Dynamical Systems and Ergodic Theory Notes

Alex Goodlad

December 17, 2020

Contents

1 Recurrence

§ Poincaré Recurrence Theorem

Definition 1.1. Let $(M, \mathcal{B}, \mu)$ be a measure space and $f\colon M \to M$ be a measurable transformation. We say that the measure μ is *invariant* under f if

$$\mu(E) = \mu f^{-1}(E)), \text{ for every measurable set } E \subset M.$$

We also say that μ is f*-invariant*, or that f*-preserves* μ to mean just the same.

Definition 1.2. We define *flows* as families of transformations $f^t\colon M \to M$, with $t \in \mathbb{R}$ satisfying the folllowing conditions:

$$f^0 = \mathrm{id} \text{ and } f^{s+t} = f^s \circ f^t \text{ for every } s, t \in \mathbb{R}.$$

In particular, each transformation f^t is invertible and the inverse is f^{-t}. Flows arise naturally in connection with differential equations of the form

$$\frac{d\gamma}{dt}(t) = X(\gamma(t))$$

in the following way: under suitable conditions on the vector field X, for each point x in the domain M there exists exactly one solution $t \mapsto \gamma_x(t)$ of the differential equation with $\gamma_x(0) = x$; then $f^t(x) = \gamma_x(t)$ defines a flow in M.

We say that a measure μ is *invariant* under a flow $(f^t)_t$ if it is invariant under each one of the transformations f^t, that is, if $\mu(E) = \mu(f^{-1}(E))$ for every measurable set $E \subset M$ and $t \in \mathbb{R}$.

Proposition 1.3. Let $f\colon f \to M$ be a measurable transformation and μ be a measure on M. Then f preserves μ if and only if

$$\int \phi d\mu = \int \phi \circ f d\mu.$$

for every μ-integrable function $\phi\colon M \to \mathbb{R}$.

Now we are going to study two versions of Poincaré's Theorem. The first one is formulated in the context of (finite) measure spaces. The second version of the recurrence theorem assumes that the ambient is a topological space with certain additional propeties

Theorem 1.4. (*(Poincaré recurrence)* Let $f\colon M \to M$ be a measurable transformation and μ be a finite measure invariant under f. Let $E \subset M$ be any measurable set with $\mu(E) > 0$. then, for μ-a.e. point $x \in E$ there exist infinitely many values of n for which $f^n(x)$ is also in E.

Poincaré recurrence implies an analogous result for continuous time systems:

Theorem 1.5. If μ is a finite invariant measure of a flow $(f^t)_t$, then for every measurable set $E \subset M$ with positive measure and for μ-almost every $x \in E$ there exists times $t_j \to +\infty$ such that $f^{t_j}(x) \in E$.

Let $f\colon M \to M$ be a measurable transformation and μ be a finite measure invariant under f. Let $E \subset M$ be any measurable set with $\mu(E) > 0$. Consider the *first-return time* function $\rho_E\colon E \to \mathbb{N} \cup \{\infty\}$, defined by

$$\rho_E(x) = \min\{n \geq 1\colon f^n(x) \in E\}.$$

if the set on the right-hand side is non-empty and $\rho_E(x) = \infty$ if, on the contrary, x has no irate in E. According to Poincaré recurrence, the second alternative occurs only on a set with zero measure.

The next result shows that this function is integrable and even provides the value of the integral. For the statement we need the following notation:

$$E_0 = \{x \in E\colon f^n(x) \notin E \text{ for every } n \geq 1\},$$
$$E_0^* = \{x \in M\colon f^n(x) \notin E \text{ for every } n \geq 0\}.$$

Theorem 1.6. *(Kač)* Let $f\colon M \to M$ be a measurable transformation, μ be a finite invariant measure and $E \subset M$ be a positive measure set. Then the function ρ_E is integrable and

$$\int_E \rho_E d\mu = \mu(M) - \mu(E_0^*).$$

Remarks: By definition, $E_n^* = f^{-n}(E) \setminus \bigcup_{k=0}^{n-1} f^{-k}(E)$. So the fact that the sum is finite implies tht the measure of E_n^* converges to zero when $n \to \infty$. This fact will be useful later.

Now let us suppose that M is a topological space, endowed with its Borel σ-algebra $\mathcal{B}$.

Definition 1.7. A point $x \in M$ is *recurrent* for a transformation $f\colon M \to M$ if there exists a sequence $n_j \to \infty$ of natural numbers such that $f^{n_j}(x) \to x$. Analogously, we say that $x \in M$ is recurrent for a flow $(f^t)_t$ if there exists a sequence $t_j \to +\infty$ of real numbers such that $f^{t_j}(x) \to x$ when $j \to \infty$.

In the next theorem, we assume that the topological space M admits a countable basis of open sets, that is, there exists a countable family $\{U_k\colon k \in \mathbb{N}\}$ of open sets such that every open subset of M may be written as a union of elements U_k of this family. This condition holds in most interesting examples.

Theorem 1.8. *(Poincaré Recurrence)* Suppose that M admits a countable basis of open sets. Let $f\colon M \to M$ be a measurable transformation and μ be a finite measure on M invariant under f. Then, μ-a.e. $x \in M$ is recurrent for f.

Let us point out that the conclusions of the earlier Poincaré Recurrence theorems and Kač Theorems are false, in general, if the measure μ is not finite.

To conclude, we present a purely topological version of the Berkhoff Recurrence Theorem, that makes no reference at all to invariant measures:

Theorem 1.9. *(Birkhoff Recurrence)* If $f\colon M \to M$ is a continuous transformation on a compact metric space M then there exists some point $x \in X$ that is recurrent for f.

§ Some Examples

Lemma 1.10. Given $x \in [0,1]$ we define the *fractional part* of $10x$ to be $10x - \lfloor 10x \rfloor$ and we define $\lfloor 10x \rfloor$ to be the integer part. Define $f\colon [0,1] \to [0,1]$ to be $f(x) = 10x - \lfloor 10x \rfloor$. Then f is measur-preserving.

Lemma 1.11. Let $f\colon M \to M$ be a measurable transformation and μ be a finite measure on M. Suppose that there exists some algebra $\mathcal{A}$ of measurable subsets of M such that $\mathcal{A}$ generates the σ-algebra $\mathcal{B}$ of M and $\mu(E) = \mu(f^{-1}(E))$ for every $E \in \mathcal{A}$. Then the latter remains true for every set $E \in \mathcal{B}$, that is, the measure μ is invariant under f.

The system we present in this section is related to another important algorithm in number theory, the *continued fraction expansion*, which plays a central role in the problem of finding the best rational approximation to any real number. Let us start with a brief presentation of this algorithm.

Given any number $x_0 \in (0,1)$, let

$$a_1 = \lfloor \frac{1}{x_0} \rfloor \text{ and } x_1 = \frac{1}{x_0} - a_1.$$

Note that a_1 is a natural number, $x_1 \in [0,1)$ and

$$x_0 = \frac{1}{a_1 + x_1}.$$

Supposing that x_1 is different from zero, we may repeat this procedure, defining

$$a_2 = \lfloor \frac{1}{x_1} \rfloor \text{ and } x_2 = \frac{1}{x_1} - a_2.$$

Then

$$x_1 = \frac{1}{a_1 + x_2} \text{ and so } x_0 = \frac{1}{a_1 + \frac{1}{a_2 + x_2}}.$$

Now we may proceed by induction: for each $n \geq 1$ such that $x_{n-1} \in (0,1)$, define

$$a_n = \lfloor \frac{1}{x_{n-1}} \rfloor \text{ and } x_n = \frac{1}{x_{n-1}} - a_n = G(x_{n-1}),$$

and observe that

$$x_0 = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n + x_n}}}}$$

It can be shown that the sequence

$$z_n = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}}} \qquad (*)$$

converges to x_0 when $n \to \infty$. This is usually expressed through the expression

$$z_n = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n + \cfrac{1}{\cdots}}}}}$$

which is called the *continued fraction expansion of* x_0.

Note that the sequence $\{z_n\}_n$ can be defined by the relation (*) consists of rational numbers.

Indeed, one can show that these are the *best rational approximations* of the number x_0, in the sense that each z_n is closer to x_0 than any other number whose denominator is smaller than or equal to the denominator of z_n (written in irreducible form).

This continued fraction algorithm is intimately related to a certain dynamical system on the interval $[0, 1]$ that we describe in the following way.

Definition 1.12. The *Gauss map* $G\colon [0, 1] \to [0, 1]$ is defined by

$$G(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor = \text{fractional part of } 1/x.$$

The graph of G can be easily sketched, starting from the following observation: for every x in each interval $I_k = (1/(k+1), 1/k]$, the integer part of $1/x$ is equal to k so $G(x) = 1/x - k$.

The continued fraction expansion of any number $x_0 \in (0, 1)$ can be obtained from the Gauss map in the following way: for each $n \geq 1$, the natural number a_n is determined by

$$G^{n-1}(x_0) \in I_{a_n},$$

and the real number x_n is simply the n-th irate $G^n(x_0)$ of the point x_0. This process halts whenever we encounter some $x_n = 0$. As explained, this can only if x_0 is a tational number.

§ Induction and Multiple Recurrence Theorems

Definition 1.12. Let $f\colon M \to M$ be a measurable transformation and μ be an invariant probability measure. Let $E \subset M$ be a measurable set with $\mu(E) > 0$ and $\rho(x) = \rho_E(x)$ be the first-return time of x to E (as defined previously). The *first-return map* to the domain E is the map g given by $g(x) = f^{\rho(x)}(x)$ whenever $\rho(x)$ is finite.

NOTE: The Poincaré recurrence theorem ensures that this is the case for μ-almost every $x \in E$ and so g is defined on a full measure subset of E.

We also denote by μ_E the restriction of μ to the measurable subsets E.

Proposition 1.13. The measure μ_E is invariant under the map $g\colon E \to E$ defined above

In an opposite direction, given any measure ν invariant under $g\colon E \to E$, we may construct a certain related measure ν_ρ that is invariant under $f\colon M \to M$. For this, g does not even have to be a first-return map.

Definition 1.14. We say that g is *induced* from f if for any map invariant map $g\colon E \to E$, we have $g(x) = f^{\rho(x)}(x)$ for measurable $\rho\colon E \to \mathbb{N}$. As before, we denote by E_k the subset of points $x \in E$ such that $\rho(x) = k$. Then we define

$$v_\rho(B) = \sum_{n=0}^{\infty} \sum_{k>n} \nu(f^{-n}(B) \cap E_k),$$

for every measurable set $B \subset M$.

Proposition 1.15. The measure ν_p defined above is invariant under f and satisfies $\nu_\rho(M) = \int_E \rho d\nu$. In particular, ν_p if finite if and only if the function ρ is integrable with respect to ν.

Corollary 1.16. If g is the first-return map of f to be a measurable subset E and $\nu = \mu|E$, then

1. $\nu_\rho(B) = \nu(B) = \mu(B)$ for every measurable set $B \subset E$.

2. $\nu_\rho(B) \le \mu(B)$ for every measurable set $B \subset M$.

Now we consider finite families of *commuting maps* $f_i\colon M \to M$, $i = 1, \dots, q$, that is, such that

$$f_i \circ f_j = f_j \circ f_i \text{ for every } i, j \in 1, \dots, q.$$

Our goal is to explain that the results earlier extend to this setting: we find points that are *simultaneously recurrent* for these transformations.

The first result in this direction generalizes the Birkhoff recurrence theorem:

Theorem 1.17. *(Birkhoff Multiple Recurrence)* Let M be a compact metric space and $f_1, \dots, f_q\colon M \to M$ be continuous commuting maps. Then there exists $a \in M$ and a sequence $(n_k)_k \to \infty$ such that

$$\lim_k f_i^{n_k}(a) = a \text{for every } i = 1, \dots, q.$$

The key point here is that the sequence $(n_k)_k$ does not depend on i: we say that the point a is *simultaneously recurrent* for all the maps $f_i, i = 1, \dots, q$. Next, we discuss the following generalization of the Poincaré recurrence theorem

Theorem 1.18. *(Poincaré Multiple Recurrence)* Let $(M, \mathcal{B}, \mu)$ be a probability space and $f_i \colon M \to M$, $i = 1, \ldots, q$ be measurable commuting maps that preserve the measure μ. Then, given any set $E \subset M$ with positive measure, there exists $n \geq 1$ such that

$$\mu(E \cap f_1^{-n}(E) \cap \cdots \cap_q^{-n} (E)) > 0.$$

In other words, for a positive measure subset of points $x \in E$, their orbits under all the maps f_i, $i = 1, \ldots, q$ return to E *simultaneously* at time n (we say that n is a *simultaneous return* of x to E): once more, the crucial point with the statement is that n does not depend on i.

2 Existence of Invariant Measures

In this chapter we prove the following result, which guarentees the existence of invariant measures for a broad class of transformations:

Theorem 2.1. *(Existence of Invariant Measures)* Let $f\colon M \to M$ be a continuous transformation on a compact metric space. Then there exists some probability measure on M invariant under f.

The main point in the proof is to introduce a certain topology in the set $\mathcal{M}_1(M)$ of probability measures on M, that we call the weak* topology. The idea is that two measures are close, with respect to this topology, if the integrals they assign to (many) bounded continuous functions are close. The precise definition and some of the properties of the weak* topology are presented in the following section.

§ The Weak* Topology and Proof of Existence

Given a measure $\mu \in \mathcal{M}_1(M)$, a finite set $\Phi = \{\phi_1, \dots, \phi_N\}$ of bounded continuous functions $\phi_i\colon M \to \mathbb{R}$ and a number $\epsilon > 0$, we define

$$V(\mu, \Phi, \epsilon) = \left\{ \nu \in \mathcal{M}_1(M)\colon \left| \int \phi_i d\nu - \int \phi_i d\mu \right| < \epsilon \text{ for every } i \right\}.$$

Note that the intersection of any two such sets contains some set of this form. Thus, the family $\{V(\mu, \Phi, \epsilon)\colon \Phi, \epsilon\}$ may be taken as a basis of neighborhoods of each $\mu \in \mathcal{M}_1$.

Definition 2.2. The *weak* topology* is the topology defined by these bases of neighborhoods. In other words, the open sets in the weak* topology are the sets $\mathcal{A} \subset \mathcal{M}_1(M)$ such that for every $\mu \in \mathcal{A}$ there exists some $V(\mu, \Phi, \epsilon)$ contained in $\mathcal{A}$.

Lemma 2.3. A sequence $(\mu_n)_{n\in\mathbb{N}}$ converges to a measure $\mu \in \mathcal{M}_1(M)$ in the weak* topology if and only if $\int \phi d\mu_n \to \int \phi d\mu$ for every bounded continuous function $\phi\colon M \to \mathbb{R}$.

Now let us discuss other useful ways of defining the weak* topology.

We find a direct variation of the definition of weak* topology is obtained by taking as the basis of neighborhoods the family of sets

$$V(\mu, \Psi, \epsilon) = \left\{ \eta \in \mathcal{M}_1(M)\colon \left| \int \psi_i d\eta - \int \psi_i d\mu \right| < \epsilon \text{ for every } i \right\},$$

where $\epsilon > 0$ and $\Psi = \{\psi_1, \ldots, \psi_N\}$ is a family of Lipschitz functions. The next definition is formulated in terms of closed subsets. Given any finite family $\mathcal{F} = \{F_1, \ldots, F_N\}$ of closed subsets of M and given any $\epsilon > 0$, consider

$$V_f(\mu, \mathcal{F}, \epsilon) = \{\nu \in \mathcal{M}_1 \colon \nu(F_i) < \mu(F_i) + \epsilon \text{ for every } i\}.$$

The next construction is analogous, just with open subsets instead of closed subsets. Given any finite family $\mathcal{A} = \{A_1, \ldots, A_N\}$ of open subsets of M and given $\epsilon > 0$, consider

$$V_a(\mu, \mathcal{A}, \epsilon) = \{\nu \in \mathcal{M}_1 \colon \nu(A_i) > \mu(A_i) - \epsilon \text{ for every } i\}.$$

We call a *continuity set* of a measure μ any Borel subset B of M whose boundary ∂B has zero measure μ. Given any finite family $\mathcal{B} = \{B_1, \ldots, B_N\}$ of continuity sets of μ and given any $\epsilon > 0$, consider

$$V_c(\mu, \mathcal{B}, \epsilon) = \{\nu \in \mathcal{M}_1 \colon |\mu(B_i) - \nu(B_i)| < \epsilon \text{ for every } i\}.$$

Theorem 2.4. *(Portmanteau)* Topologies all above are equivalent.

Given $\delta > 0$ and $B \subset M$, we define

$$d(x, B) = \inf\{d(x, y) \colon y \in B\},$$

and we call the set $B^\delta := \{x \in M \colon d(x, B) < \delta\}$ the δ-neighborhood of B.

Definition 2.5. Given $\mu, \nu \in \mathcal{M}_1(M)$, define

$$D(\mu, \nu) = \inf\{\delta \colon \mu(B) < \nu(B^\delta) + \delta \text{ and } \nu(B) < \mu(B^\delta) + \delta, \forall B \in \mathcal{B}(M)\}.$$

Lemma 2.6. D is a metric on $\mathcal{M}_1(M)$.

This distance D is called the *Levy-Prohorov metric* on $\mathcal{M}_1(M)$. In what follows we denote by $B_D(\mu, r)$ the ball of radius $r > 0$ around any $\mu \in \mathcal{M}_1(M)$.

Proposition 2.7. If M is a separable metric space then the topology induced by the Levy-Prohov distance D coincides with the weak* topology on $\mathcal{M}_1(M)$.

Theorem 2.8. The space $\mathcal{M}_1(M)$ is compact for the weak* topology.

Proposition 2.9. Every sequence $(\mu_k)_{k \in \mathbb{N}}$ is $\mathcal{M}_1(M)$ has some subsequence that converges in the weak* topology.

Definition 2.10. A set $\mathcal{M}$ of Borel measures in a topological space is *tight* if for every $\epsilon > 0$ there exists a compact set $K \subset M$ such that $\mu(K^c) < \epsilon$ for every measure $\mu \in \mathcal{M}$.

Theorem 2.11. *(Prohorov)* Let M be complete separable metric space. A set $\mathcal{K} \subset \mathcal{M}_1(M)$ is tight if and only if every sequence in $\mathcal{K}$ admits some subsequence that is convergent in the weak* topology of $\mathcal{M}_1(M)$.

Definition 2.12. Given any $f\colon M \to M$ and any measure η on M, we denote by $f_*\eta$ and call the *iterate (or image) of* η *under* f the measure defined by

$$f_*\eta(B) = \eta(f^{-1}(B))$$

for each measurable $B \subset M$. Note that the measure η is invariant under f if and only if $f_*\eta = \eta$.

Lemma 2.13. Let η be a measure and ϕ be a bounded measureable function. Then

$$\int \phi d f_*\eta = \int \phi \circ f d\eta.$$

Proposition 2.14. If $f\colon M \to M$ is continuous then $f_*\colon \mathcal{M}_1(M) \to \mathcal{M}_1(M)$ is continuous relative to the weak* topology.

Definition 2.15. A *topological vector space* is a vector space V endowed with a topology relative to which both operations of V (addition and multiplication by a scalar) are continuous. A set $K \subset V$ is said to be *convex* if $(1-t)x + ty \in K$ for every $x, y \in K$ and every $t \in [0,1]$.

Theorem 2.16. *(Schauder-Tychonoff)* Let $F\colon V \to V$ be a continuous transformation on a topological vector space V. Suppose that there exists a compact convex $K \subset V$ such that $F(K) \subset K$. Then $F(v) = v$ for some $v \in K$.

Let ν be any probability measure on M: for example, ν could be the Dirac mass at any point. Form the sequence of probability measures

$$\mu_n = \frac{1}{n}\sum_{j=0}^{n-1} f_*^j \nu, \quad (*)$$

where $f_*^j\nu$ is the image of ν under the iterate f^j. This sequence has some accumulation point, that is, there exists some subsequence $(n_k)_{n\in\mathbb{N}}$ and some probability measure $\mu \in \mathcal{M}_1(M)$ such that

$$\frac{1}{n_k}\sum_{j=0}^{n_k-1} f_*^j \nu \to \mu$$

in the weak* topology. Now we only need the following lemma.

Lemma 2.17. Every accumulation point of a sequence $(\mu_n)_{n\in\mathbb{N}}$ of the form $(*)$ is a probability measure invariant under f.

Example 2.18. Consider $f\colon (0,1] \to (0,1]$ given by $f(x) = x/2$. Suppose that f admits some invariant probability measure: we are going to show that this is actually not true. There are no recurrent points since in particular, they all converge to 0.

Example 2.19. Modifying a little the previous construction, we see that the same phenomenon may occur in compact spaces, if the transformation is not continuous. Consider $f\colon [0,1] \to [0,1]$ given by $f(x) = x/2$ if $x \neq 0$ and $f(0) = 1$. For the same reason as before, no point $x \in (0,1]$ is recurrent. So, if there exists some invariant probability measure μ then it must give full weight to the sole recurrent point $x = 0$, i.e., δ_0 is the measure but that measure is *not* invariant under f since $\{0\}$ has measure 1 but its preimage is the empty set, which has measure zero.

Here's an example of a measure that *does work*

Example 2.20. Consider $f\colon [0,1] \to [0,1]$ given by $f(x) = x/2$. This is a continuous transformation on a compact space. So, by the existence theorem, f admits some probability measure. Using the same arguments as the previous example, we find that there exists a unique invariant probability measure, namely, the Dirac mass δ_0 at the origin. Note that in this case the measure δ_0 is indeed invariant.

§ Other Functional Analysis Tidbits

§ Skew-Products, Inverse Limits, and Topological Conjugacy

Definition 2.21. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. We call a *skew-product* any measurable transformation $F\colon X \times Y \to X \times Y$ form $F(x,y) = (f(x), g(x,y))$. Represent by $\pi\colon X \times Y \to X$ the canonical production to the first coordinate.

By definition
$$\pi \circ F = f \circ \pi.$$
Let λ be a probability measure on $X \times Y$ invariant under F. We find that the $\pi_*\lambda$ of λ to X is invariant under f. The next proposition provides a partial converse.

Proposition 2.22. Let X be a complete separable metric space, Y be a compact metric space and F be continuous. Then, for every probability measure μ on X invariant under f such that there exists some probability measure λ on $X \times Y$ invariant under F such that $\pi_*\lambda = \mu$.

Next, we'll talk about inverse limits. We are going to see that, given any surjective transformation $f\colon M \to M$, one can always find an extension $\widehat{f}\colon \widehat{M} \to \widehat{M}$ that is invertible. By *extension* we mean that there exists a surjective map $\pi\colon \widehat{M} \to M$ such that $\pi \circ \widehat{f} = f \circ \pi$.

Definition 2.23. To begin with, take the *inverse limit* of $(\widehat{M}, \widehat{f})$ of (M, f) to be the set $\widehat{M}$ of all *pre-orbits* of f, that is, all sequences $(x_n)_{n \leq 0}$ indexed by the non-prositive integers and satisfying $f(x_n) = x_{n+1}$ for every $n < 0$. Consider the map $\pi\colon \widehat{M} \to M$ sending each sequence $(x_n)_{n \leq 0}$ to its term x_0 of order zero. Observe that $\pi(\widehat{M}) = M$. Findally, define $\widehat{f}\colon \widehat{M} \to \widehat{M}$ to be the shift by one unit to the left:
$$\widehat{f}(\ldots, x_n, \ldots, x_0) = (\cdots, x_n, \ldots, x_0, f(x_0)).$$

It is clear that $\widehat{f}$ is well-defined and satisfies $\pi \circ \widehat{f} = f \circ \pi$. Moreover, $\widehat{f}$ is invertible: the inverse being the shift to the right.

If M is a measurable space then we may turn $\widehat{M}$ into a measurable space by endowing it with the σ-algebra generated by the *measurable cylinders*

$$[A_k, \dots, A_0] = \{(x_n)_{n \le 0} \in \widehat{M} : x_i \in A_i \text{ for } i = k, \dots, 0\},$$

where $k \le 0$ and $A_k, \dots, A_0$ are measurable subsets of M. Then π is a measurable map, since $\pi^{-1}(A) = [A]$. Moreover, $\widehat{f}$ is meaurable if f is measurable:

$$\widehat{f}^{-1}([A_k, \dots, A_0]) = [A_k, \dots, A_{-2}, A_{-1} \cap f^{-1}(A_0)].$$

The inverse of $\widehat{f}$ is also measurable, since

$$\widehat{f}([A_k, \dots, A_0]) = [A_k, \dots, A_0, M].$$

Analogously, if M is a topological space then we may turn $\widehat{M}$ into a topological space by endowing it with the topology generated by the *open cyllinders* $[A_k, \dots, A_0]$, where $k \le 0$ and $A_k, \dots, A_0$ are open subsets of M. A similar derivation to the one above shows that the maps π and $\widehat{f}^{-1}$ are continuous, while $\widehat{f}$ is continous if f is continuous.

If M is a metric space, with distance d, then the following function is a distance on $\widehat{M}$:

$$\widehat{d}(\widehat{x}.\widehat{y}) = \sum_{n=-\infty}^{0} 2^n \min\{d(x_n, y_n), 1\},$$

where $\widehat{x} = (x_n)_{n \le 0}$ and $\widehat{y} = (y_n)_{n \le 0}$. It follows immediately from the definition that if $\widehat{x}$ and $\widehat{y}$ belong to the same pre-image $\pi^{-1}(x)$ then

$$\widehat{d}(\widehat{f}^j(\widehat{x}), \widehat{f}^j(\widehat{y})) \le 2^{-j} \widehat{d}(\widehat{x}, \widehat{y}), \text{ for every } j \ge 0.$$

So, every pre-image $\pi^{-1}(x)$ is a *stable set*, that is, a subset restricted to which the transformation $\widehat{f}$ is uniformly contracting.

Example 2.24. Given any transformation $g \colon M \to M$, consider its maximal invariant set $M_g = \bigcap_{n=1}^{\infty} g^n(M)$. Clearly, $g(M_g) \subset M_g$. Suppose that

(i) M is compact and g is continuous
(ii) $\# g^{-1}(y) < \infty$ for every y.

Then the restriction $f = (g|M_g) \colon M_g \to M_g$ is surjective. This restriction contains all the interesting dynamics of g. For example, assuming that $f^n(M)$ is a measurable set for every n, every probability measure invariant under f is also invariant under f.

A set $\Lambda \subset M$ such that $f^{-1}(\Lambda) = \Lambda$ is called an *invariant set* of f. There is a corresponding notion for the transformation $\widehat{f}$. The next proposition shows that every closed invariant set of f admits a unique lift to a closed invariant set of the transformation $\widehat{f}$:

Proposition 2.25. Assume that M is a topological space. If $\Lambda \subset M$ is a closed set under f then $\widehat{\Lambda} = \pi^{-1}(\Lambda)$ is the only closed set invariant under f and satisfying $\pi(\widehat{\Lambda}) = \Lambda$.

Now let $\widehat{\mu}$ be an invariant measure of $\widehat{f}$ and let $\mu = \pi_* \widehat{\mu}$. The property $\pi \circ \widehat{f} = f \circ \pi$ implies that μ is invariant under f.

Definition 2.26. We say that $\widehat{\mu}$ is a *lift* of μ.

Proposition 2.27. Assume that M is a complete separable metric space and $f \colon M \to M$ is continuous. Then every probability measure μ invariant under f admits a unique lift, that is, there is a unique measure $\widehat{\mu}$ on $\widehat{M}$ invariant under $\widehat{f}$ and such that $\pi_* \widehat{\mu} = \mu$.

Finally, we'll talk about topological conjugacy.

Definition 2.28. Suppose M, N are topological spaces, $f \colon M \to M, g \colon N \to N$ are continuous. If there exists a homeomorphism $\varphi \colon M \to N$ such that the diagram commutes $g = \varphi \circ f \circ^{-1}$ (see diagram below) then f and g are *topologically conjugate*.

$$\begin{tikzcd}
M \arrow[r, "f"] \arrow[d, "\varphi"'] & M \arrow[d, "\varphi"] \\
N \arrow[r, "g"'] & N
\end{tikzcd}$$

§ Arithmetic Progressions

3 Ergodic Theorems

In this chapter we present the fundamental results of ergodic theory. To motivate these results, considser a measurable set $E \subset M$ with positive measure and an arbitrary point $x \in M$. We want to analyze the set of iterates of x that visit E, that is, $\{j \geq 0\colon f^j(x) \in E\}$.

For example, the Poincaré recurrence theorem states that this set is infinite, for almost every $x \in E$. We would like to have more precise quantitative information. Let us call the *mean sojourn time* of x to E the value of

$$\tau(E, x) = \lim_{n\to\infty} \#\{0 \leq j < n\colon f^j(x) \in E\}.$$

There is an analogous notion for flows, defined by

$$\tau(E, x) = \lim_{T\to\infty} m(\{0 \leq t \leq T\colon f^t(x) \in E\}),$$

where m is the Lebesgue measure on the real line.

These questions go back to the work of the Austrian physicist Ludwig Boltzmann (1844-1906), who developed the kinetic theory of gases. Boltzmann was an emphatic supporter of atomic theory, according to which gases are formed by a large number of small moving particules, constantly colliding with each other, at a time when this theory was still highly controversial.

The proposal of the kinetic theory was, then, to try and explain the behavior of gases at the macroscopic scale as the statistical combination of the motions of all its molecules. To formulate the theory in precise mathematical terms, Boltzmann was forced to make an assumption that became known as the *ergodic hypothesis*. In modern language, the ergodic hypothesis claims that, for the kind of systems (Hamiltonian flows) that describe the motions of particles of a gas, *the mean sojourn time to any measurable set E exists and is equal to the measure of E, for almost every point x.*

Efforts to validate (or not) this hypothesis led to important developments, in mathematics (ergodic theory, dynamical systems) as well as in physics (statistical mechanics).

Denoting by φ the characteristic function of the set E, we may rewrite the expression on the right-hand side as

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)).$$

This suggest a natural generalization of the original question: does the limit exist for all integrable functions?

The ergodic theorem of von Neumann states that the limit in does exist, in the space $L^2(\mu)$, for every function $\varphi \in L^2(\mu)$. The ergodic theorem of Birkhoff goes farther, by asserting that convergence holds for μ-almost every point, for every $\varphi \in L^1(\mu)$. In particular, the limit is well defined for μ-almost every x.

§ Von Nuemann Ergodic Theorem

§ Birkhoff Ergodic Theorem

We start by stating the version of the theorem for mean sojourn times:

Theorem 3.1. *(Birkhoff Ergodic Theorem, version 1)* Let $f\colon M \to M$ be a measurable transformation and μ be a probability measure invariant under f. Given any measurable set $E \subset M$, the mean sojourn time

$$\tau(E,x) = \lim_n \frac{1}{n}\#\{j = 0, 1, \ldots, n-1 : f^j(x) \in E\}$$

exists at μ-almost every point $x \in M$. Moreover, $\int \tau(E,x)d\mu(x) = \mu(E)$.

As we obseved,

$$\tau(E,x) = \lim_n \frac{1}{n}\sum_{j=0}^{n-1}\varphi(f^j(x)), \text{ where } \varphi = \mathbb{1}_E.$$

The next statement extends **Theorem 3.1** to the case when φ is any integrable function:

Theorem 3.2. *(Birkhoff Ergodic Theorem, version 2)* Let $f\colon M \to M$ be a measurable transformation and μ be a probability measure invariant under f. Given any integrable function $\varphi\colon M \to \mathbb{R}$, the limit

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}\varphi(f^j(x))$$

exists at μ-almost every point $x \in M$. Moreover, the funtion $\overline{\varphi}$ defined in the way is integrable and satisfies

$$\int \overline{\varphi}(x)d\mu(x) = \int \varphi(x)d\mu(x).$$

The limit $\overline{\varphi}$ is called the *time avergae* or *orbital average* of φ. The next proppositions shows that time averages are constant on the orbit of μ-almost every point, which further generalizes linetheo3.1

Proposition 3.3. Let $\varphi\colon M \to \mathbb{R}$ be an integrable function. then

$$\overline{\varphi}(f(x)) = \overline{\varphi(x)},$$

for μ-almost every point $x \in M$.

Lemma 3.4. If ϕ is an integrable function then $\lim_n (1/n)\phi(f^n(x)) = 0$ for μ-almost every point $x \in M$.

Theorem 3.5. Let M be a compact metric space and $f\colon M \to M$ be a measurable map. then there exists some measurable set $G \subset M$ with $\mu(G) = 1$ such that

$$\frac{1}{n}\sum_{j=0}^{n-1} \varphi(f^j(x)) \to \overline{\varphi}(x),$$

for every $x \in G$ and every continuous function $\varphi\colon M \to \mathbb{R}$.

§ Subadditive Ergodic Theorem

Definition 3.6. A sequence of functions $\varphi_n\colon N \to \mathbb{R}$ is said to be *subadditive* for a transformation $f\colon M \to M$ if

$$\varphi_{m+n} \leq \varphi_m + \varphi_n \circ f^m \text{ for every } m, n \geq 1.$$

Example 3.7. A sequence $\varphi_n\colon M \to \mathbb{R}$ is *additive* for the transformation f if $\varphi_{m+n} = \varphi_m + \varphi_n \circ f^m$ for every $m, n \geq 1$. For example, the time sums

$$\varphi_n(x) = \sum^{n-1} \varphi(f^j(x))$$

of any function $\varphi\colon M \to \mathbb{R}$ form an additive sequence. In fact, every additive sequence is of this form, with $\varphi = \varphi_1$. Of course, additive sequences are also subadditive.

Example 3.8. Let $A\colon M \to GL(d)$ be a measurable function with values in the *linear group*, that is, the set $GL(d)$ of invertible square matrices of dimension d. Define

$$\phi^n(x) = A(f^{n-1}(x)) \cdots A(f(x))A(x),$$

for every $n \geq 1$ and $x \in M$. Then the sequence $\varphi_n(x) = \log ||\phi^n(x)||$ is subadditive. Indeed,

$$\phi^{m+n}(x) = \phi^n(f^m(x))\phi^m(x),$$

and so we have

$$\varphi_{m+n}(x) = \log ||\phi^n(f^m(x))\phi^m(x)|| \leq \log ||\phi^m(x)|| + \log ||\phi^n(f^m(x))|| = \varphi_m(x) + \varphi_n(f^m(x)),$$

for every m, n and x.

Theorem 3.9. *(Kingman)* Let μ be a probability measure invariant under a transformation $f\colon M \to M$ and let $\varphi_n\colon M \to \mathbb{R}$, $n \geq 1$ be a subadditive sequence of measurable function such that $\varphi_1^+\colon M \to \mathbb{R}$, $n \geq 1$ be a subbadditive sequence of measurable functions such that $\varphi_1^+ \in L^1(\mu)$. Then $(\varphi_n/n)_n$ converges at μ-almost every point to some function $\varphi\colon M \to [-\infty, +\infty)$ that is invariant under f. Moreover, $\varphi^+ \in L^1(\mu)$ and

$$\int \varphi d\mu = \lim_n \frac{1}{n} \int \varphi_n d\mu = \inf_n \frac{1}{n} \int \varphi_n d\mu \in [-\infty, +\infty).$$

Definition 3.10. A sequence $(a_n)_n$ in $[-\infty, +\infty)$ is said to be *subadditive* if $a_{m+n} \leq a_m + a_n$ for every $m, n \geq 1$.

Lemma 3.11. If $(a_n)_n$ is a subadditive sequence then

$$\lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n} \in [-\infty, \infty).$$

For a subadditive sequence $(\varphi_n)_n$, we shall denote

$$\varphi_-(x) = \liminf_n \frac{\varphi_n}{n}(x)$$
$$\varphi_+(x) = \limsup_n \frac{\varphi_n}{n}(x)$$

Lemma 3.12. For every $n > k \geq 1$ and μ-almost every $x \in M$.

$$\varphi_n \leq \sum_{i=0}^{n-k-1} \psi_k(f^j(x)) + \sum_{i=n-k}^{n-1} \max\{\psi_k, \phi_1\}(f^i(x)).$$

Lemma 3.13. For any fixed k,

$$\limsup_n \frac{\varphi_{kn}}{n} = k \limsup_n \frac{\varphi_n}{n}.$$

Definition 3.14. Let $f \colon M \to M$ be a measurable transformation and μ be an invariant probability measure. Consider any measurable function $\theta M \to GL(d)$ with values in the group $GL(d)$. The *cocycle* defined by θ over f is the sequence of functions defined by

$$\phi^n(x) = \theta(f^{n-1}(x)) \cdots \theta(f(x))\theta(x) \text{ for } n \geq 1 \text{ and } \theta^0(x) = \mathrm{id},$$

for every $x \in M$. We leave it to the reader to check that

$$\phi^{m+n}(x) = \phi^n(f^m(x)) \cdot \phi^m(x) \text{ for every } m, n \in \mathbb{N} \text{ and } x \in M.$$

It is also easy to check that, conversely, any sequence $(\phi^n)_n$ with this property is the coceycle defined byt $\theta = \phi^1$ over the transformation f.

Theorem 3.14. *(Furstenberg-Kesten)* If $\log^+ ||\theta|| \in L^1(\mu)$ then

$$\lambda_{max}(x) = \lim_n \frac{1}{n} \log ||\phi^n(x)||$$

exists at μ-almost every point. Moreover, $\lambda_{max}^+ \in L^1(\mu)$ and

$$\int \lambda_{max} d\mu = \lim_n \frac{1}{n} \int \log ||\phi^n|| d\mu = \inf_n \frac{1}{n} \int \log ||\phi^n|| d\mu.$$

If $\log^+ ||\theta^{-1}|| \in L^1(\mu)$ then

$$\lambda_{min}(x) = \lim_n -\frac{1}{n} \log ||\phi^n(x)^{-1}||$$

exists at μ-almost every point. Moreover, $\lambda_{min} \in L^1(\mu)$ and

$$\int \lambda_{min} d\mu = \lim_n -\frac{1}{n} \int \log ||(\phi^n)^{-1}|| d\mu = \sup_n -\frac{1}{n} \int \log ||(\phi^n)^{-1}|| d\mu.$$

4 Ergodicity

§ Definition of Ergodic Systems

The thoerems presented in the previous chapter fully establish the first part of Boltzman's ergodic hypothesis: for any measurable set E, the mean sojourn time $\tau(E,x)$ is well-defined for almost every point x. The second part of the ergodic hypothesis, that is, the claim that $\tau(E,x)$ should coincide with the measure of E for almost every x, is a statement of a different nature and is the subject of the present chapter.

In this chapter we take μ to be the probability measure invariant under some measurable transformation $f\colon M \to M$.

Definition 4.1.

1. We say that the system (f,μ) is *ergodic* if, given any measurable set E, we have $\tau(E,x)=\mu(E)$ for μ-almost every point $x \in M$.

2. A measurable function $\varphi\colon M \to \mathbb{R}$ is said to be *invariant* if $\varphi = \varphi \circ f$ at μ-almost every point. In other words, φ is invariant if it is constant on every trajectory of f outside a zero meausre subset. Moreover, we say that a measurable set $B \subset M$ is *invariant* if its characteristic function $\mathbb{1}_B$ is.

We are going to see that this is equivalent to saying that the system is dynamically indivisible, in the sense that every invariant set has either full measure or zero measure. We will also discuss other equivalent formulations.

Example 4.2. Let $f\colon [0,1] \to [0,1]$ be the decimal expansion, and μ be the Lebesgue measure. Clearly, the set $A = \mathbb{Q} \cap [0,1]$ of rational numbers is invariant. Other interesting examples are the sets of points $x = 0.a_1a_2\ldots$ in $[0,1[$ with prescribed proportions of digits a_i with each value $k \in \{0,\ldots,9\}$. More precisely, given any vector $p = (p_0,\ldots,p_9)$ such that $p_i \geq 0$ for every i and $\sum_i p_i = 1$, define

$$A_p \left\{ x\colon \lim_n \frac{1}{n} \#\{1 \leq i \leq n\colon a_i = k\} = p_k \text{ for } k = 0,\ldots,9 \right\}.$$

Observe that if $x = 0.a_1a_2\ldots$, then every point $y \in f^{-1}(x)$ may be written as $y = 0.ba_1a_2\ldots$ with $b \in \{0,\ldots,9\}$. It is clear that the extra digit b does not affect the propertion of digits with any of the values $0, \ldots, 9$ in the decimal expansion. Thus, $y \in A_p$ if and only if $x \in A_p$. This implies that A_p is indeed invariant under f.

Example 4.3. Let $\varphi\colon [0,1] \to \mathbb{R}$ be a function in $L^1(\mu)$. According to the ergodic theorem of Birkhoff, the time average $\widetilde{\varphi}$ is an invariant function. So, every level set

$$B_c = \{x \in [0,1]; \widetilde{\varphi}(x) = c\}$$

is an invariant set. Observe also that every infariant function is of this form: it is clear that if φ is invairant then it coincides with its time average $\widetilde{\varphi}$ at μ-almost every point.

Here's a few equivalent ways to define ergodicity

Proposition 4.4. Let μ be an invariant probability measure of a measurable transformation $f\colon M \to M$. The following conditions are equivalent:

(i) For every measurable set $B \subset M$ one has $\tau(B, x) = \mu(B)$ for μ-almost every point.

(ii) For every measurable set $B \subset M$ the function $\tau(B, \cdot)$ is constant at μ-almost every point.

(iii) For every integrable function $\varphi\colon M \to \mathbb{R}$ one has $\widetilde{\varphi}(x) = \int \varphi d\mu$ for μ-almost every point.

(iv) For every invariant integrable function $\varphi\colon M \to \mathbb{R}$, the time average $\widetilde{\varphi}\colon M \to \mathbb{R}$ is contant at μ-almost every point.

(v) For every invariant integrable function $\psi\colon M \to \mathbb{R}$ one has $\psi(x) = \int \psi d\mu$ for μ-almost every point.

(vi) Every invariant integrable function $\psi\colon M \to \mathbb{R}$ is constant at μ-almost every point.

(vii) For every invariant subset A we have either $\mu(A) = 0$ or $\mu(A) = 1$.

The next proposition characterizes the ergodicity property in terms of the Koopman operator $U_f(\varphi) = \varphi \circ f$:

Proposition 4.5. Let μ be an invariant probability measure of a measurable transformation $f\colon M \to M$. The following conditions are equivalent:

(i) (f, μ) is ergodic.

(ii) For any pair of measurable sets A and B one has

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \mu(f^{-j}(A) \cap B) = \mu(A)\mu(B).$$

(iii) For any functions $\varphi \in L^p(\mu)$ and $\psi \in L^p(\mu)$, with $1/p + 1/q = 1$, one has

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \int (U_f^j \varphi)\psi d\mu = \int \varphi d\mu \int \psi d\mu.$$

Corollary 4.6. Assume that the condition in previous proposition holds for every A and B in some algebra $\mathcal{A}$ that generates the σ-algebra of measurable sets. then (f,μ) is ergodic.

Corollary 4.7. Suppose

$$\lim_n \frac{1}{n}\sum_{j=0}^{n-1}\int (U_f^j\varphi)\psi = \int \varphi d\mu \int \psi d\mu,$$

holds for every φ and ψ in dense subsets of $L^p(\mu)$ and $L^q(\mu)$, respectively. then (f,μ) is ergodic.

Next, we take the transformation $f\colon M\to M$ to be fixed and we analyze the set $M_e(f)$ of proabability measures that are ergodic with respect to f as a subset of space $M_1(f)$ of all probability measures invariant under f.

Recall that a measure ν is said to be *absolutely continuous* with respect to another measure μ if $\mu(E)=0$ implies $\nu(E)=0$. Then we write $\nu\ll\mu$. This relation is transitive: if $\nu\ll\mu$ and $\mu\ll\lambda$ then $\nu\ll\lambda$. Then the first result asserts that the ergodic probability measures are minimal for this order relation:

Lemma 4.8. If μ and ν are invariant probability measures such that μ is ergodic and ν is absolutely continuous with respect to μ and $\mu=\nu$.

It is clear that if μ_1 and μ_2 are probability measures invariant under the transformation f then so is $(1-t)\mu_1+t\mu_2$, for any $t\in(0,1)$. This means that the space $M_1(f)$ of all probability measures invariant under f is *convex.* The next proposition asserts that the ergodic probability measures are the *extramal elements* of this convex set:

Proposition 4.9. An invariant probability measure μ is ergodic if and only if it is not possible to write it as $\mu=(1-t)\mu_1+t\mu_2$ with $t\in(0,1)$ and $\mu_1,\mu_2\in M_1(f)$ with $\mu_1\neq\mu_2$.

Let us also point that distict ergodic measures "live" in disjoint subsets of the space M

Lemma 4.10. Assume that the σ-algebra of M admits some countable generating subset Γ. Let $\{\mu_1\colon i\in I\}$ be an arbitrary family of ergodic probability measures, all distinct. Then these measures μ_i are mutually singular; there exist pairwise disjoint measurable subsets $\{P_i\colon i\in I\}$ invariant under f and such that $\mu_i(P_i)=1$ for every $i\in I$.

Now assume that $f\colon M\to M$ is a continuous transformation in a topological space M.

Definition 4.11. We say that f is *transitive* if there exists some $x\in M$ such that $\{f^n(x)\colon n\in\mathbb{N}\}$ is dense in M.

Recall that a topological space M is called a *Baire space* if the intersection of any countable family of open dense subsets is dense in M. Every complete metric space is a Baire spoace and the same is true for every locally compact topogoical space.

Lemma 4.12. Let M be a Baire space with a countable basis of open sets. Then $f: M \to M$ is transitive if and only if for every pair of open sets U and V there exists $k \geq 1$ such that $f^{-k}(U)$ intersects V.

Proposition 4.13. Let M be a Baire space with a countable basis of open sets. If μ is an ergodic probability measure then the restreiction of f to the support of μ is transitive.

§ Example of Ergodic Systems

§ Conservative Dynamics

As we have seen in previous sections, non-ergodic systems are quite common in the realm of Hamiltonian flows and symplectic transformations. However this fact alone is not sufficient to invalidate the ergodic hypothesis of Boltzmann in the context where it was formulated. Indeed, ideal gases are of a special class of systems and it is conceivable that ergodicity could be typical in this more restricted setting, even if it is not typical for general Hamiltonian systems.

In the 1960's, the Russian mathematician and theoretical physicist Yakov Sinai [Sin63] conjectured that Hamiltonian systems formed by spherical hard balls that hit each other elastically are ergodic. Hard ball systems (see Example 4.4.13 for a precise definition) had been proposed as a model for the behavior of ideal gases by the American scientist Josiah Willard Gibbs who, together with Boltzmann and Scottish mathematician and theoretical physicist James Clark Maxwell, created the area of statistical mechanics. The *ergodic hypothesis of Boltzmann-Sinai*, as Sinai's conjecture is often referred to, is the main topic in the present section.

In fact, we are going to discuss the problem of ergodicity for somewhat more general systems, called *billiards*, whose formal definition was first given by Birkhoff in the 1930's. In its simplest form, a billiard is given by a bounded connected domain $\Omega \subset \mathbb{R}^2$, called the *billiard table*, whose boundary $\partial\Omega$ is formed by a finite number of differentiable curves. We call the *corners* those points of the boundary where it fails to be differentiable; by hypothesis, they consistitute a finite set $\mathcal{C} \subset \partial\Omega$. One considers a point particle moving uniformly, along straight lines inside Ω, with elastic reflections on the boundary. That is, whenever the particle hits $\partial\Omega \setminus \mathcal{C}$ it is reflected in such a way that the angle of incidence equals the angle of reflection. When the particle hits some corner it is absorbed: its trajectory is not defined from then on.

Let us denote by $\mathbf{n}$ the unit vector field orthogonal to the boundary $\partial\Omega$ and pointing to the inside of Ω. It defines an orientation in $\partial\Omega \setminus \mathcal{C}$: a vector t tangent the boundary is *positive* if the basis $\{t, \mathbf{n}\}$ of $\mathbb{R}^2$ is positive. It is clear that the motion of the particle is characterized completely by the sequence of collisions with the boundary. Moreover, each such collision may be described by the position $s \in \partial\Omega$ and the angle of reflection $\theta \in (-\pi/2, \pi/2)$. Therefore, the evolution of the billiard is governed by the transformation

$$f: (\partial\Omega \setminus \mathcal{C}) \times (-\pi/2, \pi/2) \to \partial\Omega \times (-\pi/2, \pi/2), \quad (*)$$

that associates with each collision (s, θ) the subsequent one (s', θ'). See Figure 4.6 in 4.4.6 (page 135) of Viana-Oliviara.

In the example on the left-hand side of Figure 4.6 the billiard table is a polygon, that is, thye boundary consists of a finite number of straight line segments. The one trajectory represented in the figure hits one of the corners. Nearby trajectories, to their side, collide with distinct boundary segments, with very diffferent angles of incidence. In particular, it is clear that the Billiard transformation cannot be continuous. Discontinuities may occur in the absence of corners. For example, on the right-hand side of Figure 4.6 the boundary has four connected components, all of which are differentiable curves. Consider the trajectory represented in the figure, tangent to the one of the boundary components. Nearby trajectories, to either side hit with different boundary components. Consequently, the billiard map is discontinuous in this case also.

Example 4.14. *(circular billiard table)* On the left-hand side of Figure 4.7 (Viana-Oliviara 4.4.6 page 136) where there's a circular billiard table, we represent a billiard in the unit ball $\Omega \subset \mathbb{R}^2$. The corresponding billard transformation is given by $f\colon (s,\theta) \mapsto (s-(\pi-2\theta),\theta)$. The behavior of this transformation is described geometrically on the right-hand side of Figure 4.7. Observe that f preserves the area measure $dsd\theta$ and satisfies the twist condition (4.4.4). Note that f is integrable (in the sense of Section 4.4.2) and, in particular, the area measure is not ergodic. We will see in a while (Theorem 4.16) that every planar billiard preserves a natural measure equivalent to the area measure on $\partial\Omega \times (-\pi/2,\pi/2)$. Then, using the previous observations, the KAM theory allows us to prove that billiards with almost circular tables are not ergodic with respect to that invariant measure.

The definition of billiard extends immediately to bounded connected domains Ω in any Euclidean space $\mathbb{R}^d, d \geq 1$, whose boundary consists of a finite number of differentiable hypersurfaces intersecting each other along submanifolds with codiminsion larger than 1. We denote by $\mathcal{C}$ the union of the submanifolds. As before, we endow $\partial\Omega$ with the orientation induced by the unit vector $\mathbf{n}$ orthogonal to the boundary and pointing to the "inside" of Ω.

Elastic reflections on the boundary are defined by the following two conditions:

(i) The incident trajectory segment, the reflected trajectory segment and the orthogonal vector $\mathbf{n}$ are co-planar.
(ii) the angle of incidence equals the angle of reflection.

The billiard transformation is defined as in (*), having the domain

$$\{(s,v) \in (\partial\Omega \setminus \mathcal{C}) \times S^{d-1}\colon v \times \mathbf{n}(s) > 0\}.$$

Even more generally, we may take as a billiard table any bounded connected domain in a Riemannian surface, whose boundary is formed by a finite number of differentiable hypersurfaces intersecting along higher codimension submanifolds. These definitions are analogous, except that the trajectories between consecutive reflections on the boundary are given by segments of geodesics and angles are measured according to the Riemannian metric on the manifolds.

Example 4.15. *(Ideal gases and billiards)* Ideally, a gas is formed by a large number N of molecules ($N \approx 10^{27}$) that move uniformly along straight lines, between collision, and collide with each other elastically. Check the right-hand side of Figure 4.8 (page 137). For simplicity, let us assume that the molecultes are identical spheres and they are contained in the torus of dimension

$d \geq 2$. Let us also assume that all the molecules move with constant unit speed. This system can be modelled by a billiard, as follows.

For $1 \leq i \leq N$, denote by $p_i \in \mathbb{T}^d$ the position of the center of the i-th molecule M_i. Let $p > 0$ be the radius of each molecule. Then, each state of the system is entirely described by a value of $p = (p_1, \ldots, p_N)$ in the set

$$\Omega = \{p = (p_1, \ldots, p_N) \in \mathbb{T}^{Nd} \colon ||p_i - p_j|| \geq 2p \text{ for every } i \neq j\}$$

(this set is connected, as long as the radius ρ is sufficiently small).

In the absence of collisions, the point p moves along a straight line inside Ω, with constant speed. When two molecules M_i and M_j collide, $||p_i - p_j|| = 2p$ and the velocity vectors change in the following way. Let v_i and v_j be the velocity vectors of the two molecultes immediately before the collision and let R_{ij} be the straight line through p_i and p_j. the elasticity hypothesis means that the velocity vectors v_i' and v_j' immediately after the collision are given by (check the right-hand side of Figure 4.8):

(i) The components of v_i and v_i' in the direction R_{ij} are symmetric and same is true for v_j and v_j';

(ii) The components of v_i and v_i' in the direction orthogonal to R_{ij} are equal and the same is true for v_j and v_j'.

This means, precisely, that the point p undergoes elastic reflection on the hypersurface $\{p \in \partial\Omega : ||p_i - p_j|| = 2p\}$ of the boundary of Ω (see Exercise 4.4.4). Therefore, the motion of the point p corresponds exactly to the evolution of the billiard in the table Ω.

The next result places billiards well inside the domain of interest of ergodic theory. Let ds be the volume measure induced on the boundary $\partial\Omega$ by the Riemannian metric of the ambient manifold; in the planar case (that is, when $\Omega \subset \mathbb{R}^2$), ds is just the arc-length. Denote by $d\theta$ the angle measure on each hemisphere $\{v \in S^{d-1} \colon v \cdot \mathbf{n}(s) > 0\}$.

Theorem 4.16. The transformation f preserves the measure $v = \cos\theta ds d\theta$ on the domain $\{(s, v) \in \partial\Omega \times S^{d-1} \colon v \cdot \mathbf{n}(s) > 0\}$.

We call a billiard *dispersing* if the boundary of the billiard table is strictly convex at every point, when viewed from the inside. In the planar case, with the orientation conventions that we adopted, this means that the curvature κ is negative at every point. Figure 4.10 (page 139) presents two examples. In the first one, $\Omega \subset \mathbb{R}^2$ and the boundary is a connected set formed by the union of five differentiable curves. In the second example, $\Omega \subset \mathbb{T}^2$ and the boundary has three connected components, all of which are differentiable and convex.

The class of dispersing billiards was introduced by Sinai in his 1970 article [Sin70]. The denomination "dispersing" refers to the fact that in such billiards any (thin) beam of parallel trajectories becomes divergent upon reflection on the boundary, as illustrated on the left-hand side of Figure 4.10. Sinai observed that dispersing billiards are hyperbollic systems, in a non-uniform sense: invariant sub-bundles exist at *almost every* point and we have the derivative is contracting in some places and expanding along other places *asymptotically*, that is, for sufficiently large irates.

5 Ergodic Decomposition

§ Ergodic Decomposition Theorem and Disintegration

Before stating the ergodic decomposition theorem, let us analyze a couple of examples that help motivate and clarify its content:

Example 5.1. Let $f\colon [0,1] \to [0,1]$ be given by $f(x) = x^2$. the Dirac measures δ_0 and δ_1 are invariant and ergodic for f. It is also clear that $x = 0$ and $x = 1$ are the unique recurrent points for f and so every invariant probability measure μ must satisfy $\mu(\{0,1\}) = 1$. Then $\mu = \mu(\{0\})\delta_0 + \mu(\{1\})\delta_1$ is a (finite) convex combination of the ergodic measures.

Example 5.2. Let $f\colon \mathbb{T}^2 \to \mathbb{T}^2$ be given by $f(x,y) = (x+y, y)$. The Lebesgue measure m on the torus is preserved by f. Observe that every horizontal circle $H = S^1 \times \{y\}$ is invariant under f the restriction $f\colon H_y \to H_y$ is the rotation R_y. Let m_y be the Lebesgue measure on H_y. Observe that m_y is also invariant under f. Moreover, m_y is ergodic whenever y is irrational. On the other hand, by the Fubini theorem,

$$m(E) = \int m_y(E)dy \text{ for every measurable set } E. \quad (5.1.1)$$

The identity is not affected if we consider the integral restricted to the subset of irrational values of y. Then presents m as an (uncountable) convex combination of ergodic measures.

Now, let us introcue some useful terminology. In what follows, $(M, \mathcal{B}, \mu)$ is a probability space and $\mathcal{P}$ is a partition of M into measurable subsets. We denote by $\pi\colon M \to \mathcal{P}$ the canonical projection that assigns to each point $x \in M$ the element $\mathcal{P}(x)$ of the partition that contains it. This projection map endows $\mathcal{P}$ with the structure of a probability space, as follows. Firstly, by definition, a subset $\mathcal{Q}$ of $\mathcal{P}$ is measurable if and only if its pre-image

$$\pi^{-1}(\mathcal{Q}) = \text{ union of all } P \in \mathcal{P} \text{ that belong to } \mathcal{Q}$$

is a measurable subset of M. It is easy to check that this definition is consistent: The family $\widehat{\mathcal{B}}$ subsets is a σ-algebra in $\mathcal{P}$. Then, we define the *quotient measure* $\widehat{\mu}$ by

$$\widehat{\mu}(\mathcal{Q}) = \mu(\pi^{-1}(\mathcal{Q})) \text{ for every } \mathcal{Q} \in \widehat{\mathcal{B}}.$$

Theorem 5.3. *(Ergodic decomposition)* Let M be a complete separable metric space, $f\colon M \to M$ be a measruable transfomration and μ be an invariant probability measure. Then there exist a measurable set $M_0 \subset M$ with $\mu(M_0) = 1$, a partition $\mathcal{P}$ of M_0 into measurable subsets and a family $\{\mu_P \colon P \in \mathcal{P}\}$ of probability measures on M, satisfying

(i) $\mu_P(P) = 1$ for $\widehat{\mu}$-almost every $P \in \mathcal{P}$;
(ii) $P \mapsto \mu_P(E)$ is measurable, for every measureable set $E \subset M$;
(iii) μ_P is invariant and ergodic for $\widehat{\mu}$-almost every $P \in \mathcal{P}$;
(iv) $\mu(E) = \int \mu_P(E) d\widehat{\mu}(P)$, for every measurable set $E \subset M$

Definition 5.4. A *disintegration* of μ with respect to a partition $\mathcal{P}$ is a family $\{\mu_P \colon P \in \mathcal{P}\}$ of probability measures on M such that, for every measurable set $E \subset M$:

(i) $\mu_P(P) = 1$ for $\widehat{\mu}$-almost every $P \in \mathcal{P}$;
(ii) the map $\mathcal{P} \to \mathbb{R}$, defined by $P \mapsto \mu_P(E)$ is measurable;
(iii) $\mu(E) = \int \mu_P(E) d\widehat{\mu}(P)$.

Recall that the partition $\mathcal{P}$ inherits from M a natural structure of probability space, wtih σ-algebra $\widehat{\mathcal{B}}$ and a probability measure $\widehat{\mu}$. the measures μ_P are called *conditional probabilities* of μ with respect to $\mathcal{P}$.

Example 5.5. Let $\mathcal{P} = \{P_1, \ldots, P_n\}$ be a finite partition of M into measurable subsets with $\mu(P_i) > 0$ for ervery i. The quotient measure $\widehat{\mu}$ is given by $\widehat{\mu}(\{P_i\}) = \mu(P_i)$. Consider the normalized restriction μ_i of μ to each P_i:

$$\mu_i(E) = \frac{\mu(E \cap P_i)}{\mu(P_i)} \text{ for every measurable set } E \subset M.$$

then $\{\mu_1, \ldots, \mu_n\}$ is a disintegration of μ with respect to $\mathcal{P}$: it is clear that $\mu(E) = \sum_{i=1}^n \widehat{\mu}(\{P_i\})\mu_i(E)$ for every measurable set $E \subset M$.

The construction extends immediately to countable partitions. In the next example we treat an uncountable case:

Example 5.6. Let $M = \mathbb{T}^2$ and $\mathcal{P}$ be the partition of M into horizontal circles $S^1 \times \{y\}$, $y \in S^1$. Denote by m_y the Lebesgue measure (arc-length) on each horizontal circle $S^1 \times \{y\}$. By the Fubini theorem

$$m(E) = \int m_y(E) d\widehat{m}(y) \text{ for every measurable set } E \subset \mathbb{T}^2.$$

Hence, $\{m_y \colon y \in S^1\}$ is a disintegration of M with respect to $\mathcal{P}$.

The next proposition asserts that disintegrations are essentially unique, when they exist.

Proposition 5.7. Assume that the σ-algebra $\mathcal{B}$ admits some countable generator. If $\{\mu_P \colon P \in \mathcal{P}\}$ are disintegrations of μ with respect to $\mathcal{P}$, then $\mu_P = \mu'_P$ for $\widehat{\mu}$-amost every $P \in \mathcal{P}$.

On the other hand, disintegrations may fail to exist:

Example 5.8. Let $f\colon S^1 \to S^1$ be an irrational rotation and $\mathcal{P}$ be the partition of S^1 whose elements are the orbits $\{f^n(x)\colon n \in \mathbb{Z}\}$ of f. Assume that there exists a disintegration $\{\mu_P\colon P \in \mathcal{P}\}$ of the Lebesgue measure μ with respect to $\mathcal{P}$. Consider the irates $\{f_*\mu_P\colon P \in \mathcal{P}\}$ of the conditional probabilities. Since the partition elements are invariant sets, $f_*\mu_P(P) = \mu_P(P) = 1$ for $\widehat{\mu}$-almost every P. It is clear that, given every measurable set $E \subset M$,

$$P \mapsto f_*\mu_P(E) = \mu_P(f^{-1}(E))$$

is a measurable function. Moreover, since μ is an invariant measure,

$$\mu(E) = \mu(f^{-1}(E)) = \int \mu_P(f^{-1}(E)) = \int \mu_P(f^{-1}(E))d\widehat{\mu}(P) = \int f_*\mu_P(E)d\widehat{\mu}(P).$$

There turns out to be no disintegration

The theorem of Roklin states that disintegrations always exist if the partition $\mathcal{P}$ is the limit of increasing sequence of countable partitions and the space M is reasonably well behaved.

§ Measurable Partitions, Rokhlin Disintegrations Theorem, Conditional Expectations

Definition 5.9. We say that $\mathcal{P}$ is a *measurable partition* if there exists some measurable set $M_0 \subset M$ with full measure such that, restricted to M_0,

$$\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$$

for some increasing sequence $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ of countable partitions (see also Exercise 5.1.1). By $\mathcal{P}_i \prec \mathcal{P}_{i+1}$ we mean that every lement of $\mathcal{P}_{i+1}$ is contained in some element of $\mathcal{P}_i$ or, equivalently, every element of $\mathcal{P}_i$ coincides with a union of elements of $\mathcal{P}_{i+1}$. Then we say that $\mathcal{P}_i$ is *coarser* than $\mathcal{P}_{i+1}$ or, equivalently, that $\mathcal{P}_{i+1}$ is *finer* than $\mathcal{P}_i$.

We represent by $\bigvee_{n=1}^{\infty} \mathcal{P}_n$ the partition whose elements are non-empty intersections of the form $\bigcap_{n=1}^{\infty} P_n$ with $P_n \in \mathcal{P}_n$ for every n. Equivalently, this is the courser partition such that

$$\mathcal{P}_n \prec \bigvee_{n=1}^{\infty} \mathcal{P}_n \text{ for every } n.$$

It follows immediately from the definition that every countable partition is measurable. It is easy to find examples of uncountable measurable partitions:

Example 5.10. Let $M = \mathbb{T}^2$, endowed with the Lebesgue measure m, and let $\mathcal{P}$ be the partition of M into horizontal circles $S^1 \times \{y\}$. Then $\mathcal{P}$ is a measurable partition. To see that, consider

$$\mathcal{P}_n = \{S^1 \times I(i,n)\colon i = 1, \dots, 2^n\},$$

where $I(i,n)$, $1 \le i \le 2^n$ is the segment of $S^1 = \mathbb{R}/\mathbb{Z}$ corresponding to the interval $[(i-1)f/2^n, i/2^n) \subset \mathbb{R}$. The sequence $(\mathcal{P}_n)_n$ is increasing and $\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$.

Example 5.11. Let $f\colon M \to M$ be a measurable transformation and μ be an ergodic probability measure. Let $\mathcal{P}$ be the partition of M whose elements are orbits of f. Then $\mathcal{P}$ is *not* measurable, unless f exhibits an orbit with full measure. Indeed, suppose that there exists a non-decreasing sequence $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ of countable partitions such that $\mathcal{P} = \bigvee_{n=1}^{\infty} \mathcal{P}_n$ restricted to some full measure subset. This last condition implies that almost every orbit of f is contained in some element P_n of the partition $\mathcal{P}_n$. In other words, up to measure zero, every element of $\mathcal{P}_n$ is invariant under f. In other words, up to measure zero, every element of $\mathcal{P}_n$ is invariant under f. By ergodicity, it follows that for every n there exists exactly on e$P_n \in \mathcal{P}_n$ such that $\mu(P_n) = 1$. Denote $P = \bigcap_{n=1}^{\infty} P_n$. then P is an element of the partition $\bigvee_{n=1}^{\infty} \mathcal{P}_n = \mathcal{P}$, that is, P is an orbit of f, and it has $\mu(P) = 1$.

Theorem 5.12. *(Rokhlin Disintegration)* Assume that M is a coplete separable metric space and $\mathcal{P}$ is a measurable partition. Then the probability measure μ admits some disintegration with respect to $\mathcal{P}$.

We shall now talk about conditional expectations. We use $\mathcal{P}_n(x)$ to denote the element of $\mathcal{P}_n$ that contains a given point $x \in M$.

Let $\psi\colon M \to \mathbb{R}$ be any bounded measurable function. For each $n \geq 1$, define $e_n(\psi)\colon M \to \mathbb{R}$ as follows:

$$e_n(\psi, x) = \begin{cases} \frac{1}{\mu(\mathcal{P}_n(x))} \int_{\mathcal{P}_n(x)} \psi d\mu & \text{if } \mu(\mathcal{P}_n(x)) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Since the partitions $\mathcal{P}_n$ are countable, the second case of the definition corresponds to a subset of points with total measure zero. Observe also that $e_n(\psi)$ is constant on each $P_n \in \mathcal{P}_n$; let us denote by $E_n(\psi, P_n)$ the value of that constant. Then

$$\int \psi d\mu = \sum_{P_n} \int_{P_n} \psi d\mu = \sum_{P_n} \mu(P_n) E_n(\psi, P_n) = \int e_n(\psi) d\mu,$$

for every $n \in \mathbb{N}$ (the sums involve only partition elements $P_n \in \mathcal{P}_n$ with positive measure).

Lemma 5.13. given any bounded measurable function $\psi\colon M \to \mathbb{R}$, there exists a subset M_ψ of M with $\mu(M_\psi) = 1$ such that

(i) $e(\psi, x) = \lim_n e_n(\psi, x)$ exists for every $x \in M_\psi$;
(ii) $e(\psi)\colon M_\psi \to \mathbb{R}$ is measurable and constant on each $P \in \mathcal{P}$;
(iii) $\int \psi d\mu = \int e(\psi) d\mu$.

The hypothesis that the ambient space M is complete separable metric space is used in the proof of the important criterion for σ-additivity that we now state and prove:

Proposition 5.14. Let M be a complete and separable metric space and $\mathcal{A}$ be an algebra generated by a countable basis $\mathcal{U} = \{U_k \colon k \in \mathbb{N}\}$ of open sets of M. Let $\mu\colon \mathcal{A} \to [0, 1]$ be an additive function with $\mu(\emptyset) = 0$. then μ extends to a probability measure on the Borel σ-algebra of M.

First, let us outline the proof. We consider the product space $\Sigma = \{0,1\}^{\mathbb{N}}$, endowed with the topology generated by the cylinders

$$[0; a_0, \ldots, a_s] = \{(i_k)_{k\in\mathbb{N}} \colon i_0 = a_0, \ldots, 1_s = a_s\}, \quad s \geq 0.$$

Note that Σ is compact. Using the fact that M is a complete metric space, we will show that the map

$$\gamma \colon M \to \Sigma, \quad \gamma(x) = (\mathbb{1}_{U_k}(x))_{k\in\mathbb{N}}$$

is a measurable embedding of M inside Σ.

Lemma 5.15. The image $\gamma(M)$ is a Borel subset of Σ.

Corollary 5.16. The map $\gamma \colon M \to \gamma(M)$ is a measurable bijection whose inverse is also measurable.

6 Unique Ergodicity

§ Unique Ergodicity and Minimality

Definition 6.1. We say that a transformation $f: M \to M$ is *uniquely ergodic* if it admits exactly one invariant probability measure.

Proposition 6.2. 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 continuous function $\varphi: M \to \mathbb{R}$, 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 $\varphi: M \to \mathbb{R}$, the sequence of time averages $n^{-1} \sum_{j=0}^{n-1} \varphi \circ f^j$ converges uniformly to a constant.

Definition 6.3. Let $\Lambda \subset M$ be a closed invariant set of $f: M \to M$. 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.

Proposition 6.4. If $f: M \to M$ is uniquely ergodic then the support of the unique invariant probability measure μ is a minimal set.

The converse to Proposition 6.2.1 is false in general:

Proposition 6.5. *(Furstenberg)* There exists some real-analytic diffeomorphism $f: \mathbb{T}^2 \to \mathbb{T}^2$ that is minimal, preserves the Lebesgue measure m on the torus, but it is not ergodic for m. In particular, f is not uniquely ergodic.

§ Haar Measure

Fix $d \geq 1$ and a rationally independent vector $\theta = (\theta_1, \ldots, \theta_d)$. As we have seen in Section 4.2.1, the rotation $R_\theta \colon \mathbb{T}^d \to \mathbb{T}^d$ is ergodic with respect to the Lebesgue measure m on the torus. We find that R_θ is uniquely ergodic.

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

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

are continuous. In all the follows it is assumed that the topology 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.*

Example 6.6. 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 $x \sim y \iff x^{-1}y \in H$. Denote by xH the equivalence class that contains each $x \in G$. Consider the following group operation in G/H:

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

Example 6.7. The set $G = GL(d, \mathbb{R})$ of invertible real matrices of, 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* $SL(d, \mathbb{R})$, consisting of the matrices with determinant 1, and the *orthogonal group* $O(d, \mathbb{R})$, formed by the orthogonal matrices.

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

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

An *endomorphism* of G is a continuous map $\phi \colon G \to G$ 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 6.8. Let $A \in GL(d, \mathbb{Z})$; in other words, A is an invertible matrix of dimension d with integer coefficients. Then A induces an endomorphism $f_A \colon \mathbb{T}^d \to \mathbb{T}^d$. 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 compaqct neighborhood. For example, every Lie group is locally compact. On the other hand, the additive group of rational numbers, with the topology inhertied from the real line, is not locally compact.

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

Theorem 6.9. *(Haar)* Let G be a locally compact topological group. Then:

(i) There exists some Borel meausre μ_G on G that is invariant under all left-translation, 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.

The Haar measure (the measure in the theorem above) featuress some additional properties:

Theorem 6.10. Assume that G is compact. Then the Haar measure μ_G is invariant under right-translations and under every surjective endomorphism of G.

Definition 6.11. We call a distance d in a topological group G *left-invariant* if it is invarinat 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.

Lemma 6.12. 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.

Example 6.13. Given a matrix $A \in GL(d, \mathbb{R})$, denoted by $||A||$ its operator norm, that is $||A|| = \sup\{||Av|| \colon ||v|| = 1\}$. Observe that $||OA|| = ||A|| = ||AO||$ for every O in the orthogonal group $O(d, \mathbb{R})$. Define

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

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

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

for every $C \in GL(d, \mathbb{R})$. This distance is not invariant under right-translations in $GL(d, \mathbb{R})$. However, it is right-invariant (and left-invariant) restricted to the orthogonal group $O(d, \mathbb{R})$.

Theorem 6.14. 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.

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 dynmaic, 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 $X = \{0, 1, \ldots, d-1\}$, endowed with the discrite 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

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

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 $+\colon M \times M \to M$ defined in this way turns M into an abelian topological group and the distance is invariant under all the translations.

Now consider the "translation by 1" $f\colon M \to M$ defined by

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

where $k > 0$ is the smallest value of n such that $\alpha_n < d - 1$.

§ Theorem of Weyl

In this section we use ideas that were discussed previously to prove a beautiful theorem of Hermann Weyl.

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

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d.$$

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

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

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

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

Theorem 6.16. *(Weyl)* If at least one of the coefficients $a_1, a_2, \ldots, a_d$ is irrational then the sequence $z_n = P_*(n)$, $n \in \mathbb{N}$ is equidistributed.

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

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

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

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

Proposition 6.17. The Lebesgue measure on $\mathbb{T}^d$ is ergodic for f.

Proposition 6.18. The transformation f is uniquely ergodic: the Lebesgue measure on the torus is the unique invariant probability measure.

Lemma 6.19. For any probability measure μ invariant under f, the projection $\pi_* \mu$ coincides with the Lebesgue measure m_0 on $\mathbb{T}^{d-1}$.

Lemma 6.20. If $\theta_0 \in G_0(m)$ then $\{\theta_0\} \times S^1$ is contained in $G(m)$.

Corollary 6.21. 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 $\psi\colon \mathbb{T}^d \to \mathbb{R}$.

To look at the proof of Weyl's theorem, we introduce the polynomial functions $p_1, \dots, p_d$ defined by

$$p_d(x) = P(x) \text{ and } p_{j-1}(x) = p_j(x+1) - p_j(x) \text{ for } j = 2, \dots, d.$$

Lemma 6.22. The polynomial $p_j(x)$ has degree j, for every $1 \leq j \leq d$. Moreover, $p_1(x) = \alpha x + \beta$ with $\alpha = d! a_d$.

Lemma 6.23. For every $n \geq 0$,

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