Thermodynamic Formalism, Maximizing Probabilities and Large Deviations

Artur O. Lopes

Text under construction - 2017

Dynamique en Cornouaille - France - 2011

1 Introduction

The first sections of this text is part of the paper [6]

We are going to present some of the basic results of the Thermodynamic Formalism of Bowen-Ruelle-Sinai. Some of the proofs we present follow the reasoning of [68] and [28].

Let (M, d) be a compact metric space. In most of the cases here $M = \{1, 2, ..., d\}$. But we an also suppose M is a connected and compact manifold. We denote by $\mathcal{B}$ the Bernoulli space $M^{\mathbb{N}}$ of sequences represented by $x = (x_0, x_1, x_2, x_3, ...)$, where $x_i, i \geq 0$ belongs to the space (alphabet) M. By Tychonoffs Theorem of compactness, we know $\mathcal{B}$ is a compact metric space when equipped with the distance given by $d_c(x, y) = \sum_{k\geq 0} \frac{d(x_k,y_k)}{c^k}$, with $c > 1$. The topologies generated by d_{c_1} or d_{c_2} are the same. We denote d when we choose $c = 2$.

In the text when we talk about the general XY Model we consider $M = \mathbb{S}^1$, the unit circle [6]. Then, da represent Lebesgue measure on the circle, or, on the interval $[0, 1)$.

The shift σ on $\mathcal{B}$ is defined by $\sigma((x_0, x_1, x_2, x_3, ...)) = (x_1, x_2, x_3, x_4, ...)$. It is a continuous function on $\mathcal{B}$.

We refer the reader to [80], [69] or [51] for general results in Ergodic Theory.

Let $A : \mathcal{B} \to \mathbb{R}$ be an *observable* or *potential* defined in the Bernoulli space $\mathcal{B}$, i.e. a real-valued function defined on $\mathcal{B}$. The potential A describe an interaction between sites in the one-dimensional lattice $M^{\mathbb{N}}$.

The elements of M are called spins. We are interested on the equilibrium probability (over the Borel sigma-algebra of $\mathcal{B}$) for the interaction of spins described by the potential A. This concept will be carefully described later on section 4. In Statistical Mechanics this corresponds to one of the main problems: understanding the equilibrium (here on the one-dimensional lattice $\mathcal{B}$) under the action of an Hamiltonian A at a positive temperature T. The same problem at temperature zero will be considered on section 3.

The basic tool for understanding such kind of problem is the Ruelle operator. When $M = \{1, 2, ..., d\}$ the Ruelle operator $\mathcal{L}_A$ acts on functions ψ in the following way

$$\varphi(x) = \mathcal{L}_A(\psi)(x) = \sum_{a=1}^{d} e^{A(ax)}\psi(ax).$$

By this we mean $\mathcal{L}_A(\psi) = \varphi$.

We will consider bellow a more general form of this operator when there is an a priory probability da on M. In this way we will get $\mathcal{L}_A(\psi)(x) = \int_M e^{A(ax)}\psi(ax)\, da$.

In fact given a completely general a priori probability ν on the Borel sigma algebra of a metric compact space M then

$$\psi \to \mathcal{L}_A(\psi) = \phi \text{ such that for all } \; x \; \text{ we have } \phi(x) = \int_M e^{A(ax)}\psi(ax)\, d\nu(a),$$

defines an operator $\mathcal{L}_A = \mathcal{L}_A^\nu$ such that all of the results of the present section can be applied (see [63]).

We will be interested on finding a a positive eigenfunction for such operator $\mathcal{L}_A$.

The probability ν has no dynamical meaning.

In the case $M = \{1, 2, .., d\}$ and $\nu = \frac{1}{d}\sum_{j=1}^d \delta_j$, then $\mathcal{L}_A$, takes the form

$$\mathcal{L}_A(\psi)(x) = \frac{1}{d}\sum_{a=1}^d e^{A(ax)}\psi(ax).$$

All the proofs work in the same way for $\mathcal{L}_A(\psi)(x) = \sum_{a=1}^d e^{A(ax)}\psi(ax)$. This corresponds to take as a priori probability $\nu = \sum_{j=1}^d \delta_j$.

More precisely the main issue in the classical Thermodynamic Formalism is to get an eigenfunction for the Ruelle operator, that is the existence of φ and $\lambda > 0$ such that

$$\mathcal{L}_A(\varphi)(x) = \sum_{a=1}^d e^{A(ax)}\varphi(ax) = \lambda\varphi(x).$$

If given B we are able to get a solution (the a priori probability is $\frac{1}{d}\sum_{j=1}^d \delta_j$)

$$\frac{1}{d}\sum_{a=1}^d e^{B(ax)}\varphi(ax) = \lambda\varphi(x),$$

then, such φ obtained for $B = A - \log d$ we will solve the previous problem.

Examples of potentials where on can find an explicit eigenfunction appears in [17].

For most of the results we consider here we will require A to be Holder-continuous, which means there exist constants $0 < \alpha < 1$ and $Hol_A > 0$ such that $|A(x) - A(y)| \leq Hol_A d(x,y)^\alpha$. We call α the exponent of A and Hol_A the constant for A.

$\mathcal{H}_\alpha$ denotes the set of α-Holder-continuous funtions

We will be interested here in the Gibbs state μ_A associated to such A, which will be a probability measure on $\mathcal{B}$. Note that the set of probabilities on $\mathcal{B}$ is compact for the weak* topology, (which is given by a metric).

For $w \in \mathcal{H}_\alpha$, denote $|w|_\alpha = \sup_{x\neq y}\frac{|w(x)-w(y)|}{d(x,y)^\alpha}$. It is known that $\mathcal{H}_\alpha$ is a Banach space for the norm

$$||w||_\alpha = |w|_\alpha + ||w||,$$

where $||w||$ is the uniform norm of w.

When $\alpha = 1$ then we are considering the space of Lipschitz functions $\mathcal{H}_1$.

For each value $\beta = 1/T$, where T is temperature, we will be interested in the Gibbs state $\mu_{\beta A}$ which will be defined later. It represents the "Statistical

Mechanics" equilibrium probability for the interaction A and it is defined in the Borel sigma-algebra of $\mathcal{B}$.

Results about the connection of Thermodynamic Formalism with different problems in Statistical Mechanics for the one-dimensional lattice appears in [20] and [19].

As we said before we will consider an a priory probability ν on M which will be denoted by simplification as da. If M is a compact differentiable manifold, then, da can be taken as the volume form (Lebesgue measure).

By a modification of the metric a potential A which is is Holder can be considered Lipchitz. We would not bother about this distinction.

We say that a potential $A : \mathcal{B} \to \mathbb{R}$ depends just on k coordinates if

$$A(x) = A(x_0, x_1, x_2, ..., x_n, ..) = A(x_0, x_1, x_2, ..., x_{k-1}).$$

In the case A depends just on two coordinates and $M = \{1, 2\}$, that is,

$$A(x) = A(x_0, x_1),$$

assume that φ depends just in one coordinate. Then, in order to solve

$$\mathcal{L}_A(\varphi)(x) = \sum_{a=1}^{d} e^{A(ax)} \varphi(ax) = \lambda \; \varphi(x)$$

for a positive λ, we can use Peron Theorem for a matrix with all entries positive.

In this way the problem we are interested here is a generalization of Perron Theorem for matrices when the potential A does not depends on a finite number of coordinates.

2 Positive temperature: the Ruelle Theorem

Let $\mathcal{C}$ be the space of continuous functions from $\mathcal{B} = M^{\mathbb{N}}$ to $\mathbb{R}$. We are interested in the Ruelle operator on $\mathcal{C}$ associated to the Lipschitz observable $A : M^{\mathbb{N}} \to \mathbb{R}$, which gets $\psi \in \mathcal{C}$, and sends to $\mathcal{L}_A(\psi) \in \mathcal{C}$ defined by

$$\mathcal{L}_A(\psi)(x) = \int_M e^{A(ax)} \psi(ax) \, d\, a \,,$$

for any $x = (x_0, x_1, x_2, ...) \in \mathcal{B}$, where ax represents the sequence

$$(a, x_0, x_1, x_2, ...) \in \mathcal{B},$$

and $d\,a$ is the a priory probability on M. Note that $\sigma(ax) = x$.

In the case $M = \{1, 2, ..., d\}$, then,

$$\mathcal{L}_A(\psi)(x) = \int_M e^{A(ax)} \psi(ax) \, da = \sum_{a=1}^{d} e^{A(ax)} \psi(ax).$$

The operator $\mathcal{L}_A$ will help us to find the Gibbs state for A. First we will show the existence of a main eigenfunction for $\mathcal{L}_A$, when A is Lipschitz.

More explicitly we want to find ψ and λ such that $\sum_{a=1}^d e^{A(ax)}\psi(ax) = \lambda\psi(x)$ for all x.

Part of our proof follows the reasoning of section 7 in [1] (which considers $M = \{1, 2, .., d\}$) adapted to the present more general case.

The operator $\mathcal{L}_A$ acts on the continuous functions (defined on $\mathcal{B}$ taking values on $\mathbb{R}$) and also is well defined when acting on the Lipchitz functions. It is bounded in both cases. It is not compact.

We can state the eigenfunction problem in another form: find u and $\lambda > 0$ such that

$$\lambda\, e^{u(x)} = \int_M e^{A(ax)} e^{u(ax)}\, d\, a\, .$$

This requires, among other things to find the eigenvalue λ.

In the discounted approach we state another associated problem: for $0 < s < 1$, find u such that

$$e^{u(x)} = \int_M e^{A(ax)} e^{s\, u(ax)}\, d\, a\, .$$

On this kind of problem we do not need to know the eigenvalue. The idea is that, this will help to solve the main problem, that is, in some way when we consider $s \to 1$ we will get the eigenfunction solution.

We begin by defining another operator on $\mathcal{C}$. Let $0 < s < 1$, and define, for $u \in \mathcal{C}$, $\mathcal{T}_{s,A}(u)$ given by

$$\mathcal{T}_{s,A}(u)(x) = \log \left(\int_M e^{A(ax)+su(ax)}\, da \right).$$

In the case $M = \{1, 2, ..., d\}$, then,

$$\mathcal{T}_{s,A}(u)(x) = \log \left(\sum_{a=1}^{d} e^{A(ax)+su(ax)} \right).$$

Our final goal is to show the existence of λ and u such that, for $s = 1$, we get

$$\mathcal{T}_{s,A}(u) = \log \lambda + u.$$

In this case for $\psi = e^u$, we obtain: $\sum_{a=1}^d e^{A(ax)}\psi(ax) = \lambda\psi(x)$, for all x

Proposition 1. *If, $0 < s < 1$, then $\mathcal{T}_{s,A}$ is an uniform contraction map.*

Proof.:

$$|\mathcal{T}_{s,A}(u_1)(x) - \mathcal{T}_{s,A}(u_2)(x)| = \left| \log \left(\frac{\int_M e^{A(ax)+su_1(ax)}}{\int_M e^{A(ax)+su_2(ax)}} \right) \right| =$$

$$= \left| \log \left(\frac{\int_M e^{A(ax)+su_2(ax)+su_1(ax)-su_2(ax)}}{\int_M e^{A(ax)+su_2(ax)}} \right) \right| \leq$$

$$\leq \log \left(\frac{\int_M e^{A(ax)+su_2(ax)+s\|u_1-u_2\|}}{\int_M e^{A(ax)+su_2(ax)}} \right) = s\|u_1 - u_2\| \,.$$

□

Let u_s be the unique fixed point for $\mathcal{T}_{s,A}$. We have

$$\log \left(\int_M e^{A(ax)+su_s(ax)}\, da \right) = u_s(x)\,. \tag{1}$$

Note that as the above operator is a contraction we can get an approximate solution by iterating several times the operator on an initial condition. This is so, but the problem is that when $s \to 1$ the contraction will be come weaker and weaker.

Proposition 2. *The family $\{u_s\}_{0<s<1}$ is an equicontinuous family of functions.*

Proof.: Let $H_s(x,y) = u_s(x) - u_s(y)$. By (1) we have

$$\begin{aligned} e^{u_s(x)} &= \int_M e^{A(ax)+su_s(ax)} \\ &= \int_M e^{A(ay)+su_s(ay)} e^{A(ax)-A(ay)+s[u_s(ax)-u_s(ay)]} \\ &\leq e^{u_s(y)} \max_a \{ e^{A(ax)-A(ay)+s[u_s(ax)-u_s(ay)]} \}. \end{aligned}$$

Hence

$$e^{u_s(x)-u_s(y)} \leq \max_a \{ e^{A(ax)-A(ay)+s[u_s(ax)-u_s(ay)]} \},$$

and this implies

$$H_s(x,y) = u_s(x) - u_s(y) \leq \max_a [A(ax) - A(ay) + sH_s(ax,ay)].$$

Proceeding by induction we get

$$H_s(x,y) \leq \max_{\theta \in \mathcal{B}} \sum_{n=0}^{\infty} s^n [A(\theta_n....\theta_0 x) - A(\theta_n...\theta_0 y)] \leq$$

$$\leq Hol_A \max_{\theta \in \mathcal{B}} \sum_{n=0}^{\infty} s^n d((\theta_n....\theta_0 x), (\theta_n...\theta_0 y))^\alpha \leq$$

$$\leq Hol_A \sum_{n=0}^{\infty} \left(\frac{s}{2^\alpha} \right)^n d(x,y)^\alpha \leq \frac{2^\alpha}{2^\alpha - 1} Hol_A d(x,y)^\alpha \,.$$

Remark 1: This shows that u_s is Lipschitz, and, moreover, that u_s, $0 \leq s < 1$, is equicontinuous family. Note the very important point: the Lipschitz constant of u_s, is given by $\frac{2^\alpha}{2^\alpha - 1} Hol_A$, and depends only on the Holder constant for A, not depending on s.

□

Let

$$S_n(z) = S_{n,A}(z) = \sum_{k=0}^{n-1} A \circ \sigma^k(z)\,.$$

Note that iterates of the operator $\mathcal{L}_A$ can be written with the use of $S_{n,A}(z)$.

$$\mathcal{L}_A^n(w)(x) = \int_{\mathbf{a}\in M^n} e^{S_{n,A}\,(\mathbf{a}x)} w(\mathbf{a}x)\, d\mathbf{a}.$$

Theorem 3. *There exists a strict positive Lipschitz eigenfunction ψ_A for $\mathcal{L}_A : \mathcal{C} \to \mathcal{C}$ associated to a strictly positive eigenvalue λ_A. The eigenvalue is simple and it is equal to the spectral radius.*

Proof. It follows from the fixed point equation that for any x

$$-||A|| + s \min u_s \leq u_s(x) \leq ||A|| + s \max u_s.$$

Therefore, $-||A|| \leq (1-s)\min u_s \leq (1-s)\max u_s \leq ||A||$, for any s. Consider a subsequence $s_n \to 1$ such that $[\,(1-s_n)\,\max u_{s_n}\,] \to k$.

Note that if $k \neq 0$ (a case which happens in several examples) the $\max u_{s_n}$ goes to ∞ when $n \to \infty$.

The family $\{u_s^* = u_s - \max u_s\}_{0<s<1}$ is equicontinuous and uniformly bounded.

Therefore, by Arzela Ascoli $\{u_{s_n}^*\}_{n\geq 1}$ has an accumulation point in $\mathcal{C}$, which we will call u.

Observe that for any s

$$e^{u_s^*(x)} = e^{u_s(x) - \max u_s} =$$

$$e^{-(1-s)\max u_s + u_s(x) - s\max u_s} =$$

$$e^{-(1-s)\max u_s} \int e^{A(ax) + (s u_s(ax) - s\max u_s)}\, da.$$

Taking limit on n on the sequence s_n we get that u satisfies

$$e^{u(x)} = e^{-k} \int e^{A(ax) + u(ax)}\, da.$$

In this way we get a **positive** Lipschitz eigenfunction $\psi_A = e^u$ for $\mathcal{L}_A$ associated to the eigenvalue $\lambda_A = e^k$. □

Remark 2: To prove that u is Lipschitz, we just use the fact that u is the limit of a sequence of uniformly Lipschitz functions (i.e. Lipschitz functions with same Lipschitz constant). Using that u is a bounded function we have that $\psi_A = e^u$ is also Lipschitz. Note the very important point: the Lipschitz constant of $u = \log(\psi_A)$ is given by $\frac{2^\alpha}{2^\alpha - 1} Hol_A$(see Remark 1 in the end of the proof of Proposition 2).

The property that the eigenvalue is simple and maximal follows from the same reasoning as in page 23 and 24 of [68]. For example, to prove that the eigenvalue is simple we suppose there are two eigenfunctions ψ_1 and ψ_2. Let $t = \min\{\psi_1/\psi_2\}$. Then $\psi_3 = \psi_1 - t\psi_2$ is a non-negative eigenfunction which vanishes at some point $z \in \mathcal{B}$. Therefore

$$0 = \lambda_A^n \psi_3(z) = \int_{\mathbf{a} \in M^n} e^{S_{n,A}\,(\mathbf{a}z)} \psi_3(\mathbf{a}z)\, d\mathbf{a}\,,$$

which implies $\psi_3(\mathbf{a}z) = 0\ \forall\, \mathbf{a} \in M^n$, $\forall n$, which makes $\psi_3 = 0$.

$\square$

We remark that in general it is not easy to find explicit solutions for the eigenvalue and eigenfunction problem of last theorem. However, in [82], it is shown for a special class of potentials A on $\{1,2\}^{\mathbb{N}}$, that the eigenvalue λ satisfies a simple functional equation (anyway, not easy to solve). Once you have the λ then you have the explicit expression for the eigenfunction in this family of examples. Related results to this appears in [57], [18].

Another class of potentials where explicit solutions exist appears is [17].

Note that

$$\int_M \frac{e^{A(ax)} \psi_A(ax)}{\lambda_A \psi_A(x)} da = 1\,,\ \forall x \in \mathcal{B}\,. \tag{2}$$

If a continuous potential B satisfies

$$\int_M e^{B(ax)} da = 1\,,\ \forall x \in \mathcal{B}\,,$$

which means $\mathcal{L}_B(1) = 1$, we say that B is **normalized**.

For example, if $\Omega = \{1,2,..,d\}^{\mathbb{N}}$ and $\nu = \frac{1}{d}\sum_{J=1}^d \delta_j$, where δ_j is the Dirac probability on j, we get that the normalized condition means: for any $x = (x_0, x_1, ..)$

$$\frac{1}{d} \sum_j e^{B(j, x_0, x_1, ..)} = 1.$$

The potential B constant equal zero is normalized.

Let

$$\bar{A} = A + \log \psi_A - \log \psi_A \circ \sigma - \log \lambda_A,$$

where $\sigma : \mathcal{B} \to \mathcal{B}$ is the usual shift map. Equation (2) shows that $\bar{A}$ is normalized. It is also Lipschitz (Holder). In this case the main eigenvalue is 1 and the main eigenfunction is constant equal to 1 (in fact we can prove, using proposition 4, that there is only one strictly positive eigenfunction, the one associated to the maximal eigenvalue).

The expression

$$\bar{A} = A + \log \psi_A - \log \psi_A \circ \sigma - \log \lambda_A,$$

is called the main cohomological equation for A.

We also say that $\bar{A}$ and $A - \log \lambda_A$ are equal up to the coboundary $\log \psi_A - \log \psi_A \circ \sigma$.

Remember that, given $x = (x_0, x_1, x_2, ...) \in \mathcal{B}$ and $a \in M$, we denote by $ax \in \mathcal{B}$ the element $ax = (a, x_0, x_1, x_2, ...)$, i.e., any $y \in \mathcal{B}$ such that $\sigma(y) = x$ is of this form.

We define the Borel sigma-algebra $\mathcal{F}$ over $\mathcal{B}$ as the σ-algebra generated by the *cylinders*. By this we mean the sigma-algebra generated by sets of the form $B_1 \times B_2 \times \ ... \ \times B_n \times M^{\mathbb{N}}$, where $n \in \mathbb{N}$, and $B_j, j \in \{1, 2, .., n\}$, are open sets in M. Similar definitions can be considered for $\mathcal{B}_i$.

We say a probability measure μ over $\mathcal{F}$ is invariant, if for any Borel set B, we have that $\mu(B) = \mu(\sigma^{-1}(B))$. This corresponds to stationary probabilities for the underlying stochastic process X_n, $n \in \mathbb{N}$, with state space M. We denote by $\mathcal{M}_\sigma$ the set of invariant probabilities. Similar definitions can be considered for $\mathcal{B}_i$.

We present below a generalization of results considered in [68].

We define the dual operator $\mathcal{L}_A^*$ on the space of the Borel measures on $\mathcal{B}$ as the operator that sends a measure v to the measure $\mathcal{L}_A^*(v)$ defined by

$$\int_{\mathcal{B}} \psi \, d\mathcal{L}_A^*(v) = \int_{\mathcal{B}} \mathcal{L}_A(\psi) \, dv \, .$$

for any $\psi \in \mathcal{C}$.

Now we want to find an eigen-probability for $\mathcal{L}_A^*$. This will help us to find the Gibbs state for the potential A.

Suppose $\mathcal{B} = \{1, 2, ..., d\}^{\mathbb{N}}$. For a given fixed y we have that there exists exactly d^n solutions x of $T^n(x) = y$. We denoted by y_j^n, $j = 1, 2..., d^n$.

It is easy to see that for given A and n

$$\mathcal{L}_A^n(\psi)(y) = \sum_{j=1}^{d^n} e^{A(y_j^n) + A(\sigma(y_j^n)) + A(\sigma^2(y_j^n) + ... + A(\sigma^{n-1}(y_j^n))} \psi(y_j^n).$$

Now we return to the general case.

Proposition 4. *If the observable $\bar{A}$ is normalized, then there exists an unique fixed point $m = m_{\bar{A}}$ for $\mathcal{L}_{\bar{A}}^*$. Such probability measure m is σ-invariant, and for all Holder continuous function ω we have that, on the uniform convergence topology,*

$$\mathcal{L}_{\bar{A}}^n \omega \to \int_{\mathcal{B}} \omega dm \, .$$

Here $\mathcal{L}_{\bar{A}}^n$ denotes the n-th iterate of the operator $\mathcal{L}_{\bar{A}} : \mathcal{C} \to \mathcal{C}$.

Proof.: We begin by proving that the normalization property implies that the convex and compact set of Borel probability measures on $\mathcal{B}$ is preserved by the operator $\mathcal{L}^*_{\bar{A}}$: in order to see that, note that for μ a Borel probability measure on $\mathcal{B}$, we have

$$\mathcal{L}^*_{\bar{A}}(\mu)(\mathcal{B}) = \int_{\mathcal{B}} 1 \; d\mathcal{L}^*_{\bar{A}}(\mu) = \int_{\mathcal{B}} \mathcal{L}_{\bar{A}}(1) d\mu = \int_{\mathcal{B}} 1 \; d\mu = \mu(\mathcal{B}) = 1$$

where the third equality is precisely the normalization hypothesis.

By Tychonoff-Schauder theorem let m be a fixed point for the operator $\mathcal{L}^*_{\bar{A}}$.

To prove that m is σ-invariant, we begin by observing that

$$\mathcal{L}_{\bar{A}}(\psi \circ \sigma)(x) = \int_M e^{\bar{A}(ax)} \psi \circ \sigma(ax) da = \int_M e^{\bar{A}(ax)} \psi(x) da = \psi(x).$$

Note that the normalization hypothesis is used in the last equality.

Therefore, if $\psi \in \mathcal{C}$, then

$$\int_{\mathcal{B}} \psi \circ \sigma dm = \int_{\mathcal{B}} \psi \circ \sigma d\mathcal{L}^*_{\bar{A}}(m) = \int_{\mathcal{B}} \mathcal{L}_{\bar{A}}(\psi \circ \sigma) dm = \int_{\mathcal{B}} \psi dm.$$

which implies the invariance property of m.

Before finishing the proof of proposition 4, we will need two claims. The first is a special estimate which will be important in the rest of this section.

Claim: For any Holder potential A, if $\|w\|$ denotes the uniform norm of the Holder function $w : \mathcal{B} \to \mathbb{R}$, we have

$$|\mathcal{L}^n_A(w)(x) - \mathcal{L}^n_A(w)(y)| \leq \left[C_{e^A} \|w\| \left(\frac{1}{2^\alpha} + \ldots + \frac{1}{2^{n\alpha}} \right) + \frac{C_w}{2^{n\alpha}} \right] d(x,y)^\alpha,$$

where C_{e^A} is the Holder constant of e^A and C_w is the Holder constant of w.

Proof of the Claim: : We prove the claim by induction. Suppose $n = 1$. We have

$$|\mathcal{L}_A(w)(x) - \mathcal{L}_A(w)(y)| \leq$$

$$\leq \int_M |e^{A(ax)} - e^{A(ay)}| \cdot |w(ax)| da + \int_M e^{A(ay)} |w(ax) - w(ay)| da \leq$$

$$\leq (C_{e^A} \|w\| + C_w) \frac{d(x,y)^\alpha}{2^\alpha},$$

where in the last inequality we used the normalization property of A. In particular we can say that the Holder constant of $\mathcal{L}_A(w)$ is given by

$$C_{\mathcal{L}_A(w)} = \frac{C_{e^A} \|w\| + C_w}{2^\alpha}. \tag{3}$$

Now, suppose the Claim holds for n. We have

$$|\mathcal{L}^{n+1}_A(w)(x) - \mathcal{L}^{n+1}_A(w)(y)| = |\mathcal{L}^n_A(\mathcal{L}_A(w))(x) - \mathcal{L}^n_A(\mathcal{L}_A(w))(y)| \leq$$

$$\leq \left[C_{e^A} \|\mathcal{L}_A(w)\| \left(\frac{1}{2^\alpha} + ... + \frac{1}{2^{n\alpha}} \right) + \frac{C_{\mathcal{L}_A(w)}}{2^{n\alpha}} \right] d(x,y)^\alpha ,$$

and, therefore the claim is proved when we use (3) and $\|\mathcal{L}_A(w)\| \leq \|w\|$ which is consequence of the normalization property of A.

As a consequence, the set $\{\mathcal{L}^n_{\bar{A}}\omega\}_{n\geq 0}$ is equicontinuous. In order to prove that $\{\mathcal{L}^n_{\bar{A}}\omega\}_{n\geq 0}$ is uniformly bounded we use again the normalization condition which implies $\|\mathcal{L}^n_{\bar{A}}\omega\| \leq \|w\|, \forall n \geq 1$.

By the Arzela-Ascoli Theorem let $\bar{\omega}$ be an accumulation point for $\{\mathcal{L}^n_{\bar{A}}\omega\}_{n\geq 0}$, i.e., suppose there exists a subsequence $\{n_k\}_{k\geq 0}$ such that

$$\bar{\omega}(x) = \lim_{k\geq 0} \mathcal{L}^{n_k}_{\bar{A}}\omega(x) .$$

Second Claim: : $\bar{\omega}$ is a constant function.

The proof of this second claim is similar to the reasoning of page 25 [68].

Now that $\bar{\omega}$ is a constant function we can prove that

$$\bar{\omega} = \int_{\mathcal{B}} \bar{\omega} dm = \lim_k \int_{\mathcal{B}} \mathcal{L}^{n_k}_{\bar{A}}\omega dm = \lim_k \int_{\mathcal{B}} \omega d(\mathcal{L}^*_{\bar{A}})^{n_k}(m) = \int_{\mathcal{B}} \omega dm,$$

which shows that $\bar{\omega}$ does not depend on the subsequence chosen.

Indeed if two different subsequences produces $\bar{\omega}_1$ and $\bar{\omega}_2$ they both have to be equal to $\int_{\mathcal{B}} \omega dm$.

Therefore, for any $x \in \mathcal{B}$ we have

$$\mathcal{L}^n_{\bar{A}}\omega(x) \to \bar{\omega} = \int_{\mathcal{B}} \omega dm .$$

The last limit shows that the fixed point m is unique.

As x takes values in a compact set we have that $\mathcal{L}^n_{\bar{A}}\omega(x)$ converges uniformly to $\bar{\omega}$. □

Proposition 5. *Let A be a Holder, not necessarily normalized potential, and ψ_A and λ_A the eigenfunction and eigenvalue given by theorem 3. To the potential A we associate the normalized potential $\bar{A} = A + \log\psi_A - \log\psi_A \circ \sigma - \log\lambda_A$. Let m be the unique probability measure that satisfies $\mathcal{L}^*_{\bar{A}}(m) = m$, given by proposition 4.*

(a) the measure

$$\rho_A = \frac{1}{\psi_A}\, m$$

satisfies $\mathcal{L}^*_A(\rho_A) = \lambda_A \rho_A$. Therefore, ρ_A is an eigen-probability for $\mathcal{L}^*_A$.

(b) for any Holder $\phi : \mathcal{B} \to \mathbb{R}$, we have that

$$\frac{\mathcal{L}^n_A(\phi)}{\lambda^n_A} \to \psi_A \int \phi \, d\rho_A .$$

(c)

$$\lim_{n\to\infty}\frac{1}{n}\log\mathcal{L}_A^n(1)=\lambda(A).$$

Proof: (a) $\mathcal{L}^*_{\bar A}(m)=m$ implies that for any $\psi\in\mathcal{C}$, we have

$$\begin{aligned}\int\psi dm &= \int\psi d\mathcal{L}^*_{\bar A}(m)\\ &= \int\mathcal{L}_{\bar A}(\psi)dm\\ &= \int\left(\int\psi(ax)e^{\bar A(ax)}da\right)dm(x)\\ &= \int\left(\int\psi(ax)\frac{e^{A(ax)}\psi_A(ax)}{\lambda_A\psi_A(x)}da\right)dm(x)\,.\end{aligned}$$

Now, if $\varphi\in\mathcal{C}$, making $\psi=\frac{\varphi}{\psi_A}$ in the last equation we have

$$\int\frac{\varphi}{\psi_A}dm=\frac{1}{\lambda_A}\int\left(\int\varphi(ax)\frac{e^{A(ax)}}{\psi_A(x)}da\right)dm(x)\,,$$

which is equivalent to

$$\lambda_A\int\varphi d\rho_A=\int\mathcal{L}_A(\varphi)d\rho_A \tag{4}$$

or

$$\mathcal{L}^*_A(\rho_A)=\lambda_A\rho_A\,.$$

(b) We have that $A=\bar A-\log\psi_A+\log\psi_A\circ\sigma+\log\lambda_A$, and therefore

$$S_{n,A}(z)\equiv\sum_{k=0}^{n-1}A\circ\sigma^k(z)=S_{n,\bar A}(z)-\log\psi_A+\log\psi_A\circ\sigma^n+n\log\lambda_A\,,$$

which makes

$$\frac{\mathcal{L}_A^n(\phi)(x)}{\lambda_A^n}=\frac{1}{\lambda_A^n}\int_{\mathbf{a}\in M^n}e^{S_{n,A}(\mathbf{a}x)}\phi(\mathbf{a}x)d\mathbf{a}=$$

$$=\psi_A(x)\int_{\mathbf{a}\in M^n}\frac{e^{S_{n,\bar A}(\mathbf{a}x)}}{\psi_A(\mathbf{a}x)}\phi(\mathbf{a}x)d\mathbf{a}=$$

$$=\psi_A(x)\mathcal{L}^n_{\bar A}\left(\frac{\phi}{\psi_A}\right)\to\psi_A(x)\int\frac{\phi}{\psi_A}dm_{\bar A}$$

where the convergence on n in the last line comes from Proposition 4.

(c) Note that

$$\frac{\mathcal{L}_A^n(1)}{\lambda_A^n}\to\ \psi_A,$$

and therefore

$$\frac{1}{n}[\log \mathcal{L}_A^n(1) - n\log\lambda_A] = \frac{1}{n}\log\psi_A.$$

As ψ_A is bounded we get

$$\lim_{n\to\infty}\frac{1}{n}\log\mathcal{L}_A^n(1) = \lambda_A.$$

□

Note that the eigenvalue $\lambda_{\bar{A}}$ for $\mathcal{L}_{\bar{A}}$ is equal to 1.

Remark 3: From now on we will call $m_{\bar{A}}$ the eigen-probability for $\mathcal{L}_{\bar{A}}^*$. One can show that the eigen-probability $\rho_A = \frac{1}{\psi_A}\, m_{\bar{A}}$ is the unique eigen-probability for $\mathcal{L}_A^*$. Also, it is not necessarily invariant for the shift σ.

Definition 1. *Let m be the unique probability measure that satisfies $\mathcal{L}_{\bar{A}}^*(m) = m$. We call this probability the Gibbs state for A, and we denoted by either m_A or $m_{\bar{A}}$.*

This probability measure $m_{\bar{A}}$ over $\mathcal{B}$ is invariant for the shift and describes the statistics of the interaction described by A. It is usual to call the probability measure $m_{\bar{A}}$ of Gibbs state (in the Thermodynamic Formalism setting [68]) for the interaction given by A.

We point out that the probability measure ρ_A is positive on open sets of $\mathcal{B}$. Suppose the metric space $M = \mathbb{S}^1$. The projection of this probability measure on the first two coordinates $\mathbb{S}^1 \times \mathbb{S}^1$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{S}^1 \times \mathbb{S}^1$. This is so because, if B is Borel in $[0,1]^2$, then from (4) we have

$$\int I_{(x_0,x_1)\in B}\, d\rho_A = \frac{1}{\lambda_A^2}\int \mathcal{L}_{\bar{A}}^2\,(I_{(x_0,x_1)\in B})\, d\rho_A,$$

and, for any $x \in \mathcal{B}$

$$\mathcal{L}_{\bar{A}}^2(I_{(x_0,x_1)\in B})(x) = \int_M\int_M e^{S_{2,\bar{A}}\,(abx)} I_{(x_0,x_1)\in B}(abx)\, da\, db.$$

The operator $w \to K(w) = w\circ\sigma$ is called the Koopman operator.

Proposition 6. *If we consider the L^2 space of real functions defined on $\mathcal{B}$ with respect to the Gibbs probability $dm_{\bar{A}}$ for A, then, the transpose of $\mathcal{L}_{\bar{A}}$ is K, that is, for any v, w we have*

$$< K(v)\,, w > = < v, \mathcal{L}_{\bar{A}}(w) > .$$

Proof: This follows from

$$< K(v)\,, w > = \int (v\circ\sigma)(x)\, w(x) dm_{\bar{A}}(x) =$$

$$\int (v\circ\sigma)(x)\, w(x)\, [\mathcal{L}_{\bar{A}}^* dm_{\bar{A}}\,]\,(x) = \int \mathcal{L}_{\bar{A}}\, [(v\circ\sigma)(x)\, w(x)\,] dm_{\bar{A}}(x) =$$

$$\int\, v(x)\, \mathcal{L}_{\bar{A}} w(x)\, dm_{\bar{A}}(x) = < v, \mathcal{L}_{\bar{A}}(w) > .$$

□

Suppose $M = \{1, 2, .., d\}$.

We denote a cylinder set a set of the form $\overline{b_0, b_1, .., b_n} = \{x = (b_0, b_1, ..., b_n, x_{n+1}, x_{n+2}, ...)$, where $x_j \in \{1, 2, ..., d\}$, for $j \geq N = 1$.

The sigma algebra $\mathcal{F}$ of Borel on $M^{\mathbb{N}}$ is generated by the collection of all cylinder sets ($n \in \mathbb{N}$)

We point out that if a normalized potential B is such that $\mathcal{L}_B^*(\mu) = \mu$, then μ is positive on cylinders (and so in open sets). This will be the case of the probability m we get from a Holder potential A as above.

Instead of proving this for a general cylinder we will show the property in a particular example in which $M = \{1, 2\}$.

Suppose by contradiction that $\mu(\overline{212}) = 0$.

Then,

$$\int_{\overline{212}} d\mu(x) = \int_{\overline{212}} d\mathcal{L}_B^*(\mu)(x) = \int I_{\overline{212}} d\mathcal{L}_B^*(\mu)(x) = \int \mathcal{L}_B(I_{\overline{212}}) d\mu(x) =$$

$$\int \sum_{j=1,2} e^{B(j,x_0,x_1,x_2,..)} I_{\overline{212}}(j, x_0, x_1, x_2, ..) d\mu(x_0, x_1, x_2, ..) =$$

$$\int e^{B(2,1,2,x_2,..)} d\mu(x_0, x_1, x_2, ..) = \int_{\overline{12}} e^{B(2,1,2,x_2,..)} d\mu(x_0, x_1, x_2, ..).$$

This means

$$\int I_{\overline{212}} d\mu = \int I_{\overline{12}} e^{B(2,1,2,x_2,..)} d\mu = \int I_{\overline{212}}(\phi_2) e^{B(2,1,2,x_2,..)} d\mu.$$

More generally, one can show that

$$\int_{\overline{2,c_0,c_1,...c_k}} d\mu(x_0, x_1, ..) = \int_{\overline{c_0,c_1,...c_k}} e^{B(2,c_0,c_1,c_2,..,c_k,x_{k+1},x_{k_2},...)} d\mu(x_0, x_1, ..).$$

Denote by ϕ_2 the transformation from $\mathcal{B} \to \overline{2}$ which is the inverse of the shift. Note that ϕ_2 is a bijection.

From the above we see that $e^{B(2,x_0,x_1,x_2,...)}$ is the Radon-Nykodin derivative of the change of coordinates associated to ϕ_2.

This means that for any continuous function $f : \mathcal{B} \to \mathbb{R}$ we have

$$\int f d\mu = \int e^B (f \circ \phi_2) d\mu. \tag{5}$$

Now from, the fact that there exists $p > 0$ such that $e^B > p$, we get that if

$$0 = \mu(\overline{212}) = \int_{\overline{212}} d\mu(x) = \int_{\overline{12}} e^{B(2,1,2,x_2,..)} d\mu(x) >$$

$$p \int_{\overline{12}} d\mu(x_0, x_1, x_2, ..) = p\ \mu(\overline{12}),$$

we get a contradiction unless $\mu(\overline{12}) = 0$.

Applying a similar method as above we can show that $\mu(\overline{12}) = 0$ implies that $\mu(\overline{2}) = 0$. Proceeding in the same way, once more, we can show that $\mu(\overline{2}) = 0$ implies that $\mu(\mathcal{B}) = 0$ which is a contradiction.

This shows that μ is positive on cylinders.

If we denote by ϕ_1 the transformation from $\mathcal{B} \to \overline{1}$ which is the inverse of the shift. ϕ_1 is a bijection.

In the same way as above we can show that $e^{B(1,x_0,x_1,x_2,...)}$ is the Radon-Nykodin derivative of the change of coordinates associated to ϕ_1.

We denote $e^B = J$ and call J the Jacobian of the probability m such that $\mathcal{L}_B^*(m) = m$.

Consider $\psi_2 : \mathbb{R} \to \mathbb{R}$ a function that makes a change of coordinates $\psi_2(x) = y$. Given a function $f : \mathbb{R} \to \mathbb{R}$ we get that

$$\int f(y)\, dy = \int \frac{d\psi_2(x)}{dx} (f \circ \psi_2)(x)\, dx.$$

$\frac{d\psi_2(x)}{dx}$ is the Radon-Nikodyn derivative for the case of Lebesgue measure.

Note the similarity of the above equation with expression (5). In this way is quite natural to call J the Jacobian of the invariant probability m.

Later on the text we will call Jacobian $J_m = J^{-1}$ (where J is the Jacobian of this section). It is just a question of notation.

In the case the a priori probability is the counting probability $\nu = \sum_{j=1}^d \delta_j$ and B is normalized on the sense that for any $x = (x_0, x_1, ..)$

$$1 = \sum_{j=1}^{d} e^{B(j,x_0,x_1,...)} = \sum_{j=1}^{d} e^{B(j,x)} = \sum_{j=1}^{d} J(j,x),$$

the entropy of m is

$$-\int \log J d\mu.$$

Now, returning to the general case of an a priori probability ν on a compact space M, we define the entropy by

$$h(\mu) \ = \ h^\nu(\mu) \ = -\int B(x) d\mu(x),$$

where $\mathcal{L}_B^*(\mu) = \mu$.

The entropy measures how random, how complex and how chaotic is the dynamical system (σ, μ). Larger the entropy more chaotic is the system.

This entropy is a real non-positive number. It is zero only for the independent probability $\mu = \nu^{\mathbb{N}}$ (the more chaotic probability).

There is another way (similar) to define entropy of a general probability (not just an equilibrium probability m). This is described on [63].

The entropy of a probability with support on a periodic orbit has entropy equal to $-\infty$.

In another way to get entropy we would like to stress the importance of the Ruelle operator when considering lattices where the fiber M is not finite. When the Bernoulli space is $(S^1)^{\mathbb{N}}$ each point x has an uncountable number of preimages. In this way to define the entropy of an invariant probability via partitions seems a complicated task. One could imagine that the entropy of an invariant probability could be $+\infty$ - for example for the equilibrium probabilities m we get above from a given Holder potential A. This is not the case: entropy can be defined in an analytical way from the Ruelle operator for the potential constant equal 0. Indeed, given an a priori probability ν and the equilibrium state m denote

$$H(m) = \inf_{v>0,\, v \text{ Holder}} \int \log\left(\frac{\mathcal{L}_0 v(s)}{v(s)}\right) dm(s). \tag{6}$$

If $M = \{1, 2, ..., d\}$ and ν is the probability which gives mass $1/d$ for each point in M we get that

$$H(m) = h(m) - \log d \leq 0,$$

where $h(m) \geq 0$ is the Kolmogorov entropy of m.

If $M = \{1, 2, ..., d\}$ and we define the Ruelle operator - using the same formalism as above - but taking the counting measure (not a probability) as the a priori ν then we get that the value of expression (6) is non-negative and equal to the Kolmogorov entropy $h(m)$.

In this way the concept of Kolmogorov entropy can be establish via (6). Note, however, that when M is not finite we need an a priori probability (not a measure) in order to proceed in our formalism. In this way - thinking on extensions of the formalism- the entropy should be more naturally considered as a negative number.

Entropy can be derived from the Ruelle operator in a pure analytical way.

One of the main interest on the probability m (obatined from the Holder potential A) is due to the fact that maximizes

$$P(A) = \sup_{\mu \in \mathcal{M}_\sigma} \{h(\mu) + \int A(x) d\mu(x)\} = h(m) + \int A \, dm$$

where $\mathcal{M}_\sigma$ is the set of σ-invariant probabilities and $h(\mu)$ the entropy of μ.

We say that m is the equilibrium state for A. In this case the Gibbs state is the equilibrium state.

$P(A)$ is called the pressure of A and $P(A) = \log\lambda$, where λ is the main eigenvalue of $\mathcal{L}_A$.

We can get $P(A)$ without talking on entropy via a Minimax result (see [63]) and just using the Ruelle operator: given a Holder potential A

$$P(A) = \sup_{\mu\in\mathcal{M}_\sigma} \Big[\inf_{u\in\mathcal{C}^+} \Big\{ \int \log\left(\frac{\mathcal{L}_A u(x)}{u(x)}\right) d\mu(x) \Big\} \Big],$$

where $\mathcal{C}^+$ is the set of continuous positive functions.

Note that entropy is a concept for dynamically invariant probabilities. In our setting the dynamics is given by the shift $\sigma : M^{\mathbb{N}} \to M^{\mathbb{N}}$. We are defining probabilities for a special kind of dynamics (in which each point has more than one preimage). This is required in order to define the Ruelle operator.

Remark 4: Consider the symbolic space

$$\mathcal{B}_i = \{(..., x_{-2}, x_{-1} \,|\, x_0, x_1, x_2, ...) |\, x_i \in M, i \in \mathbb{Z}\},$$

and the shift $\hat{\sigma}$, such that,

$$\hat{\sigma}\,(..., x_{-2}, x_{-1} \,|\, x_0, x_1, x_2, ...) = (..., x_{-2}, x_{-1}, x_0 \,|\, x_1, x_2, ...).$$

Note that $\hat{\sigma}$ is a bijection.

When we consider a dynamical system like the shift $\hat{\sigma} : M^{\mathbb{Z}} \to M^{\mathbb{Z}}$ (which is a bijection), in order to define entropy $h(\mu)$ for a $\hat{\sigma}$ invariant probability μ, a different approach have to be used; for instance, via Brin-Katok formula to be described later on the text (see section 4).

A variational principle of the form: given Holder potential $B : \mathcal{B}_i = M^{\mathbb{Z}} \to \mathbb{R}$, take the supremum of

$$\int B d\mu + h(\mu) \tag{7}$$

among all probabilities μ which are $\hat{\sigma}$-invariant can be considered. We refer the reader to [12] for general results on this setting.

Given the Holder potential $B : \mathcal{B}_i = M^{\mathbb{Z}} \to \mathbb{R}$, we first derive (as in Proposition 1.2 [68] or, in [12]) the associated cohomologous Holder potential $A : \mathcal{B} \to \mathbb{R}$ (the Holder class can change). Now, we get a dynamical system where each point has more than one preimage.

Then, we can use the Ruelle operator approach. We proceed as above to get ρ_A over $\mathcal{B}$. Finally, we consider the natural extension $\hat{\rho}_A$ of ρ_A in $\mathcal{B}_i$ (see [69] [12]). In this way we solve the Statistical Mechanics problem for the interaction described by B in the lattice $\mathbb{Z}$: its the probability measure $\hat{\rho}_A$.

More precisely, we were able to find the solution $\hat{\rho}_A$ of the variational problem described by (7).

Note that if C is a set that depends just on the coordinates x_0, x_1, then $\rho_{\beta A}(C) = \hat{\rho}_{\beta A}(C)$. For sets $C \subset \mathcal{B}_i$, of this form, we can use indistinctly $\rho_{\beta A}(C)$ or $\hat{\rho}_{\beta A}(C)$.

Now we return to the setting of the Ruelle operator for a Holder potential $A : M^{\mathbb{N}} \to \mathbb{R}$.

A transformation $\theta : E \to F$, where E, F are Banach spaces, is called analytic if θ has derivatives of all orders and moreover,

$$\theta(x+v) = \theta(x) + \sum_{j=1}^{\infty} \frac{1}{j!} D^j \theta(x) \, (\underbrace{v, v, ..., v}_{j}).$$

In $\mathcal{H}_\alpha$ we consider the γ-Holder norm.

Consider the transformation $\theta = \theta_\varphi : \mathcal{H}_\alpha \to \mathcal{H}_\alpha$, given by $\theta(\phi) = \psi$, where, $\psi(x) = \mathcal{L}_\phi(\varphi)(x)$. Then, θ is analytic (see lemma 1.1 [74], [64], [75]).

More precisely, consider a fixed φ, then

$$\theta(\phi + \tilde{\phi})\,(\varphi) - \theta(\phi)(\varphi) = \sum_{j=1}^{\infty} \frac{1}{j!} \theta(\phi)\,(\varphi\, \underbrace{\tilde{\phi}\,\tilde{\phi}...\tilde{\phi}}_{j}) =$$

$$\sum_{j=1}^{\infty} \frac{1}{j!} \mathcal{L}_\phi(\varphi\, \tilde{\phi}^j\,) = \sum_{j=1}^{\infty} \frac{1}{j!} D^j \theta(\phi)\,(\underbrace{\tilde{\phi}, \tilde{\phi}, ..., \tilde{\phi}}_{j})\,.$$

There is a natural way to choose the eigenfunction ψ_A of the Ruelle operator $\mathcal{L}_A$ in such way that

$$A \to \psi_A$$

is analytic.

There is a natural way to choose the eigenprobability ν_A of the dual Ruelle operator $\mathcal{L}_A^*$ in such way that

$$A \to \nu_A \in \mathcal{H}_\alpha{}'$$

is analytic. Above we consider the strong norm of operators and $\mathcal{H}_\alpha{}'$ the space of continuous linear functionals with domain $\mathcal{H}_\alpha$.

Using classical results on perturbation of operators (see [68], [49], [64] and [75]) we have that $A \to \lambda_A$ is analytic on A.

In the case $\Omega = S^1$ and $T : S^1 \to S^1$ an expanding transformation a similar result is true.

In [21] it is considered the differential calculus on the space of equilibrium probabilities.

It follows from the above that if we fix a certain Holder potential A and consider a family βA of potentials, where $\beta \in \mathbb{R}$, then

$$\beta \to \psi_{\beta\, A},$$

$$\beta \to \nu_{\beta\, A},$$

$$\beta \to \lambda_{\beta\, A}$$

are analytic on β.

Proposition 7. *Consider the case $\mathcal{B} = \{1, 2, ..., d\}^{\mathbb{N}}$.*
Suppose $A = \bar{A}$ is normalized and n, y is fixed, and, denote $\mu^n_{A,y}$ by

$$\mu^n_{A,y} = \sum_{j=1}^{d^n} e^{\bar{A}(y^n_j)+\bar{A}(\sigma(y^n_j))+\bar{A}(\sigma^2(y^n_j)+...+\bar{A}(\sigma^{n-1}(y^n_j))}\delta_{y^n_j}.$$

Then, independently of y, we have, as $n \to \infty$

$$\mu^n_{A,y} \to m_{\bar{A}}.$$

Proof: It is easy to see that for the given normalized $\bar{A}$ and n we have that $\mu^n_{A,y}$ is a probability. It is not invariant.

We also know that for any ϕ Holder

$$\mathcal{L}^n_A(\phi)(y) = \sum_{j=1}^{d^n} e^{A(y^n_j)+A(\sigma(y^n_j))+A(\sigma^2(y^n_j)+...+A(\sigma^{n-1}(y^n_j))}\phi(y^n_j) \to \int \phi dm_A.$$

The result follows immediately. If ϕ is just continuous, we use the fact that the Holder functions are dense in the set continuous functions with the C^0 topology to get the final result.

□

Proposition 8. *The only Lipschitz continuous eigenfunction ψ of $\mathcal{L}_A$ which is totally positive is ψ_A (the one associated to the maximal eigenvalue λ_A).*

Proof: Suppose $\psi : \mathcal{B} \to \mathbb{R}$ is a Lipschitz continuous eigenfunction of $\mathcal{L}_A$ associated to the eigenvalue β.

It follows from the above that $\frac{\mathcal{L}^n_A(\psi)}{\lambda^n_A} \to \psi_A \int \psi d\rho_A$, when $n \to \infty$.

Therefore, if $\psi > c > 0$, then $\int \psi d\rho_A > 0$. Moreover, $\mathcal{L}^n_A(\psi) = \beta^n \psi$. This is only possible if $\beta = \lambda_A$ and $\psi = \psi_A$.

□

It is easy to see that if A is Holder with exponent α, and, denoting $\mathcal{H}_\alpha$, the set of real valued functions with Holder exponent α, then $\mathcal{L}_{\bar{A}} : \mathcal{H}_\alpha \to \mathcal{H}_\alpha$.

The α norm of Holder functions induce a norm $||H||_\alpha$ on the set of bounded operators H acting on $\mathcal{H}_\alpha$ (see [27]).

The spectral radius of a bounded operator $H : \mathcal{H}_\alpha \to \mathcal{H}_\alpha$ is

$$\inf \{||H^n||_\alpha^{1/n} \, , \, n \geq 0\}.$$

All elements of the spectra of H have norm less than the spectral radius value of H (see [27]).

We note that $\mathcal{K}_\alpha \equiv \{w \in \mathcal{H}_\alpha\,,\, ||w||_\alpha \leq 1\}$ is compact in the uniform norm as a subset of $\mathcal{C}$. To prove that, we just need to observe that the definition of the norm $||w||_\alpha$ implies that $\mathcal{K}_\alpha$ is a equicontinuous and uniformly bounded set, and then we have the result directly by using Arzela-Ascolis theorem.

We can also prove that $\mathcal{K}_\alpha^A \equiv \{w \in \mathcal{H}_\alpha\,,\, \int_w dm_A = 0\,,\, ||w||_\alpha \leq 1\}$ is compact in the uniform norm. For doing that, let $I_{m_A} : \mathcal{H}_\alpha \to \mathbb{R}$ be given by $I_{m_A}(w) = \int w dm_A$. We have that I_{m_A} is a bounded linear operator, and therefore $I_{m_A}^{-1}\{0\}$ is a closed subset of $\mathcal{H}_\alpha$. Now $\mathcal{K}_\alpha^A = \mathcal{K}_\alpha \cap I_{m_A}^{-1}\{0\}$ is compact.

Proposition 9. *Suppose $\bar{A}$ is normalized, then the eigenvalue $\lambda_{\bar{A}} = 1$ is maximal. Moreover, the remainder of the spectrum of $\mathcal{L}_{\bar{A}} : \mathcal{H}_\alpha \to \mathcal{H}_\alpha$ is contained in a disk centered at zero with radius strictly smaller than one.*

Proof. Remember that 1 is the eigenfunction associated to the eigenvalue 1. We will show that $\mathcal{L}_{\bar{A}}$ restricted to $\mathcal{K}_\alpha^{\bar{A}}$ has spectral radius strictly smaller than 1. We know from proposition 4 that $\mathcal{L}_{\bar{A}}^k$ converges to zero in the compact set $\mathcal{K}_\alpha^{\bar{A}}$.

The normalization hypothesis implies $||\mathcal{L}_{\bar{A}}^{n+1}(w)|| \leq ||\mathcal{L}_{\bar{A}}^n(w)|| \,\forall n \geq 0$. We will now prove that this monotonicity property implies that the convergence above is uniform. More precisely, we have

Claim: Given a small ϵ there exists $N = N_\epsilon \in \mathbb{N}$ such that

$$||\mathcal{L}_{\bar{A}}^n(w)|| < \epsilon \;\forall\, n \geq N\,,\, \forall w \in \mathcal{K}_\alpha^{\bar{A}}\,.$$

To prove this claim, let $C_n \equiv \{w \in \mathcal{K}_\alpha^{\bar{A}} \,:\, ||\mathcal{L}_{\bar{A}}^m(w)|| < \epsilon, \,\forall\, m \geq n\}$. The monotonicity property implies $C_n \subseteq C_{n+1}$ and also that C_n is an open set in the uniform norm, while $\mathcal{L}_{\bar{A}}^k(w) \to 0$ implies $\cup_n C_n = \mathcal{K}_\alpha^{\bar{A}}$. Therefore, compactness of $\mathcal{K}_\alpha^{\bar{A}}$ implies $\mathcal{K}_\alpha^{\bar{A}} = C_N$ for some $N \in \mathbb{N}$.

The last claim is easy to prove and can be enunciated as:

Claim: There exists $C > 0$ such that

$\forall n \in \mathbb{N}$ and $w \in \mathcal{H}_\alpha$

$$|\mathcal{L}_{\bar{A}}^n(w)|_\alpha \leq C||w|| + \frac{|w|_\alpha}{(2^\alpha)^n}.$$

Now, for any given n and k, using the last Claim we have for $w \in \mathcal{H}_\alpha$

$$|\mathcal{L}_{\bar{A}}^{n+k}(w)|_\alpha \leq C||\mathcal{L}_{\bar{A}}^k(w)|| + \frac{|\mathcal{L}_{\bar{A}}^k(w)|_\alpha}{(2^\alpha)^n} \leq C||\mathcal{L}_{\bar{A}}^k(w)|| + C\,\frac{||w||}{(2^\alpha)^n} + \frac{|w|_\alpha}{(2^\alpha)^{n+k}}.$$

Therefore, for fixed k, if ϵ is small enough and $n \geq N_\epsilon$, we have that for all $w \in \mathcal{K}_\alpha^{\bar{A}}$

$$||\mathcal{L}_{\bar{A}}^{n+k}(w)||_\alpha \leq \epsilon < 1.$$

In this case the spectral radius is smaller than $\epsilon^{\frac{1}{n+k}}$.

□

We denote $\lambda_{\bar{A}}^1 < \lambda_{\bar{A}} = 1$ the spectral radius of $\mathcal{L}_{\bar{A}}$ when restricted to the set $V = \{w \in \mathcal{H}_\alpha : \int w\, dm_{\bar{A}} = 0\}$.

Now we will show the exponential decay of correlation for Holder functions.

Proposition 10. *If $v, w \in \mathcal{L}^2(m_{\bar{A}})$ are such that w is Holder and $\int w\, dm_{\bar{A}} = 0$, then, there exists $C > 0$ such that for all n*

$$\int (v \circ \sigma^n)\, w\, dm_{\bar{A}} \leq C\, (\lambda_{\bar{A}}^1)^n$$

Proof. First note that

$$\int (v \circ \sigma)\, w\, dm_{\bar{A}} = \int v\, \mathcal{L}_{\bar{A}}(w)\, dm_{\bar{A}}.$$

Moreover,

$$\int (v \circ \sigma^n)\, w\, dm_{\bar{A}} = \int v\, \mathcal{L}_{\bar{A}}^n w\, dm_{\bar{A}}.$$

Suppose w is such that $\int w dm_{\bar{A}} = 0$.

In order to estimate the decay of correlation $\int (v \circ \sigma)\, w\, dm_{\bar{A}} \to 0$, when $n \to \infty$, we can use the the uniform convergence of $\mathcal{L}_{\bar{A}}^n(x) \to 0$.

Indeed, as $\mathcal{L}_{\bar{A}}(1) = 1$ we get that 1 is the main eigenfunction. Consider the space $V = \{w \mid < w, 1 >= 0\} \subset \mathcal{L}^2(m_{\bar{A}})$.

As, for $w \in V$ we have

$$< \mathcal{L}_{\bar{A}}(w), 1 >=< w, K(1) >=< w, 1 >= 0,$$

we have that $\mathcal{L}_{\bar{A}}(V) \subset V$.

In this way for a Holder $w \in V$ there exist a $C > 0$ such that for any n

$$|\mathcal{L}_{\bar{A}}^n(w)(x)| \leq C\, (\lambda_{\bar{A}}^1)^n.$$

Therefore, given a bounded v and $w \in V$, we have for some $c > 0$ that

$$|\int (v \circ \sigma^n)\, w\, dm_{\bar{A}}| \leq \int |v\, \mathcal{L}_{\bar{A}}^n w|\, dm_{\bar{A}} \leq c\, (\lambda_{\bar{A}}^1)^n.$$

If $\int w dm_{\bar{A}} \neq 0$, we can consider $w - \int w dm_{\bar{A}}$ and proceed in the same way as before.

□

The above proposition implies that $m_{\bar{A}}$ is mixing (same reasoning as in section 2 of [50] which considers the case of the shift on $\{1, 2, .., d\}^{\mathbb{N}}$).

A probability μ is ergodic if for any Borel set A such that $\sigma^{-1}(A) = A$ we have that $\mu(A) = 0$ or $\mu(A) = 1$.

Proposition 11. *The invariant probability measure $m_{\bar{A}}$ is ergodic.*

Proof. If a dynamical system is mixing then it is ergodic (see section 2 in [50]).

□

Given a probability m over $\mathcal{B}$, a sub-sigma algebra $\mathcal{F} \subset \mathcal{B}$ and measurable function $f : \Omega \to \mathbb{R}$ the expected value of f is the g

$$E_m(f|\mathcal{F}) = g$$

such that, g is $\mathcal{F}$-measurable and for all $A \in \mathcal{F}$

$$\int_A g\, dm = \int_A f dm.$$

This is equivalent (see [9] or [35]) to say that g is such that for any $\mathcal{F}$-measurable function ϕ is valid

$$\int \phi\, g\, dm = \int \phi\, f dm.$$

Let $\mathcal{B}$ denote the Borel sigma-algebra on $\Omega = \{1, 2, ..., d\}^{\mathbb{N}}$ and $\mathcal{X}_n = \sigma^{-n}(\mathcal{B})$.

In this case h is $\mathcal{X}_n$-measurable, if and only if, is of the form $h = v \circ \sigma^n$ for some $\mathcal{B}$-measurable function v (see [9] or [35]).

Lemma 12. *For any continuous function $f : \Omega \to \mathbb{R}$ we have that*

$$E_m(f \,|\, \mathcal{X}_n)(x) = \mathcal{L}^n_{\log J}(f)(\sigma^n(x)).$$

Proof:

Note that for any n we have $\mathcal{L}^n_{\log J}(f\,(\,\phi \circ \sigma^n)) = \phi\, \mathcal{L}^n_{\log J}(f)$

Consider a continuous function $\phi : \Omega \to \mathbb{R}$, then

$$\int (\phi \circ \sigma^n(x))\, f(x) dm(x) = \int \mathcal{L}^n_{\log J}\,[\,(\phi \circ \sigma^n(x))\, f(x)\,] dm(x) =$$

$$\int \phi(x)\, \mathcal{L}^n_{\log J}\,(f)(x)\, dm(x) = \int \phi(\sigma^n(x))\, \mathcal{L}^n_{\log J}\,(f)(\sigma^n(x))\, dm(x),$$

where in the last equality we use the fact that m is invariant for σ.

Note that the Gibbs state formalism via boundary conditions, as in [42], does not require, in principle, to talk about entropy.

In Statistical Mechanics, for a fixed interaction A under a certain temperature $T > 0$, up to a multiplicative constant, the natural potential to be considered is $\frac{1}{T} A$. We denote $\beta = \frac{1}{T}$, and, using the results above we can consider the corresponding eigenfunction $\psi_{\beta A}$, eigenvalue $\lambda_{\beta A} = \lambda_\beta$, and the Gibbs state which now will be denoted $\mu_{\beta A}$.

What happen with these two objects when $T \to 0$ (or, $\beta \to \infty$), is the purpose of the next section.

3 Zero temperature: calibrated subactions, maximizing probabilities and selection of probabilities

In this section and also in the next two sections we will consider, among other issues, questions involving selections of probabilities when the temperature goes to zero, maximizing probabilities for a given potential and existence of calibrated subactions. Among other results we will show that, under some conditions, the sequence $\{\mu_{\beta A}\}$ of Gibbs states for the potential βA converges to a measure μ_∞ which has the property of maximizing the integral $\int A d\mu$ among all invariant measures μ for the shift map. Sometimes such convergence will not occur (this is what we call non selection of probabilities).

We will also consider calibrated subactions, which is an important tool that allows one to identify the support of the maximizing measure μ_∞ (see equation (9) below), and can be used to relate the maximal eigenvalues of the Ruelle operator to the value $m(A) = \int A d\mu_\infty$ (see theorem 14). Existence of calibrated subactions are also related to the existence of large deviation principles for the convergence of $\{\mu_{\beta A}\}$ to μ_∞.

Some of the problems discussed here are usually called ergodic optimization problems (see [5][47]).

Consider a fixed Holder potential A and a real variable $\beta > 0$. We denote by $\psi_{\beta A}$ the eigenfunction for the Ruelle operator associated to βA.

Remark 5: Given β and A, the Lipschitz constant of u_β, such that $\psi_{\beta A} = e^{u_\beta}$, depends on the Holder constant for $\beta\, A$ (see Remarks 1 and 2). More precisely, the Lipschitz constant of $u_\beta = \log(\psi_{\beta A})$ is given by $\beta \frac{2^\alpha}{2^\alpha - 1} Hol_A$. Therefore, $\frac{1}{\beta}\log(\psi_{\beta A})$, $\beta > 0$, is equicontinuous. Note that it is also uniformly bounded from the reasons described below.

A possible renormalization condition for $\psi_{\beta A}$ [24] is $\int \psi_{\beta A}\, d\rho_{\beta A} = 1$, where $\rho_{\beta A}$ is the eigen-probability for $\mathcal{L}^*_{\beta A}$ (see proposition 5 and remark 3). For each $\beta > 0$ the normalization hypothesis $\int \psi_{\beta A}\, d\rho_{\beta A} = 1$ implies the existence of $x_\beta \in \mathcal{B}$ such that $\psi_\beta(x_\beta) = 1$. Here we are using the connectedness hypothesis of $\mathcal{B}$. When $\beta \to \infty$ we have that $x_{\beta_k} \to \bar{x}$, for a subsequence. Note that when we normalize $\psi_{\beta A}$ the Holder constant of $\log(\psi_{\beta A})$ remains unchanged, which assures the uniformly continuous property of the family $1/\beta \log(\psi_{\beta A})\,, \beta > 0$. Moreover, the normalization hypothesis and Remark 5 implies that $1/\beta \log(\psi_{\beta A})\,, \beta > 0$ is uniformly bounded.

Therefore, there exists a subsequence $\beta_n \to \infty$, and V Lipschitz, such that on the uniform convergence

$$V := \lim_{n\to\infty} \frac{1}{\beta_n} \log(\psi_{\beta_n A}).$$

Consider point $p_0 \in \mathcal{B}$. Another possible normalization for the eigenfunction $\psi_{\beta A}$ is to assume that $\psi_{\beta A}(p_0) = 1$. We will prefer this late form.

By selection of a function V, when the temperature goes to zero (or, $\beta \to \infty$), we mean the existence of the limit (in the uniform norm)

$$V := \lim_{\beta\to\infty} \frac{1}{\beta} \log(\psi_{\beta A}).$$

The existence of the limit when $\beta \to \infty$ (not just of a subsequence), in the general case, is not an easy question.

In this section we denote $\mu_{\beta A}$ the Gibbs state for the potential βA, i.e. the eigen-probability of $\mathcal{L}^*_{\bar{A}}$, where $\bar{A} = A + \log \psi_A - \log \psi_A \circ \sigma - \log \lambda_A$.

By selection of a measure $\tilde{\mu}_\infty$, when the temperature goes to zero (or, $\beta \to \infty$), we mean the existence of the limit (in the weak* sense)

$$\tilde{\mu}_\infty := \lim_{\beta \to \infty} \mu_{\beta A}.$$

In some sense V is what one can get in the limit, in the log-scale, from the eigenfunction (at non-zero temperature), and $\tilde{\mu}_\infty$ is the Gibbs state at temperature zero.

Nice results on selection for the case the potential depends on a finite number of coordinates are [53] [14] and [10].

Other cases of selection of probability for interesting cases of potentials is [4].

Results when the space of symbols is countable and not compact are presented for instance in [11].

Even if A is Lipschitz not always the above limit on $\mu_{\beta A}$, $\beta \to \infty$, exist. In fact there is an example due to A. C. D. van Enter and W. M. Ruszel where there is no limit for $\mu_{\beta A}$, as $\beta \to \infty$ for the XY model.

When $M = \{1, 2, .., d\}$ an example of non-selection of probabilities is presented in [16] [15].

Some theorems in this section are generalizations of corresponding ones in [62] (which consider only potentials A which depend on two coordinates). Related results appear in [40] and [41]. Results about selection (or, non selection) in the setting of Thermodynamic Formalism appear in [7] [4] [15] [53] [14] [60].

Some of the proofs and results presented in the present section are similar to other ones in Ergodic Optimization [47] and Thermodynamics Formalism, but, the main point is that we have to avoid in the proofs the concept of entropy and the variational principle of pressure.

Remember that we denote by $\mathcal{M}_\sigma$ the set of σ invariant Borel probabilities over $\mathcal{B}$. As $\mathcal{M}_\sigma$ is compact, given A, there always exists a subsequence β_n, such that $\mu_{\beta_n A}$ converges to an invariant probability measure.

We consider the following problem: given $A : \mathcal{B} \to \mathbb{R}$ Lipschitz, we want to find measures that maximize, over $\mathcal{M}_\sigma$, the value

$$\int A(x)\, d\mu(\mathbf{x}).$$

We define

$$m(A) = \max_{\mu \in \mathcal{M}_\sigma} \left\{ \int A d\mu \right\}.$$

Any of these measures will be called a maximizing probability, which is sometimes denoted by μ_∞. As $\mathcal{M}_\sigma$ is compact, there exist always at least one maximizing probability. It is also true that there exists ergodic maximizing probabilities. Indeed, the set of maximizing probabilities is convex, compact and the extreme probabilities of this convex set are ergodic (can not be express as convex combination of others [50]). Any maximizing probability is a convex combination of ergodic ones [69].

Even when A is Holder the maximizing probability μ_∞ do not have to be unique. For instance, suppose that A is Holder and has maximum value just in the union of two different fixed points (for the shift σ) $p_0 \in \mathcal{B}$ and $p_1 \in \mathcal{B}$. In this case the set of maximizing probabilities μ_∞ is $\{t\,\delta_{p_0} + (1-t)\delta_{p_1} \,|\, t \in [0,1]\}$.

Note that δ_{p_0} and δ_{p_1} are ergodic, but the other maximizing probabilities are not.

Recently G. Contreras announced that generically for potentials on the Holder class the maximizing probability has support on a unique periodic orbit [25].

Similar definitions for a potential $A : \mathcal{B}_i \to \mathbb{R}$ and maximization of $\int A\, d\hat{\mu}$, over all the $\hat{\mu}$ which are $\hat{\sigma}$-invariant probabilities, can also be considered. Questions about selection of measure also make sense.

Definition 2. *A continuous function $u : \mathcal{B} \to \mathbb{R}$ is called a* calibrated subaction *for $A : \mathcal{B} \to \mathbb{R}$, if, for any $y \in \mathcal{B}$, we have*

$$u(y) = \max_{\sigma(x)=y} [A(x) + u(x) - m(A)]. \tag{8}$$

This can also be expressed as

$$m(A) = \max_{a \in M} \{A(ay) + u(ay) - u(y)\}.$$

Note that for any $x \in \mathcal{B}$ we have

$$u(\sigma(x)) - u(x) - A(x) + m(A) \geq 0.$$

The above equation for u can be seen as a kind of discrete version of a subsolution of the Hamilton-Jacobi equation [26] [8] [34]. It can be also seen as a kind of dynamic additive eigenvalue problem [22] [23] [39].

If u is a calibrated subaction, then $u + c$, where c is a constant, is also a calibrated subaction. An interesting question is when such calibrated subaction u is unique up to an additive constant.

Remember that if ν is invariant for σ, then for any continuous function $u : \mathcal{B} \to \mathbb{R}$ we have

$$\int [u(\sigma(x)) - u(x)]\, d\nu = 0$$

Suppose μ is maximizing for A and u a calibrated subaction for A.

It follows at once (see for instance [24] [47] [76] for a similar result) that for any x in the support of μ_∞ we have

$$u(\sigma(x)) - u(x) - A(x) + m(A) = 0. \tag{9}$$

In this way if we know the value $m(A)$, then a calibrated subaction u for A help us to identify the support of maximizing probabilities μ. The above equation can be true outside the union of the supports of the maximizing probabilities μ.

Maximizing probabilities μ_∞ are natural candidates for being selected by $\mu_{\beta A}$, as $\beta \to \infty$. But, in our setting, without the maximizing principle of pressure (as one can take advantage in classical Thermodynamic Formalism) this is not so obvious. We address the question in a future section .

Lets consider for a moment the more simple case $\mathcal{B} = \{1, 2..., d\}^{\mathbb{N}}$.

For any given β we have for any y the expression

$$\sum_i e^{\beta A + \log \psi_{\beta A} - \log \psi_{\beta A} \circ \sigma - \log \lambda_{\beta A}} \, (i\, y) = 1.$$

Note that for any y and i

$$\beta A + \log \psi_{\beta A} - \log \psi_{\beta A} \circ \sigma - \log \lambda_{\beta A} \, (i\, y) < 0$$

If V is such that $V := \lim_{\beta_n \to \infty} \frac{1}{\beta_n} \log(\psi_{\beta_n A})$, as the convergence is uniform we get that for any y

$$\sum_i e^{\beta_n A + \beta_n V - \beta_n V \circ \sigma - \log \lambda_{\beta_n A}} \, (i\, y) \, \sim 1.$$

Suppose we know that for a subsequence of the above subsequence

$$\lim_{\beta_n \to \infty} \frac{1}{\beta_n} \log \lambda_{\beta_n \, A} = c.$$

The next proposition claims that this will happen for a subsequence.

Then,

$$\sum_i e^{\beta_n (A + V - V \circ \sigma - c)} \, (i\, y) \, \sim 1.$$

It is not possible that for y fixed, all the $(A + V - V \circ \sigma - c)(i\, y)$, $i = 1, 2, ..., d$ to be strictly negative. Otherwise, the sum would go to zero.

This shows that for all y

$$c = \max_i \{A(iy) + V(iy) - V(y)\}.$$

If we are able to show that $c = m(A)$, then, indeed, any limit of subsequence $V = \lim_{\beta_n \to \infty} \frac{1}{\beta_n} \log(\psi_{\beta_n A})$ is a subaction. The equality $c = m(A)$ will be show to be true in the next theorem.

All of the above can be formalized with ϵ and δ, etc...

Now, we return to the general problem.

Proposition 13. *For any β, we have $-\|A\| < \frac{1}{\beta} \log \lambda_\beta < \|A\|$.*

Proof: Fix $\beta > 0$. We choose $\bar{x}$ the maximum of $\psi_{\beta A}$ in $\mathcal{B}$ and $\tilde{x}$ the minimum of $\psi_{\beta A}$ in $\mathcal{B}$. Now, if $\|A\|$ is the uniform norm of A, we have

$$\lambda_\beta = \frac{1}{\psi_{\beta A}(\bar{x})} \int e^{\beta A(a\,\bar{x})} \psi_{\beta A}(a\,\bar{x}) da \leq \int e^{\beta A(a\,\bar{x})} da \leq e^{\beta \|A\|} \text{ and}$$

$$\lambda_\beta = \frac{1}{\psi_{\beta A}(\tilde{x})} \int e^{\beta A(a\,\tilde{x})} \psi_{\beta A}(a\,\tilde{x}) da \geq \int e^{\beta A(a\,\bar{x})} da \geq e^{-\beta \|A\|} \ ,$$

which proves the result. □

Next theorem is inspired by Theorem 1 in [1] and Theorem 3.3 in [43]. It follows from the last part of its proof that $K = m(A)$.

Theorem 14. *Given A Lipschitz there exists u Lipschitz which is a calibrated subaction for A. As a consequence, we have that*

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log \lambda_\beta = m(A).$$

Proof. Suppose $A : \mathcal{B} \to \mathbb{R}$ is Lipschitz.

Given $0 < \lambda \leq 1$, consider the operator $\hat{\mathcal{L}}_\lambda : \mathcal{C} \to \mathcal{C}$ given by,

$$\hat{\mathcal{L}}_\lambda(u)(x) = \sup_{a \in \mathbb{S}^1} [A(ax) + \lambda u(ax)].$$

Given $x \in \mathcal{B}$, we denote by $a_x \in \mathbb{S}^1$ one of the points a where the supremum is attained.

It is easy to see that for any $0 < \lambda < 1$, the transformation $\hat{\mathcal{L}}_\lambda$ is a contraction on $\mathcal{C}$ with the uniform norm. Indeed, given $x \in \mathcal{B}$

$$\sup_{a \in \mathbb{S}^1} [A(ax) + \lambda u(ax)] - \sup_{b \in \mathbb{S}^1} [A(bx) + \lambda v(bx)] \leq$$

$$[A(a_x\, x) + \lambda u(a_x\, x)] - [A(a_x\, x) + \lambda v(a_x\, x)] \leq$$

$$\lambda u(a_x\, x) - \lambda v(a_x\, x) \leq \lambda ||u - v||.$$

Denote by u_λ the corresponding fixed point in $\mathcal{C}$. We want to show that u_λ is equicontinuous. Consider $x_0, y_0 \in \mathcal{B}$. For the given x_0 we take the corresponding $a_{x_0} \in M$, and then the we get $x_1 = a_{x_0} x_0$. By induction, given x_j, get $x_{j+1} = a_{x_j} x_j$.

We can also can get a sequence $y_j \in \mathcal{B}$, $j \geq 1$, such that, $y_j = a_{x_{j-1}} ... a_{x_1} a_{x_0} y_0$. Note that for all j we have $\sigma^j(y_j) = y_0$.

As for any j we have $u_\lambda(y_j) \geq A(y_{j+1}) - \lambda u_\lambda(y_{j+1})$, then

$$u_\lambda(x_j) - u_\lambda(y_j) \leq$$

$$[A(x_{j+1}) - A(y_{j+1})] + \lambda\, [u_\lambda(x_{j+1}) - u_\lambda(y_{j+1})].$$

Therefore, given x_0, y_0

$$u_\lambda(x_0) - u_\lambda(y_0) \leq \sum_{j=0}^{\infty} \lambda^j \left[A(x_j) - A(y_j)\right] \leq$$

$$(1-\lambda) \sum_{j=0}^{\infty} \lambda^j \sum_{i=0}^{j} \left[A(x_i) - A(y_i)\right] \leq$$

$$\sup_j \sum_{i=0}^{j} \left[A(x_i) - A(y_i)\right] \leq$$

$$||A|| \sup_j \sum_{i=0}^{j} (\frac{1}{2})^j d(x_0, y_0) < ||A|| \, 2 \, d(x_0, y_0).$$

This shows that u_λ is Lipschitz, and, moreover, that u_λ, $0 \leq \lambda < 1$, is equicontinuous family. Note the very important point: the Lipschitz constant of u_λ depends on $||A||$.

Denote $u_\lambda^* = u_\lambda - \max u_\lambda$. Using Arzela-Ascoli we get the existence of a subsequence $\lambda_n \to 1$ such that $u_{\lambda_n}^* \to u$.

We claim that u is a subaction.

Indeed, given $x \in \mathcal{B}$, as $|u_\lambda(x)| \leq \lambda \, |u_\lambda(a_x \, x)| + |A(a_x \, x)| \leq \lambda \, ||u_\lambda|| + ||A(x)||$, then $(1-\lambda)||u_\lambda|| < C$, where C is a constant.

From this follows that there is a constant k, such for some subsequence (of the previous subsequence λ_n), which will be also denoted by λ_n, we have $(1-\lambda_n)||u_{\lambda_n}|| \to k$.

Note that for any λ

$$u_\lambda^*(x) = u_\lambda(x) - \max u_\lambda =$$

$$-(1-\lambda) \max u_\lambda + u_\lambda(x) - \lambda \max u_\lambda =$$

$$-(1-\lambda) \max u_\lambda + \max_{a \in \mathbb{S}^1} \{A(ax) + (\lambda u_\lambda(ax) - \lambda \max u_\lambda)\}.$$

Taking limit on n on the sequence λ_n we get

$$u(x) = -k + \max_{a \in \mathbb{S}^1} \{A(ax) + u(ax)\} = \max_{a \in \mathbb{S}^1} \{A(ax) + u(ax) - k\}.$$

Now, all we have to show is that $k = m(A)$.

a) From the above it follows at once that

$$-u(\sigma(y)) + u(y) + A(y) \leq k.$$

If ν is a σ-invariant probability measure, then,

$$\int A(y) d\nu(y) = \int [u(\sigma(y)) - u(y) + A(y)] \, d\nu(y) \leq k,$$

and, this shows that $m(A) \leq k$.

b) Now we show that $m(A) \geq k$. Note that for any x there exist $y = a_x\, x$ such that $\sigma(y) = x$, and

$$-u(\sigma(y)) + u(y) + A(y) = k.$$

Therefore, the compact set $K = \{y \,|\, -u(\sigma(y)) + u(y) + A(y) = k\}$ is such that, $K' = \cap_n \sigma^{-n}(K)$ is non-empty, compact and σ-invariant. If we consider an σ-invariant probability measure ν with support on K', we have that $\int A(y) d\nu(y) = k$. From this follows that $m(A) \geq k$.

The above also shows that the c we get before is also equal to $m(A)$.

□

Now we state a general result assuming just that A is continuous (not necessarily Lipschitz). We refer the reader to Theorem 1 in [38], Proposition 4 in [62], Theorem 2.4 in [43] for related results.

Theorem 15. *Given a potential $A \in \mathcal{C}$, we have*

$$m(A) = \inf_{f \in \mathcal{C}} \max_{(\mathbf{a},\mathbf{x}) \in \mathbb{S}^1 \times \mathcal{B}} [A(\mathbf{a}\,\mathbf{x}) + f(a\mathbf{x}) - f(\mathbf{x}))].$$

Proof: First, consider the convex correspondence $F : \mathcal{C} \to \mathbb{R}$ defined by $F(g) = \max(A + g)$. Consider also the subset

$$\mathcal{G} = \{g \in \mathcal{C} : \text{there exists } f \text{ such that } g(\mathbf{ax}) = f(a\,\mathbf{x}) - f(\mathbf{x}), \quad f \in \mathcal{C}\} \neq \emptyset.$$

Now consider the concave correspondence $G : \mathcal{C} \to \mathbb{R} \cup \{-\infty\}$ taking $G(g) = 0$, if $g \in \bar{\mathcal{G}}$, and $G(g) = -\infty$ otherwise.

Let $\mathcal{S}$ be the set of the signed measures over the Borel sigma-algebra of $\mathcal{B}$. Remember that the corresponding Fenchel transforms, $F^* : \mathcal{S} \to \mathbb{R} \cup \{+\infty\}$ and $G^* : \mathcal{S} \to \mathbb{R} \cup \{-\infty\}$, are given by

$$F^*(\hat{\mu}) = \sup_{g \in \mathcal{C}} \left[\int g(\mathbf{ax})\; d\hat{\mu}(\mathbf{ax}) - F(g) \right] \text{ , and}$$

$$G^*(\hat{\mu}) = \inf_{g \in \mathcal{C}} \left[\int g(\mathbf{ax})\; d\hat{\mu}(\mathbf{ax}) - G(g) \right].$$

Denote

$$\mathcal{S}_0 = \left\{ \hat{\mu} \in \mathcal{S} : \int f(\mathbf{ax})\; d\hat{\mu}(\mathbf{ax}) = \int f(\mathbf{x})\; d\hat{\mu}(\mathbf{x}) \;,\; \forall\, f \in \mathcal{C} \right\}.$$

We denote by $\mathcal{M}$ the set of probabilities over $\mathcal{B}$.

Given F and G as above, we claim that

$$F^*(\hat{\mu}) = \begin{cases} -\displaystyle\int_{\hat{\Sigma}} A(\mathbf{y},\mathbf{x})\; d\hat{\mu}(\mathbf{y},\mathbf{x}) & \text{if } \hat{\mu} \in \mathcal{M} \\ +\infty & \text{otherwise} \end{cases} \quad \text{and}$$

$$G^*(\hat{\mu}) = \begin{cases} 0 & \text{if } \hat{\mu} \in \mathcal{S}_0 \\ -\infty & \text{otherwise} \end{cases} .$$

We refer the reader to the [38] or [62] for a proof of this claim (which is basically the same as we need here).

Once the correspondence F is Lipschitz, the theorem of duality of Fenchel-Rockafellar [70] assures

$$\sup_{g \in \mathcal{C}} [G(g) - F(g)] = \inf_{\hat{\mu} \in \mathcal{S}} [F^*(\hat{\mu}) - G^*(\hat{\mu})] .$$

$$\sup_{g \in \mathcal{G}} \left[- \max_{(\mathbf{a},\mathbf{x}) \in \mathbb{S}^1 \times \mathcal{B}} (A+g)(\mathbf{ax}) \right] = \inf_{\hat{\mu} \in \mathcal{M}_\sigma} \left[- \int A(\mathbf{ax}) \, d\hat{\mu}(\mathbf{ax}) \right] .$$

Finally, from the definition of $\mathcal{G}$, the claim of the theorem follows.

□

4 Entropy and Pressure - Notes by M. Craizer

In this section $M = \{1, 2, ..., d\}$. We follow [28].

Let $K = \{1, 2..., d\}^{\mathbb{N}}$ and $T : K \hookleftarrow$ be a continuous and locally injective map and $\mu \in \mathcal{M}(K)$.

We will take as main example $T = \sigma$ the shift which is a d to one map.

Entropy is a concept for invariant probabilities, that is, $\mu \in \mathcal{M}(T)$. It measures the dynamical complexity in the measure theoretical sense. Bigger the entropy more complex the statistics of the orbits under iteration of the map T. The entropy is always a non-negative number, and, if $T = \sigma$ is the shift (a d to one map) the maximal value of the entropy is $\log d$. A measure which attains such value is called a measure of maximal entropy.

We denoted $h(\mu)$ the entropy of the invariant probability μ.

The entropy of a invariant probability will defined later.

We are interested in the following kind of problem.

Definition 3. *Given a continuous potential ψ we call pressure of ψ the real value*

$$P(\psi) = \sup_{\mu \in \mathcal{M}(T)} \{h_\mu + \int \psi d\mu\},$$

where h_μ is the Kolmogorov entropy, and, the supremum is among the invariant probabilities for T, that is, in the set $\mathcal{M}(T)$.

This set is not empty because the entropy is upper semicontinuous for the shift.

Definition 4. *A probability μ which realizes the supremum is called an equilibrium state for ψ.*

Note that for any given ψ we have that if $\phi = \psi + g \circ T - g + c$, then $P(\phi) = P(\psi) + c$, and, they have both the same equilibrium states.

We want to characterize the equilibrium states and we will do so by relating this problem to the results of last section. We want to relate Gibbs probabilities with equilibrium probabilities.

When the potential ψ is Holder the equilibrium probability is unique. When the potential is not Holder there are cases where can exist more than one equilibrium state. This is associated to the so called phase transition theory (see [57], [36] and [18]).

Definition 5. *Given a compact metric space K and a continuous $T : K \to K$ such that*

a) each $x \in M$ has exactly d preimages, $d \in \mathbb{N}$,

b) $\forall x, y \in M,\ d(T(x), T(y)) > \xi_0\, d(x, y)$, for a fixed real number $\xi_0 > 1$,

then we say that T is an expanding map. We also say that ξ_0 is an expanding constant for T.

The shift transformation is clearly expanding.

The shift transformation $\sigma = T$ acting on $K = \{1, 2, ..., d\}^{\mathbb{N}}$ is not injective. But, restricted to the cylinder set $\overline{i}$, $i = 1, 2, ..., d$, is bijective (even a homeomorphism).

Note that T expands distances by a uniform constant multiplicative factor strictly bigger than one. This is classical example of expanding transformation. Most of our results easily extends to expanding transformation of fixed degree d in a compact metric space.

For each $i = 1, 2, ., d$ we denote $g_i : \overline{i} \to K$, the inverse of T on $\overline{i}$, and we call it the main contractive branch of T^{-1}. Note that g_i contract distances by a uniform constant multiplicative factor strictly smaller than one.

In the same way, the shift transformation $\sigma^2 = T^2$ acting on $K = \{1, 2, ..., d\}^{\mathbb{N}}$ is not injective. But, restricted to the cylinder set $\overline{i, j}$, $i = 1, 2, ..., d$, $j = 1, 2, ..., d$, is bijective (even a homeomorphism).

For each pair (i, j) we denote $g_{i,j} : \overline{i, j} \to K$, the inverse of T^2 on $\overline{i, j}$, and we call it the (i, j) main contractive branch of T^{-2}.

For each n we can consider inverse branches g of T^n in the same way.

A function positive continuous function $F : K \to \mathbb{R}$ is Jacobian of $\mu \in \mathcal{M}(K)$ if

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

for any Borel set A such that $T\,|_A$ is injective. The Jacobian, if it exists, is unique *a.e.* and it is denoted by J_μ.

Note that if $\mu \in \mathcal{M}(T)$, then, $\mathcal{L}_{\log J_\mu}(1) = 1$.
If $\psi \in C^\circ(K)$ define $\mathcal{L}_\psi : C^\circ(K) \hookleftarrow$ by

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

for any $\varphi \in C^\circ(k)$.

Lemma - Let $\nu \in \mathcal{M}(K)$ satisfying $\mathcal{L}^*_\psi \nu = \lambda \nu$, $\lambda > 0$.
Then

$$J_\nu = \lambda e^{-\psi}.$$

Let h be continuous and strictly positive and $\mu = h\nu$. Then

$$J_\mu = \lambda e^{-\psi} \frac{h \circ T}{h}.$$

Proof: Let A be a Borel set such that $f\,|_A$ is injective.
Take a sequence $\{h_n\}_{n \geq 1}$ in $C^\circ(K)$ such that $h_n \longrightarrow \mathcal{X}_A \quad a.e. \quad [\nu]$ and $\| h_n \|_{C^\circ} \leq 2$, $\forall\, n \geq 1$. Then

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

This last expression converges to $X_{T(A)}(x) \quad a.e. \quad [\nu]$ and so, by the dominated converge Theorem

$$\int \lambda e^{-\psi} h_n d\nu = \int \mathcal{L}_\psi(e^{-\psi} h_n) d\nu \to \nu(f(A)).$$

Hence

$$\int_A \lambda e^{-\psi} d\nu = \nu(T(A)).$$

Also

$$\begin{aligned} \int \lambda e^{-\psi} \frac{h \circ T}{h} h_n d\mu &= \int \mathcal{L}_\psi(e^{-\psi} h \circ T h_n) d\nu \\ &= \int \mathcal{L}_\psi(e^{-\psi} h_n) d\mu. \end{aligned}$$

Since μ and ν are equivalent, this last expression converges to $\mu(T(A))$ and so

$$\int_A \lambda e^{-\psi} \frac{h \circ T}{h} d\mu = \mu(f(A)).$$

□

From this follow that given a Holder potential A and the corresponding normalized potential $\bar{A}$, then $J_{m(A)} = e^{-\bar{A}}$, where $m(A)$ is the Gibbs state for A

Indeed, Let ψ_A be continuous and strictly positive main eigenfunction for the Ruelle operator for A, λ the main eigenfunction, and $m(A) = \psi_A \rho_A$. Then, from last proposition

$$J_{m(A)} = \lambda e^{-A} \frac{\psi_A \circ T}{\psi_A}.$$

From now on, we will suppose that T is a topologically mixing expanding map. If $\mu \in \mathcal{M}(K)$, define the support of μ as

$$supp(\mu) = \{\overline{x \in K \mid \text{ any neighborhood } V \text{ of } x, \mu(V) > 0}\}.$$

Lemma If $\mu \in \mathcal{M}(K)$ admits a Jacobian J_μ, then $supp(\mu) = K$.

Proof: Suppose that there exists an open set V with $\mu(V) = 0$.
Cover V with borel sets $A \subset V$ such that $T \mid_A$ is injective. Then

$$\mu(T(A)) = \int_A J_\mu \, d\mu = 0.$$

Hence $\mu(T(V)) = 0$. Inductively, $\mu(T^n(V)) = 0$. But since there exists $n \in \mathbf{N}$ such that $T^n(V) = K$, this is a contradiction.

□

We denote

$$J_\mu T^n(x) = \prod_{j=o}^{n-1} J_\mu(T^j(x)).$$

Lemma - If J_μ is strictly positive and Hölder-continuous, there exists $A > 0$ such that $\forall n$, if $g : S \to K$ is a contractive branch of T^{-n}, then

$$\frac{J_\mu T^n(x)}{J_\mu T^n(y)} \leq A$$

for any x, $y \in g(S)$.

Proof - Since g is a contractive branch of T^{-n}, and x, $y \in g(S)$

$$d(T^j(x), T^j(y)) \leq \lambda^{n-j} d(T^n(x), T^n(y)) \leq \lambda^{n-j} d$$

for $0 \leq j \leq n$, where d=diameter(S). Then

$$\frac{J_\mu T^n(x)}{J_\mu T^n(y)} = \prod_{j=o}^{n-1} \frac{J_\mu(T^j x)}{J_\mu(T^j(y))} \leq \prod_{j=o}^{n-1} \frac{\mid J_\mu(T^j x) - J_\mu(T^j y) \mid}{J_\mu T(T^j y)} + 1$$

$$\leq \prod_{j=o}^{n-1} 1 + \frac{1}{c} \mid J_\mu(T^j(x)) - J_\mu(T^j(y)) \mid$$

where $c = \inf_{x\in K} J_\mu T(x) > 0$. Hence

$$\begin{aligned}\frac{J_\mu T^n(x)}{J_\mu T^n(y)} &\leq \prod_{j=o}^{n-1} 1 + \frac{C}{c} d(T^j(x), T^j(y))^\gamma \leq \prod_{j=o}^{n-1} 1 + \frac{C}{c}(\lambda^{n-j})^\gamma d \\ &\leq \prod_{j=o}^{\infty} 1 + \frac{C}{c}(\lambda^\gamma)^j d \overset{def}{=} A.\end{aligned}$$

□

Given a Holder potential A, denote

$$e^{A_{T^n}}(x) = \prod_{j=o}^{n-1} e^A(T^j(x)).$$

Lemma - Given A a Hölder-continuous potential, there exists $B > 0$ such that $\forall n$, if $g : S \to K$ is a contractive branch of T^{-n}, then

$$\frac{e^{A_{T^n}}(x)}{e^{A_{T^n}}(y)} \leq B$$

for any x, $y \in g(S)$.

Proof: It follows from the relation of $\log J_{m_A}$ and the potential A via coboundary equation (ϕ_A is bounded).

□

The above property is called the bounded distortion for a Holder potential A.

Corollary (Distortion Lemma) - There exists $B > 0$ such that for any S_1, $S_2 \subset S$

$$\frac{1}{B}\frac{\mu(g(S_1))}{\mu(g(S_2))} \leq \frac{\mu(S_1)}{\mu(S_2)} \leq B\frac{\mu(g(S_1))}{\mu(g(S_2))}.$$

Proof: Fix $x_0 \in g(S)$. Then

$$\begin{aligned}\mu(S_1) &= \int_{g(S_1)} J_\mu T^n d\mu \leq A J_\mu T^n(x_0)\mu(g(S_1)) \\ \mu(S_2) &= \int_{g(S_2)} J_\mu T^n d\mu \geq \frac{1}{A} J_\mu T^n(x_0)\mu(g(S_2)).\end{aligned}$$

Then

$$\frac{1}{A^2}\frac{\mu(S_1)}{\mu(S_2)} \leq \frac{\mu(g(S_1))}{\mu(g(S_2))}.$$

Inverting the roles of S_1 and S_2 we obtain the other inequality.

□

By definition, the dynamic ball of center x and radius ξ is

$$B(n,\xi,x) = \{y \in K \,|\, d(T^j(x), T^j(y)) < \xi, \, \forall\, j \,, 0 \leq j \leq n-1\}.$$

Remark: It is easy to see that if $\xi > 0$ is such that $B(x,\xi) = B_\xi(x) = \overline{i_1, i_2, ..., i_n}$, where $x = (i_1, i_2, ..., i_n, i_{n+1}, ..)$, then

$$T^n(B_\xi(x)) = K.$$

The corresponding main inverse branch g of T^{-n} satisfies

$$g(K) = B_\xi(x).$$

Corollary: Let ξ_0 be an expansivity constant for T, and $0 < \xi \leq \xi_0$. Then there exists $C_\xi > 0$ such that

$$\frac{1}{C_\xi} \leq \mu(B(n,\xi,x)) \cdot J_\mu T^n(x) \leq C_\xi$$

$\forall n \geq 0$, $\forall x \in K$.

Proof: From the above remark, $B(n,\xi,x) = g(B_\xi(f^n x))$, where $g : B_r(T^n(x)) \to K$ is a contractive branch of T^{-n}.

Cover K with balls $B_1 ... B_l$ of radius $\xi/3$ and let $\delta_\xi = \min_{1\leq i \leq l} \mu(B_i)$. δ_ξ is strictly positive, because by former Lemma, μ is positive on open sets. Also, if $y \in K$,

$$\mu(B_\xi(y)) \geq \delta_\xi.$$

Hence

$$\delta_\xi \leq \mu(B_\xi(T^n(x))) = \int_{g(B_\xi(T^n(x)))} J_\mu T^n d\mu \leq A J_\mu T^n(x) \mu(B(n,\xi,x)).$$

It follows that

$$\mu(B(n,\xi,x)) \cdot J_\mu T^n(x) \geq \frac{\delta_\xi}{A} \overset{def}{=} \frac{1}{C_x i}.$$

Also

$$1 \geq \mu(B_\xi(T^n x)) \geq \frac{J_\mu T^n(x)}{A} \mu(B(n,\xi,x)) \geq \frac{\mu(B(n,\xi,x)) J_\mu T^n(x)}{C_\xi}.$$

Hence

$$\mu(B(n,\xi,x)) J_\mu T^n(x) \leq C_\xi.$$

□

Theorem Brin-Katok formula. [13] Given μ a T invariant probability on the compact metric space K, there exist the limit

$$\lim_{\xi \downarrow 0} \limsup_{n\to\infty} -\frac{1}{n} \log \mu(B(n,\xi,x)) = r(x)$$

a.e. $x \in K$. This defines a certain integrable measurable r.

If μ is ergodic, then the above limit is constant a. e. μ.

Definition: Given μ and g as above we call entropy of μ the nonnegative value

$$h(\mu) = \int r \, d\mu.$$

Corollary: Suppose that μ is Gibbs for A. Then

$$h_\mu = \int \log J_\mu d\mu.$$

Proof: - The Theorem of Brin-Katok claims that

$$h_\mu = \lim_{\xi \downarrow 0} \limsup_{n\to\infty} -\frac{1}{n} \log \mu(B(n,\xi,x))$$

a.e. $x \in K$. From the above it follows that

$$\limsup_{n\to\infty} -\frac{1}{n} \log \mu(B(n,\xi,x)) = \limsup_{n\to\infty} \frac{1}{n} \log J_\mu T^n(x)$$

if $0 < \xi < \xi_0$. From Birkhoff's Theorem it results that

$$\frac{1}{n} \log J_\mu T^n(x) = \frac{1}{n} \sum_{j=0}^{n-1} \log J_\mu T(T^j x) \to \int \log J_\mu d\mu$$

a.e. $x \in K$. Hence

$$h_\mu(T) = \int \log J_\mu d\mu.$$

□

Theorem - Let $K = \{1, 2..., d\}^{\mathbb{N}}$, T the shift, and A Hölder-continuous. Then, there exist $h : K \to \mathbf{R}$ Hölder-continuous and strictly positive, $\rho \in \mathcal{M}(K)$ and $\lambda > 0$ such that

1. $\int h d\rho = 1$

2. $\mathcal{L}_A \cdot h = \lambda h$

3. $\mathcal{L}_A^* \rho = \lambda \rho$

4. $\| \lambda^{-n} \mathcal{L}_A^n \varphi - h \int \varphi d\rho \|_{C^0} \to 0$, $\forall \varphi \in C^0(K)$.

5. h is the unique positive eigen-function of $\mathcal{L}_A$, except for multiplication by scalars.

6. The probability $\mu = m_A = h\rho$ is invariant, and

$$\log \lambda = h_\mu + \int A d\mu.$$

7. For any $\hat{\mu} \in \mathcal{M}(T)$,

$$\log \lambda \geq h_{\hat{\mu}} + \int A d\hat{\mu}.$$

8. $P(A) \;=\; \log \lambda \;=\; h_\mu + \int A d\mu$.

Proof: We have just to prove (6) and (7).

From a previous lemma we have when $h = \log \psi_A$, and λ the main eigenvalue that

$$J_\mu = \lambda e^{-A} \frac{h \circ T}{h}.$$

Hence by a previous Corollary

$$h_\mu(T) = \int \log J_\mu \, d\mu = \log \lambda - \int A d\mu$$

because μ is invariant. This proves (6).

Now we will prove (7).

We want to show that the Gibbs state m_A for A is the equilibrium state for A.

We can suppose without lost of generality that A is normalized. In this case $A = -\log J_{m_A} = -\log J_\mu$, $\lambda = 1$ and $h = 1$.

Then, all we have to show is that

$$\int \log J_{\hat{\mu}} d\hat{\mu} - \int \log J_\mu d\hat{\mu} \leq 0.$$

The set of invariant probabilities that are obtained as Gibbs states for Holder potentials is dense in the set of invariant probabilities (see for instance [46], [55], or [71]).

It is also know (see results after this theorem) that the entropy is upper-semicontinuous.

In this way, if we assume by contradiction that there exists an invariant probability $\hat{\nu}$ such that

$$h(\hat{\nu}) > \int \log J_\mu d\hat{\nu}$$

then, there exists also a Gibbs probability $\hat{\mu}$ for a Holder potential satisfying the same property, that is,

$$\int \log J_{\hat{\mu}} d\hat{\mu} \; > \int \log J_\mu d\hat{\mu}.$$

Then we can suppose without lost of generality a maximization among $\hat{\mu}$ which are Gibbs and have a jacobian $J_{\hat{\mu}} = J_2$. We denote the jacobian of m_A by J_1. Note that under such notation $\mathcal{L}^*_{-\log J_2}\hat{\mu} = \hat{\mu}$.

The following inequality is well known (see for instance [68] Lemma 3.3).

$$-\sum_i q_i \log q_i + \sum_i q_i \log p_i \leq 0,$$

for any set of positive $p_i, i = 1, 2, .., d,\ q_i, i = 1, 2, .., d,$ such that $\sum_i p_i = 1, \sum q_i = 1$. Moreover, if the equality happen, then any $p_i = q_i$, $i = 1, 2, ..., d$.

From this basic inequality we have for any $x \in K$

$$\sum_i J_2^{-1}(ix) \log J_2(ix) - \sum_i J_2^{-1}(ix) \log J_1(ix) \leq 0.$$

In other words, for any x

$$\mathcal{L}_{-\log J_2} \log J_2(x) - \mathcal{L}_{-\log J_2} \log J_1(x) \leq 0.$$

Now, if we integrate with respect to $\hat{\mu}$ we get

$$\int \mathcal{L}_{-\log J_2} \log J_2 d\hat{\mu} - \int \mathcal{L}_{-\log J_2} \log J_1 \hat{\mu} \leq 0.$$

As, $\mathcal{L}^*_{-\log J_2}\hat{\mu} = \hat{\mu}$, we finally obtain

$$h(\hat{\mu}) - \int \log J_{m_A} \hat{\mu} = \int \log J_2 d\hat{\mu} - \int \log J_1 \hat{\mu} \leq 0.$$

□

Theorem: [71] For expanding systems the entropy is upper-semicontinuous at any probability $\nu \in \mathcal{M}(T)$.

Remark: The results presented above can be used to show that a measure ν is invariant for an expanding map T, if and only if,

$$h(\nu) = \inf_{\phi \in C(X)} \{P(\phi) - <\phi, \nu>\}. \tag{10}$$

All the above is true for the shift.

Remark: One can show that if A is Holder, then the function on $t \in \mathbb{R}$ defined by $P(t\,a)$ is convex in t (see next section). Moreover, it is a differentiable (even real analytic) function of t [68] [71].

Proposition 16. *Consider the case $\mathcal{B} = \{1, 2, ..., d\}^{\mathbb{N}}$.*

Suppose $A = \bar{A}$ is normalized and n is fixed. Denote by x_j^n the d^n solutions x of $\sigma^n(x) = x$. Denote $\nu_{A,y}^n$ by

$$\nu_A^n = \sum_{j=1}^{d^n} e^{\bar{A}(x_j^n)+\bar{A}(\sigma(x_j^n))+\bar{A}(\sigma^2(x_j^n)+...+\bar{A}(\sigma^{n-1}(x_j^n))}\delta_{x_j^n}.$$

Then, we have, as $n \to \infty$.

$$\nu_A^n \to m_{\bar{A}}.$$

If A is not normalized then

$$\frac{1}{\lambda^n}\sum_{j=1}^{d^n} e^{\bar{A}(x_j^n)+\bar{A}(\sigma(x_j^n))+\bar{A}(\sigma^2(x_j^n)+...+\bar{A}(\sigma^{n-1}(x_j^n))}\delta_{x_j^n} = m_A,$$

where λ is the main eigenvalue for the Ruelle operator for A

Proof:

First note that ν_A^n is an invariant probability. This so because the sum $\bar{A}(x_j^n) + \bar{A}(\sigma(x_j^n)) + \bar{A}(\sigma^2(x_j^n) + ... + \bar{A}(\sigma^{n-1}(x_j^n))$ in each point of the same $n-$periodic orbit is the same. Note that if $\sigma^n(x) = x$, the minimal period of x is not necessarily n.

The ν_A^n is a convex combination of periodic measures, therefore, invariant.

We already know from a previous section that for a fixed y we have that

$$\mu_{A,y}^n \to m_{\bar{A}}.$$

By the bounded distortion property we have that given any ν which is weak limit probability $\nu_{A,y}^n$, then ν is absolutely continuous with respect to m_A. Indeed, for a fixed y note that in each domain of injectivity K_j^n there is one and only one x_j^n and one preimage at level n of y.

Note that ν is invariant because each ν_A^n also is.

Claim: if we consider two invariant probabilities which are absolutely continuous, if one is ergodic, then they are equal.

From the claim it follows that m_A is the only possible limit of subsequence ν_A^n.

The proof of the claim follows from the fact that the Birkhoff set (the set of full measures where the ergodic theorem can be applied) of the ergodic probability intersected with the corresponding one for the other invariant measure has full measure. Note that both complements has measure zero.

In the case A is not normalized, then, for a given n, the sum $A(x_j^n) + A(\sigma(x_j^n)) + A(\sigma^2(x_j^n) + ... + A(\sigma^{n-1}(x_j^n))$ is equal (up to the constant $-n \log \lambda$) to the sum of the normalized $\bar{A}$. This is so, because due to the coboundary equation there is a canceling in the telescopic sum.

□

Remark: The above result is useful for applications because you do not have to find the normalized $\bar{A}$ to get an approximation of the Gibbs state m_A.

Note also that any Gibbs state has positive entropy, but can be approximated by invariant probabilities with zero entropy. This shows that the entropy is not continuous. But, it is upper semicontinuous in the case of the shift acting on $\{1, 2, ., d\}^{\mathbb{N}}$ [71].

5 Large Deviations in time

Some of the results presented in this section appeared in some form in [55] and [56]. General references for Large Deviation are [31] and [30].

In this paragraph 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[31] 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 6. Given ϵ greater than zero and $n \in \mathbf{N}$, then by definition $Q_n(\epsilon)$ is equal to:

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

Proposition 17. *Given ϵ,*

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

Proof: For a given value ϵ denote

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

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 18. *Given* $\epsilon > 0$

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

One would like to be sure that the convergence to zero we consider above (see a future Proposition) 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.

It is quite surprising that in the case μ is an equilibrium state this result can be obtained using properties related to the Pressure. 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 behavior of the typical trajectory given by the Theorem of Birkhoff.

More generally 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

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

Definition 7. 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 $\lim_{n\to\infty} Q_n(A) = 1$.

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:

$$\lim_{n\to\infty} \frac{1}{n}\log \mu\{z \mid \frac{1}{n}\sum_{j=0}^{n-1} f(T^j(z)) \in A\} = \lim_{n\to\infty} \frac{1}{n}\log Q_n(A) = -Inf_{v\in A}\{I(v)\},$$

when A is an interval. In this case we say that a **Large Deviation Principle** is true for the observable f.

The function I it will be called the *deviation function*. The shape of I is basically the shape of $\mid v - \int f(x)d\mu(x) \mid^2$, $v \in \mathbf{R^m}$, 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 8. Given $n \in \mathbf{N}$ 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 9. Suppose that for each $t \in \mathbf{R^n}$ and $n \in \mathbf{N}$, 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 10. 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 19. *The function c(t) is convex in* $t \in \mathbf{R^n}$.

Proof: The Hölder inequality claims that

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

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.

□

Proposition 20. Given a Holder function f and t real we have

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

$$P(tf) - \log d.$$

Proof:
From a previous section for fixed t and y

$$\lim_{n\to\infty} \frac{1}{n} \log \mathcal{L}^n_{tf}(1)(y) = P(tf).$$

Note that the Jacobian of the maximal entropy measure is constant and equal to d.

For each n denote the d^n domains of injectivity for T^n by K^n_j. This is just the partition of cylinders set of level n.

Given a fixed y denote by y^n_j the d^n preimages of y by T^n. There is one and only one $y^n_j \in K^n_j$.

The integral

$$\int e^{t\,(f(x)+f(T(x))+f(T^2(x)+...+f(T^{n-1}(x))} d\mu(x) =$$

$$\sum_{j=1}^{d^n} \int_{K^n_j} e^{t\,(f(x)+f(T(x))+f(T^2(x)+...+f(T^{n-1}(x))} d\mu(x)$$

By the bounded distortion estimate (f is Holder), up to a multiplicative constant

$$\int_{K^n_j} e^{t\,(f(x)+f(T(x))+f(T^2(x)+...+f(T^{n-1}(x))} d\mu(x) \sim$$

$$e^{t\,(f(y^n_j)+f(T(y^n_j))+f(T^2(y^n_j)+...+f(T^{n-1}(y^n_j))}\, d^{-n}.$$

Note that the above estimate is uniform on n.

As, for each n

$$\mathcal{L}^n_{tf}(1)(y) = \sum_{j=1}^{d^n} e^{t\,(f(y^n_j)+f(T(y^n_j))+f(T^2(y^n_j)+...+f(T^{n-1}(y^n_j))},$$

the results follows.

□

Remark: The above shows that the pressure $P(tA)$ is convex on t for A Holder fixed.

We want to specify the deviation function I. This will be done next.

Define $I(v)$, $v \in \mathbf{R^m}$, as the Legendre transform of the function c(t), $t \in \mathbf{R^m}$, that is

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

This I will do the job as we will see.

This 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 $c^{'}(t_0) = v$. 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.

Remark: Note that from above it follows that $c(t)$ is differentiable on t because $P(tf)$ is differentiable on t [68].

In more precise mathematical terms one should say that the deviation function I(v) of c(t), $t \in \mathbf{R^m}$, takes values v in the dual of $\mathbf{R^m}$. The dual of $\mathbf{R^m}$ is $\mathbf{R^m}$ 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. If we need to consider Legendre transforms in infinite dimensional vector 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. We need to require that the free energy be differentiable which is true for Holder potential.

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

Theorem 21. *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* $\mathbf{R^m}$

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

where I is the Legendre transform of c.

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[31].

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

Definition 11. 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 22. *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 23. *Suppose $g(y)$ defined on $y \in \mathbf{R}$ satisfies $g^{"}(y) > \alpha > 0$ 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 $g^{"}(y) > \alpha > 0, y \in \mathbf{R}$ 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.

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

Proposition 24. *Suppose $f(x)$ and $f^*(x)$ are stricly convex and differentiable for every x, then the Legendre Transform is an involution, that is, $f^{**} = f$.*

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 $(y, z(y)) = (y, f(x) + f^{'}(x)y)$ (the tangent line to the graph of f on $(x, f(x))$) with the line $x = x_0$. 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

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

From a geometric reasoning it is easy to se that $\Delta(p) \leq f(x_0)$ and the maximum on p is $f(x_0)$.

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 $g^*(x_0) = f(x_0)$.
□

Definition 12. 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 13. Suppose g is a convex function on $\mathbf{R^m}$. We say that $y \in \mathbf{R^m}$ is a sub-differential 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 sub-differentials 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 $[u_i, v_i]$.

Characterization of the subdiferential. Suppose φ is convex defined in $\mathbb{R}^n$. Then, for all $x, y \in \mathbb{R}^n$,

$$\langle x, y\rangle = \varphi(x) + \varphi^*(y) \Longleftrightarrow y \in \partial\varphi(x).$$

Given $x, y \in \mathbb{R}^n$ any,

$$\begin{aligned} \langle x, y\rangle = \varphi(x) + \varphi^*(y) &\Longleftrightarrow \forall z \in \mathbb{R}^n, \quad \langle x, y\rangle \geq \varphi(x) + \langle y, z\rangle - \varphi(z) \\ &\Longleftrightarrow \forall z \in \mathbb{R}^n, \quad \varphi(z) \geq \varphi(x) + \langle y, z - x\rangle \\ &\Longleftrightarrow y \in \partial\varphi(x). \end{aligned}$$

In the first equivalence we use the fact that $\varphi^{**} = \varphi$.

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

Proposition 25. *$y \in \delta g(x)$, 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 $< y, z > -g(z)$.

From the above reasoning that $y \in \delta g(x)$ is 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 26. *$I(v) = 0$, 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 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) [31]. In the second inequality we will use differentiability of the free energy.

Proposition 27. *(Upper large deviation inequality) Suppose c(t), $t \in \mathbf{R}$ is a well defined convex function, then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu\{x \mid \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \in K\} =$$

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n(K) \leq -Inf_{z\in K} I(z) \qquad (11)$$

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 [31] 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

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

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

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n([a,\infty)) \le - \sup_{t\ge 0}\{ta - c(t)\}. \tag{12}$$

Now we need the following claim:

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

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

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

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

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

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

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n\{K\} \le - \inf_{z\in K} I(z) \tag{13}$$

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):

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n(K) \leq \limsup_{n\to\infty} \frac{1}{n} \log(Q_n(-\infty, a] + Q_n[b, \infty)) \leq$$

$$\limsup_{n\to\infty} \frac{1}{n} (\log Q_n(-\infty, a] + \log Q_n[b, \infty)) \leq -I(a) - I(b) \leq$$

$$-Inf\{I(a), I(b)\} = Inf_{z\in K} I(z).$$

Therefore the Proposition is proved.

□

Proposition 28. *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 $I(z) \geq I(v) = 0$ 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 $n \in \mathbf{N}$:

$$\mu(\{z \,|\, |\, \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(z)) \in K\} =$$

$$\mu(\{z \mid \;\frac{1}{n} \sum_{j=0}^{n-1} f(T^j(z)) - c^{'}(0) \,|\geq \epsilon\} \leq e^{-nM}. \tag{14}$$

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 14. 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:

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

for n large enough.

Proposition 29. *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

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

□

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

Definition 15. 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).

<u>Remark</u> Note that for any interval (a, b) contained in $\mathbb{R}$,

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

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 30. *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*

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

$$\lim_{n\to\infty} \frac{1}{n} \int e^{snx} d\nu_n(x)$$

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*

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

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

$$\lim_{n\to\infty} \nu_n(B(c'(0)), \epsilon)) = 1 \tag{15}$$

(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 31. *(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

$$d\mu_t^n(x) = \frac{e^{ntx}}{Z_n(t)} d\mu^n(x). \tag{16}$$

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:

<u>Claim</u> Suppose $c'(t) = z$, then for any r:

$$\lim_{n\to\infty} \mu_t^n((z-r, z+r)) = 1 \tag{17}$$

<u>Proof of the Claim:</u>

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:

$$c_t(s) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{snx} d\mu_n^t(x)$$

$$\lim_{n\to\infty} \frac{1}{n} \log \int \frac{e^{nx(s+t)}}{e^{nc_n(t)}} d\mu^n(x) =$$

$$\lim_{n\to\infty} \frac{1}{n} \log \int e^{nx(s+t)} d\mu^n(x) - \lim_{n\to\infty} \frac{1}{n} \log e^{c_n(t)n} =$$

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

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, $n \in \mathbf{N}$), 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:

$$Q_n(A) \geq Q_n(B(z,r)) = \int_{B(z,r)} d\mu^n(x) =$$

$$Z_n(t)\int_{B(z,r)} e^{-ntx}\mu_t^n(x) \geq$$
$$e^{n(c_n(t)-tz)-rn|t|}\mu_t^n(B(z,r)).$$

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

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

That is the Large Deviation Principle is true.

References

[1] T. Bousch, La condition de Walters. *Ann. Sci. ENS*, 34, (2001)

[2] V. Bangert, Mather sets for twist maps and geodesics on tori, *Dynamics Reported* 1, 1-56, 1988.

[3] A. Baraviera, A. O. Lopes and Ph. Thieullen, A Large Deviation Principle for Gibbs states of Holder potentials: the zero temperature case. *Stoch. and Dyn.* (6), 77-96, (2006).

[4] A. Baraviera, R. Leplaideur and A. O. Lopes, Selection of measures for a potential with two maxima at the zero temperature limit, *SIAM Journal on Applied Dynamical Systems*, Vol. 11, n 1, 243-260 (2012)

[5] A. Baraviera, R. Leplaideur and A. O. Lopes, Ergodic Optimization, zero temperature and the Max-Plus algebra, 23^o Coloquio Brasileiro de mAtematica, IMPA, Rio de Janeiro, (2013)

[6] A. Baraviera, L. M. Cioletti, A. O. Lopes, J. Mohr and R. R. Souza, On the general XY Model: positive and zero temperature, selection and non-selection, *Rev. in Math Phys*, Vol. 23, N. 10, pp 1063-1113 (2011).

[7] A. Baraviera, A. O. Lopes A and J. Mengue, On the selection of subaction and measure for a subclass of Walters's potentials, Volume 33, issue 05, pp. 1338–1362, *Erg. Theo. and Dyn. Syst.* (2013)

[8] P. Bernard and G. Contreras. A Generic Property of Families of Lagrangian Systems. *Annals of Math.* Vol. 167, No. 3, 2008

[9] P. Billingsley, Probability and Measure, Wyley Ed. (2012)

[10] R. Bissacot, E. Garibaldi and Ph. Thieullen, Zero-temperature phase diagram for double-well type potentials in the summable variation class, to appear in Ergodic Theory and Dynamical Systems

[11] R. Bissacot and R. dos Santos Freire Jr., On the existence of maximizing measures for irreducible countable Markov shifts: a dynamical proof, Ergodic Theory Dynam. Systems 34 (2014), no. 4, 1103-1115

[12] R. Bowen, Gibbs States and the Ergodic Theory of Anosov Diffeomorphisms, Lecture notes in Math., volume 470, Springer-Verlag, 1975.

[13] M. Brin and A. Katok, On local entropy. Geometric dynamics, 3038, Lecture Notes in Math., 1007, Springer, Berlin, 1983

[14] J. Brémont, Gibbs measures at temperature zero. *Nonlinearity*, 16(2): 419–426, 2003.

[15] J.R. Chazottes and M. Hochman, On the zero-temperature limit of Gibbs states, *Commun. Math. Phys.*, Volume 297, N. 1, 2010

[16] J.R. Chazottes, J.M. Gambaudo and E. Ulgade, Zero-temperature limit of one dimensional Gibbs states via renormalization: the case of locally constant potentials. *Erg. Theo. and Dyn. Sys.* (2010).

[17] L. Cioletti, M. Denker, A. O. Lopes and M. Stadlbauer, Spectral Properties of the Ruelle Operator for Product Type Potentials on Shift Spaces, to appear in Journal of the London Mathematical Society

[18] L. Cioletti and A. O. Lopes, Phase Transitions in One-dimensional Translation Invariant Systems: a Ruelle Operator Approach, Journ. of Statistical Physics, 159 - Issue 6, 14241455 (2015)

[19] L. Cioletti and A. O. Lopes, Ruelle Operator for Continuous Potentials and DLR-Gibbs Measures, preprint (2017)

[20] L. Cioletti and A. O. Lopes, Interactions, Specifications, DLR probabilities and the Ruelle Operator in the One-Dimensional Lattice, preprint (2014).

[21] B. Kloeckner, P. Giulietti, A. O Lopes, and D. Marcon The calculus of thermodynamical formalism, to appear in Journ. of the European Math Society

[22] W. Chou and R. J. Duffin, An additive eigenvalue problem of physics related to linear programming, *Advances in Applied Mathematics* **8** (1987), 486-498.

[23] W. Chou and R. Griffiths, Ground states of one-dimensional systems using effetive potentials, *Physical Review B*, Vol. 34, N 9, 6219-6234, 1986

[24] G. Contreras, A. O. Lopes and Ph. Thieullen. Lyapunov minimizing measures for expanding maps of the circle, *Ergodic Theory and Dynamical Systems* Vol 21, 1379-1409, 2001.

[25] G. Contreras, Ground states are generically a periodic orbit, Invent. Math. 205 (2016), no. 2, 383412

[26] G. Contreras and R. Iturriaga. Global minimizers of autonomous Lagrangians, 22° Colóquio Brasileiro de Matemática, IMPA, 1999.

[27] J. Conway, A Course in Functional Analysis, Springer Verlag, 1990

[28] M. Craizer, Teoria Ergódica das Transformações expansoras, Master Dissertation, IMPA, Rio de Janeiro (1985)

[29] K. Deimling. Nonlinear Functional Analysis, Springer Verlag, 1985

[30] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, Springer Verlag, 1998.

[31] R. Ellis, Entropy, Large Deviations, and Statistical Mechanics, Springer Verlag, 2005

[32] A. C. D. van Enter, R. Fernandez and A. D. Sokal, Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory, *Journ. of Stat. Phys.* V.72, N 5/6 879-1187, (1993)

[33] A. C. D. van Enter and W. M. Ruszel, Chaotic Temperature Dependence at Zero Temperature, *Journal of Statistical Physics*, Vol. 127, No. 3, 567-573, (2007)

[34] A. Fathi, Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens, *Comptes Rendus de l'Académie des Sciences, Série I, Mathématique* Vol 324 1043-1046, 1997.

[35] P. Fernandez, Teoria da Medida, Proj. Euclides, IMPA

[36] A. Fisher and A. O. Lopes, Exact bounds for the polynomial decay of correlation, 1/f noise and the central limit theorem for a non-Holder Potential, Nonlinearity, Vol 14, Number 5, pp 1071–1104 (2001).

[37] Y. Fukui and M. Horiguchi, One-dimensional Chiral XY Model at finite temperature, *Interdisciplinary Information Sciences*, Vol 1, 133-149, N. 2 (1995)

[38] E. Garibaldi and A. O. Lopes, On Aubry-Mather theory for symbolic Dynamics, *Ergodic Theory and Dynamical Systems*, Vol 28 , Issue 3, 791-815 (2008)

[39] E. Garibaldi and A. O. Lopes, The effective potential and transshipment in thermodynamic formalism at temperature zero, Vol 13 - N 1, 1250009 (13 pages) (2013) *Stoch. and Dynamics*

[40] E. Garibaldi and Ph. Thieullen, Minimizing orbits in the discrete Aubry-Mather model, Nonlinearity, 24 (2011), no. 2, 563611

[41] E. Garibaldi and Ph. Thieullen, Description of some ground states by Puiseux technics, to appear *Journ. of Statis. Phys.*

[42] H.-O. Georgii, Gibbs Measures and Phase Transitions. de Gruyter, Berlin, (1988).

[43] D. A. Gomes, Viscosity solution method and the discrete Aubry-Mather problem, *Discrete and Continuous Dynamical Systems, Series A* **13** (2005), 103-116.

[44] D. A. Gomes and E. Valdinoci, Entropy Penalization Methods for Hamilton-Jacobi Equations, *Adv. Math.* 215, No. 1, 94-152, 2007.

[45] D. A. Gomes, A. O. Lopes and J. Mohr, The Mather measure and a large deviation principle for the entropy penalized method, *Communications in Contemporary Mathematics* 13 (2011), no. 2, 235-268

[46] R. B. Israel, Convexity in the Theory of Lattice Gases, Princeton Press, (1979)

[47] O. Jenkinson. Ergodic optimization, *Discrete and Continuous Dynamical Systems, Series A*, V. 15, 197-224, 2006

[48] S. Karlin. Total Positivity. Standford Univ. Press, 1968.

[49] T. Kato, Perturbation Theory of Linear Operators, Springer, 1966.

[50] G. Keller, Gibbs States in Ergodic Theory, Cambrige Press, 1998.

[51] K. Oliveira and M. Viana, Ergodic Theory, Soc. Bras. Mat. - Rio de Janeiro (2014).

[52] J. Lebowitz and A. Martin-Lf, On the Uniqueness of the Gibbs State for Ising Spin Systems, *Comm. Math. Phys.* **25**, 276-282 (1972).

[53] R. Leplaideur, A dynamical proof for the convergence of Gibbs measures at temperature zero, *Nonlinearity*, 18(6):2847–2880, 2005.

[54] A. O. Lopes, An analogy of charge distribution on Julia sets with the Brownian motion. *J. Math. Phys.* 30 9. (1989), 2120-2124.

[55] A. O. Lopes, Entropy and Large Deviation, *NonLinearity*, Vol. 3, N. 2, 527-546, 1990.

[56] A. O. Lopes, Entropy, Pressure and Large Deviation, Cellular Automata, Dynamical Systems and Neural Networks, E. Goles e S. Martinez (eds.), Kluwer, Massachusets, pp. 79-146, 1994.

[57] A. O. Lopes, The Zeta Function, Non-Differentiability of Pressure and The Critical Exponent of Transition, Advances in Mathematics, Vol. 101, pp. 133-167, 1993.

[58] A. O. Lopes, J. Mohr, R. Souza and Ph. Thieullen, Negative entropy, zero temperature and stationary Markov chains on the interval, *Bulletin of the Brazilian Mathematical Society* **40** (2009), 1-52.

[59] A. O. Lopes and J. Mengue, Zeta measures and Thermodynamic Formalism for temperature zero, *Bulletin of the Brazilian Mathematical Society* 41 (3) pp 449-480 (2010)

[60] A. O. Lopes and J. Mengue, Selection of measure and a Large Deviation Principle for the general XY model, Vol 29, Issue 1 pp 24-39, *Dynamical Systems* (2014)

[61] A. O. Lopes, E. R. Oliveira and Ph. Thieullen, The dual potential, the involution kernel and transport in ergodic optimization, Dynamics, Games and Science -International Conference and Advanced School Planet Earth DGS II, Portugal (2013), Edit. J-P Bourguignon, R. Jelstch, A. Pinto and M. Viana, Springer Verlag, 331-364

[62] A. O. Lopes, J. Mohr, R. Souza and Ph. Thieullen, Negative entropy, zero temperature and stationary Markov chains on the interval, *Bulletin of the Brazilian Mathematical Society* **40**, 1-52, 2009.

[63] A. O. Lopes, J. K. Mengue, J. Mohr and R. R. Souza, Entropy and Variational Principle for one-dimensional Lattice Systems with a general a-priori probability: positive and zero temperature, to appear in *Erg. Theory and Dyn. Systems.*

[64] R. Mañé, The Hausdorff dimension of diffeomorphisms of Surfaces, Bol. Soc. Bras. Mat, Vol. 20, N.2 (1990)

[65] R. Mañé. Generic properties and problems of minimizing measures of Lagrangian systems, *Nonlinearity*, Vol 9, 273-310, 1996.

[66] J. Mather, Action minimizing invariant measures for positive definite Lagrangian Systems, *Math. Z.*, 207 (2), pp 169-207, 1991

[67] D. H. Mayer, *The Ruelle-Araki transfer operator in classical statistical mechanics*, LNP 123, Springer Verlag 1980

[68] W. Parry and M. Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics, *Astérisque* Vol 187-188 1990

[69] M. Pollicott and M. Yuri, Dynamical systems and Ergodic Theory, Cambrige Press, 1998

[70] R. T. Rockafellar, Extension of Fenchel's duality theorem for convex functions, *Duke Math. J.*, 33, 81-89, 1966.

[71] D. Ruelle, Thermodynamic Formalism, second edition, Cambridge, 2004.

[72] H. H. Schaefer. Banach Lattices and Positive Operators, Springer Verlag, 1974.

[73] B. Simon, The Statistical Mechanics of Lattice Gases, Princeton Univ Press, 1993

[74] E. A. da Silva, Dimensão de Hausdorff de Ferraduras, Diss. Mestrado UNB, Brasilia (2010)

[75] E. A. da Silva, R. R. da Silva and R. R. Souza, The Analyticity of a Generalized Ruelle's Operator, *Bull. Braz. Math. Soc.*, 45 (1), 53-72 (2014)

[76] R. R. Souza, Sub-actions for weakly hyperbolic one-dimensional systems, *Dynamical Systems* 18 **(2)**, 165-179 (2003).

[77] Y. Velenik, Phase Separation as a Large Deviations Problem. Phd Thesis. Lausanne. (2003).

[78] A. Taylor, Introduction to Functional Analysis, Krieger Pub Co, (1986).

[79] F. Spitzer. A Variational characterization of finite Markov chains. *The Annals of Mathematical Statistics.* (43): N.1 303-307, 1972.

[80] C. Walkden, Ergodic Theory, Lecture Notes University of Manchester (2014)

[81] A. N. Shiryaev, Probability. Second Edition, Springer (1984).

[82] P. Walters, A natural space of functions for the Ruelle operator theorem, *Ergodic Theory and Dynamical Systems.* 27, 1323-1348 (2007).