ENTROPY, PRESSURE AND LARGE DEVIATION

ARTUR LOPES
Instituto de Matemática
Universidade Federal do Rio Grande do Sul
91500 Porto Alegre RS
Brasil

ABSTRACT. We present a brief introduction to Ergodic Theory and equilibrium states of Thermodynamic Formalism. We also analyze Large deviation properties of the equilibrium states defined in Thermodynamic Formalism. Several problems related to Statistical Mechanics are consider.

1. Introduction

Our purpose in the first paragraphs of this text is to present the basic concepts of Ergodic Theory in the most simple way. We introduce the Ergodic Theorem of Birkhoff and the concept of *entropy* and *pressure*. Our final goal is to analyze several important problems related to Statistical Mechanics in the setting of Ergodic Theory.

We hope to present some of the main ideas of Ergodic Theory without too many technicalities. The relation between the concepts of *pressure* and *entropy* with the *free-energy* of Large Deviation Theory will be explored in the last paragraphs.

Given a space X, a probability P on X is a law that associates to each subset B of X a real value P(B). The value P(X) is assumed to be one. We also assume in the definition of probability that for any sequence $B_n, n \in \mathbb{N}$ of disjoint subsets of X (that is, $B_n \cap B_m = \emptyset$ for m different from n), the union of such sets, $\cup_{n\in\mathbb{N}} B_n$, satisfies $P(\cup_{n\in\mathbb{N}} B_n) = \sum_{n=0}^{\infty} P(B_n)$. Finally we require that P(A-B) = P(A)-P(B) for any subsets A and B of X, such that B is contained in A.

Unfortunately, in most cases one can not have all the above properties defined for all subsets of X. Therefore we define the probability P on a smaller family of

E. Goles and S. Martínez (eds.), Cellular Automata, Dynamical Systems and Neural Networks, 79–146.

subsets of X. In the present text this family of subsets is a σ-algebra $\mathcal{A}$. We refer the reader to any book on Real Analysis [16] for the precise definition of σ-algebra. In all the situations we will face in this text, the subsets B of X for which we want to assign a value P(B), will be elements of the family $\mathcal{A}$. Therefore we will not have problems with sets B whose probability P(B) is not well defined.

Ya. Sinai define Ergodic Theory in the following way "The basic problems in Ergodic Theory consist of the study of the statistical properties of the groups of motions of non-random objects". The group of motions we are interested in this text is the set of iterates of a map T from a metrical space X into itself, that is $T, T^2, T^3, ..., T^n, ...$ What properties one can expect for the iterates of a general point x, in other words, what results can be stated for the set $\{x, T(x), T^2(x), T^3(x), ..., T^n(x), ...\}$? We will suppose there exist a certain probability P involved in the problem and we will be interested in properties that are true for every x in X outside a negligible set A of probability zero (that is, P(A) = 0).

2. Birkhoff's Ergodic Theorem

Let $\Omega = \{0,1\}^{\mathrm{N}}$ be the set of sequences of 0's and 1's, that is, $z \in \Omega$ if $z = (z_0, z_1, z_2, ..., z_n, ...)$ where $z_i \in \{0,1\}$ for all $i \in \mathrm{N}$.

We call this set the Bernoulli space. We can think of this set as the set of events of tossing a coin infinitely many times, in which we associate head with 0 and tail with 1. For example, (0,1,0,1,0,1,...) is the event in which we have alternatedly head and tail, beginning with a tail at time 0, that is, $z_0 = 0$.

A cylinder (or a parallepiped) A is a subset of Ω defined by a finite specification of elements; the set $A = \{(0,1,1,0,1, z_5, z_6, ..., z_n, ...) \mid z_i \in \{0,1\},\ i \geq 5\}$, for example, is a cylinder, which we denote by $(\overline{0,1,1,0,1})$. In general a cylinder is given by

$$(\overline{a_0, a_1, ..., a_m}) = \{(a_0, a_1, ..., a_m, z_{m+1}, z_{m+2}, ..., z_{m+n}, ...) \mid z_i \in \{0,1\}, i \geq m+1\},$$

where n is fixed and $a_0, a_1, ..., a_m$ belonging to $\{0,1\}$ are also fixed. We should think of $(\overline{0,1,1,0,1})$ as the event of tossing a coin and have successively head, tail, tail, head and tail and no specification about the rest of the other tossings.

We would like to define a probability on the set $X = \Omega$. First we will assign values P(A) for the elementary subsets: the cylinders A. After that we will extend this probability to more complicated subsets B, as countable unions and intersection of cylinders A, and then to more general and elaborated specifications. The family of subsets B for which we will be able to assign the value P(B) will be called later the σ- algebra $\mathcal{A}$.

The probability of having in order head, tail, tail, head and tail when we toss the coin depends on the probability of having head or tail at each time.

Suppose that p_0, p_1 are two numbers such that $p_0, p_1 \geq 0$ and $p_0 + p_1 = 1$. Suppose that each time we toss the coin we have probability p_0 of having head (or 0) and probability p_1 of having tail (or 1). If we suppose the tossings are independent, the probability of having head, tail, tail, head and tail is $p_0^2 p_1^3$. Therefore it is natural to give probability $p_0^2 p_1^3$ to the set $A = (\overline{0,1,1,0,1})$.

In the same way we can define $P(\overline{a_0, a_1, ..., a_n}) = p_0^q p_1^m$ where q is the number of 0's in the sequence $\{a_0, a_1, ..., a_n\}$ and m is the number of 1's in the sequence $\{a_0, a_1, ..., a_n\}$. In this way we obtain a well defined measure on any cylinder. We define cylinders more generally by a finite number of specifications but perhaps not in sequence, for instance $\{(0, z_1, 1, z_3, 0, z_5, z_6, z_7, ...) \mid z_1 \in \{0,1\}, z_3 \in \{0,1\}, z_i \in \{0,1\}, i \geq 5\}$ is a cylinder. We will present the precise definition of the general cylinder later. Using well known ideas of measure theory one can extend this probability P to the σ-algebra generated by all cylinders (see[13]).

In this way if we denote this σ-algebra by $\mathcal{A}$ and the probability by P we have $(\Omega, \mathcal{A}, P)$ as a well defined measure space. Note that $P(\Omega) = 1$, because $1 = p_0 + p_1 = P(\overline{0}) + P(\overline{1}) = P(\Omega)$. Remember that $(\overline{0}) = \{(0, z_1, z_2, z_3, ...) \mid z_i \in \{0,1\}$ for $i \geq 1\}$ and $(\overline{1}) = \{(1, z_1, z_2, z_3, ...) \mid z_i \in \{0,1\}$ for $i \geq 1\}$. We say that the coin is fair if $p_0 = 0.5$ and $p_1 = 0.5$. It is a well known observable fact that if we toss the coin a very large number of times, like 200 times, we will obtain more or less half times head and half times tail. It is also reasonable to suppose that if the coin has probability p_0 to obtain head (or 0) and p_1 of having a tail (or 1) then if we toss the coin 200 times, we will obtain more or less 200 p_0 times head and 200 p_1 times tail. In Probability Theory this is known as the Law of Large Numbers [1].

The Ergodic Theorem of Birkhoff is a quite general theorem that will assure that the above result is true. We explain now more carefully the meaning of the Ergodic Theorem.

Note first that P depends on p_0 and p_1. The Birkhoff Ergodic Theorem (it will be formally stated later) claims that there exists a set A such that $P(A) = 1$, and such that for all $z \in A$, where $z = (z_0, z_1, z_2, ..., z_n, ...)$, we have that

$$p_0 = \lim_{n\to\infty} \frac{1}{n} \text{ (cardinal of heads among } z_0, z_1, ..., z_{n-1})$$

and

$$p_1 = \lim_{n\to\infty} \frac{1}{n} \text{ (cardinal of tails among } z_0, z_1, ..., z_{n-1}).$$

The above result claims that the mean value of heads that appears in tossing the coin n times converges to p_0. Before we state the Birkhoff Ergodic Theorem in precise mathematical terms we need to introduce the concepts of shift and invariant measure.

The shift map σ from Ω to Ω is the map such that for

$$z = (z_0, z_1, z_2, ..., z_n, ...)$$

we have

$$\sigma(z) = (z_1, z_2, z_3, ..., z_{n+1}, ...).$$

Therefore we can express the number of tails we have tossing the coin n times (as expressed by $z \in \Omega$) by

$$\sum_{j=0}^{n-1} I_A(\sigma^j(z)),$$

where I_A is the indicator of $A = (\overline{1})$, that is, $I_A(z) = 1$ for $z \in (\overline{1})$ and $I_A(z) = 0$ for $z \notin (\overline{1})$; in other terms, $I_A(z) = 1$ if $z_0 = 1$ and $I_A(z) = 0$ if $z_0 = 0$. In the same way,

$$\sum_{j=0}^{n-1} I_B(\sigma^j(z))$$

is the number of heads we have for the event z of tossing the coin n times; here $I_B(z)$ is the indicator of the set $B = (\overline{0})$, that is, $I_B(z) = 0$ if $z \notin (\overline{0})$ and $I_B(z) = 1$ if $z \in (\overline{0})$.

In this way we can see that the shift helps us to formulate the number of heads and tails in a simple expression.

Definition 2.1. The set $\{z, \sigma(z), \sigma^2(z), ..., \sigma^n(z), ...\}$ is called the *orbit* of z under the shift map σ. The element $\sigma^n(z)$ is called the $n^{\underline{th}}$ *iterate* of z.

We will call the Borel σ-Algebra of Ω the σ-Algebra generated by the cylinders. The Borel σ-Algebra of $\mathbf{R}$ is the σ-Algebra generated by the finite intervals (see [16]). We say that f from X to $\mathbf{R}$ is measurable if for each set A in the σ-Algebra of Borel of $\mathbf{R}$, $f^{-1}(A)$ is in the σ-Algebra of Borel of Ω.

Given a certain measurable map $\phi : \Omega \to \mathbf{R}$, the mean value of ϕ on z up to the $n^{\underline{th}}$ iterate is

$$\frac{1}{n}\sum_{j=0}^{n-1} \phi(\sigma^j(z)).$$

In this way, for $\phi = I_{(\overline{0})}$, the mean value of $I_{(\overline{0})}$ on z, up to the $n^{\underline{th}}$-iterate is the mean value of times we obtain a head, tossing the coin n times. In the case of the fair coin, that is, $p_0 = 0.5 = p_1$, and $\phi = I_{(\overline{0})}$, one should expect that the mean number of heads should converge to 0.5 when n goes to infinity.

We will be interested in obtaining the limit of these mean values as n goes to infinity, that is,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \phi(\sigma^j(z))$$

for P-almost all points z.

First we need to introduce the concept of invariant measure.

Definition 2.2. Given $(X, \mathcal{A}, \mu)$, where X is a set, $\mathcal{A}$ is a σ-algebra on X and μ is a measure on this σ-algebra, we consider T a measurable map from X to X (that is $T^{-1}(A) \in \mathcal{A}$ for all $A \in \mathcal{A}$), and say that μ is *invariant* for T if for all measurable sets $A \in \mathcal{A}$, $\mu(T^{-1}(A)) = \mu(A)$.

Invariant measures appear very naturally in several areas of Mathematics as for instance, in Hamiltonian Mechanics, Geometry and Number Theory.

We now show that the probability P (depending on p_0 and p_1) introduced before is invariant for the shift.

Proposition 2.1. The probability P is always invariant for the shift map $\sigma : \Omega \to \Omega$.

Proof. It is enough to show that $P(T^{-1}(A)) = P(A)$ for the sets A that are generators (the cylinders) of the σ-algebra.

Consider $A = (\overline{a_0, a_1, ..., a_n})$ a cylinder, then

$$\begin{aligned}
P(T^{-1}(A)) &= P((\overline{0, a_0, a_1, ..., a_n}) \cup (\overline{1, a_0, a_1, ..., a_n})) \\
&= P((\overline{0, a_0, a_1, ..., a_n})) + P((\overline{1, a_0, a_1, ..., a_n})) \\
&= p_0\, p_0^{\sum_{j=0}^{n-1} I_{(\overline{0})}\sigma^j(z)}\, p_1^{\sum_{j=0}^{n-1} I_{(\overline{1})}\sigma^j(z)} + p_1\, p_0^{\sum_{j=0}^{n-1} I_{(\overline{0})}\sigma^j(z)}\, p_1^{\sum_{j=0}^{n-1} I_{(\overline{1})}\sigma^j(z)} \\
&= (p_0 + p_1)\, p_0^{\sum_{j=0}^{n-1} I_{(\overline{0})}\sigma^j(z)}\, p_1^{\sum_{j=0}^{n-1} I_{(\overline{1})}\sigma^j(z)} \\
&= p_0^{\sum_{j=0}^{n-1} I_{(\overline{0})}\sigma^j(z)}\, p_1^{\sum_{j=0}^{n-1} I_{(\overline{1})}\sigma^j(z)} = P((\overline{a_0, a_1, ..., a_n})) = P(A). \quad \blacksquare
\end{aligned}$$

Notation. We introduce the following notation: $\mathcal{M}(T)$ is the set of all invariant probabilities μ for the measurable map $T : X \to X$.

Therefore $\mathcal{M}(\sigma)$ denotes the set of all invariant probabilities for σ. For each p_0, p_1, such that $p_0 + p_1 = 1$, $p_0, p_1 \geq 0$, we have that the corresponding P belongs to $\mathcal{M}(\sigma)$ as was shown in the proposition above. There exist of course other probabilities $\mu \in \mathcal{M}(\sigma)$ that are not of the form P.

The set of probabilities $\mathcal{M}(T)$ is a convex simplex in the set of all measures on the σ-algebra $\mathcal{A}$ of the set X. It is well known in Convex Analysis that the points in the corners of the convex play a very important role.

Definition 2.3. A point x in a convex set $\mathcal{C}$ is called *extremal* if x cannot be expressed as $x = \lambda y + (1 - \lambda)z$, where y and z are in $\mathcal{C}$, x different from y and z and $0 < \lambda < 1$.

It is possible to show that the probability measures that are extremals for the set of invariant probabilities $\mathcal{C} = \mathcal{M}(T)$ are the ergodic probabilities.

We define ergodic measures however by a different property.

Definition 2.4. We say that $\mu \in \mathcal{M}(T)$ is *ergodic* if for all $A \in \mathcal{A}$ such that $T^{-1}(A) = A$ either $\mu(A) = 0$ or $\mu(A) = 1$.

The above definition means that for an ergodic measure the action of the measurable map T on any non trivial set $A \in \mathcal{A}$ (a trivial set being equal to $\emptyset$ or X up to a set of μ-measure zero) is so random that it can not leave the set A invariant; in other words the set A has to spread around the set X under iteration of T.

Note that the empty set $\emptyset$ and the total set Ω are always invariant, but they have respectively measure 0 and 1.

Remark. It can be shown that the shift with the invariant probability P defined above is ergodic [18].

In Ergodic Theory, most of the proofs of general results follow the recipe: first prove the result for ergodic measures and then use the ergodic decomposition theorem [13] to extend the result for other kind of measures.

Notation. Given a probability μ on the set X, we will say that a property happen *μ-almost everywhere*, if there exist a subset A contained in X, such that $\mu(\mathrm{A}) = 1$ and the property is true, for all z in the set A.

Notation. We will denote by $\mathcal{L}^1(\mu)$ the set of measurable functions f from X to $\mathbf{R}$ such that $\int f(z) d\mu(z)$ exist and is finite.

Now we can state Birkhoff's Ergodic Theorem.

Theorem 2.1. (Birkhoff) - Let $(X, \mathcal{A}, \mu)$ be a probability space and $T : X \to X$ a measurable transformation that preserves μ, that is, $\mu \in \mathcal{M}(T)$ and suppose that μ is ergodic. Then for any $f \in \mathcal{L}^1(\mu)$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(z)) = \int f(x) d\mu(x) \tag{1}$$

for $z \in X$, μ-almost everywhere.

The above result essentially claims that for ergodic measures, spatial mean (the right hand side of (1)) is equal to temporal mean (the left hand side of (1)) for almost every point z. Therefore, in this case, in order to compute an integral, one has to estimate the value of a series. In several practical situations this property brings a simplification to the problem of estimating an integral.

When we consider $T = \sigma$, $P = \mu$ and $X = \Omega$ in the Bernoulli shift example we mentioned before, then considering $f(x) = I_{(\overline{0})}(x)$, we get

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} I_{(\overline{0})}(\sigma^j(z)) = \int I_{(\overline{0})}(x) d\,P(x) = \mu(\overline{0}) = p_0,$$

(for P-almost every z), which we mentioned before in our reasoning. This theorem therefore is a very general result that can, as a particular case, assure the validity of the Strong Law of Large Numbers.

In the case $p_0 = 0.5 = p_1$, the fair coin, the event of obtaining head every time from 0 to infinity (that is, (1,1,1,1,1,...)) is rare (has P-measure zero). For a set A of measure one the events $(z_0, z_1, ..., z_n, ..) \in A$ are such that head and tail appear with the same frequency.

The questions that people in Probability and Ergodic Theory are concerned with are not of deterministic nature. The statements that are relevant and pertinent, are the ones about events that happen with probability one. In other words, the statements abouts sets A such that $\mu(A) = 1$. Sets of measure zero are consider negligible.

The Birkhoff Ergodic Theorem is one of the most celebrated theorems of Mathematics and was inspired by Statistical Mechanics, more specifically by the billiard ball model, which is a model for a particle reflecting on the walls of a closed compartment [13].

We now state a more general version of Birkhoff's Ergodic Theorem, without the assumption that the measure is ergodic.

Theorem 2.2. (Birkhoff) - Let $(X, \mathcal{A}, \mu)$ be a probability space and $\mu \in \mathcal{M}(T)$, where T is measurable, $T : X \to X$. Then for any $f \in \mathcal{L}^1(\mu)$ the limit

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

exists for $z \in X$, μ-almost everywhere. If the limit is denoted by

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

then it is also true that

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

Note that the difference of the last result to the previous one is that in the case the measure is ergodic $\tilde{f}$ is constant μ - almost everywhere.

The Bernoulli space Ω can be equipped with a distance $d_\theta : \Omega \times \Omega \to \mathbf{R}$ in the following way: for a fixed value θ with $0 < \theta < 1$, we define the metric $d_\theta(x, y) = \theta^N$, (where N is the largest natural number such that $x_i = y_i$, $|i| < N$) if x is different from y. When x is equal to y then we define the distance to be zero. If we define open sets Ω in the usual way (product topology) we have that the σ-algebra generated by the cylinders is the σ-algebra of Borel, since the cylinders form a basis for the topology of Ω.

As an example consider $\theta = 0.3$, z=(1,1,0,1,0,0,1,...) and $\epsilon = 0.0081 = 0.3^4$, then is easy to see that $B(z, \epsilon)$ (the open ball of center z and radius ϵ) is equal to the cylinder $(\overline{1, 1, 0, 1})$.

Note that the indicator function I_A is continuous if A is a cylinder.

In the rest of this text we will consider a certain fixed value θ and denote by d the metric associated with it.

Definition 2.5. A map T from a metric space (X, d) into itself is *expanding* if there exist $\lambda > 1$ such that for any x, there exist $\epsilon > 0$ such that $\forall y \in B(x, \epsilon)$, $d(T(x), T(y)) \geq \lambda d(x, y)$.

Note that if $d_\theta(x, y) = \alpha$, $x, y \in \Omega$, then $d_\theta(\sigma(x), \sigma(y)) = \alpha\theta^{-1} = \theta^{-1} d_\theta(x, y)$. Therefore the Bernoulli shift σ is expanding with the value $\lambda = \theta^{-1}$ in the notation of above definition.

It is also necessary to introduce the two-sided Bernoulli shift as the set $\Omega = \{0,1\}^{\mathbb{Z}}$ of elements of the form

$$z = (\ldots, z_{-n}, \ldots, z_{-3}, z_{-2}, z_{-1}, z_0, z_1, z_2, \ldots, z_n, \ldots).$$

The shift $\sigma : \Omega \to \Omega$ is defined in the same way,

$$\sigma(z_i) = (z_{i+1})$$

when $z = (z_i)$. For example for $z = (z_i)$ where $z_i = 1$ for i even and $z_i = 0$ for i odd, $\sigma(z) = (z_{i+1}) = (y_i)$ where $y_i = 1$ for i odd and $y_i = 0$ for i even. Note that $\sigma^2(z) = z$ in this case.

Definition 2.6. For a general map $T : X \to X$, the *orbit of* x is the set $\{x, T(x), T^2(x), \ldots, T^n(x), \ldots\}$. We say x is *periodic* of period n if $n \geq 1$ is the smallest possible natural number such that $T^n(x) = x$.

Therefore in the example given above z is a periodic point of period 2. The orbit of z in this case is $\{z, T(z)\}$. Note that the shift in the one-sided Bernoulli space is not one-to-one, but the shift in the two sided Bernoulli space is.

Consider a finite set (an alphabet) of k symbols $\{0, 1, \ldots, k-1\}$ and a probability μ_0 on this finite set, that is,

$$\mu_0(i) = p_i$$

and

$$\sum_{i=0}^{k-1} p_i = 1.$$

Consider also the set of sequences of these symbols, that is, the set of sequences $z = (z_0, z_1, z_2, \ldots, z_n, \ldots)$ where $z_i \in \{0, 1, \ldots, k-1\}$. We will again denote by Ω the set of all these sequences. Sometimes we denote by $z : \mathbb{N} \to \{0, 1, \ldots, k-1\}$ an element of Ω and $z(n)$ by z_n. The shift on Ω is defined in the same way as before, $\sigma : \Omega \to \Omega$ is such that for $z = (z_0, z_1, z_2, \ldots, z_{n+1}, \ldots) \in \Omega$, $\sigma(z) = (z_1, z_2, \ldots, z_n, \ldots) \in \Omega$.

Definition 2.7. Given finite subsets $A_0, A_1, ..., A_m$ of $\{0, 1, ..., k-1\}$ and $j \in \mathrm{N}$, we define the *cylinder* $C(j, A_0, ..., A_m)$ by

$$C(j, A_0, ..., A_m) = \{x \in \Omega \mid x(j+i) \in A_i,\ 0 \leq i \leq m\}.$$

Disjoint unions of cylinders form an algebra that generates a σ-algebra $\mathcal{A}$ on Ω. Moreover, given the probability μ_0 on $\{0, 1, ..., k-1\}$, there exists a unique probability P on the σ-algebra $\mathcal{A}$ (the product measure associated to μ_0) such that for every cylinder:

$$P(C(j, A_0, ..., A_m)) = \prod_{i=0}^{m} \mu_0(A_i).$$

The above definition is the precise definition of a general cylinder we promised before.

We define in Ω a metric in the same way as always: for a fixed θ, $0 < \theta < 1$, we define $d_\theta(x, y) = \theta^N$ where N is the largest N such that $x_i = y_i$ for all $0 \leq i \leq N$ for x different from y and zero otherwise. It is easy to see that d_θ has all the properties of a metric.

These definitions, of course, extend the previous ones defined for the shift in two symbols. The system defined above is also called the one-sided Bernoulli shift on

$$\Omega = B(p_0, p_1, ..., p_{k-1})$$

with probability $P(p_0, p_1, ..., p_{k-1})$ on Ω.

The two-sided shift is the set of all functions $z : \mathrm{Z} \to \{0, 1, ..., k-1\}$ and in the same way as before $\sigma(x)(i) = x(i+1)$ is by definition the shift map on this space. The cylinders are defined in a similar way: given subsets $A_0, ..., A_n$ of $\{0, 1, ..., k-1\}$ and $j \in \mathrm{Z}$ (remember that $j \in \mathrm{N}$ in the one-sided shift case)

$$C(j, A_0, ..., A_m) = \{z \in \Omega \mid z(j+i) \in A_i,\ 0 \leq i \leq m\}.$$

In the same way as before we consider the σ-algebra generated by the cylinders. Moreover, given a probability μ_0 on $\{0, 1, ..., k-1\}$ such that $\mu_0(i) = p_i$, $i \in \{0, ..., k-1\}$, $\sum_{i=0}^{k-1} p_i = 1$, then we define $P(C(j, A_0, ..., A_m)) = \Pi_{i=0}^{m} \mu_0(A_i)$. For

$0 < \theta < 1$ fixed, the metric we will consider on Ω is $d_\theta(x,y) = \theta^N$ where N is the largest N such that $x_i = y_i$ for all i such $|i| \leq$ N if x is different from y and zero otherwise.

We will call such system the two-sided Bernoulli shift on

$$\Omega = B(p_0, p_1, ..., p_{k-1})$$

with probability $P(p_0, p_1, ..., p_{k-1})$ on Ω.

The main difference between the one-sided shift and the two-sided shift is that the latter is one-to-one. With the one-sided shift, any $z \in \Omega = B(p_0, p_1, ..., p_{k-1})$ has k preimages, that is, if $z = (z_0, z_1, ..., z_n, ...)$, then

$$x_0 = (0, z_0, z_1, ..., z_n, ...),$$

$$x_1 = (1, z_0, z_1, ..., z_n, ...),$$

$$...$$

and

$$x_{k-1} = (k-1, z_0, z_1, ..., z_n, ...)$$

are such that $\sigma(x_i) = z, i \in \{0, ..., k-1\}$, that is, $\sigma^{-1}(z) = \{x_0, ..., x_{k-1}\}$.

More generally for $z = (z_0, z_1, ...)$, the set of solutions x of $\sigma^n(x) = z$ is the set of points x of the form

$$x = (x_0, x_1, ..., x_{n-1}, z_0, z_1, ...)$$

0, (k-1)

where $x_0, x_1, ..., x_{n-1} \in \{1, 2, ...k\}$ are arbitrary. Therefore the cardinality of the set of such solutions x is k^n.

Notation. We call the set of such points, the *pre-images* of z by σ.

Periodic orbits for σ are also easy to find. The set of all periodic orbits of period n is obtained in the following way: take $z_0, z_1, .., z_{n-1}$ in all possible ways such that $z_i \in \{0, 1\}$, $i \in \{1, 2, ..., n-1\}$. For each one of these $z_0, z_1, ...z_{n-1}$

repeat the block infinitely many times, in order to obtain the set of all x such that $\sigma^n(x) = x$, where

$$x = (z_0, z_1, ..., z_{n-1}, z_0, z_1, ..., z_{n-1}, z_0, z_1, ..., z_{n-1}, ...).$$

Remark. Note that the cardinality of the set of solutions z of $\sigma^n(z) = z$ and the cardinality of the set of solutions x of $\sigma^n(x) = z$ is the same and equal to k^n. In fact, the procedure of finding the set of solutions is quite similar in both cases.

Proposition 2.2. The set of all periodic points for the shift is dense in Ω with the d_θ metric.

Proof. Given $z = (z_i)_{i \in \mathbb{N}}, z_i \in \{0, ..., k-1\}$, and $\xi > 0$, take N such that $\theta^N < \xi$. Now define x as the successive repetition of the string $(z_0, z_1, ..., z_N)$, that is,

$$x = (z_0, z_1, z_2, ..., z_N, z_0, z_1, z_2, ..., z_N, z_0, z_1, z_2, ..., z_N, ...).$$

Then $d_\theta(z, x) < \theta^N < \xi$ and $T^{N+1}(x) = x$, that is, x is a periodic point of period at most $(N+1)$ and ξ close to z. This proves the proposition. ∎

Remark. A similar result for the preimages of a certain point z can be obtained (the proof is basically the same), that is: any $y \in X$ can be aproximated by preimages of z.

Note that the temporal mean $\tilde{f}(z)$ of f (in Birkhoff's Theorem) at a point z belonging to a periodic orbit, is the mean value of f in the orbit of z. Therefore, in most cases (but not all cases, as we can see below), the periodic orbits have to be excluded from the set A of measure one mentioned in Birkhoff's Theorem.

In an extensive number of cases in Dynamical Systems the periodic orbits are dense in the region where the dynamics is concentrated [6]. Periodic orbits are extremely important for understanding the dynamics and the ergodic properties of a measure μ even if they can have μ - measure zero.

There exist invariant probabilities that are finite sums of Dirac measures in $\mathcal{M}(T)$, but they have to be concentrated on periodic orbits because of the invariance.

For example, the measure μ such that:

$$\mu((001001001001...)) = 1/3$$

$$\mu((010010010010...)) = 1/3$$

$$\mu((100100100100...)) = 1/3$$

is invariant and has support on a periodic orbit of period 3.

The space X we consider in this text will always be a compact metric space with metric d. We also denote by $C(X)$ the set of continuous functions on X taking values in $\mathbf{R}$. We will consider in $C(X)$ the supremum norm, that is, $\|f\| = \sup\{|\, f(x) \,|\mid x \in X\}$.

Notation. We will denote by $\mathcal{M}(X)$ the set of all probabilities on the Borel σ-algebra of X.

Notation. A law η such that for each set A in the σ-Algebra of Borel of X, $\eta(A)$ is a real number (not necessarily positive) or is equal to ∞, and such that:

$$\text{a)}\quad \eta(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \eta(A_i)$$

when the A_i are disjoint (that is $A_i \cap A_j = \emptyset$ for $i \neq j$),

$$\text{b)}\quad \eta(\emptyset) = 0$$

$$\text{c)}\quad \eta(A - B) = \eta(A) - \eta(B)$$

when $B \subset A$, is called a signed measure. We denote by $\mathcal{S}(X)$ the set of all signed measures on the Borel σ-algebra of X.

Example. For the set $X = \mathbf{R}$, given a continuous function $\Phi(x)$ (not necessarilly positive and not necessarily integrable), the law $\eta(A) = \int_A \Phi(x)dx$ is a signed-measure on X.

There exist signed-measures on X= $\mathbf{R}$ that are not of the above form.

Given a certain normed space V, the dual of V, denoted by V^*, is the set of all continuous linear functionals on V, that is, the set of all functionals $\mathcal{L} : V \to \mathbf{R}$ that are linear and continuous. The following theorem claims that the dual of the set $C(X)$ is the space $\mathcal{S}(X)$ [16].

Theorem 2.3. (Riesz) - Let $\mathcal{L} : C(X) \to \mathbf{R}$ be a continuous linear functional. Then there exists a unique $v \in \mathcal{S}(X)$ such that $\mathcal{L}(f) = \int f dv(u)$ for any $f \in C(X)$.

Corollary 2.1. If $\mathcal{L}$ is positive (that is, for any $f \in C(X)$, $\mathcal{L}(f) \geq 0$ if $f \geq 0$) and if $\mathcal{L}(1) = 1$, then there exists a unique probability $\mu \in \mathcal{M}(X)$ such that $\mathcal{L}(f) = \int f d\mu$ for any $f \in C(X)$.

Definition 2.8. Given $T : X \to Y$ measurable and $v \in \mathcal{M}(X)$, we define $T^*(v) = \omega$ as the unique measure $w \in \mathcal{M}(Y)$ such that $\int (f \circ T)(x) dv(x) = \int f(x) dw(x)$ for any $f \in C(X)$.

The measure w always exists and is well defined by Riesz's Theorem applied to $\mathcal{L}(f) = \int (f \circ S)(x) dv(x)$. The measure w is usually called *the pull back of the measure* v by the map S.

It is easily obtained from well known properties about approximation of continuons functions by step functions (finite sums of indicators with different weights) and vice-versa [16] that 1) and 2) below are equivalent:

1) for any Borel set A,

$$v(S^{-1}(A)) = \int (I_A \circ S)(x) dv(x) = \int I_A(x) dw(x) = w(A).$$

2) for any $f \in C(X)$,

$$\int (f \circ S)(x) dv(x) = \int f(x) dw(x).$$

A particular important case is when $X = Y$ and $T : X \to X$. In this case $w = T^*(v)$ is also a measure on X.

From the above considerations we can state:

Proposition 2.3. - $\mu \in \mathcal{M}(T)$ if and only if $T^*(\mu) = \mu$.

One would like to say that a sequence of measures μ_n converges to μ if and only if, for any Borel set A, the sequence $\mu_n(A)$ converges to $\mu(A)$. This is almost true. One has to suppose that the boundary of the set A has μ - measure zero and then the claim is true [16]. The more usefull definition of convergence is in terms of the action of the measures on the continuous functions:

Definition 2.9. We say that a sequence $\mu_n \in \mathcal{M}(X)$ *converges weakly* to a probability μ if for any continuous function $f : X \to \mathbf{R}$ we have that

$$\lim_{n\to\infty} \int f(x)d\mu_n(x) = \int f(x)d\mu(x).$$

If X is a compact metric space, the space $\mathcal{M}(X)$ of all probabilities is weakly sequentially-compact, that is, any sequence $\mu_n \in \mathcal{M}(X)$ has a convergent subsequence to an element $\mu \in \mathcal{M}(X)$ [16]. The set $\mathcal{M}(T)$ is also weakly sequentially-compact.

Definition 2.10. The *Dirac Delta measure* at the point z is by definition the probability measure that associates measure one to each Borel set that contains z and has measure zero otherwise. We will denote such a probability by δ_z.

It is well known that for a continuous function f and $z \in X$, the value $\int f(x)d\delta_z(x) = f(z)$.

Given the above definitions, the Ergodic Theorem of Birkhoff can be stated in the following way:

Theorem 2.4. Let $(X, \mathcal{A}, \mu)$ be a probability space, $T : X \to X$ a measurable transformation that preserves μ and suppose μ is ergodic. Then

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j(z)} \tag{2}$$

for μ almost every z.

Definition 2.11. The right hand side of the above equality is called the *empirical measure* . [7]

Definition 2.12. The *support of a measure* μ defined on X is the set of points $x \in X$ such that for any ϵ greater than zero the measure $\mu(B(x,\epsilon))$ of the ball of center x and radius ϵ is strictly positive.

Given a measure μ on X, in terms of Birkhoff's Theorem, there is no important information outside the support of the measure.

The above result shows that the support of two different ergodic measures have to be disjoint.

3. Entropy

Let X be a compact metric space with a metric $d: X \times X \to \mathbf{R}$ and $T: X \to X$ a transformation preserving the measure $\mu \in \mathcal{M}(T)$ defined on the Borel σ-algebra of X.

The dynamic ball $B(z,n,\xi)$, for $z \in X$, $n \in \mathrm{N}$ and $\xi > 0$ is by definition the set $B(z,n,\xi) = \{y \in X \mid d(T^j(z), T^j(y)) < \xi$ for all $0 \leq j \leq n-1\}$. One could think that one has a microscope that is able to detect that two points $x, y \in X$ are distinct if they are ξ appart, that is, $d(x,y) > \xi$. Therefore $B(z,n,\xi)$ is the set of points we are not able to distinguish from z performing n iterations. The value $\mu(B(z,n,\xi))$ gives the amount of indeterminacy after we perform $n-1$ iterations of the map T on the point z.

For z and ξ fixed and increasing n, the sets $B(z,n,\xi)$ decrease, that is, for $m \geq n$, $B(z,n,\xi) \supset B(z,m,\xi)$. When n goes to infinity, $B(z,n,\xi)$ converges to the set $\{z\}$ in the nice cases. In this case, if also $\mu(\{z\}) = 0$, then $\mu(B(z,n,\xi))$ will converge to zero, when n goes to infinity. One would like in this case to express the exponential velocity of decreasing in the form $\mu(B(z,n,\xi)) \approx \lambda^n$ for a certain value λ with $0 < \lambda < 1$, when ξ is very small. Writing λ as $e^{-h(\mu)}$, $h(\mu)$ will be what we call later the *entropy* of μ. The entropy of a measure will determine therefore the exponential velocity of decreasing of the indeterminacy of the system after iterations of the map T.

Theorem 3.1. (Brin-Katok) [4] - Suppose μ is ergodic for the transformation T on $(X, \mathcal{A}, \mu)$ and consider d a metric on the compact set X. Then the two limits

$$\lim_{\xi\to 0}\left(-\limsup_{n\to\infty}\frac{1}{n}\log\,\mu(B(z,n,\xi))\right)=\lim_{\xi\to 0}\left(-\liminf_{n\to\infty}\frac{1}{n}\log\,\mu(B(z,n,\xi))\right) \qquad (3)$$

exist and do not depend on z for μ- almost every point z in X.

Remark. By definition, given a sequence $(a_n), n \in \mathrm{N}$ of real numbers we call $\limsup_{n-\infty} a_n$, the supremum of the set of limits of convergent subsequences of the sequence (a_n). The definition of liminf is analogous. The reason for introducing this definition is that not all sequences (a_n) converge (therefore $\lim_{n\to\infty} a_n$ has no meaning), but for bounded sequences, the limsup and liminf will always exist. A sequence (a_n) converge, if and only if, the limsup and the liminf are equal (and off course, they are equal to the limit). In the above theorem, not always the sequence $a_n = \frac{1}{n}\log\mu(B(z,n,\epsilon))$ will converge. The limsup and liminf will exist in any case.

Definition 3.1. For an invariant ergodic measure $\mu \in \mathcal{M}(T)$ we define the *entropy* of μ as the value

$$h(\mu) = -\lim_{\xi\to 0}\left(\limsup_{n\to\infty}\frac{1}{n}\log\mu(B(z,n,\xi))\right),$$

where z was chosen in a set of measure one satisfying the above Theorem.

Note that we could define alternatively the entropy by the lim inf (see Theorem 3.1).

Later on we will define the entropy of a measure $\mu \in \mathcal{M}(T)$ when μ is not ergodic.

Note that the larger the entropy of the measure, the faster will be the decreasing of indeterminacy of the system. Therefore larger the entropy, more chaotic the system is.

Example. A trivial example where we can compute the entropy is the following: consider a periodic point x of period n, and the probability $\mu = \sum_{j=0}^{n-1} \frac{1}{n}\delta_{T^j(x)}$. It is easy to see that this measure μ is ergodic and that the entropy $h(\mu) = 0$.

The above example is in fact not exactly random or chaotic, but, in some sense, totally deterministic.

Proposition 3.1. The entropy of the probability $\mu = P(p_0, p_1)$, with $p_0, p_1 > 0$ and $p_0 + p_1 = 1$, invariant for the shift on $\Omega = B(p_0, p_1)$, is

$$-p_0 \log p_0 - p_1 \log p_1. \tag{4}$$

Proof. As we mention before, it can be shown that the probability $P(p_0, p_1)$ under the action of the shift is ergodic (see Remark after Definition 2.4).

Consider $z \in \Omega$ in a set A of P-measure one satisfying the Birkhoff Ergodic Theorem. That is, for any $f \in C(X)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(\sigma^j(z)) = \int f(x) d\,P(x).$$

The intersection of A with the set of full measure of the Definition 3.1 will also have measure one. Without loss of generality we can suppose that z is in this intersection.

Fix $\xi > 0$. Remember that we consider on Ω the metric

$$d_\theta(x, y) = \theta^N$$

where N is the largest integer such that $x_i = y_i$ for any $0 \le i \le N$. Let n_0 be such that $\theta^{n_0} < \xi$ and assume n_0 is the smallest possible such. Then, for $n > n_0$ we have $B(z, n, \xi) = \{y \in \Omega \mid d_\theta(\sigma^j(z), \sigma^j(y)) < \xi, 0 \le j \le n-1\} = (\overline{z_0, z_1, z_2, ..., z_{n+n_0-1}})$ and therefore

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log\, P(B(z, n, \xi)) \\
&= \limsup_{n\to\infty} \frac{1}{n} \log\, p_0^{\sum_{j=0}^{n+n_0-1} I_{(\overline{0})}(\sigma^j(z))}\, p_1^{\sum_{j=0}^{n+n_0-1} I_{(\overline{1})}(\sigma^j(z))} \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n+n_0-1} I_{(\overline{0})}(\sigma^j(z)) \log\, p_0 + \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n+n_0-1} I_{(\overline{0})}(\sigma^j(z)) \log\, p_1.
\end{aligned}$$

The limits in the last expression exist see because z was chosen satisfying Birkhoff's Ergodic Theorem, and therefore

$$\lim_{n\to\infty} \frac{1}{n+n_0-1} \sum_{j=0}^{n+n_0-1} I_{(\overline{0})}(\sigma^j(z)) = \int I_{(\overline{0})}(x)dP(x) = p_0$$

and

$$\lim_{n\to\infty} \frac{1}{n+n_0-1} \sum_{j=0}^{n+n_0-1} I_{(\overline{1})}(\sigma^j(z)) = \int I_{(\overline{1})}(x)dP(x) = p_1.$$

Therefore

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log P(B(z,n,\xi)) \\
&= p_0 \log p_0 \lim_{n\to\infty} \frac{n+n_0-1}{n} + p_1 \log p_1 \lim_{n\to\infty} \frac{n+n_0-1}{n} \\
&= p_0 \log p_0 + p_1 \log p_1.
\end{aligned}$$

Finally,

$$-\lim_{\xi\to 0}\left(\limsup_{n\to\infty} \frac{1}{n} \log P(B(z,n,\xi))\right) = -p_0 \log p_0 - p_1 \log p_1$$

and therefore

$$h(P) = -p_0 \log p_0 - p_1 \log p_1. \quad \blacksquare$$

The next result can be obtained using a similar argument to the one used in the proof of last Theorem:

Theorem 3.2. For the probability $P(p_1, p_2, ..., p_n)$, invariant for the shift σ in n symbols, the entropy is:

$$h(P) = -\sum_{i=1}^{n} p_i \log p_i.$$

Note that from the definition of entropy in principle the value $h(\mu)$ could depend on the metric d we are using.

Theorem 3.3. (Brin-Katok) [4] - Suppose $\mu \in \mathcal{M}(T)$ is an invariant measure, not necessarily ergodic on X and consider d a metric on the compact set X. Then there exist for $\mu - a.e.$ point $z \in X$ the two limits:

$$h(z) = \lim_{\xi \to 0} \left(- \limsup_{n \to \infty} \frac{1}{n} \log \mu(B(z,n,\xi)) \right) = \lim_{\xi \to 0} \left(- \liminf_{n \to \infty} \frac{1}{n} \log \mu(B(z,n,\xi)) \right).$$

The function $h(z)$ is integrable.

The difference between this result and the previous one for ergodic measures is the function h(z). When the measure is not ergodic the "limit sup" can change from point to point even in a set of full measure. When μ is ergodic $h(z)$ is constant for all z μ-almost everywhere.

Definition 3.2. The *entropy* of $\mu \in \mathcal{M}(T)$ is the integral $\int h(z) d\mu(z)$ where $h(z)$ is defined in the above theorem.

This definition generalizes the previous one for ergodic measures.

Note that the concept of entropy was defined only for invariant probabilities on $\mathcal{M}(T)$ and not for the general probability on $\mathcal{M}(X)$. The entropy of an invariant measure is always a non-negative number.

4. Topological Pressure

The entropy of a system (T, X, μ) measures the randomness of the system. The larger the entropy, the more chaotic the system is.

The concept of entropy appears in Physics and is associated with the principle that Nature tends to maximize entropy. That is, if one considers particles of a gas concentrated at a corner of a closed box, at an initial time t_o, then after some time the particles will tend to an equilibrium where the particles are spread in a totally random way. This means that after some time the gas will have a uniform density in the box. As the velocity of the particles is very large, in fact, this is the state that will be observed. Therefore the state that will ocurr in Nature will be the one that is most randomic among all possible states.

A system of particles is much more random (has more entropy) if it is uniformly spread in the box than if it is concentrated at a corner of the box. Therefore equilibrium is attained in maximum entropy arrangements.

The definition of entropy by Shannon was introduced with relation to Information Theory. If one wants to transmit a message through a channel using an alphabet with n symbols $\{1, 2, ..., n\}$, each one with a certain probability $p_1, ..., p_n$, $\sum_{i=1}^{n} p_i = 1$ of being used, then the entropy of this system is the entropy of the Bernoulli shift $B(p_1, p_2, ..., p_n)$. The entropy in this case is a very important information of practical use (see [2]).

Historically the concept of entropy in Physics was defined in a different way than the one introduced much later in 1948 by Shannon.

Our motivation here is associated with a more recent approach of Bowen-Ruelle-Sinai, who, around 1960, proposed to use Shannon's entropy as a mathematical tool for understanding Statistical Mechanics in one-dimensional lattices. Soon we will show that this program includes the study of the Topological Pressure (see definition below) for the shift. In fact these mathematicians proposed to study a more general problem that includes not only the shift but also a larger class of maps. This theory is known nowadays as the Thermodynamic Formalism [17]. The Ruelle-Perron-Frobenius Operator (see next chapter) was introduced for a certain class of maps (the expanding maps in the case one consider one-dimensional dynamics) in order to handle the problem of finding the measure of maximal pressure (see [17]). Several important results in different areas of Mathematics as Geometry, Number Theory, Dimension of Fractals, etc..., were obtained using results related to the above mentioned operator [17], which is a natural generalization (to the space of continuous functions) of a Perron-Frobenius matrix acting on $\mathbf{R}^n$ (see [17] or example after Theorem 7.5). In the context of Physics the Ruelle-Perron-Frobenius Operator corresponds to the Transfer Operator of Statistical Mechanics [17].

Now we will follow the beautiful and simple motivation of the subject presented in Bowen. [3]

Consider a physical system with possible states $1, 2, ..., m$ and let the energies of these states be $E_1, E_2, ..., E_m$, respectively. Suppose that our system is put in contact with a much larger "heat source", which is at temperature T. Energy is thereby allowed to pass between the original system and the heat source, and the temperature T of the source remains constant, as it is so much larger than our system. As the energy of our system is not considered fixed, any array of the states can occur. The physical problem we are considering is not deterministic, and we can only speack of the probability that a certain state, let's say j, ocurr. That is, if one performs a sequence of observations, let's say 1000, it will be observed that for a certain proportion of these observation the state j will ocurr. The relevant question is to know for each j, the value of this proportion (probability) when the number of observations goes to infinity. It has been known from Statistical Mechanics for a long time that the probability P_j that the state j occurs is given by the Gibbs distribution:

$$P_j = \frac{e^{-BE_j}}{\sum_{i=1}^{m} e^{-BE_i}}, \; j \in \{1, 2, ..., m\},$$

where $B = \frac{1}{kT}$ and k is a physical constant.

A mathematical formulation of the above consideration in a variational way can be obtained as follows: consider

$$\tilde{F}(p_1, p_2, ..., p_m) = \sum_{i=1}^{m} -p_i \log p_i - \sum_{i=1}^{m} p_i B E_i,$$

defined over the simplex in $\mathbf{R}^m$ given by

$$\left\{ (p_1, p_2, ..., p_m) \mid p_i \geq 0, \; i \in \{1, 2, ..., m\} \text{ and } \sum_{i=1}^{m} p_i = 1 \right\}.$$

Using Lagrange multipliers, it is easy to show that the maximum of $\tilde{F}$ in the simplex is obtained at

$$P_j = \frac{e^{-BE_j}}{\sum_{i=1}^{m} e^{-BE_j}}, \; j \in \{1, 2, ..., m\},$$

in accordance with the P_j above.

The quantity

$$H(p_1, p_2, ..., p_m) = \sum_{i=1}^{m} -p_i \log p_i$$

is called the entropy of the distribution $(p_1, p_2, ..., p_m)$. Let $-\sum_{i=1}^{m} p_1 E_i$ denote the average energy $E(p_1, p_2, ..., p_m)$.

Then we can say that the Gibbs distribution maximizes

$$H(p_1, p_2, ..., p_m) - BE(p_1, p_2, ..., p_m).$$

The expression $BE - H$ is called, in this context, free energy (in fact, there exist several different concepts in Mathematics and Physics also called free energy).

Therefore we can say that Nature minimizes free energy. When the temperature $T = \infty$, that is, $E = 0$, nature maximizes entropy. In this case the Gibbs state is the most random probability, namely, $P_j = 1/m$, $j \in \{1, 2, ..., m\}$. Again, using analogy with Classical Mechanics, E plays the role of potential energy and H plays the role of kinetic energy.

Now, let us return to Gibbs measures. Generalizing the above considerations, Ruelle proposed the following model: consider the one-dimensional lattice Z. Here one has for each integer a physical system with possible states $1, 2, ..., m$. A configuration of the system consists of assigning an $x_i \in \{1, 2, ..., m\}$ for each $i \in \mathbf{Z}$.

Thus a configuration is a point $x = \{x_i\}_{i \in \mathbf{Z}} \in \{1, 2, ..., m\}^Z = \Omega$

Considering now on the space Ω the shift map

$$\sigma : \begin{pmatrix} \Omega \\ x_{i\ i \in \mathbf{Z}} \end{pmatrix} \to \begin{pmatrix} \Omega \\ x_{i+1\ i \in \mathbf{Z}} \end{pmatrix},$$

and $\mathcal{M}(\sigma)$ the space of probabilities v such that for any Borel set A

$$v(A) = v(\sigma^{-1}(A)),$$

one obtains the well-known Bernoulli shift model.

A continuous function $\phi : \Omega \to \mathbf{R}$, in this setting, contains the information of energy and temperature.

The problem here is to find a way to obtain the Gibbs distribution in the infinite one-dimensional lattice in a similar way as it was obtained before for the finite case.

For instance, for Spin-lattices, one can consider a positive spin $+$ and a negative spin $-$ in each site of the one-dimensional lattice $\mathbf{Z}$ and consider a certain probability p of arrangements. In this case we have to consider the Bernoulli space in two symbols $\Omega = \{+,-\}^{\mathbf{Z}}$ and probabilities p on Ω.

Note that it is natural to consider just probabilities $p \in M(\sigma)$, because there is no natural reason to consider a certain distinguished point of the lattice as the origin 0 in $\mathbf{Z}$.

Given a certain continuous function $\phi : \Omega \to \mathbf{R}$ (ϕ will contain the information of temperature, energy, magnetic-field, etc...), consider the following variational problem:

Definition 4.1. For a continuous function ϕ consider

$$\left(\sup_{p \in \mathcal{M}(\sigma)}\right) \left\{ h(p) + \int \phi(z) dp(z) \right\},$$

where $h(p)$ is the entropy of the probability p. We call such a supremum the *Topological Pressure* (a better name would be Free Energy, but we follow here the terminology of Ruelle) associated with ϕ and denote it by $P(\phi)$.

Remark. There exist an analogous definition of Pressure for invariant measures for T instead of σ.

Example. A good example to have in mind is the following: consider $\Omega = \{+,-\}^{\mathbf{N}}$ (+ is positive spin and - negative spin) and ϕ is constant in each one of the four cylinders $\overline{(+,+)}, \overline{(+,-)}, \overline{(-,-)}$ and $\overline{(-,+)}$. Consider $q_0, q_1 > 0, q_0 + q_1 = 1$ and define ϕ in the following way:

a) $\phi(z) = q_0, \forall z \in \overline{(+,+)}$

b) $\phi(z) = q_1, \forall z \in \overline{(+,-)}$

c) $\phi(z) = 1, \forall z \in \overline{(-,+)}$ and

d) $\phi(z) = 0, \forall z \in \overline{(-,-)}$.

In this case, we assume that in the lattice $\mathbf{Z}$ there exist a probabilility q_0 of having a + at the right of a + and a probability q_1 of having a - at the right of a +. We also assume by c) and d) that at the right of a - there exist always a spin +. One would like to find a probability μ, defined in the all space Ω such that the above mentioned property happen. This probability μ will be called later the equilibrium state associated to the potential ϕ. The equilibrium state μ will be defined by means of a variational formula (see Definition 4.2). In the case of the present example, the solution can be obtained by means of the theory of Markov Chains and Perron-Frobenius operator (note that we introduce a stochastic matrix) and this will be explained in section 7 (see example after Theorem 7.5).

The solution for the case of a general ϕ (not constant in cylinders) will require a more sofisticated version of the Perron-Frobenius theorem that will be presented on section 7.

Most of the time we will use the word pressure instead of topological pressure. It is natural to ask which properties does a probability p which attain such supremum have.

Definition 4.2. We will call the probability μ that attains the above supremum (in the case there exists one such μ) the *Gibbs state* (or *equilibrium probability for* ϕ) for the one-dimensional lattice with potential function ϕ. In other words:

$$h(\mu) + \int \phi(z) d\mu(z) = P(\phi)$$

or

$$h(\mu) + \int \phi(z) d\mu(z) \geq h(\nu) + \int \phi(z) d\nu(z)$$

for any $\nu \in \mathcal{M}(T)$.

Notation. Sometimes we will denote this probability μ by μ_ϕ in order to express the dependence of μ on ϕ.

For expanding systems the probability that attains the above supremum is unique, and therefore equilibrium states do exist (see paragraph 7). Non-uniqueness of the probability that attains the supremum is related with Phase Transition of spin-lattices [9],[10],[12]. D.Ruelle [17] was able to obtain a certain function ϕ that represents interactions of a certain special kind and such that the probability that attains the above supremum $P(\phi)$ is exactly the "Gibbs state" in the lattice **Z** that, with other procedures, people in Physics already knew a long time ago. Therefore the terminology of Gibbs state that we introduced above is quite proper.

The analogy of the above setting in the lattice **Z** with the finite case we mention before is transparent.

If we assume a wall effect, then we have to consider the lattice **N**, that is the one-sided shift.

The setting we presented above is suitable for analyzing problems in Statistical Mechanics of the one-dimensional lattice Z. For the two-dimensional case $\mathbf{Z}^2$ (or for the three-dimensional case $\mathbf{Z}^3$), one should consider actions of $\mathbf{Z}^2$ (or $\mathbf{Z}^3$) and the situation is much more complicated (see [17] for references).

Entropy is defined for measures and Pressure for continuous functions. The set of measures and the set of continuous functions are dual one of the other. In fact these two concepts are related one to the other by means of a Legendre Transform [8]. Some of these properties will be consider in the last part (see section 7) of these notes.

We refer the reader to [7] [8] [5] [11] for a complete description of the above results.

When two different ϕ and ψ determine the same equilibrium state μ ? That is, when $\mu_\phi = \mu_\psi$? This is an important question that will be analized more carefully later. The following proposition is an easy consequence of the properties of the probabilities $\nu \in \mathcal{M}(T)$.

Proposition 4.1. Criterium of Homology - Suppose ϕ and ψ are two continuous functions such that there exist a continuous function g and a constant k satisfying $\phi - \psi = g \circ T - g + k$, then $\mu_\phi = \mu_\psi$.

Proof. For any $\nu \in \mathcal{M}(T)$, $\int (g \circ T(z) - g(z)) d\nu(z) = 0$ by definition, therefore $h(\nu) + \int \phi(z) d\nu(z) = h(\nu) + \int \psi(z) d\nu(z) + k$ for any $\nu \in \mathcal{M}(T)$. Therefore $P(\phi) = P(\psi) + k$ and $\mu_\phi = \mu_\psi$.

Note that if k=0, then $P(\phi) = P(\psi)$. ■

Definition 4.3. In the case $\phi = 0$, we have

$$P(\phi) = \sup_{p \in \mathcal{M}(T)} h(p),$$

and this value $P(0)$ is called the *topological entropy* of T . We will denote such value by h(T).

We refer the reader to [3] [15] [17] [18] for results about Pressure and Thermodynamic Formalism.

In the case $T = \sigma$ it can be shown that $h(\sigma) = \log d$ (see Definition 4.3) if (σ, Ω) is the shift in d symbols.

More generally, if an expanding map T has the property that for any $a \in X, \#\{T^{-1}(a)\} = d$, then $h(T) = \log d$.

From Theorem 3.2 the entropy of the shift σ of d symbols, under the probability $P(1/d, 1/d, ..., 1/d)$ is equal to $\log d$. Therefore, in this case we can identify very easily the equilibrium state for $\phi = 0$, it is the probability $\mu_0 = P(1/d, 1/d, ..., 1/d)$. This measure will be called later the *maximal entropy measure* .

In paragraph 7 we will consider very precise results on the existence of equilibrium states for expanding maps.

5. Large Deviation

In this paragraph and in the next one, we will consider T a continuous map from a compact metric space (X,d) into itself, μ an ergodic invariant measure on $(X, \mathcal{A})$ and f a continuous function from X to $\mathbf{R^m}$. Some of the proofs will be done for $m = 1$ in order to simplify the notation.

The Ergodic Theorem of Birkhoff claims that for an ergodic measure $\mu \in \mathcal{M}(T)$ and a continuous function f from X to $\mathbf{R^m}$, for μ-almost every point $z \in X$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(z)) = \int f(x) d\mu(x).$$

The typical example of application of the Ergodic Theorem, as we said before, is the situation where we toss a fair coin 1000 times. One can observe that among these 1000 tossings, more or less 500 times appears a head and the same happens for tails. The event of obtaining head all the 1000 times is possible, and has P-probability $(0.5)^{1000}$. This number is very small but is not zero. This event is a deviation of the general behaviour of the typical trajectory. It is very relevant in several problems in Probability, in Mathematics and in Physics to understand what happens with the trajectories that deviate of the mean. We will show later mathematical examples of such phenomena.

For each time n the data $\frac{1}{n}\sum_{j=0}^{n-1} I_{\overline{0}}(\sigma^j(z))$ are spread around the mean value 1/2, but when n goes to infinity, the data are more and more concentrated (in terms of probability) around the mean value. The main question is: how to estimate deviating behaviour? For the fair coin, the typical trajectory will produce, in the limit as n goes to infinity, the temporal mean 1/2. Suppose we estipulate that a mistake of $\epsilon = 0.01$ is tolerable for the distance of the finite temporal mean to the spatial mean

$$\mid \frac{1}{n}\sum_{j=0}^{n-1} I_{\overline{0}}(\sigma^j(z)) - \int I_{\overline{0}}(x)dP(x) \mid,$$

but not more than that.

For n=1000, there exists a set $B_n(\epsilon)$ with small P=P(1/2,1/2) probability such that the temporal mean of orbits has a temporal mean outside the tolerance level. For example the cylinder with the first 1000 elements equal to 0 is contained in $B_n(\epsilon)$, because

$$\frac{1}{n}\sum_{j=0}^{999} I_{\overline{0}}(\sigma^j(z)) - \frac{1}{2} = 1 - \frac{1}{2} = \frac{1}{2} \geq 0.01.$$

for z in the above mentioned cylinder.

We will be concerned here with the problem of estimating the velocity with which $\mu(B_n(\epsilon))$ goes to zero when n goes to infinity.

From a practical point of view, the Ergodic Theorem would not be very useful, if $\mu(B_n(\epsilon))$ goes to zero too slowly. For a given ϵ of tolerance and a fixed n (any practical experiment is finite), we choose at random a point z in X, according to $P(1/2,1/2)$. If the velocity of convergence to zero of the sequence $\mu(B_n(\epsilon))$ is very slow, then there is a very large probability of choosing the point z in the *bad* set $B_n(\epsilon)$.

The area of Mathematics where such kind of problems are tackled is known as Large Deviation Theory (see [7] for a very nice and general reference).

Let's return now to the general case of a measurable map T from X to X, leaving invariant a measure μ. We will be more precise about what we want to measure.

Definition 5.1. Given ϵ greater than zero and $n \in \mathrm{N}$, then by definition $Q_n(\epsilon)$ is equal to:

$$Q_n(\epsilon) = \mu\{z \mid |\frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z)) - \int f(x)d\mu(x) | \geq \epsilon\}.$$

Proposition 5.1. Given ϵ,

$$\lim_{n\to\infty} Q_n(\epsilon) = 0.$$

Proof. For a given value ϵ denote

$$A_n = \{z \,||\frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z)) - \int f(x)d\mu(x) | \geq \epsilon\}.$$

We will show that $\lim_{n\to\infty} \mu(A_n) = 0$.

Consider the set $Y = \cap_{n=1}^{\infty} \cup_{i\geq n}^{\infty} A_i$. For each $z \in Y$, the sequence $a_n = \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z))$ has a subsequence with distance more than ϵ from $\int f(x)d\mu(x)$. Therefore, for any $z \in Y$ the above defined sequence a_n does not converge to $\int f(x)d\mu(x)$, and hence Y has measure zero by the Ergodic Theorem of Birkhoff.

As the sequence $D_n = \cup_{i\geq n}^{\infty} A_i$ is decreasing and $\mu(Y) = 0$, then

$$\lim_{n\to\infty} \mu(A_n) \leq \lim_{n\to\infty} \mu(D_n) = \mu(Y) = 0$$

Therefore the proposition is proved. ■

Corollary 5.1. Given $\epsilon > 0$

$$\lim_{n\to\infty} \mu\{z \,||\frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z)) - \int f(x)d\mu(x) | \leq \epsilon\} = 1$$

One would like to be sure that the convergence to zero we consider above in Proposition 5.1 is at least exponential, that is: for any ϵ, there exists a positive M such that for every n

$$\mu\{z \mid | \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z)) - \int f(x)d\mu(x) | \geq \epsilon\} \leq e^{-Mn}$$

Under suitable assumptions we will show that this property will be true (see Prop. 6.8).

It is quite surprising that in the case μ is an equilibrium state (see Def. 4.2) this result can be obtained using properties related to the Pressure (see paragraph 7 and 8). We will return to this fact later, but first we need to explain some of the basic properties of Large Deviation Theory.

The relevant question here is how fast, in logarithmic scale the value $Q_n(\epsilon)$ goes to zero, that is, how to find the value

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(\epsilon).$$

The above value is an important information about the asymptotic value of the μ -measure of the set of trajectories that deviate up to ϵ of the behaviour of the typical trajectory given by the Theorem of Birkhoff.

More generaly speaking, for a certain subset A of $\mathbf{R^m}$ one would like to know, for a certain fixed value of n, when the values z are such that:

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z)) \in A.$$

In the situation we analyze before (corollary 5.1)

$$A = \{y \in \mathbf{R^m} \mid | y - \int f(x)d\mu(x) | \geq \epsilon\}$$

Definition 5.2. Given a subset A of $\mathbf{R^m}$ and $n \in \mathbf{N}$ we denote

$$Q_n(A) = \mu\{z \mid \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z)) \in A\}.$$

In the same way as before one would like to know the value

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(A).$$

Remark. If the set A is an open interval that contains the mean value $\int f(x)d\mu(x)$, then the above limit is zero because $\lim_{n\to\infty} Q_n(A) = 1$ (see corollary 5.1).

First, we will try to give a general idea of how the solution of this problem is obtained, and then later we will show the proofs of the results we will state now.

There exists a *magic* function I(v) defined for $v \in \mathbf{R}^m$ (the set where the function f takes its values) such the the above limit is determined by:

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(A) = -\inf_{v\in A}\{I(v)\},$$

when A is an interval.

The function I it will be called the *deviation function*. The shape of I is basically the shape of $| v - \int f(x)d\mu(x) |^2$, $v \in \mathbf{R}^m$, that is, I(v) is a non-negative continuous function that attains a minimum equal to zero at the value $\int f(x)d\mu(x)$.

The properties we mentioned before are not always true for the general T, μ and f, but under reasonable assumptions the above mentioned properties will be true. This will be explained very soon.

The natural question is: how can one obtains such a function I? The function I(v), $v \in \mathbf{R}^m$ is obtained as the Legendre Transform (we will present the general definition later) of the *free energy* c(t), $t \in \mathbf{R}^m$ to be defined below.

Definition 5.3. Given $n \in \mathrm{N}$ and $t \in \mathbf{R}^m$ we denote

$$c_n(t) = \frac{1}{n} \log \int e^{<t,(f(x)+f(T(x))+f(T^2(x)+\ldots+f(T^{n-1}(x))>} d\mu(x).$$

Definition 5.4. Suppose that for each $t \in \mathbf{R}^n$ and $n \in \mathrm{N}$, the value $c_n(t)$ is finite, then we define c(t), the *free energy* , as the limit:

$$c(t) = \lim_{n\to\infty} c_n(t),$$

in the case the above limit exists.

Remark. Note that c(0) = 0.

Remark. The function c(t) is also known in Probability as the moment generating function. For people familiar with Probability Theory and Stochastic Processes, we would like to point out that the random variables $f(T^n(z)), n \in \mathbf{N}$ are not independent in general.

Definition 5.5. A function g(t) is *convex* if for any $s, t \in \mathbf{R}^m$ and $0 < \lambda < 1$,

$$g(\lambda s + (1-\lambda)t) \leq \lambda g(s) + (1-\lambda)g(t)$$

We say g is *strictly convex* , if for any $0 < \lambda < 1$ the above expression is true with $<$ instead of $\leq$.

It is easy to see that a differentiable function $g(t)$ such that its second derivative satisfies $g''(t) \geq 0$ for all $t \in \mathbf{R}$ is convex.

Proposition 5.2. The function c(t) is convex in $t \in \mathbf{R}^n$.

Proof. The Hölder inequality [16] claims that

$$\int | hk | \, d\mu(x) \leq (\int | h(x) |^p \, d\mu(x))^{1/p} (\int | k(x) |^q \, d\mu(x))^{1/q},$$

where h and k are respectively on $\mathcal{L}_p(\mu)$ and $\mathcal{L}_q(\mu)$ and p and q are such that 1/p + 1/q =1.

Consider $s, t \in \mathbf{R}^n$, $h(x) = e^{<\lambda s, f(x)+\ldots+f(T^{n-1}(x))>}$, $k(x) = e^{<(1-\lambda)s, f(x)+\ldots+f(T^{n-1}(x))>}$, $\lambda \in (0,1)$, and then define p=1/λ and q=1/(1- λ). Now, using the Hölder inequality:

$$\int e^{<\lambda s+(1-\lambda)t, f(x)+f(T(x))+\ldots+f(T^{(n-1)}(x)>} d\mu(x) \leq$$

$$(\int e^{<s, f(x)+\ldots+f(T^{(n-1)}(x))>} d\mu(x))^{\lambda} (\int e^{<t, f(x)+\ldots+f(T^{(n-1)}(x))>} d\mu(x))^{1-\lambda}.$$

Therefore, taking $\frac{1}{n}\log$ in each side of the above inequality, one obtains that:

$$c_n(\lambda s + (1-\lambda)t) \leq \lambda c_n(s) + (1-\lambda)c_n(t),$$

and hence c(t) is convex, because it is the limit of convex functions. ∎

Definition 5.6. The *deviation function* I(v), $v \in \mathbf{R}^{\mathbf{m}}$, is by definition the Legendre transform of the function c(t), $t \in \mathbf{R}^{\mathbf{m}}$, that is

$$I(v) = \sup_{t\in\mathbf{R}^{\mathbf{m}}} \{< t, v > -c(t)\}.$$

The deviation function I is well defined in the case c(t) is strictly convex.

In order to simplify the argument, let's consider the one dimensional case m=1. When c is differentiable, then it is easy to see that

$$I(v) = \sup_{t\in\mathbf{R}}\{tv - c(t)\} = t_0 v - c(t_0),$$

where t_0 is such that $c'(t_0) = v$ (see proposition 6.1). Such a t_0 is well defined if c is strictly convex and differentiable. In this case the deviation function I(v) is also differentiable in v, as it is easy to see. If c(t) is piecewise differentiable (with left and right derivatives), then I(z) has also this property.

In more precise mathematical terms one should say that the deviation function I(v) of c(t), $t \in \mathbf{R}^{\mathbf{m}}$, takes values v in the dual of $\mathbf{R}^{\mathbf{m}}$. The dual of $\mathbf{R}^{\mathbf{m}}$ is $\mathbf{R}^{\mathbf{m}}$ itself, and therefore, in the finite dimensional case (m finite) there is no problem to define the Legendre transform in the way we did above. We will need to consider Legendre transforms in infinite dimensional vector spaces soon. This will require some small changes in the definition of Legendre Transform. Before that, we will consider the main properties that are true in the finite dimensional case. The key property is the differentiability of the free energy c(t). Assuming piecewise differentiability (with the existence of right and left derivatives for c(t), $t \in \mathbf{R}$), most results we will state below will be true (Theorem 6.2 and Proposition 6.8 require that the free energy be differentiable).

The main result we want to prove in the next paragraph is:

Theorem 5.1. Assume the free energy c(t), $t \in \mathbf{R^m}$ is well defined and also that c is differentiable, then for an open paralepiped A contained in $\mathbf{R^m}$

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(A) = -\inf_{z\in A}\{I(v)\}.$$

The above result is true for much more general sets A contained in $\mathbf{R^m}$, but we will state and prove the general result later.

The main results for the finite dimensional case will be proved for n=1. The general case is not very much different from the case n=1. The infinite dimensional case is however much more difficult than the finite dimensional case [7].

6. Free Energy and the Deviation Function

We will need to develop some elementary properties of Legendre Transforms in order to prove the Theorem we stated above.

Definition 6.1. Given a convex piecewise differentiable map g(y), $y \in \mathbf{R^m}$, the *Legendre transform* of g, denoted by $g^*(p), p \in \mathbf{R^m}$, is by definition

$$g^*(p) = \sup_{y\in\mathbf{R^m}} \{< p, y > - g(y)\}.$$

Proposition 6.1. Suppose $g(y)$ is defined for all $y \in \mathbf{R}$ and that the second derivative is continuous. If there exists $\alpha > 0$ such that, $g''(y) > \alpha > 0$, $y \in \mathbf{R}$, then $g^*(p) = py_0 - g(y_0)$ where $g'(y_0) = p$.

Proof. In the case there exists a value y_0 such that $g'(y_0) = p$, then clearly $g^*(p) = y_0 p - g(y_0)$. Therefore, all we have to show is that $g'(y)$ is a global diffeomorphism from $\mathbf{R}$ to $\mathbf{R}$.

Note that for a positive h, $g'(x+h) - g'(x) = \int_x^{x+h} g''(y)dy > \alpha h$. Therefore the map g' is injective. The map g' is open (that is, the image $g'(A)$ of each open set A is open) because $g'(x+h) - g'(x) > \alpha h$. The map g' is closed (that is, the image $g'(K)$ of each closed set K is closed), because it is continuous. We claim that g' is sobrejective. This is easy too see: the image by g' of the open and closed set $\mathbf{R}$, is an open and closed interval and therefore equal to $\mathbf{R}$. The conclusion is that g' is bijective from $\mathbf{R}$ to itself. ■

Proposition 6.2. Suppose $g(y)$ defined on $y \in \mathbf{R}$ satisfies $g"(y) > 0$ for all $y \in \mathbf{R}$, then g^* satifies $g^{*"}(p) > 0$ for all $p \in \mathbf{R}$.

Proof. We will use the following notation: for each value p denote $y(p)$ the only value y such that $\frac{dg}{dy}(y(p)) = p$. As we saw in the last proposition $g^*(p) = y(p)p - g(y(p))$. Taking derivatives with respect to p,

$$\frac{dg^*}{dp}(p) = \frac{dy}{dp}(p)p + y(p) - \frac{dg}{dy}(y(p))\frac{dy}{dp}(p) = \frac{dy}{dp}(p)p + y(p) - p\frac{dy}{dp}(p) = y(p).$$

Hence $g^{*"}(p) = y'(p)$

Now, as for any p, $p = \frac{dg}{dy}(y(p))$, taking derivatives in both sides with respect to p, $1 = g''(y(p))y'(p) = g''(y(p))g^{*''}(p)$. Thus $g^{*''}$ is positive, if g'' is positive. ■

Remark. We will assume that all maps g to which we apply the Legendre transform satisfy the condition $g"(y) > \alpha, y \in \mathbf{R}$ for a certain fixed positive value α. When we consider piecewise differentiable maps (with left and right derivatives), then we will also suppose that the left and right derivatives satisfy the same condition in α.

The geometric interpretation of the Legendre transform of g in terms of the graphic of g is shown in fig 1.

Now we will prove a key result in the Theory of Legendre Transforms:

Proposition 6.3. Suppose $f(x)$ and $f^*(x)$ are stricly convex and differentiable for every x, then the Legendre Transform is an involution, that is, $f^{**} = f$.

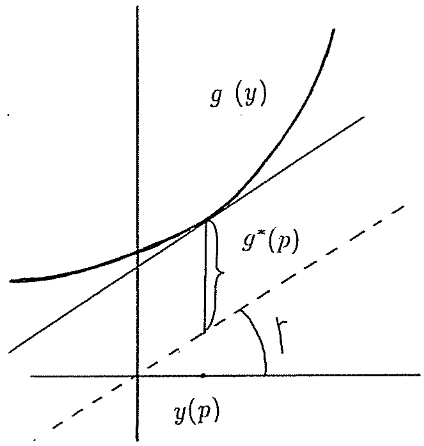

Figure 1.

Proof. We will show that if g denotes f^*, then $g^* = f$.

For a given $p \in \mathbf{R}$ denote by $x(p)$ the value x such that $\sup_{x\in\mathbf{R}}\{px - f(x)\}$ attains the supremum. Since $f^* = g$, then $\frac{df}{dx}(x(p)) = p$ and $g(p) = px(p) - f(x(p))$.

For a certain fixed value x_0 and for each $x \in \mathbf{R}$ define $\Delta(x)$ as the value Δ obtained by the intersection of the line $(y, z(y)) = (y, f(x) + f'(x)y)$ with the line $x = x_0$ (see fig 2). It is easy to see that $\frac{f(x)-\Delta}{x-x_0} = f'(x)$, and therefore

$$\Delta(x) = f(x) - xf'(x) + f'(x)x_0.$$

Given p, $g(p) = px(p) - f(x(p))$ where x(p) is such that $\frac{df}{dx}$ (x(p)) = p. Therefore, if we write Δ in terms of p, then

$$\Delta(p) = \Delta(x(p)) = \Delta(x) = f(x(p)) - x(p)p + px_0 = -g(p) + px_0.$$

Note that

$$\sup_{p\in\mathbf{R}} \Delta(p) = \sup_{p\in\mathbf{R}}\{px_0 - g(p)\} = g^*(x_0).$$

From fig 2 one can easily see that $\sup \Delta(p)$ is attained when $p = f'(x_0)$ and the supremum value of Δ is $f(x_0)$. Therefore we conclude that $g^*(x_0) = f(x_0)$. ∎

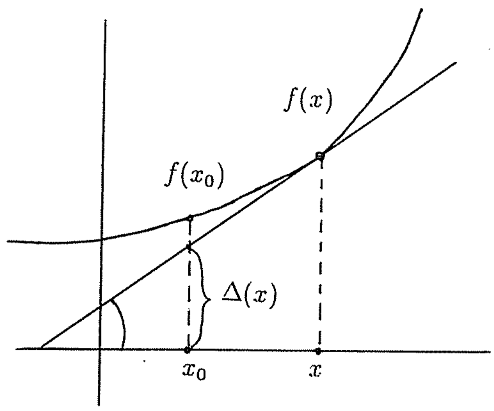

Figure 2.

Definition 6.2. We say that f is *conjugated* to g if $f^* = g$.

The last result claims that if f is conjugated to g, then g is also conjugated to f.

Definition 6.3. Suppose g is a convex function on $\mathbf{R^m}$. We say that $y \in \mathbf{R^m}$ is a *subdifferential* of g in the value x, if $g(z) \geq g(x)+ < y, z - x >$ for any $z \in \mathbf{R^m}$. We denote the set of all subdifferentials of g in the value x by $\delta g(x)$.

This definition allows one to deal with the case c(t), $t \in \mathbf{R}$, piecewise differentiable (it is differentiable up to a finite set of points $t_i, i \in \{1, 2, ..., n\}$). In the values t where c is differentiable there is a unique subdifferential $c^{'}(t) = \delta c(t)$, but in the values t_i where c(t) has left and right derivatives (we assume in the definition that this property is true) respectively equal to u_i and v_i, then $\delta c(t_i)$ is the interval $[\ u_i, v_i\ ]$.

The next result shows a duality between the subdiferentials of conjugated functions.

Proposition 6.4. y $\in \delta g(x)$ if and only if x $\in \delta g^*(y)$.

Proof. By definition y$\in$ $\delta g(x)$ is equivalent to

$$g(z) \geq g(x)+ < y, z - x >$$

for all $z \in \mathbf{R}$.

The last expression is equivalent to

$$< y, z > -g(z) \leq < y, x > -g(x)$$

for all z $\in \mathbf{R}$.

Therefore $y \in \delta g(x)$ is equivalent to say that x realizes the supremum of $< y, z > -g(z)$.

We also obtain from the above reasoning that $y \in \delta g(x)$ is equivalent to $g^*(y) =< y, x > - g(x)$, and thus equivalent to $< x, y >= g^*(y) + g(x)$.

Applying the same result for $g = g^*$, and interchanging the role of x and y, that is, x=y and y=x, we conclude that $x \in \delta g^*(y)$ is equivalent to $< y, x >= g^{**}(x) + g^*(y)$. The last expression is equivalent to $< y, x >= g(x) + g^*(y)$, because from the last proposition $g^{**} = g$.

Hence $y \in \delta g(x)$ is equivalent to $x \in \delta g^*(y)$ ■

Using this proposition one can show the following result:

Proposition 6.5. $I(v) = 0$, if and only if, $v \in \delta c(0)$. The function I is non-negative and has minimum equal zero in the set $\delta c(0)$.

Proof. First note that as $I = c^*$, then from the last proposition $v \in \delta c(0)$, if and only if, $0 \in \delta I(v)$. In this case,

$$I(z) \geq I(v) + < 0, z - v >= I(v) = 0$$

for any $z \in \mathbf{R}$. Therefore, I(z) has infimum in the set $\delta c(0)$.

Proposition 6.4 claims that $< t, v >= c(t) + c^*(v) = c(t) + I(v)$, if and only if, $v \in \delta c(t)$. Now, using this proposition for the case $t = 0$, one obtain $I(v) = -c(0) = 0$. The final conclusion is that $I(z) \geq I(v) = 0$ for $v \in \delta c(0)$ and $z \in \mathbf{R}$. ■

The proof of the main Theorem 5.1 is done in two separated parts: the upper large deviation inequality and the lower large deviation inequality. First we will show the upper large deviation inequality. This inequality is true in a quite general context, even without the hypothesis of full differentiability of c(t) [7]. In the second inequality we will use differentiability of the free energy.

Proposition 6.6. (Upper large deviation inequality) Suppose c(t), $t \in \mathbf{R}$ is a well defined convex function, then

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu\{x \mid \sum_{j=0}^{n-1} f(T^j(x)) \in K\} = \limsup_{n\to\infty} \frac{1}{n} \log Q_n(K) \leq - \inf_{z\in K} I(z) \quad (5)$$

$\frac{A}{n}$

where K is a closed set in $\mathbf{R}$.

Proof. Let's first recall Tchebishev's inequality: let g be a measurable function from X in $\mathbf{R}$ and h from $\mathbf{R}$ to $\mathbf{R}$ a non-negative, nondecreasing function such that $\int h(g(x))d\mu(x)$ is finite. In this case, for any value d such that h(d) is positive

$$\mu\{x \mid g(x) \geq d\} \leq \frac{\int h(g(x))d\mu(x)}{h(d)}.$$

We refer the reader to [7] for the proof of Tchebishev's inequality.

Denote $\delta c(0) = [u_0, v_0]$ (it is very easy to see that $\delta c(0)$ is an interval).

We will show first the claim of the theorem for semi intervals $[a, \infty,)$ where a is larger than the right derivative v_0 of c at t=0. For such a and any $t > 0$, Tchebishev's inequality for

$$h(y) = e^{nty},\ g(x) = \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x)),\ d = a,$$

(Remark- we require $t > 0$ in order h(y) being non-decreasing) implies that

$$Q_n([a,\infty)) \leq e^{-nta} \int e^{t\sum_{j=0}^{n-1} f(T^j(x))} \mu(x) = e^{-n(ta - c_n(t))}.$$

Therefore taking limits when n goes to infinity, one concludes that

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n([a,\infty)) \leq - \sup_{t\geq 0}\{ta - c(t)\}. \quad (6)$$

Now we need the following claim:

Claim. $\sup_{t\geq 0}\{at - c(t)\} = I(a) = \sup_{t\in\mathbf{R}}\{at - c(t)\}$.

Proof of the Claim. c(t) is convex, hence u_0, the left derivative of c at 0, satisfies $u_0 \leq \frac{c(t)}{t}, t < 0$. Therefore,

$$ta - c(t) = t(a - \frac{c(t)}{t}) \leq t(a - u_0).$$

The last term is negative because $a \geq v_0 \geq u_0$.

The conclusion, is that $I(a) = \sup_{t\in\mathbf{R}}\{ta - c(t)\} = \sup_{t>0}\{ta - c(t)\}$.

Hence the claim is proved.

Before we return to the proof of Theorem, we will need first to prove another claim.

Claim. $I(a) = \inf_{z\geq a} I(z)$.

Proof of the Claim. From Proposition 6.5, $I(z)$ is equal to 0 on $[u_0, v_0] = \delta c(0)$. We claim that for $z > v_0$ the function I is monotone nondecreasing. This is so because, if there exist two values z_1 and z_2 larger than v_0, such that $I(z_1) = I(z_2)$, then there exists $z \in [z_1, z_2]$ with $0 \in \delta I(z)$ (this follows at once from the convexity and the definition 6.3 but do not require differentiability).

This means, by proposition(6.5), that $z \in \delta c(0)$, but this is false because z is not in $[u_0, v_0]$. Therefore $I(a) = \inf_{z>a} I(z)$, and the second claim is also proved.

Now, from equation (6) and using the two claims stated above, we obtain the desired conclusion

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n\{K\} \leq -\inf_{z\in K} I(z) \tag{7}$$

when $K = [a, \infty)$ and a larger than v_0, the rigth derivative of c at 0.

The proof for intervals K of the form $(-\infty, a]$, $a < u_0$ is similar.

Now we will prove the claim of the theorem for a general closed set K.

First note that if K intersects the set $\delta c(0) = [u_0, v_0]$, then the claim is trivial because $\inf_{z\in K} I(z) = 0$ (remember that $v \in \delta c(0)$, if and only if, I(v) = 0, by proposition 6.5).

Hence, we will suppose that K does not intersect the set $[u_0, v_0]$.

Consider a, b two real values such that $(-\infty, a] \cup [b, \infty)$ is the smallest possible set such that $K \subset (-\infty, a] \cup [b, \infty)$. As the set K is closed, then ($a = -\infty$ or $a \in K$) and ($b = \infty$ or $b \in K$). Suppose for simplification of the notation that $a, b \in K$ (the other case can be easily handled by the reader). From the first part we know that $\inf_{z\in(-\infty,a]} I(z) = I(a)$ and $\inf_{z\in[b,\infty)} I(z) = I(b)$. Therefore $\inf_{z\in K} I(z) = \min\{I(a), I(b)\}$, because $a, b \in K$.

Finally from the first part(7):

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log Q_n(K) \le \limsup_{n\to\infty} \frac{1}{n} \log(Q_n(-\infty, a] + Q_n[b, \infty)) \le \\
&\limsup_{n\to\infty} \frac{1}{n}(\log Q_n(-\infty, a] + \log Q_n[b, \infty)) \le -I(a) - I(b) \le \\
&-\inf\{I(a), I(b)\} = \inf_{z\in K} I(z).
\end{aligned}$$

Therefore the Proposition is proved. ■

Proposition 6.7. If c(t) is differentiable at t=0, then $c'(0) = \int f(x) d\mu(x)$.

Proof. We know from the last proposition that $I(z) \ge I(v) = 0$ for $z \in \mathbf{R}$ and $v \in \delta c(0) = \{c'(0)\}$.

Note that if c is differentiable at 0, we have uniqueness of the z such that $I(z) = 0$, this value being equal to $v = c'(0)$.

The proof will be done by contradiction. Suppose $c'(0)$ is different from $\int f(x) d\mu(x)$. Given $\epsilon = \frac{|c'(0) - \int f(x) d\mu(x)|}{2} > 0$, consider

$$K = (-\infty, c'(0) - \epsilon] \cup [c'(0) + \epsilon, \infty)$$

and $M = \inf_{z\in K} I(z) > 0$. Proposition 6.6 assures that for suficiently large $n \in \mathbf{N}$:

$$\mu(\{z \mid | \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(z)) \in K\} = \mu(\{z \mid \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(z)) - c'(0) \mid \ge \epsilon\} \le e^{-nM}. \quad (8)$$

From the last inequality, μ-almost every point z has the property that its temporal mean converges to $c'(0)$, and from the Theorem of Birkhoff, this value $c'(0)$ has to be the spatial mean $\int f(x)d\mu(x)$. Hence we obtain a contradiction and the proposition is proved. ∎

Definition 6.4. We say that the μ-integrable function f from X to $\mathbf{R}$ has the *exponential convergence property*, if for any $\epsilon > 0$, there exist $M > 0$ such that:

$$\mu\{y \mid |\sum_{j=0}^{n-1} f(T^j(y)) - \int f(x)d\mu(x) | \geq \epsilon\} \leq e^{-nM}$$

for n large enough. $\frac{1}{n}$

Proposition 6.8. Suppose c is differentiable at $t = 0$, then f has the exponential convergence property.

Proof. As we have just shown that $c'(0) = \int f(x)d\mu(x)$ and $v = c'(0)$ is the only value that $I(v) = 0$, then given ϵ, there exists

$$M = \inf_{z \in [\int f(x)d\mu(x) - \epsilon, \int f(x)d\mu(x) + \epsilon]} I(z),$$

such that

$$\mu\{y \mid |\sum_{j=0}^{n-1} f(T^j(y)) - \int f(x)d\mu(x) | \geq \epsilon\} \leq e^{-nM}. \quad ∎$$

We will need the very well known definition of distribuition in order to simplify the notation in the proof of the next theorem:

Definition 6.5. Given a μ-integrable function $f : X \to \mathbf{R}$, (a random variable) then the measure μ^f defined on the real line $\mathbf{R}$, such that for any continuous function $g : \mathbf{R} \to \mathbf{R}$

$$\int g \circ f d\mu(z) = \int g(x)d\mu^f(x)$$

is called the *distribuition function* of the μ-integrable function f.

Such a measure μ^f always exists (using the notation of the first chapter $f : X \to Y$ (or $f : X \to \mathbf{R}$), then μ_f it is the pull-back of the measure μ by the map f as introduced in Definition 2.8).

Remark. Note that for any interval (a, b) contained in $\mathbf{R}$,

$$\mu^f((a,b)) = \mu\{y \mid f(y) \in (a,b)\}.$$

As a practical rule, remember that each time one wants to integrate $\int g(x) d\mu^f(x)$, one substitutes the variable x by f(z) and integrates with respect to μ, that is: $\int g(f(z)) d\mu(z)$.

The proofs of all results we obtained before are quite general and can be easily extented (the proofs being absolutely the same) to the following case:

Theorem 6.1. For each value $n \in \mathbf{N}$, let X_n be a μ-integrable function on X such that $\frac{X_n(z)}{n} \in \mathbf{R}, z \in X$ has $\nu_n(x), x \in \mathbf{R}$ as distribuition function, that is, using the notation that we introduced above for distribuition function, $\nu_n = \mu^{\frac{X_n}{n}}$. Define

$$c(s) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{sX_n}(z) d\mu(z) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{snx} d\nu_n(x)$$

the *free energy* of the sequence $\frac{X_n}{n}$.

Suppose $c(s)$ is differentiable at $s = 0$, then there exists a positive M such that

$$\mu(\{z \mid | \frac{X_n}{n}(z) - c'(0) | \geq \epsilon\} \leq e^{-nM} \tag{9}$$

for n large enough.

The value M is obtained in the following way:

$$M = \inf_{l \in (-\infty, c'(0)-\epsilon) \cup (c'(0)+\epsilon, \infty)} I(l),$$

where for each value l, $I(l) = \sup_{l\in\mathbf{R}}\{sl - c(s)\}$, is the Legendre transform of c(s).

Remark. Note that it follows from the above theorem that

$$\lim_{n\to\infty} \nu_n((-\infty, c'(0) - \epsilon] \cup [c'(0) + \epsilon, \infty)) = 0$$

and therefore that

$$\lim_{n\to\infty} \nu_n(B(c'(0)), \epsilon)) = 1 \tag{10}$$

(see last remark and the definition of distribuition function).

The last theorem can be seen as a generalization of the results we obtained before by making the measurable function $X_n(z)$ defined above play the role of the function $\sum_{j=0}^{n-1} f(T^j(z))$ that we previously considered.

Now we will use this last result to prove the lower large deviation inequality:

Theorem 6.2. (Lower large deviation inequality) Suppose that the free energy c(t) is differentiable for every $t \in R$, then for any open set A:

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(A) \geq -\inf_{z\in A} I(z).$$

Proof. We will assume that for any real value $z \in \mathbf{R}$ there exists a value t such that $c'(t) = z$. If we suppose that $c"(t) > \alpha > 0$, then this assumption is satisfied as we saw in Proposition 6.1.

The above hypothesis is not necessary for the proof of the theorem, but in order to avoid too many technicalities, we will prove the result under this assumption.

Consider z in the open set A and r such that $B(z,r) = (z-r, z+r)$ is contained in A. Denote by t a value such that $c'(t) = z$ (there exists such a t by hypothesis).

Now we will need to use the concept of distribuition of a μ-measurable function that we introduce before.

We will denote by μ^n the distribuition on $\mathbf{R}$ such that $\mu^n = \mu^{\frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z))}$ (see the notation introduced after definition 6.5).

Therefore, given a set $(a, b) \subset \mathbf{R}$,

$$\int_{(a,b)} d\mu^n(x) = \mu^n((a,b)) = \mu\{z \mid \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z)) \in (a,b)\} = Q_n((a,b)).$$

Denote $Z_n(t) = \int e^{tnx} d\mu^n(x) = e^{nc_n(t)}$ (see definition 5.3 and remember the practical rule mentioned in the remark after the definition 6.5 of distribuition). The reader familiar with Statistical Mechanics will recognize the Partition function in the definition we introduced.

For each value $t \in \mathbf{R}$ and $n \in \mathbf{N}$, we will now denote by μ_t^n the probability on $\mathbf{R}$ given by

$$d\mu_t^n(x) = \frac{e^{ntx}}{Z_n(t)} d\mu^n(x). \tag{11}$$

Note that for a fixed t and n,

$$Z_n(t) = e^{nc_n(t)} = \int e^{nt\frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x))} d\mu(x) = \int e^{tnx} d\mu^n(x),$$

and therefore the term $Z_n(t) = e^{c_n(t)}$ appears only as a normalization term in the definition of the probability μ_t^n (it does not depend on x).

This one-parameter family of probabilities $\mu_t^n, t \in \mathbf{R}$, will play a very important role in the proof of the theorem.

One should think of the measure μ_t^n in the following way: for t=0 the measure $\mu^n = \mu_t^n$. From the Theorem of Birkhoff, the measure $\mu^n = \mu_0^n$ focalizes on (or has mean value) $v = \int f(x) d\mu(x) = c'(0)$, that is,

$$\limsup_{n \to \infty} \mu^n((c'(0) - \epsilon, c'(0) + \epsilon)) = \limsup_{n \to \infty} Q_n((c'(0) - \epsilon, c'(0) + \epsilon)) = 1.$$

For the given value $z \in A$, we choose t such that $c'(t) = z$, and then the measure μ_t^n, will focalize on (or has mean value) $z = c'(t)$ as will be shown:

Claim. Suppose $c'(t) = z$, then for any r:

$$\lim_{n \to \infty} \mu_t^n((z - r, z + r)) = 1 \tag{12}$$

Proof of the Claim. For the value t and $n \in \mathbf{N}$, let X_n be a measurable functions such that $\frac{X_n}{n}$ has distribuition function μ_t^n (such measurable functions always exist by trivial arguments). Now we will use the last theorem and the fact that $z = c'(0)$. Define the new free energy

$$c_t(s) = \lim_{n \to \infty} \frac{1}{n} \log \int e^{sX_n(z)}(z) d\mu(z) = \lim_{n \to \infty} \frac{1}{n} \log \int e^{snx} d\mu_t^n(x)$$

as was done in the last theorem.

One can obtain $c_t(s)$ from $c(s)$ in the following way:

$$\begin{aligned} c_t(s) &= \lim_{n\to\infty} \frac{1}{n} \log \int e^{snx} d\mu_n^t(x) \lim_{n\to\infty} \frac{1}{n} \log \int \frac{e^{nx(s+t)}}{e^{nc_n(t)}} d\mu^n(x) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \int e^{nx(s+t)} d\mu^n(x) - \lim_{n\to\infty} \frac{1}{n} \log e^{c_n(t)n} \\ &= c(t+s) - c(t). \end{aligned}$$

Hence, if c is differentiable on t, then $c_t(s)$ is differentiable at $s = 0$ and $\frac{dc}{dt}(t) = \frac{dc_t}{ds}(0)$. Now, as the hyphotesis of differentiability of the last theorem is satisfied, the conclusion follows (see remark after theorem 6.1):

$$\lim_{n\to\infty} \mu_n^t(B(c_t^{'}(0), r)) = 1$$

Using the fact that we choose t in such manner that $c_t^{'}(0) = c^{'}(t) = z$, we conclude that:

$$\lim_{n\to\infty} \mu_n^t((B(z,r)) = 1$$

and the claim is proved.

Note that introducing the parameter t in our problem (defining the one-parameter family of measures μ_t^n, $n \in \mathbb{N}$), has the effect of translating by t the free energy $c(s)$ (on the parameter s), that is,

$$c_t(s) = c(t+s) - c(t).$$

In other words we adapt the measure μ_t^n in such way that this new measure has mean value z.

Now we will return to the proof of the theorem.

For any point $x \in B(z,r), -tz- \mid t \mid r \leq -tx$. Therefore:

$$\begin{aligned} Q_n(A) \geq Q_n(B(z,r)) &= \int_{B(z,r)} d\mu^n(x) \\ &= Z_n(t) \int_{B(z,r)} e^{-ntx} \mu_t^n(x) \geq e^{n(c_n(t)-tz)-rn|t|} \mu_t^n(B(z,r)). \end{aligned}$$

Hence

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(A) \geq c(t) - tz - r \mid t \mid + \liminf_{n\to\infty} \frac{1}{n} \log \mu_t^n(B(z,r))$$

From the claim we know that the last term in the right hand side of the above expression is zero. Hence, as $c(t) - tz = -I(z)$, because $c'(t) = z$, then

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(A) \geq -I(z) - r \mid t \mid .$$

As r was arbitrary and positive, we conclude finally that

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(A) \geq -I(z).$$

Now as z was arbitrary in the open set A, we obtain that

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(A) \geq -\inf_{z\in A} I(z),$$

and this is the end of the proof of the theorem. ∎

As $I(z)$ is assumed to be continuous (because $c(t)$ is assumed to be differentiable), the final conclusion is:

Theorem 6.3. Suppose $c(t)$ is differentiable in t, then for a given interval C (open or closed)

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(C) = -\inf_{z\in C} I(z).$$

Now we will want to relate the results we obtained above with the Pressure of Thermodynamic Formalism.

7. The Ruelle Operator

In this chapter we will present several results related to the pressure of expanding maps. For such a class of maps the Ruelle Operator will produce a complete solution for the problem of existence and uniqueness of equilibrium states. Theorem 7.2 will explain how to obtain in a constructive way the equilibrium states. We point out that the Bernoulli shift is a very important case where the results we will present can be applied. In this section we will consider only maps that have the property that for each point $z \in X$, $\{T^{-1}(z)\}$ is equal to a fixed value $d > 1$, independent of z. Therefore the results will aply directly to the one-sided shift but not for the two-sided shift (see section 2 for definitions). The results presented here can be extented to the two-sided shift, but this require a certain proposition that we will not present here (see [15]).

Recall the definition:

Definition 7.1. A map T from a compact metric space (X, d) to itself is *expanding* if there exist $\lambda > 1$ such that, for any $x \in X$ there exist $\epsilon > 0$ such that $\forall y \in B(x, \epsilon), d(T(x), T(y)) > \lambda d(x, y)$.

Example. Consider $a_0 = 0 < a_1 < a_2 < a_3 < ... < a_{n-1} < a_n = 1$ a sequence of distict numbers on the interval [0,1]. Suppose T is a differentiable (C^∞) by part map from [0,1] to itself such that $|T'(x)| > \lambda > 1$, for all z different from $a_0, a_1, ...a_n$. Suppose also that for each $i \in \{0, 1, 2, ..n-1\}$, $T([a_i, a_{i+1}]) = [0, 1]$. We will also suppose that T has a C^∞ extension to the values $a_i, i \in \{0, 1, 2, ..., n\}$ with the same properties. This map is expanding and is one of the possible kinds of maps where the results we will present in this section can apply. In fig 3 we show the graph of a map T where all the above properties happen.

Notation. We will use the following notation: for $\phi \in C(X)$ and $\nu \in \mathcal{M}(X)$ or $(\mathcal{S}(X))$ we denote the value $\int \phi(x) d\nu(x)$ by $< \phi, \nu >$.

Definition 7.2. For a given operator $\mathcal{L}$ from $C(X)$ to itself, the *dual of* $\mathcal{L}$ is the operator $\mathcal{L}^*$ defined from the dual space $C(X)^* = \mathcal{S}(X)$ (the space of signed measures) to itself defined in the following way: $\mathcal{L}^*$ is the only operator from $\mathcal{S}(X)$ to itself such that for any $\phi \in C(X)$ and $\nu \in \mathcal{S}(X)$

$$< \mathcal{L}(\phi), \nu >=< \phi, \mathcal{L}^*(\nu) > .$$

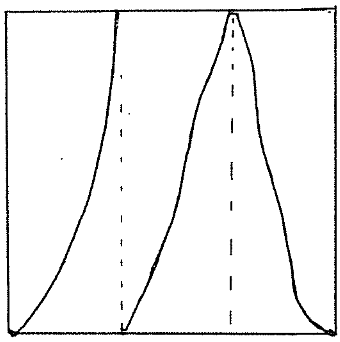

Figure 3.

Remark. Such an operator $\mathcal{L}^*$ is well defined by the Riesz Theorem. This is so because for a given fixed $\nu \in \mathcal{S}(X)$ the operator $\mathcal{H}$ from $C(X)$ to $\mathbf{R}$ given by $\mathcal{H}(\phi) =< \mathcal{L}(\phi), \nu >= \int \mathcal{L}\phi(x) d\nu(x)$ satisfies the hypothesis of the Riesz Theorem. Therefore, there exists a signed-measure μ such that $\int \mathcal{L}\phi(x) d\nu(x) = \mathcal{H}(\phi) = \int \phi(x) d\mu(x) =< \phi, \mu >$. Hence, by definition, $\mathcal{L}^*(\nu) = \mu$.

We will assume in the next theorem that the map T has a fixed degree d, that is, that for any $a \in X$, $\#\{T^{-1}(a)\} = d$. For such a map kind $h(T) = \log d$ (see definition 4.3).

Definition 7.3. Define $\mu_n(x) \in \mathcal{M}(X)$ by

$$\mu_n(x) = \frac{1}{d^n} \sum_{T^n(y)=x} \delta_y,$$

where $d = \#T^{-1}(a)$ independs on $a \in X$.

Theorem 7.1. Let $T : X \hookleftarrow$ be an expanding map of degree d. There exists $\mu \in \mathcal{M}(T)$ such that $\mu = \lim_{n\to\infty} \mu_n(x)$ for any $x \in X$. Moreover μ satisfies:

(1) μ is ergodic and positive on open sets;

(2) $h(\mu) = \log d$;

(3) $h(\eta) < \log d$ for any $\eta \in \mathcal{M}(T)$, $\eta \neq \nu$.

Remark. Remember that $P(0) = \log d = h(T)$ and therefore μ is the equilibrium state for $\psi = 0$ (see definition 4.3). The maximal measure for the one-sided shift in d symbols can be obtained also as the Probability $P(1/d, 1/d, ..., 1/d)$ (see definition 2.7 and remark in the end of section 4).

Definition 7.4. The above defined measure μ is called the *maximal measure.*

Definition 7.5. Suppose that $T : X \hookleftarrow$ is a continuous map and $\psi : X \to \mathbf{R}$ is a continuous function. Remember that we denote by $C(X)$ the space of continuous functions on X. Define $\mathcal{L}_\psi : C(X) \hookleftarrow$ by

$$\mathcal{L}_\psi \phi(x) = \sum_{y \in T^{-1}x} e^{\psi(y)} \phi(y)$$

for any $\phi \in C(M)$ and $x \in M$. We call this operator the *Ruelle-Perron-Frobenius Operator (Ruelle Operator for short).*

It is quite easy to see that:

$$\mathcal{L}_\psi^n \phi(x) = \sum_{y \in T^n(x)} e^{\psi(y)+\psi(T(y))+\psi(T^2(y))+...+\psi(T^{n-1}(y))} \phi(y). \tag{13}$$

A function ψ is called Holder-continuous is there exist $\gamma > 0$ such that $\forall x, y \in X$; $d(T(x), T(y)) < d(x, y)^\gamma$. We will require in the next theorem that the function ψ be Holder and without this hypothesis about ψ the results stated in the theorem will not be necessarily true (see [10] for a counter-example).

Now we will state a fundamental theorem in Thermodynamic Formalism.

Theorem 7.2. (see [3] for a proof) - Let $T : X \hookleftarrow$ be an expanding map and $\psi : X \to \mathbf{R}$ be Hölder-continuous. Then there exist $h : X \to \mathbf{R}$ Hölder-continuous and strictly positive, $\nu \in \mathcal{M}(X)$ and $\lambda > 0$ such that:

(1) $\int h d\nu = 1$;

(2) $\mathcal{L}_\psi h = \lambda h$;

(3) $\mathcal{L}_\psi^* \nu = \lambda \nu$;

(4) $\| \lambda^{-n} \mathcal{L}_\psi^n \phi - h \int \phi d\nu \|_{C(X)} \to 0$ for any $\phi \in C(X)$. ;

(5) h is the unique positive eigenfunction of $\mathcal{L}_\psi$, except for multiplication by scalars ;

(6) The probability $\mu_\psi = h\nu$ is T-invariant (that is, $\mu_\psi \in \mathcal{M}(T)$), ergodic, has positive entropy, is positive on open sets and satisfies

$$\log \lambda = h(\mu_\psi) + \int \psi d\mu.;$$

(7) For any $\eta \in \mathcal{M}(T)$, $\eta \neq \mu_\psi$;

$$\log \lambda > h(\eta) + \int \psi d\eta;$$

In order to explain how one can obtain the equilibrium states μ_ψ associated to ψ in a more appropriate way, we will need to consider a series of remarks.

Remark. It follows from (6) and (7) of Theorem 7.2 that $P(\psi) = \log \lambda$ and that μ_ψ is the unique equilibrium state for ϕ. Therefore the pressure is equial to $\log \lambda$, where λ is an eigenvalue of the Ruelle Operator. In fact, it can be shown that λ is the largest eigenvalue of the operator $\mathcal{L}_\psi$ [3] [15]. The remaider of the spectrum of $\mathcal{L}_\psi$ is contained in a disc (on $\mathbf{C}$) of radius strictly smaller than λ. The multiplicity of the eigenvalue λ is one.

Note that $\mu_\psi \in \mathcal{M}(T)$, but ν is not necessarily in this set.

Remark. The value $P(\psi)$ can be computed in the following way: fix a certain point $x_0 \in X$ and consider ϕ constant and equal to 1 in (4) of Theorem 7.2. As h is bounded (being continuous on a compact space) then from (4) Theorem 7.2

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\mathcal{L}_\psi^n 1(x_0)}{\lambda^n} = 0$$

that is,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathcal{L}_\psi^n 1(x_0) = \log \lambda = P(\psi) \tag{14}$$

or

$$\lim_{n\to\infty}\frac{1}{n}\log\sum_{y\in T^n(x_0)} e^{\psi(y)+\psi(T(y))+\ldots+\psi(T^{n-1})(y)} = P(\psi). \tag{15}$$

Remark. The ëigenfunction h can be obtained with the following procedure: consider ϕ constant equal 1 in (4), then

$$h(x) = \lim_{n\to\infty}\frac{\mathcal{L}_\psi^n 1(x)}{\lambda^n} \tag{16}$$

Remark. In order to obtain μ, we just need to obtain ν. The probability ν can be obtained from Theorem 7.2 (4): consider a certain value x_0 and δ_{x_0}, then from (4)

$$h(x_0)\nu = \lim_{n\to\infty}\frac{\mathcal{L}_\psi^{n*}(\delta_{x_0})}{\lambda^n} = \lim_{n\to\infty}\sum_{T^n(x)=x_0}\frac{e^{\psi(x)+\psi(T(x))+\ldots+\psi(T^{n-1}(x))}}{\lambda^n}\delta_x \tag{17}$$

Therefore ν can be obtained in the above mentioned way.

In this way we can obtain ν by means of the limit of a sequence of finite sum of Dirac measures on the preimages of the point x. In the case of the maximal measure ($\psi = 0, P(0) = \log d, \lambda = d, h = 1, \nu = \mu = \mu_\psi$), the weights in the points x such that $T^n(x) = x_0$ are evenly distribuited and equal to d^{-n}. For the general Holder continuous map ψ, it is necessary to distribute the weights in a different form as above. There is a more apropriate way to obtain directly the equilibrium measure μ_ψ, that will be presented later.

Remark. If one is interested in finding an invariant measure μ for the map T, given in the example after Definition 7.1, and that has also a density ρ with respect to dx, that is $d\mu(x) = \rho(x)dx$, then one should consider the potential $\psi(x) = -\log|T'(x)|$. In this case, it is not difficult to check that Theorem 7.2 gives $\lambda = 1$ and $h(x) = \rho(x)$ (see [13]). The equilibrium probability $d\mu$ (satisfying (6) Theorem 7.2) will be in this case the measure $\rho(x)dx$.

Let us see now how Theorem 7.1 follows from Theorem 7.2. Take $\psi \equiv 0$ and let λ, h and ν be given by Theorem 7.2. Then

$$\mathcal{L}_\psi 1(x) = \sum_{y\in T^{-1}x} 1(y) = d\,1.$$

Because of part (5) of Theorem 1.2, $d = \lambda$ and $h \equiv 1$. Also, part (4) of theorem 7.2 shows that

$$\frac{1}{d^n} \sum_{y \in T^{-n}x} \varphi(y) \longrightarrow \int \varphi d\nu$$

for any $\varphi \in C(X)$. This proves Theorem 7.1.

Definition 7.6. A continuous function $J : X \to \mathbf{R}$ is the *Jacobian* of T with respect to $\mu \in \mathcal{M}(X)$ if

$$\mu(T(A)) = \int_A J d\mu$$

for any Borel set A such that $T \mid_A$ is injective.

It is easy to prove that such a J exists (use the Radon-Nykodin Theorem) and it is unique (in the apropriate sense). The Jacobian is the local rate of variation of the measure μ by means of forward iteration of the map. Some ergodic properties of μ can be analysed through J.

Theorem 7.3. Suppose that J (the Jacobian of an invariant measure μ) is Hölder-continuous and strictly positive. Then

(a) $h(\mu) = \int \log J d\mu$;

(b) μ is ergodic.

Consider now the question of finding a T-invariant probability with a given Jacobian $J > 1$. It is easy to prove that every function $J > 1$ that is the Jacobian of T with respect to some T-invariant probability must satisfy

$$\sum_{T(x)=y} \frac{1}{\mid J(x) \mid} = 1 \tag{18}$$

for any $y \in X$. This condition is also sufficient.

Theorem 7.4. Let $T : X \hookleftarrow$ be an expanding map and $J : X \to \mathbf{R}$ strictly positive and Hölder-continuous, the Jacobian of $\eta \in \mathcal{M}(T)$. Consider $\psi = -\log J$, then the equilibrium state $\mu_\psi = \eta$, h is constant equal 1 and $P(-\log J) = 0$.

Proof. From (18) and condition (2) of Theorem 7.2, $h \equiv 1$ and $\lambda = 1$ in the last theorem. Hence $P(-\log J) = 0$. ■

Theorem 7.5. Suppose ψ is Hölder continuous, μ_ψ is the equilibrium state associated with ψ and h is the eigenfunction associated with λ in Theorem 7.2, then the Jacobian J_ψ of the probability μ_ψ is given by:

$$J_\psi(x) = \lambda e^{-\psi(x)} \frac{h \circ T(x)}{h(x)} \tag{19}$$

Remark. It follows from the last expression that

$$\psi(x) - (-\log J_\psi(x) = \log(h \circ T(x)) - \log h(x) + \lambda \tag{20}$$

That is ψ and $-\log J_\psi$ satisfies the homology criterium (Proposition(4.1)) and therefore they determine the same equilibrium state, that is $\mu_\psi = \mu_{-\log J_\psi}$. Remember that $P(-\log J_\psi) = \log \lambda = \log 1 = 0$.

It follows from the last claims and from

$$\lim_{n \to \infty} \mathcal{L}^n_{-\log J_\psi}(\phi) = \int \phi(x) d\mu_\psi(x)$$

(see (4) in Theorem 7.2) that the equilibrium state μ_ψ can be obtained in the following way:

$$\mu_\psi = \lim_{n \to \infty} \sum_{T^n(y)=x} e^{-\log J_\psi(y) - \log J_\psi(T(y)) - \ldots - \log J_\psi(T^{n-1}(y))} \delta_y \tag{21}$$

$$= \lim_{n \to \infty} \sum_{T^n(y)=x} (J_\psi(y) J_\psi(T(y)) \ldots J_\psi(T^{n-1}(y)))^{-1} \delta_y \tag{22}$$

Hence from λ and h one can obtain μ_ψ as the limit of a sum of weights placed in the preimages of a point $x \in X$ (J_ψ is given by (19)).

Example. We will consider now the example mentioned in section 4, just after Definition 4.1. In fact we can analyze a more general example where we will be able to show explicitely the equilibrium probability. Consider p(+,+), p(+,-), p(-,+) and p(-,-) non-negative numbers such that p(+,+) + p(+,-) = 1 and p(-,+)

+ p(-,-) = 1. These numbers p(i,j), i,j $\in \{+,-\}$ express the probability of having spin j at the right of spin i in the lattice **Z**.

Consider the matrix

$$A = \begin{pmatrix} p(+,+) & p(-,+) \\ p(+,-) & p(-,-) \end{pmatrix}$$

It can be shown [15] that this matrix A has the value 1 as the larger eigenvalue (this result is known in the usual textbooks on Matrix Theory as the Perron-Frobenius Theorem) and we will denote by (p(+),p(-)) the normalized eigenvalue associated to the eigenvalue 1, that is:

$$A(p(+),p(-)) = (p(+),p(-)) \ , \ p(+) + p(-) = 1.$$

Now we can define a measure μ on cylinders (and then extend to the more general class of Borel sets) by:

$$\mu(\overline{i_0, i_1, i_2, ..., i_n}) = p(i_0, i_1)p(i_1, i_2)...p(i_{(n-1)}, i_n)p(i_n),$$

$n \in \mathrm{N}$, $i_0, i_1, i_2, ..., i_n \in \{+,-\}$. It quite easy to see that considering in Theorem 7.2 the potential ψ constant in each one of the four cylinders given by:

a) $\psi(z) = \log p(+,+)\ \forall z \in (\overline{+,+})$,

b) $\psi(z) = \log p(+,-)\ \forall z \in (\overline{+,-})$,

c) $\psi(z) = \log p(-,+)\ \forall z \in (\overline{-,+})$ and

d) $\psi(z) = \log p(-,-)\ \forall z \in (\overline{-,-})$,

then the eigenfunction h is constant equal 1 and λ equal 1. It is not difficult to see that the measure μ given above satisfies the equation (3) in Theorem 7.2 (see also Definition 7.2), that is $\mathcal{L}^*_\psi \mu = \mu$ (first show that $\mathcal{L}^*_\psi \mu(B) = \mu(B)$, for the cylinders B depending on the two first coordinates, and then depending on three coordinates, and so on...). Therefore μ is the equilibrium state for the ψ given above.

This example shows that the Ruelle Operator is in fact an extension of the Perron-Frobenius Operator of Matrix Theory (finite dimension) to the infinite dimensional space of functions.

The Jacobian of the measure μ is constant by parts and is constant in each cilinder (see Theorem 7.5)

$$J(z) = e^{-\psi(z)} = p(i,j)^{-1}, \forall z \in \overline{(i,j)}, i,j \in \{+,-\}.$$

The above described example includes the one we mention before in section 4.

Theorem 7.6. Suppose T is a continuous map from X to X, X is a compact metric space and $h(T)$ is finite. Consider ν a finite signed measure on the Borel σ-algebra of X. Then the following properties are equivalent:

$$\text{(a) } \nu \in \mathcal{M}(T)$$

and

$$\text{(b) } \forall \phi \in C(X)\ , < \phi, \nu > \leq P(\phi). \tag{23}$$

Proof. (a) $\to$ (b)

By definition of Pressure, $< \phi, \nu > \leq P(\phi)$, because $\nu \in \mathcal{M}(T)$ and $h(\nu) \geq 0$.

(b) $\to$ (a)

Suppose ν satisfies (b), then we will show first that ν is a measure, that is, for any non-negative continuous function $\phi, < \phi, \nu > \geq 0$.

Consider $\phi \in C(X)$ such that $\phi(x) \geq 0, \forall x \in X$, then given $n \in \mathrm{N}$ and $\delta > 0$

$$\int (\phi(x) + \delta) n d\nu(x) \geq -P(-(\phi + \delta)n)$$

by assumption (b). By definition of pressure and from the fact that ϕ is nonegative

$$-P(-(\phi+\delta)n) = -\sup_{\mu \in \mathcal{M}(T)} \{h(\mu) - \int (\phi(x) + \delta) n d\mu(x)\} \geq$$

$$-(h(T) - \inf_{x \in X} \{(\phi(x) + \delta)n\}) \geq -h(T) + n\delta$$

For large n the last expression is positive. As δ was arbitrary, it follows that $\int \phi(x) d\nu(x) =< \phi, \nu > \geq 0$. Hence, ν is a measure. Now we will show that ν is a probability, that is that, $\nu(X) = 1$.

For $n \in \mathbf{Z}$ $n \int d\nu(x) = \int n d\nu(x) \leq P(n) = h(T) + n$, therefore $\nu(X) \leq \frac{h(T)}{n} + 1$, if $n > 0$, and $\nu(X) \geq \frac{h(T)}{n} + 1$, if $n < 0$.

Now letting n go to ∞ in the first expression and n to $-\infty$ in the second we conclude that $\nu(X) = 1$.

This means that ν is a probability. Finally we will show that $\nu \in \mathcal{M}(T)$, that is, we will show that for any $\phi \in C(X)$, $\int \phi(x) d\nu(x) = \int \phi(T(x)) d\nu(x)$. In other words we have to show that $< \phi \circ T - \phi, \nu >= 0$.

For a given $n \in \mathbf{Z}$, $n < \phi \circ T - \phi, \nu > \leq P(\phi \circ T - \phi) n$ by assumption (b). Now using the criteria of the homology we have that the last term is $P(0) = h(T)$. Hence $< \phi \circ T - \phi, \nu > \leq \frac{h(T)}{n}$, if $n > 0$, and $< \phi \circ T - \phi, \nu > \geq \frac{h(T)}{n}$, if $n < 0$.

Now letting n go to ∞ in the first expression and n to $-\infty$ in the last expression we conclude that $< \phi \circ T - \phi, \nu >= 0$. Thus the Theorem is proved. ■

The Pressure $P(\psi)$ is a continuous function of ψ (see[18]); one could ask if the entropy $h(\nu)$ is continuous in $\nu \in \mathcal{M}(T)$, that is, wheter

$$w_n \in \mathcal{M}(T)$$

converging weakly to ν (see definition 2.9) implies $\lim_{n \to \infty} h(w_n) = h(\nu)$.

An equilibrium state μ_ψ can be obtained as a limit of finite sums of Dirac measures on periodic orbits of arbitrarily large period. We did not prove this fact, but from the expression (17) in this paragraph (in fact expression (17) is for preimages and not for periodic orbits) it is quite reasonable to believe that the above claim is true (see remark before Proposition 2.2).

Another reason supporting the above claim is the fact that for an expanding map the periodic orbits are dense in the support of any invariant measure(see [13]) (see proposition 2.2 for a proof in the case T is the shift).

The entropy of an invariant measure with support on a periodic orbit is zero (see example after Theorem 3.1), therefore as the entropy of an equilibrium state is positive (theorem 7.2 (6)), one concludes that the entropy is not continuous. The entropy can *jump up* in the limit.

The entropy however can not *jump down* in the limit as it is stated in Theorem 7.8. We need first to state more precisely what we mean by that.

Definition 7.7. A function F on a space $\mathcal{M}$ is *upper-semicontinuous* at ν if for any convergent sequence $w_n \in \mathcal{M}, n \in \mathrm{N}$ such that $\lim_{n\to\infty} w_n = \nu \in \mathcal{M}$, then

$$\lim_{n\to\infty} F(w_n) \lneqq F(\nu).$$

Theorem 7.7. (see [18] for a proof) Suppose T is a continuous map from X to X, where X is a compact metric space, and that $h(T) = \sup_{\nu\in\mathcal{M}(T)}\{h(\nu)\}$ is finite. For a given probability $\nu \in \mathcal{M}(T)$ the following statements are equivalent:

(a) $h(\nu) = \inf_{\phi\in C(X)}\{P(\phi) - <\phi, \nu>\}$;

(b) the entropy is upper-semicontinuous at ν.

Theorem 7.8. (see [18] for a proof) For expanding systems the entropy is upper-semicontinuous at any probability $\nu \in \mathcal{M}(T)$.

Remark. From the two results presented above one can conclude that a measure ν is invariant for an expanding map T, if and only if

$$h(\nu) = \inf_{\phi\in C(X)}\{P(\phi) - <\phi, \nu>\} = -\sup_{\phi\in C(X)}\{<\phi, \nu> - P(\phi)\}. \tag{24}$$

Therefore the entropy is minus the Legendre Transform of the Pressure.

Remember that the dual of $C(X)$ is $\mathcal{S}(X)$ and that Pressure is defined for continuous functions and entropy for elements of $\mathcal{M}(T) \subset \mathcal{S}(X)$.

Proposition 6.3 claims that in the finite dimensional case the Legendre transform is an involution, that is, $f^{**} = g$. Therefore, one could also expect that the Legendre transform of minus the entropy should be the P ressure. This is so because, by definition,

$$P(\psi) = \sup_{\nu\in\mathcal{M}(T)}\{<\psi, \nu> - (-h(\nu))\}.$$

The disturbing point in the above expression is that we are taking supremum in a smaller set $\mathcal{M}(T)$ and not in the dual of $C(X)$, that is in the set $\mathcal{S}(X)$. If we define the entropy of a signed measure η by

$$h(\eta) = \inf_{\psi\in C(X)}\{P(\psi) - <\psi, \eta>\}$$

as in (24), then $h(\eta) < 0$ for $\eta \in \mathcal{S}(X) - \mathcal{M}(T)$ (see theorem 7.6).

Hence we finally can state that:

$$P(\psi) = \sup_{\nu \in \mathcal{S}(X)} \{< \psi, \nu > -(-h(\nu))\}, \tag{25}$$

because the entropy of non-invariant measures will not interfere in the supremum and the analogy with the finite dimensional case is complete.

For results about Large Deviation properties in this setting (level-2 large deviation) we refer the reader to [8]. In the next paragraph we will consider large deviation properties, but in another setting (the level-1 large deviation). The terminology of level-1 and level-2 is explained in more detail in [7]. The reference [7] is an excelent source of results for large-deviation, but does not consider the entropy (Kolmogorov-Shanon entropy) and presure as we are doing here.

We will repeat definition 6.3 but now for the infinite dimensional case.

Definition 7.8. For a given convex function K from $C(X)$ to $\mathbf{R}$, we call a signed measure $\mu \in \mathcal{S}(X)$ (the dual of $C(X)$) a *subdifferential* of K at the value η and write $\mu = \delta K(\eta)$, if the following is true: for any $\psi \in C(X)$,

$$K(\psi) \geq K(\eta) + < \psi - \eta, \mu > .$$

Notation. As the pressure $P(\psi)$ is convex in ψ we can consider the above definition for the pressure and we will denote the subset of signed-measures μ that are subdifferential of P at the value η by $t(\eta)$. In other words,

$$t(\eta) = \delta P(\eta) = \{\mu \in \mathcal{S}(x) \,|\, P(\psi) \geq P(\eta) + \int (\psi(x) - \eta(x)) d\mu(x), \forall \psi \in C(X)\}. \tag{26}$$

Remember that for a continuous function ψ, the set of probabilities μ such that $P(\psi) = h(\mu) + \int \psi(x) d\mu(x)$ is called the set of equilibrium measures. The main Theorem stated in the beginning of this section is that for an expanding map T and a Hölder continuous function ψ, equilibrium states exist and are unique.

Theorem 7.9. (see [18]) Suppose T is an expanding map such that $h(T)$ is finite. If ψ is a continuous function on X, then $t(\psi)$ is the set of equilibrium states for ψ. The set $t(\psi)$ is not empty.

The next result improves the claim that for expanding systems the subdifferential of the pressure P at ψ is μ_ψ (that is, $\delta P(\psi) = \mu_\psi$).

Theorem 7.10. Suppose that T is an expanding map. Given f and g Holder continuous functions, the function

$$p(t) = P(f + tg)$$

is convex and real analytic in t. The value $p'(t)$ is equal to $\int g(x) d\mu_{f+tg}(x)$.

Proof. We refer the reader to [15] [17] for the proof of the differentiability of $p(t)$. We will assume that p is differentiable and we will show that $p'(t) = \int g(x) d\mu_{f+tg}$.

We will reduce the question to its simplest form in order to simplify the argument.

First note that it is enough to show that $\frac{d}{dt} P(f + tg)|_{t=0} = \int g d\mu_f$. For the general case consider $P((f + tg) + sg)$ and take derivative at $s = 0$.

Another simplification is that we can substitute f by $-\log J$ where J is the Jacobian of μ_f. In fact (see the Remark after theorem 7.5)

$$(f + tg) - (-\log J + tg) = P(f) + \log(h \circ T) - \log h,$$

and therefore $f + tg$ and $\log J + tg$ are homologous. Hence $\mu_{f+tg} = \mu_{-\log J+tg}$ and furthermore $P(f + tg) = P(-\log J + tg) + P(f)$. Taking derivative with repect to t in both sides of the last expression:

$$\frac{d}{dt} P(-\log J + tg) = \frac{d}{dt} P(f + tg).$$

Note that from (22), for any ϕ the integral

$$\int \phi d\mu_{-\log J} = \lim_{n \to \infty} \sum_{T^n(y) = x_0} \phi(y) e^{-\sum_{j=0}^{n-1} \log J(T^j(y))} \tag{27}$$

$$= \lim_{n \to \infty} \mathcal{L}^n_{-\log J} \phi(x_0) \tag{28}$$

where x_0 is a certain point in X.

We will use the above property very soon.

One of the Remarks after Theorem 7.2 states that (see (15))

$$P(-\log J+tg)=\lim_{n\to\infty}\frac{1}{n}\log\sum_{T^n(y)=x_0} e^{\sum_{j=0}^{n-1}(-\log J+tg)(T^j(y))},$$

hence, derivativing term by term (the fact that this is possible is a crucial step that will not be proved here [15][17]) one obtains:

$$\frac{d}{dt}P(-\log J+tg)=\lim_{n\to\infty}\frac{1}{n}\frac{\sum_{T^n(y)=x_0}\sum_{j=0}^{n-1} g(T^j(y))e^{\sum_{j=0}^{n-1}(-\log J+tg)(T^j(y))}}{\sum_{T^n(y)=x_0} e^{\sum_{j=0}^{n-1}(-\log J+tg)(T^j(y))}}.$$

Now in the last expression considering $t=0$ we obtain

$$\frac{d}{dt}P(-\log J+tg)|_{t=0}=\lim_{n\to\infty}\frac{1}{n}\frac{\sum_{T^n(y)=x_0}\sum_{j=0}^{n-1} g(T^j(y))e^{-\sum_{j=0}^{n-1}\log J(T^j(y))}}{\sum_{T^n(y)=x_0} e^{-\sum_{j=0}^{n-1}\log J(T^j(y))}} \tag{29}$$

Claim. $\sum_{T^n(y)=x_0} e^{-\sum_{j=0}^{n-1}\log J(T^j(y))}=1, \forall n\in \mathrm{N}, \forall x_0\in X$

Proof of the Claim. The proof is by induction. The claim is true for $n=1$ by (18). Suppose the claim is true for n, then we will prove that the claim is true for n+1.

In fact

$$\sum_{T^{n+1}(y)=x_0} e^{-\sum_{j=0}^{n}\log J(T^j(y))}=$$

$$\sum_{T(z)=x_0} e^{-\log J(z)}\sum_{T^n(y)=z} e^{-\sum_{j=0}^{n-1}\log J(T^j(y))}=\sum_{T(z)=x_0} e^{-\log J(z)}\,1=1.$$

In the last two equalities we used the fact that the claim is true for n and 1.

This is the end of the proof of the claim.

Now, we return to the proof of the Theorem. It follows from the claim and (29)(27)(28) (taking $\phi=g\circ T^j$) that:

$$\frac{d}{dt}P(-\log J+tg)|_{t=0}=\lim_{n\to\infty}\frac{1}{n}\sum_{T^n(y)=x}\sum_{j=0}^{n-1} g(T^j(y))e^{\sum_{j=0}^{n-1}-\log J(T^j(y))}$$

$$=\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\mathcal{L}^n_{-\log J}(g(T^j))(x_0) \tag{30}$$

As the convergence in Theorem 7.2 (4) is uniform (and the eigenfunction h of theorem 7.2 is constant equal 1 for $\psi = -\log J$ by Theorem 7.4), then for an y ϵ, there exist $N > 0$ such that for any $n \in \mathrm{N}, n > N$ and $z \in X$,

$$|\mathcal{L}_{-\log J}\, g(z) - \int g(x) d\mu_{-\log J}(x)| \leq \epsilon.$$

Therefore, from (30), considering z varying under the form $T^j(x_0)$

$$\frac{d}{dt} P(-\log J + tg)|_{t=0} = \int g(x) d\mu_{-\log J}(x).$$

Finally, we conclude that:

$$\frac{d}{dt} P(f + tg)|_{t=0} = \frac{d}{dt} P(-\log J + tg)|_{t=0} = \int g(x) d\mu_{-logJ}(x) = \int g(x) d\mu_f(x) \tag{31}$$

and this is the end of the proof of the Theorem. ■

Theorem 7.11. (see [18]) Suppose T is an expanding map on X and $h(T)$ is finite, then there exists a dense subset B of $C(X)$, such that for ψ in B, there exists just one equilibrium state for ψ, that is, the cardinal of $t(\psi)$ is 1.

8. Pressure and Large Deviation

In this paragraph we will show a result relating large deviation with pressure. It is possible to obtain very precise results about the deviation function for Holder functions and the maximal measure of an expanding map.

Notation. Let z_0 be a point of X, and for each $n \in \mathrm{N}$, denote by $z(n, i, z_0)$, $i \in \{1, 2, 3, ..., d^n\}$ the d^n solutions of the equation

$$T^n(z) = z_0.$$

We know that the maximal entropy measure (see theorem 7.1) μ can be obtained as

$$\mu = \lim_{n \to \infty} d^{-n} \sum_{i=1}^{d^n} \delta_{z(n,i,z_0)}.$$

Notation. In this section we will will denote by μ the maximal entropy measure (see theorem 7.1).

Given $0 < \gamma < 1$, denote by $C(\gamma)$ the space of Hölder-continuous real-valued functions in X endowed with the metric

$$\| g \| = \| g \|_0 + \sup_{x \neq y} \frac{| g(x) - g(y) |}{| x - y |^\gamma}$$

where $\| g \|_0$ is the usual supremum norm.

Theorem 8.1. Let T be an expanding map, and $g \in C(\gamma)$, then

$$P(g) = \lim_{n \to \infty} n^{-1} \left[\log \int \exp\left(\sum_{j=0}^{n-1} g(T^j(z)) \right) d\mu(z) \right] + \log d.$$

where μ is the maximal measure.

Proof. Let g be a Hölder-continuous function on the compact set X.

Let us consider a fixed $z_0 \in X$ and denote by $z(n, i)$ the $z(n, i, z_0)$, $n \in \mathrm{N}$ and $i \in \{1, 2, 3, ..., d^n\}$.

For a given $n \in \mathrm{N}$,

$$\int \exp\left(\sum_{j=0}^{n-1} g(T^j(z)) \right) d\mu(z) = \lim_{m \to \infty} d^{-m} \sum_{i=1}^{d^m} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(m, i))) \right)$$

$$= \lim_{m \to \infty} d^{-m} \sum_{k=1}^{d^{m-n}} \sum_{i=1}^{d^n} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(n, i, (z(m-n, k))))) \right).$$

From [3] (in this moment the hypothesis about expansivity and Holder-continuous are essential), there exist constants C_1, c_1 such that for n large enough

$$c_1 \sum_{i=1}^{d^n} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(n, i, z))) \right) \leq \sum_{i=1}^{d^n} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(n, i))) \right)$$

$$\leq C_1 \sum_{i=1}^{d^n} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(n, i, z))) \right) \qquad (1)$$

for any $z \in X$.

Therefore,

$$c_1 d^{-m} d^{m-n} \sum_{i=1}^{d^n} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(n,i))) \right)$$

$$\leq d^{-m} \sum_{k=1}^{d^{m-n}} \sum_{i=1}^{d^n} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(n,i,(z(m-n,k))))) \right)$$

$$\leq C_1 d^{-m} d^{m-n} \sum_{i=1}^{d^n} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(n,i))) \right).$$

From this, it follows that

$$\lim_{n\to\infty} n^{-1} \log \int \exp\left(\sum_{j=0}^{n-1} g(T^j(z)) \right) d\mu(z)$$

$$= \lim_{n\to\infty} n^{-n} \log \sum_{i=1}^{d^n} \exp\left(\sum_{j=0}^{n-1} g(T^j(z(n,i))) \right) - \log d.$$

Now from the expression of the pressure that appears as a Remark after theorem 7.2 (see expression 7.15) the claim of the theorem is proved. ■

Remark. Consider the free-energy $c(t)$ of a continuous function g and the maximal measure μ. Suppose g is Hölder-continuous, then from the definition 5.3 of the free-energy $c(t), t \in \mathbf{R}$ one concludes from the last theorem that $P(tg) = c(t) + \log d$. Remember that the free-energy depends on the function and on the measure we are considering.

Theorem 8.2. The free-energy c(t) for a Hölder-continuous function g and the maximal measure μ satisfies

$$c(t) = P(tg) - \log d. \tag{32}$$

Therefore c(t) is differentiable and g has the exponential convergence property.

Proof. If c(t) is differentiable, then g has the exponential convergence property for μ (see proposition 6.8). Since $c(t) = P(tg) + \log d$ (from last theorem) and $P(tg)$ is differentiable (theorem 7.10), the results follows. ■

It is quite natural to ask if one can obtain the deviation function

$$I(v) = \sup_{t \in \mathbf{R}}\{tv - c(t)\}$$

from results of Thermodynamic Formalism. The next theorem answers this question.

Theorem 8.3. Suppose g is Hölder-continuous, μ is the maximal measure and $p(t) = P(tg), t \in \mathbf{R}$. Then the deviation function is

$$I(v) = \log d - h(\mu_{t_0 g}), \tag{33}$$

where $\mu_{t_0 g} = \mu_\psi$ is the equilibrium state for $\psi = t_0 g$ and t_0 satisfies $p'(t_0) = v$.

Proof. By definition

$$I(v) = \sup_{t \in \mathbf{R}}\{tv - c(t)\} = \sup_{t \in \mathbf{R}}\{tv - (P(tg) - \log d)\} = \sup_{t \in \mathbf{R}}\{tv - p(t)\} + \log d.$$

It is easy to see that p(t) is convex and from theorem 7.10 p(t) is also differentiable. Suppose t_0 is the unique value such that $p'(t_0) = v$, then from last theorem and the definition of pressure

$$\begin{aligned} I(v) &= \sup_{t \in \mathbf{R}}\{tv - p(t)\} + \log d = t_0 v - p(t_0) + \log d \\ &= t_0 v - h(\mu_{t_0 g}) - \int t_0 g(x) d\mu_{t_0 g}(x) + \log d. \end{aligned}$$

Now from Theorem 7.10 $v = p'(t_0) = \int g(x) d\mu_{t_0 g}(x)$, and the claim of the Theorem follows. ■

In conclusion, for $g \in C(\gamma)$ and the maximal measure μ one can obtain the value of $I(v)$, $v \in \mathbf{R}$ by $I(v) = \log d - h(\mu_{t_0 g})$ where t_0 satisfies $p'(t_0) = v$.

Remark. More general results about large deviations and free-energy of Hölder functions g and equilibrium states μ_g can be obtained, but we will not consider such questions here. We refer the reader to [5],[8],[9] for interesting results in this subject. Theorem 3 in [8] is not correctly stated, but is not necessary for the proof of Theorem 7, the main result of [8].

References

[1] Billingsley, P., Probability and Measure, John Wiley (1979).

[2] Billingsley, P., Ergodic Theory and Information, John Wiley (1965).

[3] Bowen, R., Equilibrium States and Ergodic Theory of Anosov Diffeomorphisms, *Lecture Notes in Math.*, Springer Verlag **470** (1975).

[4] Brin, M., A. Katok, On Local Entropy, *Geometric Dynamics - Lecture Notes in Math.*, Springer Verlag **1007**, 30-38 (1983).

[5] Collet, P., J. Lebowitz, A. Porzio, The Dimension Spectrum of Some Dynamical Systems, *Journal of Statistical Physics* **47**, 609-644 (1987).

[6] Devaney, R., An Introduction to Chaotic Dynamical Systems, Benjamin (1986).

[7] Ellis, K., Entropy, Large Deviation and Statistical Mechanics, Springer-Verlag (1985).

[8] Lopes, A., Entropy and Large Deviation, *NonLinearity* **3**(2), 527-546 (1990).

[9] Lopes, A., Dimension Spectra and a Mathematical Model for Phase Transitions, *Advances in Applied Mathematics* **11**(4), 475-502 (1990).

[10] Lopes, A., A First Order Level-Two Phase Transition, *Journal of Statistical Physics* **60**(3/4), 395-411 (1990).

[11] Lopes, A., The Dimension Spectrum of the Maximal Measure, *SIAM Journal of Matematical Analysis* **20**(5), 1243-1254 (1989).

[12] Lopes, A., The Zeta Function, Non-Differentiability of Pressure and the Critical Exponent of Transition, *Advances in Mathematics*, to appear, preprint 1990.

[13] Mañe, R., Ergodic Theory and Differentiable Dynamics, Springer-Verlag (1987).

[14] Mañe, R., On the Hausdorff Dimension of the Invariant Probabilities of Rational Maps, *Lecture Notes in Math.* **1331**, 86-116, Springer-Verlag (1990).

[15] Parry, W., M. Pollicott, Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics, *Asterisque* **187-188** (1990).

[16] Rudin, W., Real and Complex Analysis, McGraw-Hill (1974).

[17] Ruelle, D., Thermodynamic Formalism, Addison-Wesley (1978).

[18] Walters, P., An Introduction to Ergodic Theory, Springer-Verlag (1981).