(x, y) : y \in H\} = \min\{d(gx, gy) : y \in H\} \\ &= \min\{d(gx, z) : z \in H\} = \varphi(gx) \quad \text{for every } x \in G. \end{aligned}$$

Since H is closed, $\varphi(x) = 0$ if and only if $x \in H$. If $H \neq G$ then $\mu_G(G \setminus H) > 0$, as the Haar measure is positive on open sets. In that case, the function φ is not constant at μ_G -almost every point, which implies that L_g cannot be ergodic with respect to μ_G .

Finally, to prove that (iii) implies (i), let us show that if μ is a probability measure invariant under L_g then $\mu = \mu_G$. For that, it suffices to check that μ is invariant under every left-translation in G . Fix $h \in G$. Since μ is invariant under L_g ,

$$\int \varphi(x) d\mu(x) = \int \varphi(g^n x) d\mu(x)$$

for every $n \in \mathbb{N}$ and every continuous function $\varphi : G \rightarrow \mathbb{R}$. On the other hand, the hypothesis ensures that there exists a sequence of natural numbers $n_j \rightarrow \infty$ such that $g^{n_j} \rightarrow h$. Given any (uniformly) continuous function $\varphi : G \rightarrow \mathbb{R}$ and any $\varepsilon > 0$, fix $\delta > 0$ such that $|\varphi(x) - \varphi(y)| < \varepsilon$ whenever $d(x, y) < \delta$. If j is sufficiently large,

$$d(g^{n_j}x, hx) = d(g^{n_j}, h) < \delta \quad \text{for every } x \in G.$$

Hence, $|\varphi(g^{n_j}x) - \varphi(hx)| < \varepsilon$ for every x and, consequently,

$$\left| \int (\varphi(x) - \varphi(hx)) d\mu \right| = \left| \int (\varphi(g^{n_j}x) - \varphi(hx)) d\mu \right| < \varepsilon.$$

Since ε is arbitrary, it follows that $\int \varphi d\mu = \int \varphi \circ L_h d\mu$ for every continuous function φ and every $h \in G$. This implies that μ is invariant under L_h for every $h \in G$, as claimed.

6.3.4 Odometers

Odometers, or *adding machines*, are mathematical models for the mechanisms that register the distance (number of kilometers) travelled by a car or the amount of electricity (number of energy units) consumed in a house. They come with a dynamic, which consists in advancing the counter by one unit each time. The main difference with respect to real-life odometers is that our idealized counters allow for an infinite number of digits.

Fix any number basis $d \geq 2$, for example $d = 10$, and consider the set $X = \{0, 1, \dots, d - 1\}$, endowed with the discrete topology. Let $M = X^{\mathbb{N}}$ be the set of all sequences $\alpha = (\alpha_n)_n$ with values in X , endowed with the product topology. This topology is metrizable: it is compatible, for instance, with the distance defined in M by

$$d(\alpha, \alpha') = 2^{-N(\alpha, \alpha')} \quad \text{where} \quad N(\alpha, \alpha') = \min\{j \geq 0 : \alpha_j \neq \alpha'_j\}. \quad (6.3.7)$$

Observe also that M is compact, being the product of compact spaces (theorem of Tychonoff).

Let us introduce in M the following operation of “sum with transport”: given $\alpha = (\alpha_n)_n$ and $\beta = (\beta_n)_n$ in M , define $\alpha + \beta = (\gamma_n)_n$ as follows. First,

- if $\alpha_0 + \beta_0 < d$ then $\gamma_0 = \alpha_0 + \beta_0$ and $\delta_1 = 0$;
- if $\alpha_0 + \beta_0 \geq d$ then $\gamma_0 = \alpha_0 + \beta_0 - d$ and $\delta_1 = 1$.

Next, for every $n \geq 1$,

- if $\alpha_n + \beta_n + \delta_n < d$ then $\gamma_n = \alpha_n + \beta_n + \delta_n$ and $\delta_{n+1} = 0$;
- if $\alpha_n + \beta_n + \delta_n \geq d$ then $\gamma_n = \alpha_n + \beta_n + \delta_n - d$ and $\delta_{n+1} = 1$.

The auxiliary sequence $(\delta_n)_n$ corresponds precisely to the transports. The map $+ : M \times M \rightarrow M$ defined in this way turns M into an abelian topological group and the distance (6.3.7) is invariant under all the translations (Exercise 6.3.8).

Now consider the “translation by 1” $f : M \rightarrow M$ defined by

$$f((\alpha_n)_n) = (\alpha_n)_n + (1, 0, \dots, 0, \dots) = (0, \dots, 0, \alpha_k + 1, \alpha_{k+1}, \dots, \alpha_n, \dots)$$

where $k \geq 0$ is the smallest value of n such that $\alpha_n < d - 1$; if there exists no such k , that is, if $(\alpha_n)_n$ is the constant sequence equal to $d - 1$, then the image $f((\alpha_n)_n)$ is the constant sequence equal to 0. We leave it to the reader to check that this transformation $f : M \rightarrow M$ is uniquely ergodic (Exercise 6.3.9).

It is possible to generalize this construction somewhat, in the following way. Take $M = \prod_{n=0}^{\infty} \{0, 1, \dots, d_n - 1\}$, where $(d_n)_n$ is any sequence of integer numbers larger than 1. Just as in the previous particular case, this set has the

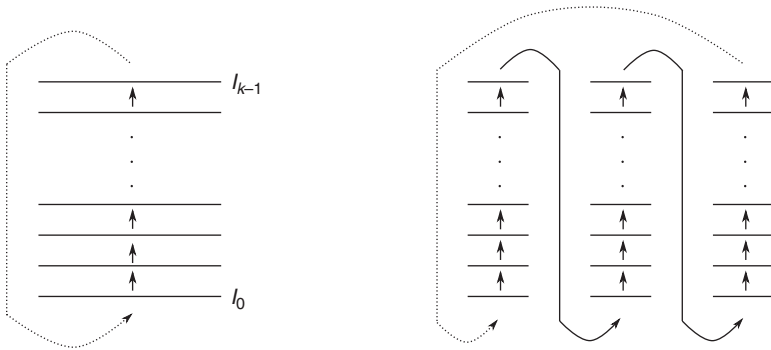


Figure 6.1. Example of the piling method

structure of a metrizable compact abelian group and the “translation by 1” is uniquely ergodic.

Example 6.3.9. A (simple) pile in an interval¹ I is an ordered family \mathcal{S} of pairwise disjoint subintervals I_0, \dots, I_{k-1} with the same length and whose union is I . Write $I_k = I_0$. We associate with \mathcal{S} the transformation $f : I \rightarrow I$ whose restriction to each I_j is the translation mapping I_j to I_{j+1} . Graphically, we represent the subintervals “piled up” on top of each other in order: from the bottom I_0 to the top I_{k-1} . Then f is nothing but the translation “upwards”, except at the top of the pile. See the left-hand side of Figure 6.1.

Let us consider a sequence $(\mathcal{S}_n)_n$ of piles in the same interval I , constructed as follows. Fix any integer number $d \geq 2$. Take $\mathcal{S}_0 = \{I\}$. For each $n \geq 1$, take as \mathcal{S}_n the pile obtained by dividing \mathcal{S}_{n-1} into d columns, all with the same width, and piling them up on top of each other. This procedure is described on the right-hand side of Figure 6.1 for $d = 3$. Let $f_n : I \rightarrow I$ be the transformation associated with each \mathcal{S}_n . We leave it to the reader to show (Exercise 6.3.10) that the sequence $(f_n)_n$ converges at every point to a transformation $f : I \rightarrow I$ that preserves the Lebesgue measure. Moreover, this transformation f is uniquely ergodic.

This is only one of the simplest applications of the so-called *piling method*, which is a very effective tool to produce examples with interesting properties. The reader may find a detailed discussion of this method in Section 6 of Friedman [Fri69]. Another application, a bit more elaborate, will be given in Example 8.2.3.

Example 6.3.10 (Substitutions). We are going to mention briefly a construction of a combinatorial nature that generalizes the definition of odometer and provides several other interesting examples of minimal and even uniquely ergodic systems. For more information, including about the relations between

¹ For definiteness, take all intervals to be closed on the left and open on the right.

such systems and the odometer, we recommend the book of Queffélec [Que87] and the paper of Ferenczi, Fisher and Talet [FFT09].

We call a *substitution* in a finite alphabet \mathcal{A} any map associating with each letter $\alpha \in \mathcal{A}$ a word $s(\alpha)$ formed by a finite number of letters of \mathcal{A} . A few examples, for $\mathcal{A} = \{0, 1\}$: *Thue–Morse substitution* $s(0) = 01$ and $s(1) = 10$; *Fibonacci substitution* $s(0) = 01$ and $s(1) = 0$; *Feigenbaum substitution* $s(0) = 11$ and $s(1) = 10$; *Cantor substitution* $s(0) = 010$ and $s(1) = 111$; and *Chacon substitution* $s(0) = 0010$ and $s(1) = 1$. We may iterate a substitution by defining $s^1(\alpha) = s(\alpha)$ and

$$s^{k+1}(\alpha) = s(\alpha_1) \cdots s(\alpha_n) \quad \text{if} \quad s^k(\alpha) = \alpha_1 \cdots \alpha_n.$$

We call a substitution s *primitive* (or *aperiodic*) if there exists $k \geq 1$ such that for any $\alpha, \beta \in \mathcal{A}$ the word $s^k(\alpha)$ contains the letter β .

Let \mathcal{A} be endowed with the discrete topology and $\Sigma = \mathcal{A}^{\mathbb{N}}$ be the space of all sequences in \mathcal{A} , endowed with the product topology. Denote by $S : \Sigma \rightarrow \Sigma$ the map induced in that space by a given substitution s : the image of each $(a_0, \dots, a_n, \dots) \in \Sigma$ is the sequence of the letters that constitute the word obtained when one concatenates the finite words $s(a_0), \dots, s(a_n), \dots$. Suppose that there exists some letter $\alpha_0 \in \mathcal{A}$ such that the word $s(\alpha_0)$ has length larger than 1 and starts with the letter α_0 . That is the case for all the examples listed above. Then (Exercise 6.3.11), S admits a unique fixed point $x = (x_n)_n$ with $x_0 = \alpha_0$.

Consider the restriction $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ of the shift map $\sigma : \Sigma \rightarrow \Sigma$ to the closure $\mathcal{X} \subset \Sigma$ of the orbit $\{\sigma^n(x) : n \geq 0\}$ of the point x . If the substitution s is primitive then $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ is minimal and uniquely ergodic (see Section 5 in [Que87]). That holds, for instance, for the Thue–Morse, Fibonacci and Feigenbaum substitutions.

6.3.5 Exercises

- 6.3.1. Let G be a manifold and \cdot be a group operation in G such that the map $(g, h) \mapsto g \cdot h$ is of class C^1 . Show that $g \mapsto g^{-1}$ is also of class C^1 .
- 6.3.2. Let G be a compact topological space such that every point admits a countable basis of neighborhoods and let \cdot be a group operation in G such that the map $(g, h) \mapsto g \cdot h$ is continuous. Show that $g \mapsto g^{-1}$ is also continuous.
- 6.3.3. Show that the distance d in Example 6.3.7 is not right-invariant.
- 6.3.4. Prove part (iii) of Theorem 6.3.4: a locally compact group G is compact if and only if its Haar measure is finite.
- 6.3.5. Identify $\text{GL}(1, \mathbb{R})$ with the multiplicative group $\mathbb{R} \setminus \{0\}$. Check that the measure μ defined on $\text{GL}(1, \mathbb{R})$ by

$$\int_{\text{GL}(1, \mathbb{R})} \varphi \, d\mu = \int_{\mathbb{R} \setminus \{0\}} \frac{\varphi(x)}{|x|} \, dx$$

is both left-invariant and right-invariant. Find a measure invariant under all the translations of $\text{GL}(1, \mathbb{C})$.

6.3.6. Identify $GL(2, \mathbb{R})$ with $\{(a_{11}, a_{12}, a_{21}, a_{22}) \in \mathbb{R}^4 : a_{11}a_{22} - a_{12}a_{21} \neq 0\}$, in such a way that $\det(a_{11}, a_{12}, a_{21}, a_{22}) = a_{11}a_{22} - a_{12}a_{21}$. Show that the measure μ defined by

$$\int_{GL(2, \mathbb{R})} \varphi d\mu = \int \frac{\varphi(x_{11}, x_{12}, x_{21}, x_{22})}{|\det(x_{11}, x_{12}, x_{21}, x_{22})|^2} dx_{11} dx_{12} dx_{21} dx_{22}$$

is both left-invariant and right-invariant. Find a measure invariant under all the translations of $GL(2, \mathbb{C})$.

6.3.7. Let G be a compact metrizable group and let $g \in G$. Check that the following conditions are equivalent:

- (1) L_g is uniquely ergodic;
- (2) L_g is transitive: there is $x \in G$ such that $\{g^n x : n \in \mathbb{Z}\}$ is dense in G ;
- (3) L_g is minimal: $\{g^n y : n \in \mathbb{Z}\}$ is dense in G for every $y \in G$.

6.3.8. Show that the operation $+$: $M \times M \rightarrow M$ defined in Section 6.3.4 is continuous and endows M with the structure of an abelian group. Moreover, every translation in this group preserves the distance defined in (6.3.7).

6.3.9. Let $f : M \rightarrow M$ be an odometer, as defined in Section 6.3.4, with $d = 10$. Given b_0, \dots, b_{k-1} in $\{0, \dots, 9\}$, denote by $[b_0, \dots, b_{k-1}]$ the set of all sequences $\beta \in M$ with $\beta_0 = b_0, \dots, \beta_{k-1} = b_{k-1}$. Show that

$$\lim_n \frac{1}{n} \#\{0 \leq j < n : f^j(x) \in [b_0, \dots, b_{k-1}]\} = \frac{1}{10^k}$$

for every $x \in M$. Moreover, this limit is uniform. Conclude that f admits a unique invariant probability measure and calculate that measure explicitly.

6.3.10. Check the claims in Example 6.3.9.

6.3.11. Prove that if s is a substitution in a finite alphabet \mathcal{A} and $\alpha \in \mathcal{A}$ is such that $s(\alpha)$ has length larger than 1 and starts with the letter α , then the transformation $S : \Sigma \rightarrow \Sigma$ defined in Example 6.3.10 admits a unique fixed point that starts with the letter $\alpha \in \mathcal{A}$.

6.4 Theorem of Weyl

In this section we use ideas that were discussed previously to prove a beautiful theorem of Hermann Weyl [Wey16] about the distribution of polynomial sequences.

Consider any polynomial function $P : \mathbb{R} \rightarrow \mathbb{R}$ with real coefficients and degree $d \geq 1$:

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_dx^d.$$

Composing P with the canonical projection $\mathbb{R} \rightarrow S^1$, we obtain a polynomial function $P_* : \mathbb{R} \rightarrow S^1$ with values on the circle $S^1 = \mathbb{R}/\mathbb{Z}$. Define

$$z_n = P_*(n), \quad \text{for every } n \geq 1.$$

We may think of z_n as the fractional part of the real number $P(n)$. We want to understand how the sequence $(z_n)_n$ is distributed on the circle.

Definition 6.4.1. We say that a sequence $(x_n)_n$ in S^1 is *equidistributed* if, for any continuous function $\varphi : S^1 \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \varphi(x_j) = \int \varphi(x) dx.$$

According to Exercise 6.4.1, this is equivalent to saying that, for every segment $I \subset S^1$, the fraction of terms of the sequence that are in I is equal to the corresponding length $m(I)$.

Theorem 6.4.2 (Weyl). *If at least one of the coefficients a_1, a_2, \dots, a_d is irrational then the sequence $z_n = P_*(n), n \in \mathbb{N}$ is equidistributed.*

In order to develop some intuition about this theorem, let us start by considering the special case $d = 1$. In this case the polynomial function reduces to $P(x) = a_0 + a_1x$. Let us consider the transformation

$$f : S^1 \rightarrow S^1, \quad f(\theta) = \theta + a_1.$$

By assumption, the coefficient a_1 is irrational. Therefore, as we have seen in Section 6.3.1, this transformation admits a unique invariant probability measure, which is the Lebesgue measure m . Consequently, given any continuous function $\varphi : S^1 \rightarrow \mathbb{R}$ and any point $\theta \in S^1$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \varphi(f^j(\theta)) = \int \varphi dm.$$

Take $\theta = a_0$. Then $f^j(\theta) = a_0 + a_1j = z_j$. Hence, the previous relation yields

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \varphi(z_j) = \int \varphi dm.$$

This is precisely what it means to say that z_j is equidistributed.

6.4.1 Ergodicity

Now we extend the previous arguments to any degree $d \geq 1$. Consider the transformation $f : \mathbb{T}^d \rightarrow \mathbb{T}^d$ defined on the d -dimensional torus \mathbb{T}^d by the following expression:

$$f(\theta_1, \theta_2, \dots, \theta_d) = (\theta_1 + \alpha, \theta_2 + \theta_1, \dots, \theta_d + \theta_{d-1}), \tag{6.4.1}$$

where α is an irrational number to be chosen later. Note that f is invertible: the inverse is given by

$$f^{-1}(\theta_1, \theta_2, \dots, \theta_d) = (\theta_1 - \alpha, \theta_2 - \theta_1 + \alpha, \dots, \theta_d - \theta_{d-1} + \dots + (-1)^{d-1}\theta_1 + (-1)^d\alpha).$$

Note also that the derivative of f at each point is given by the matrix

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{pmatrix},$$

whose determinant is 1. This ensures that f preserves the Lebesgue measure on the torus (recall Lemma 1.3.5).

Proposition 6.4.3. *The Lebesgue measure on \mathbb{T}^d is ergodic for f .*

Proof. We are going to use a variation of the Fourier series expansion argument in Proposition 4.2.1. Let $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$ be any function in $L^2(m)$. Write

$$\varphi(\theta) = \sum_{n \in \mathbb{Z}^d} a_n e^{2\pi i n \cdot \theta}$$

with $\theta = (\theta_1, \dots, \theta_d)$, $n = (n_1, \dots, n_d)$ and $n \cdot \theta = n_1\theta_1 + \dots + n_d\theta_d$. The L^2 -norm of φ is given by

$$\sum_{n \in \mathbb{Z}^d} |a_n|^2 = \int |\varphi(\theta)|^2 d\theta_1 \cdots d\theta_d < \infty. \tag{6.4.2}$$

Observe that

$$\begin{aligned} \varphi(f(\theta)) &= \sum_{n \in \mathbb{Z}^d} a_n e^{2\pi i(n_1(\theta_1 + \alpha) + n_2(\theta_2 + \theta_1) + \dots + n_d(\theta_d + \theta_{d-1}))} \\ &= \sum_{n \in \mathbb{Z}^d} a_n e^{2\pi i n_1 \alpha} e^{2\pi i L(n) \cdot \theta}, \end{aligned}$$

where $L(n) = (n_1 + n_2, n_2 + n_3, \dots, n_{d-1} + n_d, n_d)$. Suppose that the function φ is invariant, that is, $\varphi \circ f = \varphi$ at almost every point. Then

$$a_n e^{2\pi i n_1 \alpha} = a_{L(n)} \quad \text{for every } n \in \mathbb{Z}^d. \tag{6.4.3}$$

This implies that a_n and $a_{L(n)}$ have the same absolute value. On the other hand, the integrability relation (6.4.2) implies that there exists at most a finite number of terms with any given absolute value different from zero. It follows that $a_n = 0$ for every $n \in \mathbb{Z}^d$ whose orbit $L^j(n)$, $j \in \mathbb{Z}$ is infinite. Observing the expression of L , we deduce that $a_n = 0$ except, possibly, if $n_2 = \dots = n_d = 0$. For the remaining values of n , that is, for every $n = (n_1, 0, \dots, 0)$, one has that $L(n) = n$ and, thus, the relation (6.4.3) becomes

$$a_n = a_n e^{2\pi i n_1 \alpha}.$$

Since α is irrational, the last factor is different from 1 whenever n_1 is non-zero. Therefore, this relation implies that $a_n = 0$ also for $n = (n_1, 0, \dots, 0)$ with $n_1 \neq 0$. In this way, we have shown that if φ is an invariant function then all the terms

in its Fourier series vanish except, possibly, the constant term. This means that φ is constant at almost every point, which proves that the Lebesgue measure is ergodic for f .

6.4.2 Unique ergodicity

The last step in the proof of Theorem 6.4.2 is the following result:

Proposition 6.4.4. *The transformation f is uniquely ergodic: the Lebesgue measure on the torus is the unique invariant probability measure.*

Proof. The proof is by induction on the degree d of the polynomial P . The case of degree 1 was treated previously. Therefore, we only need to explain how the case of degree d may be deduced from the case of degree $d - 1$. For that, we write $\mathbb{T}^d = \mathbb{T}^{d-1} \times S^1$ and

$$f : \mathbb{T}^{d-1} \times S^1 \rightarrow \mathbb{T}^{d-1} \times S^1, \quad f(\theta_0, \eta) = (f_0(\theta_0), \eta + \theta_{d-1}), \tag{6.4.4}$$

where $\theta_0 = (\theta_1, \dots, \theta_{d-1})$ and $f_0(\theta_0) = (\theta_1 + \alpha, \theta_2 + \theta_1, \dots, \theta_{d-1} + \theta_{d-2})$. By induction, the transformation

$$f_0 : \mathbb{T}^{d-1} \rightarrow \mathbb{T}^{d-1}$$

is uniquely ergodic. Let us denote by $\pi : \mathbb{T}^d \rightarrow \mathbb{T}^{d-1}$ the projection $\pi(\theta) = \theta_0$.

Lemma 6.4.5. *For any probability measure μ invariant under f , the projection $\pi_*\mu$ coincides with the Lebesgue measure m_0 on \mathbb{T}^{d-1} .*

Proof. Given any measurable set $E \subset \mathbb{T}^{d-1}$,

$$(\pi_*\mu)(f_0^{-1}(E)) = \mu(\pi^{-1}f_0^{-1}(E)).$$

Using that $\pi \circ f = f_0 \circ \pi$ and the fact that μ is f -invariant, we get that the expression on the right-hand side is equal to

$$\mu(f^{-1}\pi^{-1}(E)) = \mu(\pi^{-1}(E)) = (\pi_*\mu)(E).$$

Therefore, $(\pi_*\mu)(f_0^{-1}(E)) = (\pi_*\mu)(E)$ for every measurable subset E , that is, $\pi_*\mu$ is invariant under f_0 . It is clear that $\pi_*\mu$ is a probability measure. Since f_0 is uniquely ergodic, it follows that $\pi_*\mu$ coincides with the Lebesgue measure m_0 on \mathbb{T}^{d-1} .

Now suppose that μ , besides being invariant, is also ergodic for f . By Theorem 3.2.6, and by ergodicity, the set $G(\mu) \subset M$ of all points $\theta \in \mathbb{T}^d$ such that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\theta)) = \int \varphi d\mu \quad \text{for any continuous function } \varphi : \mathbb{T}^d \rightarrow \mathbb{R} \tag{6.4.5}$$

has full measure. Let $G_0(\mu)$ be the set of all $\theta_0 \in \mathbb{T}^{d-1}$ such that $G(\mu)$ intersects $\{\theta_0\} \times S^1$. In other words, $G_0(\mu) = \pi(G(\mu))$. It is clear that $\pi^{-1}(G_0(\mu))$ contains $G(\mu)$ and, thus, has full measure. Hence, using Lemma 6.4.5,

$$m_0(G_0(\mu)) = \mu(\pi^{-1}(G_0(\mu))) = 1. \tag{6.4.6}$$

For the same reasons, this relation remains valid for the Lebesgue measure:

$$m(G_0(m)) = m(\pi^{-1}(G_0(m))) = 1. \tag{6.4.7}$$

The identities (6.4.6) and (6.4.7) imply that the intersection between $G_0(\mu)$ and $G_0(m)$ has full measure for m_0 . So, in particular, these two sets cannot be disjoint. Let θ_0 be any point in the intersection. By definition, $G(\mu)$ intersects $\{\theta_0\} \times S^1$. But the next result asserts that $G(m)$ contains $\{\theta_0\} \times S^1$:

Lemma 6.4.6. *If $\theta_0 \in G_0(m)$ then $\{\theta_0\} \times S^1$ is contained in $G(m)$.*

Proof. The crucial observation is that the measure m is invariant under every transformation of the form

$$R_\beta : \mathbb{T}^{d-1} \times S^1 \rightarrow \mathbb{T}^{d-1} \times S^1, \quad (\zeta, \eta) \mapsto (\zeta, \eta + \beta).$$

The hypothesis $\theta_0 \in G_0(m)$ means that there exists some $\eta \in S^1$ such that $(\theta_0, \eta) \in G(m)$, that is,

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(F^j(\theta_0, \eta)) = \int \varphi \, dm$$

for every continuous function $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$. Any other point of $\{\theta_0\} \times S^1$ may be written as $(\theta_0, \eta + \beta) = R_\beta(\theta_0, \eta)$ for some $\beta \in S^1$. Recalling (6.4.1), we see that

$$f(R_\beta(\tau_0, \zeta)) = (\tau_1 + \alpha, \tau_2 + \tau_1, \dots, \tau_{d-1} + \tau_{d-2}, \zeta + \beta + \tau_{d-1}) = R_\beta(f(\tau_0, \zeta))$$

for every $(\tau_0, \zeta) \in \mathbb{T}^{d-1} \times S^1$. Hence, by induction,

$$f^j(\theta_0, \eta + \beta) = f^j(R_\beta(\theta_0, \eta)) = R_\beta(f^j(\theta_0, \eta))$$

for every $j \geq 1$. Therefore, given any continuous function $\varphi : \mathbb{T}^d \rightarrow \mathbb{R}$,

$$\begin{aligned} \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(\theta_0, \eta + \beta)) &= \lim_n \frac{1}{n} \sum_{j=0}^{n-1} (\varphi \circ R_\beta)(f^j(\theta_0, \eta)) \\ &= \int (\varphi \circ R_\beta) \, dm = \int \varphi \, dm. \end{aligned}$$

This proves that $(\theta_0, \eta + \beta)$ is in $G(m)$ for every $\beta \in S^1$, as stated.

It follows from what we said so far that $G(\mu)$ and $G(m)$ intersect each other at some point of $\{\theta_0\} \times S^1$. In view of the definition (6.4.5), this implies that the two measures have the same integral for every continuous function. According to Proposition A.3.3, this implies that $\mu = m$, as we wanted to prove.

Corollary 6.4.7. *The orbit of every point $\theta \in \mathbb{T}^d$ is equidistributed on the torus \mathbb{T}^d , in the sense that*

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \psi(f^j(\theta)) = \int \psi \, dm$$

for every continuous function $\psi : \mathbb{T}^d \rightarrow \mathbb{R}$.

Proof. This follows immediately from Propositions 6.1.1 and 6.4.4.

6.4.3 Proof of the theorem of Weyl

To complete the proof of Theorem 6.4.2, we introduce the polynomial functions p_1, \dots, p_d defined by

$$\begin{aligned} p_d(x) &= P(x) \quad \text{and} \\ p_{j-1}(x) &= p_j(x+1) - p_j(x) \quad \text{for } j = 2, \dots, d. \end{aligned} \tag{6.4.8}$$

Lemma 6.4.8. *The polynomial $p_j(x)$ has degree j , for every $1 \leq j \leq d$. Moreover, $p_1(x) = \alpha x + \beta$ with $\alpha = d!a_d$.*

Proof. By definition, $p_d(x) = P(x)$ has degree d . Hence, to prove the first claim it suffices to show that if $p_j(x)$ has degree j then $p_{j-1}(x)$ has degree $j - 1$. In order to do that, let

$$p_j(x) = b_j x^j + b_{j-1} x^{j-1} + \dots + b_0,$$

where $b_j \neq 0$. Then

$$\begin{aligned} p_j(x+1) &= b_j(x+1)^j + b_{j-1}(x+1)^{j-1} + \dots + b_0 \\ &= b_j x^j + (j b_j + b_{j-1}) x^{j-1} + \dots + b_0. \end{aligned}$$

Subtracting one expression from the other, we get that

$$p_{j-1}(x) = (j b_j) x^{j-1} + b'_{j-2} x^{j-2} + \dots + b'_0$$

has degree $j - 1$. This proves the first claim in the lemma. This calculation also shows that the main coefficient of $p_{j-1}(x)$ (the coefficient of the term with highest degree) can be obtained multiplying by j the main coefficient of $p_j(x)$. Consequently, the main coefficient of p_1 must be equal to $d!a_d$, as claimed in the last part of the lemma.

Lemma 6.4.9. *For every $n \geq 0$,*

$$f^n(p_1(0), p_2(0), \dots, p_d(0)) = (p_1(n), p_2(n), \dots, p_d(n)).$$

Proof. The proof is by induction on n . Since the case $n = 0$ is obvious, we only need to treat the inductive step. Recall that f was defined in (6.4.1). If

$$f^{n-1}(p_1(0), p_2(0), \dots, p_d(0)) = (p_1(n-1), p_2(n-1), \dots, p_d(n-1))$$

then $f^n(p_1(0), p_2(0), \dots, p_d(0))$ is equal to

$$(p_1(n-1) + \alpha, p_2(n-1) + p_1(n-1), \dots, p_d(n-1) + p_{d-1}(n-1)).$$

Using the definition (6.4.8) and Lemma 6.4.8, we find that this expression is equal to

$$(p_1(n), p_2(n), \dots, p_d(n)),$$

and that proves the lemma.

Finally, we are ready to prove that the sequence $z_n = P_*(n)$, $n \in \mathbb{N}$ is equidistributed. We treat two cases separately.

First, suppose that the main coefficient a_d of $P(x)$ is irrational. Then the number α in Lemma 6.4.8 is irrational and, thus, the results in Section 6.4.2 are valid for the transformation $f : \mathbb{T}^d \rightarrow \mathbb{T}^d$. Let $\varphi : S^1 \rightarrow \mathbb{R}$ be any continuous function. Consider $\psi : \mathbb{T}^d \rightarrow \mathbb{R}$ defined by

$$\psi(\theta_1, \theta_2, \dots, \theta_d) = \varphi(\theta_d).$$

Fix $\theta = (p_1(0), p_2(0), \dots, p_d(0))$. Using Lemma 6.4.9 and Corollary 6.4.7, we get that

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \varphi(z_n) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \psi(f^j(\theta)) = \int \psi \, dm = \int \varphi \, dx.$$

This ends the proof of Theorem 6.4.2 in the case when a_d is irrational.

Now suppose that a_d is rational. Write $a_d = p/q$ with $p \in \mathbb{Z}$ and $q \in \mathbb{N}$. It is clear that we may write z_n as a sum

$$z_n = x_n + y_n, \quad x_n = a_d n^d \quad \text{and} \quad y_n = Q_*(n)$$

where $Q(x) = a_0 + a_1x + \dots + a_{d-1}x^{d-1}$ and $Q_* : \mathbb{R} \rightarrow S^1$ is given by $Q_* = \pi \circ Q$. To begin with, observe that

$$x_{n+q} - x_n = \frac{p}{q}(n+q)^d - \frac{p}{q}n^d$$

is an integer, for every $n \in \mathbb{N}$. This means that the sequence x_n , $n \in \mathbb{N}$ is periodic (with period q) in the circle \mathbb{R}/\mathbb{Z} . In particular, it takes no more than q distinct values. Observe also that, since a_d is rational, the hypothesis of the theorem implies that some of the coefficients a_1, \dots, a_{d-1} of Q are irrational. Hence, by induction on the degree, the sequence y_n , $n \in \mathbb{N}$ is equidistributed. More than that, the subsequences

$$y_{qn+r} = Q_*(qn+r), \quad n \in \mathbb{Z}$$

are equidistributed for every $r \in \{0, 1, \dots, q-1\}$. In fact, as the reader may readily check, these sequences may be written as $y_{nq+r} = Q_*^{(r)}(n)$ for some polynomial $Q^{(r)}$ that also has degree $d-1$ and, thus, the induction hypothesis applies to each one of them as well. From these two observations it follows

that every subsequence z_{qn+r} , $n \in \mathbb{Z}$ is equidistributed. Consequently, z_n , $n \in \mathbb{N}$ is also equidistributed. This completes the proof of Theorem 6.4.2.

6.4.4 Exercises

6.4.1. Show that a sequence $(z_j)_j$ is equidistributed on the circle if and only if

$$\lim_n \frac{1}{n} \#\{1 \leq j \leq n : z_j \in I\} = m(I)$$

for every segment $I \subset S^1$, where $m(I)$ denotes the length of I .

6.4.2. Show that the sequence $(\sqrt{n} \bmod \mathbb{Z})_n$ is equidistributed on the circle. Does the same hold for the sequence $(\log n \bmod \mathbb{Z})_n$?

6.4.3. Koksma [Kok35] proved that the sequence $(a^n \bmod \mathbb{Z})_n$ is equidistributed on the circle for Lebesgue-almost every $a > 1$. That is not true for every $a > 1$. Indeed, consider the *golden ratio* $a = (1 + \sqrt{5})/2$. Check that the sequence $(a^n \bmod \mathbb{Z})_n$ converges to $0 \in S^1$ when $n \rightarrow \infty$; in particular, it is not equidistributed on the circle.