ICTP-ESF School and Conference on Dynamical Systems

Krerley Oliveira

Contents

Chapter 1

Invariant Measures

[1]

1.1 σ-algebras and Measures

Given a set X and a family $\mathcal{B}$ of subsets of X, we say that $\mathcal{B}$ is a *σ-algebra* if it is closed for the elementary operations of sets and contains X. That is

- $X \in \mathcal{B}$
- $A \in \mathcal{B}$ implies $A^c \in \mathcal{B}$
- $A_j \in \mathcal{B}$ for $j = 1, 2, \ldots$ implies $\bigcup_{j=1}^{\infty} A_j \in \mathcal{B}$.

It is easy to see that a σ-algebra $\mathcal{B}$ is also closed for countable intersection: in fact, if $A_j \in \mathcal{B}$ for $j = 1, 2, \ldots, n, \ldots$ then $\cap_{j=1}^{\infty} A_j = \left(\cup_{j=1}^{\infty} A_j^c\right)^c$ is also in $\mathcal{B}$.

Definition 1.1. A *measurable space* is a pair $(X, \mathcal{B})$ where X is a set and $\mathcal{B}$ is a σ-algebra of subsets of X. The elements of $\mathcal{B}$ are called *measurable sets*.

We give some examples of σ-algebras.

Example 1.2. Let X be any set.

1. Denote by 2^X the family of all the subsets of X. Then $\mathcal{B} = 2^X$ is clearly a σ-algebra.

[1]These notes are intended to help the audience in the ICTP-ESF School and Conference on Dynamical Systems, held in May, 2012 at Trieste-Italy. Along the preparation of this material, I used part of a ongoing project of a book, written in colaboration with Professor Marcelo Viana, whose I am deeply indebt. I would like to thanks also Xiaochuan Liu for the help with the translation from the original portuguese to english. Any left mistake or missprint is my entire responsability.

The author is thankful to the organizers of the School and Conference an Dynamical Systems 2012, held at Trieste for its exceptional scientific environment. Special thanks goes to Marcelo Viana and Stefano Luzatto for encouragement, and to PRONEX-Dynamical Systems/CNPq-FAPERJ and ICTP-ESF for financial support.

2. $\mathcal{B} = \{\emptyset, X\}$ is also a σ-algebra.

Note that if $\mathcal{B}$ is an algebra of X then $\{\emptyset, X\} \subset \mathcal{B} \subset 2^X$. Therefore $\{\emptyset, X\}$ is is the smallest algebra and 2^X is the largest algebra of X.

Proposition 1.3. *Consider any non-empty family $\{\mathcal{B}_i : i \in \mathcal{I}\}$ of σ-algebra ($\mathcal{I}$ is any set, just to index the elements of the family). Then the intersection $\mathcal{B} = \cap_{i \in \mathcal{I}} \mathcal{B}_i$ is also a σ-algebra.*

Proof. Check it! □

Now, given any set $\mathcal{E}$ of subsets of X, we can apply Proposition 1.3 to the family of all σ-algebras containing $\mathcal{E}$. Note that this family is not empty, since it contains the σ-algebra 2^X. According to previous observation, the intersection of all these σ-algebras, is also an σ-algebra, which surely contains $\mathcal{E}$. Moreover, the way it is constructed, it is contained in all σ-algebras containing E. Therefore it is the smallest σ-algebra that contains $\mathcal{E}$. This leads to the following definition:

Definition 1.4. The *generating σ-algebra* of a family $\mathcal{E}$ of subsets of X is the smallest σ-algebra $\sigma(\mathcal{E})$ which contain the family $\mathcal{E}$, i.e., is the intersection of all the σ-algebras that contains $\mathcal{E}$.

Example 1.5. The *Borel σ-algebra* of a topological space is the σ-algebra $\sigma(\tau)$ generated by topology τ, i.e., the smallest σ-algebra that contains all the open subsets. In this case, the measurable sets are called *Borel sets.*

Now we introduce the concept of measure.

Definition 1.6. A (positive) *measure* on a measurable space $(X, \mathcal{B})$ is a (positive) function, which satisfies:

1. $\mu(\emptyset) = 0$;
2. $\mu(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j)$ for any $A_j \in \mathcal{B}$ pairwise disjoint.

The triple $(X, \mathcal{B}, \mu)$ is called *measure space.* When $\mu(X) < \infty$ holds we say that μ is a *finite* measure and if $\mu(X) = 1$ we say that μ is a *probability measure*. In the later case, $(X, \mathcal{B}, \mu)$ is a *probability space.*

1.1.1 Bernoulli measures

Given two measure spaces $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$, it is possible to construct a σ-algebra on Cartesian product $X_1 \times X_2$ using $\mathcal{A}_1$ and $\mathcal{A}_2$ in a natural way. Consider in $X_1 \times X_2$ the σ-álgebra generated by the family of all the products of the form $A_1 \times A_2$ with $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. It is called *product σ-algebra* and is denoted by $\mathcal{A}_1 \otimes \mathcal{A}_2$. Then, define

$$(\mu_1 \times \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),$$

for $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. There exists a unique measure $\mu_1 \times \mu_2$ defined on the σ-algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ which extend the previous equality. This measure is called the *product measure* of μ_1 by μ_2.

This construction can be generalized to the product of any finite family of measure spaces $X_1, X_2, \ldots, X_n$. For our purpose it is useful to go further and define the product of a countable family of measure spaces.

Consider measure spaces $(X_i, \mathcal{B}_i, \mu_i)$, com $i \in \mathcal{I}$, where the set of indices can be either $\mathcal{I} = \mathbb{N}$ or $\mathcal{I} = \mathbb{Z}$. Consider the Cartesian product

$$\Sigma = \prod_{i \in \mathcal{I}} X_i = \{(x_i)_{i \in \mathcal{I}} : x_i \in X_i\}. \tag{1.1}$$

A *cylinder* of Σ is any set of the form

$$[m; A_m, \ldots, A_n] = \{(x_i)_{i \in \mathcal{I}} : x_i \in A_i \text{ for } m \leq i \leq n\} \tag{1.2}$$

where $m \in \mathcal{I}$, $n \in \mathbb{N}$ and $A_i \in \mathcal{B}_i$ for $m \leq i \leq n$. By definition, the *product σ-algebra* in Σ is the σ-algebra $\mathcal{B}$ generated by the family of all the cylinders. Note that the family $\mathcal{B}_0$ of finite unions of pairwise cylinders is an algebra (which generates $\mathcal{B}$). Now, we define the *product measure* μ on Σ. Firstly, consider the map μ defined on the family of cylinders by

$$\mu([m; A_m, \ldots, A_n]) = \prod_{i=m}^{n} \mu_i(A_i). \tag{1.3}$$

Then extend μ to the algebra $\mathcal{B}_0$, stipulating that the measure of any finite union of pairwise disjoint cylinders is equal to the sum of the measures of the cylinders. This extension is well defined and is σ-additive.

At last, we may extend μ to a measure on $(\Sigma, \mathcal{B})$. The measure space $(\Sigma, \mathcal{B}, \mu)$ constructed in this way is called *product* of the spaces $(X_i, \mathcal{B}_i, \mu_i)$, $i \in \mathcal{I}$.

A very important case of the above construction occurs when the spaces $(X_i, \mathcal{B}_i, \mu_i)$ are all equal to a given $(X, \mathcal{C}, \nu)$. These systems model sequences of identical random experiments in which the outcome of each experiment is independent with the others. Assume that each experiment takes values in the set X, with probability distribution equal to ν. In this case the measure μ is given by

$$\mu([m; A_m, \ldots, A_n]) = \prod_{i=m}^{n} \nu(A_i). \tag{1.4}$$

It is called *Bernoulli measure* defined by ν. Note that (1.4) means that the probability measure of $\{x_m \in A_m, \ldots, x_n \in A_n\}$ is the product of the probability measures of the events $x_i \in A_i$. Therefore, it translates that independence of the successive experiments.

In the case that $X = \{1, \ldots, d\}$ is a finite set, equipped with the σ-algebra 2^X, this construction becomes a little simpler. For example, it suffices to only consider elementary cylinders

$$[m; a_m, \ldots, a_n] = \{(x_i)_{i \in \mathcal{I}} \in X : x_m = a_m, \ldots, x_n = a_n\}, \tag{1.5}$$

where each A_j consists of a unique point a_j. In fact, each cylinder is a unique finite union of pairwise disjoint elementary cylinders. Moreover, the relation (1.4) can be written as

$$\mu([m; a_m, \dots, a_n]) = p_{a_m} \cdots p_{a_n} \tag{1.6}$$

where $p_a = \nu(\{a\})$ for any $a \in X$.

1.1.2 Bernoulli shifts

Let $(X, \mathcal{C}, \nu)$ be any probability measure space. In this section we consider the product space $\Sigma = X^{\mathbb{N}}$, equipped with the product σ-algebra $\mathcal{B} = \mathcal{C}^{\mathbb{N}}$ and the product measure $\mu = \nu^{\mathbb{N}}$, which were defined in Section 1.1.1. This means that M is the set of all the sequences $(x_n)_{n\in\mathbb{N}}$ with $x_n \in X$ for any n. By definition, $\mathcal{B}$ is the σ-algebra generated by the cylinders

$$[m; A_m, \dots, A_n] = \{(x_i)_{i\in\mathcal{I}} : x_i \in A_i \text{ for } m \leq i \leq n\}$$

where $m \leq n$ and each A_i is an element of $\mathcal{C}$. Moreover, μ is characterized by

$$\mu([m; A_m, \dots, A_n]) = \prod_{i=m}^{n} \nu(A_i). \tag{1.7}$$

You can think of elements of Σ as representing the results of sequences of experiments governed by the same probability distribution ν:

Given any measurable set $A \subset X$, the probability of getting $x_i \in A$ is equal to $\nu(A)$, mo matter what i is. Moreover, the results of successive experiments are independent: in fact the relation (1.7) means that the probability of $x_i \in A_i$ for any $m \leq i \leq n$ is the product of the probability of each of the events $x_i \in A_i$ separately.

In this section we induce a map $\sigma : \Sigma \to \Sigma$ in the space Σ, called *Bernoulli shift*, which preserves the measure μ. It is worth noticing that you can replace $\mathbb{N}$ by $\mathbb{Z}$ in the construction, that is, you can consider Σ as being the space of the bilateral sequences $(\dots, x_{-n}, \dots, x_0, \dots, x_n, \dots)$. With minor adjustments, we leave this task to the reader, everything that will be said then remains valid in this case. Moreover, in the bilateral case the Bernoulli shift is a invertible map.

The *Bernoulli shift* is a pair (σ, μ) where $\sigma : \Sigma \to \Sigma$ is the map defined by

$$\sigma\big((x_n)_n\big) = (x_{n+1})_n.$$

In other words, σ sends the sequence $(x_0, x_1, \dots, x_n, \dots)$ to the sequence $(x_1, \dots, x_n, \dots)$.

When X is a topological space, and $\mathcal{C}$ is its Borel σ-algebra, we can endow Σ with the *product topology* which is, by definition, the topology generated by the cylinders $[m; A_m, \dots, A_n]$ where the sets $A_m, \dots, A_n$ are open sets of X. the property (1.10) implies that the shift $\sigma : \Sigma \to \Sigma$ is continuous for this topology. Tychonoff's theorem (see [Dug66]) states that Σ is compact if X is compact.A particular important case occurs when X is s finite set equipped with the discrete topology, in which each subset is open set.

In this case, the topology is metrizable, i.e., we may define a metric such that the topology generated by the balls is exactly the topology defined by the cylinders. In order to do it, we proceed as follows:

Fix any number $0 < \theta < 1$. Given two sequences $x = (x_0, x_1, \dots)$ and $y = (y_0, y_1, \dots)$, consider $k(x,y) = \min\{\ell \geq 0; x_\ell \neq y_\ell\}$. Define:

$$d(x,y) = \theta^{k(x,y)},$$

and $d(x,x) = 0$. One can show that:

- The function d defined as above is a metric on Σ is a metric;
- If the symbol space X is finite, the metric space (Σ, d) is compact;
- If $X = \mathbb{N}$, the metric space (Σ, d) is not compact;
- The topology induced by d is the product topology.
- Similar properties hold true for the two-sided case ($X^{\mathbb{Z}}$ and d defined in a similar fashion).

We leave the proof of the claims above as exercises for the reader. We say that a transformation $f : M \to M$ is *transitive* if there exists $x \in M$ whose trajectory $f^n(x)$, $n \geq 0$ is dense in Σ.

Proposition 1.7. *Let X be a finite set and $\Sigma = X^{\mathbb{N}}$ or $\Sigma = X^{\mathbb{Z}}$. Then the shift $\sigma : \Sigma \to \Sigma$ is a continuous map and transitive. Moreover, the set of periodic points of σ is dense in Σ.*

Proof. We leave the proof of the next result to the reader.

□

1.2 Invariant Measures

Let $(M, \mathcal{B}, \mu)$ be a measure space, i.e., a set M with a collection of sets $\mathcal{B}$, closed by countable union and intersections (called σ-algebra) and a positive function $\mu : \mathcal{B} \to \mathbb{R}$. The sets of $\mathcal{B}$ are called *measurable sets.*

A *measurable map* is a function $f : M \to M$ such that $f^{-1}(B) \in \mathcal{B}$ for all $B \in \mathcal{B}$. Let us start by defining the invariant measure of a measurable transformation $f : M \to M$.

Definition 1.8. The measure μ is invariant under f (we also say f preserves μ) if

$$\mu(E) = \mu(f^{-1}(E)) \quad \text{for all measurable set } E \subset M. \tag{1.8}$$

Heuristically, this means that the probability of being in a given set and the likelihood that the image of it is in this set are equal. Note that the definition makes sense, since the pre-image of a measurable set of a transformation is still a measurable set.

One very first question is about the existence of an invariant measure. For continuous maps, the best result that we can get is:

Theorem 1.9. *Let $f : M \to M$ be a continuous transformation on a compact metric space. Then there exists at least one probability measure in M which is invariant under f.*

The compactness and continuity are essencial in the statement of the previous theorem. To see it, consider the function $f : (0,1] \to (0,1]$ given by $f(x) = x/2$. Suppose that f admits some probability invariant measure; the goal is to show that this can't happen. By the recurrence theorem, for this probability almost every point of $(0,1]$ is recurrent. But it is clear that there exists no recurrent point: the orbit of any point $x \in (0,1]$ converges to zero and, in particular, doesn't converge to the starting point x. Therefore, f is an example of continuous transformation on a not compact space which does not admit any probability measure.

Modifying the example slightly, we can show that the same phenomenona can occur in compact spaces, if the transformation is not continuous. Consider $f : [0,1] \to [0,1]$ given by $f(x) = x/2$ if $x \neq 0$ e $f(0) = 1$. By the same reasons as before, no point $x \in (0,1]$ is recurrent. therefore, if there exists some invariant probability measure μ it has to give the total weight to the single recurrent point, which is $x = 0$. In other words, μ must be the Dirac measure δ_0 supported at zero, which is defined by

$$\delta_0(E) = 1 \text{ if } 0 \in E \quad \text{and} \quad \delta_0(E) = 0 \text{ if } 0 \notin E.$$

But the measure δ_0 is not invariant under f: Taking $E = \{0\}$ we have that E have measure 1 but its pre-image $f^{-1}(E)$ is the empty set, which have measure zero. Therefore, This transformation does not admit any invariant probability measure neither.

Our third example is a little different in nature. Consider $f : [0,1] \to [0,1]$ given by $f(x) = x/2$. It is a continuous transformation on a compact space. Then, the theorem that we will prove, admit some invariant probability measure. By the same arguments we used in the previous case, we conclude that in fact there is a unique invariant probability measure, which is the Dirac measure δ_0 supported on point zero. Note that in this case the measure δ_0 is in fact invariant.

We mentioned this last case to emphasize the limitations of Theorem 1.9 (which are inherent to its great generality): the measures which it ensures to exist can be quite trivial; for example, in this case when it comes "almost all points" we are referring only to the point $x = 0$. Therefore, an important objective in Ergodic Theory is to find a invariant measure which is more sophisticated, with additional properties (for example, to be equivalent to Lebesgue measure) that makes them more interesting.

In order to get a better understanding of invariant measures, one useful proposition that characterizes the invariance is given by

Proposition 1.10. *Let $f : M \to M$ be a measurable transformation and μ is a measure on M. Then f preserves μ, if and only if,*

$$\int \phi \, d\mu = \int \phi \circ f \, d\mu. \tag{1.9}$$

for all μ-integral function $\phi : M \to \mathbb{R}$.

Proof. Suppose that the measure is invariant. We will show that the relation (1.9) is valid for characteristic functions. First, note that by hypothesis $\mu(B) = \mu(f^{-1}(B))$ for every measurable set B. Since,

$$\int \mathcal{X}_B \, d\mu = \mu(B) \quad \text{and} \quad \mu(f^{-1}(B)) = \int (\mathcal{X}_B \circ f) \, d\mu,$$

this means (1.9) holds for all the characteristic functions. Then, by linearity of the integral, (1.9) also holds for simple functions. Then we will use an approximation argument to conclude that (1.9) holds for all integrable functions. Given any integrable function $\phi : M \to \mathbb{R}$, consider a sequence $(s_n)_n$ of simple functions converging to ϕ for which $|s_n| \leq |\phi|$ for all n. Then, using the dominated convergence theorem twice:

$$\int \phi \circ f \, d\mu = \lim \int s_n \circ f \, d\mu = \lim \int (s_n \circ f) \, d\mu = \int (\phi \circ f) \, d\mu.$$

This shows that (1.9) holds for every integrable function and it is invariant. The converse also follows immediately, taking ϕ the characteristic function of a measurable set B.

□

Now, we discuss a very interesting result, *the Poincaré Recurrence Theorem.* It states that for any finite invariant measure, almost every point of any measurable set E returns to E for a infinite number of times:

Theorem 1.11 (Poincaré Recurrence Theorem)**.** *Let $f : M \to M$ be a measurable transformation and μ is a finite measure which is invariant under f. Let $E \subset M$ be any measurable set with $\mu(E) > 0$. Then, for μ-almost all point $x \in E$, there exist infinite numbers of n for which $f^n(x)$ also lie in E.*

Proof. Denote by E_0 the set of points $x \in E$ which never returns E We first prove this set is with measure zero. For this, we begin by observing that their pre-images $f^{-n}(E_0)$ are pairwise disjoint. In fact, suppose that there are $m > n \geq 1$ such that $f^{-m}(E_0)$ intersect $f^{-n}(E_0)$. Let x be a point in the intersection and let $y = f^n(x)$. Then $y \in E_0$ and $f^{m-n}(y) = f^m(x) \in E_0$, which is contained in E. This means that y comes back at least once into E, which contradicts with the definition of E_0. This contradiction proves that the pre-images of E_0 are disjoint, as stated.

Observe that $\mu(f^{-n}(E_0)) = \mu(E_0)$ for all $n \geq 1$, since μ is invariant, we conclude

$$\mu\Big(\bigcup_{n=0}^{\infty} f^{-n}(E_0)\Big) = \sum_{n=0}^{\infty} \mu(f^{-n}(E_0)) = \sum_{n=0}^{\infty} \mu(E_0).$$

As we assume that the measure is finite, the expression on the left hand side is finite. On the other hand, we have the right hand side in a sum of infinite terms, all equal. The only way this sum is finite is that each term is simply zero. Therefore, we have $\mu(E_0) = 0$, as promised.

Now denote by F the set of points $x \in E$ which returns E only a finite number of times. As a direct consequence of the definition, we have that every point $x \in F$ has some iteration $f^k(x)$ in E_0. That is,

$$F \subset \bigcup_{k=0}^{\infty} f^{-k}(E_0)$$

Since $\mu(E_0) = 0$ and μ is invariant, we have:

$$\mu(F) \leq \mu(\cup_{k=0}^{\infty} f^{-k}(E_0)) \leq \sum_{k=0}^{\infty} \mu(f^{-k}(E_0)) = \sum_{k=0}^{\infty} \mu(E_0) = 0$$

Therefore, $\mu(F) = 0$ as we want to prove.

□

Now, we will illustrate these concepts with few examples.

Proposition 1.12. *The Bernoulli measure μ is invariant under the shift map σ.*

Proof. Observe that the pre-image of any cylinder is still a cylinder:

$$\sigma^{-1}([m; A_m, \ldots, A_n]) = [m+1; A_m, \ldots, A_n]. \tag{1.10}$$

It follows that σ is measurable with respect to the σ-algebra $\mathcal{B}$. Moreover,

$$\mu\big(\sigma^{-1}([m; A_m, \ldots, A_n])\big) = \nu(A_m) \cdots \nu(A_n) = \mu\big([m; A_m, \ldots, A_n]\big)$$

and, observing that the cylinders generate the product σ-algebra, this ensures that the measure μ is invariant under σ.

□

1.2.1 Decimal expansion

Our first example is the transformation defined in the interval $[0, 1]$ as follows

$$f : [0, 1] \to [0, 1], \quad f(x) = 10x - [10x]$$

where $[10x]$ represents the largest integer less than or equal to $10x$. In other words, f associates with each $x \in [0, 1]$ the fractional part of $10x$.

We claim that the Lebesgue measure μ on the interval is invariant for the transformation f, i.e., it satisfies the condition

$$\mu(E) = \mu(f^{-1}(E)) \quad \text{for all measurable set } E \subset M. \tag{1.11}$$

This fact can be verified as follows. We first assume that E is an interval, then, the pre-image $f^{-1}(E)$ consists of ten intervals, each one which is ten times shorter than E. Then, the Lebesgue measure of $f^{-1}(E)$ is equal to the Lebesgue measure of E. This shows that (1.11) is satisfied in the case of intervals. As a consequence, this relationship is fulfilled when E is a finite union of intervals.

Now, the family of finite unions of intervals is a algebra which generates the Borel σ-algebra of $[0, 1]$. Therefore, in order to complete the proof, simply use the following general fact:

Lemma 1.13. *Let $f : M \to M$ be a measurable transformation and μ is a finite measure on M. Suppose that there exists a algebra $\mathcal{A}$ an algebra 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 all $E \in \mathcal{A}$. Then the same holds for all the set $E \in \mathcal{B}$, i.e., the measure μ is invariant under f.*

Proof. Let us first prove that $\mathcal{C} = \{E \in \mathcal{B} : \mu(E) = \mu(f^{-1}(E))\}$ is a monotone class. For this, let $E_1 \subset E_2 \subset \dots$ be a sequence of elements of $\mathcal{C}$ and let $E = \cup_{i=1}^{\infty} E_i$. We have

$$\mu(E) = \lim_{i\to\infty} \mu(E_i) \quad \text{and} \quad \mu(f^{-1}(E)) = \lim_{i\to\infty} \mu(f^{-1}(E_i)).$$

Then, using the fact that $E_i \in \mathcal{C}$,

$$\mu(E) = \lim_{i\to\infty} \mu(E_i) = \lim_{i\to\infty} \mu(f^{-1}(E_i)) = \mu(f^{-1}(E)).$$

Then $E \in \mathcal{C}$ and this proves that $\mathcal{C}$ is in fact a monotone class.

Now it is easy to get the conclusion of the lemma. Note that $\mathcal{C}$ contains $\mathcal{A}$, by hypothesis. Therefore, using the monotone class theorem, it follows that $\mathcal{C}$ contains the σ-algebra $\mathcal{B}$ generated by $\mathcal{A}$. This is precisely what we wanted to prove. □

Now we will explain how, from the fact that the Lebesgue measure is invariant under the transformation f, we can obtain interesting conclusions using the Poincaré recurrence theorem.The function f has a direct relationship with the algorithm of the decimal expansion: if x is given by

$$x = 0, a_0 a_1 a_2 a_3 \cdots$$

wieh $a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, then its image is given by

$$f(x) = 0, a_1 a_2 a_3 \cdots .$$

Thus, this makes it easy to write an expression of the nth iteration for any $n \geq 1$:

$$f^n(x) = 0, a_n a_{n+1} a_{n+2} \cdots \tag{1.12}$$

Now, let E be the subset of $x \in [0,1]$ whose decimal expansion begins with the digit 7, that is, with $a_0 = 7$. According to the Theorem 1.11, almost all the elements of E has infinite iterations which are also in E. Taking into account the expression (1.12), this means there are infinitely many of n such that $a_n = 7$. Therefore, we proved that *almost all numbers x whose expansion decimal begins by 7 has infinite number of digits that are equal to 7.*

Obviously, instead of 7 we can consider any other digit. In addition, we can also consider multi-digit blocks.Now, we see a much stronger result: for almost all number $x \in [0,1]$, every digit appears with frequency $1/10$ in their expansion decimal.

Open Question 1.14. Let $f : S^1 \to S^1$ and $g : S^1 \to S^1$ given by $f(x) = 2x \mod 1$ and $g(x) = 3x \mod 1$. In a completely similar fashion, one can prove that the Lebesgue measure is preserved by f and g. Is it true that Lebesgue measure is the unique non-atomic measure preserved *simultaneously* by f and g?

The former question was proposed by Furstenberg, and so far, still unsolved. For a recent account and comments on the subject, we suggest [Mat].

1.3 Ergodicity

In the following, we state the famous Birkhoff Ergodic Theorem and Kingman Subadditive Ergodic Theorem, which is a general version of Birkhoff's Theorem. They have several important applications in Ergodic Theory.

Theorem 1.15. *Let $f : M \to M$ be a measurable transformation and μ is an invariant probability measure of f. Given any integrable function $\varphi : M \to \mathbb{R}$, the limit*

$$\tilde{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) \tag{1.13}$$

exist at μ-almost every point $x \in M$. Moreover, the function $\tilde{\varphi}$ thus defined is integrable and satisfies

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

The limit $\tilde{\varphi}$ is called *time mean* of φ: The proposition below shows that the time means are constant along the orbits at μ-almost every point, which generates the equality

$$\tilde{\varphi}(f(x)) = \tilde{\varphi}(x). \tag{1.14}$$

Another key concept in Ergodic Theory is the notion of *ergodicity.*

Definition 1.16. We say that an invariant probability is *ergodic*, if given any integrable function φ, the time average $\tilde{\varphi}$ is constant almost everywhere (therefore, equal to $\int \phi \, d\mu$).

One can show that this is equivalent to any set A such that $f^{-1}(A) = A$, satisfy $\mu(A) = 0$ or $\mu(A) = 1$. We call these sets as *invariant sets.* In the same fashion, we say that a function $\phi : M \to \mathbb{R}$ is f-invariant, if $\phi(f(x) = \phi(x)$ for μ-almost every point $x \in M$.

If μ is ergodic, any invariant function is constant at μ-almost every point. We may conclude from Birkhoff's Ergodic Theorem that any ergodic probability μ and integrable function ϕ, the time average is constant μ-almost everywhere.

Proposition 1.17. *Any Bernoulli shift (σ, μ) is ergodic.*

Proof. Let A be any invariant measurable set. We want to show that $\mu(A) = 0$ or $\mu(A) = 1$. We will use the following fact:

Lemma 1.18. *If B and C are finite unions of pairwise disjoint cylinders then we have*

$$\mu(B \cap \sigma^{-j}(C)) = \mu(B)\mu(\sigma^{-j}(C)) = \mu(B)\mu(C),$$

for any large enough j.

Proof. To begin with, suppose that B and C are both cylinders: $B = [k; B_k, \dots, B_l]$ and $C = [m; C_m, \dots, C_n]$. Then,

$$\sigma^{-j}(C) = [m + j; C_m, \dots, C_n] \quad \text{for each } j.$$

Consider any large enough j so that $m + j > l$. Then,

$$\begin{aligned} B \cap \sigma^{-j}(C) &= \{(x_n)_n : x_k \in B_k, \dots, x_l \in B_l, x_{m+j} \in C_m, \dots, x_{n+j} \in C_n\} \\ &= [k; B_k, \dots, B_l, X, \dots, X, C_m, \dots, C_n], \end{aligned}$$

where X appears exactly $m+j-l-1$ times. By definition (1.7), this gives that

$$\mu(B \cap \sigma^{-j}(C)) = \prod_{i=k}^{l} \nu(B_i)\, 1^{m+j-l-1} \prod_{i=m}^{n} \nu(C_i) = \mu(B)\mu(C).$$

this proves the conclusion of the lemma when the involved sets are cylinders. The general case follows immediately, because μ is finitely additive. □

Suppose firstly that the invariant set A belongs to algebra $\mathcal{B}_0$ of the finite unions of disjoint cylinders. In this case, we can apply the previous lemma with $B = C = A$. We conclude that $\mu(A \cap \sigma^{-j}(A)) = \mu(A)^2$ whenever we take j sufficiently large. But, as A is invariant, the left hand side of this inequality is $\mu(A)$. Thus we obtain that $\mu(A) = \mu(A)^2$, which can only happen when $\mu(A) = 0$ ou $\mu(A) = 1$.

Now let's prove the case when A is any invariant measurable set. The idea is to approximate the invariant set by elements of the algebra $\mathcal{B}_0$: given any $\varepsilon > 0$ there exists $B \in \mathcal{B}_0$ such that $\mu(A \Delta B) < \varepsilon$. Fix j such that

$$\mu(B \cap \sigma^{-j}(B)) = \mu(B)\mu(\sigma^{-m}(B)) = \mu(B)^2. \tag{1.15}$$

Observe que the symmetric difference $(A\cap\sigma^{-j}(A))\Delta(B\cap\sigma^{-j}(B))$ is contained in

$$(A\Delta B)\cup(\sigma^{-j}(A)\Delta\sigma^{-j}(B)) = (A\Delta B)\cup\sigma^{-j}(A\Delta B).$$

This, together with the fact that μ is invariant under f, implies that

$$\left|\mu(A\cap\sigma^{-j}(A)) - \mu(B\cap\sigma^{-j}(B))\right| \le 2\mu(A\Delta B) < 2\varepsilon. \tag{1.16}$$

Moreover,

$$\left|\mu(A)^2 - \mu(B)^2\right| \le 2\left|\mu(A) - \mu(B)\right| < 2\varepsilon. \tag{1.17}$$

Jointing the relations (1.15), (1.16), (1.17), we conclude that $|\mu(A)-\mu(A)^2| < 4\varepsilon$. As ε is arbitrary, we deduce that $\mu(A) = \mu(A)^2$ and, therefore, either $\mu(A) = 0$ or $\mu(A) = 1$. □

The Birkhoff Ergodic Theorem can be obtained as a special case of a more general result, called *the subadditive ergodic theorem.* To state this theorem, we need to define what means a subadditive sequence of functions:

We say that a sequence of functions $\varphi_n : M \to \mathbb{R}$ is *subadditive* for a transformation $f : M \to M$ if

$$\varphi_{m+n} \le \varphi_m + \varphi_n \circ f^m \quad \text{for any } m, n \ge 1. \tag{1.18}$$

Example 1.19. The sequence $\varphi_n : M \to \mathbb{R}$ is said to be *additive* if the equality in (1.18) holds, i.e., if $\varphi_{m+n} = \varphi_m + \varphi_n \circ f^m$ for all $m, n \ge 1$. For example, all the orbital sums

$$\varphi_n(x) = \sum_{j=0}^{n-1} \varphi(f^j(x))$$

constitute an additive sequence. It is easy to verify that all the additive sequence is in this form, with $\varphi = \varphi_1$.

The next example need the notion of *norm* of a square matrix, which is defined in the following way. Let A be a square matrix of dimension $d \ge 2$. Then

$$\|A\| = \sup\left\{\frac{\|Av\|}{\|v\|} : v \in \mathbb{R}^d \setminus \{0\}\right\} \tag{1.19}$$

It follows directly from the definition that the norm of the product of two matrices is smaller than or equal to the product of the norms of these two matrices:

$$\|AB\| \le \|A\|\,\|B\|\,. \tag{1.20}$$

Example 1.20. Let $\theta : M \to \mathrm{GL}(d)$ be a measurable function taking values in the set $\mathrm{GL}(d)$ of the invertible square matrices of dimension d. Define $\phi^n(x) = \theta(f^{n-1}(x))\cdots\theta(f(x))\theta(x)$ for any $n \ge 1$ and $x \in M$. Then the sequence $\varphi_n(x) = \log\|\phi^n(x)\|$ is subadditive. In fact,

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

and therefore, using (1.20),

$$\begin{aligned}\varphi_{m+n}(x) &= \log \|\phi^n(f^m(x))\phi^m(x)\| \\ &\leq \log \|\phi^m(x)\| + \log \|\phi^n(f^m(x))\| = \varphi_m(x) + \varphi_n(f^m(x)).\end{aligned}$$

for any m, n and x.

Remember that, given a function $\varphi : M \to \mathbb{R}$ represent by $\varphi^+ : M \to \mathbb{R}$ the function defined by $\varphi^+(x) = \max\{\varphi(x), 0\}$.

Theorem 1.21 (Kingman). *Let $\varphi_n : M \to \mathbb{R}$, $n \geq 1$ be a subadditive sequence of measurable functions such that $\varphi_1^+ \in L^1(\mu)$. Then the sequence $(\varphi_n/n)_n$ converges at μ-almost every point to a measurable function $\varphi : M \to [-\infty, +\infty)$. Moreover, $\varphi^+ \in L^1(\mu)$ and*

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

For a proof of Theorem 1.21 we suggest [AB].

1.3.1 Linear automorphisms on the torus

In what follows we assume that $\mathbb{T}^d$ is equipped with the plane Riemannian metric, making it locally isometric to Euclidean space $\mathbb{R}^d$. Let m be the Lebesgue measure associated to this Riemannian metric.

Let A be the $d \times d$ matrix with integer coefficients and nonvanish determinant. Then $A(\mathbb{Z}^d) \subset \mathbb{Z}^d$ and, consequently, A induces a transformation

$$f_A : \mathbb{T}^d \to \mathbb{T}^d, \quad f_A([x]) = [A(x)]$$

where $[x]$ denote the equivalence class which contains $x \in \mathbb{R}$. We call such transformations *linear endomorphisms* of the torus. Note that f_A is differentiable and the derivative $Df_A(x)$ at each point is canonically identified with A. In particular, the Jacobian determinant $\det Df_A([x])$ is a constant, denoted as $\det A$. It is clear that A^{-1} has integer coefficients if, and only if, $|\det A| = 1$. Thus, in this case, f_A is invertible and its inverse is the transformation $f_{A^{-1}}$ induced by the inverse matrix A^{-1}. We say that f_A is an *automorphism*.

We claim that f_A preserves the Lebesgue measure in $\mathbb{T}^d$. This can be seen as follows. As f_A is a local diffeomorphism, the pre-image of any measurable set D with diameter small enough consists of $|\det A|$ (= degree of f_A) disjoint parts, D_i, each of which is sent diffeomorphically to D. By the change of variable formula, $m(D) = |\det A| \, m(D_i)$ for any i. This proves that $m(D) = m(f^{-1}(D))$ for any small enough domain D. Then f preserves the measure m.

1.3.2 Hopf argument

In this section we present a geometric method to prove the ergodicity of certain automorphism in the torus. This method can be applied when $|\det A| = 1$ and the matrix A is hyperbolic, that is, it has no eigenvalues of modulo 1. The big advantage of this method is that it can be extended into more general differential systems, not necessarily linear.

The hypothesis that the matrix A is hyperbolic means that the space $\mathbb{R}^d$ can be written as a direct sum $\mathbb{R}^d = E^s \oplus E^u$ such that:

1. $A(E^s) = E^s$ and all the eigenvalues of $A \mid E^s$ have modulus smaller than 1;

2. $A(E^u) = E^u$ and all the eigenvalues of $A \mid E^u$ have modulus greater than 1.

Then exists a constant $C > 0$ and $\lambda < 1$ such that

$$\begin{aligned} \|A^n(v^s)\| &\leq C\lambda^n \|v^s\| \quad \text{for any } v^s \in E^s \text{ and any } n \geq 0, \\ \|A^{-n}(v^u)\| &\leq C\lambda^n \|v^u\| \quad \text{for any } v^u \in E^u \text{ and any } n \geq 0. \end{aligned} \tag{1.21}$$

Example 1.22. Consider $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. its eigenvalues are

$$\lambda_u = \frac{3+\sqrt{5}}{2} > 1 > \lambda_u = \frac{3-\sqrt{5}}{2} > 0$$

and the respective eigenspace are:

$$E^u = \{(x,y) \in \mathbb{R}^2 : y = \frac{\sqrt{5}-1}{2}x\} \quad \textit{and} \quad E^s = \{(x,y) \in \mathbb{R}^2 : y = -\frac{\sqrt{5}+1}{2}x\}.$$

The family of all the affine subspaces of $\mathbb{R}$ of the form $v + E^s$, with $v \in \mathbb{R}^d$, define a partition $\mathcal{F}^s$ of $\mathbb{R}^d$, which is called *stable foliation* whose elements are called *stable leaves* of A. It is invariant under A, in other words, the image of any stable leaf is also a stable leaf. Moreover, by property (1.21), the transformation A contracts distances uniformly on each leaf. Similarly, the set of all the affine subspaces of $\mathbb{R}$ of the form $v + E^u$ with $v \in \mathbb{R}^d$ defines a partition $\mathcal{F}^u$ of $\mathbb{R}^d$, called *unstable foliation.* This foliation is also invariant and the transformation A expands distances along its leaves.

Projecting $\mathcal{F}^s$ and $\mathcal{F}^u$ by the canonical projection $\pi : \mathbb{R}^d \to \mathbb{T}^d$ we obtain foliations $\mathcal{W}^s$ and $\mathcal{W}^u$ in the torus which is called *stable foliation* and *unstable foliation* of the transformation f_A. The previous observations show that these foliations are invariant under f_A. Additionally:

(a) given any two points x and y in the same stable leaf, we have that $d(f_A^j(x), f_A^j(y)) \to 0$ when $n \to +\infty$;

(b) given any two points y and z in the same unstable leaf, we have that $d(f_A^j(y), f_A^j(z)) \to 0$ when $n \to -\infty$.

We use the geometry information to prove (f_A, m) is ergodic. For this, consider any continuous function $\varphi : \mathbb{T} \to \mathbb{R}$ and consider the time mean

$$\varphi^+(x) = \lim \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f_A^j(x)) \quad \text{and} \quad \varphi^-(x) = \lim \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f_A^{-j}(x)),$$

defined by m-almost every $x \in \mathbb{T}^d$. By Birkhoff Theorem aplied to f and f^{-1}, there exists a set $X \subset \mathbb{T}^d$ with full measure such that

$$\varphi^+(x) = \varphi^-(x) \quad \text{for any } x \in X. \tag{1.22}$$

Denote by $\mathcal{W}^s(x)$ and $\mathcal{W}^u(x)$, respectively, the stable leaf and the unstable leaf of f_A through each point $x \in \mathbb{T}^d$.

Lemma 1.23. *The function φ^+ is constant in each leaf of $\mathcal{W}^s$: if $\varphi^+(x)$ exists and $y \in \mathcal{W}^s(x)$ then $\varphi^+(y)$ exists and is equal to $\varphi^+(x)$. Similarly, φ^- is constant on each leaf of $\mathcal{W}^u$.*

Proof. According to the property (a) above, $d(f_A^j(x), f_A^j(y))$ converges to zero when $j \to \infty$. As φ is continuous (thus uniformly continuous, since the domain is compact) this implies that

$$\varphi(f_A^j(x)) - \varphi(f_A^j(y)) \to 0 \quad \text{when } j \to \infty.$$

or further, the Cesaro limit

$$\lim \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f_A^j(x)) - \varphi(f_A^j(y))$$

is also zero. This implies that $\varphi^+(y)$ exists and is equal to $\varphi^+(x)$. The argument for φ^- is entirely analogous. □

Given an open subset R of the torus and given $x \in R$, denote by $\mathcal{W}^s(z, R)$ the connected component of $\mathcal{W}^s(x) \cap R$ which contains x and by $\mathcal{W}^u(z, R)$ the connected component of $\mathcal{W}^u(x) \cap R$ which contains x. R is called *rectangle* if $\mathcal{W}^s(x, R)$ intersects $\mathcal{W}^u(y, R)$ at one unique point, for any x and y in R.

Lemma 1.24. *Given any rectangle $R \subset \mathbb{T}^d$, there exists a measurable set $Y_R \subset X \cap R$ such that $m(R \setminus Y_R) = 0$ and, given any x and y in Y_R, there exist points x' and y' in $X \cap R$ such that $x' \in \mathcal{W}^s(x, R)$ and $y' \in \mathcal{W}^s(y, R)$ and $y' \in \mathcal{W}^u(x')$.*

Proof. Represent by m_x^s the Lebesgue measure in the stable leaf $\mathcal{W}^s(x)$ of a given point $x \in \mathbb{T}^d$. Note that $m(R \setminus X) = 0$, since X has full measure in $\mathbb{T}^d$. Then, using the Fubini theorem,

$$m_x^s\big(\mathcal{W}^s(x, R) \setminus X\big) = 0 \quad \text{for } m\text{-almost every point } x \in R.$$

Define $Y_R = \big\{x \in X \cap R : m_x^s\big(\mathcal{W}^s(x, R) \setminus X\big) = 0\big\}$. Then Y_R has full measure in R. Given $x, y \in R$ consider the map

$$\pi : \mathcal{W}^s(x, R) \to \mathcal{W}^s(y, R), \quad \pi(x') = \text{intersection of } \mathcal{W}^u(x', R) \text{ and } \mathcal{W}^s(y, R).$$

This map is affine and, therefore, has the following property, which is called *absolutely continuous*:

$$m_x^s(E) = 0 \quad \Leftrightarrow \quad m_y^s(\pi(E)) = 0.$$

In particular, the image of $\mathcal{W}^s(x, R) \cap X$ has full measure in $\mathcal{W}^s(y, R)$ and, consequently, it intersects $\mathcal{W}^s(y, R) \cap X$. In other words, there exists $x' \in \mathcal{W}^s(x, R) \cap X$ whose image $y' = \pi(x')$ is in $\mathcal{W}^s(y, R) \cap X$. Observing that x' and y' are in the same unstable leaf, by definition of π, we see that these points satisfy the the conditions of the conclusions of the lemma. □

Consider any rectangle R. Given any x, y in Y_R, consider the points x', y' in X given by Lemma 1.24. Using also Lemma 1.23, we obtain:

$$\varphi^-(x) = \varphi^+(x) = \varphi^+(x') = \varphi^-(x') = \varphi^-(y') = \varphi^+(y') = \varphi^+(y) = \varphi^-(y).$$

this shows that the functions φ^+ and φ^- coincide with each other and are constant in Y_R. Now let $R_1, \ldots, R_N$ be a finite cover by rectangles. Consider the set

$$Y = \bigcup_{j=1}^{N} Y_j, \qquad \text{where } Y_j = Y_{R_j}.$$

Observe that $m(Y) = 1$, since $Y \cap R_j \supset Y_j$ has full measure of R_i for any j. We claim that $\varphi^+ = \varphi^-$ is constant at the whole Y. In fact, given any $k, l \in \{1, \ldots, N\}$ we can find $j_0 = k, j_1, \ldots, j_{n-1}, j_n = l$ such that each R_{j_i} intersects $R_{j_{i-1}}$ (this is a simple consequence of the path connectivity of torus). Remember that R_j is open and X_j is a subset with full measure, we obtain that each X_{j_i} intersects $X_{j_{i-1}}$. Then, $\varphi^+ = \varphi^-$ is constant in the union of all the X_{j_i}. This proves our statement.

So, we show that the time mean $\varphi^\pm$ of any continuous function φ is constant at m-almost every point. Consequently, the system (f_A, m) is ergodic.

Chapter 2

Entropies

2.1 Metric Entropy

2.1.1 Entropy of a Partition

Let $(M, \mathcal{B}, \mu)$ be a probability space. In this chapter, *partition* is always understood as a family $\mathcal{P}$ of finite or countable pairwise disjoint measurable subsets of M whose union is with full measure. Denote by $\mathcal{P}(x)$ the element of the partition which contains the point x.

Definition 2.1. The *sum* $\mathcal{P} \vee \mathcal{Q}$ of two partitions $\mathcal{P}$ and $\mathcal{Q}$ is the partition whose elements are the intersections $P \cap Q$ with $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$.

More generally, given any countable family of partitions $\mathcal{P}_n$, define

$$\bigvee_n \mathcal{P}_n = \{\bigcap_n P_n : P_n \in \mathcal{P}_n \text{ for any } n\}$$

Given a partition $\mathcal{P}$ we may associate its *information function* $I_{\mathcal{P}}$ setting

$$I_{\mathcal{P}} : M \to \mathbb{R}, \quad I_{\mathcal{P}}(x) = -\log \mu(\mathcal{P}(x)). \tag{2.1}$$

It is clear that the function $I_{\mathcal{P}}$ is measurable.

Definition 2.2. The *entropy*, or, *average information*, of the partition $\mathcal{P}$ is the number

$$H_\mu(\mathcal{P}) = \int I_{\mathcal{P}}\, d\mu = \sum_{P \in \mathcal{P}} -\mu(P) \log \mu(P). \tag{2.2}$$

As is usual in the theory of the Lebesgue integral, we make the convention that $0 \log 0 = \lim_{x \to 0} x \log x = 0$.

We say that two partitions $\mathcal{P}$ and $\mathcal{Q}$ are independent if $\mu(P \cap Q) = \mu(P)\mu(Q)$ for any $P \in \mathcal{P}$ and any $Q \in \mathcal{Q}$. In this case, $I_{\mathcal{P} \vee \mathcal{Q}} = I_{\mathcal{P}} + I_{\mathcal{Q}}$ and, therefore, $H_\mu(\mathcal{P} \vee \mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$. In general, the inequality $\leq$ holds as we shall see. This is related to the fact that the function $\phi(x) = -x \log x$ is concave: $\phi'' < 0$ and, consequently,

$$t\phi(x) + (1-t)\phi(y) \le \phi\big(tx + (1-t)y\big) \quad \text{for any } x, y \text{ and any } t \in [0,1].$$

Example 2.3. Consider $M = [0,1]$ equipped with the Lebesgue measure. for each $n \ge 1$ consider the partition $\mathcal{P}^n$ into subintervals $\big((i-1)/10^n, i/10^n\big]$ with $1 \le i \le 10^n$. Then

$$H_\mu(\mathcal{P}^n) = \sum_{i=1}^{10^n} -10^{-n} \log 10^{-n} = n \log 10.$$

Example 2.4. Let $M = \{1, \ldots, d\}^{\mathbb{N}}$ equipped with a product measure $\mu = \nu^{\mathbb{N}}$. Denote by $p_i = \nu(\{i\})$ for each $i \in \{1, \ldots, d\}$. For any $n \ge 1$, let $\mathcal{P}^n$ be the partition of M into cylinders $[0; a_1, \ldots, a_n]$ of length n. The entropy of $\mathcal{P}^n$ is

$$\begin{aligned}
H_\mu(\mathcal{P}^n) &= \sum_{a_1,\ldots,a_n} -p_{a_1} \ldots p_{a_n} \log\big(p_{a_1} \ldots p_{a_n}\big) \\
&= \sum_j \sum_{a_1,\ldots,a_n} -p_{a_1} \ldots p_{a_j} \ldots p_{a_n} \log p_{a_j} \\
&= \sum_j \sum_{a_j} -p_{a_j} \log p_{a_j} \sum_{a_i, i \ne j} p_{a_1} \ldots p_{a_{j-1}} p_{a_{j+1}} \ldots p_{a_n}.
\end{aligned}$$

The last sum is equal to 1, since $\sum_i p_i = 1$. Therefore,

$$H_\mu(\mathcal{P}^n) = \sum_{j=1}^{d} \sum_{a_j=1}^{d} -p_{a_j} \log p_{a_j} = \sum_{j=1}^{d} \sum_{i=1}^{d} -p_i \log p_i = -n \sum_{i=1}^{d} p_i \log p_i.$$

Lemma 2.5. *Any finite partition has finite entropy. In fact, $H_\mu(\mathcal{P}) \le \log \#\mathcal{P}$ and the equality holds if and only if $\mu(P) = 1/\#\mathcal{P}$ for any $P \in \mathcal{P}$.*

Proof. Let $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$ and consider the numbers $t_i = 1/n$ and $x_i = \mu(P_i)$. By Jensen's inequality (Theorem 5.2):

$$\frac{1}{n} H_\mu(\mathcal{P}) = \sum_{i=1}^{n} t_i \phi(x_i) \le \phi\big(\sum_{i=1}^{n} t_i x_i\big) = \phi\big(\frac{1}{n}\big) = \frac{\log n}{n}.$$

Therefore, $H_\mu(\mathcal{P}) \le \log n$. Moreover, the equality occurs if and only if $\mu(P_i) = 1/n$ for any $i = 1, \ldots, n$. □

The following example shows that countable partition can have infinite entropy. From now on we always consider partitions (finite or countable) with finite entropy.

Example 2.6. Consider $M = [0,1]$ equipped with the Lebesgue measure μ. Observe that the series $\sum_{k=1}^{\infty} 1/(k(\log k)^2)$ is convergent. Let c be the value of the sum. Then we can decompose $[0,1]$ into intervals P_k with $\mu(P_k) = 1/(ck(\log k)^2)$ for all k. Let $\mathcal{P}$ be the partition formed by these intervals. Then,

$$H_\mu(\mathcal{P}) = \sum_{k=1}^{\infty} \frac{\log c + \log k + 2\log\log k}{ck(\log k)^2}.$$

By the ratio criterion for the convergence, the series on the right-hand side has the same behavior with the series $\sum_{k=1}^{\infty} 1/(k \log k)$ which, as we know, is divergent (use the integral criterion). Therefore, $H_\mu(\mathcal{P}) = \infty$.

The *conditional entropy* of a partition $\mathcal{P}$ with respect to a partition $\mathcal{Q}$ is the number

$$H_\mu(\mathcal{P}/\mathcal{Q}) = \sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} -\mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu(Q)}. \tag{2.3}$$

Intuitively, it measures the additional information provided by the partition $\mathcal{P}$ once we know the information of the partition $\mathcal{Q}$. It is clear that $H_\mu(\mathcal{P}/\mathcal{M}) = H_\mu(\mathcal{P})$ for any $\mathcal{P}$, where $\mathcal{M}$ denote the trivial partition $\mathcal{M} = \{M\}$. Moreover, if $\mathcal{P}$ and $\mathcal{Q}$ are independent then $H_\mu(\mathcal{P}/\mathcal{Q}) = H_\mu(\mathcal{P})$. In general, the inequality $\leq$ as we will see soon.

Given two partitions, $\mathcal{P}$ and $\mathcal{Q}$ we say that $\mathcal{P}$ is *coarser* than $\mathcal{Q}$, and we write $\mathcal{P} \prec \mathcal{Q}$, if any element of $\mathcal{Q}$ is contained in some element of $\mathcal{P}$, except a measure zero. The sum $\mathcal{P}\vee\mathcal{Q}$ is, precisely, the coarsest one of the partitions $\mathcal{R}$ such that $\mathcal{P}\prec\mathcal{R}$ and $\mathcal{Q}\prec\mathcal{R}$.

Lemma 2.7. *Let $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{R}$ be partitions with finite entropy. Then,*

(a) $H_\mu(\mathcal{P}\vee\mathcal{Q}/\mathcal{R}) = H_\mu(\mathcal{P}/\mathcal{R}) + H_\mu(\mathcal{Q}/\mathcal{P}\vee\mathcal{R})$;

(b) if $\mathcal{P}\prec\mathcal{Q}$ then $H_\mu(\mathcal{P}/\mathcal{R}) \leq H_\mu(\mathcal{Q}/\mathcal{R})$ and $H_\mu(\mathcal{R}/\mathcal{P}) \geq H_\mu(\mathcal{R}/\mathcal{Q})$.

(c) $\mathcal{P}\prec\mathcal{Q}$ if and only if $H_\mu(\mathcal{P}/\mathcal{Q}) = 0$.

Proof. By definition,

$$\begin{aligned}
H_\mu(\mathcal{P}\vee\mathcal{Q}/\mathcal{R}) &= \sum_{P,Q,R} -\mu(P\cap Q\cap R)\log\frac{\mu(P\cap Q\cap R)}{\mu(R)} \\
&= \sum_{P,Q,R} -\mu(P\cap Q\cap R)\log\frac{\mu(P\cap Q\cap R)}{\mu(P\cap R)} \\
&\qquad + \sum_{P,Q,R} -\mu(P\cap Q\cap R)\log\frac{\mu(P\cap R)}{\mu(R)}.
\end{aligned}$$

The sum of the right-hand side can be written as

$$\begin{aligned}
&\sum_{S\in\mathcal{P}\vee\mathcal{R},Q\in\mathcal{Q}} -\mu(S\cap Q)\log\frac{\mu(C\cap Q)}{\mu(S)} + \sum_{P\in\mathcal{P},R\in\mathcal{R}} -\mu(P\cap R)\log\frac{\mu(P\cap R)}{\mu(R)} \\
&\qquad = H_\mu(\mathcal{Q}/\mathcal{P}\vee\mathcal{R}) + H_\mu(\mathcal{P}/\mathcal{R}).
\end{aligned}$$

This proves the item (a). Now observe that if $\mathcal{P} \prec \mathcal{Q}$ then

$$H_\mu(\mathcal{P}/\mathcal{R}) = \sum_P \sum_R \sum_{Q\subset P} -\mu(Q\cap R)\log\frac{\mu(P\cap R)}{\mu(R)}$$
$$\le \sum_P \sum_R \sum_{Q\subset P} -\mu(Q\cap R)\log\frac{\mu(Q\cap R)}{\mu(R)} = H_\mu(\mathcal{Q}/\mathcal{R}).$$

This proves the first part of item (b). For the second part, note that for any $P\in\mathcal{P}$ and $R\in\mathcal{R}$, we have

$$\frac{\mu(R\cap P)}{\mu(P)} = \sum_{Q\subset P}\frac{\mu(Q)}{\mu(P)}\frac{\mu(R\cap Q)}{\mu(Q)}.$$

It is clear that $\sum_{Q\subset P}\mu(Q)/\mu(P) = 1$. Then, by Jensen's inequality (Theorem 5.2),

$$\phi\Big(\frac{\mu(R\cap P)}{\mu(P)}\Big) \ge \sum_{Q\subset P}\frac{\mu(Q)}{\mu(P)}\phi\Big(\frac{\mu(R\cap Q)}{\mu(Q)}\Big)$$

for any $P\in\mathcal{P}$ and $R\in\mathcal{R}$. Consequently,

$$H_\mu(\mathcal{R}/\mathcal{P}) = \sum_{P,R}\mu(P)\phi\Big(\frac{\mu(R\cap P)}{\mu(P)}\Big) \ge \sum_{P,R}\mu(P)\sum_{Q\subset P}\frac{\mu(Q)}{\mu(P)}\phi\Big(\frac{\mu(R\cap Q)}{\mu(Q)}\Big)$$
$$= \sum_{Q,R}\mu(Q)\phi\Big(\frac{\mu(R\cap Q)}{\mu(Q)}\Big) = H_\mu(\mathcal{R}/\mathcal{Q}).$$

Finally, it follows from the definition (2.3) that $H_\mu(\mathcal{P}/\mathcal{Q}) = 0$ if and only if, for any $P\in\mathcal{P}$ and any $Q\in\mathcal{Q}$,

$$\mu(P\cap Q) = 0 \quad \text{otherwise} \quad \frac{\mu(P\cap Q)}{\mu(Q)} = 1.$$

In other words, either Q is disjoint with P (except for a set with measure zero) or Q is contained in P (except for a set with measure zero). This means that $H_\mu(\mathcal{P}/\mathcal{Q}) = 0$ if and only if $\mathcal{P} \prec \mathcal{Q}$. □

In particular, take $\mathcal{Q} = \mathcal{M}$ in the item (b) of the lemma we obtain that

$$H_\mu(\mathcal{R}/\mathcal{P}) \le H_\mu(\mathcal{R}) \quad \text{for any partitions } \mathcal{R} \text{ and } \mathcal{P}. \tag{2.4}$$

Moreover, take $\mathcal{R} = \mathcal{M}$ in item (a), we have that

$$H_\mu(\mathcal{P}\vee\mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}/\mathcal{P}) \le H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}). \tag{2.5}$$

We also need the following property of continuity:

Lemma 2.8. *Given $k\ge 1$ and $\varepsilon > 0$ there exists $\delta > 0$ such that, for any finite partition $\mathcal{P} = \{P_1,\dots,P_k\}$ and $\mathcal{Q} = \{Q_1,\dots,Q_k\}$,*

$$\mu(P_i\Delta Q_i) < \delta \text{ for any } i = 1,\dots,k \quad \Rightarrow \quad H_\mu(\mathcal{Q}/\mathcal{P}) < \varepsilon.$$

Proof. Fix $\varepsilon > 0$ and $k \geq 1$. By the continuity of the fuction $\phi : [0,1] \to \mathbb{R}$, $\phi(x) = -x \log x$, there exists $\rho > 0$ such that $\phi(x) < \varepsilon/k^2$ for any $x \in [0, \rho) \cup (1-\rho, 1]$. Take $\delta = \rho/k$. Given partitions $\mathcal{P}$ and $\mathcal{Q}$ as in the statement, denote by $\mathcal{R}$ the partition whose elements are the intersections $P_i \cap Q_j$ with $i \neq j$, and also the set $\cup_{i=1}^k P_i \cap Q_i$. Note that $\mu(P_i \cap Q_j) \leq \mu(P_i \Delta Q_i) < \delta < \rho$ for any $i \neq j$ and

$$\mu\Big(\bigcup_{i=1}^{k} P_i \cap Q_i\Big) \geq \sum_{i=1}^{k}(\mu(P_i) - \mu(P_i \Delta Q_i)) > \sum_{i=1}^{k}(\mu(P_i) - \delta) = 1 - \rho$$

Therefore,

$$H_\mu(\mathcal{R}) = \sum_{R \in \mathcal{R}} \phi(\mu(R)) < \#\mathcal{R}\frac{\varepsilon}{k^2} \leq \varepsilon.$$

It is clear from the definition that $\mathcal{P} \vee \mathcal{Q} = \mathcal{P} \vee \mathcal{R}$. Then, using (2.5) and (2.4),

$$\begin{aligned} H_\mu(\mathcal{Q}/\mathcal{P}) &= H_\mu(\mathcal{P} \vee \mathcal{Q}) - H_\mu(\mathcal{P}) = H_\mu(\mathcal{P} \vee \mathcal{R}) - H_\mu(\mathcal{P}) \\ &= H_\mu(\mathcal{R}/\mathcal{P}) \leq H_\mu(\mathcal{R}) < \varepsilon. \end{aligned}$$

This proves the lemma. □

2.1.2 Entropy of a dynamical system

Let $f : M \to M$ be a measurable transformation preserving a probability measure μ.Given a partition $\mathcal{P}$ of M with finite entropy, denote

$$\mathcal{P}^n = \bigvee_{i=0}^{n-1} f^{-i}(\mathcal{P}) \quad \text{for any } n \geq 1.$$

Observe that the element $\mathcal{P}^n(x)$ which contain $x \in M$ is given by:

$$\mathcal{P}^n(x) = \mathcal{P}(x) \cap f^{-1}(\mathcal{P}(f(x))) \cap \cdots \cap f^{-n+1}(\mathcal{P}(f^{n-1}(x))).$$

It is clear that the sequence $\mathcal{P}^n$ is non-decreasing, i.e., $\mathcal{P}^n \prec \mathcal{P}^{n+1}$ for all n. Therefore, the sequence of the entropies $H_\mu(\mathcal{P}^n)$ is also non-decreasing.

Observe that in Examples 2.3 and 2.4 the entropy of $\mathcal{P}^n$ grow linearly with n. This motivate the following lemma:

Lemma 2.9. *$H_\mu(\mathcal{P}^{m+n}) \leq H_\mu(\mathcal{P}^m) + H_\mu(\mathcal{P}^n)$ for any $m, n \geq 1$.*

Proof. By definition, $\mathcal{P}^{m+n} = \vee_{i=0}^{m+n-1} f^{-i}(\mathcal{P}) = \mathcal{P}^m \vee f^{-m}(\mathcal{P}^n)$. Therefore, using (2.5),

$$H_\mu(\mathcal{P}^{m+n}) \leq H_\mu(\mathcal{P}^m) + H_\mu(f^{-m}(\mathcal{P}^n)). \tag{2.6}$$

On the other hand,

$$H_\mu(f^{-1}(\mathcal{Q})) = H_\mu(\mathcal{Q}) \quad \text{for any partition } \mathcal{Q}, \tag{2.7}$$

since the measure μ is invariant under f and, therefore, $\mu(f^{-1}(Q)) = \mu(Q)$ for any $Q \in \mathcal{Q}$. In particular, $H_\mu(f^{-m}(\mathcal{P}^n)) = H_\mu(\mathcal{P}^n)$ for any m, n. Replacing this fact in (2.6) we obtain the conclusion of the lemma. □

Entropy of f with respect to the measure μ and the partition $\mathcal{P}$ is the limit

$$h_\mu(f,\mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \inf_n \frac{1}{n} H_\mu(\mathcal{P}^n). \tag{2.8}$$

Observe that this entropy is greater when the partition is finer. In fact, if $\mathcal{P} \prec \mathcal{Q}$ then $\mathcal{P}^n \prec \mathcal{Q}^n$ for all n. Using Lemma 2.7, it follows that $H_\mu(\mathcal{P}^n) \le H_\mu(\mathcal{Q}^n)$ for any n. Consequently,

$$\mathcal{P} \prec \mathcal{Q} \quad \Rightarrow \quad h_\mu(f,\mathcal{P}) \le h_\mu(f,\mathcal{Q}) \tag{2.9}$$

Finally, the *entropy* of the system (f,μ) is defined by

$$h_\mu(f) = \sup_{\mathcal{P}} h_\mu(f,\mathcal{P}), \tag{2.10}$$

where the supremum is taken over all the partitions with finite entropy. A useful observation is that the definition is not affected if we only consider the supremum over the finite partitions.

Example 2.10. Suppose that the invariant measure μ is supported on a periodic orbit. In other words, there exists x in M and $k \ge 1$ such that $f^k(x) = x$ and the measure μ is given by

$$\mu = \frac{1}{k}\big(\delta_x + \delta_{f(x)} + \cdots + \delta_{f^{k-1}(x)}\big).$$

In this case the measure only takes finite number of values. Consequently, the entropy $H_\mu(\mathcal{P})$ also only takes a finite number of values when we consider all the countable partitions $\mathcal{P}$. In particular, $\lim_n n^{-1} H_\mu(\mathcal{P}^n) = 0$ for all partition $\mathcal{P}$. This proves that, in this case, $h_\mu(f) = 0$.

Example 2.11. Consider the transformation $f:[0,1]\to[0,1]$ given by $f(x) = 10x - [10x]$ equipped with the Lebesgue measure. Let $\mathcal{P}$ be the partition of $[0,1]$ into subintervals of the form $\big((i-1)/10, i/10\big]$ with $i = 1,\ldots,10$. Then $\mathcal{P}^n$ is the partitions into the intervals of the form $\big((i-1)/10^n, i/10^n\big]$ with $i = 1,\ldots,10^n$. Using the calculation in Example 2.3, we obtain that

$$h_\mu(f,\mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \log 10.$$

Using the theory that will be developed in Section 2.1.3 we see that this is also the uniquevalue of the entropy $h_\mu(f)$, i.e., $\mathcal{P}$ realizes the supreme in the definition (2.10).

Open Question 2.12. Let $f : S^1\times(-\alpha,\alpha) \to S^1\times(-\alpha,\alpha)$ be the cilinder map $f(x,y) = (10x \mod 1, a_0 - y^2 + b\sin(2\pi x))$, where a_0 and b are conveniently choosen.

1. f has a (non-zero) finite number of maximal entropy measures?

2. If g is close to f in the C^3 topology, g must have a (non-zero) finite number of maximal entropy measures?

Example 2.13. Consider the shift $f : M \to M$ in the set $M = \{1, \dots, d\}^{\mathbb{N}}$ (or $M = \{1, \dots, d\}^{\mathbb{Z}}$), equipped with a Bernoulli measure $\mu = \nu^{\mathbb{N}}$ (respectively, $\mu = \nu^{\mathbb{N}}$). Let $\mathcal{P}$ be the partition of M into cylinders $[0; a]$ with $a = 1, \dots, d$. Then $\mathcal{P}^n$ is the partition into cylinders $[0; a_1, \dots, a_n]$ of length n. Using the calculation in Example 2.4 we conclude that

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \sum_{i=1}^{d} -p_i \log p_i.$$

The theory we present in Section 2.1.3 allow to conclude that this is also the value of entropy $h_\mu(f)$.

Lemma 2.14. *$h_\mu(f, \mathcal{Q}) \leq h_\mu(f, \mathcal{P}) + H_\mu(\mathcal{Q}/\mathcal{P})$ for any partitions $\mathcal{P}$ and $\mathcal{Q}$ with finite entropy.*

Proof. By Lemma 2.7, for any $n \geq 1$ the following holds

$$\begin{aligned} H_\mu\big(\mathcal{Q}^{n+1}/\mathcal{P}^{n+1}\big) &= H_\mu\big(\mathcal{Q}^n \vee f^{-n}(\mathcal{Q})/\mathcal{P}^n \vee f^{-n}(\mathcal{P})\big) \\ &\leq H_\mu\big(\mathcal{Q}^n/\mathcal{P}^n\big) + H_\mu\big(f^{-n}(\mathcal{Q})/f^{-n}(\mathcal{P})\big) \end{aligned}$$

The last term is equal to $H\mu(\mathcal{Q}/\mathcal{P})$, because the measure μ is invariant under f. Therefore, the previous relation proves that

$$H_\mu\big(\mathcal{Q}^n/\mathcal{P}^n\big) \leq n H_\mu\big(\mathcal{Q}/\mathcal{P}\big) \quad \text{for any } n \geq 1. \tag{2.11}$$

Using Lemma 2.7 once again, it follows that

$$H_\mu(\mathcal{Q}^n) \leq H_\mu(\mathcal{P}^n \vee \mathcal{Q}^n) = H_\mu(\mathcal{P}^n) + H_\mu(cQ^n/\mathcal{P}^n) \leq H_\mu(\mathcal{P}^n) + n H_\mu(\mathcal{Q}/\mathcal{P}).$$

Dividing by n and passing to the limit as $n \to \infty$ we get the conclusion of the lemma. □

Lemma 2.15. *$h_\mu(f, \mathcal{P}) = \lim_n H_\mu(\mathcal{P}/\bigvee_{j=1}^n f^{-j}(\mathcal{P}))$ for any partition $\mathcal{P}$ with finite entropy.*

Proof. Using Lemma 2.7(a) and the fact that the measure μ is invariant:

$$\begin{aligned} H_\mu\big(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big) &= H_\mu\big(\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\big) + H_\mu\big(\mathcal{P}/\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\big) \\ &= H_\mu\big(\bigvee_{j=0}^{n-2} f^{-j}(\mathcal{P})\big) + H_\mu\big(\mathcal{P}/\bigvee_{j=1}^{n-1} f^{-j}(\mathcal{P})\big) \end{aligned}$$

for all n. By recurrence, it follows that

$$H_\mu\big(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big) = H_\mu(\mathcal{P}) + \sum_{k=1}^{n-1} H_\mu\big(\mathcal{P}/\bigvee_{j=1}^{k} f^{-j}(\mathcal{P})\big).$$

Therefore, $h_\mu(f,\mathcal{P})$ is given by the Cesàro limit

$$h_\mu(f,\mathcal{P}) = \lim_n \frac{1}{n} H_\mu\big(\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big) = \lim_n \frac{1}{n} \sum_{k=1}^{n-1} H_\mu\big(\mathcal{P}/\bigvee_{j=1}^{k} f^{-j}(\mathcal{P})\big).$$

On the other hand, Lemma 2.7(b) guarantees that the sequence $H_\mu(\mathcal{P}/\vee_{j=1}^{n} f^{-j}(\mathcal{P}))$ is decreasing. In particular, $\lim_n \vee_{j=1}^{n} f^{-j}(\mathcal{P})$ exist and, consequently, coincide with the Cesàro limit in the previous inequality. □

Recall that $\mathcal{P}^n = \vee_{j=0}^{n-1} f^{-j}(\mathcal{P})$. When $f : M \to M$ is invertible, also consider $\mathcal{P}^{\pm n} = \vee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$.

Lemma 2.16. *If $\mathcal{P}$ is a partition with finite entropy then $h_\mu(f,\mathcal{P}) = h_\mu(f,\mathcal{P}^k)$ for any $k \geq 1$. If f is invertible, we also have $h_\mu(f,\mathcal{P}) = h_\mu(f,\mathcal{P}^{\pm k})$ for any $k \geq 1$.*

Proof. Observe that, given any $n \geq 1$,

$$\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P}^k) = \bigvee_{j=0}^{n-1} f^{-j}\big(\bigvee_{i=0}^{k-1} f^{-i}(\mathcal{P})\big) = \bigvee_{l=0}^{n+k-1} f^{-l}(\mathcal{P}) = \mathcal{P}^{n+k}.$$

Therefore,

$$h_\mu\big(f,\mathcal{P}^k\big) = \lim_n \frac{1}{n} H_\mu\big(\mathcal{P}^{n+k}\big) = \lim_n \frac{1}{n} H_\mu\big(\mathcal{P}^n\big) = h_\mu\big(f,\mathcal{P}\big).$$

This proves the first part of the lemma. To prove the second part, note that:

$$\bigvee_{j=0}^{n-1} f^{-j}(\mathcal{P}^{\pm k}) = \bigvee_{j=0}^{n-1} f^{-j}\big(\bigvee_{i=-k}^{k-1} f^{-i}(\mathcal{P})\big) = \bigvee_{l=-k}^{n+k-1} f^{-l}(\mathcal{P}) = f^{-k}\big(\mathcal{P}^{n+2k}\big)$$

for any n and any k. Therefore,

$$h_\mu\big(f,\mathcal{P}^{\pm k}\big) = \lim_n \frac{1}{n} H_\mu\big(f^{-k}(\mathcal{P}^{n+2k})\big) = \lim_n \frac{1}{n} H_\mu\big(\mathcal{P}^{n+2k}\big) = h_\mu\big(f,\mathcal{P}\big).$$

(the second equality use the fact that μ is invariant under f). □

Proposition 2.17. *We have that $h_\mu(f^k) = k h_\mu(f)$ for any $k \in \mathbb{N}$. If f is invertible then $h_\mu(f^k) = |k| h_\mu(f)$ for any $k \in \mathbb{Z}$.*

Proof. Consider $g = f^k$ and let $\mathcal{P}$ be any partition of M with finite entropy. Recall that $\mathcal{P}^k = \mathcal{P} \vee f^{-1}(\mathcal{P}) \vee \cdots \vee f^{-k+1}(\mathcal{P})$, we see that

$$\mathcal{P}^{km} = \bigvee_{i=0}^{km} f^{-j}(\mathcal{P}) = \bigvee_{i=0}^{m} f^{-ki}(\mathcal{P}^k) = \bigvee_{i=0}^{m} g^{-i}(\mathcal{P}^k).$$

Therefore,

$$kh_\mu(f,\mathcal{P}) = \lim_m \frac{1}{m} H_\mu(\mathcal{P}^{km}) = \lim_m \frac{1}{m} H_\mu\big(\bigvee_{i=0}^{m} g^{-i}(\mathcal{P}^k)\big) = h_\mu(g,\mathcal{P}^k).$$

Using Lemma 2.16, we see that $kh_\mu(f,\mathcal{P}) = h_\mu(g,\mathcal{P})$. Taking the supremum over these partitions $\mathcal{P}$ we see that $kh_\mu(f) = h_\mu(g)$, as stated.

Now suppose that f is invetible. Let $\mathcal{P}$ be any partition of M with finite entropy. For any $n \geq 1$,

$$H_\mu\big(\vee_{j=0}^{n-1} f^{-j}(\mathcal{P})\big) = H_\mu\big(f^{-n+1}\big(\vee_{i=0}^{n-1} f^{i}(\mathcal{P})\big)\big) = H_\mu\big(\vee_{i=0}^{n-1} f^{i}(\mathcal{P})\big),$$

since the measure μ is invariant. Dividing by n and passing to the limit as $n \to \infty$, we obtain that $h_\mu(f,\mathcal{P}) = h_\mu(f^{-1},\mathcal{P})$. Taking the supreme over these partitions $\mathcal{P}$, we see that $h_\mu(f) = h_\mu(f^{-1})$. Replacing f by f^k and use item (a), it follows that $h_\mu(f^{-k}) = h_\mu(f^k) = kh_\mu(f)$ for any $k \in \mathbb{N}$. □

2.1.3 Kolmogorov-Sinai Theorem

In general, the main difficulty in calculating the entropy lies in the calculation of the supreme in the definition (2.10). The methods we develop in this section allow to simplify this task for many interesting cases, identifying certain partitions $\mathcal{P}$ that carries the supreme, i.e., such that $h_\mu(f,\mathcal{P}) = h_\mu(f)$. The main result is the following:

Theorem 2.18. *Let $\mathcal{P}_1 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ be a non-decreasing sequence of partitions with finite entropy such that $\cup_{n=1}^{\infty}\mathcal{P}_n$ generates the σ-algebra of the measurable sets. Then*

$$h_\mu(f) = \lim_n h_\mu(f,\mathcal{P}_n).$$

Proof. The limit always exists, because the property (2.9) implies that the sequence $h_\mu(f,\mathcal{P}_n)$ is non-decreasing. We need the following fact:

Lemma 2.19. $\lim_n H_\mu(\mathcal{Q}/\mathcal{P}_n) = 0$ *for any finite partition* $\mathcal{Q}$.

Proof. Write $\mathcal{Q} = \{Q_1,\dots,Q_k\}$. Given any $\varepsilon > 0$, fix $\delta > 0$ as in Lemma 2.8. Let $\mathcal{A}$ be the algebra formed by finite unions of the elements of $\cup_n \mathcal{P}^n$. By hypotheses, $\mathcal{A}$ generates the σ-algebra of all the measurable sets. Then, by the approximation theorem (Theorem 5.10), for each $i = 1,\dots,s$ there exists $A_i \in \mathcal{A}$ such that

$$\mu(Q_i \Delta A_i) < \delta/(4k). \tag{2.12}$$

The fact that Q_i's cover M guarantees that A_i's are also very close of covering M:

$$\mu\big(A_i \cap (\cup_{j\neq i} A_j)\big) \leq \mu\big(\cup_{j=1}^{n} (A_j \setminus Q_j)\big) < \delta/4 \quad \text{for any } i, \tag{2.13}$$

$$\text{and} \quad \mu\big(M \setminus \cup_{i=1}^{k} A_i\big) \leq \mu\big(\cup_{i=1}^{k} (Q_i \setminus A_i)\big) < \delta/4. \tag{2.14}$$

Then, define

$$Q'_i = \begin{cases} A_1 & \text{for } i = 1 \\ A_i \setminus \cup_{j=1}^{i-1} A_j & \text{for } 1 < i < k \\ M \setminus \cup_{j=1}^{k-1} A_j & \text{for } i = k \end{cases}$$

Then $\mathcal{Q}' = \{Q'_1, \ldots, Q'_k\}$ is a partition of M. We claim that

$$\mu(A_i \Delta Q'_i) < \delta/2 \quad \text{for any } i = 1, \ldots, k. \tag{2.15}$$

This is trivial for $i = 1$. For $i > 1$ we have that $A_i \setminus Q'_i$ is contained in $A_i \cap (\cup_{j<i} A_j)$. Then, using (2.13), we obtain that $\mu(A_i \setminus Q'_i) < \delta/4$. This proves the claim for any $1 < i < k$, since in this case $Q'_i \setminus A_i = \emptyset$. Finally, for $i = k$, we have that $Q'_k \setminus A_k$ is contained in the complement of $\cup_{i=1}^k A_i$. Thus, using (2.14), we see that $\mu(Q'_k \setminus A_k) < \delta/4$. Adding this estimation with the previous one, we see that $\mu(A_k \Delta Q'_k) < \delta/2$. This completes the proof of the claim (2.15).

Combining the inequalities (2.12) and (2.15), we obtain that $\mu(Q_i \Delta Q'_i) < \delta$ for any $i = 1, \ldots, k$, and, it is clear that $Q'_i \in \mathcal{A}$ for all i. Then, as it is a finite family, we can find $m \geq 1$ such that each Q'_i is a union of the elements of $\mathcal{P}_m$. In other words, the partition $\mathcal{Q}' = \{Q'_1, \ldots, Q'_k\}$ is coarser than $\mathcal{P}_m$. Then, by the lemmas 2.7 and 2.8,

$$H_\mu(\mathcal{Q}/\mathcal{P}_n) \leq H_\mu(\mathcal{Q}/\mathcal{P}_m) \leq H_\mu(\mathcal{Q}/\mathcal{Q}') < \varepsilon \quad \text{for any } n \geq m.$$

This completes the proof of the lemma. □

By lemma 2.14, we also have that

$$h_\mu(f, \mathcal{Q}) \leq h_\mu(f, \mathcal{P}_n) + H_\mu(\mathcal{Q}/\mathcal{P}_n) \quad \text{for any } n.$$

Passing to the limit as $n \to \infty$ and then taking the supremum over all the finite partitions $\mathcal{Q}$ we obtain the conclusion of the theorem. □

2.1.4 Generating partitions

Now we deduce several useful consequences of the theorem.

Corollary 2.20. *Let $\mathcal{P}$ be a partition with finite entropy such that the union of its iterates $\mathcal{P}^n = \vee_{j=0}^{n-1} f^{-j}(\mathcal{P})$, $n \geq 1$ generates the σ-algebra of the measurable sets. Then $h_\mu(f) = h_\mu(f, \mathcal{P})$.*

Proof. Just apply Theorem 2.18 to the sequence $\mathcal{P}^n$, remembering that $h_\mu(f, \mathcal{P}^n) = h_\mu(f, \mathcal{P})$ for all n, according to Lemma 2.16. □

Corollary 2.21. *Suppose that the system (f, μ) is invertible. Let $\mathcal{P}$ be a partition with finite entropy such that the uinion of the iterates $\mathcal{P}^{\pm n} = \vee_{j=-n}^{n-1} f^{-j}(\mathcal{P})$, $n \geq 1$ generates the σ-algebra of the measurable sets. Then $h_\mu(f) = h_\mu(f, \mathcal{P})$.*

Proof. Just apply Theorem 2.18 to the sequence $\mathcal{P}^{\pm n}$, remember that $h_\mu(f, \mathcal{P}^{\pm n}) = h_\mu(f, \mathcal{P})$ for all n, according to Lemma 2.16. □

In particular, the Corollaries 2.20 and 2.21 completes the calculation of the entropy of the decimal expansion transformation and the Bernoulli shift, which was stated in Examples 2.11 and 2.13, respectively.

In both cases in Corollaries 2.20 and 2.21 we say that $\mathcal{P}$ is a *generating partition*, or a *generator* of the system. Note, however, that this contains some abuse of language, since the conditions of the corollaries are not equivalent. For example, if we take $M = \{1, \ldots, d\}^{\mathbb{Z}}$ then the partition $\mathcal{P}$ into cylinders $\{[0; a] : a = 1, \ldots, d\}$ is such that the union of the bilateral iterates $\mathcal{P}^{\pm n}$ generates the σ-algebra but the union of the unilateral iterates $\mathcal{P}^n$ does not.

In this respect we also observe that certain *invertible* systems admit partitions that are generating in the sense of Corollary 2.20, i.e., such that the union of the unilateral iterates $\mathcal{P}^n$ generates the σ-algebra of the measurable sets. For example, if $f : S^1 \to S^1$ is an irrational rotation and $\mathcal{P} = \{I, S^1 \setminus I\}$ is a partition of the circle into two complementary intervals, then $\mathcal{P}$ is a generating partition. However, this type of behavior is only possible for systems with zero entropy:

Corollary 2.22. *Suppose that $f : M \to M$ is invertible and there exists some partition $\mathcal{P}$ with finite entropy such that $\cup_{n=1}^{\infty} \mathcal{P}^n$ generates the σ-algebra of the measurable sets of M. Then $h_\mu(f) = 0$.*

Proof. Combining Lemma 2.15 and Corollary 2.20:

$$h_\mu(f) = h_\mu(f, \mathcal{P}) = \lim_n H_\mu(\mathcal{P}/f^{-1}(\mathcal{P}_n)).$$

As $\cup_n \mathcal{P}^n$ generates the σ-algebra $\mathcal{B}$ of the measurable sets, $\cup_n f^{-1}(\mathcal{P}^n)$ generates the σ-algebra $f^{-1}(\mathcal{B})$. But $f^{-1}(\mathcal{B}) = \mathcal{B}$, since f is invertible. Thus, Theorem 2.18 implies that $H_\mu(\mathcal{P}/f^{-1}(\mathcal{P}_n))$ converges to zero as $n \to \infty$. It follows that $h_\mu(f) = 0$. □

Suppose that M is a metric space, equipped with its Borel σ-algebra.

Corollary 2.23. *Let $\mathcal{P}_1 \prec \cdots \prec \mathcal{P}_n \prec \cdots$ be a non-decreasing sequence of partitions with finite entropy such that* $\operatorname{diam} \mathcal{P}_n(x) \to 0$ *at μ-almost every $x \in M$. Then*

$$h_\mu(f) = \lim_n h_\mu(f, \mathcal{P}_n).$$

Proof. Let U be any open set of M. The hypothesis guarantees that for any x there exists $n(x)$ such that the set $P_x = \mathcal{P}_{n(x)}(x)$ is contained in U. It is clear that P_x belongs to the algebra $\mathcal{A}$ generated by $\cup_n \mathcal{P}_n$. Observe also that this algebra is countable, since it is formed by the finite unions of the elements of the partitions of $\mathcal{P}_n$. In particular, the set of the values taken by P_x is countable. It follows that $U = \cup_{x \in U} P_x$ is also in this algebra $\mathcal{A}$. This proves that the σ-algebra generated by $\mathcal{A}$ contains all the open sets and, therefore, contains all the Borel sets. Now, the conclusion follows from the direct application of Theorem 2.18. □

Example 2.24. Let $f : S^1 \to S^1$ be a homeomorphism and let μ be any invariant probability measure. Given a finite partition $\mathcal{P}$ of S^1 into subintervals, denote by $x_1, \dots, x_m$ their extreme points. For any $j \geq 1$, the partition $f^{-j}(\mathcal{P})$ is formed by subintervals of S^1 determined by the points $f^{-j}(x_i)$. This implies that, for any $n \geq 1$, the elements of $\mathcal{P}^n$ have their extreme points in the set

$$\{f^{-j}(x_i) : j = 0, \dots, n-1 \text{ and } i = 1, \dots, m\}.$$

In particular, $\#\mathcal{P}^n \leq mn$. Then, using Lemma 2.5,

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) \leq \lim_n \frac{1}{n} \#\mathcal{P}^n = \lim_n \frac{1}{n} \log kn = 0.$$

It follows that $h_\mu(f) = 0$: for this it suffices to consider any sequence of finite partitions with diameter going to zero and apply Corollary 2.23.

Corollary 2.25. *Let $\mathcal{P}$ be a partition with finite entropy such that, for μ-almost every $x \in M$, we have that* $\operatorname{diam} \mathcal{P}^n(x) \to 0$. *Then $h_\mu(f) = h_\mu(f, \mathcal{P})$.*

Proof. Just apply Corollary 2.23 to the sequence $\mathcal{P}^n$, remembering that $h_\mu(f, \mathcal{P}^n) = h_\mu(f, \mathcal{P})$ for all n. □

Similarly, if f is invertible and $\mathcal{P}$ is a partition with finite entropy such that $\operatorname{diam} \mathcal{P}^{\pm n}(x) \to 0$ for μ-almost every $x \in M$, then $h_\mu(f) = h_\mu(f, \mathcal{P})$.

Example 2.26. We say that a continuous transformation $f : M \to M$ in a compact metric space M is *expansive* if there exists $\varepsilon > 0$ such that two distinct orbits can not remain within a distance less than ε at all the iterates:

$$d(f^j(x), f^j(y)) \leq \varepsilon \text{ for any } j \geq 0 \quad \Rightarrow \quad x = y.$$

Let $\mathcal{P}$ be any partition of M such that $\operatorname{diam} \mathcal{P} = \sup\{\operatorname{diam} P : P \in \mathcal{P}\}$ is smaller than ε. Then, $\operatorname{diam} \mathcal{P}^n(x) \to 0$ for any $x \in M$, as the reader can verify easily. Therefore, the partition $\mathcal{P}$ is generating.

2.2 Topological Entropy

Initially, we present the definitions of Adler-Konheim-McAndrew and Bowen-Dinaburg and prove that they are equivalent when the map is defined in a compact metric space.

2.2.1 Definition via open coverings

The original definition of topological entropy is very similar to the definition of Kolmogorov-Sinai entropy, using open coverings instead of partitions.

Let M be a compact topological space. We call *open cover* of M any family α of open sets whose union is M. By compactness, all open cover admits a *subcover* (i.e., a subfamily that is still a cover) with a finite number of elements. We call *entropy* of the cover α the number

$$H(\alpha) = \log N(\alpha) \tag{2.16}$$

where $N(\alpha)$ is the smallest number such that α admits a finite subcover with that number of elements.

Given two open covers α and β, we say that α is *finer* than β, and write $\alpha \prec \beta$ if every element of β is contained in some element of α. For example, if β is a subcover α then $\alpha \prec \beta$. We can check that,

$$\alpha \prec \beta \quad \Rightarrow \quad H(\alpha) \leq H(\beta). \tag{2.17}$$

Given covers $\alpha_1, \ldots, \alpha_n$, we denote by $\alpha_1 \vee \cdots \vee \alpha_n$ its *sum*, i.e., the cover whose elements are the intersections $A_a \cap \cdots \cap A_n$ with $A_j \in \alpha_j$ for each j. Observe that $\alpha_j \prec \alpha_1 \vee \cdots \vee \alpha_n$ for all j.

Let $f : M \to M$ be a continuous map. If α is an open cover of M then $f^{-j}(\alpha) = \{f^{-j}(A) : A \in \alpha\}$ is also an open cover. For each $n \geq 1$, we denote by

$$\alpha^n = \alpha \vee f^{-1}(\alpha) \vee \cdots \vee f^{-n+1}(\alpha).$$

We can check that

$$H(\alpha^{m+n}) = H(\alpha^m \vee f^{-m}(\alpha^n)) \leq H(\alpha^m) + H(f^{-m}(\alpha^n)) \leq H(\alpha^m) + H(\alpha^n)$$

for all $m, n \geq 1$. In other words, $H(\alpha^n)$ is an subadditive sequence. Consequently,

$$h(f, \alpha) = \lim_n \frac{1}{n} H(\alpha^n) \tag{2.18}$$

always exists. It is called entropy of f with respect to the cover α. The relation (2.17) implies that

$$\alpha \prec \beta \quad \Rightarrow \quad h(f, \alpha) \leq h(f, \beta). \tag{2.19}$$

Finally, we define the *topological entropy* of f as

$$h(f) = \sup\{h(f, \alpha) : \alpha \text{ is open cover of } M\}. \tag{2.20}$$

In particular, if β is a subcover of α then $h(f, \alpha) \leq h(f, \beta)$. Therefore, the definition (2.20) does not change when we restrict the supremum to the set of *finite* open covers.

Observe that the topological entropy $h(f)$ is a non-negative number, maybe infinity.

Example 2.27. Let $f : S^1 \to S^1$ be any homeomorphism (for example, a rotation R_θ) and let α be a cover formed by a finite number of open intervals.

Let $\partial\alpha$ be the set of endpoints of these intervals. For each $n \geq 1$, the cover α^n is given by intervals, whose endpoints belong to

$$\partial\alpha^n = \partial\alpha \cup f^{-1}(\partial\alpha) \cup \cdots \cup f^{-n+1}(\partial\alpha).$$

Observe that $\#\alpha^n \le \#\partial\alpha^n \le n\#\partial\alpha$. Thus,

$$h(f,\alpha) = \lim_n \frac{1}{n} H(\alpha^n) \le \liminf_n \frac{1}{n} \log \#\alpha^n \le \liminf_n \frac{1}{n} \log n = 0.$$

We shall see in Proposition 2.36 that $h(f) = \lim_k h(f,\alpha_k)$ for any sequence of open covers α_k with $\operatorname{diam}\alpha_k \to 0$. Therefore, considering open covers by intervals of length less than 1/ k, we concluded by the above calculation that $h(t) = 0$ for all homeomorphism circle.

Example 2.28. Consider $M = X^{\mathbb{N}}$, where $X = \{1,\dots,d\}$, and let α be the cover of M by cylinders $[0;a]$, $a \in X$. Consider the shift map $f : M \to M$. For each n, the cover α^n is given by cylinders of length n:

$$\alpha^n = \{[0; a_0, \dots, a_{n-1}] : a_j \in X\}.$$

Thus, $H(\alpha^n) = \log \#\alpha^n = \log d^n$ and, consequently, $h(f,\alpha) = \log d$. Observe that $\operatorname{diam}\alpha^n$ converges to zero when $n \to \infty$. It follows from the corollary 2.37, that we gonna prove in a while, that $h(f) = h(f,\alpha) = \log d$. The same goes for the shift $f : M \to M$ defined in $M = X^{\mathbb{Z}}$.

Example 2.29. The topological entropy of the shift $f : M \to M$, defined on the space $M = [0,1]^{\mathbb{Z}}$ is infinity.

Proof. By the same reasoning as in Example 2.13, we show that for any finite set $X = \{a_1,\dots,a_n\} \subset [0,1]$ with n elements, the shift map restricted to the set $M_n = X^{\mathbb{Z}}$ has topological entropy equal to $\log n$. Since the topological entropy of f in M is greater or equal to the topological entropy of f when restricted to an invariant compact set of M, we have that $h(f) \ge \log n$ for all $n \in \mathbb{N}$. □

Now we show that entropy is a topological invariant for topological equivalence. Let $f_1 : M_1\ toM_1$ and $f_2 : M_2\ toM_2$ continuous transformations on compact topological spaces M_1 and M_2. We say that f_2 is a *factor* of f_1 if there is a continuous surjective map $h : M_1 \to M_2$ such that $h \circ f_1 = f_2 \circ h$. If h can be chosen invertible (homeomorphism), we say that the two transformations are *topologically equivalent* or *topologically conjugated* and call h a *topological conjugacy* between f_1 and f_2.

Proposition 2.30. *If f_2 is a factor of f_1 then $h(f_2) \le h(f_1)$. In particular, if f_1 and f_2 are topologically equivalent, then $h(f_1) = h(f_2)$.*

Proof. Let $h : M_1 \to M_2$ be a surjective continuous map such that $h \circ f_1 = f_2 \circ h$. Given a cover α of M_2, a family

$$h^{-1}(\alpha) = \{h^{-1}(A) : A \in \alpha\}$$

is an open cover of M_1. Given subsets $A_0, A_1, \dots, A_{n-1} \in \alpha$, we have that:

$$\begin{aligned} h^{-1}\big(A_0 \cap f_1^{-1}(A_1) \cap \cdots \cap f_1^{-n+1}(A_{n-1})\big) &= h^{-1}\big(A_0\big) \cap h^{-1}\big(f_1^{-1}(A_1)\big) \cap \cdots \cap h^{-1}\big(f_1^{-n+1}(A_{n-1})\big) \\ &= h^{-1}(A_0) \cap f_2^{-1}\big(h^{-1}(A_1)\big) \cap \cdots \cap f_2^{-n+1}\big(h^{-1}(A_{n-1})\big). \end{aligned}$$

By definition, $h^{-1}(\alpha^n)$ is given by the sets in the left-hand side of this equation, and the sets in the right-hand side are $h^{-1}(\alpha)^n$. Therefore, $h^{-1}(\alpha^n) = h^{-1}(\alpha)^n$. Since h is surjective, a family $\gamma \subset \alpha^n$ covers M_2 if, and only if, $h^{-1}(\gamma)$ covers M_1. Thus, $H(h^{-1}(\alpha)^n) = H(h^{-1}(\alpha^n)) = H(\alpha^n)$. As n is arbitrary, it follows that $h(f_1, h^{-1}(\alpha)) = h(f_2, \alpha)$. Thus, taking the supremum over all covers α of M_2:

$$h(f_2) = \sup_\alpha h(f_2, \alpha) = \sup_\alpha h(f_1, h^{-1}(\alpha)) \leq h(f_1).$$

This prove the first part of the proposition. The second part is a immediate consequence, since in this case f_1 is also a factor of f_2. □

The converse of the Proposition 2.30 is false, in general. For example, all circle homeomorphisms have zero topological entropy (remember the Example 2.27) but not need to be topologically conjugate (for instance, the identity and a rotation R_θ with nonzero θ).

2.2.2 Generating Sets and Separated Sets

Now, we present the definition of topological entropy given by Bowen-Dinaburg. Let $f : M \to M$ be a continuous map in a metric space M, not necessarily compact, and let $K \subset M$ be a compact subset. When M is compact, we may consider $K = M$, as we can see in (2.23).

Given $\varepsilon > 0$ and $n \in \mathbb{N}$, we say that a set $E \subset M$ is a *(n, ε)-generator* of M, if for all $x \in K$ there exist $a \in E$ such that $d(f^i(x), f^i(y)) < \varepsilon$ for all $i \in \{0, \ldots, n-1\}$. In other words,

$$K \subset \bigcup_{a \in E} B(a, n, \varepsilon),$$

where $B(a, n, \varepsilon) = \{x \in M : d(f^i(x), f^i(y)) < \varepsilon \text{ for all } i = 0, \ldots, n-1\}$ is the *dynamical ball* with center a, length n and radius ε. Observe that $\{B(x, n, \varepsilon) : x \in K\}$ is an open cover of K. Then, by compactness, exist finite (n, ε)-generators for K . We denote by $g_n(f, \varepsilon, K)$ the minimum of the cardinality of a (n, ε)-generator of K. Define

$$g(f, \varepsilon, K) = \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, K).$$

Observe that $\varepsilon \mapsto g(f, \varepsilon, K)$ is non-increasing monotone function. Indeed, it is clear from the definition that if $\varepsilon_1 < \varepsilon_2$ the all (n, ε_1)-generator set is also a (n, ε_2)-generator. Thus, $g_n(f, \varepsilon_1, K) \geq g_n(f, \varepsilon_2, K)$ for all $n \geq 1$ and, taking the limit, $g(f, \varepsilon_1, K) \geq g(f, \varepsilon_2, K)$. In particular, this give us, that

$$g(f, K) = \lim_{\varepsilon \to 0} g(f, \varepsilon, K)$$

exist. Finally, define

$$g(f) = \sup\{g(f, K) : K \subset M \text{ compact}\}. \tag{2.21}$$

Now, we introduce the dual notion. Given $\varepsilon > 0$ and $n \in \mathbb{N}$, we say that a set $E \subset K$ is (n,ε)-*separated* if given $x, y \in E$, there exist $i \in \{0, \dots, n-1\}$ such that $d(f^j(x), f^j(y)) \geq \varepsilon$. In other words, if $x \in E$ then $B(x, n, \varepsilon)$ does not contain another point of E. Denote by $s_n(f, \varepsilon, K)$ the maximum of the cardinality of a (n,ε)-separated set. Define

$$s(f, \varepsilon, K) = \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, K).$$

It is clear that if $0 < \varepsilon_1 < \varepsilon_2$, then every (n, ε_2)-separated set is a also (n, ε_1)-separated. Thus, $s_n(f, \varepsilon_1, K) \geq s_n(f, \varepsilon_2, K)$ for all $n \geq 1$ and, taking the limit, $s(f, \varepsilon_1, K) \geq s(f, \varepsilon_2, K)$. In particular,

$$s(f, K) = \lim_{\varepsilon \to 0} s(f, \varepsilon, K)$$

always exists. Finally, define

$$s(f) = \sup\{s(f, K) : K \subset M \text{ compact}\}. \tag{2.22}$$

Its clear that $g(f, K_1) \leq g(f, K_2)$ e $s(f, K_1) \leq s(f, K_2)$ if $K_1 \subset K_2$. In particular:

$$g(f) = g(f, M) \quad \text{e} \quad s(f) = s(f, M) \quad \text{if } M \text{ is compact.} \tag{2.23}$$

Another interesting observation is that the definitions (2.21) and (2.22) does not change when we restricted the supremum to compact sets with diameter small.

Proposition 2.31. *We have that $g(f, K) = s(f, K)$ for all compact set $K \subset M$. Consequently, $g(f) = s(f)$.*

Proof. We need the following lemma:

Lemma 2.32. *$g_n(f, \varepsilon, K) \leq s_n(f, \varepsilon, K) \leq g_n(f, \varepsilon/2, K)$ for all $n \geq 1$, every $\varepsilon > 0$ and all compact $K \subset M$.*

Proof. Initially, let $E \subset K$ be a (n, ε)-separated set with maximal cardinality. Then, given every $y \in K \setminus E$, we have that $E \cup \{y\}$ is not (n, ε)-separated. Thus, there exists $x \in E$ and $i \in \{0, \dots, n-1\}$ such that $d(f^i(x), f^i(y)) < \varepsilon$. This proves that E is a (n, ε)-generating set of K and, on the other hand, this imply that $g_n(f, \varepsilon, K) \leq \#E = s_n(f, \varepsilon, K)$.

To prove the another inequality, let $E \subset K$ be a (n, ε)-separated set and let $F \subset M$ be a $(n, \varepsilon/2)$-generating set of K. By the hypothesis, we have that for all $x \in E$ there exists some point $y \in F$ such that $d(f^i(x), f^i(y)) < \varepsilon/2$ for all $i \in \{0, \dots, n-1\}$. Define $\phi : E \to F$ setting $\phi(x)$ as any point y as before. we claim that ϕ is injective. Indeed, assume that $x, z \in E$ are such that $\phi(x) = y = \phi(z)$. Then,

$$d(f^i(x), f^i(z)) \leq d(f^i(x), f^i(y)) + d(f^i(y), f^i(z)) < /vep/2 + \varepsilon/2$$

for all $i \in \{0, \dots, n-1\}$. Since E is (n, ε)-separated, this imply that $x = z$. So, ϕ é injective, as we claimed. It follows that $\#E \leq \#F$. As E and F are arbitrary, this prove that $s_n(f, \varepsilon, K) \leq g_n(f, \varepsilon/2, K)$. $\square$

Then, given any $\varepsilon > 0$ and a compact $K \subset M$,

$$g(f,\varepsilon,K) = \limsup_n \frac{1}{n}\log g_n(f,\varepsilon,K) \le \limsup_n \frac{1}{n}\log s_n(f,\varepsilon,K)$$
$$\le \limsup_n \frac{1}{n}\log g_n(f,\varepsilon/2,K) = g(f,\varepsilon/2,K).$$

Taking the limit when $\varepsilon \to 0$, we obtain that

$$g(f,K) = \lim_{\varepsilon\to 0} g(f,\varepsilon,K) \le \lim_{\varepsilon\to 0} s(f,\varepsilon,K) = s(f) \le \lim_{\varepsilon\to 0} g(f,\varepsilon/2,K) = g(f,K).$$

This proves the first part of the proposition. The second part is a immediate consequence. □

Proposition 2.33. *If M is a compact metric space then $h(f) = g(f) = s(f)$.*

Proof. By the Proposition 2.31, it is enough to show that $s(f) \le h(f) \le g(f)$.

Fix $\varepsilon > 0$ and $n \ge 1$. Let $E \subset M$ be a subset (n,ε)-separated and let α be any open cover of M with diameter less than ε. If x and y are in the same element of α^n then

$$d(f^i(x), f^i(y)) \le \operatorname{diam}\alpha < \varepsilon \quad \text{for all } i = 0,\dots,n-1.$$

In particular, each element of α^n contains at most one element of E and follow from this that $\#E \le N(\alpha^n)$. Taking E with maximal cardinality, we may conclude that $s_n(f,\varepsilon,M) \le N(\alpha^n)$ for all $n \ge 1$. So,

$$s(f,\varepsilon,M) = \limsup_n \frac{1}{n}\log s_n(f,\varepsilon,M) \le \lim_n \frac{1}{n}\log N(\alpha^n) = h(f,\alpha) \le h(f).$$

Taking $\varepsilon \to 0$ we obtain that $s(f) = s(f,M) \le h(f)$.

Given any open cover α of M, let $\varepsilon > 0$ be a Lebesgue number of α, this means, a positive number ε such that any ball with radius ε is contained in some element of α. Let $E \subset M$ be a (n,ε)-generator set of M with minimal cardinality. For each $x \in E$ and $i = 0,\dots,n-1$, there exists $A_{x,i} \in \alpha$ such that $B(f^i(x),\varepsilon)$ is contained in $A_{x,i}$. Then

$$B(x,n,\varepsilon) \subset \bigcap_{i=0}^{n-1} f^{-i}(A_{x,i}).$$

Thus, since E is a generating set, we have that $\gamma = \{\cap_{i=0}^{n-1} f^{-i}(A_{x,i}) : x \in E\}$ is an open cover of M. As $\gamma \subset \alpha^n$, it follows that $N(\alpha^n) \le \#E = g_n(f,\varepsilon,M)$ for all n. So,

$$\begin{aligned} h(f,\alpha) = \lim_n \frac{1}{n}\log N(\alpha^n) &\le \liminf_n \frac{1}{n}\log g_n(f,\varepsilon,M) \\ &\le \limsup_n \frac{1}{n}\log g_n(f,\varepsilon,M) = g(f,\varepsilon,M). \end{aligned} \tag{2.24}$$

Taking $\varepsilon \to 0$, we have that $h(f,\alpha) \le g(f,M) = g(f)$. Since α is an arbbitrary open cover, we have that $h(f) \le g(f)$. □

We define the *topological entropy* of a continuous transformation $f: M \to M$ in a metric space M as $h(f) = g(f) = s(f)$. By Proposition 2.33, we have that this definition is compatible with the definition introduced at 2.2.1 for continuous maps in compact topological spaces. However, we may observe that in the non-compact case, the topological entropy may depend on the distance function (and not only on the topology) of M. .

Example 2.34. Assume that $f: M \to M$ do not expand distance, i.e., assume that $d(f(x), f(y)) \leq d(x, y)$ for all $x, y \in M$. Then, the topological entropy of f is zero. Indeed, this assumption implies that $B(x, n, \varepsilon) = B(x, \varepsilon)$ for all $n \geq 1$. So, a set E is a (n, ε)-generator if, and only if, it is $(1, \varepsilon)$-generator. In particular, the sequence $g_n(f, \varepsilon, K)$ does not depend on n and, therefore, $g(f, \varepsilon, K) = 0$ for all $\varepsilon > 0$ and all compact set K. Taking $\varepsilon \to 0$ and considering the supremum over K, we obtain that $g(f) = 0$. Analougouslly, $s(f) = 0$.

Among them there are two important special cases: *contractions* such that exists $\lambda < 1$ satisfying $d(f(x), f(y)) \leq \lambda d(x, y)$ for all $x, y \in M$; and *isometries*, such that $d(f(x), f(y)) = d(x, y)$ for all $x, y \in M$. Every metrizable compact group admits a distance for which any translation is an isometry. So, it also follows from the foregoing that any translation of a metrizable topological group has topological entropy zero.

Remember that $g(f) = g(f, M)$ and $s(f) = s(f, M)$. We can easily see that the conclusion of Proposition 2.33can be rewriten as follows:

$$h(f) = \lim_{\varepsilon \to 0} \limsup_n \frac{1}{n} \log g_n(f, \varepsilon, M) = \lim_{\varepsilon \to 0} \limsup_n \frac{1}{n} \log s_n(f, \varepsilon, M).$$

But, from the proof of the proposition we can obtain the following related equality:

Corollary 2.35. *If $f: M \to M$ is a continuous map in a compact metric space, then*

$$h(f) = \lim_{\varepsilon \to 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M) = \lim_{\varepsilon \to 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M).$$

Proof. The equation (2.24) give that

$$h(f, \alpha) \leq \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M)$$

if $\varepsilon > 0$ is a Lebesgue number for the cover α. Taking $\varepsilon \to 0$, we may conclude that

$$h(f) \leq \lim_{\varepsilon \to 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M). \tag{2.25}$$

The first inequality in Lemma 2.32 imply that

$$\lim_{\varepsilon \to 0} \liminf_n \frac{1}{n} \log g_n(f, \varepsilon, M) \leq \lim_{\varepsilon \to 0} \liminf_n \frac{1}{n} \log s_n(f, \varepsilon, M). \tag{2.26}$$

It is clear that

$$\lim_{\varepsilon\to 0}\liminf_n \frac{1}{n}\log s_n(f,\varepsilon,M) \leq \lim_{\varepsilon\to 0}\limsup_n \frac{1}{n}\log s_n(f,\varepsilon,M). \tag{2.27}$$

As just noted, the right-hand expression is equal to $h(f)$. Therefore, the inequalities (2.25)-(2.27) imply the corollary. □

2.2.3 Properties and Computations

The next proposition and its corollary substantially simplify the calculation of topological entropy in concrete examples. When M is a metric space, we call *diameter* of an open cover the supreme of the diameters of its elements.

Proposition 2.36. *Assume that M is a compact metric space and let $(\beta_k)_k$ be any sequence of open covers of M such that* $\operatorname{diam}\beta_k$ *converges to zero. Then,*

$$h(f) = \sup_k h(f,\beta_k) = \lim_k h(f,\beta_k).$$

Proof. Given any open cover α, consider $\varepsilon > 0$ a Lebesgue number of α. Take $n \geq 1$ such that $\operatorname{diam}\beta_k < \ vep$ for all $k \geq n$. By definition of the Lebesgue number, it follows that every element of β_k is contained in some element of α. In other words, $\alpha \prec \beta_k$ and, therefore, $h(f,\beta_k)\ geh(f,\alpha)$. Recalling the definition (2.20) This proves that

$$\liminf_k h(f,\beta_k) \geq h(f).$$

It is clear from the definitions that $h(f) \geq \sup_k h(f,\beta_k) \geq \limsup_k h(f,\beta_k)$. Combining these two remarks we obtain the lemma. □

Corollary 2.37. *Let M be a compact metric space and β an open cover such that*

(1) The diameter of $\beta^k = \bigvee_{j=0}^{k-1} f^{-j}(\beta)$ goes to zero when $k \to \infty$, or

(2) $f : M \to M$ is a homeomorphism and the diameter of $\beta^{\pm k} = \bigvee_{j=-k}^{k-1} f^{-j}(\beta)$ goes to zero when $k \to \infty$,

then $h(f) = h(f,\beta)$.

Proof. In (1), by the Proposition 2.36, we have that:

$$h(f) = \lim_k h(f,\beta^k) = h(f,\beta).$$

The proof of (2) is analougous. □

Now, we gonna check that the topological entropy behaves as we wish with respect to positive iterates, at least when f is uniformly continuous:

Proposition 2.38. *If $f : M \to M$ is an uniformly continuous map in a metric space, then $h(f^k) = kh(f)$ for all $k \in \mathbb{N}$.*

Proof. Fix $k \geq 1$ and consider $K \subset M$ any compact set. For any $n \geq 1$ and $\varepsilon > 0$, we have that if $E \subset M$ is a (nk, ε)-generator of K for f, then E is also a (n, ε)-generator of K for f^k. So, $g_n(f^k, \varepsilon, K) \leq g_{nk}(f, \varepsilon, K)$. Therefore,

$$g(f^k, \varepsilon, K) = \lim_n \frac{1}{n} g_n(f^k, \varepsilon, K) \leq \lim_n \frac{1}{n} g_{nk}(f, \varepsilon, K) = kg(f, \varepsilon, K).$$

Taking $\varepsilon \to 0$ and taking the supremum on K, we have that $h(f^k) \leq kh(f)$.

The other inequality use the uniform continuity of f. Indeed, consider $\delta > 0$ such that $d(x, y) < \delta$ imply $d(f^j(x), f^j(y)) < \varepsilon$ for all $j \in \{0, \ldots, n-1\}$. If $E \subset M$ is a (n, δ)-generator de K for f^k then E is a (nk, ε)-generator of K for f. Thus, $g_{nk}(f, \varepsilon, K) \leq g_n(f^k, \delta, K)$. This imply that $kg(f, \varepsilon, K) \leq g(f, \delta, K)$. Taking ε and δ going to zero, we have that $kg(f, K) \leq g(f^k, K)$ for any compact set K. Thus, $kh(f) \leq h(f^k)$. □

In particular, the Proposition 2.38 holds true for any continuous map in a compact metric space. On the other hannd, in the case of homemorphisms in a compact metric space the conclusion can be extended to negative iterates:

Proposition 2.39. *Let M be a compact metric space and $f : M \to M$ a homemorphism. Then, $h(f^{-1}) = h(f)$. Consequently, $h(f^n) = |n|h(f)$ for all $n \in \mathbb{Z}$.*

Proof. Let α be a open cover of M. For any $n \geq 1$, denote by

$$\alpha_+^n = \alpha \vee f^{-1}(\alpha) \vee \cdots \vee f^{-n+1}(\alpha) \quad \text{e} \quad \alpha_-^n = \alpha \vee f(\alpha) \vee \cdots \vee f^{n-1}(\alpha)$$

Observe that $\alpha_-^n = f^{n-1}(\alpha_+^n)$. Even more, γ is an finite subcover of α_+^n if, and only if, $f^{n-1}(\gamma)$ is an finite subcover of α_-^n. Since these two covers have the same cardinality, it follows that $H(\alpha_+^n) = H(\alpha_-^n)$. So,

$$h(f, \alpha) = \lim_n \frac{1}{n} H(\alpha_+^n) = \lim_n \frac{1}{n} H(\alpha_-^n) = h(f^{-1}, \alpha).$$

Since α is arbitrary, this prove that $h(f) = h(f^{-1})$. The second part of the statement follows from this and from the Proposition 2.38. □

The claim in the Proposition 2.39 is false, in general, for the non-compact case:

Example 2.40. Consider $M = \mathbb{R}$ equipped with the usual distance and $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x$. Now, we check that $h(f) \neq h(f^{-1})$. In order to do it, let $K = [0, 1]$ and , given $n \geq 1$ e $\varepsilon > 0$, suppose that $E \subset \mathbb{R}$ is a (n, ε)-generator set for K. In particular, every point of $f^{n-1}(K) = [0, 2^{n-1}]$ is at most ε away from the set $f^{n-1}(E)$. Thus,

$$2\varepsilon \# E = 2\varepsilon \# f^{n-1}(E) \geq 2^{n-1}.$$

This proves that $g_n(f, \varepsilon, K) \geq 2^{n-2}/\varepsilon$ for all n and, therefore, $g(f, \varepsilon, K) \geq \log 2$. Follows immediatelly that $h(f) \geq h(f, K) \geq \log 2$. On the other hand, f^{-1} is a contraction and, from Example 2.34 we have that its entropy $h(f^{-1})$ is zero.

2.3 Examples

Let us illustrated some of the ideas the we discussed so far, using some examples.

2.3.1 Expansive Maps

A continuous map $f : M \to M$ in a compact metric space is *expansive* if exists $\varepsilon_0 > 0$ such that $d(f^j(x), f^j(y)) < \varepsilon_0$ for all $j \in \mathbb{N}$ implies that $x = y$. When $f : M \to M$ is invertible, we say that f is (two-sided) expansive, if there exists $\varepsilon_0 > 0$ such that $d(f^j(x), f^j(y)) < \varepsilon_0$ for all $j \in \mathbb{Z}$ implies that $x = y$. In these two cases, ε_0 is called *expansivity constant* of f.

Proposition 2.41. *If $\varepsilon_0 > 0$ is a expansivity constant for f then*

(a) $h(f) = h(f, \alpha)$ for every open cover α with diameter ε_0;

(b) $h(f) = g(f, \varepsilon, M) = s(f, \varepsilon, M)$ for all $\varepsilon < \varepsilon_0/2$.

In particular, $h(f) < \infty$.

Proof. Let α be any open cover of M with $\operatorname{diam} \alpha < \varepsilon_0$. We claim that $\lim_k \operatorname{diam} \alpha^k = 0$. Indeed, suppose that this is not true. It is clear that the diameter sequence is not increasing. Then, there exists $\delta > 0$ and for every $k \geq 1$ there exist points x_k and y_k at the same element of α^k such that $d(x_k, y_k) \geq \delta$. By compactness, we may take $x, y \in M$ and a subsequence $(k_j)_j$ such that $x = \lim_j x_{k_j}$ e $y = \lim_j y_{k_j}$. Observe that $x \neq y$, since $d(x, y) \geq \delta$. On the other hand, the choice of x_k and y_k at the same element of α^k imply that

$$d(f^i(x_k), f^i(y_k)) \leq \operatorname{diam} \alpha \quad \text{for all } 0 \leq i < k.$$

Taking the limit, we have that $d(f^i(x), f^i(y)) \leq \operatorname{diam} \alpha < \varepsilon_0$ for all $i \geq 0$. This is a contradiction, since ε_0 is a expansivity constant of f. This contradiction proves our claim. Using the Proposition 2.36, we have that $h(f) = h(f, \alpha)$, as we claim on (a).

To prove (b), let α be the cover of M by balls with radius ε. Note that α^n contains any dynamical ball $B(x, n, \varepsilon)$:

$$B(x, n, \varepsilon) = \bigcap_{j=0}^{n-1} f^{-j}\big(B(f^j(x), \varepsilon)\big) \quad \text{and} \quad B(f^j(x), \varepsilon) \in \alpha.$$

If E is a (n, ε)-generator set of M then $\{B(a, n, \varepsilon) : a \in E\}$ is an open cover of M; By the previous reasoning, it is a subcover of α. Thus (remember the Lemma 2.32),

$$N(\alpha^n) \leq g_n(f, \varepsilon, M) \leq s_n(f, \varepsilon, M) \quad \text{for all } n.$$

Taking the limit, we obtain that $h(f, \alpha) \leq g(f, \varepsilon, M) \leq s(f, \varepsilon, M)$. Remember that $s(f, \varepsilon, M) \leq s(f, M) = h(f)$. Since $\operatorname{diam} \alpha < \varepsilon_0$, we have by the first part of the proposition that $h(f) = h(f, \alpha)$. This imply the item (b).

The last claim in the proposition is a immediate consequence, since $g(f, \varepsilon, K)$, $s(f, \varepsilon, K)$ and $h(f, \alpha)$ are always finite. □

Now, we prove that for any expansive transformation the topological entropy is a upper bound for the growth rate of periodic points. We denote by $\mathrm{Fix}(f^n)$ the set of points $x \in M$ such that $f^n(x) = x$.

Proposition 2.42. *If M is a compact metric space and $f : M \to M$ is expansive then*

$$\limsup_n \frac{1}{n} \log \# Fix(f^n) \le h(f).$$

Proof. Let α be a cover of M with $\operatorname{diam}\alpha < \varepsilon_0$, where ε_0 is a expansitivy constant of f. We claim that each element of α^n cointains at most one point of $\mathrm{Fix}(f^n)$. Indeed, if $x, y \in \mathrm{Fix}(f^n)$ belong to the same element of α^n, then $d(f^i(x), f^i(y) < \operatorname{diam}\alpha < \varepsilon_0$ for all $i = 0, \dots, n-1$. Since $f^n(x) = x$ and $f^n(y) = y$, it follows that $d(f^i(x), f^i(y) < \varepsilon_0$ for all $i \ge 0$. By expansivity, this imply that $x = y$, proving our claim. We have that

$$\limsup_{\frac{1}{n}} \log \# \mathrm{Fix}(f^n) \le \limsup_n \frac{1}{n} \log \# N(\alpha^n) = h(f, \alpha).$$

Taking the limit when the diameter of α goes to zero, we obtain the proof od the proposition. □

In some interesting situations, we may prove that the entropy is actually *equal* to the growth rate of periodic points:

$$\lim_n \frac{1}{n} \log \# \mathrm{Fix}(f^n) = h(f). \tag{2.28}$$

This is the case, for instance, when we consider the subshifts of finite type, studed in the Section 2.3.2. In general, the Equation (2.28) holds true if $f : M \to M$ is an expanding map in a compact space.

2.3.2 Subshifts of Finite Type

Let $X = \{1, \dots, d\}$ be a finite set and $A = (A_{i,j})_{i,j}$ a $d \times d$ matrix with entries 0 or 1, and such that any row is null: for all i there exists j such that $A_{i,j} = 1$. This kind of matrix is called *transition matrix*. Consider the subset Σ_A of $M = X^{\mathbb{N}}$ of all sequences $(x_n)_n$ that are A-*admissible*, i.e., such that

$$A_{x_n, x_{n+1}} = 1 \quad \text{for all } n \in \mathbb{N}. \tag{2.29}$$

It is clear that Σ_A is an invariant set of the shift map $f : M \to M$, meaning that $f(\Sigma_A) \subset \Sigma_A$. Moreover, we have that Σ_A is a closed set of M and, therefore, Σ_A is a compact metric space.

We denote by $f_A : \Sigma_A \to \Sigma_A$ the restriction of the shift map f to Σ_A. This map is called *One-sided Subshift of finite type* associated to A. The two-sided shift of finite type associated to a transition matrix A is defined is a entire analougous way, taking $M = X^{\mathbb{Z}}$ and requiring that (2.29) holds true for all

$n \in \mathbb{Z}$. In this case, we also require that the columns (not only the rows) of A are non-zero.

The restriction of a Bernoulli shift $f : M \to M$ to the support of any Markov measure is a subshift of finite type:

Example 2.43. Given an stochastic matrix $P = (P_{i,j})_{i,j}$ define $A = (A_{i,j})_{i,j}$ by

$$A_{i,j} = \begin{cases} 1 & \text{if } P_{i,j} > 0 \\ 0 & \text{if } P_{i,j} = 0. \end{cases}$$

Observe that A is a transition matrix. We see that the a sequence is A-admissible if, and only if, it is P-admissible. Let μ be the Markov measure given by a probability vector $p = (p_j)_j$ with positive entries and such that $P^* p = p$. The support of μ coincides with the set $\Sigma_A = \Sigma_P$ of all admissible sequences.

It is very useful to associate a oriented graph to a transition matrix.

$$G_A = \{(a, b) \in X \times X : A_{a,b} = 1\}.$$

In other words, G_A is a graph for which vertices are points of $X = \{1, \ldots, d\}$ and such that there exists an edge from a to b if, and only if, $A_{a,b} = 1$.

We call *path with length* l on G_A to any sequence $a_0, \ldots, a_l$ in X such that $A_{a_{i-1},a_i} = 1$ for all i, i.e., such that always exists an edge connecting a_{i-1} to a_i. Given $a, b \in X$ an $l \geq 1$, what is the number $A^l_{a,b}$ of paths with length l beginning at a and ending at b (this meas: with $a_0 = a$ and $a_l = b$)? In order to answer this question, observe that

1. $A^1_{a,b} = 1$ if there exists an edge connecting a to b and $A^1_{a,b} = 0$ otherwise. In other words, $A^1_{a,b} = A_{a,b}$ for all a, b.
2. A path with length $l+m$ beginning at a and ending at b is the concatenations of a path with length m beginning at a and ending at b some point $z \in X$ with a path with length l beginning at z and ending at b. Therefore,

$$A^{l+m}_{a,b} = \sum_{z=1}^{d} A^l_{a,z} A^m_{z,b} \quad \text{for all } a, b \in X \text{ and all } l, m \geq 1.$$

From this two remarks it follows immediately that $A^l_{a,b}$ is, in fact, the entry in the row a and column b of the matrix A^l.

In the next proposition we compute the topological entropy of these transformations. We need some facts about transition matrices.

The *spectral radius* $\rho(B)$ of a linear transformation $B : \mathbb{R}^d \to \mathbb{R}$ (this means, the maximum of the absolute values of engeinvalues of B) is given by

$$\rho(B) = \lim_n \|B^n\|^{1/n}, \tag{2.30}$$

where $\|\cdot\|$ denotes any norm in the vector space of linear transformations; all norms are equivalent, since the domain has finite dimension. In general, we consider the *operator norm* $\|B\| = \sup\{\|Bv\|/\|v\| : v \neq 0\}$ but, sometimes it is useful to consider the norm defined by

$$\|B\|_s = \sum_{i,j=1}^{d} |B_{i,j}|.$$

Now, assume that A is a transition matrix. Since A has non-negative entries, we may use the Perron-Frobenius Theorem (Theorem 5.1), to conclude that the spectral radius λ_A of A is an eigenvalue of A. By the definition of transition matrix, we also have that the rows of A are not null. Then, we have the same for A^n, for any $n \geq 1$. This imply that all entries of the vector $A^n(1,\ldots,1)$ are (inteiros) positive and, therefore,

$$\|A^n\| \geq \frac{\|A^n(1,\ldots,1)\|}{\|(1,\ldots,1)\|} \geq 1 \quad \text{for all } n \geq 1.$$

Using (2.30), we deduce that $\lambda_A = \rho(A) \geq 1$ for all trasition matrix A.

Proposition 2.44. *The topological entropy of a subshift of finite type $f_A : \Sigma_A \to \Sigma_A$ is given by $h(f_A) = \log \lambda_A$ where λ_A is the biggest eingenvalue of A.*

Proof. Let us prove the proposition for one-sided subshifts. The two-sided case is analougous and we leave it for the reader. We consider the open cover α of Σ_A given by the restrictions

$$[0;a]_A = \{(x_j)_j \in \Sigma_A : x_0 = a\}$$

of cylinders $[0;a]$ of M. For each $n \geq 1$, the open cover α^n is given by the restrictions

$$[0;a_0,\ldots,a_{n-1}]_A = \{(x_j)_j \in \Sigma_A : x_j = a_j \text{ para } j = 0,\ldots,n-1\}.$$

of all cylinders with length n. Observe that this set is non-empty if, and only if, $a_0,\ldots,a_{n-1}$ is a path (with length $n-1$) in the graph G_A: it is clear that this condition is necessary; To prove that the condition is suficient, remember that by the definition, for all i there exists j such that $A_{i,j} = 1$. Since cylinders are disjoint two-by-two, this remark show that $N(\alpha^n)$ is equal to the total number of paths with length $n-1$ in the graph G_A, i.e.,

$$N(\alpha^n) = \sum_{i,j=1}^{d} A_{i,j}^{n-1} = \|A^{n-1}\|_s.$$

By caracterization of the spectral radius given in (2.30), it follows that

$$h(f,\alpha) = \lim_n \frac{1}{n} \log N(\alpha^n) = \lim_n \frac{1}{n} \log \|A^{n-1}\|_s = \log \rho(A) = \log \lambda_A.$$

Finally, since $\operatorname{diam} \alpha^n \to 0$, the Corollary 2.37 gives that $h(f_A) = h(f_A,\alpha)$. $\square$

Proposition 2.45. *If $f_A : \Sigma_A \to \Sigma_A$ is a subshift of finite type, then*

$$h(f_A) = \lim_n \frac{1}{n} \log \# Fix(f_A^n).$$

Proof. Let us prove the proposition for one-sided subshifts. The two-sided case is analougous and we leave it for the reader. Observe that $(x_k)_k \in \Sigma_A$ is a fixed pointt of f_A^n if, and only if, $x_k = x_{k-n}$ for all $k \geq n$. In particular, each cylinder $[0; a_0, \ldots, a_{n-1}]_A$ contains at most one element of $\mathrm{Fix}(f_A^n)$. Moreover, there exists a fixed point in the cylinder if, and only if, $a_0, \ldots, a_{n-1}, a_0$ is a path (with length n) in the graph G_A. This proves that

$$\#\mathrm{Fix}(f_A^n) = \sum_{i=1}^{d} A_{i,i}^n = \text{trace of } A^n$$

for all n. Consequently,

$$\lim_n \frac{1}{n} \log \#\mathrm{Fix}(f_A^n) = \sum_{i=1}^{d} A_{i,i}^n = \lim_n \frac{1}{n} \log\left(\text{trace of } A^n\right) = \log \rho(A).$$

Now, the proof follows from the previous proposition. □

2.3.3 Differentiable Maps

In this section, we suppose that M is a Riemannian manifold, i.e., a differentiable manifold with finite dimension with a inner product in the tangent space T_xM at every point x, such that the inner product depends smoothly on $x \in M$.

If $f : M \to M$ is a differentiable map, its derivative at x is a linear map $Df(x) : T_xM \to T_{f(x)}M$. The (operator) norm of $Df(x)$ is given by

$$\|Df(x)\| = \sup\left\{\frac{\|Df(x)v\|}{\|v\|} : v \in T_xM \text{ e } v \neq 0\right\}.$$

Our goal is to show that this norm is a upper bound for the topological entropy:

Proposition 2.46. *Let $f : M \to M$ be a differentiable map in a d dimensional Riemannian manifold M such that $\|Df\|$ is bounded. Then*

$$h(f) \leq d \log^+ \sup \|Df\| < \infty.$$

Proof. Consider $L = \sup\{\|Df(x)\| : x \in M\}$. By the mean value inequality, we have that

$$d(f(x), f(y)) \leq L d(x, y) \quad \text{for all } x, y \in M.$$

If $L \leq 1$ then, as we discussed before in the Example 2.34, the entropy of f is zero, as we wished to prove.

Thus, we may assume that $L > 1$. Let $\mathcal{A}$ be an atlas of M given by differentiable charts $\varphi : (-2, 2)^d \to M$. Given any compact set $K \subset M$, we may find a finite family $\mathcal{A}_K \subset \mathcal{A}$ such that

$$\left\{\varphi((-1, 1)^d) : \varphi \in \mathcal{A}_K\right\}$$

covers K. Fix $B > 0$ such that $d(\varphi(u), \varphi(v)) \leq Bd(u, v)$ for all $u, v \in [-1, 1]^d$ and all $\varphi \in \mathcal{A}_K$. Given $n \geq 1$ e $\varepsilon > 0$, fix $\delta = (\varepsilon / B\sqrt{d})L^{-n}$. We represent by $\delta\mathbb{Z}^d$ the set of points $(\delta k_1, \ldots, \delta k_d)$ with $k_j \in \mathbb{Z}$ for all $j = 1, \ldots, d$. Let $E \subset M$ be the union of images $\varphi(\delta\mathbb{Z}^d \cap (-1, 1)^d)$, with $\varphi \in \mathcal{A}_K$.

Observe that every point of $(-1, 1)^d$ is within a distance of $\delta\sqrt{d}$ from some point of $\delta\mathbb{Z}^d \cap (-1, 1)^d$. So, given any $\varphi \in \mathcal{A}_K$, every $x \in \varphi((-1, 1)^d)$ is within a distance $B\delta\sqrt{d}$ from some $a \in \varphi(\delta\mathbb{Z}^d \cap (-1, 1)^d)$. Then, by the choice of δ,

$$d(f^j(x), f^j(a)) \leq L^j B\delta\sqrt{d} < L^n B\delta\sqrt{d} = \varepsilon$$

for all $j = 0, \ldots, n-1$. This proves that E is (n, ε)-generator for K. By the other hand, by construction:

$$\#E \leq \#\mathcal{A}_K \#\big(\delta\mathbb{Z}^d \cap (-1, 1)^d\big) \leq \#\mathcal{A}_K (2/\delta)^d \leq \#\mathcal{A}_K (2B\sqrt{d}L^n/\varepsilon)^d.$$

So, the right-hand side expression is a upper bound for $g_n(f, \varepsilon, K)$ and , consequently,

$$g(f, \varepsilon, K) \leq \limsup_n \frac{1}{n} \log(2B\sqrt{d}L^n/\varepsilon)^d = d \log L.$$

Taking $\varepsilon \to 0$ and taking the supremum over K, we have that $h(f) \leq d \log L$. □

A central statement in the theory of topological entropy is the following conjecture proposed by Mike Shub [Shu74]:

Open Question 2.47 (Entropy Conjecture)**.** Let $f : M \to M$ be a C^1 diffeomorphism in a d dimensional Riemannian manifold, then

$$h(f) \geq \max_{1 \leq k \leq d} \log \rho(f_k) \tag{2.31}$$

where $\rho(f_k)$ is the spectral radius of the action $f_k : H_k(M) \to H_k(M)$ induced by f in the homology of dimension k with real coefficients.

Although several partial answers and related results are known, positive and negative, the complete statement of this conjecture remains open. Let's summarize some of the known facts.

It is known that the inequality (2.31) is true for an open subset and dense space of homeomorphisms on any variety with dimension $d \neq 4$. Furthermore it is true for all classes homeomorphism in certain varieties, such as spheres and infranilvariedades [MP77b, MP77a, MP08]. Moreover, Shub [Shu74] exhibited a Lipschitz homeomorphism with entropy zero that does not satisfy the condition (2.31).

A useful way to analyze (2.31) is to compare the entropy with each of spectral radii $\rho(f_k)$. The case $k = d$ is relatively easy. In fact, for any f continuous map in a manifold of dimension d, the spectral radius $\rho(f_d)$ equals the absolute value $|\deg f|$ of the degree of f. In particular, the inequality $h(f) \geq \log \rho(f_d)$ is trivial in the case of homeorphisms. For non-invertible discontinuous maps, the topological entropy can be smaller than the logarithm of the absolute value of the degree. But, it was shown in [MP77b] that for differentiable maps $h(f) \geq \log|\deg f|$ always holds true.

Manning cite Mann75 proved that the inequality $h(f) \geq \log\rho(f_1)$ is true for any homeomorphism in a manifold of any dimension d. It follows that $h(f) \geq \log\rho(f_{d-1})$, since the Poincaré duality theorem implies

$$\rho(f_k) = \rho(f_{d-k}) \quad \text{for all } 0 < k < d.$$

In particular, Manning's theorem together with previous observations show that the conjecture is valid for any homeomorphisms on a manifold of dimension $d \leq 3$.

Bowen [Bow78] proved that for any homeomorphism in a manifold $h(f)$ is greater than or equal to the logarithm of the growth rate of the fundamental group. It is shown that this growth rate is greater than or equal to the spectral radius $\rho(f_1)$. Therefore, this Bowen's result implies the theorem of Manning, just mentioned.

The main result concerning the stability conjecture is the Yomdin theorem [Yom87], under which the conjecture is true for all diffeomorphisms of class C^∞. The crucial ingredient of the proof is a relationship between the topological entropy $h(f)$ and rate $v(f)$ of volume growth (in all dimensions 1 *lek led*) when we iterate the diffeomorphism. For C^∞ diffeomorphisms these two numbers coincide (this is false in general if f is C^k only, with k finite). The conjecture of entropy is a consequence, because $\log\rho(f_*) \leq v(f)$ for all C^1 diffeomorphism.

In the case of the C^1 diffeomorphisms, it is also known that the inequality (2.31) is true for all Axiom A diffeomorphism without cycles [SW75], for some partially hyperbolic diffeomorphisms [SX10] and more generally for any diffeomorphism away from homoclinic tangencies, by a recent work of Liu, Viana and Yangang.

2.3.4 Linear Endomorphisms

Given a real number $x > 0$, we denote by $\log^+ x = \max\{\log x, 0\}$. In this section, we prove the following result:

Proposition 2.48. *Let $f_A : \mathbb{T}^d \to \mathbb{T}^d$ be the induced endomorphism on $\mathbb{T}^d$ by some hyperbolic invertible matrix A with integer coefficients. Let μ be the Haar (Lebesgue) measure on $\mathbb{T}^d$. Then*

$$h_\mu(f_A) = \sum_{i=1}^{d} \log^+ |\lambda_i|.$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of A, with multiplicity.

Proof. First, assume that A is diagonalizable. Let $v_1, \ldots, v_d$ be linear independent unitary vectors of $\mathbb{R}^d$ such that $Av_i = \lambda_i v_i$ for each i. Let u be the number of eigenvalues of A with modulus bigger than one. we may assume that the eigenvalues are numbered in such way that $|\lambda_i| > 1$ if, and only if, $i \le u$. Given $x \in \mathbb{T}^d$, each point y in a neighbourhood of x maybe written as

$$y = x + \sum_{i=1}^{d} t_i v_i$$

since $t_1, \ldots, t_d$ are close to zero. Given $\varepsilon > 0$, denote by $D(x, \varepsilon)$ the set of points y with $|t_i| < \varepsilon$ for every $i = 1, \ldots, d$. Moreover, for each $n \ge 1$, consider

$$D(x, n, \varepsilon) = \{y \in \mathbb{T}^d : f_A^j(y) \in D(f_A^j(x), \varepsilon) \text{ for every } j = 0, \ldots, n-1\}.$$

Observe that $f_A^j(y) = f_A^j(x) + \sum_{i=1}^d t_i \lambda_i^n v_i$ for every $n \ge 1$. Therefore,

$$D(x, n, \varepsilon) = \big\{x + \sum_{i=1}^{d} t_i v_i : |\lambda_i^n t_i| < \varepsilon \text{ for } i \le u \text{ and } |t_i| < \varepsilon \text{ para } i > u\big\}.$$

Thus, there exists a constant $C_1 > 1$ that depends only on A, such that

$$C_1^{-1} \varepsilon^d \prod_{i=1}^{u} |\lambda_i|^{-n} \le \mu(D(x, n, \varepsilon)) \le C_1 \varepsilon^d \prod_{i=1}^{u} |\lambda_i|^{-n}$$

for every $x \in \mathbb{T}^d$, $n \ge 1$ and $\varepsilon > 0$. It is clear that there exists $C_2 > 1$ depending only on A, such that

$$B(x, C_2^{-1}\varepsilon) \subset D(x, \varepsilon) \subset B(x, C_2\varepsilon)$$

for every $x \in \mathbb{T}^d$ and every $\varepsilon > 0$ small. Thus, $B(x, n, \varepsilon/C_2) \subset D(x, n, \varepsilon) \subset B(x, n, C_2\varepsilon)$ for every $n \ge 1$. Combining these two remarks, and taking $C = C_1 C_2^d$, we have that:

$$C^{-1} \varepsilon^d \prod_{i=1}^{u} |\lambda_i^{-n}| \le \mu(B(x, n, \varepsilon)) \le C \varepsilon^d \prod_{i=1}^{u} |\lambda_i^{-n}|$$

for every $x \in \mathbb{T}^d$, $n \ge 1$ and $\varepsilon > 0$. Then,

$$h_\mu^+(f, \varepsilon, x) = h_\mu^-(f, \varepsilon, x) = \lim_n \frac{1}{n} \log \mu\big(B(x, n, \varepsilon)\big) = \sum_{i=1}^{u} \log |\lambda_i|$$

for every $x \in \mathbb{T}$ and every $\varepsilon > 0$ small. Thus, using Brin-Katok Theorem

$$h_\mu(f) = h_\mu(f, x) = \sum_{i=1}^{u} \log |\lambda_i|$$

for μ-almost every point x. This proves the Proposition 2.48 in the diagonalizable case.

The general case can be proved in a similar way, using Jordan's canonical form for A.

□

Now, we prove that the topological entropy of a torus automorphism is in fact equal to the entropy of the Haar measure.

Proposition 2.49. *Let $f_A : \mathbb{T}^d \to \mathbb{T}^d$ be the induced endomorphism on the d-dimensional torus $\mathbb{T}^d$ by some invertible matrix A with integer entries. Then,*

$$h(f_A) = \sum_{j=1}^{d} \log^+ |\lambda_j|. \tag{2.32}$$

where $\lambda_1, \ldots, \lambda_d$ are the eigeinvalues of A, counted with multiplicity.

We proved at the Proposition 2.48 that the entropy of f_A with respect to the Haar measure μ is equal to the right-hand side of (2.32). By the variational principle, that we will show at Section 3.2, the topological entropy is greater of equal to the metric entropy of any invariant probability. This imply that

$$h(f_A) \geq h_\mu(f) = \sum_{j=1}^{d} \log^+ |\lambda_j|.$$

Here, we focus on the proof of the oposite inequality:

$$h(f_A) \leq \sum_{j=1}^{d} \log^+ |\lambda_j|. \tag{2.33}$$

Suppose that there exists a basis $v_1, \ldots, v_d$ of $\mathbb{R}^d$ with $Av_i = \lambda_i v_i$ for each i. Without loss of generatily, we may assume that $\|v_i\| = 1$ for all i. Moreover, renumbering the eigenvalues, if necessary, we may assume that $|\lambda_i| > 1$ for $1 \leq i \leq u$ and $|\lambda_i| \leq 1$ for all $i > u$. Let $e_1, \ldots, e_d$ be the canonical basis of $\mathbb{R}^d$ and let $P : \mathbb{R}^d \to \mathbb{R}^d$ be the linear isomorphism defined by $P(e_i) = v_i$ for each i. Then $P^{-1}AP$ is diagonal matrix. Fix $L > 0$ big enough such that $P((0, L)^d)$ contains some unit cube $\prod_{i=1}^{d}[b_i, b_i + 1]^d$. Let $\pi : \mathbb{R}^d \to \mathbb{T}^d$ be the canonical projection. Then, $\pi P((0, L)^d)$ contains the torus $\mathbb{T}^d$.

Given $n \geq 1$ and $\varepsilon > 0$, fix $\delta > 0$ such that $\|P\|\delta\sqrt{d} < \varepsilon$. Moreover, for each $i = 1, \ldots, d$, take

$$\delta_i = \begin{cases} \delta|\lambda_i|^{-n} & \text{se } i \leq u \\ \delta & \text{se } i > u \end{cases}$$

consider the set

$$E = \pi P\big(\{(k_1\delta_1, \ldots, k_d\delta_d) \in (0, L)^d : k_1, \ldots, k_d \in \mathbb{Z}\}\big)$$

Observe that, for any $j \geq 1$,

$$f^j(E) \subset \pi P\big(\{(k_1\lambda_1^j\delta_1, \ldots, k_d\lambda_d^j\delta_d) : k_1, \ldots, k_d \in \mathbb{Z}\}\big).$$

consider $0 \leq j < n$. By construction, $|k_i\lambda_i^j\delta_i| \leq \delta$ for all $i = 1, \ldots, d$. So, every point of $\mathbb{R}^d$ is within a distance less or equal than $\delta\sqrt{d}$ from some point

$(k_1\lambda_1^j\delta_1, \dots, k_d\lambda_d^j\delta_d)$. Then, for each $x \in \mathbb{T}^d$ we may find $a \in E$ such that $d(f^j(x), f^j(a)) \le \|P\|\delta\sqrt{d}$ for all $0 \le j < n$. This show that E is a (n, ε)-generating set of $\mathbb{T}^d$. From the another hand,

$$\#E \le \prod_{i=1}^{d} \frac{L}{\delta_i} = \left(\frac{L}{\delta}\right)^d \prod_{i=1}^{u} |\lambda_i|^n.$$

this show that $g_n(f_A, \varepsilon, \mathbb{T}^d) \le (L/\delta)^d \prod_{i=1}^{u} |\lambda_i|^n$ for all $n \ge 1$ and every $\varepsilon > 0$. Then,

$$h(f) = \lim_{\varepsilon \to 0} \limsup_{n} \frac{1}{n} g_n(f_A, \varepsilon, \mathbb{T}^d) \le \sum_{i=1}^{u} \log |\lambda_i| = \sum_{i=1}^{d} \log^+ |\lambda_i|.$$

This show the Proposition 2.49 if A is diagonalizable.

The general case can be proved in a similar way, using the Jordan form for A.

Chapter 3

Equilibrium States and Pressure

On this class, we will discuss two main things. The first one, is the relation between metric entropy and topological entropy. This relation is given by the following result:

Theorem 3.1 (Variational Principle)**.** *Let $f : M \to M$ be a continuous transformation on a compact metric space M. Then, its topological entropy $h(f)$ coincides with the supremum of the metric entropy $h_\mu(f)$ among all invariant probabilities.*

This was demonstrated originally by Dinaburg [Din70, Din71], Goodman [Goo71a] and Goodwin [Goo71b]. In the Section 3.2 we obtain as a particular case of a broader result, called *the variational principle for the pressure*, which is due to Walters [Wal75].

The *pressure* $P(f, \phi)$ is a weighted version of the topological entropy $h(f)$, where the *weights* are determined by a continuous function $\phi : M \to \mathbb{R}$, which is called *potential*. We will study these concepts and their properties in Section 3.1. The topological entropy corresponds to the particular case in which the potential is identically zero. The idea of pressure was brought from Statistical Mechanics to Ergodic theory by the mathematician and theoretical physicist David Ruelle, one of creators of differentiable ergodic theory, and was later extended by the British mathematician Peter Walters.

The variational principle can be extended to the pressure, as discussed in Section 3.2: for every continuous function ϕ,

$$P(f, \phi) = \sup \big\{ h_\mu(f) + \int \phi \, d\mu : \mu \text{ is invariant under} f \big\}. \tag{3.1}$$

The second main thing is the notion of equilibrium state. A invariant probability μ is called *equilibrium* for the potential ϕ if it achieves the supremum in (3.1), i.e., if $h_\mu(f) + \int \phi \, d\mu = P(f, \phi)$. The properties of equilibrium states will be studied in Section 3.3.

3.1 Pressure

In this section we introduce a important generalization of the topological entropy, called *pressure* (or *topological pressure*), and we study its main properties.

3.1.1 Definition by open covers

Let $f : M \to M$ be a continuous map in a compact metric space. We call *potential* on M any continuous function $\phi : M \to \mathbb{R}$. For each $n \in \mathbb{N}$, define $\phi_n : M \to \mathbb{R}$ by $\phi_n = \sum_{i=0}^{n-1} \phi \circ f^i$. Moreover, given any non-empty set $C \subset M$, denote

$$\phi_n(C) = \sup\{\phi_n(x) : x \in C\}. \tag{3.2}$$

Given a open cover α of M define

$$P_n(f, \phi, \alpha) = \inf \big\{ \sum_{U \in \gamma} e^{\phi_n(U)} : \gamma \text{ is a finite subcover of } \alpha^n \big\}. \tag{3.3}$$

This is a subadittive sequence and, therefore, the limit

$$P(f, \phi, \alpha) = \lim_n \frac{1}{n} \log P_n(f, \phi, \alpha) \tag{3.4}$$

is well defined. Finally, we define the *pressure* of ϕ with respect to f to any limit $P(f, \phi)$ of $P(f, \phi, \alpha)$ when the diameter of α goes to zero. The existence of this limit is given by the following lemma:

Lemma 3.2. *There exists* $\lim_{\operatorname{diam} \alpha \to 0} P(f, \phi, \alpha)$*, i.e., there exists* $P(f, \phi) \in [0, \infty]$ *such that*

$$\lim_k P(f, \phi, \alpha_k) = P(f, \phi)$$

for every sequence $(\alpha_k)_k$ *of open covers with* $\operatorname{diam} \alpha_k \to 0$*.*

Proof. Let $(\alpha_k)_k$ and $(\beta_k)_k$ be sequences of open covers with diameter going to zero. Given any $\varepsilon > 0$ fix $\delta > 0$ such that $|\phi(x) - \phi(y)| \le \varepsilon$, if $d(x, y) \le \delta$. By hypothesis, $\operatorname{diam} \alpha_k < \delta$ for all k big enough. For any fixed k, let $\rho > 0$ the Lebesgue number for α_k. By hypothesis, $\operatorname{diam} \beta_l < \rho$ for all l big enough. By the definition of Lebesgue number, any $B \in \beta_l$ is contained in some $A \in \alpha_k$. Observe that $\phi_n(A) \le n\varepsilon + \phi_n(B)$ for all $n \ge 1$, since $\operatorname{diam} \alpha_k < \delta$. This imply that

$$P_n(f, \phi, \alpha_k) \le e^{n\varepsilon} P_n(f, \phi, \beta_l) \quad \text{for all } n \ge 1$$

and, therefore, $P(f, \phi, \alpha_k) \le \varepsilon + P(f, \phi, \beta_l)$. Taking $l \to \infty$ and $k \to \infty$, we obtain that

$$\limsup_k P(f, \phi, \alpha_k) \le \varepsilon + \liminf_l P(f, \phi, \beta_l).$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\limsup_k P(f, \phi, \alpha_k) \le \liminf_l P(f, \phi, \beta_l)$. Changing the role of these covers, we conclude that there exist the limits $\lim_k P(f, \phi, \alpha_k)$ and $\lim_l P(f, \phi, \beta_l)$ are they are equal. ☐

Before continue, we observe some easy consequences of the definition of pressure. The very first one, is that the pressure of the null potenntial is the topological entropy. Indeed, it is immediate from (3.3) that $P_n(f,0,\alpha) = N(\alpha^n)$ for all $n \geq 1$ and, thus, $P(f,0,\alpha) = h(f,\alpha)$ for any open cover α. So,

$$P(f,0) = h(f). \tag{3.5}$$

Given any constant $c > 0$, we have that $P_n(f,\phi+c,\alpha) = e^{cn}P_n(f,\phi,\alpha)$ for all $n \geq 1$ and, $P(f,\phi+c,\alpha) = P(f,\phi,\alpha)+c$ for every open cover α. So,

$$P(f,\phi+c) = P(f,\phi+c). \tag{3.6}$$

Analogously, if $\phi \leq \psi$ then $P_n(f,\phi,\alpha) \leq P_n(f,\psi,\alpha)$ for all $n \geq 1$ and, therefore, $P(f,\phi,\alpha) = P(f,\psi,\alpha)$ for every open cover α. It means that

$$\phi \leq \psi \quad \Rightarrow \quad P(f,\phi) \leq P(f,\psi). \tag{3.7}$$

In particular, since $\inf\phi \leq \phi \leq \sup\phi$, we have that

$$h(f) + \inf\phi \leq P(f,\phi) \leq h(f) + \sup\phi \tag{3.8}$$

for any potencial ϕ. An interesting corollary show that if $h(f)$ is finite, then $P(f,\phi) < \infty$ for any potencial ϕ and otherwise, $P(f,\phi) = \infty$ for all potentials ϕ. An example of this can be seen at Example 2.29.

Another direct consequence of the definition is that the pressure is an invariant for the topological equivalence:

Proposition 3.3. *Let $f_i : M_i \to M_i$, $i = 1,2$ be continuous maps in compact metric spaces. If there exists an homeomorphism $h : M_1 \to M_2$ such that $h \circ f_1 = f_2 \circ h$ then $P(f_2,\phi) = P(f_1,\phi\circ h)$ for any potencial ϕ in M_2.*

Remark 3.4. It is possible to replace the supremum by the infimum in (3.2), however the definition is a little bit more complicated. To see this, consider:

$$\underline{P}_n(f,\phi,\alpha) = \inf\big\{\sum_{U\in\gamma} e^{\underline{\phi}_n(U)} : \gamma \text{ is a finite subcover of } \alpha^n\big\}$$

where $\underline{\phi}_n(C) = \inf\{\phi_n(x) : x \in C\}$. The sequence $\underline{P}_n(f,\phi,\alpha)$ can not be subadditive. From the other hand, since ϕ is (uniformly) continous, given any $\varepsilon > 0$ there exists $\delta > 0$ such that $\underline{\phi}_n(C) \leq \phi_n(C) \leq n\varepsilon + \underline{\phi}_n(C)$ if $\operatorname{diam} C \leq \delta$. Then,

$$\underline{P}_n(f,\phi,\alpha) \leq P_n(f,\phi,\alpha) \leq e^{n\varepsilon}\underline{P}_n(f,\phi,\alpha)$$

for every open cover of α with $\operatorname{diam}\alpha \leq \delta$. Consequently,

$$\begin{aligned} P(f,\phi) &= \lim_{\operatorname{diam}\alpha\to 0} \limsup_n \frac{1}{n}\log \underline{P}_n(f,\phi,\alpha) \\ &= \lim_{\operatorname{diam}\alpha\to 0} \liminf_n \frac{1}{n}\log \underline{P}_n(f,\phi,\alpha). \end{aligned} \tag{3.9}$$

3.1.2 Generating sets and separated sets

Noow we present two alternative definitions for the pressure, using generating sets and separated sets, respectively. As before, $f : M \to M$ is a continuous map in a compact metric space and $\phi : M \to \mathbb{R}$ is a continuous function.

Given $n \geq 1$ and $\varepsilon > 0$, define

$$\begin{aligned} G_n(f,\phi,\varepsilon) &= \inf\Big\{\sum_{x\in E} e^{\phi_n(x)} : E \text{ is a } (n,\varepsilon)\text{-generating set of } M\Big\} \quad \text{e} \\ S_n(f,\phi,\varepsilon) &= \sup\Big\{\sum_{x\in E} e^{\phi_n(x)} : E \text{ is a } (n,\varepsilon)\text{-separated set of } M\Big\}. \end{aligned} \tag{3.10}$$

Now, define

$$\begin{aligned} G(f,\phi,\varepsilon) &= \limsup_n \frac{1}{n}\log G_n(f,\phi,\varepsilon) \quad \text{e} \\ S(f,\phi,\varepsilon) &= \limsup_n \frac{1}{n}\log S_n(f,\phi,\varepsilon) \end{aligned} \tag{3.11}$$

and

$$G(f,\phi) = \lim_{\varepsilon\to 0} G(f,\phi,\varepsilon) \quad \text{e} \quad S(f,\phi) = \lim_{\varepsilon\to 0} S(f,\phi,\varepsilon) \tag{3.12}$$

(the limits above exists, since the functions are monotone in ε).

Note that $G_n(f,0,\varepsilon) = g_n(f,\varepsilon)$ and $S_n(f,0,\varepsilon) = s_n(f,\varepsilon)$ for all $n \geq 1$ and every $\varepsilon > 0$. Then, (Proposition 2.33), $G(f,0) = g(f)$ and $S(f,0) = s(f)$ are equal to the topological entropy $h(f)$. In general, we have:

Proposition 3.5. *$P(f,\phi) = G(f,\phi) = S(f,\phi)$ for any potencial ϕ on M.*

Proof. Consider $n \geq 1$ and $\varepsilon > 0$. It is clear from the definitions that any (n,ε)-separated maximal set is a (n,ε)-generating set. Then,

$$\begin{aligned} S_n(f,\phi,\varepsilon) &= \sup\Big\{\sum_{x\in E} e^{\phi_n(x)} : E \text{ is a } (n,\varepsilon)\text{-separated set}\Big\} \\ &= \sup\Big\{\sum_{x\in E} e^{\phi_n(x)} : E \text{ is a } (n,\varepsilon)\text{-separated maximal set}\Big\} \\ &\geq \inf\Big\{\sum_{x\in E} e^{\phi_n(x)} : E \text{ is a } (n,\varepsilon)\text{-generating set}\Big\} = G_n(f,\phi,\varepsilon). \end{aligned} \tag{3.13}$$

This imply thar $G(f,\phi,\varepsilon) \leq S(f,\phi,\varepsilon)$. Taking the limit when $\varepsilon \to 0$, we obtain that $G(f,\phi) \leq S(f,\phi)$.

Now, we prove that $S(f,\phi) \leq P(f,\phi)$. Let ε and δ be positive numbers such that $d(x,y) \leq \delta$ implies $|\phi(x) - \phi(y)| \leq \varepsilon$. Let α be any open cover of M with $\operatorname{diam}\alpha < \delta$. Let $E \subset M$ be any (n,δ)-separated set. Given any subcover γ of α^n, it is clear that any point of E is contained in some element of γ. From the other hand, the hypothesis that E is a (n,δ)-separated set implies that any element of γ contains at most one element of E. So,

$$\sum_{x\in E} e^{\phi_n(x)} \leq \sum_{U\in\gamma} e^{\phi_n(U)}.$$

Taking the supremum on E and the infimum on γ, we obtain that

$$S_n(f,\phi,\delta) \le P_n(f,\phi,\alpha). \tag{3.14}$$

It follows that $S(f,\phi,\delta) \le P(f,\phi,\alpha)$. Taking $\delta \to 0$ (thus, $\operatorname{diam}\alpha \to 0$), we conclude that $S(f,\phi) \le P(f,\phi)$, as we claimed before.

Finally, we proved that $P(f,\phi) \le G(f,\phi)$. Let ε and δ be positive numbers such that $d(x,y) \le \delta$ implies $|\phi(x)-\phi(y)| \le \varepsilon$. Let α be any open cover of M with $\operatorname{diam}\alpha < \delta$ and $\rho > 0$ a Lebesgue number for α. If $E \subset M$ is a (n,ρ)-generating set, for each $x \in E$ and $i = 0,\dots,n-1$, there exists $A_{x,i} \in \alpha$ such that $B(f^i(x),\rho)$ is contained in $A_{x,i}$. Denote by

$$\gamma(x) = \bigcap_{i=0}^{n-1} f^{-i}(A_{x,i}).$$

Observe that $\gamma(x) \in \alpha^n$ and that $B(x,n,\rho) \subset \gamma(x)$. So, the hypothesis that E is a (n,ρ)-generating set implies that $\gamma = \{\gamma(x) : x \in E\}$ is a subcover of α. Observe that

$$\phi_n(\gamma(x)) \le n\varepsilon + \phi_n(x) \quad \text{for all } x \in E,$$

since $\operatorname{diam} A_{x,i} < \delta$ for all i. Consequently,

$$\sum_{U\in\gamma} e^{\phi_n(U)} \le e^{n\varepsilon} \sum_{x\in E} e^{\phi_n(x)}.$$

This proves that $P_n(f,\phi,\alpha) \le e^{n\varepsilon} G_n(f,\phi,\rho)$ for all $n \ge 1$ and, consequently,

$$P(f,\phi,\alpha) \le \varepsilon + \liminf_n \frac{1}{n} G_n(f,\phi,\rho) \le \varepsilon + G(f,\phi,\rho). \tag{3.15}$$

Taking $\rho \to 0$ we have that $P(f,\phi,\alpha) \le \varepsilon + G(f,\phi)$. Thus, taking ε, δ e $\operatorname{diam}\alpha$ going to zero, $P(f,\phi) \le G(f,\phi)$. □

The conclusion of the Proposition 3.5 maybe rewrite as:

$$P(f,\phi) = \lim_{s\to 0}\limsup_n \frac{1}{n} G_n(f,\phi,s) = \lim_{s\to 0}\limsup_n \frac{1}{n} S_n(f,\phi,s), \tag{3.16}$$

since the equations (3.15) and (3.13) in the proof also give that

$$P(f,\phi) \le \lim_{s\to 0}\liminf_n \frac{1}{n} G_n(f,\phi,s) \le \lim_{s\to 0}\liminf_n \frac{1}{n} S_n(f,\phi,s).$$

Comparing with (3.16), we have that:

$$P(f,\phi) = \lim_{s\to 0}\liminf_n \frac{1}{n} G_n(f,\phi,s) = \lim_{s\to 0}\liminf_n \frac{1}{n} S_n(f,\phi,s). \tag{3.17}$$

3.1.3 Properties

First, we prove a version of Corollary 2.37 for the pressure:

Proposition 3.6. *Let $f : M \to M$ be a continuous map in a compact metric space. Let β be a open cover of M such that*

(1) $\operatorname{diam}\beta^k$ *converges to zero when $k \to \infty$, or*

(2) $f : M \to M$ is a homeomorphism and $\operatorname{diam}\beta^{\pm k} \to 0$ *when $k \to \infty$.*

Then $P(f,\phi) = P(f,\phi,\beta)$ for any potential ϕ on M.

Proof. First, we prove the following lemma:

Lemma 3.7. *$P(f,\phi,\alpha^k) = P(f,\phi,\alpha)$ for every open cover α and every $k \geq 1$.*

Proof. By definition, for any $n \geq 1$:

$$P_n(f,\phi,\alpha^k) = \inf\{\sum_{U\in\gamma} e^{\phi_n(U)} : \gamma \text{ finite subcover of } (\alpha^k)^n\} \quad \text{e}$$

$$P_{n+k}(f,\phi,\alpha) = \inf\{\sum_{U\in\gamma} e^{\phi_{n+k}(U)} : \gamma \text{ finite subcover of } \alpha^{n+k}\}.$$

It is clear that $(\alpha^k)^n = \alpha^{n+k}$. Then, denoting by $L = \sup|\phi|$,

$$e^{-kL} P_n(f,\phi,\alpha^k) \leq P_{n+k}(f,\phi,\alpha) \leq e^{kL} P_n(f,\phi,\alpha^k)$$

for all $n \geq 1$. This implies that $P(f,\phi,\alpha^k) = P(f,\phi,\alpha)$. □

Lemmas 3.2 and 3.7 give us that $P(f,\phi) = \lim_k P(f,\phi,\beta^k) = P(f,\phi,\beta)$. This proves the item (1) of Proposition 3.6. For item (2) we also make use of the following fact:

Lemma 3.8. *Let f be a homeomorphism, then $P(f,\phi,f^{-1}(\alpha)) = P(f,\phi,\alpha)$ for every open cover α.*

Proof. By definition, given any $n \geq 1$,

$$P_n(f,\phi,\alpha) = \inf\{\sum_{U\in\gamma} e^{\phi_n(U)} : \gamma \text{ finite subcover of } \alpha^n\} \quad \text{e}$$

$$P_n(f,\phi,f^{-1}(\alpha)) = \inf\{\sum_{V\in\delta} e^{\phi_n(V)} : \delta \text{ finite subcover of } (f^{-1}(\alpha))^n\}.$$

Fix $L = \sup|\phi|$. Note that $(f^{-1}(\alpha))^n = f^{-1}(\alpha^n)$ and every finite subcover of $f^{-1}(\alpha^n)$ can be written as $f^{-1}(\gamma)$ for some open subcover γ of α^n. Moreover,

$$\phi_n(U) - 2L \leq \phi_n(f^{-1}(U)) \leq \phi_n(U) + 2L$$

For any $U \subset M$. It follows that,

$$e^{-2L} P_n(f,\phi,\alpha) \leq P_n(f,\phi,f^{-1}(\alpha)) \leq e^{-2L} P_n(f,\phi,\alpha)$$

for all $n \geq 1$. So, $P(f,\phi,f^{-1}(\alpha)) = P(f,\phi,\alpha)$. □

Corollary 3.9. *If f is a homeomorphism then $P(f,\phi,\alpha^{\pm k}) = P(f,\phi,\alpha)$ for any open cover α and every $k \geq 1$.*

Proof. It is clear from the definition that $\alpha^{\pm k} = f^{-k}(\alpha^{2k})$. So, the corrolary follows directly from 3.7 and 3.8. □

To finish the proof of Proposition 3.6, we just observe that the item (2) of 3.6 is a direct consequence of Lemma 3.2 and Corollary 3.9:

$$P(f,\phi) = \lim_k P(f,\phi,\beta^{\pm k}) = P(f,\phi,\beta).$$

This conclude the proof. □

Proposition 3.10. *Let $f : M \to M$ be a continuous map in a compact metric space and let ϕ be a potential on M. Then:*

(1) $P(f^k,\phi_k) = kP(f,\phi)$ for all $k \geq 1$.

(2) If f is a homeomorphism $P(f^{-1},\phi) = P(f,\phi)$.

Proof. Let α be a open cover of M and define $\beta = \alpha^k$. Given a potential ϕ on M, denote $\psi = \phi_k$. Observe that $\beta^n = \alpha^{kn}$ for each n and that $\psi_n = \phi_{kn}$, where $\psi_n = \sum_{j=0}^{n-1} \psi \circ f^k$. Then,

$$\begin{aligned} P_n(f^k,\psi,\beta) &= \inf\{\sum_{U\in\gamma} e^{\psi_n(U)} : \gamma \subset \beta^n\} \\ &= \inf\{\sum_{U\in\gamma} e^{\phi_{kn}(U)} : \gamma \subset \alpha^{kn}\} = P_{kn}(f,\phi,\alpha). \end{aligned}$$

Consequently, $P(f^k,\psi,\beta) = kP(f,\phi,\alpha)$ for every α. Taking $\operatorname{diam}\alpha \to 0$ (observe that $\operatorname{diam}\beta \to 0$), we deduce that $P(f^k,\psi) = kP(f,\phi)$. This prove item (1).

To prove item (2), let α be an open cover of M. For any $n \geq 1$, denote

$$\alpha_-^n = \alpha \vee f(\alpha) \vee \cdots \vee f^{n-1}(\alpha) \quad \text{e} \quad \phi_n^- = \sum_{j=0}^{n-1} \phi \circ f^{-j}.$$

Observe that $\alpha_-^n = f^{n-1}(\alpha^n)$ and that γ is a finite subcover of α^n if, and only if, $\delta = f^{n-1}(\gamma)$ is a finite subcover of α_-^n. Moreover,

$$\phi_n(U) = \phi_n^-(f^{n-1}(U)),$$

for any $U \subset M$. Putting these two facts together, we obtain that

$$\begin{aligned} P_n(f,\phi,\alpha) &= \inf\big\{\sum_{U\in\gamma} e^{\phi_n(U)} : \gamma \text{ finite subcover of } \alpha^n\big\} \\ &= \inf\big\{\sum_{V\in\delta} e^{\phi_n^-(V)} : \delta \text{ finite subcover of } \alpha_-^n\big\} = P_n(f^{-1},\phi,\alpha) \end{aligned}$$

for all $n \geq 1$. Then $P(f,\phi,\alpha) = P(f^{-1},\phi,\alpha)$ and, taking $\operatorname{diam}\alpha \to 0$, we have that $P(f,\phi) = P(f^{-1},\phi)$. □

In the following, we fix $f : M \to M$ and we consider $P(f, \cdot)$ as a function defined on $C^0(M, \mathbb{R})$, the space of continuous functions, with the supremum norm

$$\|\varphi\| = \sup\{|\varphi(x)| : x \in M\}.$$

As we observed in (3.8), if the topological entropy $h(f)$ is infinity, then the pressure must be is identically equal to ∞. In what follows, we shall assume that $h(f)$ is finite. Then, $P(f, \phi)$ is finite for any potential ϕ.

Proposition 3.11. *The pressure function is Lipschitz:*

$$|P(f, \phi) - P(f, \psi)| \leq \|\phi - \psi\|$$

for any ϕ and ψ.

Proof. It is clear that $\phi \leq \psi + \|\phi - \psi\|$. So, by (3.6) and (3.7), we have that $P(f, \phi) \leq P(f, \psi) + \|\phi - \psi\|$. Exchanging the roles of ϕ and ψ we obtain the other inequality. □

In general, the pressure is not a differentiable function.

Proposition 3.12. *The pressure function is convex:*

$$P(f, (1-t)\phi + t\psi) \leq (1-t)P(f, \phi) + tP(f, \psi)$$

for any ϕ and ψ on M and for all $0 \leq t \leq 1$.

Proof. Write $\xi = (1-t)\phi + t\psi$. Then $\xi_n = (1-t)\phi_n + t\psi_n$ for all $n \geq 1$ and, therefore, $\xi_n(U) \leq (1-t)\phi_n(U) + t\psi_n(U)$ for all $U \subset M$. So, by Hölder inequality (Theorem 5.16),

$$\sum_{U \in \gamma} e^{\xi_n(U)} \leq \big(\sum_{U \in \gamma} e^{\phi_n(U)}\big)^{1-t} \big(\sum_{U \in \gamma} e^{\psi_n(U)}\big)^t$$

for any finite family γ of subsets of M. This imply that given any finite open cover α,

$$P_n(f, \xi, \alpha) \leq P_n(f, \phi, \alpha)^{1-t} P_n(f, \psi, \alpha)^t$$

for all $n \geq 1$ and, therefore, $P(f, \xi, \alpha) \leq (1-t)P(f, \phi, \alpha) + tP(f, \psi, \alpha)$. Taking the limit when $\operatorname{diam} \alpha \to 0$, we finish the proof of the proposition. □

We say that two potentials $\phi, \psi : M \to \mathbb{R}$ are *cohomologous* if there exists a continuous function $u : M \to \mathbb{R}$ such that $\phi = \psi + u \circ f - u$. Observe that this is a equivalence relation in the space of potentials. .

Proposition 3.13. *Let $f : M \to M$ be a continuous map on a compact metric space. If $\phi, \psi : M \to \mathbb{R}$ are cohomologous potentials, then $P(\phi, f) = P(\psi, f)$.*

Proof. If $\psi = \phi + u \circ f - u$ then $\psi_n(x) = \phi_n(x) + u(f^n(x)) - u(x)$ for all $n \in \mathbb{N}$. Take $K = \sup |u|$. Then $|\psi_n(C) - \phi_n(C)| \leq 2K$ for any $C \subset M$. So, for any open cover γ,

$$e^{-2K} \sum_{U \in \gamma} e^{\phi_n(U)} \leq \sum_{U \in \gamma} e^{\psi_n(U)} \leq e^{2K} \sum_{U \in \gamma} e^{\phi_n(U)}.$$

This imply that, given any open cover α of M,

$$e^{-2K} P_n(f, \phi, \alpha) \leq P_n(f, \psi, \alpha) \leq e^{2K} P_n(f, \phi, \alpha)$$

for all n. So, $P(f, \phi, \alpha) = P(f, \psi, \alpha)$ for all α and, consequently, $P(f, \phi) = P(f, \psi)$. □

3.1.4 Some Comments on Statistical Mechanics

Let us pause to explain the connection between the mathematical concept of pressure and the questions of Physics that are the source of much the theory presented in this chapter and previous chapter. The reader interested in pursuing this issue should consult the classic work of Ruelle [Rue].

The (equilibrium) Statistical Mechanics aims to describe the Physical properties of the systems formed by a large number of units (or particles, or sites) that interact with each other. The *Avogadro constant* 6.022 *times*10^{23} gives an idea of what is meant by 'great' in specific examples.

The biggest challenge is trying to understand the physical phenomena of *phase transition* (i.e., change of state) as, for instance, the passage of water from liquid to solid state. The mathematical methods developed for this purpose have been very fruitful in other contexts, such as quantum field theory and nearest the scope of this book, the Ergodic Theory of the Uniformly Hyperbolic Dynamical Systems(Bowen [Bow75]).

For the purposes of mathematical modeling, it is convenient to assume that the set M of units that form the system is infinite. Examples are the most studied lattices $M = \mathbb{Z}^d$, such models are often called *crystal lattices.* Furthermore, it is usual to assume that each unit $x \in M$ admits a finite set Ω_x of possible values. For instance, $\Omega_x = \{-1, +1\}$ in the case of emph spin systems (-1 means that particle x has spin down, $+1$ means that the spin points up) and $\Omega = \{0, 1\}$ in the case of emph lattice gases (1 means that the site x is occupied by a molecule of gas, 0 means that the site is empty).

Then the *configuration space* is a subset of the system Ω the product $\prod_{x\ inM} \Omega_x$ and a *state* of the system is a probability measure on Ω. A *equilibrium state* describe a macroscopic configuration of the system that can be physically observed, ie, what actually occurs in nature. And a phase transition corresponds to the coexistence of more than a equilibrium state.

According to *variational principle* of Statistical Mechanics, which dates back to the *law of least effort* of Maupertuis, states that an equilibrium is characterized by minimizing a certain fundamental quantity such as, for instance, the

Gibbs free energy or pressure. In the case of crystal lattices, it is usual to assume that the configuration space Ω is invariant by the shifts

$$(x_n^1, \ldots, x_n^{i-1}, x_n^i, x_n^{i+1}, \ldots, x_n^d) \mapsto (x_n^1, \ldots, x_n^{i-1}, x_{n+1}^i, x_n^{i+1}, \ldots, x_n^d).$$

This allows you to connect such systems with the Ergodic Theory. In particular, we prove that under appropriate assumptions, the equilibrium states are measures of a certain type, called *Gibbs states*, which are invariant by shifts.

We will describe the concept of Gibbs state and detail some earlier ideas in the context of crystal lattices. Before that, just to motivate, we consider a particularly simple model, formed by a single unidade with set Ω finite. The *entropy* of a probability $\mu \in \Omega$ is the number

$$S(\mu) = \sum_{\xi \in \Omega} -\mu(\xi) \log \mu(\xi).$$

To each state $\xi \in \Omega$ we correspond a number $E(\xi)$ for the energy. Denote by $E(\mu)$ the average energy:

$$E(\mu) = \sum_{\xi \in \Omega} \mu(\xi) E(\xi).$$

We assume that the system is at an absolute temperature T which is kept constant throughout time. The *Gibbs free energy* is defined by:

$$G(\mu) = E(\mu) - \kappa T S(\mu)$$

where $\kappa = 1.380 \times 10^{-23} m^2 kg s^{-2} K^{-1}$ is called the *Boltzmann constant.* In other words, denoting by $\beta = 1/(\kappa T)$, we have that

$$-\beta G(\mu) = \sum_{\xi \in \Omega} \mu(\xi)[-\beta E(\xi) - \log \mu(\xi)]. \tag{3.18}$$

It is easy to show that (3.18) is maximal (and, therefore, the Gibbs free energy $G(\mu)$ is minimal) if, and only if:

$$\mu(\xi) = \frac{e^{-\beta E(\xi)}}{\sum_{\eta \in \Omega} e^{-\beta E(\eta)}} \quad \text{for all } \xi \in \Omega \tag{3.19}$$

(see Lemma 3.17 below). Therefore, emphGibbs distribution μ given by (3.19) is the unique equilibrium state of the system.

3.2 Variational Principle

The variational principle for the pressure, was proved originally by Ruelle [Rue73], in a more restrict setting, and by Walters [Wal75], in the following setting:

Theorem 3.14 (Variational Principle). *Let $f : M \to M$ be a continuous map on a compact metric space, and let $\mathcal{M}(f)$ be the set of all invariant probabilities of f. Then, for any continuous function $\phi : M \to \mathbb{R}$,*

$$P(\phi, f) = \sup\{h_\nu(f) + \int \phi\, d\nu : \nu \in \mathcal{M}(f)\}.$$

The Theorem 3.1 corresponds to the case when $\phi \equiv 0$. In particular, it follows that f has zero topological entropy if, and only if, $h_\nu(f) = 0$ for every invariant probability ν. This is the case, for instance, of circle homeomorphisms (Example 2.27) and transformations in metrizable compact groups (Example 2.34).

In Sections 3.2.1 and 3.2.2 we present a proof of Theorem 3.14 by Misiurewicz [Mis76]. Before that, let us discuss some consequences.

Corollary 3.15. *Let $f : M \to M$ be a continuous map on a compact metric space and let $\mathcal{M}_e(f)$ be the set of all ergodic invariant probabilities. Then,*

$$P(\phi, f) = \sup\{h_\nu(f) + \int \phi\, d\nu : \nu \in \mathcal{M}_e(f)\}.$$

Proof. Given any $\nu \in \mathcal{M}(f)$, consider $(\nu_P)_P$ its ergodic decomposition. By the Ergodic Decomposition Theorem,

$$h_\nu(f) + \int \phi\, d\nu = \int \left(h_{\nu_P}(f) + \int \phi\, d\nu_P\right) d\hat{\mu}(P).$$

This imply that

$$\sup\{h_\nu(f) + \int \phi\, d\nu : \nu \in \mathcal{M}(f)\} \le \sup\{h_\nu(f) + \int \phi\, d\nu : \nu \in \mathcal{M}_e(f)\}.$$

The oposity inequality is trivial, since $\mathcal{M}_e(f) \subset \mathcal{M}(f)$. Now, we just apply the Theorem 3.14. □

Another interesting consequence is that, for maps with finite topological entropy, the pressure function determines the set of invariant proabiilities:

Corollary 3.16 (Walters). *Let $f : M \to M$ be a continuous map on a compact metric space with finite topological entropy. Let η be a finite measure with sign on M.Then, η is a invariant probability by f if, and only if, $\int \phi\, d\eta \le P(f, \phi)$ for every continuous function $\phi : M \to \mathbb{R}$.*

Proof. The 'only if' part is a direct consequence of Theorem 3.14: if η is an invariant probability

$$P(f, \phi) \ge h_\eta(f) + \int \phi\, d\eta \ge \int \phi\, d\eta$$

for every continuous function ϕ. In the following, we prove the converse.

Let η be a finite measure with sign such that $\int \phi \, d\eta \le P(f,\phi)$ for every ϕ. Consider any $\phi \ge 0$. For every $c > 0$ and $\varepsilon > 0$,

$$c \int (\phi + \varepsilon) \, d\eta = - \int -c(\phi + \varepsilon) \, d\eta \ge -P(f, -c(\phi + \varepsilon)).$$

By (3.8), we have that

$$P(f, -c(\phi + \varepsilon)) \le h(f) + \sup\big(-c(\phi+\varepsilon)\big) = h(f) - c \inf(\phi + \varepsilon).$$

However, $c \int (\phi + \varepsilon) \, d\eta \ge -h(f) + \inf(\phi + \varepsilon)$. When $c > 0$ is big enough, the right-hand side of this inequality is positive. So, $\int (\phi + \varepsilon) \, d\eta > 0$. Since $\varepsilon > 0$ is arbitrary, this imply that $\int \phi \, d\eta \ge 0$ for any potential $\phi \ge 0$. Thus, η is a positive measure.

The next step is to show that η is a probability. By hypothesis,

$$\int c \, d\eta \le P(f, c) = h(f) + c$$

for all $c \in \mathbb{R}$. For $c > 0$ this imply that $\eta(M) \le 1 + h(f)/c$. Taking the limit when $c \to +\infty$, we obtain that $\eta(M) \le 1$. Analogously, considering $c < 0$ and taking the limit when $c \to -\infty$, we have that $\eta(M) \ge 1$. Therefore, η is a probability, as we claimed before.

Now, we just need to prove that η is invariant by f. By hypothesis, given $c \in \mathbb{R}$ and any ϕ,

$$c \int (\phi \circ f - \phi) \, d\eta \le P(f, c(\phi \circ f - \phi)).$$

By Proposition 3.13, the expression at the right-hand side is equal to $P(f, 0) = h(f)$. For $c > 0$, this imply that

$$\int (\phi \circ f - \phi) \, d\eta \le \frac{h(f)}{c}$$

and, taking the limit when $c \to +\infty$, it follows that $\int (\phi \circ f - \phi) \, d\eta \le 0$. The same argument aplied to $-\phi$ give us that $\int (\phi \circ f - \phi) \, d\eta \ge 0$. So, $\int \phi \circ f \, d\eta = \int \phi \, d\eta$ for all potential ϕ. This imply that $f_* \eta = \eta$. □

3.2.1 Upper Bound Proof

In this section we prove that, given any invariant probability ν,

$$h_\nu(f) + \int \phi \, d\nu \le P(f, \phi). \tag{3.20}$$

In order to do it, let $\mathcal{P} = \{P_1, \dots, P_s\}$ be any finite partition. Let us show that if α is an open cover of M with sufficient small diameter, depending only on $\mathcal{P}$, then

$$h_\nu(f, \mathcal{P}) + \int \phi \, d\nu \le \log 4 + P(f, \phi, \alpha). \tag{3.21}$$

Taking $\operatorname{diam}\alpha \to 0$, we have that $h_\nu(f,\mathcal{P}) + \int \phi\, d\nu \le \log 4 + P(f,\phi)$ for any finite partition $\mathcal{P}$. So, $h_\nu(f)+\int \phi\, d\nu \le \log 4+P(f,\phi)$. Now, replace f by f^k and the potential ϕ by ϕ_k. Observe that $\int \phi_k\, d\nu = k\int \phi\, d\nu$, since ν is f-invariant. Using Propositions 2.17 and 3.10, we have that

$$kh_\nu(f,\mathcal{P}) + k\int \phi\, d\nu \le \log 4 + kP(f,\phi)$$

for all $k \ge 1$. Dividing by k and taking the limit when $k \to \infty$ we obtain the inequality (3.20).

Now, it is enough to show (3.21). We use the following basic fact:

Lemma 3.17. *Let $a_1,\dots,a_k$ be real numbers and let $p_1,\dots,p_k$ be non-negative numbers such that $p_1+\dots+p_k = 1$. Seja $A = \sum_{i=1}^k e^{a_i}$.Then,*

$$\sum_{i=1}^k p_i(a_i - \log p_i) \le \log A.$$

Moreover, the equality holds true if, and only if, $p_j = a_j/A$ for all j.

Proof. Write $t_i = e^{a_i}/A$ e $x_i = p_i/e^{a_i}$. Note que $\sum_{i=1}^k t_i = 1$. By convexity,

$$\sum_{i=1}^k t_i\phi(x_i) \le \phi(\sum_{i=1}^k t_ix_i).$$

Observe that $t_i\phi(x_i) = (p_i/A)(a_i - \log p_i)$ and that $\sum_{i=1}^k t_ix_i = 1/A$. So, the previous inequality may be rewritten as:

$$\sum_{i=1}^k \frac{p_i}{A}(a_i - \log p_i) \le \frac{1}{A}\log A.$$

Multiplying by A, we obtain the inequality as in the lemma. Moreover, we have the if, and only if, x_i are all equal, i.e., if exists c such that $p_i = ce^{a_i}$ for all i. Summing over $i = 1,\dots,k$ we see in this case that $c = 1/A$, as we claimed before in the statement. □

Since ν is a regular measure, given $\varepsilon > 0$ we may find compact sets $Q_i \subset P_i$ such that $\nu(P_i \setminus Q_i) < \varepsilon$ for all $i = 1,\dots,s$. Let Q_0 the complement of $\cup_{i=1}^s Q_i$ and let $P_0 = \emptyset$. Then, $\mathcal{Q} = \{Q_0,Q_1,\dots,Q_s\}$ is a finite partition of M satisfying $\nu(P_i\Delta Q_i) < s\varepsilon$ for all $i = 0,1,\dots,s$. Then, by Lemma 2.8,

$$H_\nu(\mathcal{P}/\mathcal{Q}) \le \log 2$$

if $\varepsilon > 0$ is small enough(depending only on s). Fix ε and $\mathcal{Q}$ and assume that the cover α is such that

$$\operatorname{diam}\alpha < \min\{d(Q_i,Q_j) : 1 \le i < j \le s\}. \tag{3.22}$$

By Lemma 2.14, we have that $h_\nu(f,\mathcal{P}) \le h_\nu(f,\mathcal{Q}) + H_\nu(\mathcal{P}/\mathcal{Q}) \le h_\nu(f,\mathcal{Q}) + \log 2$. Thus, to prove that (3.21) it is enough to show that

$$h_\nu(f,\mathcal{Q}) + \int \phi \, d\nu \le \log 2 + P(f,\phi,\alpha). \tag{3.23}$$

In order to do it, observe that

$$H_\nu(\mathcal{Q}^n) + \int \phi_n \, d\nu \le \sum_{Q\in\mathcal{Q}^n} \nu(Q)\big(-\log\mu(Q) + \phi_n(Q)\big)$$

for all $n \ge 1$. Then, by Lemma 3.17,

$$H_\nu(\mathcal{Q}^n) + \int \phi_n \, d\nu \le \log\big(\sum_{Q\in\mathcal{Q}^n} e^{\phi_n(Q)}\big). \tag{3.24}$$

Let γ be any finite subcover of α^n. For each $Q \in \mathcal{Q}^n$ consider x_Q a point in the closure of Q such that $\phi_n(x_Q) = \phi_n(Q)$ (remember that $\phi_n(Q)$ denote the supremum of ϕ_n on the set Q). Consider $U_Q \in \gamma$ such that $x_Q \in U_Q$. Then, $\phi_n(Q) \le \phi_n(U_Q)$ for all $Q \in \mathcal{Q}^n$. The condition (3.22) imply that each element of α intersects the closure of no more than two elements of $\mathcal{Q}$. Therefore, each element of α^n intersects the closure of, at most, 2^n elements of $\mathcal{Q}^n$. In particular, for each $U \in \gamma$ there exist no more than 2^n elements Q of $\mathcal{Q}^n$ such that $U_Q = U$. So:

$$\sum_{Q\in\mathcal{Q}^n} e^{\phi_n(Q)} \le 2^n \sum_{U\in\gamma} e^{\phi_n(U)}, \tag{3.25}$$

for any finite subcover γ of α^n. Combining (3.24) and (3.25):

$$H_\nu(\mathcal{Q}^n) + \int \phi_n \, d\nu \le n\log 2 + \log P_n(f,\phi,\alpha).$$

Dividing by n and taking the limit when $n \to \infty$, we obtain (3.23). This completes the proof of the upper bound (3.20).

3.2.2 Aproximating the pressure

To finish the proof of Theorem 3.14, we show that for any $\varepsilon > 0$ there exists a f-invariant probability μ such that

$$h_\mu(f) + \int \phi \, d\mu \ge S(f,\phi,\varepsilon) \tag{3.26}$$

Clearly, this imply that the supremum of $h_\nu(f) + \int \phi \, d\nu$ when ν belongs to $\mathcal{M}(f)$ is greater or equal to $S(f,\phi) = P(f,\phi)$.

For each $n \ge 1$, let E be a (n,ε)-separated set such that

$$\sum_{y\in E} e^{\phi_n(y)} \ge \frac{1}{2} S_n(f,\phi,\varepsilon). \tag{3.27}$$

Denote by A the left-hand side expression of this inequality. Consider probability measures ν_n and μ_n defined on M by:

$$\nu_n = \frac{1}{A}\sum_{x\in E} e^{\phi_n(x)}\delta_x \quad \text{e} \quad \mu_n = \sum_{j=0}^{n-1} f_*^j \nu_n.$$

Since the set of all invariant probabilities is compact, and observing that by definition (3.11), we may choose a subsequence $(n_j)_j \to \infty$ such that

1. $\dfrac{1}{n_j}\log S_{n_j}(f,\phi,\varepsilon)$ converges to $S(f,\phi,\varepsilon)$ e

2. μ_{n_j} converges to some probability μ in the weak* topology.

We show that the probability μ is f invariant and satisfies (3.26). To make our job easier, we split the argument in four steps.

Step 1: We prove that μ is invariant. Let $\varphi : M \to \mathbb{R}$ be any real continuous function. For each $n \ge 1$,

$$\int \varphi\, d(f_*\mu_n) = \frac{1}{n}\sum_{j=1}^{n}\int \varphi\circ f^j\, d\nu_n = \int \varphi\, d\mu_n + \frac{1}{n}\Big(\int \varphi\circ f^n\, d\nu_n - \int \varphi\, d\nu_n\Big)$$

and, consequently,

$$\Big|\int \varphi\, d(f_*\mu_n) - \int \varphi\, d\mu_n\Big| \le \frac{2}{n}\sup|\varphi|.$$

Restricting to $n = n_j$ and taking the limit when $j\to\infty$, we see that $\int \varphi\, df_*\mu = \int \varphi\, d\mu$ for every continuous function $\varphi : M \to \mathbb{R}$. So, $f_*\mu = \mu$ as we claimed.

Step 2: We estimate the entropyy with respect to ν_n. Let $\mathcal{P}$ be any finite partition of M such that $\operatorname{diam}\mathcal{P} < \varepsilon$ and $\mu(\partial\mathcal{P}) = 0$. The first condition implies that each element of $\mathcal{P}^n$ contains at most one element of E. From the other hand, every element of E is contained is some element of $\mathcal{P}^n$, obviously. So,

$$\begin{aligned} H_{\nu_n}(\mathcal{P}^n) &= \sum_{x\in E} -\nu_n(\{x\})\log\nu_n(\{x\}) = \sum_{x\in E} -\frac{1}{A}e^{\phi_n(x)}\log\left(\frac{1}{A}e^{\phi_n(x)}\right) \\ &= \log A - \frac{1}{A}\sum_{x\in E} e^{\phi_n(x)}\phi_n(x) = \log A - \int \phi_n\, d\nu_n \end{aligned} \tag{3.28}$$

(the last inequality follows directly from the definition of ν_n).

Step 3: We compute the entropy with respect to μ_n. Consider $1 \le k < n$. For each $r \in \{0, \dots, k-1\}$, let $q_r \ge 0$ be the biggest integer such that $r + kq_r \le n$. In other words, $q_r = [(n-r)/k]$. So,

$$\mathcal{P}^n = \mathcal{P}^r \vee \big[\bigvee_{j=0}^{q_r-1} f^{-(kj+r)}(\mathcal{P}^k) \big] \vee f^{-(kq_r+r)}(\mathcal{P}^{n-(kq_r+r)})$$

(the first term does not exists if $r = 0$ and the last one does not exists if $n = kq_r + r$). So,

$$H_{\nu_n}(\mathcal{P}_n) \le \sum_{j=0}^{q_r-1} H_{\nu_n}(f^{-(kj+r)}(\mathcal{P}^k)) + H_{\nu_n}(\mathcal{P}^r) + H_{\nu_n}(f^{-(kq_r+r)}(\mathcal{P}^{n-(kq_r+r)})).$$

It is clear that $\#\mathcal{P}^r \le (\#\mathcal{P})^k$. Using the Lemma 2.5, it follows that $H_{\nu_n}(\mathcal{P}^r) \le k \log \#\mathcal{P}$. By the same reason, the last expression in the previous inequality is bounded by $k \log \#\mathcal{P}$. Thus,

$$H_{\nu_n}(\mathcal{P}_n) \le \sum_{j=0}^{q_r-1} H_{f_*^{(kj+r)}\nu_n}(\mathcal{P}^k) + 2k \log \#\mathcal{P} \tag{3.29}$$

for all $r \in \{0, \dots, k-1\}$. Now, it is clear that every number $i \in \{0, \dots, n-1\}$ may be uniquely written as $i = kj + r$ with $0 \le j \le q_r - 1$. Then, summing (3.29) over all values of r,

$$kH_{\nu_n}(\mathcal{P}_n) \le \sum_{i=0}^{n-1} H_{f_*^i\nu_n}(\mathcal{P}^k) + 2k^2 \log \#\mathcal{P}. \tag{3.30}$$

The concavity of the function $\phi(x) = -x \log x$ iimply that

$$\frac{1}{n} \sum_{i=0}^{n-1} H_{f_*^i\nu_n}(\mathcal{P}^k) \le H_{\mu_n}(\mathcal{P}^k).$$

Combining this inequality with (3.30), we have that

$$\frac{1}{n} H_{\nu_n}(\mathcal{P}_n) \le \frac{1}{k} H_{\mu_n}(\mathcal{P}^k) + \frac{2k}{n} \log \#\mathcal{P}.$$

On the other hand, by the definition of μ_n,

$$\frac{1}{n} \int \phi_n \, d\nu_n = \frac{1}{n} \sum_{j=0}^{n-1} \int \phi \circ f^j \, d\nu_n = \int \phi \, d\mu_n.$$

Summing with the previous equation, we have that

$$\frac{1}{n} H_{\nu_n}(\mathcal{P}_n) + \frac{1}{n} \int \phi_n \, d\nu_n \le \frac{1}{k} H_{\mu_n}(\mathcal{P}^k) + \int \phi \, d\mu_n + \frac{2k}{n} \log \#\mathcal{P}. \tag{3.31}$$

Step 4: Relate the previous estimates for the limit measure μ. Combining (3.28) and (3.31), we have that

$$\frac{1}{k}H_{\mu_n}(\mathcal{P}^k) + \int \phi\, d\mu_n \geq \frac{1}{n}\log A - \frac{2k}{n}\log \#\mathcal{P}.$$

By the choice of E on (3.27), it follows that

$$\frac{1}{k}H_{\mu_n}(\mathcal{P}^k) + \int \phi\, d\mu_n \geq \frac{1}{n}\log S_n(f,\phi,\varepsilon) - \frac{1}{n}\log 2 - \frac{2k}{n}\log \#\mathcal{P}. \tag{3.32}$$

The choice of the partition $\mathcal{P}$ implies that $\mu(\partial\mathcal{P}^k) = 0$ for all $k \geq 1$, since

$$\partial\mathcal{P}^k \subset \mathcal{P} \cup f^{-1}(\mathcal{P}) \cup \cdots \cup f^{-k+1}(\mathcal{P}).$$

Then, $\mu(P) = \lim_j \mu_{n_j}(P)$ for all $P \in \mathcal{P}^k$ and, therefore,

$$H_\mu(\mathcal{P}^k) = \lim_j H_{\mu_{n_j}}(\mathcal{P}^k).$$

Since ϕ is continuous, we also have $\int \phi\, d\mu = \lim_j \int \phi\, d\mu_{n_j}$. Then, restricting (3.32) to the subsequence $(n_j)_j$ and taking the limit when $j \to \infty$,

$$\frac{1}{k}H_\mu(\mathcal{P}^k) + \int \phi\, d\mu \geq S(f,\phi,\varepsilon).$$

Taking the limit when $k \to \infty$, we have that

$$h_\mu(f,\mathcal{P}) + \int \phi\, d\mu \geq S(f,\phi,\varepsilon).$$

Now, taking $\varepsilon \to 0$ (and, consequently, $\operatorname{diam}\mathcal{P} \to 0$), we obtain (3.26). This complete the proof of the Variational Principle (Theorem 3.14).

3.3 Equilibrium States

Let $f : M \to M$ be a continuous maps on a compact metric space. An invariant probability μ is an *equilibrium state* for a potential $\phi : M \to M$ if it is a maximum for the pressure, i.e., if

$$h_\mu(f) + \int \phi\, d\mu = P(\phi, f) = \sup\{h_\nu(f) + \int \phi\, d\nu : \nu \in \mathcal{M}(f)\}.$$

In the particular case of $\phi \equiv 0$, we also say that μ is a *maximal entropy measure* or just *maximal measure*. In this section, we study basic properties of the set $\mathcal{E}(f,\phi)$ of all Equilibrium States. Let us begin with some basic examples

Example 3.18. If $f : M \to M$ has zero topological entropy, every invariant probability μ is a maximal entropy measure, since $h_\mu(f) = 0 = h(f)$. For any potential $\phi : M \to \mathbb{R}$,

$$P(f,\phi) = \sup\{\int \phi\, d\nu : \nu \in \mathcal{M}(f)\}.$$

Thus, ν is a equilibrium state if, and only if, ν maximize the integral of ϕ. Since the function $\nu \mapsto \int \phi \, d\nu$ is continuous and $\mathcal{M}(f)$ is compact with respect to the weak* topology, there exists at least one maximum for any potential ϕ.

Example 3.19. Let $f_A : M \to M$ be the linear endomorphism induced on the $\mathbb{T}^d$ by some invertible matrix A with integer coeficients. Let μ be the Haar measure on $\mathbb{T}^d$. By Proposition 2.48 and 2.49,

$$h_\mu(f_A) = \sum_{i=1}^{d} \log^+ |\lambda_i| = h(f)$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of A. In particular, the Haar measure is the maximal entropy measure for f.

Example 3.20. Let $f : M \to M$ be the shift defined on $M = \{1, \ldots, d\}^{\mathbb{N}}$ (or on $M = \{1, \ldots, d\}^{\mathbb{Z}}$) and let μ be the Bernoulli measure given by a probability vector $p = (p_1, \ldots, p_d)$. As in Example 2.13,

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}^n) = \sum_{i=1}^{d} -p_i \log p_i.$$

We leave to the reader the duty of check that this function attains it maximum when all p_i are equal to $1/d$. Moreover, in this case $h_\mu(f) = \log d$. Remember that (Example 2.28) $h(f) = \log d$. Thus, the Bernoulli measure given by the probability vector $p = (1/d, \ldots, 1/d)$ é is the unique measure of maximal entropy among all Bernoulli measure. Indeed, we may show that this measure is the unique maximal entropy measure (among all invariant probabilities).

Proposition 3.21. *Assume that $h(f) < \infty$. Then, for any potential $\phi : M \to \mathbb{R}$, the set of Equilibrium States is a convex subset of $\mathcal{M}(f)$: more precisely, given $t \in (0,1)$ and given $\mu_1, \mu_2 \in \mathcal{M}(f)$,*

$$(1-t)\mu_1 + t\mu_2 \in \mathcal{E}(f, \phi) \quad \Leftrightarrow \quad \{\mu_1, \mu_2\} \subset \mathcal{E}(f, \phi).$$

Moreover, an invariant probability μ belongs to $\mathcal{E}(f, \phi)$ if, and only if, almost every ergodic component of μ belong to $\mathcal{E}(f, \phi)$.

Proof. This hypothesis that the topological entropy is finite give us that $P(f, \phi) < \infty$ for every potential ϕ. The functional $\Psi : \mu \mapsto h_\mu(f) + \int \phi \, d\mu$ is convex linear:

$$\Psi((1-t)\mu_1 + t\mu_2) = (1-t)\Psi(\mu_1) + t\Psi(\mu_2)$$

for every $t \in (0,1)$ and every $\mu_1, \mu_2 \in \mathcal{M}(f)$. Then $\Psi((1-t)\mu_1 + t\mu_2)$ is equal to the supremum ofΨ (i.e., it is equal to $P(f, \phi)$) if, and only if, $\Psi(\mu_1)$ and $\Psi(\mu_2)$ are equal to this supremum. This proves the first part of the proposition.

The proof of the second part is analogous. Let $(\mu_P)_P$ be the ergodic decomposition of μ. We have that

$$\Psi(\mu) = \int \Psi(\mu_P) \, d\hat{\mu}(P).$$

So, $\Psi(\mu) = \sup \Psi$ if, and only if, $\Psi(\mu_P) = \sup \Psi$ for $\hat{\mu}$-almost every P. □

Corollary 3.22. *If $\mathcal{E}(f,\phi)$ is non-empty then it contains ergodic invariant probabilities. Moreover, the extremal elements of the convex $\mathcal{E}(f,\phi)$ are precisely the ergodic measures on $\mathcal{E}(f,\phi)$.*

Proof. To prove the first claim, we just need to consider the ergodic components of any element of $\mathcal{E}(f,\phi)$. If $\mu \in \mathcal{E}(f,\phi)$ is ergodic then μ is a extremal element of $\mathcal{M}(f)$. Henceforth, μ is a extral element of $\mathcal{E}(f,\phi)$. Conversely, if $\mu \in \mathcal{E}(f,\phi)$ is not ergodic, we may write

$$\mu = (1-t)\mu_1 + t\mu_2, \quad \text{com } 0 < t < 1 \text{ e } \mu_1, \mu_2 \in \mathcal{M}(f).$$

By Proposition 3.21 we have that $\mu_1, \mu_2 \in \mathcal{E}(f,\phi)$ and, therefore, μ is not an extremal element of $\mathcal{E}(f,\phi)$. □

In general, Equilibrium States set *may* be empty, as can be seen in [Wal82], page 192.

But, there exists a big class of maps such that there exist equilibrium states for every potential:

The entropy function $\mu \to h_\mu(f)$ does not need to be continuous. For instance, if f is a complete (or Markov) shift with finite number d of symbols, given a Bernoulli measure it can be approximated by a sequence Dirac measures on periodic orbits. For instance, the measure associated to the vector $p = (1/d, \dots, 1/d)$ is an accumulation point of Dirac measures on periodic orbits. This imply that the entropy is *not* continuous: the entropy of μ is equal to $\log p$.

Lemma 3.23. *If the entropy function of f is upper-semicontinuous, then $\mathcal{E}(f,\phi)$ is compact (with respect to the weak* topology) and non-empty, for any potential ϕ.*

Proof. Let $(\mu_n)_n$ be a sequence in $\mathcal{M}(f)$ such that

$$h_{\mu_n}(f) + \int \phi \, d\mu_n \text{ converges to } P(f,\phi).$$

By compacness of $\mathcal{M}(f)$, the sequence admits some accumulation point μ. The semicontinuity of entropy, with the continuity of integral, imply that

$$h_\mu(f) + \int \phi \, d\mu \geq \liminf_n h_{\mu_n}(f) + \int \phi \, d\mu_n = P(f,\phi).$$

So, μ is an equilibrium state, as we claim. Analogously, taking any sequence $(\nu)_n$ in $\mathcal{E}(f,\phi)$ we have that any accumulation point ν is an equilibrium state. This show that $\mathcal{E}(f,\phi)$ is closed and, consequently, compact. □

Corollary 3.24. *Suppose that $f : M \to M$ is a expansive continuous map on a compact metric space M. Then, every potential $\phi : M \to \mathbb{R}$ admits some equilibrium state.*

The uniqueness is a more complex problem. It is not difficult to exhibt transformations with infinite equilibrium states. For instance, let $f : S^1 \to S^1$ a homeomorphism of the circle with infinitely many fixed points. The Dirac measures supported on these points are ergodic invariant probabilities. Since the topological $entropyh(f) = 0$, these measures are equilibrium states for any potential with maximum on these fixed points. (Example 3.18).

There is a special class of maps, called *expanding maps*, for which we have the uniqueness of the equilibrium state for any Hölder continuous potential. In particular, these maps are *intrinsically ergodic*, i.e., they admits a unique measure with maximal entropy.

Chapter 4

Lyapunov Exponents and Ruelle Inequality

Here, we always will assume that $f : M \to M$ is a (at least C^1) differentiable map in a compact Riemannian manifold M. The goal of this section is to introduce the notion of Lyapunov exponent of a point x in the direction $v \in T_xM$.

This number measures the (exponential) speed of the growth in norm of the vector under the action of the derivative $Df^n(x)$ and is defined by

$$\lambda_v(x) = \lim_{n\to\infty} \frac{1}{n} \log \|Df^n(x)v\|,$$

if it exists. This (exponential) growth rate does not need to be defined at every point, as the reader may guess. A major result in the theory is the Oseledets Theorem, that asserts this growth rate is well-defined at almost every point, in a specific sense.

To begin with, we say that a set A has *full probability*, if given an f-invariant probability μ, then $\mu(A) = 1$. From the the Ergodic Decomposition Theorem, we have that

Proposition 4.1. *A has full probability if, and only if, $\mu(A) = 1$ for every ergodic measure.*

To illustrate and motivate the definition of Lyapunov exponent, let us consider the (simpler) case of one dimensional C^1 map $f : S^1 \to S^1$.

Example 4.2. In order to understand the assymptotics of $|Df^n(x)|$, we want to use the Birkhoff's Ergodic Theorem, averaging the function $\varphi(x) = \log |Df(x)|$. Fix an invariant measure μ and assume that $\varphi(x) = -\log |Df(x)|$ is μ-integrable. In this case, Birkhoff's Ergodic Theorem give us that

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

is well-defined at μ-almost every point and

$$\int \lambda \, d\mu = \int \log |Df(x)| \, d\mu.$$

On the other hand, by the chain rule we have that

$$\lambda(x) = \lim_{n\to\infty} \frac{1}{n} \log |Df^n(x)|.$$

Thus, the average growth rate of the derivative is well defined at *almost* every point.

Observe that the integrability of φ is a crucial assumption in order to apply the Birkhoff's Theorem in the argument above. We may extend a little bit further this argument for multidimensional maps, making use of the Kingman's Subadditive Ergodic Theorem 1.21:

Theorem 4.3. *[Furstenberg-Kesten Theorem for Cocycles in $GL(2,\mathbb{R})$] Assume that $f : M \to M$ preserves a probability μ. Let $A : M \to GL(2,\mathbb{R})$ be a measurable map such that $\int \log \|A(x)\| \, d\mu$ is finite. Define*

$$A^n(x) = A(f^{n-1}(x))A(f^{n-1}(x)) \dots A(x).$$

Then, for μ-almost every point, the function

$$\lambda(x) = \lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\|$$

is well-defined, measurable and f-invariant. Moreover, λ is integrable and its integral is given by

$$\int \lambda \, d\mu = \lim_{n\to\infty} \frac{1}{n} \int \log \|A^n\| \, d\mu.$$

The proof of Theorem 4.3 above is a direct consequence of Kingman's Subadditive Ergodic Theorem. In fact,

Proof. By Example 1.20, the sequence $\varphi_n(x) = \log \|A^n(x)\|$ is subadditive. Therefore, since $\log \|A(x)\|$ is μ-integrable, we may apply the Kingman's Subadditive Ergodic Theorem to obtain that

$$\lambda(x) = \lim_{n\to\infty} \frac{\varphi_n(x)}{n}$$

is well-defined, measurable and f-invariant. Moreover, λ is integrable and its integral is given by

$$\int \lambda \, d\mu = \lim_{n\to\infty} \frac{1}{n} \int \log \|A^n\| \, d\mu.$$

□

A interesting application of the Furstenberg-Karsten Theorem occurs when we consider a differentiable map $f : M \to M$ and take $A(x) = Df(x)$ (with a slightly abuse of notation, writing $Df(x)$ in convenient coordinates). In this case, the theorem above give us the existence of the upper Lyapunov exponent at x with respect to f.

Now, we discuss a general version of the result above, due to to Oseledets.

4.1 Oseledets Theorem: invertible version

Definition 4.4. Let M be a Riemannian manifold and $f : M \to M$ a C^1 diffeomorphism. We say that x has a *Lyapunov structure*, if there are real numbers $\lambda_1(x) > ... > \lambda_l(x)$ and a decomposition in subspaces of the tangent space

$$T_xM = E_1(x) \oplus \cdots \oplus E_l(x)$$

such that

1. $Df(x)E_i(x) \subset E_i(f(x))$.
2. $\lim_{n\to\pm\infty} \frac{1}{n} \log \| Df^n(x)u \| = \lambda_j(x)$, for any $u \in E_j \setminus \{0\}$ and $1 \le j \le l$.
3. $\lim\limits_{n\to\infty} \frac{1}{n} \log \sin \angle\big(\bigoplus\limits_{i\in I} Df^n(x)E_i(x), \bigoplus\limits_{j\in J} Df^n(x)E_j(x)\big) = 0$, whenever $I \cap J = \emptyset$.

In this case, we say that x is a regular point of f.

Observe that $\{\lambda_1, ..., \lambda_l\}$ coincide with

$$\{\lambda \in \mathbb{R} : \lambda = \lim_{n\to\pm\infty} \frac{1}{n} \log \| Df^n(x)u \| = \lambda \text{ for some } u \in T_xM \setminus \{0\}\}$$

and we also have that

$$E_j = \{u \in T_xM : \frac{1}{n} \log \| Df^n(x)u \| \to \lambda_j, \text{ when } n \to \pm\infty\}$$

.

Thus, when they exists, the real numbers $\lambda_1(x) > ... > \lambda_l(x)$ and subspaces $E_1(x), ..., E_l(x)$ are uniquely determined. The number$\lambda_j(x)$ and the subspace $E_j(x)$ are called, respectively, Lyapunov exponents and Lyapunov spaces of f at x.

Denoting by $R(f)$ the set of all regular points of f, it is not difficult to see that $f(R(f)) = R(f)$, i.e., the set $R(f)$ is invariant by f with, $\lambda j(f(x)) = \lambda j(x)$ and $Df(x)E_j(x) = E_jf(x)$ for all $1 \le j \le l$ e $x \in R(f)$.

Example 4.5. If $p \in M$ is a fixed point with eigenvalues $\sigma_1(p) > ... > \sigma_l(p)$ and respective invariant subspaces $E_1(p), ..., E_l(p)$, one can easily check that p

is regular, taking $\lambda_i = \log \|\sigma_i\|$ as its Lyapunov exponents. In this case, we also have that

$$\sum_{i=1}^{l(p)} \lambda_i(p) \dim E_i(p) = \log|\det Df(p)|.$$

Then, p is regular point for f. A completely similar statement can be made for any periodic point of f.

Remark 4.6. One can show that if $x \in M$ is regular, then

$$\sum_{i=1}^{l(x)} \lambda_i(x) = \lim_{n\to\infty} \frac{1}{n} \log|\det Df^n(x)|.$$

From the topological point of view, the subset $R(f)$ could be *small* (maybe Baire first category or even a point). However, from the pointview of Ergodic Theory, we have that $R(f)$ is always a relevant set. This is the content of

Theorem 4.7 (Oseledets Theorem for diffeomorphisms)**.** *Let M be a Rimannian manifold with finite dimension. Then, the set of regular points for a C^1 diffeomorphism, $f : M \to M$, has full probability. Moreover, we have that if $\lambda_1(x) > ... > \lambda_l(x)$ are the Lyapunov exponents with invariant subspaces $E_1(x), ..., E_l(x)$, then*

$$\int \sum_{i=1}^{l(x)} \lambda_i(x)\, dim E_i(x)\, d\mu = \int \log|\det Df(x)|\, d\mu.$$

Proof. For a proof of this theorem, we recommend [Via]. □

Proof. See [Mañ87] □

Example 4.8. (Linear Automorphisms on $\mathbb{T}^2$) As we discussed in 1.3.1, if we equip the 2-dimensional torus $\mathbb{T}^2$ with the plane Riemannian metric and consider a 2×2 hyperbolic matrix A with integer coefficients and $\det A = 1$, the map f_A preserves the Lebesgue measure m associated to this Riemannian metric and it is an ergodic measure with respect to f_A.

By Oseledets theorem, there exists Lyapunov structure at (almost) every point $x \in \mathbb{T}^2$ Actually, in this example, every point admits the same Lyapunov exponent λ^+ and λ^-, and they coincide with the the logaritm of the eigeinvalues of A.

Observe also that one Lyapunov exponent is positive and the other Lyapunov expoents is negative, since $\lambda^+ + \lambda^- = \log|\det A| = 0$.

This is an (very first) example of *Anosov map*, part of a very interesting (and developed) theory of hyperbolic dynamics. In this situation, the splitting $E^+(x)$ and $E^-(x)$ are, in fact, defined everywhere and depends continuously with the point. In particular, the angle between E^+ and E^- is bounded away from zero.

4.2 Ruelle's Inequality

Now, we describe a interesting and useful result by Ruelle that give an upper bound on the metric entropy of μ. One could ask if the Lyapunov exponents of a measure μ give some information about the entropy of this measure. From one hand, the entropy can be regarded as the rate of growth of the number of (special dynamically defined) balls. From the other hand, Lyapunov exponents gives the *way* that balls deform under the action of f. For instance, one can point out the example that we discussed before:

Example 4.9. We proved in Proposition 2.48 that if $f_A : \mathbb{T}^d \to \mathbb{T}^d$ is the induced endomorphism on $\mathbb{T}^d$ by some hyperbolic invertible matrix A with integer coefficients, and μ is the Haar (Lebesgue) measure on $\mathbb{T}^d$. Then

$$h_\mu(f_A) = \sum_{i=1}^{d} \log^+ |\sigma_i|.$$

where $\sigma_1, \dots, \sigma_d$ are the eigenvalues of A, with multiplicity.

Observe that for any eigenvalue of A, the numbers $\log|\sigma_i|$ are the Lyapunov exponents of f_A with respect to μ. One could guess that this equality always holds true. However, this is not the case. We easily can make up an example:

Example 4.10. Let f be any C^1 map with some hyperbolic fixed point of the saddle type. Define the *Dirac* measure at p as the measure defined at $A \subset M$ as $\delta_p(A) = 1$, if $p \in A$ and $\delta_p(A) = 0$, otherwise. Since $\delta_p(A) = 0$ or 1 for every $A \subset M$, we have that

$$H_{\delta_p}(\mathcal{P}) = 0$$

for every partition $\mathcal{P}$. In particular,

$$h_{\delta_p}(f) = 0.$$

In particular, if $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$ are the lyapunov exponents of f at p, counted with multiplicity, then

$$h_\mu(f_A) < \sum_{\lambda^i > 0} \lambda_i.$$

A completely similar argument give us that $h_\mu(f) = 0$, if

$$\mu = \frac{1}{n}(\delta_p + \cdots + \delta_{f^{n-1}(p)},$$

where $f^n(p) = p$.

Now, let us state the main theorem of this section:

Theorem 4.11. *(Ruelle Inequality) Let $f : M \to M$ be a C^1 map and M a compact Riemannian manifold. For any invariant probability μ we have that*

$$h_\mu f \leq \int \sum_{i:\lambda_i(x)>0} \lambda_i(x) d\mu(x), \tag{4.1}$$

where $\lambda_i(x)$, are the Lyapunov exponents of x counted with multiplicity.

Proof. See [Mañ87] □

One particular case occurs when μ is an ergodic measure. Since λ_i is an invariant function, we have that they arre constant at almost every point and we may rewrite the Equation 4.1 as:

$$h_\mu f \leq \sum_{i:\lambda_i>0} \lambda_i.$$

As a direct consequence of the Ruelle Inequality, we get that

Proposition 4.12. *If $f : M \to M$ is a C^1 map in a compact manifold M and μ is an ergodic measure with positive entropy then has some positive Lyapunov exponent.*

Definition 4.13. We say that a (ergodic) measure is *hyperbolic*, if all Lyapunov exponents of μ are non-zero.

Another consequence of 4.1 is

Proposition 4.14. *If $f : M^2 \to M^2$ is a C^1 diffeomorphism in a compact surface M^2 with positive topological entropy, then f preserves some hyperbolic measure.*

Proof. Observe that if $h_{top}(f) > 0$ then, by the Variational Principle there must exists some measure μ with positive mettric entropy. Thus, μ has a positive Lyapunov exponent λ^+. On the other hand, since

$$h_\mu(f) = h_\mu(f^{-1}) > 0,$$

we have that f^{-1} has also an positive exponent $-\lambda^- > 0$. Since the Lyapunov exponents of μ with respect to f are the additive inverse of the exponents of μ with respect to f^{-1}, then we have that $\lambda^- < 0$ is also a Lyapunov exponent of f, and it finishes the proof. □

It is worthwile to remark that when $f \in C^{1+\alpha}$ in the setting above preserves an hyperbolic measure μ, then there exists an sequence of hyperbolic sets Λ_n, invariant by f with $\mu(\Lambda_n) \to 1$, that *approximate* well μ. This was proved by Katok (see in [Kat80]).

Chapter 5

Useful Facts

5.1 Perron-Frobenius Theorem

Theorem 5.1 (Perron-Frobenius)**.** *Let A be a $d \times d$ matrix with nonnegative entries. Then there exists $\lambda \geq 0$ and some vector $v \neq 0$ with nonnegative entries such that $Av = \lambda v$ and $\lambda \geq |\gamma|$ for any eigenvalue γ of A.*

If A admits some power whose entries are all positive then $\lambda > 0$ and there exists some eigenvector v with positive entries. In fact, $\lambda > |\gamma|$ for any other eigenvalue γ of A. Moreover, the eigenvalue λ has multiplicity 1 and is the unique eigenvalue of A that admits some eigenvector with nonnegative entries.

A proof of Perron-Frobenius theorem can be found in the book of Meyers [Mey00], among others. Applying this theorem to the matrix $A = P^*$, we conclude that there exists $\lambda \geq 0$ and $p \neq 0$ with $p_i \geq 0$ for any i, such that

$$\sum_{i=1}^{d} p_i P_{i,j} = \lambda p_j, \quad \text{for any } 1 \leq j \leq d.$$

Summing over $i = 1, \ldots, d$ we obtain that

$$\sum_{j=1}^{d} \sum_{i=1}^{d} p_i P_{i,j} = \lambda \sum_{j=1}^{d} p_j.$$

Using the property (ii) of invariant matrix, the left hand side of this equality can be written as

$$\sum_{i=1}^{d} p_i \sum_{j=1}^{d} P_{i,j} = \sum_{i=1}^{d} p_i.$$

Comparing these two last equalities, and remember that the sum of the entries of p is a positive number, we conclude that $\lambda = 1$.

5.2 Jensen Inequality

We say that a function $\phi : I \to \mathbb{R}$ defined in an interval I of the real line is *convex* if for any $x, y \in I$ e $t \in [0,1]$ the following holds:

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y).$$

Moreover, we say that ϕ is *concave* if $-\phi$ is convex. For twice differentiable functions we have the following practical criteria: ϕ is convex if $\phi''(x) > 0$ for any $x \in I$ and is concave if $\phi''(x) < 0$ for any $x \in I$.

Theorem 5.2 (Jensen's inequality)**.** *Let $\phi : I \to \mathbb{R}$ be a convex function. If μ is a probability measure in X and $f \in L^1(\mu)$ such that $\int f\,d\mu \in I$, then:*

$$\phi\left(\int f\,d\mu\right) \leq \int \phi \circ f\,d\mu.$$

Example 5.3. For any probability measure μ and any positive integrable function f we have that

$$\log \int f\,d\mu \geq \int \log f\,d\mu.$$

In fact, this corresponds to Jensen's inequality for the function $\phi : (0,\infty) \to \mathbb{R}$ given by $\phi(x) = -\log x$. Observe that ϕ is convex: $\phi''(x) = 1/x^2 > 0$ for any x.

Example 5.4. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function, let $(\lambda_i)_i$ be a sequence of nonnegative real numbers satisfying $\sum_{i=1}^{\infty} \lambda_i \leq 1$ and let $(a_i)_i$ be a bounded sequence of real numbers. Then,

$$\phi\left(\sum_{i=1}^{\infty} \lambda_i a_i\right) \leq \sum_{i=1}^{\infty} \lambda_i \phi(a_i). \tag{5.1}$$

This can be seen in the following way. Consider $X = [0,1]$ equipped with the Lebesgue measure μ. Let $f : [0,1] \to \mathbb{R}$ be a function of the form $f(x) = \sum_{i=1}^{\infty} a_i \mathcal{X}_{E_i}$ where E_i are pairwise disjoint measurable sets such that $\mu(E_i) = \lambda_i$. Applying Jensen's inequality to the function f gives precisely the relation (5.1).

5.3 Approximation and Extension of Measures

A very useful result in the construction of measures is

Theorem 5.5 (Extension)**.** *Let $\mathcal{B}_0$ be an algebra of subsets of X and let $\mu_0 : \mathcal{B}_0 \to [0,+\infty]$ be a σ-additive function with $\mu(\emptyset) = 0$ and $\mu(X) < \infty$. Then there exists a unique measure defined in the σ-algebra $\mathcal{B}$ generated by $\mathcal{B}_0$ which is a extension of μ_0, i.e., for all element $B \in \mathcal{B}_0$ we have $\mu_0(B) = \mu(B)$.*

The theorem 5.5 is applicable with the same conclusion when the measure μ_0 in question is only σ-finite. In addition, if μ_0 is only finite additive there is still a measure that extends μ_0 to $\mathcal{B}$. However, in this case we can't guarantee that such extension is unique.

In general, in order to show that a function defined in a σ-algebra is a measure, the most difficult part is to check the σ-additivity. The criterion most commonly used for this purpose is expressed in the following result:

Theorem 5.6 (Continuity at the empty set)**.** *Let $\mathcal{B}$ be a algebra of subsets of set X and let $\mu : \mathcal{B} \to [0, +\infty)$ be a finite additive function with $\mu(X) < \infty$. Then μ is σ-additive if and only if,*

$$\lim_{n\to\infty} \mu(\bigcap_{j=1}^{n} A_j) = 0 \tag{5.2}$$

for any measurable sets $A_1 \supset \cdots \supset A_j \supset \cdots$ with $\cap_{j=1}^{\infty} A_j = \emptyset$.

Using similar arguments, it is possible to obtain other characterizations for the σ-additivity of a finitely additive function:

Theorem 5.7 (Continuity from above and from below)**.** *Let $\mathcal{B}$ be a algebra of subsets of a set X and let $\mu : \mathcal{B} \to [0, +\infty)$ be a finitely additive function with $\mu(X) < \infty$. The following conditions are equivalent:*

1. *μ is σ-additive;*
2. *for all sequence $A_1 \supset \cdots \supset A_j \supset \cdots$ of measurable sets we have*
$$\lim_{n\to\infty} \mu(A_n) = \mu(\cap_{j=1}^{\infty} A_j); \tag{5.3}$$
3. *for all sequence $A_1 \subset \cdots \subset A_j \subset \cdots$ of measurable sets we have*
$$\lim_{n\to\infty} \mu(A_n) = \mu(\cup_{j=1}^{\infty} A_j). \tag{5.4}$$

Another related result, which will be useful for our study, is the monotone class theorem, which we state below.

Definition 5.8. We say that a collection $\mathcal{C}$ of non-empty measurable subsets which contain X is a *monotone class*, if $\mathcal{C}$ which is closed under countable monotone unions and intersections,i.e., if

- for all subsets $A_1 \subset A_2 \subset \cdots \in \mathcal{C}$, $\cup_{n\geq 1} A_n \in \mathcal{C}$ and
- for all subsets $A_1 \supset A_2 \supset \cdots \in \mathcal{C}$, we have $\cap_{n\geq 1} A_n \in \mathcal{C}$.

Clearly, the families $\{\emptyset, X\}$ and 2^X are monotone classes. Moreover, if $\{\mathcal{C}_i : i \in \mathcal{I}\}$ is any family of monotone class, then $\cap_{i\in\mathcal{I}} \mathcal{C}_i$ is a monotone class. Therefore, given a subset $\mathcal{A}$ of 2^X, we can always consider the smallest monotone class which contains $\mathcal{A}$.

Theorem 5.9 (Monotone classes)**.** *The smallest monotone class that which contains an algebra $\mathcal{A}$ coincides with the σ-algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$.*

Another important fact about σ-algebras, which will be useful later, states that all the element B of σ-algebra generated by an algebra can be approximated by some element B_0 of the algebra, in the sense that the measure of symmetric difference

$$B \Delta B_0 = (B \setminus B_0) \cup (B_0 \setminus B)$$

can be as small as one wants.

Theorem 5.10 (Approximation)**.** *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $\mathcal{B}_0$ be an algebra which generate the σ-algebra $\mathcal{B}$. Then for any $\varepsilon > 0$ and any $B \in \mathcal{B}$ there exits a $B_0 \in \mathcal{B}_0$ such that $\mu(B \Delta B_0) < \varepsilon$.*

Definition 5.11. A measure space is said to be *complete* if any subset of a measurable set with measure zero is also measurable.

It is possible to transform any measure space $(M, \mathcal{B}, \mu)$ into a complete space, as follows. The family $\bar{\mathcal{B}}$ of all the subsets $A \subset M$ such that $\mu(A \Delta B) = 0$ for some $B \in \mathcal{B}$ is a σ-algebra which contains $\mathcal{B}$. Consider $\bar{\mu} : \bar{\mathcal{B}} \to [0, +\infty]$ given by $\bar{\mu}(A) = \mu(B)$ if $\mu(A \Delta B) = 0$ for some $B \in \mathcal{B}$. This function is well defined and is actually a measure on $\bar{\mathcal{B}}$, whose restriction in $\mathcal{B}$ coincides with μ. By construction, $(M, \bar{\mathcal{B}}, \bar{\mu})$ is a complete measure space. We often say that a function defined on M measurable if it is measurable for the σ-algebra completion $\bar{\mathcal{B}}$.

5.4 $L^p(\mu)$ with $1 \leq p < \infty$

In this section we define spaces consisting of functions which have special propertities of integrability. These spaces are normed vector spaces which are complete with respect to this norm, i.e., Banach spaces. We also state some inequalities involving the norms of these spaces. Throughout this section $(X, \mathcal{B}, \mu)$ is always a measure space.

Given any $p \in [1, \infty)$, we say that $f : X \to \mathbb{R}$ is a *p-the power integrable function* if the function $|f|^p$ is integrable.

Definition 5.12. Denote by $L^p(\mu)$ the set of the functions which are equal at μ-almost every point to a p-th power integrable function, modulus the equivalence relation which identifies any two function that agree at μ-almost every point.

Note that if μ is finite, which occur frequently in our examples, all the bounded measurable functions are in $L^p(\mu)$, since

$$\int |f|^p \, d\mu \leq (\sup |f|)^p \, m(X).$$

In particular, if X is compact then all the continuous functions are in $L^p(\mu)$. in other words,

$$C^0(X) \subset L^p(\mu) \quad \text{for any } p.$$

For each function $f \in L^p(\mu)$, define

$$\|f\|_p = \left(\int |f|^p \, d\mu \right)^{\frac{1}{p}}.$$

The next theorem shows that $L^p(\nu)$ is equipped with the structure of a Banach space:

Theorem 5.13. *The set $L^p(\mu)$ is a real vector space. Moreover, $\|f\|_p$ is a norm in $L^p(\mu)$ and this norm is complete.*

The most interesting part of the proof of this theorem is to prove the triangle inequality which, in this case is known as *Minkowski inequality*:

Theorem 5.14 (Minkowski inequality)**.** *Let $f, g \in L^p(\mu)$. Then:*

$$\left(\int |f+g|^p \, d\mu \right)^{\frac{1}{p}} \leq \left(\int |f|^p \, d\mu \right)^{\frac{1}{p}} + \left(\int |g|^p \, d\mu \right)^{\frac{1}{p}}.$$

5.5 Holder Inequality

The case $p = 2$ deserves our particular attention. The reason is that in this case the norm $\|\cdot\|_2$ defined above comes from a inner product, namely,

$$\langle f, g \rangle = \int fg \, d\mu. \tag{5.5}$$

If follows from the properties of the integral that this expression actually define an inner product in $L^2(\mu)$. This product is related with the norm $\|\cdot\|_2$ by:

$$\|f\|_2 = \sqrt{\langle f, f \rangle}.$$

In particular, the following Cauchy-Schwartz inequality holds: :

Theorem 5.15 (Cauchy-Schwartz inequality)**.** *Given $f, g \in L^2(\mu)$, then $fg \in L^1(\mu)$ and the inequality holds:*

$$\left| \int fg \, d\mu \right| \leq \sqrt{\int |f|^2 \, d\mu} \sqrt{\int |g|^2 \, d\mu}.$$

This inequality has the following interesting consequence. Suppose that the measure μ is finite and given any $f \in L^2(\mu)$. Then, take $g \equiv 1$,

$$\int |f| \, d\mu = \int |fg| \, d\mu \leq \sqrt{\int |f|^2 \, d\mu} \sqrt{\int 1 \, d\mu} < \infty.$$

This shows that al the functions in $L^2(\mu)$ are in $L^1(\mu)$.

For any values of p, the following generalization of the Cauchy-Schwartz inequality holds:

Theorem 5.16 (Hölder inequality)**.** *Given* $1 < p < \infty$ *consider* q *defined by relation* $\frac{1}{p} + \frac{1}{q} = 1$*. Then, for any* $f \in L^p(\mu)$ *and* $g \in L^q(\mu)$ *we have that* $fg \in L^1(\mu)$ *and the inequality holds:*

$$\int |fg| \, d\mu \leq \left(\int |f|^p \, d\mu \right)^{\frac{1}{p}} \left(\int |g|^q \, d\mu \right)^{\frac{1}{q}} .$$

Bibliography

[AB] A. Avila and J. Bochi. Proof of the subadditive ergodic theorem. Preprint www.mat.puc-rio.br/∼jairo/.

[Bow75] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lect. Notes in Math.* Springer Verlag, 1975.

[Bow78] R. Bowen. Entropy and the fundamental group. In *The Structure of Attractors in Dynamical Systems*, volume 668 of *Lecture Notes in Math.*, pages 21–29. Springer- Verlag, 1978.

[Din70] E. Dinaburg. A correlation between topological entropy and metric entropy. *Dokl. Akad. Nauk SSSR*, 190:19–22, 1970.

[Din71] E. Dinaburg. A connection between various entropy characterizations of dynamical systems. *Izv. Akad. Nauk SSSR Ser. Mat.*, 35:324–366, 1971.

[Dug66] J. Dugundji. *Topology.* Allyn and Bacon Inc., 1966.

[Goo71a] T. Goodman. Relating topological entropy and measure entropy. *Bull. London Math. Soc.*, 3:176–180, 1971.

[Goo71b] G. Goodwin. Optimal input signals for nonlinear-system identification. *Proc. Inst. Elec. Engrs.*, 118:922–926, 1971.

[Kat80] A. Katok. Lyapunov exponents, entropy and periodic points of diffeomorphisms. *Publ. Math. IHES*, 51:137–173, 1980.

[Mañ87] R. Mañé. *Ergodic theory and differentiable dynamics.* Springer Verlag, 1987.

[Mat] C. Matheus. mathblog. http://matheuscmss.wordpress.com/2009/02/19/furstenbergs-2x-3x-mod-1-problem/.

[Mey00] C. Meyer. *Matrix analysis and applied linear algebra.* Society for Industrial and Applied Mathematics (SIAM), 2000.

[Mis76] M. Misiurewicz. A short proof of the variational principle for a Z_+^N action on a compact space. *Asterisque*, 40:147–187, 1976.

[MP77a] M. Misiurewicz and F. Przytycki. Entropy conjecture for tori. *Bull. Pol. Acad. Sci. Math.*, 25:575–578, 1977.

[MP77b] M. Misiurewicz and F. Przytycki. Topological entropy and degree of smooth mappings. *Bull. Pol. Acad. Sci. Math.*, 25:573–574, 1977.

[MP08] W. Marzantowicz and F. Przytycki. Estimates of the topological entropy from below for continuous self-maps on some compact manifolds. *Discrete Contin. Dyn. Syst. Ser.*, 21:501–512, 2008.

[Rue]

[Rue73] D. Ruelle. Statistical Mechanics on a compact set with Z^ν action satisfying expansiveness and specification. *Trans. Amer. Math. Soc.*, 186:237–251, 1973.

[Shu74] M. Shub. Dynamical systems, filtrations and entropy. *Bull. Amer. Math. Soc.*, 80:27–41, 1974.

[SW75] M. Shub and R. Williams. Entropy and stability. *Topology*, 14:329–338, 1975.

[SX10] R. Saghin and Z. Xia. The entropy conjecture for partially hyperbolic diffeomorphisms with 1-D center. *Topology Appl.*, 157:29–34, 2010.

[Via] M. Viana. Lectures on lyapunov exponents. Preprint www.impa.br/∼viana/out/lle.pdf.

[Wal75] P. Walters. A variational principle for the pressure of continuous transformations. *Amer. J. Math.*, 97:937–971, 1975.

[Wal82] P. Walters. *An introduction to ergodic theory.* Springer Verlag, 1982.

[Yom87] Y. Yomdin. Volume growth and entropy. *Israel J. Math.*, 57:285–300, 1987.