PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Thermodynamic Formalism and Dimension Theory

Autor:
Felipe PÉREZ

Supervisor:
Dr. Godofredo IOMMI

Tesis presentada a la Facultad de Matemáticas de la Pontificia Universidad Católica de Chile para optar al grado de Magíster en Matemática

Comisión informante:
Carlos Vásquez (Pontificia Universidad Católica de Valparaíso)
Jairo Bochi (Pontificia Universidad Católica de Chile)

Junio 2016
Santiago, Chile

Contents

Introduction

One of the basic aims of ergodic theory is the study of asymptotic behavior of the orbits of a dynamical system. Even deterministic systems may exhibit a chaotic behavior, so a statistical point of view is the most plausible approach. For this, we may use the measure and probability theory, and invariant measures arise as the natural objects to address this problem. The existence of invariant measures on a system provides non-trivial results about the recurrence of almost every point, and if the measure satisfy the ergodic hypothesis, we even have a quantitative result, namely, the Birkhoff's ergodic theorem. Under very general conditions, the systems have plenty of invariant measures, so the problem becomes what invariant measure is the most natural to observe the system with. Ya. Sinai and D. Ruelle gave an answer to this question, and realized that many ideas of the thermodynamic of equilibrium states theory and statistical mechanics can be translated into an abstract setting and obtained analogue results to the ones obtained by Gibbs, among other physicists. The notions of entropy and pressure of a system, as well as variational principles were formulated for abstract systems and successfully applied in other mathematical theories, most notably in the theory of Anosov diffeomorphisms. One of the most surprising aspects of thermodynamic formalism is its connection with dimension theory, masterfully exemplified by Bowen's formula [Bow79]. This formula allowed the computation of Hausdorff dimensions (denoted by $\dim_H$) of a huge amount of sets and study its regularity and stability under perturbations of the sets. In this work, we studied the relation between thermodynamic formalism and dimension theory with emphasis in the transfer operator technique. This relation is then applied in a number theoretic problem in order to obtain measures of the complexity of certain sets of relevance for this theory.

The Borel-Bernstein Theorem (see chapter 4) implies that the sets

$$E(B) = \{x :\in [0,1] : a_n(x) \geq B^n \text{ for infinitely many } n \in \mathbb{N}\}$$

where $a_n(x)$ is the $n-$th digit of the continuous fraction expansion, have zero Lebesgue measure for every $B > 1$. In 1941 Good [Goo41] proved that the set

$$S(B) = \{x \in [0,1] : a_n(x) \geq B^n \text{ for every } n \in \mathbb{N}\}$$

has constant Hausdorff dimension equal to $1/2$. Recently, Wang and Wu [WW08] proved the the Hausdorff dimension of $E(B)$ depends continuously on the parameter $B \in (1, \infty)$. The strategy used by the authors shows that the number $\dim_H E(B)$ can be computed as the solution of the limit of a uniformly convergent sequence of continuous functions, so the regularity of the function $B \mapsto \dim_H E(B)$ cannot be improved by this method. Noting that $d_B = \dim_H E(B)$ satisfies a *Bowen like* equation

$$P(G, -d_B \log |BG'|) = 0$$

where $P(G, \cdot)$ is the topological pressure with the respect to the Gauss map G, it is possible to use the methods of thermodynamic formalism to conclude that $E(B)$ actually depends *real analytically* in B. In fact, we prove

Theorem. *The function $B \in (1, \infty) \mapsto \dim_H E(B) \in \mathbb{R}_+$ is real analytic, strictly decreasing and satisfies $\lim_{B \to 1} d_B = 1$ and $\lim_{B \to \infty} d_B = 1/2$.*

This monograph is organized in four chapters. The only prerequisite to read this work is basic knowledge on measure theory and functional analysis. The first chapter covers the basic ergodic theory, setting most of the measure theory language that will be used through this work. Some essential examples are presented, as well as some crucial theorems of the classic theory. Also, the basics of dimension theory are presented in this chapter, including the fundamental example of the geomertic construction via Moran covers. Chapter two is devoted to introduce basic notions of thermodynamic formalism, including entropy (metric and topological) and pressure. Some important results are stated without proof, such as variational principles. We include a reduced case proof of the Bowen's formula to stress the resemblance with the geometric construction done in the first chapter. The third chapter is the fundamental technical core of this work, in which we make an extensive use of the transfer and Ruelle's Operators to deduce dynamical properties of a system from the spectral properties of such operators. We also dedicate a couple of pages to study an explicit form for the transfer operator when the system is regular enough. Ruelle's Theorem is the most important result of this chapter, sinces it characterizes the spectrum of Ruelle's operator and relates it with the topological pressure of the geometric potential associated to the system. We prove many dynamical consequences of this fact. Most of the work is done in the finite state setting, and the infinite state system (Gauss map) results will be cited from Mayer's works [May76], [May90]. Finally, in the fourth chapter we study the work of Wang and Wu [WW08] and prove the Bowen equation for $E(B)$, allowing us to establish the regularity of the function coding the dimension of the Borel-Bernstein sets. Again, we include part of the proof done by Wang and Wu to highlight certain similarities with the geometric constructions done before.

Chapter 1

Background Theory

This chapter is based on [Wal82], [VO15], [DK02], [EW13] (first two sections), [Fal97], [Bar08] and [Pes08] (third section).

1.1 Invariant Measures

Dynamical systems are our main objects of study, and we will use measure theory as the main tool to analyze them. The basic object of (finite) measure theory is the probability space, and the structure-preserving morphisms are the measure-preserving functions:

Definition 1.1. Let $(X_i, \mathcal{B}_i, \mu_i)$ be probability spaces, for $i = 1, 2$. A measurable function $T : X_1 \to X_2$ is *measure-preserving* if $\mu_1(T^{-1}(B)) = \mu_2(B)$ for every $B \in \mathcal{B}_2$. If T is invertible and T^{-1} is a measure-preserving function, we say that T is an *isomorphism* of measure spaces, and the underlying spaces are said *isomorphic*.

Remark. Sometimes we will require that two measure spaces are isomorphic *modulo zero*, that is, there exists two subsets $E_i \subset X_i$ having zero measure and such that $X_1 \setminus E_1$ is isomorphic to $X_2 \setminus E_2$.

Remark. Under the notation of the previous definition, we say that T is μ-invariant or that μ is $T-$invariant. In both cases, we will simply say invariant if the measure or the function are obvious.

We are mainly interested in the case $(X_1, \mathcal{B}_1, \mu_1) = (X_2, \mathcal{B}_2, \mu_2)$.

Remark. The identity of a measure space is a measure-preserving function, and the composition of two measure-preserving functions is again a measure-preserving function. Hence, measure spaces and measure-preserving functions form a category. Two isomorphic measure spaces share the same measure theoretic properties.

Definition 1.2. A (discrete) *measurable dynamical system* is a measure space $(X, \mathcal{B}, \mu)$, called the *underlying space* together with a measurable function $T : X \to X$ called the *dynamic* of the system. We will always assume the underlying space is of finite measure (and without lost of generality, a probability space). We say that two dynamical systems $(X_1, \mathcal{B}_1, \mu_1, T_1), (X_2, \mathcal{B}_2, \mu_2, T_2)$ are *equivalent* if, up to zero measure sets, there exists a measurable isomorphism $h : X_1 \to X_2$ such that the following diagram commutes:

$$\begin{tikzcd} X_1 \arrow[r, "T_1"] \arrow[d, "g"'] & X_2 \arrow[d, "g"] \\ X_2 \arrow[r, "T_2"] & X_2 \end{tikzcd}$$

and h is measure preserving.

It is a natural problem to study the equivalence of measurable dynamical systems under measure preserving isomorphisms.

Remark. Given a measure-preserving function $T : (X_1, \mathcal{B}_1, \mu_1) \to (X_2, \mathcal{B}_2, \mu_2)$, it naturally extends to a measure-preserving function on the completions of both spaces, $\overline{T} : (X_1, \overline{\mathcal{B}}_1, \overline{\mu}_1) \to (X_2, \overline{\mathcal{B}}_2, \overline{\mu_2})$.

As usual in measure theory, properties holding for the generating semi-algebras, extend to the whole sigma-algebra.

Theorem 1.3. *Suppose $T : (X_1, \mathcal{B}_1, \mu_1) \to (X_2, \mathcal{B}_2, \mu_2)$ is a measurable function, and $\mathcal{S}_i$ are semi-algebras such that $\mathcal{B}_i = \sigma(\mathcal{S}_i)$. If $\mu_1(T^{-1}(S)) = \mu_2(S)$ for every $S \in \mathcal{S}_2$, then T is measure-preserving.*

Proof. Define $\mathcal{C} = \{B \in \mathcal{B}_2 : \mu_1(T^{-1}(B)) = \mu_2(B)\}$ and use prove that it is a sigma-algebra containing $\mathcal{S}_2$. ∎

Now we have an integral version of the invariance property:

Theorem 1.4. *A measurable transformation $T : X \to X$ is $\mu-$ invariant if and only if*

$$\int_X \phi \, \mathrm{d}\mu = \int_X \phi \circ f \, \mathrm{d}\mu$$

for every $\phi \in L^1(\mu)$.

Proof. The property for indicator functions is the the definition of invariance and it extends by linearity of the integral for simple functions. Then it follows for arbitrary functions by dominated convergence theorem. ∎

The set of probability measures on a compact metric space, together with the weak-* topology is a compact space. Using the Schauder-Tychonoff, it is possible to prove the existence of invariant measures. In fact, we have

Theorem 1.5. *Let $T : X \to X$ be a continuous transformation on a compact metric space. Then, there exists a $T-$invariant probability measure on X.*

Proof. See [VO15]. ■

Example (Circle rotations)**.** Let $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\alpha \in S^1$. Define $R_\alpha : S^1 \to S^1$ by $R_\alpha(z) = z \cdot e^{2\pi i\alpha}$. Then the restriction of the 2-dimensional Lebesgue measure to S^1 is R_α invariant. There is an alternative representation for this dynamical system: given $\alpha \in (0, 1)$ define $R_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ by $T(x) = x + \alpha \mod 1$. Again, the measure induced in $\mathbb{R}/\mathbb{Z}$ is invariant for R_α. We will see later that the arithmetic properties of α are reflected in the dynamical properties of the system.

Example (m-adic expansions, full shifts)**.** For a given natural number $m > 1$, every real number $x \in [0, 1)$ admits an expansion in base m in the form

$$x = \frac{a_1}{m^1} + \frac{a_2}{m^2} + \frac{a_3}{m^3} + ... + \frac{a_k}{m^k} + ...$$

with $a_k \in \{0, ...m - 1\}$. We denote $x = [a_1, a_2, a_3, ...]$. Note that the expansion is unique for every irrational number, and for rational numbers it is unique unless $a_k = m - 1$ for every $k > N$ for some natural number $N \geq 1$. In that case, $[a_1, ..., a_N, a_{N+1}, ...] = [a_1, ..., a_N + 1, 0, 0, ...]$. We will follow the convention of writing the number in the second form.

Now we define a related dynamical system. For a fixed base m, define $T : [0, 1) \to [0, 1)$ by $T(x) = mx \mod 1$. If x is represented as $[a_1, a_2, ...]$ then $\sigma(x) = [a_2, a_3, ...]$, that is, T acts as a shift in the m-base expansion of real numbers.

Since T is piecewise affine with m branches with slope m, it is possible to see that the Lebesgue measure is invariant by T. This dynamical system can be represented in an abstract setting by considering an m symbol set $M = \{0, ..., m - 1\}$ and constructing the set of sequences with entries on M, this is, $\Sigma^+ = M^{\mathbb{N}}$, and σ the left-shift as in the m-base expansion. We endow Σ^+ with a topology taking the collection of cylinders $C_{i_0,...i_m} = \{(x_n) \in \Sigma^+ : x_0 = i_0, ..., x_m = i_m\}$ as a basis. Note that this is equivalent to give the discrete topology to M and then the product topology to Σ^+. This topology can also be seen as the one induced by the metric d_θ given by

$$d_\theta((x_n), (y_n)) = \theta^{\min\{n : x_n \neq y_n\}}$$

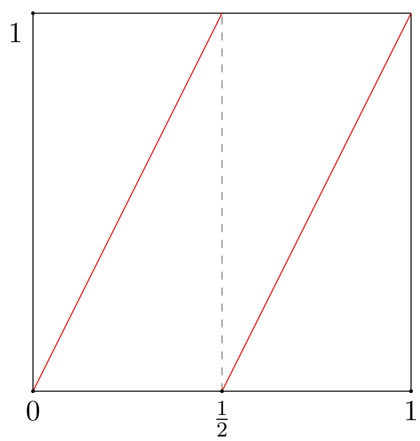

FIGURE 1.1: Plot of T for $m = 2$.

with $\theta \in (0, 1)$.

We construct a family of $\sigma-$invariant Borel measures as follow: given a probability vector $p = (p_1, ..., p_m)$, take μ_p the Bernoulli measure generated by p.

The same construction can be done if we take the two-sided sequences in $M^{\mathbb{Z}}$ just by *mutatis mutandis*.

Example (Markov Shifts)**.** We now proceed to define a more general class of dynamical systems. Given a $m \times m$ matrix $A = (a_{ij})$ with entries in $\{0, 1\}$, consider the subset $\Sigma_A = \{(x_i) \in \Sigma : A_{x_i x_{i+1}=1}\}$. If we restrict the shift σ to Σ_A, we obtain a dynamical system called a *Subshift of Finite Type* of Σ. Note that we can also define the cylinders $C_{i_0,...,i_m}$ as in the full shift, but the may be empty. The set Σ_A can be equipped with a topological structure by taking the subspace topology, which coincides with the topology induced by the restriction of the metric d_θ.

Example (Continued Fractions and Gauss Map)**.** This example is fundamental through the course of this work, so we will treat it with more details. Motivated by rational approximation of irrational numbers, we introduce the Continued Fraction Expansion of a real number. Given a rational number $x \in (0, 1)$, by the Euclid's Algorithm, there exist natural numbers $a_1, ..., a_n$ such that

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots + \cfrac{1}{a_n}}}}$$

which we abbreviate as $[a_1(x), a_2(x), ..., a_n(x)] = [a_1, ..., a_n]$ and call it the *Continued Fraction Expansion* of x. Note that $[a_1, a_2, ..., a_{n-1}, a_n] = [a_1, a_2, ..., a_{n-1}, a_n - 1, 1]$, so every rational number has two different continued fraction expansion. Obviously an irrational number $x \in (0,1)$ cannot admit such expansion. Consider the function

$$G : [0,1] \to [0,1]$$
$$x \mapsto \begin{cases} \dfrac{1}{x} - \left[\dfrac{1}{x}\right] & \text{for } x \in (0,1] \\ 0 & \text{for } x = 0 \end{cases}$$

which clearly defines a dynamic in $[0,1]$. Now we investigate the action of G on the continued fraction expansion of a rational number $x = [a_1, ..., a_n] \in [0,1]$:

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots + \cfrac{1}{a_n}}}} \quad , \quad \frac{1}{x} = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots + \cfrac{1}{a_n}}}$$

$$\left[\frac{1}{x}\right] = a_1 \quad , \quad G(x) = \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \ddots + \cfrac{1}{a_n}}}}$$

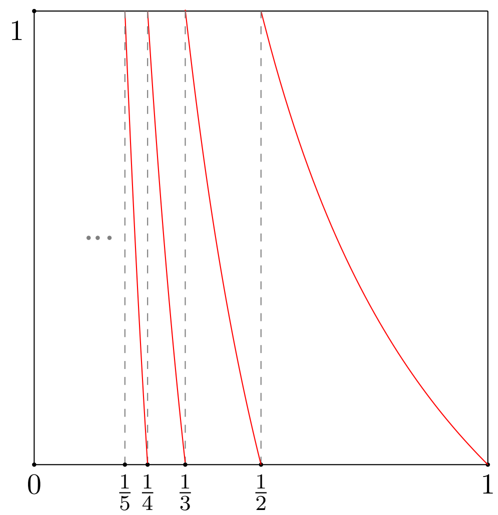

FIGURE 1.2: Plot of the Gauss Map G.

so the Gauss map acts as a shift in the continued fraction expansion of rational numbers. This is the base point to define the continued fraction expansion for irrational numbers. In fact, let $x \in (0,1) \setminus \mathbb{Q}$, and define for every $n \in \mathbb{N}$

$$a_n(x) := \left[\frac{1}{G^n x}\right].$$

Then, we have that

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots + \cfrac{1}{a_n + G^n(x)}}}}$$

Theorem 1.6. *The sequence $(a_1, a_2, ..., a_n, ...)$ defines a sequence of rational numbers $p_n/q_n = [a_1, ..., a_n]$ which converges to the irrational number x.*

Proof. See [EW13]. ■

Thus, we can adopt the notation $x = [a_1, a_2, ..., a_n, ...]$ for $\lim_{n\to\infty}[a_1, ..., a_n]$. It is possible to prove that this limit exists and that every irrational number has a unique continued fraction expansion constructed in this way. With this notation, we can construct rational approximations of irrational numbers by truncating the continued fraction expansion, this is, we approximate an irrational number $x = [a_1, ..., a_n, ...]$ by the sequence of rational numbers $\frac{p_n}{q_n} := [a_1, ..., a_n]$. It is possible to prove that this sequence of rational numbers is the best way to approximate x with a given complexity.

We list some properties of the continued fraction expansion that will be used later.

Theorem 1.7. *For every finite sequence $(b_1, ..., b_n) \in \mathbb{N}^n$, the set*

$$I(b_1, ..., b_n) = \{x \in [0,1] : a_i(x) = b_i \textit{ for } i = 1, ..., n\}$$

is an interval with endpoints

$$\frac{p_n}{q_n} \textit{ and } \frac{p_n + p_{n-1}}{q_n + q_{n-1}}.$$

For every $s \in \mathbb{N}$, we have that the set $I(b_1, ..., b_n, s)$ is an interval with endpoints

$$\frac{(s+1)p_n + p_{n-1}}{(s+1)q_n + q_{n-1}} \textit{ and } \frac{sp_n + p_{n-1}}{sq_n + q_{n-1}}.$$

Proof. See [EW13]. ■

Observe now that G does not preserve Lebesgue measure m. In fact, we have that

$$G^{-1}\left(\left((0, \frac{1}{2})\right)\right) = \bigcup_{n=1}^{\infty} \left(\frac{1}{n+\frac{1}{2}}, \frac{1}{n}\right)$$

and hence

$$m\left(G^{-1}\left(0, \frac{1}{2}\right)\right) = \sum_{n=1}^{\infty}\left(\frac{1}{n} - \frac{1}{n+\frac{1}{2}}\right)$$
$$= 2 - \log 4 \neq \frac{1}{2} = m\left(\left(0, \frac{1}{2}\right)\right).$$

In [Gau03] Gauss introduced an invariant measure for the map G. It is not known how did he find that measure. Given a Borel set $A \subset [0,1]$, we define its *Gauss Measure* μ by

$$\mu(A) = \frac{1}{\log 2}\int_A \frac{\mathrm{d}x}{1+x}.$$

Since G has countable branches given by $\frac{1}{x} - n$ with $n \in \mathbb{N}$, for $(a,b) \subset [0,1]$ we have that

$$G^{-1}((a,b)) = \bigcup_{n=1}^{\infty}\left(\frac{1}{n+b}, \frac{1}{n+a}\right),$$

so its Gauss measure is

$$\mu(G^{-1}(a,b)) = \sum_{n=1}^{\infty}\mu\left(\frac{1}{n+b}, \frac{1}{n+a}\right)$$
$$= \frac{1}{\log 2}\sum_{n=1}^{\infty}\log\left(\frac{\frac{1}{n+a}+1}{\frac{1}{n+b}+1}\right)$$
$$= \frac{1}{\log 2}\log\left(\frac{b+1}{a+1}\right) = \mu((a,b)).$$

By Theorem 1.3, we conclude that G is $\mu-$invariant. Note that the Gauss measure is absolutely continuous with respect to the Lebesgue Measure. Even more, there exist constants $c_1, c_2 > 0$ such that $c_1 m(A) \leq G(A) \leq c_2 m(A)$ for every measurable set A, so every property holding for almost every $x \in (0,1)$ with respect to the Gauss measure, also holds for almost every $x \in (0,1)$ with respect to the Lebesgue measure and vice versa.

Suppose we are given a probability space $(X, \mathcal{B}, \mu)$ with a topology, compatible with the measure structure, and a measurable function $T : X \to X$. One of the main subjects of study of ergodic theory, is the *recurrence*. We can distinguish recurrence at different levels.

Definition 1.8. A point $x \in X$ is said to be *fixed* by T if $T(x) = x$. It is said to be *periodic* with respect to T if there exists an integer $n \geq 1$ such that $T^n(x) = x$. For each $A \in \mathcal{B}$, a point $x \in A$ is said to be *A-recurrent* if there are infinitely many n such that $T^n(x) \in A$.

Remark. Note that A-recurrence is weaker than periodicity, and periodicity is weaker than being fixed.

In many situations, we are interested in studying the dynamic of a pair (X, T) where X is a set and $T : X \to X$ a function preserving some structure on X. Poincare's recurrence theorem guarantees that the existence of a single invariant probability measure μ, implies that almost every point is $A-$recurrent for every non-trivial A, that is $\mu(A) > 0$. Note that the choice of a measure involves an implicit choice of a sigma-algebra structure $\mathcal{B}$ on X, so if μ_1 and μ_2 are two different invariant measures on X, they provide us of different information of the dynamic on the system. The choice of the right invariant measures is one of the main problems on ergodic theory, and thermodynamic formalism arises as a tool to ease that choice.

Theorem 1.9 (Poincare's Recurrence Theorem)**.** *Let $T : X \to X$ be a measurable function on a measurable space, and μ is an invariant probability measure for T. Suppose that A is a measurable set with $\mu(A) > 0$. Then μ-almost every point $x \in A$ is A-recurrent.*

Proof. See [VO15]. ■

Note that Poincare's theorem does not provide information on the return frequency of every point x to the set A. The right hypothesis are given on the next section.

Example. Let $A \subset X$ be a subset of positive measure and $T : X \to X$ a measurable $\mu-$invariant transformation. Then, by Poincare's Recurrence Theorem, the function $\rho : A \to \mathbb{N}$ given by $\rho(a) = \min\{n : T^n(a) \in A\}$ is finite for almost every $a \in A$. We may define, up to a zero measure set, a dynamic $T' : A \to A$ in A by $T'(a) = T^{\rho(a)}(a)$. The dynamic T' is called the *induced system* , and the function ρ is called the *Return time* .

1.2 Ergodicity

Consider the dynamic $T : [0, 1] \to [0, 1]$ of the Figure 1.3.

Note that if we set $A_1 := [0, 1/2]$ and $A_2 := [1/2, 1]$ then the restriction of T to every subset A_i gives us a new dynamic system. Hence, the original system is divided into two pieces which are independent dynamical systems by themselves. We will study then, dynamical systems which are *irreducible* in this sense: they do not have a proper subsystem. This leads to the notion of *ergodicity*.

Definition 1.10. Let $(X, \mathcal{B}, \mu)$ a probability space. A measure-preserving function $T : X \to X$ is said to be *ergodic* with respect to μ if for $E \in \mathcal{B}$ we have that $T^{-1}(E) = E$ implies $\mu(E) = 0$ or $\mu(E) = 1$.

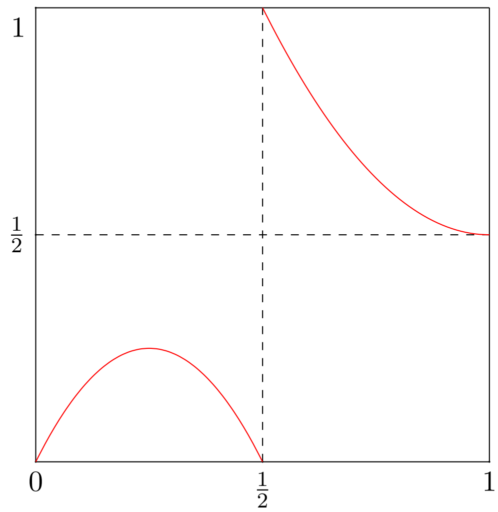

FIGURE 1.3: The system can be decomposed in two different subsystems.

Remark. Sometimes we will change the point of view, fixing the function T and saying that the measure μ is ergodic with respect to T.

Now we have the following characterization of ergodicity in terms of functions, analogue to the given in Theorem 1.4:

Theorem 1.11. *Let $(X, \mathcal{B}, \mu)$ a probability space and T a measure-preserving function on X. The following are equivalent*

1. *T is ergodic.*
2. *Whenever f is measurable and $f \circ T(x) = f(x)$ for every x implies that f is constant a.e.*
3. *Whenever f is measurable and $f \circ T(x) = f(x)$ a.e. implies that f is constant a.e.*
4. *Whenever $f \in L^2$ and $f \circ T(x) = f(x)$ for every x implies that f is constant a.e.*
5. *Whenever $f \in L^2$ and $f \circ T(x) = f(x)$ a.e. implies that f is constant a.e.*

Proof. See [Wal82]. ■

Remark. The above theorem remains true if we change L^2 by L^p for arbitrary $p \geq 1$, including the case $p = \infty$.

The previous theorem is remarkable since it allows us to understand ergodicity as the simplicity of the eigenvalue 1 of the composition operator (Koopman operator) $U_T : f \mapsto f \circ T$ acting on several function spaces, for example, in the Hilbert space L^2. This will be crucial in the next chapters, when we introduce the transfer operator as the dual operator of the Koopman operator. We will return to this topic in Chapter 3.

Example (Irrational Circle Rotations)**.** The Lebesgue measure in S^1 is ergodic with respect to the transformation $R_\alpha : S^1 \to S^1$ given by $R_\alpha([x]) = [x+\alpha]$ for irrational choices of α. In fact, suppose that $\phi \in L^2$ is invariant under composition with R_α. Consider the Fourier expansion of ϕ: there exist a sequence $(a_k)_{k\in\mathbb{Z}}$ of complex numbers such that

$$\phi(x) = \sum_{k\in\mathbb{Z}} a_k e^{2\pi i k x}.$$

Composing with R_α, we obtain

$$\phi(R_\alpha(x)) = \sum_{k\in\mathbb{Z}} a_k e^{2\pi i k(x+\alpha)}.$$

Using the uniqueness of the Fourier expansion and comparing coefficients of ϕ and $\phi \circ R_\alpha$, we obtain that $a_k e^{2\pi i k\alpha} = a_k$. By the irrationality of α, $e^{2\pi i k\alpha} \neq 1$ for every $k \neq 0$, so $a_k = 0$ for every $k \neq 0$, and hence, $\phi(x) = a_0$. By theorem 1.11 we conclude that the Lebesgue measure is ergodic.

Example (Full Shift)**.** Now we prove that the Lebesgue measure m on $[0,1)$ is ergodic with respect to the transformation $T : [0,1) \to [0,1)$ given by $T(x) = 2x \mod 1$. Our two main tools will be the Lebesgue Density Theorem and a *bounded distortion* property of the map. Suppose $A \subset [0,1]$ is an invariant set. The Lebesgue Density Theorem states that almost every $a \in A$ is a density point, that is,

$$\liminf_{\varepsilon\to 0} \left\{ \frac{m(I\cap A)}{m(I)} : I \text{ interval such that } a \in I \subset B(a,\varepsilon) \right\} = 1.$$

Note that T^k is an affine bijection to $(0,1)$ when restricted to the intervals

$$I(k,m) = \left(\frac{m-1}{2^k}, \frac{m}{2^k} \right), \qquad m = 1, ..., 2^k.$$

The set of points on the boundary of the intervals is countable and therefore, has zero Lebesgue measure. Given a density point $a \in A \setminus \{m/2^k : k \in \mathbb{N},\ 0 \le m \le 2^k\}$ and $k \in \mathbb{N}$, there exists a unique $0 \le m_k \le 2^k$ such that $a \in I(k, m_k)$. By the density theorem, we have that

$$\frac{m(I(k,m_k)\cap A)}{m(I(k,m_k))} \to 1 \quad \text{when } k \to \infty.$$

Now, since T^k is an affine bijection of $I(k, m_k)$ to $(0, 1)$, we have the following *bounded distortion* property

$$\frac{m(T^k(E_1))}{m(T^k(E_2))} = \frac{m(E_1)}{m(E_2)}$$

for every pair of measurable sets $E_1, E_2 \subset I(k, m_k)$. Taking $E_1 = I(k, m_k) \cap A$ and $E_2 = I(k, m_k)$ we obtain

$$\frac{m(T^k(I(k, m_k) \cap A))}{m(0, 1)} = \frac{m(I(k, m_k) \cap A)}{m(I(k, m_k))}.$$

Now, since A is invariant, we have that $T^k(I(k, m_k) \cap A) \subset A$, and hence

$$m(A) \geq \frac{m(I(k, m_k \cap A))}{m(I(k, m_k))} \to 1$$

so we conclude that $m(A) = 1$ as we wanted. The same argument proves the ergodicity for the full shift map in m symbols.

Example (Gauss Map)**.** The ergodicity of the Gauss measure μ with respect to the Gauss map G is proved with essentially the same technique used to prove the ergodicity of the Lebesgue measure with respect to the full shift. The only subtle difference in the argument is that the pair (G, μ) satisfies a weaker version of the bounded distortion property satisfied by the full shift system. In fact, as in the previous example, for every $k \in \mathbb{N}$ there exists a collection of intervals $I(k, m)$ indexed by $m \in \mathbb{N}$ such that G^k is a bijection from $I(k, m)$ to $(0, 1)$. There exists a constant $K > 1$ such that for every $k, m \in \mathbb{N}$ and $E_1, E_2 \subset I(k, m)$ measurable sets,

$$\frac{\mu(G^k(E_1))}{\mu(G^k(E_2))} \leq K \frac{\mu(E_1)}{\mu(E_2)}.$$

This *bounded distortion* property follows from the metric properties of the Gauss map. As we did in the previous exercise, we take an invariant set $A \subset (0, 1)$ and a density point $a \in A$ not contained in the boundary of the intervals $I(k, m)$. Then for every $k \in \mathbb{N}$ there exists a unique $m_k \in \mathbb{N}$ such that $a \in I(k, m_k)$. Using $E_1 = I(k, m_k) \cap A$ and $E_2 = I(k, m_k)$ on the above inequality and the Lebesgue Density Theorem, we conclude that $m(A) = 1$. We prove now the bounded distortion property. By computing derivatives, it is possible to see that for every $x \in (0, 1]$ we have

$$|G'(x)| \geq 1 \ , \ |(G^2)'(x)| \geq 2 \ , \ |G''(x)/G'(x)^2| \leq 2.$$

Let g a local inverse of G, defined in some interval where G is bijective and such that $G(g(z)) = z$. Then

$$|(\log |G' \circ g(z)|)'| = \Big|\frac{G''(g(z))g'(z)}{G'(g(z))}\Big| = \Big|\frac{G''(g(z))}{G'(g(z))^2}\Big| \leq 2.$$

This implies that the function $\log |G' \circ g(z)|$ admits 2 as a Lipschitz constant for every choice of g. Now, by the chain rule, we have for every $x, y \in I(k, m)$ that

$$\begin{aligned} \log \frac{|(G^k)'(x)|}{|(G^k)'(y)|} &= \sum_{j=0}^{k-1} \log |G'(G^j(x))| - \log |G'(G^j(y))| \\ &= \sum_{j=1}^{k} \log |G' \circ g_j(G^j(x))| - \log |G' \circ g_j(G^j(y))|, \end{aligned}$$

where g_j is a local inverse for G, defined on the interval $[G^j(x), G^j(y)]$ Using that 2 is Lipschitz constant for $\log |G' \circ g(z)|$, we obtain

$$\begin{aligned} \log \frac{|(G^k)'(x)|}{|(G^k)'(y)|} &= \sum_{j=1}^{k} \log |G' \circ g_j(G^j(x))| - \log |G' \circ g_j(G^j(y))| \\ &\leq 2 \sum_{j=1}^{k} |G^j(x) - G^j(y)| = \sum_{i=0}^{k-1} |G^{k-i}(x) - G^{k-i}(y)|. \end{aligned}$$

By the Mean Value Theorem and the estimates for $|G'|$ and $|(G^2)'|$, we have that

$$|G^k(x) - G^k(y)| \geq 2^{[i/2]} |G^{k-i}(x) - G^{k-i}(y)|$$

for every $i \in \{0, ..., k\}$. This leads to

$$\log \frac{|(G^k)'(x)|}{|(G^k)'(y)|} \leq 2 \sum_{i=0}^{k-1} 2^{-[i/2]} |G^k(x) - G^k(y)| \leq 8 |G^k(x) - G^k(y)| \leq 8.$$

Integrating with respect to the Lebesgue Measure m, we obtain

$$\frac{m(G^k(E_1))}{m(G^k(E_2))} = \frac{\int_{E_1} |(G^k)'| \, \mathrm{d}m}{\int_{E_2} |(G^k)'| \, \mathrm{d}m} \leq e^8 \frac{m(E_1)}{m(E_2)}.$$

Finally, since Gauss and Lebesgue measures are comparable, we have that

$$\frac{m(G^k(E_1))}{m(G^k(E_2))} \leq c \frac{m(G^k(E_1))}{m(G^k(E_2))} \leq c' \frac{m(E_1)}{m(E_2)} \leq K \frac{\mu(E_1)}{\mu(E_2)}.$$

Corollary 1.12. *For every finite sequence $(a_1, ..., a_n) \in \mathbb{N}^n$ and $x \in I(a_1, ..., a_n)$, we have that* $\operatorname{diam} I(a_1, ..., a_n) \asymp |(G^n)'(x)|$.

In 1931 George David Birkhoff proved one of the most important theorems on ergodic theorey, the so called Ergodic Theorem. Earlier that year, John Von Neumann proved independently a weaker version of the theorem, which we also present.

Theorem 1.13 (Von Neumann Ergodic Theorem)**.** *Let $T : X \to X$ be a $\mu-$invariant transformation. Then, for every $f \in L^2$ there exists $\hat{f} \in L^2$ invariant under composition with T (that is, $\hat{f} \circ T = \hat{f}$) such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to \hat{f},$$

where the convergence is in the L^2 norm. Moreover, if μ is ergodic, then $\hat{f} = \int f \; \mathrm{d}\mu$.

Proof. Proceeding as in Theorem 1.4, it is possible to see that if T is $\mu-$invariant, then $\|f \circ T\|_2 = \|f\|_2$. Let $f \in M = \{g - g \circ T : g \in L^2\}$, then

$$\|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\|_2 = \frac{1}{n}\|g \circ T^n - g\|_2 \leq \frac{2}{n}\|g\|_2 \to 0$$

so we conclude that the theorem is true for functions in M. Now suppose that $f \in \overline{M}$, that is, for every $\varepsilon > 0$ there exists $f_\varepsilon \in M$ such that $\|f - f_\varepsilon\|_2 < \varepsilon$. Take N_0 large enough such that

$$\|\frac{1}{n}\sum_{k=0}^{n-1} f_\varepsilon \circ T^k\|_2 < \varepsilon$$

for every $n \geq N_0$. Then, for every $n \geq N_0$ we have that

$$\begin{aligned}\|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\|_2 &\leq \|\frac{1}{n}\sum_{k=0}^{n-1}(f - f_\varepsilon) \circ T^k\|_2 + \|\frac{1}{n}\sum_{k=0}^{n-1} f_\varepsilon \circ T^k\|_2 \\ &\leq \frac{1}{n}\sum_{k=0}^{n-1}\|(f - f_\varepsilon) \circ T^k\|_2 + \varepsilon \\ &\leq 2\varepsilon\end{aligned}$$

so we conclude that $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to 0$ in L^2. Now we can can decompose L^2 as $L^2 = \overline{M} \oplus \overline{M}^\perp$. We claim that $\overline{M}^\perp = \{f \in L^2 : \; f = f \circ T\}$. In fact, take $f \in \overline{M}^\perp$, then

$$\begin{aligned}\|f - f \circ T\|_2^2 &= \langle f - f \circ T, f - f \circ T\rangle \\ &= \|f\|_2^2 - 2\langle f, f \circ T\rangle + \|f \circ T\|_2^2 \\ &= 2\|f\|_2^2 - 2\langle f, f - (f - f \circ T)\rangle \\ &= 2\|f\|_2^2 - 2\|f\|_2^2 = 0\end{aligned}$$

since $\langle f, (f - f \circ T) \rangle = 0$, so $f = f \circ T$ a.e. Finally, we conclude that the theorem holds in L^2. ■

The previous theorem can be formulated in purely operator theoretic terms (see [VO15]).

Corollary 1.14. *Let T be a $\mu-$invariant transformation. Then, for every $f \in L^1$ there exists $\hat{f} \in L^1$ invariant under composition with T such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to \hat{f}$$

in the L^1 norm.

Proof. For $f \in L^\infty \subset L^2$, by Von Neumman's Ergodic there exists $\hat{f} \in L^2$ invariant under composition with T such that $f_n/n = (f + f \circ T + ... + f \circ T^{n-1})/n$ converges to $\hat{f}$ in the L^2 norm. Note that $\|f_n/n\|_\infty \leq \|\hat{f}\|_\infty$, hence

$$|\langle f_n/n, 1_B \rangle|_2 \leq \|\hat{f}\|_\infty \mu(B)$$

for every measurable set B. Since $f_n/n \to \hat{f}$ in L^2, we obtain

$$|\langle \hat{f}, 1_B \rangle|_2 \leq \|\hat{f}\|_\infty \mu(B)$$

and thus $\|\hat{f}\|_\infty \leq \|f\|_\infty$. Then, $\hat{f} \in L^\infty$, and since $\|\cdot\|_1 \leq \|\cdot\|_2$, we conclude that $f_n/n \to \hat{f}$ in the L^1 norm. With this, we established the convergence for the dense set L^∞. A standard approximation argument completes the proof for arbitrary functions in L^1. ■

Theorem 1.15 (Birkhoff Ergodic Theorem)**.** *Let $T : X \to X$ be a $\mu-$invariant transformation. Then, for every $f \in L^1$ there exists $\hat{f} \in L^1$ invariant under composition (that is, $\hat{f} \circ T = \hat{f}$) such that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to \hat{f},$$

where the convergence is a.e., and

$$\int_X f^* \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu.$$

Moreover, if μ is ergodic, then $\hat{f} = \int f \, \mathrm{d}\mu$.

Proof. In order to prove the theorem, we need several lemmas:

Lemma 1.16 (Maximal Inequality). *For $n \geq 1$, set $f_n = f + f \circ T + ... + f \circ T^{n-1}$, $f_0 = 0$ and $F_N = \max_{0 \leq n \leq N} f_n$. Then*

$$\int_{\{x : F_N(x) > 0\}} f \, \mathrm{d}\mu \geq 0$$

for all $N \geq 1$.

Proof. For each $N \geq 1$, $F_N \in L^1$. Since the composition with T preserves the real number ordering, we have that

$$F_N \circ T \geq f_n \circ T$$

for every $0 \leq n \leq N$, hence

$$F_N \circ T + f \geq f_n \circ T + f = f_{n+1}$$

and then

$$F_N \circ T + f \geq \max_{1 \leq n \leq N} f_n.$$

For $x \in P = \{x : F_N(x) > 0\}$ we have

$$F_N(x) = \max_{0 \leq n \leq N} f_n(x) = \max_{1 \leq n \leq N} f_n(x)$$

and therefore

$$\begin{aligned} F_N \circ T(x) + f(x) &\geq F_N(x) \\ f(x) &\geq F_N(x) - F_N \circ T(x) \end{aligned}$$

for every $x \in P$. Integrating this inequality, we obtain

$$\begin{aligned} \int_P f \, \mathrm{d}\mu &\geq \int_P F_N \, \mathrm{d}\mu - \int_P F_N \circ T \, \mathrm{d}\mu \\ &= \int_X F_N \, \mathrm{d}\mu - \int_P F_N \circ T \, \mathrm{d}\mu \\ &\geq \int_X F_N \, \mathrm{d}\mu - \int_X F_N \circ T \, \mathrm{d}\mu \\ &= \|F_N\|_1 - \|F_N \circ T\|_1 \geq 0 \end{aligned}$$

since the Koopman's Operator has norm less or equal to 1. ∎

Lemma 1.17 (Maximal Ergodic Theorem). *For $g \in L^1$ and $\alpha \in \mathbb{R}$, let*

$$E_\alpha = \Big\{x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x) > \alpha \Big\}.$$

Then

$$\alpha\mu(E_\alpha) \leq \int_{E_\alpha} g \, \mathrm{d}\mu \leq \|g\|_1.$$

Moreover, $\alpha\mu(E_\alpha \cap A) \leq \int_{E_\alpha \cap A} g \, \mathrm{d}\mu$ whenever $T^{-1}A = A$.

Proof. Let $f = (g - \alpha)$, so E_α becomes

$$E_\alpha = \bigcup_{N=0}^{\infty} \{x : F_N(x) > 0\}.$$

It follows that $\int_{E_\alpha} f \, \mathrm{d}\mu \geq 0$ and then $\int_{E_\alpha} g \, \mathrm{d}\mu \geq \alpha\mu(E_\alpha)$. For the second part of the lemma, apply the same argument to $f = (g - \alpha)$ to the restriction of the system to A. ■

Now we proceed to prove Birkhoff's Ergodic Theorem. For $x \in X$, define

$$f^*(x) = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x),$$
$$f_*(x) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x).$$

Note that

$$\frac{n+1}{n} \left(\frac{1}{n+1} \sum_{i=0}^{n} f(T^i x) \right) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(Tx)) + \frac{1}{n} f(x).$$

By taking the limit along a subsequence for which the left-hand side converges to the limsup, we obtain that $f^* \leq f^* \circ T$. Analogously, we obtain that $f^* \geq f^* \circ T$, and then $f^* = f^* \circ T$. The same argument shows that $f_* = f_* \circ T$, this is, f^* and f_* are invariant by T.

Now, for $\alpha, \beta \in \mathbb{Q}$ with $\alpha > \beta$, define

$$E_\alpha^\beta = \{x \in X : f_*(x) < \beta \text{ and } f^*(x) > \alpha\}.$$

Since f_* and f^* are $T-$invariant, we have that E_α^β are also $T-$invariant. By the Maximal Ergodic Theorem, we have that

$$\int_{E_\alpha^\beta} f \,\mathrm{d}\mu \geq \alpha\mu(E_\alpha^\beta).$$

Replacing f by $-f$, we obtain

$$\int_{E_\alpha^\beta} f \,\mathrm{d}\mu \leq \beta\mu(E_\alpha^\beta)$$

and hence $\mu(E_\alpha^\beta) = 0$. Therefore

$$\mu(\{x : f_*(x) < f^*(x)\}) = \mu\left(\bigcup_{\alpha<\beta,\alpha,\beta\in\mathbb{Q}} E_\alpha^\beta\right) = 0$$

and consequently $f_* = f^*$ almost everywhere and hence

$$g_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) \to f^*(x)$$

for almost every $x \in X$. By Corollary 1.14 we have that $g_n \to f' \in L^1$ in the L^1 norm, and the convergence is almost everywhere for some subsequence g_{n_k}. This implies that $f^* = f' \in L^1$ and that $g_n \to f^*$ in the L^1 norm. Finally

$$\int_X f \,\mathrm{d}\mu = \int_X g_n \,\mathrm{d}\mu = \int_X f^* \,\mathrm{d}\mu.$$

$\blacksquare$

Remark. If we take an indicator function on Birkhoff's theorem, we obtain that the frequency of return to a given set is essentially equal to the measure of that set. This is, larger sets are visited more frequently than smaller sets. Note that this is a quantitative version of Poincare's recurrence theorem:

$$\lim_{n\to\infty} \frac{\mathrm{card}\{n \in \{0, ..., n-1\} : T^n(x) \in A\}}{n} = \mu(A)$$

for almost every $x \in X$.

Using the ergodic theorem, we can formulate an alternative characterization of ergodicity

Corollary 1.18. *Let $(X, \mathcal{B}, \mu)$ a probability space and T a measure preserving transformation on X. Then T is ergodic if and only if*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B)$$

for every $A, B \in \mathcal{B}$.

Proof. Assume T is ergodic. By Birkhoff Ergodic Theorem with $f = 1_A$ we get

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_A(T^k(x)) = \mu(A)$$

for almost every $x \in X$. Multiplying by 1_B and using the dominated convergence theorem, we get that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) = \mu(A)\mu(B).$$

Conversely, suppose there exists $E \in \mathcal{B}$ such that $T^{-1}E = E$. Then

$$\mu(E) = \frac{1}{n} \sum_{k=0}^{n-1} \mu(E) \to \mu(E)^2$$

so $\mu(E) = 0$ or $\mu(E) = 1$. ∎

The above characterization is remarkable since relates the notion of ergodicity with the independence of the sets A and $T^{-i}(B)$ when i becomes large. Hence ergodicity can be seen as *independence of means*. In the next section we formulate a stronger notion, sometimes easier to prove than ergodicity.

There are several generalizations of the Ergodic Theorems. We begin by observing that if we set $\varphi_n(x) = \sum_{j=0}^{n-1} \varphi(T^j(X))$ then we have the following identity

$$\varphi_{m+n} = \varphi_m + \varphi_n \circ T^m$$

for every $m, n \geq 1$. A sequence of functions $\varphi_n : X \to \mathbb{R}$ for which the previous equality holds, is called *additive*. More generally, it is called *sub-additive* if the following identity holds

$$\varphi_{m+n} \leq \varphi_m + \varphi_n \circ T^m$$

for every $m, n \geq 1$.

Theorem 1.19 (Kingman Sub-additive Ergodic Theorem). *Let $(X, \mathcal{B}, \mu)$ a probability space, $T : X \to X$ a $\mu-$invariant transformation and $\varphi_n : X \to \mathbb{R}$ a sub-additive sequence of measurable functions such that $\varphi_1^+ = \max\{\varphi_1, 0\} \in L^1(\mu)$. Then the sequence $(\varphi_n/n)_n$ converges $\mu-$a.e. to a $T-$invariant function $\varphi : X \to [-\infty, \infty)$. Moreover, $\varphi^+ \in L^1(\mu)$ and*

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

Remark. If $A : X \to GL(d)$ is a measurable function, define $\phi^n(x) = \Pi_{k=0}^{n-1} A(T^k(x))$ for every $n \geq 1$ and $x \in X$. Then the sequence $\varphi_n(x) = \log \|\phi^n(x)\|$ is sub-additive.

Remark. The proof presented in [AB09] does not use the Birkhoff's Ergodic Theorem, so it is possible to deduce it as a consequence of the Kingman's Ergodic Theorem.

We present now some number theoretical applications of the Ergodic Theorems.

Definition 1.20. We say that a real number x is *normal* in base k, if for every $m \in \mathbb{N}$ and $(a_1, ..., a_m) \in \{0, ..., k-1\}^m$, we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{[a_1,...,a_m]}(x) = \frac{1}{m}.$$

Note that the expression of the left side quantifies the frequency of appearance of the string $a_1...a_m$ in the base-k expansion of the number x, and the normality implies that every string appears with the same frequency.

Theorem 1.21 (Borel Normal Numbers Theorem). *Let $k \geq 2$ be a positive integer. Then, m-a.e. $x \in [0, 1]$ is normal in base k.*

Proof. We present the proof of the simple normality of almost every $x \in [0, 1]$. The proof of the normality is just an extension of the argument.

Fix an integer $k \geq 2$ and a digit $0 \leq i \leq k-1$. Consider the transformation $T : [0, 1) \to [0, 1)$ given by $T(x) = kx \mod 1$. Every number $x \in [0, 1]$ has a expansion in base k, say $x = (x_0, x_1, ...)$ This transformation is ergodic with respect to the Lebesgue measure, so we can use the Ergodic Theorem for the function $\chi_{i,k}$, the indicator function of the interval $(\frac{i}{k}, \frac{i+1}{k})$, and conclude that

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{card}\{k :\in \{0, ..., n-1\} : x_k = i\} = \int_{(0,1)} \chi_{i,k} \, \mathrm{d}m = \frac{1}{k}.$$

■

No explicit number is know to be normal in every base. There are several examples of normal numbers in a fixed base, see [Cha33], [CE46] and [DE52].

Remark. The notion of normality can be extended to continued fraction expansion using the measure of cylinders given by the Gauss measure.

The Ergodic Theorem allowed us to formulate an alternative characterization of ergodicity in terms of correlations, so now we formulate two stronger notions of decay of correlations.

Definition 1.22. Let $(X, \mathcal{B}, \mu)$ a probability space, and T a measure-preserving transformation on X. The transformation T is *weak-mixing* if

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0$$

for all $A, B \in \mathcal{B}$, and it is *strong-mixing* (or mixing) if

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$$

for all $A, B \in \mathcal{B}$.

The definition of mixing can be understood in terms of covariance of random variables.

Proposition 1.23. *A measure μ is mixing if and only if*

$$\lim_{n\to\infty} \Big| \int_X (f \circ T^n) \cdot g \,\mathrm{d}\mu - \int_X f \,\mathrm{d}\mu \int_X g d\mu \Big| = 0$$

for every $f, g \in L^1$.

Remark. Trivially strong-mixing implies mixing and mixing implies ergodicity.

1.3 Dimension Theory

As observed in the previous sections, Birkhoff's Ergodic Theorem implies that almost every real number is normal in every base. In particular, each digit has a frequency of $1/m$ on the m-ary expansion of almost every real number. Thus, the sets of numbers having a non-generic behavior in terms of these frequencies are invisible to Lebesgue measure. The dimension theory arises as a powerful tool to measure the complexity of the sets having non-generic behaviors. We begin by introducing three notions of dimension on metric spaces.

Definition 1.24. Consider a set $X \subset \mathbb{R}^n$. We define the *lower* and *upper box dimension* of X as the following limits

$$\underline{\dim}_B X = \liminf_{\varepsilon \to 0} \frac{\log N(X, \epsilon)}{-\log \varepsilon} \quad , \quad \overline{\dim}_B X = \limsup_{\varepsilon \to 0} \frac{\log N(X, \epsilon)}{-\log \varepsilon}$$

where $N(X, \varepsilon)$ is the least number of balls of radius ε needed to cover X.

By definition, we have the inequality

$$\underline{\dim}_B X \leq \overline{\dim}_B X.$$

When both limits coincide, we call the common value the *box dimension* of X and denote it by $\dim_B(X)$.

Remark. Note that the previous definition depends only on the metric structure of X, so the box dimension definitions can be carried out to an arbitrary metric space.

We use the box dimension of sets to define the box dimension of a measure.

Definition 1.25. Let μ be a finite measure in X. The *lower* and *upper box dimension* of μ are defined respectively by

$$\underline{\dim}_B \mu = \lim_{\delta \to 0} \inf_Z \{\underline{\dim}_B(Z) : \mu(Z) \geq \mu(X) - \delta\},$$
$$\overline{\dim}_B \mu = \lim_{\delta \to 0} \inf_Z \{\overline{\dim}_B(Z) : \mu(Z) \geq \mu(X) - \delta\}.$$

Again, we have the following inequality

$$\underline{\dim}_B \mu \leq \overline{\dim}_B \mu.$$

Recall the diameter of a set $U \subset \mathbb{R}^n$ is given by

$$\operatorname{diam} U = \sup\{d(x, y) : x, y \in U\}.$$

For a cover $\mathcal{U}$ of a set $X \subset \mathbb{R}^n$, its diameter is given by

$$\operatorname{diam} \mathcal{U} = \sup\{\operatorname{diam} U : U \in \mathcal{U}\}.$$

Definition 1.26. Given $X \subset \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, the *$\alpha-$dimensional Hausdorff measure* of X is given by

$$m(X, \alpha) = \lim_{\delta \to 0} \inf_{\mathcal{U}} \sum_{U \in \mathcal{U}} (\operatorname{diam} U)^\alpha,$$

where the infimum is taken over finite or countable covers $\mathcal{U}$ of X with $\operatorname{diam}\mathcal{U} \leq \delta$.

It is possible to prove that there exists a number $s \in [0, \infty]$ such that $m(X, \alpha) = \infty$ for $t < s$ and $m(X, \alpha) = 0$ for $t > s$, since $m(X, \alpha)$ is decreasing in α for a fixed set X.

Definition 1.27. The unique number

$$s = \inf\{\alpha \in [0, \infty] : m(X, \alpha) = 0\}$$

is called the *Hausdorff dimension* of X.

Example. The Cantor Set K, defined as the set of numbers in $[0, 1]$ having just 0 and 2 on their ternary expansion, has Hausdorff Dimension equal to $\frac{\log 2}{\log 3}$. Later we will deduce this as a consequence of a more general construction.

As before, we extend the notion of Hausdorff dimension to arbitrary measures on X.

Definition 1.28. Let μ be a finite measure on X. The *Hausdorff dimension* of μ is defined by

$$\dim_H \mu = \inf\{\dim_H(Z) : \mu(X \setminus Z) = 0\}.$$

Proposition 1.29. *For every set $X \subset \mathbb{R}^n$ and measure μ in X we have the following inequalities*

$$\dim_H X \leq \underline{\dim}_B X, \tag{1.1}$$

$$\dim_H \mu \leq \underline{\dim}_B \mu. \tag{1.2}$$

Proof. See [Bar08]. ■

Now we define the pointwise dimension which takes into account how concentrated is the measure on each point.

Definition 1.30. The *lower* and *upper pointwise dimensions* of the measure μ at a point $x \in X$ is given by

$$\underline{d}_\mu(x) = \liminf_{r \to 0} \frac{\log \mu(B(x, r))}{\log r}\ ,\ \overline{d}_\mu(x) = \limsup_{r \to 0} \frac{\log \mu(B(x, r))}{\log r}.$$

When both limits coincide, we call the common value the *pointwise dimension* of μ at x and denote it by $d_\mu(x)$.

The following proposition is useful to obtain bounds for the Hausdorff dimension.

Proposition 1.31. *Let $X \subset \mathbb{R}^n$ and $\alpha \in (0, \infty]$, then*

1. *If $\underline{d}_\mu(x) \geq \alpha$ for $\mu-$almost every $x \in X$, then $\dim_H \mu \geq \alpha$;*
2. *If $\underline{d}_\mu(x) \leq \alpha$ for every $x \in X$, then $\dim_H X \leq \alpha$*

Proof. See [Bar08]. ∎

In order to estimate the Hausdorff measure of a set, we use measures supported on that set.

Theorem 1.32 (Mass distribution Principle). *Let μ be a probability measure on a compact metric space X and suppose there exist numbers $d, K, r > 0$ such that*

$$\mu(B) \leq K(\operatorname{diam} B)^d \tag{1.3}$$

for every measurable set $B \subset X$ with $\operatorname{diam} B < r$. Then if $\mu(A) > 0$ we have $m(A, d) > 0$ and hence $\dim_H A \geq d$.

Proof. Without lost of generality, we prove it for $A = X$. Let $\mathcal{U}$ a cover of X with $\operatorname{diam} \mathcal{U} < r$, then

$$\frac{\mu(U)}{K} \leq (\operatorname{diam} U)^d$$

for every $U \in \mathcal{U}$. Summing over the cover, we obtain

$$\frac{\mu(X)}{K} \leq \sum_{U \in \mathcal{U}} \frac{\mu(U)}{K} \leq \sum_{U \in \mathcal{U}} (\operatorname{diam} U)^d,$$

hence

$$\frac{\mu(X)}{K} \leq \inf_{\mathcal{U}} \sum_{U \in \mathcal{U}} (\operatorname{diam} U)^d.$$

Then $0 < \mu(X) \leq Km(X, d)$, so $\dim_H X \geq d$ ∎

Remark. The Hausdorff and pointwise dimensions (among other quantities) can be understood as particular cases of a more general notion called $u-$dimension, developed by Pesin.

We calculate now the dimensions of a geometric construction. Consider a set of disjoint closed intervals $\Delta_1, ..., \Delta_p$ of the real line, with lengths $\lambda_1, ..., \lambda_p$. For each k, we choose p disjoint closed intervals $\Delta_{k1}, ..., \Delta_{kp} \subset \Delta_k$ with length $\lambda_k\lambda_1, ..., \lambda_k\lambda_p$. Iterating this process, for each $n \in \mathbb{N}$ we get p^n disjoint closed intervals $\Delta_{i_1 \dots i_n}$ with length $\lambda_{i_1} \dots \lambda_{i_n}$. The limit set is defined by

$$F = \bigcap_{n=1}^{\infty} \bigcup_{i_1, \dots, i_n} \Delta_{i_1 \dots i_n}.$$

Proposition 1.33. *Let s be the unique solution of the equation*

$$\sum_{k=1}^{p} \lambda_k^s = 1. \tag{1.4}$$

Then

$$\dim_H F = \underline{\dim}_B F = \overline{\dim}_B F = s.$$

Moreover, $0 < m(F, s) < \infty$.

Proof. It is possible to check the existence of a unique solution of the equation defining s, by noticing that the function $f(x) = \sum_{k=1}^{p} \lambda_k^x$ behaves asymptotically as an exponential function. Now we proceed to compute the Hausdorff dimension of F. Obtaining an upper bound is straightforward, since F is equipped with a natural cover: for each n, the sets $\Delta_{i_1 \dots i_n}$ form a cover and we have that

$$\sum_{i_1 \dots i_n} (\operatorname{diam} \Delta_{i_1 \dots i_n})^s = \sum_{i_1 \dots i_n} (\lambda_{i_1} \dots \lambda_{i_n})^s = \left(\sum_{k=1}^{p} \lambda_k^s \right)^n = 1,$$

thus $m(F, s) \leq 1$ since $\operatorname{diam} \Delta_{i_1 \dots i_n} \to 0$. Consequently, $\dim_H F \leq s$.

To obtain a lower bound we need to work more. Define a probability measure μ in F by setting

$$\mu(\Delta_{i_1 \dots i_n}) = (\lambda_{i_1} \dots \lambda_{i_n})^s.$$

Now we construct a *Moran cover* of F: given $\omega = (i_1, i_2, \dots) \in \{1, \dots, p\}^{\mathbb{N}}$ and $r \in (0, 1)$, there exists a unique integer $n(\omega, r)$ such that

$$\lambda_{i_1} \dots \lambda_{i_{n(\omega,r)}} < r \leq \lambda_{i_1} \dots \lambda_{i_{n(\omega,r)-1}}.$$

For fixed r, the set $\{\Delta(\omega, r) = \Delta_{i_1, \dots, i_{n(\omega,r)}}\}$ forms a cover by pairwise disjoint sets. We have the following estimation for the diameter of $\Delta(\omega, r)$

$$r \min_{k=1 \dots p} \{\lambda_k\} \leq \operatorname{diam} \Delta(\omega, r) < r.$$

Setting $c = (\min\{\lambda_1, ..., \lambda_p\})^{-1}$ we get that for any interval I with $\operatorname{diam} I = r$, there is at most c of the sets $\Delta(\omega, r)$ that intersect I. Then

$$\mu(I) \leq \sum_{\Delta(\omega,r)\cap I\neq\emptyset} \mu(\Delta(\omega, r)) < \sum_{\Delta(\omega,r)\cap I\neq\emptyset} r^s \leq cr^s.$$

Take a countable cover $\mathcal{U}$ of F and a set $U \in \mathcal{U}$. Then U is contained in an interval I_U with $\operatorname{diam} I_U = \operatorname{diam} U$, so by the previous inequality we obtain

$$\mu(U) \leq \mu(I_U) \leq c(\operatorname{diam} U)^s.$$

Summing over the cover

$$1 = \mu(F) = \sum_{U\in\mathcal{U}} \mu(U) \leq c \sum_{U\in\mathcal{U}} (\operatorname{diam} U)^s,$$

which implies that $m(F, s) \geq 1$ and consequently, $\dim_H F = s$.

By Proposition 1.29, it is enough to find an upper bound for the upper box dimension. For this purpose, we consider the Moran Cover constructed above for a fixed $r \in (0, 1)$. By the compactness of F, there exists a finite subcover, namely $\hat{\Delta}_1, ..., \hat{\Delta}_{N(r)}$. By construction, $\operatorname{diam} \hat{\Delta}_j < r$ for $j = 1, ..., N(r)$ and

$$(\operatorname{diam} \Delta_{i_1...i_m})^s = \sum_{i_{m+1}...i_n} (\operatorname{diam} \Delta_{i_1...i_n})^s,$$

hence

$$\sum_{j=1}^{N(r)} (\operatorname{diam} \hat{\Delta}_j) = 1.$$

This, together with the inequality $\frac{r}{c} \leq \operatorname{diam} \hat{\Delta}_j$ imply that $N(r) \leq (c/r)^s$ (the sets $\hat{\Delta}_j$ are pairwise disjoint). Then

$$N(F, r) \leq N(r) \leq (c/r)^s$$

and

$$\overline{\dim}_B F \leq \limsup_{r\to 0} \frac{s \log(c/r)}{-\log r} = s$$

which completes the proof. ■

Remark. Note that this calculation immediately implies that the Hausdorff dimension of the Cantor set is equal to $\frac{\log 2}{\log 3}$.

Chapter 2

Thermodynamic Formalism

This chapter is based on [Wal82], [VO15] and [Fal97].

2.1 Entropy

There are several motivations to define the entropy of a dynamical system. It was originally introduced by Kolmogorov as an invariant of dynamical systems able to distinguish between the two-symbol and three-symbol shifts.

The entropy arose for the first time measure of the complexity of a thermodynamic system. In fact, Boltzmann proposed it as a logarithmic measure of the total possible microscopic states of a particle system. The second law of thermodynamics is classically formulated as *The entropy function of an isolated systems tends to its maximum*. For a many particle system, the second law implies that the thermodynamic equilibrium is reached when the systems achieves its maximum complexity. This statement is not only true for classical systems, yet for quantum systems.

Inspired in the thermodynamic works of Boltzmann, the American mathematician Claude Shannon published in 1949 a celebrated article, *A Mathematical Theory of Communication*, in which he started the information theory. In this article, Shannon introduced the notion of *information entropy*, as a measure of the uncertainty of a message sent between two information sources.

Interested in problems on information theory and dimension of functional spaces, Andrey Kolmogorov developed the notion of *metric entropy* for Bernoulli systems and then for quasi-regular dynamical systems, which allowed him to distinguish the 2-shift and the 3-shift as non metrically isomorphic. Later, this notion was expanded for the whole class of dynamical systems by Yakov Sinai. The entropy became a robust invariant of dynamical systems, and in some cases, even a classifying object.

Definition 2.1. Let $(X, \mathcal{B}, \mu)$ a probability space. A *measurable partition* of X,is a countable or finite subset $\mathcal{P}$ of $\mathcal{B}$ such that $\mu(P \cap P') = 0$ for every $P \neq P' \in \mathcal{P}$ and $\mu(\bigcup_{\mathcal{P}} P) = 1$. For a countable family of partitions $\{\mathcal{P}_n\}_n$, we define its *joint* as the partition $\bigvee_n \mathcal{P}_n := \{\bigcap_n P_n : P_n \in \mathcal{P}_n\}$. The *entropy* of the partition $\mathcal{P}$ is defined as

$$H_\mu(\mathcal{P}) := \sum_{P \in \mathcal{P}} -\mu(P) \log \mu(\mathcal{P})$$

with the convention $0 \log 0 = 0$.

Remark. In what follows, we restrict to partitions with finite entropy. Every finite partition has finite entropy.

Remark. Note that $H_\mu(\mathcal{P}) \leq \log(\operatorname{card} \mathcal{P})$ and the equality holds if and only if $\mu(P) = 1/\operatorname{card} \mathcal{P}$ for every $P \in \mathcal{P}$.

The following definition is based in the probabilistic notion of independence and conditional information.

Definition 2.2. The *conditional entropy* of a partition $\mathcal{P}$ with respect to the partition $\mathcal{Q}$ is the number

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

We have defined the notion of entropy associated to a fixed partition of the space. Now we apply the dynamic to the partition to obtain dynamical partitions and compute their entropy.

Let $T : X \to X$ be a measurable function preserving the probability measure μ. For a fixed partition $\mathcal{P}$, we denote for $n \geq 1$

$$\mathcal{P}^n := \bigvee_{i=0}^{n-1} T^{-i}(\mathcal{P}) = \{P_{i_0} \cap T^{-1}(P_{i_1}) \cap ... \cap T^{-n+1}(P_{i_{n-1}}) : P_{i_j} \in \mathcal{P}\}.$$

Now consider the sequence of entropies $\{H_\mu(\mathcal{P}^n)\}$. We have the following lemma, fundamental to ensure the existence of dynamically defined quantities:

Lemma 2.3 (Fekete lemma)**.** *Let $\{a_n\}$ a subadditive sequence, that is, $a_{n+m} \leq a_n + a_m$. Then the limit*

$$\lim_{n \to \infty} \frac{a_n}{n}$$

exists (it can be $-\infty$) and it is equal to

$$\inf_n \frac{a_n}{n}.$$

Proof. Let $L = \inf_n a_n/n \in [-\infty, \infty)$ and B any real number with $B > L$. There exists $k \geq 1$ such that $\frac{a_k}{k} < B$. For $n > k$, write $n = kp + q$ with p, q integer numbers such that $p \geq 1$ and $1 \leq q \leq k$. Then, $a_n \leq a_{kp} + a_q \leq pa_k + a_q \leq pa_k + \alpha$ where $\alpha = \max a_i : 1 \leq i \leq k$. Dividing by n in the last inequality, we obtain

$$\frac{a_n}{n} \leq \frac{pk}{n}\frac{a_k}{k} + \frac{\alpha}{n}.$$

For n large enough, we get $L \leq \frac{a_n}{n} < B$. By letting $B \to L$, we conclude that

$$\lim_{n\to\infty} \frac{a_n}{n} = L = \inf_n \frac{a_n}{n}.$$

■

With the previous lemma, we call the *entropy* of T with respect to the measure μ and the partition $\mathcal{P}$ to the limit

$$h_\mu(T, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{P}^n)$$

and finally, the (metric) *entropy of T* is defined by

$$h_\mu(T) = \sup_{\mathcal{P}} h_\mu(T, \mathcal{P})$$

where the supremum is taken over all partitions with finite entropy.

We have the following criteria for partitions reaching the supremum in the definition of the entropy of a mapping.

Theorem 2.4 (Kolmogorov-Sinai)**.** *Let $\mathcal{P}$ be a partition with finite entropy such that $\bigcup_n \mathcal{P}^n$ generates the $\sigma-$algebra. Then*

$$h_\mu(T) = h_\mu(T, \mathcal{P}).$$

Proof. See [VO15]. ■

Example (Circle rotations)**.** We will just use the fact that the rotation R_α is an homeomorphism of S^1 and that the Lebesgue measure m on S^1. Consider a partition $\mathcal{P}$ of S^1 defined by points $x_1, ..., x_m$, then for every $k \geq 1$ the partition $R_\alpha^{-k}(\mathcal{P})$ is defined by the points $R_\alpha^{-k}(x_i)$. Hence, we have that $\operatorname{card}\mathcal{P}^n \leq mn$, and by the remark of the definition of entropy, we have that

$$h_m(R_\alpha, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H_m(\mathcal{P}^n) \leq \lim_{n\to\infty} \frac{1}{n} \log(\operatorname{card}\mathcal{P}^n) = \lim_{n\to\infty} \frac{1}{n} \log mn = 0.$$

By the Kolmogorov-Sinai Theorem, we conclude that R_α has zero entropy with respect to the Lebesgue measure.

Example (Full Shift)**.** Recall the transformation $T : [0,1) \to [0,1)$ given by $T(x) = mx \mod 1$. Consider the (measurable) partition of $[0,1)$ given by $\mathcal{P} = \{(\frac{k}{10}, \frac{k+1}{10}) : k = 0, ..., m-1\}$. Then, $\bigcup_n \mathcal{P}^n$ generates the $\sigma-$algebra of measurable sets of $[0,1)$. By the Kolmogorov-Sinai theorem, we get that $h_m(T) = \log m$, where the entropy is calculated with respect to the Lebesgue measure m.

Example (Gauss Map)**.** This calculation is more subtle. It is possible to prove that

$$h_\mu(G) = \int \log |G'| \, \mathrm{d}\mu,$$

where μ is the Gauss measure. The integral of the right hand side reduces to

$$\int \log |G'| \, \mathrm{d}\mu = \int_0^1 \frac{-2\log x \ \mathrm{d}x}{(1+x)\log 2} = \frac{\pi^2}{6\log 2}.$$

To show the first equality, consider the partition $(0,1)$ given by $\mathcal{P} = \{1/(1+m), 1/m\}$ for $m \geq 1$ and the dynamical iterates $\mathcal{P}^n$. By the Kolmogorov-Sinai Theorem, we may use $\mathcal{P}$ to compute to entropy $h_\mu(G)$. Recall that G^n is a diffeomorphic bijection from every element of $\mathcal{P}^n$ to $(0,1)$. Note that the entropy function associated to the partition $\mathcal{P}^n$ can be written as

$$H_\mu(\mathcal{P}^n) = \sum_{P_n \in \mathcal{P}^n} -\mu(P_n) \log \mu(P_n) = \int -\log \mu(\mathcal{P}^n(x)) \, \mathrm{d}\mu(x).$$

The Gauss measure and Lebesgue measure are comparable: there exist constants $c_1, c_2 > 0$ such that

$$c_1 m(\mathcal{P}^n(x)) \leq \mu(\mathcal{P}^n(x)) \leq c_2 m(\mathcal{P}^n(x))$$

for every $x \in (0,1)$. Since G^n is a diffeomorphic bijection from $\mathcal{P}^n(x)$ to $(0,1)$, for every x there exists $y \in \mathcal{P}^n(x)$ such that $\log(\mathcal{P}^n(x)) = -\log |(G^n)'(y)|$. By the bounded distortion property, there exists $C > 0$ such that

$$\frac{|(G^n)'(y)|}{|(G^n)'(x)|} \leq C$$

for every $x, y \in P_n \in \mathcal{P}^n$. follows that

$$-\log(Cc_1) \geq -\log \mu(\mathcal{P}^n(x)) - \log |(G^n)'(x)| \geq -\log(C/c2).$$

Integrating with respect to μ, we obtain

$$-\log(Cc_1) \geq H_\mu(\mathcal{P}^n) - \int \log |(G^n)'(x)| \,\mathrm{d}\mu(x) \geq -\log(C/c2).$$

Finally, since μ is $G-$invariant, we have that

$$\int \log |G'| \,\mathrm{d}\mu = \frac{1}{n} \sum_{j=0}^{n-1} \int \log |G'| \circ G^j \,\mathrm{d}\mu = \frac{1}{n} \int \log |(G^n)'| \,\mathrm{d}\mu$$

and hence

$$h_\mu(G) = h_\mu(G, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{P}^n) = \int \log |G'| \,\mathrm{d}\mu.$$

Inspired in the metric entropy of Kolmogorov and Sinai, Adler, Konheim and McAndrew defined a topological notion of entropy for continuous transformations on compact spaces. Later, Dinaburg and Bowen gave a definition for continuous transformations on metric spaces, and then extended it to non-compact spaces. It turns out that both definition agree for compact metric spaces, but Bowen's definition is more intuitive.

Definition 2.5. Let X be a compact topological space, and α an open cover of X. By compactness, there exists a finite subcover of α. Let $N(\alpha)$ the least number such that α admits a subcover with $N(\alpha)$ elements. We define the *entropy* of the cover α to be the number

$$H(\alpha) = \log N(\alpha).$$

As we did a dynamical version of partition of a measure space, we now do a dynamical version of covers of a topological space. Let $T : X \to X$ be a continuous transformation, for $n \geq 1$ we denote

$$\alpha^n := \alpha \vee T^{-1}(\alpha) \vee ... \vee T^{-n+1}(\alpha).$$

It is possible to prove that the sequence $\{H(\alpha^n)\}$ is subadditive, and hence we have

$$h(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} H(\alpha^n) = \inf_n \frac{1}{n} H(\alpha^n).$$

Finally, the *topological entropy* of T is defined as

$$h(T) = \sup h(T, \alpha),$$

where the supremum is taking over all open covers of X.

It is possible to see that the entropy is a topological invariant. In fact, we have

Theorem 2.6. *If $S : Y \to Y$ is a topological factor of $T : X \to X$ (there exists a surjective continuous mapping $U : X \to Y$ such that $U \circ T = S \circ U$) then $h(S) \leq h(T)$. In particular, if T and S are topologically conjugated, then $h(T) = h(S)$.*

Proof. See [VO15]. ■

We turn now to the definition of the Bowen-Dinaburg topological entropy. As before, we consider a continuous transformation $T : X \to X$ on a metric space, this time, not necessarily compact. Fix a compact set $K \subset X$, given $\varepsilon > 0$ and $n \in \mathbb{N}$ we say that a set $E \subset M$ is $(n, \varepsilon)-$generator of K if

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

where $B(a, e, \varepsilon) = \{x \in M : d(T^i(x), T^i(a)) < \varepsilon \text{ for } i \in \{0, ..., n-1\}\}$ is the *dynamic ball* of center a, length n and radius ε. By the compactness of K, there exists a finite $(n, \varepsilon)-$generator set. Call $g_n(T, \varepsilon, K)$ the least number of elements of an $(n, \varepsilon)-$generator set of K, and define

$$g(T) = \sup_K \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log g_n(T, \varepsilon, K).$$

With the same setting as above, we say that a set $F \subset K$ is $(n, \varepsilon)-$separated if for every $x \in F$, $B(x, n, \varepsilon) \cap F = \{x\}$. Denote by $s_n(T, \varepsilon, K)$ the largest number of elements of an $(n, \varepsilon)-$separating set , and define

$$s(T) = \sup_K \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log s_n(T, \varepsilon, K).$$

Theorem 2.7. *We have $g(T) = s(T)$. If X is compact, then $h(T) = s(T) = g(T)$*

Proof. See [Wal82]. ■

Definition 2.8. For a continuous function on a metric space, we define its *topological entropy* as the number $s(T) = g(T)$. Note that this definition is compatible with the open cover definition for compact spaces.

Analogously to the metric entropy, it is possible to compute the topological entropy using certain open covers of the space.

Theorem 2.9. *If X is a compact metric space and α an open cover of X such that* $\operatorname{diam} \alpha^n \to 0$ *if $n \to \infty$, then $h(T) = h(T, \alpha)$.*

Proof. See [VO15]. ■

For a certain kind of transformations, it is possible to understand the entropy as a measure of growth of the complexity of the system.

We say that a continuous transformation $T : X \to X$ is *expansive* if there exist $\sigma > 1$ and $\rho > 0$ such that for every $p \in X$, $T(B(p, \rho))$ contains a neighborhood of $B(T(p), \rho)$ and

$$d(T(x), T(y)) \geq \sigma d(x, y)$$

for every $x, y \in B(p, \rho)$.

Theorem 2.10. *For every expansive transformation $T : X \to X$, we have*

$$h(T) = \limsup_{n\to\infty} \frac{1}{n} \log \# \operatorname{Fix}(T^n).$$

Proof. See [VO15]. ■

The above expression allows us to understand the entropy as the exponential growth of the periodic points of the map, that is, other way to measure the complexity of the map.

Example (Circle rotations)**.** Applying theorem 2.9 and a cover analogue to the partition used in the calculation of the metric entropy for circle rotations we conclude that $h(R_\alpha) = 0$.

Example (Full shift)**.** The transformation $T : [0, 1) \to [0, 1)$ given by $Tx = mx \mod 1$ is clearly expansive. Note that the equation $T^n x = mx$ has m^n solutions. Applying theorem 2.10 we get

$$h(T) = \limsup_{n\to\infty} \frac{1}{n} \# \operatorname{Fix}(T^n) = \lim_{n\to\infty} \frac{1}{n} \log m^n = \log m.$$

Example (Gauss map)**.** Note that we may restrict the map to invariant subsets with arbitrarily large entropy, namely

$$K_m = \bigcap_{k=0}^{\infty} G^{-k}\left(\left[\frac{1}{m+1}, 1\right)\right).$$

Then $G|_{K_m} : K_m \to K_m$ behaves like a full shift in m symbols and has entropy equal to $\log m$, so $G : [0,1] \to [0,1]$ has infinite entropy.

The notions of metric entropy and topological entropy do not seem to be related at first glance, but they are indeed.

Theorem 2.11 (Variational Principle, first version)**.** *If $T : X \to X$ is a continuous transformation in a compact metric space, then*

$$h(T) = \sup h_\mu(T),$$

where the supremum is taken over all $T-$invariant probability measures μ on m.

Proof. See [Wal82]. ■

This variational principle is a consequence of a more general result due to Ruelle and Walters.

2.2 Topological Pressure

We begin this section introducing some notation: let $T : X \to X$ a continuous function on a compact metric space, and a continuous function $\phi : X \to \mathbb{R}$ which we will call a *potential* of the system. For every $n \in \mathbb{N}$, set $\phi_n(x) = \sum_{k=0}^{n-1} \phi \circ T^k(x)$. For a non-empty set $C \subset M$, we denote

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

Now we proceed to construct the topological pressure: given an open cover α of X, we define

$$P_n(T, \phi, \alpha) = \inf\{\sum_{U \in \gamma} \exp \phi_n(U)\},$$

where the infimum is taken over all finite subcovers γ of α^n. Observe that if we take $\phi \equiv 0$, $P_n(T, \phi, \alpha)$ is actually the number of elements of the cover γ. The sequence

$P_n(T, \phi, \alpha)$ is sub-multiplicative, and hence, $\log P_n(T, \phi, \alpha)$ is sub-additive. By Lemma 2.3, the limit

$$P(T, \phi, \alpha) = \lim_{n \to \infty} \frac{1}{n} \log P_n(T, \phi, \alpha)$$

exists.

Lemma 2.12. *For every sequence $(\alpha_k)_k$ of open covers of X such that $\operatorname{diam} a_k \to 0$, the limit*

$$P(T, \phi) := \lim_{k \to \infty} P(T, \phi, \alpha_k)$$

exists and it is independent of the choice of the sequence of covers.

Definition 2.13. The quantity $P(T, \phi)$ of the previous lemma is the *pressure* of T relative to the potential ϕ.

From the definition, we observe that the pressure is a generalization of the entropy. In fact, if we use the potential $\phi \equiv 0$ we recover the topological entropy of f.

Remark. It is possible to define the pressure using $\underline{\phi}_n(U) = \inf\{\phi_n(x) : x \in U\}$ instead of $\phi_n(U)$ and both definitions agree.

We have an analogue for Theorem 2.9 which allows us to calculate the pressure using certain open covers of the space:

Theorem 2.14. *Let $T : X \to X$ be a continuous map in a compact metric space, $\phi : X \to \mathbb{R}$ a continuous potential and β an open cover of X such that $\operatorname{diam} \beta^k \to 0$. Then $P(T, \phi) = P(T, \phi, \beta)$.*

Proof. See [VO15]. ∎

From the topological nature of the definition, we see that $P(T, \phi)$ is invariant under topological conjugation. Now we list some other properties of the pressure.

Theorem 2.15. *Regard $P(T, \cdot)$ as a function defined on $C^0(X, \mathbb{R})$ with the supremum norm, then*

(a) *$P(T, \cdot)$ is Lipschitz continuous, with Lipschitz constant equal to 1;*

(b) *$P(T, \phi + c) = P(T, \phi) + c$ for every $c \in \mathbb{R}$;*

(c) *if $\phi \leq \psi$ then $P(T, \phi) \leq P(T, \psi)$;*

(d) *$P(T, \cdot)$ is convex;*

(e) *$P(T,\cdot)$ is constant in every cohomology class, that is, $P(T,\phi) = P(T,\phi+u\circ f-u)$ for every $u \in C^0(X,\mathbb{R})$;*

As we stated before, the variational principle for the topological entropy admits a generalization.

Theorem 2.16 (Variational Principle). *Let $T : X \to X$ be a continuous transformation on a compact metric space. Then, for every continuous function $\phi : X \to \mathbb{R}$, we have that*

$$P(T,\phi) = \sup\{h_\mu(T) + \int_X \phi \, \mathrm{d}\mu\},$$

where the supremum is taken over all $T-$ invariant probability measures μ on X.

Proof. See [VO15]. ■

Definition 2.17. A measure attaining μ the supremum of the previous theorem is said to be an *equilibrium state* for the potential ϕ.

It is possible to prove the existence of equilibrium states in a very general setting.

Theorem 2.18. *Suppose $T : X \to X$ is a continuous expansive transformation in a compact metric space X. Then for every continuous potential $\phi : X \to \mathbb{R}$ there exists an equilibrium state.*

Proof. See [VO15]. ■

As in the case of entropy, it is also possible to characterize pressure in terms of periodic points. Thus, pressure is a way to measure the exponential growth of periodic points with an assigned weight for each point.

Definition 2.19. A function $T : X \to X$ is said to be *topologically exact* if for every open set $U \subset X$, there exists $N \in \mathbb{N}$ such that $T^N(U) = X$.

Theorem 2.20. *Let $T : X \to X$ a topologically exact expansive transformation and $\phi : X \to \mathbb{R}$ a Hölder potential. Then,*

$$P(T,\phi) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{x\in \mathrm{Fix}(T^n)} \exp(\phi_n(x)).$$

Proof. See [VO15]. ■

Now we observe that for the case of Subshifts of Finite Type, the pressure can be written in a more gentle form. Suppose $\Sigma \subset \Sigma_A^+$ is a compact set invariant under the action of the shift, and $\varphi : \Sigma \to \mathbb{R}$, the the topological pressure of φ is given by

$$P_\Sigma(\phi) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{i_1,\dots,i_n} \exp \sup_{C_{i_1,\dots,i_n}} \left(\sum_{k=0}^{n-1} \varphi \circ \sigma^k \right),$$

where $C_{i_1,\dots,i_n} = \{(x_1, x_2, \dots) \in \Sigma_A^+ : (x_1, \dots, x_n) = (i_1, \dots, i_n)\}$.

Remark. When there is no possibility of confusion, we write $P(T, \phi) = P(\phi)$.

Definition 2.21. Suppose μ is a σ-invariant probability measure in Σ^+ and $\varphi : \Sigma \to \mathbb{R}^+$ a continuous function. Then μ is called a *Gibbs Measure* if there exist constants $D_1, D_2 > 0$ such that

$$D_1 \leq \frac{\mu(C_{i_1,\dots,i_n})}{\exp(-nP(\varphi) + \sum_{k=0}^{n-1} \varphi(\sigma^k \omega))} \leq D_2$$

for every $n \in \mathbb{N}$ and $\omega \in C_{i_1,\dots,i_n}$.

We finish this section with an example that suggests a relation between dimension theory and thermodynamic formalism. This connection will be formalized in the next section.

Example. Let Σ^+ be the full shift in p symbols, and take numbers $\lambda_1, \dots, \lambda_p \in (0, 1)$. Consider the function $\varphi : \Sigma_A^+ \to \mathbb{R}$ given by

$$\varphi((i_1, i_2, \dots)) = \log \lambda_{i_1}.$$

Then we have

$$\begin{aligned} P(s\varphi) &= \lim_{n\to\infty} \frac{1}{n} \log \sum_{i_1\dots i_n} \exp \left(s \sum_{k=1}^{n} \log \lambda_{i_k} \right) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \left(\sum_{k=1}^{p} \lambda_i^s \right)^n \\ &= \log \sum_{k=1}^{p} \lambda_i^s. \end{aligned}$$

So the equation 1.4 deduced in chapter 1 to calculate the Hausdorff dimension of the geometric construction, is equivalent to the equation

$$P(s\varphi) = 0.$$

This result is part of a much more general phenomena, called the Bowen Equation.

2.3 Bowen Equation

We finish this chapter with the celebrated Bowen Equation, giving a first strong link between Thermodynamic Formalism and Dimension Theory. We state the theorem in a general multidimensional setting, but a sketch of proof in a one-dimensional two-branched case. A very elegant proof of the full general case can be found in [VO15].

Theorem 2.22 (Bowen's equation). *Let $D, D_1, ..., D_N \subset \mathbb{R}^n$ compact convex sets such that $D_i \subset D$ and $D_i \cap D_j$ for $i \neq j$. Set $D_* = D_1 \cup ... \cup D_N$ and suppose that*

$$vol(D \setminus D_*) > 0.$$

Suppose there exists a $\mathcal{C}^1$ function $T : D_* \to D$ such that the restriction to every D_i is an homeomorphism. Set

$$\Lambda = \bigcap_{k=0}^{\infty} T^{-k}(D_*).$$

We make the following hypothesis for T:

1. *T is expansive in D_*,*
2. *$\log |DT|$ is Hölder in D_*,*
3. *T is conformal, this is, $\|DT(x)\|\|DT(x)^{-1}\| = 1$ for every $x \in D_*$*

The, the Hausdorff Dimension of Λ is sn, where s is the unique solution of the equation

$$P(-s \log |\det DT|) = 0.$$

Reduced case proof. Suppose $D_1, D_2 \subset \mathbb{R}$ are two subintervals of the closed interval $D = [0, 1] \subset \mathbb{R}$, T is Hölder with continuous derivative T' such that $|T'(x)| > 1$ in D_*. For $(i_1, ..., i_n) \in \{0, 1\}^n$, set $D(i_1, ..., i_n) = \{x \in D_* : T^k(x) \in D_{i_k}\}$. First, we prove that there exists a unique solution to the equation

$$P(-t \log |T'|) = 0.$$

In fact, it is possible to see that the system is conjugated to a full shift on 2 symbols, so it has topological entropy equal to $\log 2 > 0$, and this is precisely the value of the left hand side of the equation at $t = 0$. On the other side, by the Variational Principle,

$$P(-t\log|T'|) = \sup\{h_\mu(f) + \int(-t\log|T'|)\,\mathrm{d}\mu\}$$
$$\leq \log 2 - t\sup\{\log|T'(x)| : x \in \Lambda\},$$

so letting $t \to \infty$, we get that $P(-t\log|T'|) \to -\infty$. By the Intermediate Value Theorem, we conclude that there exists a root of the equation $P(-t\log|T'|) = 0$. The uniqueness follows from the fact that P is monotonous, so $P(-t\log|T'|)$ is decreasing in t. We call this unique root by t_0.

Note that the restriction of T^n to each set $D(i_1, ..., i_n)$ is a bijection to $[0,1]$, and denote by $T_{i_1,...,i_n}$ its inverse. From the definition, we have that $T_{i_1,...,i_n} = T_{i_1} \circ .... \circ T_{i_n}$. Since T_1 and T_2 are defined in compact sets, there exist constants $C_1, C_2 > 0$ such that $C_1 \leq |T_1'(x)|, |T_2'(y)| \leq C_2$ for every $x \in D_1, y \in D_2$. By the Mean Value Theorem, we obtain

$$C_1|x-y| \leq |T_i(x) - T_i(y)| \leq C_2|x-y|$$

for every $x, y \in \Lambda$ and $i \in \{1,2\}$. Applying this several times we obtain

$$\operatorname{diam} D(i_1, ...i_n) = \operatorname{diam} T_{i_1,...,i_n}(D) \leq C_2^n.$$

Recall that we assumed that $\log|T'|$ is Hölder, and so is $\phi(x) = -\log|T'|$: there exist constants $a, \alpha > 0$ such that

$$|\phi(x) - \phi(y)| \leq a|x-y|^\alpha.$$

Then, for $x, y \in D(i_1, ...i_n)$ we have that

$$|\phi(T^j x) - \phi(T^j y)| \leq a|T^j x - T^j y|^\alpha \leq a(\operatorname{diam} D(i_{j+1}, ..., i_n))^\alpha \leq aC_2^{\alpha(n-j)}.$$

Therefore,

$$|\phi_n(x) - \phi_n(y)| = |\sum_{i=0}^{n-1}\phi(T^i x) - \sum_{i=0}^{n-1}\phi(T^i y)| \leq \sum_{i=0}^{n-1} aC_2^{\alpha(n-i)} \leq a\frac{C_2^\alpha}{1-C_2^\alpha} := b.$$

Note that this argument is valid for every ϕ Hölder continuous, and can be written equivalently as

$$\exp(-b) \leq \frac{\exp(\phi_n(x))}{\exp(\phi_n(y))}\exp(b)$$

for every $(i_1, ..., i_n) \in \{0,1\}^n$ and $x, y \in D(i_1, ...i_n)$. For the particular case of $\phi = -\log|T'|$ the previous inequality takes the form

$$\exp(-b) \leq \frac{|(T^n)'(x)|}{|(T^n)'(y)|} \leq \exp(b)$$

for every $(i_1, ..., i_n) \in \{0,1\}^n$ and $x, y \in D(i_1, ...i_n)$. This is a bounded distortion property. By the Mean Value Theorem, there exists $z \in D(i_1, ..., i_n)$ such that

$$\operatorname{diam}[0,1] = \operatorname{diam}(D(i_1, ..., i_n)) \cdot |(T^n)'(z)|.$$

Combining this conclusion with the bounded distortion property, we obtain

$$\exp(-b) \leq (\operatorname{diam} D(i_1, ...i_n)) \cdot |(T^n)'(x)| \leq \exp(b)$$

for every $x \in D(i_1, ..., i_n)$. Using this estimate, we can write the pressure function for T as following

$$\begin{aligned} P(-t\log|T'|) &= \lim_{n\to\infty} \frac{1}{n} \log \sum_{x \in \operatorname{Fix}(T^n)} \exp(S_n(-t\log|T'(x)|)) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \sum_{i_1,...,i_n} (\operatorname{diam} D(i_1, ..., i_n))^t. \end{aligned}$$

This equality can be rewritten as

$$\sum_{i_1,...,i_n} (\operatorname{diam} D(i_1, ..., i_n))^t \asymp \exp(nP(-t\log|T'|)).$$

Now, for t_0 there exists a Gibbs measure μ_{t_0} associated to the potential $-t\log|T'|$ so there exists $C > 0$ such that

$$\frac{1}{C} \leq \frac{\mu_{t_0}(D(i_1, ...i_n))}{\exp(-nP(-t_0\log|T'|) + S_n(-t\log|T'|))} = \frac{\mu_{t_0}(D(i_1, ...i_n))}{|(T^n)'(x)|^{t_0}} \leq C$$

for every $n \in \mathbb{N}$ and $x \in D(i_1, ..., i_n)$. Using the estimation previously proved, we get

$$\frac{1}{A} \leq \frac{\mu_{t_0}(D(i_1, ...i_n))}{(\operatorname{diam} D(i_1, ..., i_n))^{t_0}} \leq A$$

for every $(i_1, ..., i_n) \in \{0,1\}^n$ and for some $A > 0$. As in chapter 1, we construct a Moran Cover of the set Λ: for every $r \in (0,1)$ and $\omega = (i_1, i_2, ...) \in \{0,1\}^{\mathbb{N}}$ there exists a unique integer $n(\omega, r)$ such that

$$\operatorname{diam} D(i_1, ..., i_{n(\omega,r)}) < r \leq \operatorname{diam} D(i_1, ..., i_{n(\omega,r)-1}).$$

For fixed r, the set $\{D(\omega, r) = D_{i_1,\ldots,i_{n(\omega,r)}}\}$ forms a cover by pairwise disjoint sets. As in the geometric construction, we have the following estimation for the diameter of $D(\omega, r)$

$$\frac{r}{c} \leq \operatorname{diam} D(\omega, r) < r$$

where $c = (\min\{\operatorname{diam} D_1, \operatorname{diam} D_2\})^{-1}$, so we get that for any interval I with $\operatorname{diam} I = r$, there is at most c of the sets $D(\omega, r)$ that intersect I. Then

$$\mu_{t_0}(I) \leq \sum_{D(\omega,r)\cap I\neq\emptyset} \mu_{t_0}(D(\omega, r) \leq \sum_{D(\omega,r)\cap I\neq\emptyset} A(\operatorname{diam} D(\omega, r))^{t_0} \leq Acr^{t_0} = Ac(\operatorname{diam} I)^{t_0}.$$

Now, given an arbitrary cover $\mathcal{U}$ of Λ with $\operatorname{diam}\mathcal{U} < r$, every set $U \in \mathcal{U}$ is contained in an interval I_U with $\operatorname{diam} U = \operatorname{diam} I_U$, so applying the previous calculations, we get

$$0 < \mu_{t_0}(\Lambda) = \sum_{U\in\mathcal{U}} \mu_{t_0}(U) \leq \sum_{U\in\mathcal{U}} \mu_{t_0}(I_U) \leq \sum_{U\in\mathcal{U}} Ac(\operatorname{diam} U)^{t_0}$$

so $m(\Lambda, t_0) > 0$ and then $\dim_H \Lambda \geq t_0$. The upper bound follows essentially the same technique used in the geometric construction of the chapter 1, and we omit its calculation. ■

In the general case we have to deal with the contraction of the sets D_i in more than one direction, contrary to the one-dimensional case. Here is when the conformality hypothesis takes relevance.

Remark. Note that this construction actually recovers the calculation done in chapter 1.

2.4 Pressure function

We finish this chapter by introducing the *Pressure function*. In the last section, we noticed that the equation

$$P(-t \log |T'|) = 0$$

establishes a connection between thermodynamic formalism and dimension theory. Then the question about the regularity of the function $P(t) = P(-t \log |T'|)$ arises naturally. From the properties of the topological pressure, we deduced that $P(t)$ is continuous, convex and that $\lim_{t\to\infty} P(t) = -\infty$. In the next chapter, we make use of Functional Analysis machinery to prove that in certain cases, the function $P(t)$ is actually analytic.

Chapter 3

Transfer Operator

This chapter is based on [VO15], [Bal00], [Sar99] and [PP90].

3.1 Transfer Operator

In chapter 1 we saw that some dynamical properties of a system are reflected into spectral properties of the associated operators. For instance, by Theorem 1.11 the ergodicity of a measure μ turns out to be equivalent to the simplicity of the eigenvalue 1 for the Koopman operator $U_T : L^2(\mu) \to L^2(\mu)$ given by $U_T(\phi) = \phi \circ T$. In this chapter, we explore more in depth the connection between the dynamical properties of a system, and the spectral properties of some operators. Eventually, we will be able to deduce the analiticity of the Pressure Function by applying perturbation theory to Ruelle Operator.

We begin this section introducing the basic object of study of this chapter.

Definition 3.1. Suppose $(X, \mathcal{B}, \mu)$ is a probability space and $T : X \to X$ a measurable transformation. We say that T is *non-singular* if $\mu(T^{-1}A) = 0$ if and only if $\mu(A) = 0$, for $A \in \mathcal{B}$.

Suppose $T : X \to X$ is a non-singular measurable transformation, and $f \in L^1(\mu)$ the density with respect to μ of a probability measureμ_f, that is, μ_f is the probability measure defined by

$$\mu_f(A) = \int_A f \, \mathrm{d}\mu.$$

We want to describe distribution of the probability measure μ_f after the one iteration of the dynamic T. We have then for every measurable set E

$$\begin{aligned}\mu_f(T^{-1}E) &= \int_X (1_E \circ T) f \,\mathrm{d}\mu \\ &= \int_X 1_E \,\mathrm{d}(\mu_f \circ T^{-1}) \\ &= \int_E \left(\frac{\mathrm{d}\mu_f \circ T^{-1}}{\mathrm{d}\mu} \right) \mathrm{d}\mu\end{aligned}$$

where 1_E is the indicator function of E. The existence of the Radon-Nikodym derivative is guaranteed by the non-singularity of T. Hence, Radon-Nikodym theorem implies that the map

$$\begin{aligned}\hat{T} : L^1(\mu) &\to L^1(\mu) \\ f &\mapsto \frac{\mathrm{d}\mu_f \circ T^{-1}}{\mathrm{d}\mu}\end{aligned}$$

is well defined.

Definition 3.2. The map $\hat{T}$ is the *transfer operator* for the system $(X, \mathcal{B}, \mu, T)$.

Note that we do not require μ to be $T-$invariant. Indeed, a measure μ_f is $T-$invariant if and only the density f is a fixed point of the transfer operator $\hat{T}$, and consequently, μ is $T-$invariant if and only if the function constant equal to 1 is a fixed point of $\hat{T}$.

The next proposition characterizes the transfer operator as a formal adjoint of the Koopman operator U_T acting on the space of essentially bounded functions.

Proposition 3.3. *For every $f \in L^1(\mu)$ and $\varphi \in L^\infty(\mu)$, we have that*

$$\int_X \varphi \cdot (\hat{T} f) \,\mathrm{d}\mu = \int_X (\varphi \circ T) f \,\mathrm{d}\mu.$$

Moreover, $\hat{T} f$ is the unique element of $L^1(\mu)$ satisfying the previous equation for every $\varphi \in L^\infty$.

Proof. We first check that the identity holds for every $\varphi \in L^\infty$:

$$\begin{aligned}\int_X \varphi \cdot (\hat{T} f) \,\mathrm{d}\mu &= \int_X \varphi \cdot \frac{d(\mu_f \circ T^{-1})}{d\mu} \,\mathrm{d}\mu \\ &= \int_X \varphi \,\mathrm{d}(\mu_f \circ T^{-1}) \\ &= \int_X (\varphi \circ T) \,\mathrm{d}\mu_f \\ &= \int_X (\varphi \circ T) f \,\mathrm{d}\mu.\end{aligned}$$

Suppose now that there exist $h_1, h_2 \in L^1$ such that

$$\int_X \varphi h_i \, \mathrm{d}\mu = \int_X (\varphi \circ T) h_i \, \mathrm{d}\mu$$

for every $\varphi \in L^\infty$, then, taking $\varphi = \mathrm{sgn}(h_1 - h_2)$, then

$$\begin{aligned}
\int_X |h_1 - h_2| \, \mathrm{d}\mu &= \int_X \varphi(h_1 - h_2) \, \mathrm{d}\mu \\
&= \int_X \varphi h_1 \, \mathrm{d}\mu - \int_X \varphi h_2 \, \mathrm{d}\mu \\
&= \int_X \varphi \circ T \, \mathrm{d}\mu - \int_X \varphi \circ T \, \mathrm{d}\mu = 0
\end{aligned}$$

so we conclude that $h_1 = h_2$ in L^1 as desired. ■

From the previous proposition, it is easy to show that $\hat{T^n} f = (\hat{T})^n f$. Now we prove some useful properties of $\hat{T}$:

Proposition 3.4. *$\hat{T}$ is a positive linear bounded operator with norm equal to 1 acting on $L^1(\mu)$. For every $g \in L^\infty$, we have that $\hat{T}[(g \circ T) f] = g\hat{T} f$ almost everywhere.*

Proof. This is a consequence of the adjunction property from the previous proposition. In fact, it is equivalent to prove that

$$\int_X \varphi \cdot (g\hat{T} f) \, \mathrm{d}\mu = \int_X (\varphi \circ T)[(g \circ T) \cdot f] \, \mathrm{d}\mu$$

which is a restatement of the adjunction relation for $\hat{T} f$:

$$\int_X (\varphi \cdot g)(\hat{T} f) \, \mathrm{d}\mu = \int_X ((\varphi \cdot g) \circ T) f \, \mathrm{d}\mu.$$

■

Example (Full shift)**.** A direct computation on the adjunction relation for $\hat{T}$ yields the following formula for $\hat{T}$ when $T : [0, 1) \to [0, 1)$ is given by $Tx = 2x \mod 1$:

$$\hat{T} f(x) = \frac{1}{2} \left[f\left(\frac{x}{2}\right) + f\left(\frac{x+1}{2}\right) \right].$$

Example (Gauss map)**.** For the Gauss map $T : (0, 1] \to (0, 1]$, $Tx = \frac{1}{x} \mod 1$, the transfer operator takes the following form:

$$\hat{T} f(x) = \sum_{n=1}^{\infty} \left(\frac{1}{x+n}\right)^2 f\left(\frac{1}{x+n}\right).$$

It is possible to generalize the previous examples to a very ample class of functions.

Definition 3.5. Let $T : X \to X$ a measurable transfomation. We say that T is *locally invertible* there exists a countable measurable sets cover $\{U_k : k \in \mathbb{N}\}$ such that T restricted to each U_k is a bijection. We call the sets U_k an *injectivity domain* for T.

Now we extend the notion of Jacobian by analogy with the change of variables formula.

Definition 3.6. Let μ be a probability measure in X. We say that a function $\xi : X \to [0, \infty)$ is a *Jacobian* of T with respect to μ if the restriction of ξ to any injectivity domain U_k is integrable and satisfies

$$\mu(T(A)) = \int_A \xi \, \mathrm{d}\mu.$$

Example (Diffeomorphisms)**.** Note that for local diffeomorphisms, $|\det DT(x)|$ is Jacobian for T.

Example (Full shift)**.** Consider the shift space in m symbols (Σ, σ) equipped with a Bernoulli probability μ with probability vector $p = (p_1, ..., p_m)$. Note that the restriction of σ to each cylinder defines a bijection to its image. Also, noting that $\mu(\sigma(I(a_1, ..., a_n))) = p_{a_2}...p_{a_n} = \dfrac{1}{p_{a_1}}\mu(I(a_1, ..., a_n))$ we conclude that the function $\xi : \Sigma \to \mathbb{R}$ given by

$$\xi((x_n)_n) = \frac{1}{p_{x_1}}$$

is a Jacobian for σ with respect to μ.

Remark. Note that the definition implies that if T has a Jacobian, then it is a non-singular transformation. The converse is also true, as stated by the following theorem.

Theorem 3.7. *Let $T : X \to X$ be a locally invertible transformation, μ a Borel probability measure on X, non-singular with respect to T. Then there exists a Jacobian for T with respect to μ, and it is unique a.e.*

Proof. The idea is that T behaves well in each injectivity domain U_k so we can produce a Jacobian for each element of the cover and then paste them to get a Jacobian for the whole space. If $\{U_k : k \in \mathbb{N}\}$ is a cover by injectivity domains, define $\mathcal{P} = \{P_k : k \in \mathbb{N}\}$ with $P_1 = U_1$ and $P_k = U_k \setminus (U_1 \cup ... \cup U_{k-1})$ for $k \geq 2$. Then $\mathcal{P}$ is a measurable partition by injectivity domains. For each $P_k \in \mathcal{P}$ define a measure μ_k by $\mu_k(A) = \mu(T(A))$, which is absolutely continuous with respect to the restriction of μ to P_k. By Radon-Nykodim theorem, there exists $\xi_k \in L^1$ such that

$$\mu(T(A)) = \mu_k(A) = \int_A \xi_k \, \mathrm{d}\mu$$

for every measurable set $A \subset P_k$. Define $\xi : X \to [0, \infty)$ by ξ_k when restricted to P_k, and the property holds. Now suppose η is a Jacobian for T, different to ξ in a positive measure set B. Without lost of generality, we may suppose $\xi < \eta$ in B and $B \subset U_k$ for some k. Then

$$\mu(T(B)) = \int_B \xi \, \mathrm{d}\mu < \int_B \eta \, \mathrm{d}\mu = \mu(T(B))$$

a contradiction. ∎

Remark. Now that we know that the Jacobian exists and it is unique a.e., we may speak of *the* Jacobian of T, and note it by $J_\mu T$.

The following theorem is essential to obtain a general formula for the transfer operator associated to locally invertible non-singular transformations.

Theorem 3.8. *Let $T : X \to X$ a locally invertible transformation, non-singular with respect to the probability measure μ. Then, for every $\psi \in L^\infty$ we have that*

$$\int_X \psi \, \mathrm{d}\mu = \int_X \sum_{z \in T^{-1}(x)} \frac{\psi}{J_\mu T}(z) \, \mathrm{d}\mu(x).$$

Proof. We need several lemmas to get the result.

Lemma 3.9. *For every injectivity domain $A \subset X$ and measurable $\phi : T(A) \to \mathbb{R}$ such the integrals of the following equation are defined, we have that*

$$\int_{T(A)} \phi \, \mathrm{d}\mu = \int_A (\phi \circ T) J_\mu T \, \mathrm{d}\mu.$$

Proof. First suppose $\phi = 1_E$ for some measurable set $E \subset T(A)$, so there exists $B \subset A$. such that $T(B) = E$. Then

$$\mu(T(B)) = \mu(T(A \cap B)) = \int_{A \cap B} J_\mu T \, \mathrm{d}\mu = \int_A 1_B \cdot J_\mu T \, \mathrm{d}\mu = \int_A (1_E \circ T) J_\mu T \, \mathrm{d}\mu.$$

On the other side,

$$\mu(T(B)) = \int_X 1_{A \cap B} \, \mathrm{d}\mu = \int_{T(A)} 1_{T(B)} \, \mathrm{d}\mu = \int_{T(A)} 1_E \, \mathrm{d}\mu$$

so the equality holds for indicator functions. By an approximation argument, the equality follows for arbitrary functions in L^∞. ∎

Lemma 3.10. *Under the same hypothesis as the previous lemma, we have that*

$$\int_A \psi \, \mathrm{d}\mu = \int_{T(A)} \left(\frac{\psi}{J_\mu T}\right) \circ (T|_A)^{-1} \, \mathrm{d}\mu.$$

Proof. Taking $\phi = \left(\frac{\psi}{J_\mu T}\right) \circ (T|_A)^{-1}$ in the previous lemma, we obtain

$$\int_{T(A)} \left(\frac{\psi}{J_\mu T}\right) \circ (T|_A)^{-1} \, \mathrm{d}\mu = \int_A (\phi \circ T) J_\mu T \, \mathrm{d}\mu = \int_A \psi \, \mathrm{d}\mu.$$

■

Lemma 3.11. *For every injectivity domain P and bounded measurable ϕ function we have*

$$\int_{P_0} \sum_{z \in T^{-1}(y)} \frac{\phi}{J_\mu T}(z) \, \mathrm{d}y = \int_{T^{-1} P_0} \phi \ \mathrm{d}\mu.$$

Proof. Note that $T|_{P_0} : P_0 \to T(P_0)$ is a bijection and hence we have

$$\int_{P_0} \sum_{z \in T^{-1}(y)} \frac{\phi}{J_\mu T}(z) \, \mathrm{d}y = \sum_{P \in \mathcal{P}} \int_{T(P) \cap P_0} \left(\frac{\phi}{J_\mu T}\right) \circ (T|_P)^{-1} \, \mathrm{d}\mu(y).$$

By the previous lemma, the later quantity is equal to

$$\sum_{P \in \mathcal{P}} \int_{P \cap T^{-1} P_0} \phi \, \mathrm{d}\mu = \int_{T^{-1} P_0} \phi \, \mathrm{d}\mu.$$

■

Now we are ready to prove the theorem: summing over $P_0 \in \mathcal{P}$ on the last lemma, we obtain

$$\sum_{P_0 \in \mathcal{P}} \int_{P_0} \sum_{z \in T^{-1}(y)} \frac{\phi}{J_\mu T}(z) \, \mathrm{d}y = \sum_{P_0 \in \mathcal{P}} \int_{T^{-1} P_0} \phi \ \mathrm{d}\mu$$
$$\int_X \sum_{z \in T^{-1}(y)} \frac{\phi}{J_\mu T}(z) \, \mathrm{d}y = \int_X \phi \ \mathrm{d}\mu$$

were we conclude the theorem. ■

Corollary 3.12. *Let $T : X \to X$ locally invertible and μ a non-singular probability measure. Then μ is $T-$invariant if and only if*

$$\sum_{z \in T^{-1}(x)} \frac{1}{J_\mu T(z)} = 1$$

for almost every $x \in X$.

With the previous theorem we are able to prove a general formula for the transfer operator.

Theorem 3.13 (Transfer operator formula)**.** *Let $T : X \to X$ be a locally invertible transformation, non-singular with respect to the probability measure μ. Then the transfer operator $\hat{T}$ satisfies*

$$\hat{T}\psi(x) = \sum_{z \in T^{-1}(x)} \frac{\psi}{J_\mu T}(z)$$

for every $\psi \in L^1$.

Proof. By Theorem 3.8, we have that

$$\int_X \psi \, \mathrm{d}\mu = \int_X \sum_{z \in T^{-1}(x)} \frac{\psi}{J_\mu T}(z) \, \mathrm{d}\mu(x).$$

Changing $\psi \mapsto \psi \cdot (1_A \circ T)$, we get

$$\begin{aligned} \int_X (1_A \circ T) \cdot \psi \ \mathrm{d}\mu &= \int_X \sum_{z \in T^{-1}(x)} \frac{\psi(z) \cdot (1_A \circ T(z))}{J_\mu T(z)} \, \mathrm{d}\mu(x) \\ &= \int_X 1_A \sum_{z \in T^{-1}x} \frac{\phi(z)}{J_\mu T(z)} \, \mathrm{d}\mu(x). \end{aligned}$$

By linearity and approximation by simple functions, we obtain

$$\int_X (\phi \circ T) \cdot \psi \ \mathrm{d}\mu = \int_X \phi \sum_{z \in T^{-1}x} \frac{\psi(z)}{J_\mu T(z)} \ \mathrm{d}\mu.$$

By Proposition 3.3, we conclude that

$$\hat{T}\psi(x) = \sum_{z \in T^{-1}x} \frac{\psi(z)}{J_\mu T(z)}.$$

■

Many dynamical properties can be translated into the asymptotic behavior of the transfer operator.

Proposition 3.14. *Suppose that $T : X \to X$ is a non-singular transformation, then*

(a) If $\hat{T}^n f \to h \int f \, \mathrm{d}\mu$ weakly in $L^1(\mu)$ for some non-negative $0 \neq f \in L^1(\mu)$, then T has an absolutely continuous invariant probability with density h.

(b) *If $\hat{T}^n f \to \int f \,\mathrm{d}\mu$ in $L^1(\mu)$ for all $f \in L^1(\mu)$, then T is a mixing probability mixing map.*

(c) *If $\hat{T}^n f \to \int f \,\mathrm{d}\mu$ in the $L^1(\mu)$ norm, then*

$$|\operatorname{Cor}(f, \varphi \circ T^n)| := \Big| \int_X f \cdot \varphi \circ T^n \,\mathrm{d}\mu - \int_X f \,\mathrm{d}\mu \int_X \varphi \,\mathrm{d}\mu \Big| \leq \|\hat{T}^n f - \int f \,\mathrm{d}\mu\|_1 \|\varphi\|_\infty$$

Proof. (a) We may assume without lost of generality that $\int f \,\mathrm{d}\mu = 1$. Then, the hypothesis becomes $\hat{T}^n f \to h$ weakly. Hence, we have for every $\varphi \in L^\infty$

$$\begin{aligned}\int_X \varphi h \,\mathrm{d}\mu &= \lim_{n\to\infty} \int_X \varphi \hat{T}^{n+1} f \,\mathrm{d}\mu \\ &= \lim_{n\to\infty} \int_X (\varphi \circ T) \hat{T}^n f \,\mathrm{d}\mu \\ &= \int_X (\varphi \circ T) h \,\mathrm{d}\mu \\ &= \int_X \varphi \hat{T} h \,\mathrm{d}\mu\end{aligned}$$

so by the adjuntion property of $\hat{T}$ we conclude that $\hat{T}h = h$. If we define now the measure $\mu_h = h\mu$ the previous calculation implies that μ_h is invariant:

$$\mu_h(T^{-1}E) = \int_{T^{-1}E} h \,\mathrm{d}\mu = \int_X (1_E \circ T) h \,\mathrm{d}\mu = \int_X 1_E(\hat{T}h) \,\mathrm{d}\mu = \int_E h \,\mathrm{d}\mu = \mu_h(E)$$

for every measurable set E.

(b) By the previous part, μ is $T-$invariant. Now, for every measurable sets E, F we have that

$$\mu(E \cap T^{-n}F) = \int_X 1_E(1_F \circ T^n) \,\mathrm{d}\mu = \int_X (\hat{T}^n 1_E) 1_F \,\mathrm{d}\mu \to \int_X \left(\int_X 1_E \,\mathrm{d}\mu \right) 1_F \,\mathrm{d}\mu = \mu(E)\mu(F)$$

so μ is mixing.

(c) Recall that $\hat{T}^n$ satisfies

$$\int_X \varphi \cdot \hat{T}^n f \,\mathrm{d}\mu = \int_X f \cdot \varphi \circ T^n \,\mathrm{d}\mu.$$

Thus, we have

$$\begin{aligned}|\operatorname{Cor}(f, \varphi \circ T^n)| &= \Big| \int_X \varphi \cdot \hat{T}^n f \,\mathrm{d}\mu - \int_X f \,\mathrm{d}\mu \int \varphi \,\mathrm{d}\mu \Big| \\ &= \Big| \int_X \varphi \cdot (\hat{T}^n f - \int_X f \,\mathrm{d}\mu) d\mu \Big| \\ &\leq \|\varphi\|_\infty \|\hat{T}^n f - \int_X f \,\mathrm{d}\mu\|_1.\end{aligned}$$

■

Note that the decay of correlations depends on the speed of convergence of $\hat{T}^n f$. We will study this phenomena in more detail. The spectra of the transfer operator is closely related with the dynamics of the mapping T, in fact we have:

Proposition 3.15. *Suppose that $T : X \to X$ is a non-singular transformation, then*

(a) All the eigenvalues of the transfer operator have modulus less or equal to 1. The non-negative functions $h \in L^1(\mu)$ such that $\hat{T}h = h$ and $\int hd\mu = 1$ are precisely the densities of the absolutely continuous invariant probabilities with respect to μ.

(b) If 1 is a simple eigenvalue of $\hat{T}$, then the corresponding associated eigenfunction is the density of an ergodic probability measure.

(c) If 1 is a simple eigenvalue of $\hat{T}$ and all the rest of the spectrum is contained in a disk of modulus strictly smaller than 1, then the corresponding associated eigenfunction is the density of a mixing measure.

(d) If T is a mixing probability measure preserving, then $\hat{T}$ has exactly one eigenvalue on the unit circle, equal to 1, and this eigenvalue is simple.

Proof. (a) The first assertion follows from the fact that $\|\hat{T}\| = 1$. If $h \in L^1$ is such that $\hat{T}h = h$ and $\int h \, d\mu = 1$, define the measure $\mu_h(A) = \int_A h \, d\mu$ in X. The second hypothesis ensures that μ_h is a probability measure, and the first hypothesis, together with the adjunction relation imply that for every $\phi \in L^\infty(\mu_h)$

$$\int_X (\phi \circ T)h \ \mathrm{d}\mu = \int_X \phi h \ \mathrm{d}\mu$$
$$\int_X (\phi \circ T) \ \mathrm{d}\mu_h = \int_X \phi \ \mathrm{d}\mu_h.$$

By Theorem 1.4 from Chapter 1, we conclude that μ_h is an invariant measure for T. The previous equations imply the converse.

(b) By (a) $\mu_h = h\mu$ defines a $T-$invariant probability measure. Since $\hat{T}h = h$ we have that

$$\int_X \varphi h \, \mathrm{d}\mu = \int_X \varphi \hat{T} h \, \mathrm{d}\mu = \int_X \varphi \circ T h \, \mathrm{d}\mu$$

for every $\varphi \in L^\infty$. By density, we conclude the same result for arbitrary functions $\varphi \in L^1$. Now for $\varphi \in L^\infty$, we have that

$$\int_X \varphi \hat{T}(fh) \, \mathrm{d}\mu = \int_X (\varphi \circ T) f h \, \mathrm{d}\mu = \int_X (\varphi \circ T)(f \circ T) h \, \mathrm{d}\mu = \int_X \varphi f h \, \mathrm{d}\mu.$$

Since φ was arbitrary, we conclude that $\hat{T}(fh) = fh$. By the uniqueness property of the transfer operator we conclude that $\hat{T}fh = fh$. By the simplicity of the eigenvalue 1, we conclude that there exists $\lambda \in \mathbb{R}$ such that $fh = \lambda h$. Note that $\mu_h(\{x : h(x) = 0\}) = 0$ so $f = \lambda\ \mu_h-$a.e., and hence μ_f is ergodic.

(c) Again, suppose $\hat{T}h = h$ and define μ_h as above. Now, define the operators $\hat{P} : L^1 \to L^1$ by

$$\hat{P}f = \mu(f)h,$$

which is a projection on the subspace $\mathbb{C}h$ and

$$\hat{N} = \hat{T} - \hat{P}.$$

Note that the eigenvalues of $\hat{N}$ are strictly less than 1, and that $\hat{N}\hat{P} = \hat{P}\hat{N} = 0$ hence $\hat{N}^n = \hat{T}^n - \hat{P}^n$. For functions $f, g \in L^1$, their correlation is given by

$$\begin{aligned}
\Big| \int_X f \cdot g \circ T^n \,\mathrm{d}\mu_h - \int_X f \,\mathrm{d}\mu_h \int_X g \,\mathrm{d}\mu_h \Big| &= \Big| \int_X f \cdot g \circ T^n \cdot h \,\mathrm{d}\mu - \int_X fh \,\mathrm{d}\mu \int_X gh \,\mathrm{d}\mu \Big| \\
&= \Big| \int_X \hat{T}^n(fh) g \,\mathrm{d}\mu - \int_X \hat{P}^n(fh) g \,\mathrm{d}\mu \Big| \\
&= \Big| \int_X \hat{N}^n(fh) g \,\mathrm{d}\mu \Big| \\
&\leq \|g\|_1 \int_X |\hat{N}^n(fh)| \,\mathrm{d}\mu.
\end{aligned}$$

By Gelfand's Formula, for $\varepsilon = (1 - \rho(\hat{N}))/2 > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\|\hat{N}^n\| < (\rho(\hat{N}) + \varepsilon)^n$$

for every $n \geq n_0$. Then, we obtain that

$$\begin{aligned}
\Big| \int_X f \cdot g \circ T^n \,\mathrm{d}\mu_h - \int_X f \,\mathrm{d}\mu_h \int_X g \,\mathrm{d}\mu_h \Big| &\leq \|g\|_\infty \int_X |\hat{N}(fh)| \,\mathrm{d}\mu \\
&\leq \|g\|_\infty \|f\|_1 \|\hat{N}^n\| \\
&\leq \|g\|_\infty \|f\|_1 (\varepsilon + \rho(\hat{N}))^n.
\end{aligned}$$

Since $\varepsilon + \rho(\hat{N}) < 1$, we conclude that μ_h is mixing by Proposition 1.23.

(d) Suppose there exists $\lambda \neq 1$ and φ_λ with $|\lambda| = 1$ and $\hat{T}\varphi_\lambda = \lambda\varphi_\lambda$. For simplicity, we suppose that $h > 0$. Note that $L^1(\mu)$ can be decomposed as a direct sum by $L^1 = V \oplus \mathbb{C}h$ with $V = \{f \in L^1 : \int_X f d\mu = 0\}$. For this, note that every $f \in L^1$

can be written as

$$f = (f - h\int_X f\,\mathrm{d}\mu) + h\int_X f\,\mathrm{d}\mu.$$

Thus, we may suppose $\int_X \varphi_\lambda d\mu = 0$. Then, for every $\psi \in L^\infty$ we have that

$$\begin{aligned}\int_X (\psi\circ T^n)\left(\frac{\varphi_\lambda}{h}\right)\mathrm{d}\mu_h &= \int_X (\psi\circ T^n)\frac{\varphi_\lambda}{h}h\,\mathrm{d}\mu\\ &= \int_X (\psi\circ T^n)\varphi_\lambda\,\mathrm{d}\mu\\ &= \int_X \psi(\hat{T}^n\varphi_\lambda)\,\mathrm{d}\mu\\ &= \lambda^n\int_X \psi\varphi_\lambda\,\mathrm{d}\mu\end{aligned}$$

and this integral is non-vanishing for well chosen φ. If μ_h is mixing, then

$$\begin{aligned}\int_X (\psi\circ T^n)\left(\frac{\varphi_\lambda}{h}\right)\mathrm{d}\mu_h &\to \int_X \psi\,\mathrm{d}\mu_h\int_X\left(\frac{\varphi_\lambda}{h}\right)\mathrm{d}\mu_h\\ &= \int_X \psi h\,\mathrm{d}\mu\int_X \varphi_\lambda\,\mathrm{d}\mu = 0\end{aligned}$$

but this is impossible since the first integral has constant non-zero absolute value. Then, $|\lambda| < 1$ as desired.

■

The rate of the decay of correlations strongly depends on the regularity of the functions, and some properties can be deduced by studying the spectra of the transfer operator. In general, the transfer operator is not compact, so the key is to find a *good* space in which it acts nicely. We start by considering the special case of the doubling map.

Let $\mathrm{Lip} \subset L^1([0,1])$ the vector space of Lipschitz functions in $[0,1]$ and

$$|f|_{\mathrm{Lip}} = \sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|}.$$

We see that $|\cdot|_{\mathrm{Lip}}$ is a seminorm in Lip, but it fails to be a norm (it vanishes for all constant functions). This problem can be fixed by adding the supremum norm

$$\|f\|_{\mathrm{Lip}} = \|f\|_\infty + |f|_{\mathrm{Lip}}$$

is in fact a norm on Lip. It is possible to prove that $(\mathrm{Lip}, \|\cdot\|_{\mathrm{Lip}})$ is a Banach space. The next lemma gives a fundamental relation between the transfer operator and Lip

Lemma 3.16. *For every $f \in \mathrm{Lip}$, we have $|\hat{T}f|_{\mathrm{Lip}} \leq \frac{1}{2}|f|_{\mathrm{Lip}}$.*

Proof. In fact, using the explicit formula for the transfer operator we have that

$$
\begin{aligned}
|\hat{T}f|_{\mathrm{Lip}} &= \sup_{x\neq y} \frac{|\hat{T}f(x)-\hat{T}f(y)|}{|x-y|} \\
&= \sup_{x\neq y} \frac{1}{2}\frac{|f(\frac{x}{2})+f(\frac{x+1}{2})-f(\frac{y}{2})-f(\frac{y+1}{2})|}{|x-y|} \\
&\leq \frac{1}{4}\left[\sup_{x\neq y}\frac{|f(\frac{x}{2})-f(\frac{y}{2})|}{|x-y|}+\sup_{x\neq y}\frac{|f(\frac{x+1}{2})-f(\frac{y+1}{2})|}{|x-y|}\right] \\
&\leq \frac{1}{2}|f|_{\mathrm{Lip}}.
\end{aligned}
$$

■

Note that the space Lip can be decomposed as $\mathrm{Lip} = \mathbf{1}\mathbb{C}\oplus H$ with $H = \{f \in \mathrm{Lip} \mid \int f dx = 0\}$ just by writing $f = \mathbf{1}\int f dx + (f - \mathbf{1}\int f dx)$ where $\mathbf{1}$ is the function constant equal to 1. This decomposition is invariant under the action of the transfer operator, moreover, $\hat{T}\mathbf{1} = \mathbf{1}$ reflecting the fact that Lebesgue measure is already an invariant measure for the doubling map. For $h \in H$ we have the estimate on the supremum norm: $\|f\|_\infty \leq |f|_{\mathrm{Lip}}$.

Now, using the decomposition of Lip we address the problem of decay of correlations: for $f = \mathbf{1}\int f dx + (f - \mathbf{1}\int f dx) \in \mathbf{1}\mathbb{C}\oplus H, \varphi \in L^1$ we have that

$$
\begin{aligned}
\mathrm{Cor}(f,\varphi\circ T^n) &= \Big|\int_X f(\varphi\circ T^n)\,\mathrm{d}x - \int_X f\,\mathrm{d}x\int_X \varphi\,\mathrm{d}x\Big| \\
&= \Big|\int_X (\hat{T}^n f)\varphi\,\mathrm{d}x - \int_X f dx\int_X \varphi\,\mathrm{d}x\Big| \\
&= \Big|\int_X (\underbrace{\hat{T}^n\mathbf{1}}_{=\mathbf{1}}\int_X f\,\mathrm{d}x)\varphi\,\mathrm{d}x + \int_X \hat{T}^n(f-\mathbf{1}\int_X f\,\mathrm{d}x)\varphi\,\mathrm{d}x - \int_X \varphi\,\mathrm{d}x\int_X f\,\mathrm{d}x\Big| \\
&= \Big|\int_X \hat{T}^n(f-\mathbf{1}\int_X f\,\mathrm{d}x)\varphi\,\mathrm{d}x\Big|.
\end{aligned}
$$

Since $h := (f - \mathbf{1}\int f\,\mathrm{d}x) \in H$, we have that $\|\hat{T}^n\|_\infty \leq |\hat{T}^n h|_{\mathrm{Lip}} \leq 2^{-n}|h|_{\mathrm{Lip}}$, hence,

$$
\mathrm{Cor}(f,\varphi\circ T^n) \leq 2^{-n}|h|_{\mathrm{Lip}}\|\varphi\|_1.
$$

In other words, for sufficiently regular functions, we have an exponential decay of correlations. Note that $|\cdot|_{\mathrm{Lip}}$ becomes a norm when restricted to H and in fact, it is equivalent to $\|\cdot\|_{\mathrm{Lip}}$, since $|h|_{\mathrm{Lip}} \leq \|h\|_{\mathrm{Lip}} = \|h\|_\infty + |h|_\infty \leq 2|h|_{\mathrm{Lip}}$. The fundamental ingredients of the previous proof are

- 1 is an eigenvalue of $\hat{T}$,

- the rest of the spectra of $\hat{T}$ is controlled by the inequality $|\hat{T}|_{\text{Lip}} \leq |h|_{\text{Lip}}$ which is essentially a bound of the spectral radius of $\hat{T}$ restricted to H:

$$\rho(\hat{T}|_H) = \lim_{n\to\infty} (\|\hat{T}^n\|_{\text{Lip}})^{\frac{1}{n}} \leq \|\hat{T}\|_{\text{Lip}} = \sup_{f\in H\setminus\{0\}} \frac{|\hat{T}f|_{\text{Lip}}}{|f|_{\text{Lip}}} \leq \frac{1}{2}$$

The previous conditions can be summarized by a geometric description of the spectra of $\hat{T}$: 1 is an isolated eigenvalue and the rest of the spectra lies in the circle $\{z : |z| \leq \frac{1}{2}\}$. It is also possible to prove that 1 is a simple eigenvalue. This set of conditions on an operator is known as *spectral gap*, and it will be the fundamental property of the operators we are going to use from now on. Although we did this analysis just for the doubling map, on the next section we will consider a more general version of this problem.

3.2 Ruelle Operator

In the previous section, we studied the action of the transfer operator on certain function spaces. For most purposes, it takes the form

$$Tg(x) = \sum_{Ty=x} \frac{g(y)}{|T'(y)|} = \sum_{Ty=x} \exp(-\log|T'(y)|)g(y).$$

We introduce a generalized version of the transfer operator.

Definition 3.17. Let $T : X \to X$ be a measurable transformation such that $T^{-1}(x)$ is at most countable for every $x \in X$, and $f : X \to \mathbb{C}$ a function such that $\sum_{Ty=x} \exp(f(y))$ is convergent of every $x \in X$. *The Ruelle transfer operator* L_f is defined on functions $g : X \to \mathbb{C}$ by

$$(L_f g)(x) = \sum_{Ty=x} \exp(f(y))g(y).$$

Remark. The domain and codomain of the Ruelle operator are not specified, since it is possible to define it several function spaces. Given the case, we will specify the domain and codomain.

Remark. There is no convention on the form of the Ruelle transfer operator. Some authors (as Baladi in [Bal00]) write it as

$$(L_f g)(x) = \sum_{Ty=x} f(y)g(y).$$

The results are essentially the same for both versions of the operator. We stick to the convention used by Bowen in [Bow75] and by Parry and Pollicott in [PP90].

The iterates of the transfer operator can be written as

$$(L_f^n g)(x) = \sum_{T^n y = x} \exp(S_n f(y)) g(y),$$

with $S_n f(x) = \sum_{k=0}^{n} f(T^k x)$.

Analogously to Proposition 3.4, we have the following property for Ruelle operator:

Proposition 3.18. *For very $g, h \in \mathcal{C}(X)$ we have that*

$$((L_f g) \cdot h)(x) = L_f(g \cdot (h \circ \sigma))(x).$$

We study now the case when T is a subshift of finite type. In particular, the sum of the Ruelle operator always converges, and it is possible to define it for $\varphi \in \mathcal{C}(X)$ as a bounded linear operator on $\mathcal{C}(X)$. For the case of Gauss Map, it is still possible to define Ruelle operator in a more reduced class of functions.

As with the transfer operator, we need to find a suitable function space where the spectrum of Ruelle operator behaves nicely.

For a function $f \in \mathcal{C}(\Sigma_A)$, we define its $k-$th variation by

$$\mathrm{var}_k\, f = \sup\{|f(x) - f(y)| \ : x_i = y_i \ \forall \ |i| \leq k\}.$$

Analogously, we may define var_k^+ for functions $f \in \mathcal{C}(\Sigma_A^+)$. For a given transition matrix A and $\theta \in (0,1)$, define $F_\theta = F_\theta(\Sigma_A) = \{f : \Sigma_A \to \mathbb{R}$ continuous st $\mathrm{var}_n\, f \leq C\theta^n$ for some C$, n \in \mathbb{N}\}$. Analogously, we may define $F_\theta^+ = F_\theta(\Sigma_A^+)$. For $f \in F_\theta$, define $|f|_\theta = \sup_n\{\mathrm{var}_n\, f/\theta^n\}$ and $\|f\|_\theta = \|f\|_\infty + |f|_\theta$. With this norm, F_θ and F_θ^+ are Banach spaces.

We present now the Ruelle-Perron-Frobenius Theorem, the fundamental result of this section. It will allow us to conclude the regularity properties of the pressure functions.

Theorem 3.19 (Ruelle-Perron-Frobenius Theorem)**.** *Suppose that A is an aperiodic matrix and $f \in F_\theta^+$. Then*

(a) There is a simple maximal positive eigenvalue λ of L_f acting on F_θ^+ with corresponding strictly positive eigenfunction $h \in F_\theta^+$;

(b) *The reminder of the spectrum of $L_f : F_\theta^+ \to F_\theta^+$ is contained in a disc of radius strictly smaller than λ;*

(c) *There is a unique probability measure ν such that $L_f^*\nu = \lambda\nu$ and $\int h d\nu = 1$;*

(d) $\dfrac{1}{\lambda^n} L_f^n v \to h \int v d\nu$ *uniformly for all $v \in \mathcal{C}(X^+)$.*

Remark. Recall that every Borel probability measure μ defines a functional $\mu : C(X) \to \mathbb{R}$ by $\mu(h) = \int_X h d\mu$. By the Riesz Representation theorem, we may identify $\mathcal{C}(X)^*$ with the set of Borel probability measures in X and adopt the notation $\mu \in \mathcal{C}(X)^*$.

With this identification, the dual operator L_f^* can be defined acting on the set of Borel probability measures by

$$\int_X g \,\mathrm{d}(L_f^*\eta) = \int_X (L_f g) \,\mathrm{d}\eta$$

for every $g \in \mathcal{C}(X^+)$ and η probability measure in X. Then, it turns out that the property $L_f^*\nu = \lambda\nu$ is equivalent to

$$\lambda \int_X g \,\mathrm{d}\nu = \int_X L_f g \,\mathrm{d}\nu$$

for every continuous function g. If we introduce the notation

$$\eta(f) = \int_X f \,\mathrm{d}\eta$$

for a probability measure η and an integrable function f, the above properties can be written as

$$(L_f^*\eta)(g) = \eta(L_f g)$$
$$\nu(L_f g) = \lambda\nu(g)$$

for every continuous function g. Iterating, we get that

$$\nu(g) = \lambda^{-n}\nu(L_f^n g)$$

for every $n \in \mathbb{N}$.

Note that the measure ν satisfies $L_f^* v = \lambda\nu$ if and only if

$$\int_X g \,\mathrm{d}\nu = \int_X \lambda^{-1} L_f g \,\mathrm{d}\nu$$

for every continuous function g. Define the measure $\nu\hat{\circ}T$ by

$$\nu\hat{\circ}T(A) = \sum_{a\in S} \nu[T(A\cap C_a)],$$

where S is the set of symbols of the Subshift of Finite Type.

Remark. In general, $\nu\hat{\circ}T$ is not equal to the usual composition $\nu \circ T$, since the sets $T(A\cap C_a)$ are not pairwise disjoint.

Then, it is possible to prove that ν is absolutely continuous with respect to $\nu\hat{\circ}T$ and the following integration formula

$$\int_X g\,\mathrm{d}\nu\hat{\circ}T = \sum_{a\in S}\int_{TC_a} g(ax)\,\mathrm{d}\nu(x).$$

Using this formula, note that $L_f^*\nu = \lambda\nu$ if and only if

$$\begin{aligned}
\int_X g\,\mathrm{d}\nu &= \int_X \lambda_f^{-1} g\,\mathrm{d}\nu \\
&= \int_X \lambda^{-1}\sum_{Ty=x}\exp(f(y))g(y)\,\mathrm{d}\nu(x) \\
&= \sum_{a\in S}\int_{TC_a}\lambda^{-1}\exp(f(ax))g(ax)\,\mathrm{d}\nu(x) \\
&= \int g\lambda^{-1}\exp(f)\,\mathrm{d}\nu\hat{\circ}T.
\end{aligned}$$

This equations leads us to the following definition

Definition 3.20. A measure ν in X satisfying

$$\frac{\mathrm{d}\nu}{\mathrm{d}\nu\hat{\circ}T} = \lambda^{-1}\exp(f)$$

is called $f-$*conformal* (see [Sar99]).

We will breafly sketch the proof of the existence of the eigenmeasure ν. Denote by $\mathcal{P}(X)$ the set of probability measures on X. By the Banach-Alaoglu Theorem, $\mathcal{P}(X)$ is compact in the weak-* topology. If λ is the eigenvalue associated to the eigenmeasure ν, then neccesarily $\lambda = L_f^*1 > 0$. Consider the operator $V : \mathcal{P}(X) \to \mathcal{P}(X)$ given by

$$V\mu = \frac{L_f^*\mu}{(L_f^*\mu)(1)}.$$

It is possible to prove that V maps $\mathcal{P}(X)$ in $\mathcal{P}(X)$ and that it is continuous in the weak-* topology. By the Schauder-Tychonoff Theorem, there exists a fixed point $\nu \in \mathcal{P}(X)$ for the map V, that is

$$\frac{L_f^*\nu}{(L_f^*\nu)(1)} = \nu$$
$$L_f^*\nu = \lambda\nu$$

which is the desired eigenmeasure.

Now we derive some consequences of Ruelle-Perron-Frobenius theorem

Proposition 3.21. *Define a measure $\mu = h\nu$ with h, ν as in Theorem 3.19. Then μ is an invariant probability measure.*

Proof. Note that $\int hd\nu = 1$ ensures that μ is a probability measure. Now, for a given continuous function g we have that

$$\begin{aligned}
\mu(g) &= \nu(hg) \\
&= \nu(\lambda^{-1}L_f h \cdot g) \\
&= \lambda^{-1}\nu(L_f(h \cdot (g \circ \sigma)) \\
&= \lambda^{-1}(L_f^*\nu)(h \cdot (g \circ \sigma)) \\
&= \nu(h \cdot (g \circ \sigma)) \\
&= \mu(g \circ \sigma)
\end{aligned}$$

where we used Proposition 3.1. ■

Now we prove that the measure just constructed satisfies a modified version of the Gibbs property (which will turn out to be equivalent to the Gibbs property):

Proposition 3.22. *There exist constants $C_1, C_2 > 0$ such that*

$$C_1 \leq \frac{\mu(C(i_0, ..., i_{m-1}))}{\exp(-m \log \lambda + S_m f(x))} \leq C_2$$

for every $m \geq 0$ and $x \in C_{i_0,...,i_{m-1}}$.

Proof. First we make some estimations on the Birkhoff sums for elements belonging to a fixed cylinder $C_{i_0,...,i_{m-1}}$. If $x, y \in C_{i_0,...,i_{m-1}}$ then $|\sigma^k(x) - \sigma^k(y)| \leq \theta^{m-k}$ for every $k = 0, ...m$. On the other side, from the definition of $|f|_\theta$, we get that

$$|S_m f(x) - S_m f(y)| \leq \sum_{k=0}^{m-1} |f|_\theta \theta^{m-k} \leq \frac{|f|_\theta}{1-\theta}.$$

Now note that for every $x, z \in C_{i_0,\dots,i_{m-1}}$

$$\begin{aligned}\mu(C_{i_0,\dots,i_{m-1}}) &= \mu(1_{C_{i_0,\dots,i_{m-1}}}) = \nu(h \cdot 1_{C_{i_0,\dots,i_{m-1}}}) \\ &= \lambda^{-m}\nu(L_f^m(h \cdot 1_{C_{i_0,\dots,i_{m-1}}})) \\ &\le \lambda^{-m}\nu(\exp(S_m f(z))h(z)) \\ &\le \exp(-m\log\lambda + S_m f(x)) \underbrace{\exp\left(\frac{|f|_\theta}{1-\theta}\right)\|h\|_\infty}_{=C_2}.\end{aligned}$$

To obtain a lower bound for the measure of the cylinder, we use that A is aperiodic, so there exists $M > 0$ such that for a given $y \in \Sigma_A^+$ there exists $z \in C_{i_0,\dots,i_{m-1}}$ with $\sigma^{m+M}(z) = y$. Then

$$\begin{aligned}\mu(C_{i_0,\dots,i_{m-1}}) &= \lambda^{-(m+M)}\nu(L_f^{m+M}(h \cdot 1_{C_{i_0,\dots,i_{m-1}}})) \\ &\ge \lambda^{-(m+M)}\exp(S_{m+M}f(z))\inf h \\ &\ge \lambda^{-(m+M)}\exp(S_m f(z) - M\|f\|_\infty)\inf h \\ &\ge \exp(-m\log\lambda + S_m f(x)) \underbrace{\lambda^{-M}\exp(\frac{-|f|_\theta}{1-\theta}) - M\|f\|_\infty}_{=C_1}\end{aligned}$$

which completes the proof. ■

Note that the above proposition suggests a relation between the maximal eigenvalue of L_f and the pressure function of f.

Proposition 3.23. *Let λ be the maximal eigenvalue of L_f given by Theorem 3.19. Then*

$$P(f) = \log\lambda.$$

Proof. Let $C_{i_0,\dots,i_{m-1}}$ be a fixed cylinder, then by the previous proposition we have that

$$C_1 \sup_{y\in C_{i_0,\dots,i_{m-1}}} \exp(S_m f(y)) \le \exp(m\log\lambda)\mu(C_{i_0,\dots,i_{m-1}}) \le C_2 \sup_{y\in C_{i_0,\dots,i_{m-1}}} \exp(S_m f(y))$$

$$\begin{aligned}C_1 \sum_{i_0,\dots,i_{m-1}} \sup_{y\in C_{i_0,\dots,i_{m-1}}} \exp(S_m f(y)) &\le \exp(m\log\lambda) \\ &\le C_2 \sum_{i_0,\dots,i_{m-1}} \sup_{y\in C_{i_0,\dots,i_{m-1}}} \exp(S_m f(y))\end{aligned}$$

$$\begin{aligned}\frac{\log C_1}{m} + \frac{1}{m}\log \sum_{i_0,\dots,i_{m-1}} \sup_{y\in C_{i_0,\dots,i_{m-1}}} \exp(S_m f(y)) &\le \log\lambda \\ &\le \frac{\log C_2}{m} + \frac{1}{m}\log \sum_{i_0,\dots,i_{m-1}} \sup_{y\in C_{i_0,\dots,i_{m-1}}} \exp(S_m f(y))\end{aligned}$$

so we obtain that

$$P(f) \leq \log \lambda \leq P(f).$$

■

Corollary 3.24. *The measure μ constructed above is a Gibbs measure,*

$$C_1 \leq \frac{\mu(C(i_0, ..., i_{m-1}))}{\exp(-mP(f) + S_m f(x))} \leq C_2$$

hence, an equilibrium state for f,

$$P(f) = h_\mu(\sigma) + \int_{\Sigma_A^+} f d\mu.$$

In conclusion, the Ruelle-Perron-Frobenius Theorem implies that the operator L_f has a positive maximal isolated eigenvalue λ, with associated positive eigenfunction h and an eigenmeasure ν for the dual operator L_f^*. The logarithm of the eigenvalue λ is equal to the pressure of the potential f. The measure $\mu = h\nu$ is a Gibbs state with respect to f. The fact that λ is isolated from the rest of the spectrum is the key for the next section.

3.3 Perturbation Theory

In this section, we just borrow a result of Kato's Perturbation Theory of Operators (see [Kat12]) and use it to conclude the analiticity of the pressure function for subshifts of finite type. Recall the notion of differentiability of functions in Banach spaces:

Definition 3.25. Let E, F be Banach spaces, $U \subset E$ an open subset, and $f : U \to F$ a function. We say that f is *analytic* if it is continuous and for every affine line $L \subset E$ and every continuous linear functional $\alpha : F \to \mathbb{C}$, the mapping $L \cap U \to \mathbb{C}$ given by $\alpha \circ f|_{U \cap L}$ is analytic.

Theorem 3.26 (Perturbation Theorem). *Let $B(V)$ denote the Banach algebra of bounded linear operators on a Banach space V. If $L_0 \in B(V)$ has a simple isolated eigenvalue α_0 with corresponding eigenvector v_0 then for any $\varepsilon > 0$ there exists $\delta > 0$ such that if $L \in B(V)$ with $\|L - L_0\| < \delta$ then L has a simple isolated eigenvalue $\alpha(L)$ and corresponding eigenvector $v(L)$ with $\alpha(L_0) = \alpha_0$, $v(L_0) = v_0$ and*

(a) $L \mapsto \alpha(L)$, $L \mapsto v(L)$ *are analytic for* $\|L - L_0\| < \delta$;

(b) *for* $\|L - L_0\| < \delta$ *we have* $|\alpha(L) - \alpha_0| < \varepsilon$, *and* $spec(L) - \{\alpha(L)\} \subset \{z : |z - \alpha_0| > \varepsilon\}$.

Note that by the results of the previous section show that the transfer operator L_f associated to a potential $f \in F_\theta^+$ satisfies the hypothesis of Theorem 3.26 with maximal eigenvalue $\lambda = \exp P(f)$. With this, we may complete the full description of the pressure function $t \mapsto P(-t \log |DT|)$ for Subshifts of Finite Type.

Corollary 3.27. *The function $f \in F_\theta^+ \mapsto P(f) \in \mathbb{R}$ is analytic. In particular, the pressure function $t \in \mathbb{R} \mapsto P(-t \log |DT|)$ is real analytic at all points of its domain.*

Proof. It is enough to prove that the function $f \in F_\theta^+ \mapsto L_f \in B(F_\theta^+)$ is analytic. To see this, consider the map $F_\theta^+ \to F_\theta^+ \to B(F_\theta^+) \to B(F_\theta^+)$ given by $f \mapsto \exp(f) \mapsto M \mapsto L_i \circ M$ where M is the operator $\phi \mapsto \phi \exp(f)$ and L_i the operator given by

$$L_i w(x) = \begin{cases} w(ix), & \text{if } A_{i,x_0} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

All these compositions can be seen analytic, so we conclude. ∎

The typical plot of the function $t \mapsto P(-t \log |DT|)$ is then as shown in the picture below:

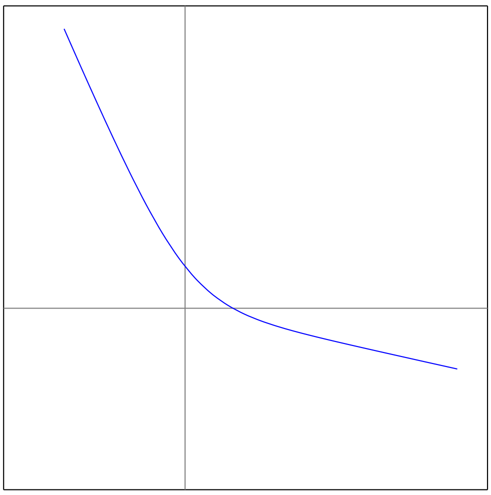

FIGURE 3.1: A generic plot of the pressure function $t \mapsto P(-t \log |DT|)$.

We finish this section by proving a differentiation formula for the pressure function.

Proposition 3.28 (Deriviative of pressure)**.** *Suppose A is aperiodic and $f, g \in F_\theta^+$. Then*

$$\frac{\mathrm{d}}{\mathrm{d}s} (P(f + sg)) \Big|_{s=0} = \int_{\Sigma_A^+} g \, \mathrm{d}\mu_f,$$

where μ_f is the equilibrium measure of f.

Proof. For each s, we have the eigenvalue equation

$$L_{f+sg}w(s) = \exp(P(f+sg))w(s)$$

for some analytic function w (by Theorem 3.19). Differentiating both sides with respect to s at $s=0$ we obtain

$$L_f(w(0)g) + L_f\left(\frac{\mathrm{d}w}{\mathrm{d}s}\Big|_{s=0}\right) = \exp(P(f))w(0)\frac{\mathrm{d}}{\mathrm{d}s}(P(f+sg))\Big|_{s=0} + \exp(P(f))\frac{\mathrm{d}w}{\mathrm{d}s}\Big|_{s=0}.$$

Note that taking $s=0$ gives $L_f w(0) = \exp(P(f))w(0)$ so $w(0)$ is the density of the measure $\mu_f = w(0)\nu_f$ constructed in the corollaries of Theorem 3.19, hence it satisfies

$$\int_{\Sigma_A^+} w(0)\,\mathrm{d}\nu_f = 1.$$

Integrating with respect to the measure ν_f and recalling the property $\nu_f(L_f h) = \lambda\nu(h)$ where $\lambda = \exp(P(f))$, we have that

$$\int_{\Sigma_A^+} L_f(w(0)g)\,\mathrm{d}\nu_f + \int_{\Sigma_A^+} L_f\left(\frac{\mathrm{d}w}{\mathrm{d}s}\Big|_{s=0}\right)\mathrm{d}\nu_f = \exp(P(f))\left(\int_{\Sigma_A^+} w(0)g\,\mathrm{d}\nu_f + \int_{\Sigma_A^+}\frac{\mathrm{d}w}{\mathrm{d}s}\Big|_{s=0}\mathrm{d}\nu_f\right)$$

and hence we have

$$\int_{\Sigma_A^+} w(0)gd\nu_f + \int_{\Sigma_A^+}\frac{dw}{ds}\Big|_0 d\nu_f = \frac{d}{ds}P(f+sg)\Big|_{s=0}\int_{\Sigma_A^+} w(0)d\nu_f + \int_{\Sigma_A^+}\frac{dw}{ds}\Big|_{s=0}d\nu_f,$$
$$\frac{d}{ds}P(f+sg)\Big|_{s=0} = \int_{\Sigma_A^+} w(0)gd\nu_f = \int_{\Sigma_A^+} gd\mu_f.$$

∎

3.4 Transfer Operator for the Gauss Map

In the last sections, we studied the techniques that allowed us to deduce the regularity properties of the pressure function $t \mapsto P(-t\log|T'|)$ for a Subshift of Finite Type. The symbolic model associated to this dynamic is no longer a Subshift of Finite Type, since the set of states is countable. This fact has serious implications in the proofs of the theorems from the previous sections. The finiteness of the set of states for the Subshifts of Finite Type is reflected in the compactness of the spaces of sequences of states. For the Gauss Map, this space is non-compact, so the techniques used before will not work. The map has infinite topological entropy and pressure may take infinity as value for a wide class of potentials.

In this section we state (without proof) some results of Mayer (see [May76], [May90]) which characterize the behavior of the pressure function for the Gauss Map.

For $\phi \in C([0,1],\mathbb{R})$, the Ruelle transfer operator $L_\phi : C([0,1]) \to C([0,1])$ is defined by

$$(L_\phi w)(x) = \sum_{Gy=x} \exp(\phi(y))w(y). \tag{3.1}$$

Taking the potential $\phi = -t\log|G'|$, the Ruelle operator takes the form

$$(L_t w)(x) = \sum_{n=1}^{\infty} \left(\frac{1}{x+n}\right)^{2t} f\left(\frac{1}{x+n}\right).$$

Theorem 3.29 (D. Mayer)**.** *Let $D = \{z \in \mathbb{C} : |z-1| < \frac{3}{2}\}$ and $A_\infty(D)$ be the Banach space of holomorphic functions in D. Then, for $t > 1/2$, the operator $L_t : A_\infty(D) \to A_\infty(D)$*

$$(L_t w)(x) = \sum_{n=1}^{\infty} \left(\frac{1}{x+n}\right)^{2t} f\left(\frac{1}{x+n}\right)$$

has an isolated maximal simple eigenvalue $\lambda_1(t)$, and it satisfies $P(-t\log|G'|) = \log\lambda_1(t)$. For every $t > 1/2$ there exists an equilibrium measure μ_t satisfying the Gibbs property and giving positive measure to every open set of $[0,1]$.

It is possible to prove that for $t \leq 1/2$, the pressure function $P(-t\log|G'|)$ is not finite. Using the Perturbation Theorem 3.26, we may conclude the following

Theorem 3.30. *The function*

$$\begin{aligned} P : [0,\infty) &\to \mathbb{R} \\ t &\mapsto P(-t\log|G'|) \end{aligned}$$

has the following properties

(a) *Is infinite for $t \in [0,1/2]$ and finite for $t \in (1/2,\infty)$.*

(b) *Is real analytic in $(1/2,\infty)$.*

(c) *Is strictly decreasing and convex in $(1/2,\infty)$.*

(d) *For every $t \in (1/2,\infty)$, there exists an equilibrium measure μ_t satisfying the Gibbs property.*

With this information, we have a full description of the pressure function for the Gauss Map:

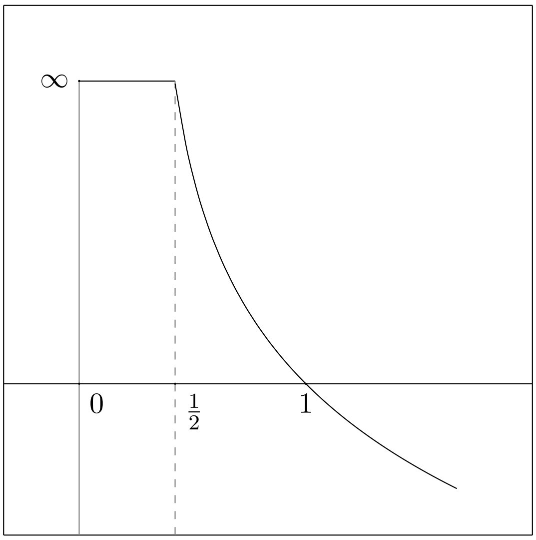

FIGURE 3.2: The plot of the function $t \mapsto P(-t \log |G'|)$.

The analiticity of the pressure function is the key to prove the analiticity of the function encoding the Hausdorff dimension of certain sets, which will be the topic of the next chapter.

Chapter 4

The Hausdorff Dimension Function For Borel-Bernstein Sets

4.1 The setting

Recall that every irrational number $x \in (0,1)$ can be written uniquely as a continued fraction of the form

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}} = [a_1(x)a_2(x)a_3(x)\dots] = [a_1a_2a_3\dots],$$

where $a_i \in \mathbb{N}$. For a general account on continued fractions see [HW79], [Khi63]. The $n-th$ approximant $p_n(x)/q_n(x)$ of the number $x \in [0,1]$ is the rational number defined by

$$\frac{p_n(x)}{q_n(x)} = [a_1(x)a_2(x)\dots a_n(x)].$$

The metric theory of continued fractions study the set of numbers for which the sequence $(a_n(x))_n$ satisfies certain properties. One of the first results in this direction is the Borel-Bernstein theorem [Bor12], [Ber11] proved in 1912.

Theorem 4.1 (Borel-Bernstein Theorem). *Let φ be a positive function $\varphi : \mathbb{N} \to \mathbb{R}$ and*

$$E(\varphi) = \{x \in (0,1) : a_n(x) \geq \varphi(n) \text{ infinitely often}\}. \tag{4.1}$$

Then $m(E(\varphi))$ is zero or one depending as the series $\sum_{n=1}^{\infty} \frac{1}{\varphi(n)}$ converges or not, where m denotes the Lebesgue measure.

Proof. First suppose $\sum \frac{1}{\varphi(n)} = \infty$ and let $x > 1$, $n, m \in \mathbb{N}$ and $1 \leq j \leq n$. Recall that the length any fundamental intervals is given by

$$|I(a_1, ..., a_n)| = \frac{1}{q_n(q_n + q_{n-1})}.$$

where q_n is the denominator of the reduced fraction associated to the number $[a_1, ..., a_n]$. If we add a new symbol, the length of the interval is

$$|I(a_1, ..., a_n, s)| = \left| \frac{sp_n + p_{n-1}}{sq_n + q_{n-1}} - \frac{(s+1)p_n + p_{n-1}}{(s+1)q_n + q_{n-1}} \right| = \frac{1}{s^2} \frac{1}{q_n + \frac{q_{n-1}}{s}} \frac{1}{\frac{q_{n-1}}{s} + (1 + \frac{1}{s} q_n)}.$$

Hence we obtain the following estimate for the ratio of the lengths

$$\frac{1}{3s^2} < \frac{|I(a_1, ..., a_n, s)|}{|I(a_1, ..., a_n)|} < \frac{2}{s^2}.$$

Then,

$$\begin{aligned} \sum_{s<x} |I(a_1, ..., a_m, a_{m+1}, ..., a_{m+j}, s)| &= \sum_{s \geq 1} |I(a_1, ..., a_{m+j}, s)| - \sum_{s \geq x} |I(a_1, ..., a_{m+j}, s)| \\ &\leq \left(1 - \frac{1}{3} \sum_{s \geq x} \frac{1}{s^2} \right) |I(a_1, ..., a_m, a_{m+1}, ..., a_{m+j})| \\ &\leq \left(1 - \frac{1}{3(1+x)} \right) |I(a_1, ..., a_{m+j})| \end{aligned}$$

for every choice of a_i, and where the last inequality is obtained by comparing with the integral of $1/x^2$. Let

$$F_{m,j} = \bigcup_{\substack{a_i \in \mathbb{N}, \\ a_{m+1} < \phi(m+i), i=1,...,j}} I(a_1, ..., a_{m+j}).$$

Now, taking successively $x = \varphi(m+n), \varphi(m+n-1), ..., \varphi(m+1)$ in the inequality above, we obtain

$$\begin{aligned} m(F_{m,n}) &= \sum_{\substack{a_i \in \mathbb{N} \\ i \leq m}} \sum_{a_{m+1} < \varphi(m+1)} \cdots \sum_{a_{m+n} < \varphi(m+n)} |I(a_1, ..., a_m, a_{m+1}, ..., a_{m+n})| \\ &\leq \sum_{a_i \in \mathbb{N}} \left(1 - \frac{1}{3(1+\varphi(m+1))} \right) \cdots \left(1 - \frac{1}{3(1+\varphi(m+n))} \right) |I(a_1, ..., a_m)| \\ &\leq \prod_{j=m+1}^{n} \left(1 - \frac{1}{3(1+\varphi(m+j))} \right). \end{aligned}$$

From the divergence of $\sum_n \frac{1}{\phi(n)}$, we conclude that for every m the later product goes to 0 as n goes to infinity. Hence, for every m, $m(F_{m,n})$ tends to 0 as n tends to infinity.

Finally, observe that the set $B_m = \{x : a_{m+i} < \varphi(m+i)\ \forall i \geq 1\}$ is contained in every $F_{m,n}$, so $m(B_m) = 0$ and $m(E(\varphi)) = m\left([0,1] \setminus \bigcup_m B_m\right) = 1$.

For the second part of the theorem, for $n \in \mathbb{N}$ denote $V_n = \{x \in [0,1] : a_n \geq \varphi(n)\}$. Then for any choice of $a_1, ..., a_n \in \mathbb{N}$, by the estimate $|I(a_1, ..., a_n, s)| < \frac{2}{s^2}|I(a_1, ..., a_n)|$, we obtain

$$\sum_{s \geq \varphi(n+1)} |I(a_1, ..., a_n, s)| < \sum_{s \geq \varphi(n+1)} \frac{2}{s^2}|I(a_1, ..., a_n)| \leq \frac{4}{\varphi(n+1)}|I(a_1, ..., a_n)|,$$

where the last inequality is again obtained by comparing with the integral of $1/s^2$. Noting that

$$V_n = \bigcup_{s \geq \varphi(n)} I(a_1, ..., a_{n-1}, s),$$

we obtain that $m(V_n) < \frac{4}{\varphi(n)}$. Finally, since $E(\varphi) = \bigcap_n \bigcup_{m \geq n} V_n$ and $\sum_n m(V_n) < \sum_n \frac{4}{\varphi(n)} < \infty$, by Borel-Cantelli Lemma we conclude that $m(E(\varphi)) = 0$. ■

Remark. The intervals $I(a_1, ..., a_n)$ have zero Lebesgue measure intersection for different choices of a_i.

We call the sets $E(\varphi)$ defined as in the equation 4.1 *Borel-Bernstein* sets. This result can be understood as a Borel-Cantelli Lemma. A natural question is to determine the Hausorff dimension of the set $E(\varphi)$ when it has Lebesgue measure zero. Several results have been obtained in this direction, for instance, in 1941 Good [Goo41] proved that

Proposition 4.2 (Good)**.**

$$\dim_H(\{x \in (0,1) : a_n(x) \to \infty \text{ as } n \to \infty\}) = \frac{1}{2}.$$

Actually, the following subset satisfies the same property, for any $B > 1$

$$\dim_H(\{x \in (0,1) : a_n(x) \geq B^n) = \frac{1}{2}.$$

The Hausdorff dimension of the set

$$E(B) = \{x \in (0,1) : a_n(x) \geq B^n \text{ infinitely often}\}$$

was recently computed by Wang and Wu [WW08] and depending on the value of B it ranges between $(1/2, 1)$. Let us stress that the main difference between the sets studied by Good and $E(B)$ is that in the later we only require a condition to be satisfied *infinitely often* and not for *all* values of $n \in \mathbb{N}$. Let us note that once the Hausdorff dimension of the sets $E(B)$ is established, it is possible to compute the Hausdorff dimension of $E(\varphi)$ for arbitrary φ (see [WW08], Theorem 4.2). Wang and Wu [WW08] also proved the following result:

Theorem 4.3 (Wang, Wu)**.** *The function $B \mapsto \dim_H(E(B))$ defined on the interval $(1, \infty)$ is continuous.*

In this section we make use of the theory of dynamical systems to study this function. First note that the Gauss map $G : [0, 1] \to [0, 1]$, defined by

$$G(x) = \frac{1}{x} - \left[\frac{1}{x}\right],$$

is closely related to the continued fraction expansion. Indeed, for $0 < x < 1$ with $x = [a_1a_2a_3...]$ we have that $a_1 = [1/x], a_2 = [1/Gx], ..., a_n = [1/G^{n-1}x]$. In particular, the Gauss map acts as the shift map on the continued fraction expansion,

$$G^n(x) = [a_{n+1}(x)a_{n+2}(x)...].$$

It is possible to define a symbolic model for the Gauss Map. In fact, consider $\Sigma = \mathbb{N}^{\mathbb{N}}$ equipped with the shift $\sigma : \Sigma \to \Sigma$ and the semi-conjugacy map $\chi : \Sigma \to (0, 1]$ sending each sequence to the real number having the given sequence as continued fraction expansion digits. Thus, we have the following commutative diagram

$$\begin{tikzcd} \Sigma \arrow[r, "\sigma"] \arrow[d, "\chi"'] & \Sigma \arrow[d, "\chi"] \\ {[0,1]} \arrow[r, "G"] & {[0,1]}\,. \end{tikzcd}$$

Note that the sets $\Sigma_m = \{1, ..., m\}$ can be identified as compact $\sigma-$invariant subsets of Σ.

It is clear from the above that studying the dynamics of the Gauss map it is possible to obtain results on the distribution of digits on the continued fraction expansion. We prove the following result:

Theorem 4.4. *The function $B \mapsto \dim_H(E(B))$ defined on the interval $(1, \infty)$ satisfies the following properties:*

1. *It is real analytic;*
2. *It is strictly decreasing;*
3. $\lim_{B\to 1} \dim_H(E(B)) = 1$;
4. $\lim_{B\to\infty} \dim_H(E(B)) = \frac{1}{2}$;

Our main technical tool is the thermodynamic formalism of the Gauss map studied in the previous chapters. We review some of the ideas that will be used in the course of this chapter.

For the Gauss Map, topological pressure can be understood as the exponential growth of periodic points with an assigned weight for each point:

$$P(\phi) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{x\in \operatorname{Fix} G^n} \exp \sum_{k=0}^{n-1} \phi(G^k x).$$

Denote by $\mathcal{M}$ the set of $G-$invariant probability measures. The topological pressure satisfies the Variational Principle

Theorem 4.5 (Variational Principle for the Gauss Map)**.**

$$P(\psi) = \sup\{h(\mu) + \int \psi d\mu : -\int \psi d\mu < \infty \text{ and } \mu \in \mathcal{M}\}, \tag{4.2}$$

where $h(\mu)$ denotes the entropy of μ.

A measure μ is called an equilibrium measure if it achieves the supremum of the previous theorem.

Recall from Theorem 3.30 that the function $t \to P(-t \log |G'|)$ is infinite in $[0, 1/2]$ and finite, real analytic, strictly decreasing and convex in $(1/2, \infty)$. Even more, for every $t \in (1/2, \infty)$, there exists an equilibrium measure μ_t which also has the Gibbs property.

We also have the following approximation property for the topological pressure

Theorem 4.6 (Approximation property)**.** *If ϕ has summable variation then*

$$P(\phi) = \sup\{P_K(\phi) : K \subset (0, 1) : K \text{ compact and invariant}\}.$$

Theorem 4.4 is a consequence of the previously listed properties of pressure function (theorem 3.30), together with the following result

Theorem 4.7 (Main Theorem)**.** *Let $d_B \in \mathbb{R}$ be the unique real number such that*

$$P(-d_B \log |BG'(x)|) = 0, \tag{4.3}$$

then $\dim_H(E(B)) = d_B$.

The above Theorem should be understood as a Bowen type formula. The proof of this result relies in the main theorem of [WW08] which can be stated as

Theorem 4.8 (Wang, Wu)**.** *The Hausdorff dimension $s_B = \dim_H E(B)$ of $E(B)$ is characterized by the following construction: for $\alpha \in \mathbb{N}$, let $s_{n,B}(\alpha) = \inf\{\rho \geq 0 : f_{n,\alpha}(\rho) \leq 1\}$ where $f_{n,\alpha} : \mathbb{R}_+ \to \mathbb{R}$ is given by*

$$f_{n,\alpha}(\rho) = \sum_{a_1,\ldots,a_n \in \{1,\ldots,\alpha\}} \frac{1}{(B^n q_n^2)^\rho}.$$

Then $s_B = \lim_{\alpha\to\infty} \lim_{n\to\infty} s_{n,B}(\alpha)$.

So if we prove that $s_B = d_B$, the proof of 4.7 is completed. We will also give a direct sketch of proof for 4.7 emulating the strategy of Wang and Wu, and making use of the limsup structure of the Borel-Bernstein sets.

4.2 Proof of Theorem 4.7

The proof of theorem 4.7 is divided in three parts. First, we show that $s_B = d_B$. As noted above, this proves Theorem 4.4. Then we show that $\dim_H E(B) \leq d_B$, and finally sketch the construction done in [WW08] to prove that $d_B \leq \dim_H E(B)$. We give just a sketch and prove some reduced facts to illustrate the procedure.

First part Recall the equation satisfied by $s_{n,B}(\alpha)$:

$$\sum_{a_1,\ldots,a_n \in \{1,\ldots,\alpha\}} \frac{1}{|Bq_n^2|^{s_{n,B}(\alpha)}} = 1.$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \log \sum_{a_1,\ldots,a_n \in \{1,\ldots,\alpha\}} \frac{1}{|Bq_n^2|^{s_B(\alpha)}} = 0$$

and

$$\lim_{\alpha\to\infty} \lim_{n\to\infty} \frac{1}{n} \log \sum_{a_1,\ldots,a_n \in \{1,\ldots,\alpha\}} \frac{1}{|Bq_n^2|^{s_B}} = 0.$$

Now, from the approximation $q_n^2(x) \asymp |(G^n)'(x)|^{-1}$ we obtain that

$$\lim_{\alpha\to\infty} \lim_{n\to\infty} \frac{1}{n} \log \sum_{a_1,\ldots,a_n \in \{1,\ldots,\alpha\}} \frac{1}{|B(G^n)'(x)|^{s_B}} = 0,$$

where x is the periodic sequence given by $(a_1, ..., a_n, a_1, ...)$. Note that this expression is equivalent to

$$\lim_{\alpha\to\infty} P|_{\Sigma_\alpha}(-s_B \log|BG'|) = 0$$

Finally, by the approximation property of Pressure, we obtain that the equation satisfied by s_B is equivalent to

$$P(-s_B \log|BG'|) = \lim_{\alpha\to\infty} P|_{K_\alpha}(-s_B \log|BG'|) = 0$$

but the only solution to this equation is by definition d_B, hence $d_B = s_B$.

The upper bound. The proof of the upper bound relies on the limsup structure of $E(B)$. Indeed, as noticed in [WW08]

$$\begin{aligned} E(B) &= \bigcap_{N=1}^{\infty} \bigcup_{n\geq N} \{x \in (0,1) : a_{n+1}(x) \geq B^{n+1}\} \\ &= \bigcap_{N=1}^{\infty} \bigcup_{n\geq N} \bigcup_{a_1,\ldots,a_n} \{x \in (0,1) : a_i(x) = a_i, 1 \leq i \leq n, a_{n+1}(x) \geq B^{n+1}\}. \end{aligned}$$

Let

$$I(a_1, ..., a_n) = \{x \in (0,1) : a_i(x) = a_i,\ 1 \leq i \leq n\}, \tag{4.4}$$

$$J(a_1, ..., a_n) = \{x \in (0,1) : a_i(x) = a_i,\ 1 \leq i \leq n,\ a_{n+1}(x) \geq B^{n+1}\}. \tag{4.5}$$

Therefore

$$E(B) = \bigcap_{N=1}^{\infty} \bigcup_{n\geq N} \bigcup_{a_1,\ldots,a_n} J(a_1, ..., a_n).$$

For the above sets, we have the following estimates on their Lebesgue measures (see [Khi63] and [WW08])

$$\begin{aligned} \frac{1}{2q_n(x)^2} \leq &|I(a_1, ..., a_n)| = \frac{1}{q_n(x)(q_n(x) + q_{n-1}(x))} \leq \frac{1}{q_n(x)^2}, \\ &|J(a_1, ..., a_n)| \leq \frac{1}{B^{n+1} q_n(x)^2} \end{aligned}$$

for $x \in I(a_1, ..., a_n)$. Now we proceed to prove the upper bound. Let $\epsilon > 0$, then

$$\mathcal{H}^{s_B+\epsilon}(E(B)) \leq \liminf_{N\to\infty} \sum_{n\geq N} \sum_{a_1,...,a_n} |J(a_1, ..., a_n)|^{s_B+\epsilon} \leq$$

$$\liminf_{N\to\infty} \sum_{n\geq N} \sum_{a_1,...,a_n} \left(\frac{1}{B^{n+1}q_n^2}\right) \leq \liminf_{N\to\infty} \sum_{n\geq N} \sum_{a_1,...,a_n} \left(\frac{2}{B^{n+1}|(G^n)'(x)|}\right)^{s_B+\epsilon}.$$

Let μ_ϵ the equilibrium measure associated to the potential $-(s_B+\epsilon)\log|BG'|$. Then, by the Gibbs property, there exists $C > 0$ such that

$$\frac{1}{(B^n(G^n)'(x))^{s_B+\epsilon}} \leq \mu_\epsilon(I(a_1, ..., a_n)) \exp(nCP(-(s_B+\epsilon)\log|BG'|))$$

for every $x \in I(a_1, ..., a_n)$, thus

$$\begin{aligned}
\mathcal{H}^{s_B+\epsilon}(E(B)) &\leq \liminf_{N\to\infty} \sum_{n\geq N} \sum_{a_1,...,a_n} \left(\frac{2}{B}\right)^{s_B+\epsilon} \mu_\epsilon(I(a_1, ..., a_n)) \exp(nCP(-(s_B+\epsilon)\log|BG'|)) \\
&\leq \liminf_{N\to\infty} \left(\frac{2}{B}\right)^{s_B+\epsilon} \sum_{n\geq N} \exp(nCP(-(s_B+\epsilon)\log|BG'|)) \\
&\leq \liminf_{N\to\infty} \left(\frac{2}{B}\right)^{s_B+\epsilon} \sum_{n\geq N} \exp(nP),
\end{aligned}$$

where $P < 0$. Therefore

$$\mathcal{H}^{s_B+\epsilon}(E(B)) \leq \liminf_{N\to\infty} \sum_{n\geq N} \exp(nP) = 0.$$

That is

$$\dim_H(E(B)) \leq s_B.$$

The lower bound. We first introduce some notation. Recall the definition of d_B as the unique solution of the equation

$$P(-t\log|BG'|) = 0.$$

For every $\alpha \geq 1$, let $d_B(\alpha)$ the solution of the equation

$$\lim_{n\to\infty} \frac{1}{n} \log \sum_{a_1,...,a_n \in\{1,...,\alpha\}} \frac{1}{(Bq_n^2)^t} = P|_{\Sigma_\alpha}(-t\log|BG'|, G|_{\chi(\Sigma_\alpha)}) = 0.$$

Denote by $d_{n,B}(\alpha)$ the solution of the equation

$$\sum_{a_1,\dots,a_n\in\{1,\dots,\alpha\}} \frac{1}{(Bq_n^2)^t} = 1$$

and by $d_B(\alpha) = \lim_{n\to\infty} d_{n,B}(\alpha)$. Then $d_B(\alpha) \to d_B$ as $\alpha \to \infty$ by the approximation property of pressure.

To prove the lower bound for the Hausdorff dimension of $E(B)$, we construct a sequence of subsets $E_\alpha(B) \subset E(B)$ such that $\dim_H E(B) \geq \dim_H E_\alpha(B) \geq d_B(\alpha)$ for α large enough. This implies that $\dim_H E(B) \geq d_B$ which completes the proof.

Start the construction by picking a sequence $n_k \in \mathbb{N}$ such that $n_1 = 1$ and

$$n_1 + \dots + n_k \leq \frac{1}{k+1} n_{k+1}$$

for $k \geq 1$. Let

$$E_\alpha(B) = \{x : [B^{n_k}] + 1 \leq a_{n_k}(x) \leq 2[B^{n_k}] \text{ for all } k \geq 1, \text{ and } 1 \leq a_j(x) \leq \alpha, \text{ for all } j \neq n_k\}.$$

Similarly, define

$$\begin{aligned} D_n = \{(\sigma_1, \dots, \sigma_n) \in \mathbb{N}^n : [B^{n_k}] + 1 \leq \sigma_{n_k} \leq 2[B^{n_k}] \text{ for all } k \geq 1, \text{ and}\\ 1 \leq \sigma_j \leq \alpha, \text{ for all } 1 \leq j \neq n_k \leq n\}. \end{aligned}$$

Then we have

$$E_\alpha(B) = \bigcap_{n\geq 1} \bigcup_{(\sigma_1,\dots,\sigma_n)\in D_n} J(\sigma_1, \dots, \sigma_n),$$

where $J(\sigma_1, \dots, \sigma_n)$ is the set defined by 4.5. We analyze now the structure of $E_\alpha(B)$. For each $n \in \mathbb{N}$, the union $\bigcup_{(\sigma_1,\dots,\sigma_n)\in\mathbb{N}^n} J(\sigma_1, \dots, \sigma_n)$ is a disjoint union of intervals. For a fixed interval $J(\sigma_1, \dots, \sigma_n)$, call $\tilde{G}(\sigma_1, \dots, \sigma_n)$ the minimum distance to other interval $J(\sigma_1', \dots, \sigma_n')$. Then it satisfies [WW08, see (19), (20)])

$$\begin{aligned} \tilde{G}(\sigma_1, \dots, \sigma_n) &\geq \frac{1}{2\alpha} |J(\sigma_1, .., \sigma_n)| \quad \text{if } n \neq n_k - 1,\\ \tilde{G}(\sigma_1, \dots, \sigma_n) &\geq \frac{1}{2} |J(\sigma_1, .., \sigma_n)| \quad \text{if } n = n_k - 1. \end{aligned}$$

This estimate is fundamental, since allows us to determine how many intervals J are contained in a given ball, information needed to estimate the measure of such ball. Now we define a measure supported on $E_\alpha(B)$ which satisfies the hypothesis of the mass distribution principle. Let $m_1 = 0$ and $m_k = n_k - n_{k-1} - 1$ for $k \geq 2$. We define a measure μ on every set of the form $J(\sigma)$ with $\sigma \in \bigcup D_n$.

For $\sigma_1 \in D_1$, let

$$\mu(J(\sigma_1)) = \frac{1}{[B]}.$$

For $(\sigma_1, ..., \sigma_{n_2-1}) \in D_{n_2-1}$, let

$$\mu(J(\sigma_1, ..., \sigma_{n_2-1})) = \mu(J(\sigma_1)) \left(\frac{1}{B^{m_2} q_{m_2}^2(\sigma_2, ..., \sigma_{n_2-1})} \right)^{d_{m_2,B}(\alpha)}$$

and for $(\sigma_1, ..., \sigma_{n_2}) \in D_{n_2}$ let

$$\mu(J(\sigma_1, ..., \sigma_{n_2})) = \frac{1}{[B^{n_2}]} \mu(J(\sigma_1, ..., \sigma_{n_2-1})).$$

For $1 < n < n_2 - 1$ and $(\sigma_1, .., \sigma_n) \in D_n$, let

$$\mu(J(\sigma_1, .., \sigma_n)) = \sum_{\sigma_n, ..., \sigma_{n_2-1} \in \{1, ..., \alpha\}} \mu(J(\sigma_1, .., \sigma_n, \sigma_{n+1}, ..., \sigma_{n_2-1}).$$

Thus, we have defined the measure μ for intervals $J(\sigma_1, ..., \sigma_n)$ when $n \in \{n_1, ..., n_2\}$. Inductively, we may define μ in $J(\sigma_1, ..., \sigma_n)$ for $n \geq n_2$ using the same procedure: define it for $J(\sigma_1, ..., \sigma_{n_{k+1}-1})$ in terms of $\mu(J(\sigma_1, ..., \sigma_{n_k}))$ by

$$\mu(J(\sigma_1, ..., \sigma_{n_{k+1}-1})) = \mu(J(\sigma_1, ..., \sigma_{n_k})) \left(\frac{1}{B^{m_k+1} q_{m_{k+1}}^2(\sigma_{n_k+1}, ..., \sigma_{n_{k+1}-1})} \right)^{d_{m_{k+1},B}(\alpha)}.$$

This allows us to define for example μ in the intervals of the form $J(\sigma_1, ..., \sigma_{n_3-1})$ from the definition of $\mu(J(\sigma_1, ..., \sigma_{n_2}))$. Now we define it for $J(\sigma_1, ..., \sigma_{n_{k+1}})$ by

$$\mu(J(\sigma_1, ..., \sigma_{n_{k+1}})) = \frac{1}{[B^{n_k}]} \mu(J(\sigma_1, ..., \sigma_{n_{k+1}-1})).$$

Finally, we define μ for any $n_k < n < n_{k+1} - 1$ by

$$\mu(J(\sigma_1, ..., \sigma_n)) = \sum_{\sigma_{n+1}, ..., \sigma_{n_{k+1}-1} \in \{1, ..., \alpha\}} \mu(J(\sigma_1, ..., \sigma_n, \sigma_{n+1}, ..., \sigma_{n_{k+1}-1})).$$

This completes the inductive definition of μ on every interval of the form $J(\sigma)$ with $\sigma \in \bigcup_n D_n$. By the Kolmogorov extension theorem, μ extends to a probability measure supported on $E_\alpha(B)$. Note that with this definition, it is possible to write $\mu(J(\sigma_1, ..., \sigma_n))$

in terms of $\mu(J(\sigma_1, ..., \sigma_{n_k}))$ with $n_{k-1} < n \leq n_k$, which has a simple form given by

$$\mu(J(\sigma_1, ..., \sigma_{n_k})) = \prod_{j=1}^{k} \frac{1}{[B^{n_j}]} \left(\frac{1}{q_{m_j}^2 B^{m_j}} \right)^{d_{m_j,B}(\alpha)}.$$

Let $0 < t < d_B(\alpha)$ and $\varepsilon = \frac{d_B(\alpha)-t}{2}$. Then, according to the calculations of [WW08, see (33), (34), (35)], there exists a constant $c_0 \in \mathbb{R}_+$ such that

$$\mu(J(\sigma_1, ..., \sigma_n)) \leq c_0 |J(\sigma_1, ..., \sigma_n)|^{t-2\varepsilon}$$

for every n large enough.

Now we estimate $\mu(B(x,r))$ for small $r > 0$. Recall the bounds for the intervals $J(\sigma_1, ...\sigma_n)$ given by

$$\tilde{G}(\sigma_1, ..., \sigma_n) \geq \frac{1}{2\alpha} |J(\sigma_1, .., \sigma_n)| \quad \text{if } n \neq n_k - 1,$$
$$\tilde{G}(\sigma_1, ..., \sigma_n) \geq \frac{1}{2} |J(\sigma_1, .., \sigma_n)| \quad \text{if } n = n_k - 1.$$

For k_0 large enough, take $r_o = \min_{1 \leq j \leq n_{k_0}} \min_{(\sigma_1, ..., \sigma_j) \in D_j} \tilde{G}(\sigma_1, ..., \sigma_j)$. Fix $x \in E_\alpha(B)$ and $0 < r < r_0$. The construction is similar to geometric construction done in chapter one, were we constructed Moran Covers: there exists a unique sequence $(\sigma_1, \sigma_2, ...) \in \mathbb{N}^{\mathbb{N}}$ such that $x \in J(\sigma_1, \sigma_2, ..., \sigma_k)$ for every $k \geq 1$ and such that

$$\tilde{G}(\sigma_1, ..., \sigma_{n-1}, \sigma_n) \leq r < \tilde{G}(\sigma_1, ..., \sigma_{n-1})$$

for some n large enough. This says that $B(x,r)$ can intersect just one interval $J(\sigma)$ with $\sigma \in D_n$, in this case, $J(\sigma_1, ..., \sigma_n)$.

In order to obtain a sharp bound for $\mu(B(x,r))$, it is necessary to analyze the cases when $n = n_k - 1$, $n = n_k - 2$ and $n \neq n_k - 1, n_k - 2$. This is due to the lack of uniformity of the definition of μ on the intervals $J(\sigma)$ for different lengths of σ.

We study now the case $n = n_k - 1$ for some k large enough. There are two subcases:

If $r \leq |I(\sigma_1, ..., \sigma_{n_k})|$: in this case, $B(x,r)$ can intersect at most four intervals $J(\sigma)$ with $\sigma \in D_{n_k}$. These intervals are $I(\sigma_1, ..., \sigma_{n_k} - 1)$, $I(\sigma_1, ..., \sigma_{n_k})$, $I(\sigma_1, ..., \sigma_{n_k+1})$ and $I(\sigma_1, ..., \sigma_{n_k} + 2)$. Then, we have that

$$\mu(B(x,r)) \leq 4\mu(J(\sigma_1, ..., \sigma_{n_k-1})).$$

Using the estimates for $\mu(J(\sigma))$ and $\tilde{G}(\sigma)$, we get

$$\begin{aligned}\mu(B(x,r)) &\leq 4 \cdot c_0 \cdot |J(\sigma_1, ..., \sigma_{n_k-1})|^{t-2\varepsilon} \\ &\leq 8 \cdot c_0 \cdot \alpha \cdot |\tilde{G}(\sigma_1, ..., \sigma_{n_k-1})|^{t-2\varepsilon} \\ &\leq 8 \cdot c_0 \cdot \alpha \cdot r^{t-2\varepsilon}.\end{aligned}$$

In the other cases, it is also possible to prove that

$$\mu(B(x,r)) \leq c_1 \cdot r^{t-2\varepsilon},$$

which implies that

$$\liminf_{r\to 0} \frac{\log \mu(B(x,r))}{\log r} \geq t - 2\varepsilon.$$

By Proposition 1.31, we conclude that $\dim_H E_\alpha(B) \geq t - 2\varepsilon = 2t - d_B(\alpha)$. Since $t < d_B(\alpha)$ is arbitrary, letting $t \to d_B(\alpha)$ we get that $\dim_H E(B) \geq \dim_H E_\alpha(B) \geq d_B(\alpha)$. Finally, letting $\alpha \to \infty$, we conclude $\dim_H E(B) \geq d_B$.

4.3 Proof of theorem 4.4

As a corollary of Theorem 4.7 and 3.30, we finally obtain:

Corollary 4.9. *The function $B \mapsto \dim_H(E(B))$ defined in the interval $(1/2, 1)$ is real analytic.*

Proof. The function $(t, B) \mapsto P(-t\log|BG'|)$ is real analytic on each variable on the range $t > 1/2$ and $B > 1$. The result follows from the implicit function theorem, once the non-degeneracy condition is verified. But note that applying the formula for the derivative of the pressure function we obtain

$$\frac{\partial}{\partial t} P(-t\log|BG'|)|_{(t_0,B_0)} = \int -t_0 \log|B_0G'| d\mu_{t_0,B_0} \neq 0,$$

where μ_{t_0,B_0} is the equilibrium measure of the potential $-t_0\log|B_0G'|$. ■

Corollary 4.10. *The function $B \mapsto \dim_H(E(B))$ is strictly decreasing.*

Proof. Just note that $\dim_H(E(B)) = s_B$ is defined as the solution of the equation

$$P(-t\log|G'|) = t\log B.$$

The result follows from the fact that the function $t \mapsto P(-t\log|G'|)$, when finite, is strictly decreasing. ■

Corollary 4.11. *We have that*

$$\lim_{B\to 1} s_B = \frac{1}{2} \text{ and } \lim_{B\to\infty} s_B = 1.$$

Proof. The result follows from the fact that $\lim_{t\to 1/2} P(-t\log|G'|) = \infty$ and that $P(-\log|G'|) = 0$ ■

This completes the description of the function $B \to \dim_H E(B)$.

4.4 What to do next?

We finish this work presenting possible future directions to improve this work and establishing additional connections with other problems.

The first objective is to find a proof of the bound $\dim_H E(B) \geq d_B$ in terms of the limit measures obtained from the equilibrium states of the approximated pressure. The idea should be to produce a measure supported in $E(B)$ by taking equilibrium measures in the procedure done in the first part of the proof of Theorem 4.4 and then taking a limit under certain conditions. We believe that this should produce a cleaner proof.

In a recent article, Seuret and Wang [SW15] considered the sets[1]

$$A(f) = \{x :\in \phi_{\omega_1} \circ ... \circ \phi_{\omega n}([0,1]) : |x - (\phi_{\omega_1 \circ ... \circ \phi_{\omega_n}})^{-1}(x)| < e^{-S_n f(x)} \text{ i.o.}\}$$

where $\Phi = \{\phi_i : i \in \Lambda\}$ is a conformal countable iterated function system and $f : [0,1] \to \mathbb{R}^+$ a function satisfying a bounded distortion property, and they showed that the Hausdorff dimension satisfy a Bowen type equation

$$d = \dim_H A(f) = \inf\{t \geq 0 : P(-t(\log|\Phi^{-1}| + f)) \leq 0\}.$$

In our setting, we get the sets

$$A'(B) = \{x \in [0,1] : |x - G^n(x)| < \frac{1}{B^n (G^n)'(x)} \text{ i.o.}\},$$

[1]The paper actually treats the multidimensional case, but for our purpose it is enough to consider the one dimensional case.

having dimension

$$d = \dim_H A'(B) = \inf\{t \geq 0 : P(-t \log |BG'|) \leq 0\} = d_B.$$

As noted in [SW15], the points in $A(f)$ are infinitely recurrent with weight f, and the result can be understood as a quantitative version of the Poincare's Recurrence Theorem. It is worth noting that the set $A'(B)$ have the same dimension as the set $E(B)$, so the natural question arises: is there any natural relation between these two sets? We intend to work on this question.

Bibliography

[AB09] Artur Avila and Jairo Bochi. On the subadditive ergodic theorem. *preprint*, 2009. www.mat.uc.cl/~jairo.bochi/docs/kingbirk.pdf.

[Bal00] Viviane Baladi. *Positive transfer operators and decay of correlations*, volume 16. World scientific, 2000.

[Bar08] Luis Barreira. *Dimension and recurrence in hyperbolic dynamics*, volume 272. Birkhauser, 2008.

[Ber11] Felix Bernstein. Über eine anwendung der mengenlehre auf ein aus der theorie der säkularen störungen herrührendes problem. *Mathematische Annalen*, 71(3):417–439, 1911.

[Bor12] Émile Borel. Sur un probleme de probabilités relatif aux fractions continues. *Mathematische Annalen*, 72(4):578–584, 1912.

[Bow75] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470. Springer, 1975.

[Bow79] Rufus Bowen. Hausdorff dimension of quasi-circles. *Publications Mathématiques de l'IHÉS*, 50:11–25, 1979.

[CE46] Arthur H. Copeland and Paul Erdös. Note on normal numbers. *Bulletin of the American Mathematical Society*, 52(10):857–860, 1946.

[Cha33] David G. Champernowne. The construction of decimals normal in the scale of ten. *J. London Math. Soc*, 8(4):254–260, 1933.

[DE52] Harald Davenport and Paul Erdös. Note on normal decimals. *Canadian J. Math*, 4:58–63, 1952.

[DK02] Karma Dajani and Cor Kraaikamp. *Ergodic theory of numbers*. Number 29. Cambridge University Press, 2002.

[EW13] Manfred Einsiedler and Thomas Ward. *Ergodic Theory*. Springer, 2013.

[Fal97] Kenneth Falconer. *Techniques in fractal geometry*, volume 3. Wiley Chichester (W. Sx.), 1997.

[Gau03] Carl F. Gauss. Tagebuch 1796–1814. *Math. Ann*, 57:1–34, 1903.

[Goo41] Irving J. Good. The fractional dimensional theory of continued fractions. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 37, pages 199–228. Cambridge Univ Press, 1941.

[HW79] Godfrey H. Hardy and Edward M. Wright. *An introduction to the theory of numbers*. Oxford University Press, 1979.

[Kat12] Tosio Kato. *Perturbation theory for linear operators*. Springer Science & Business Media, 2012.

[Khi63] Aleksandr Y. Khintchine. *Continued fractions*. 1963.

[May76] Dieter H Mayer. On a zeta function related to the continued fraction transformation. *Bulletin de la Societe Mathematique de France*, 104:195–203, 1976.

[May90] Dieter H Mayer. On the thermodynamic formalism for the gauss map. *Communications in mathematical physics*, 130(2):311–333, 1990.

[Pes08] Yakov B Pesin. *Dimension theory in dynamical systems: contemporary views and applications*. University of Chicago Press, 2008.

[PP90] William Parry and Mark Pollicott. *Zeta functions and the periodic orbit structure of hyperbolic dynamics*. Société Mathématique de France, 1990.

[Sar99] Omri M Sarig. Thermodynamic formalism for countable markov shifts. *Ergodic Theory and Dynamical Systems*, 19(06):1565–1593, 1999.

[SW15] Stéphane Seuret and B-W Wang. Quantitative recurrence properties in conformal iterated function systems. *Advances in Mathematics*, 280:472–505, 2015.

[VO15] Marcelo Viana and Krerley Oliveira. *Foundations of Ergodic Theory*. Cambridge University Press, 2015.

[Wal82] Peter Walters. *An introduction to ergodic theory*, volume 79. Springer-Verlag New York, 1982.

[WW08] Bao-Wei Wang and Jun Wu. Hausdorff dimension of certain sets arising in continued fraction expansions. *Advances in Mathematics*, 218(5):1319 – 1339, 2008.

Index