Ergodic Optimization, Zero Temperature Limits and the Max-Plus Algebra

Publicações Matemáticas

Ergodic Optimization, Zero Temperature Limits and the Max-Plus Algebra

Alexandre Baraviera
UFRGS

Renaud Leplaideur
U. de Bretagne Occidentale

Artur Lopes
UFRGS

29º Colóquio Brasileiro de Matemática

- Análise em Fractais – Milton Jara
- Asymptotic Models for Surface and Internal Waves – Jean-Claude Saut
- Bilhares: Aspectos Físicos e Matemáticos – Alberto Saa e Renato de Sá Teles
- Controle Ótimo: Uma Introdução na Forma de Problemas e Soluções – Alex L. de Castro
- Eigenvalues on Riemannian Manifolds – Changyu Xia
- Equações Algébricas e a Teoria de Galois – Rodrigo Gondim, Maria Eulalia de Moraes Melo e Francesco Russo
- **Ergodic Optimization, Zero Temperature Limits and the Max-Plus Algebra – Alexandre Baraviera, Renaud Leplaideur e Artur Lopes**
- Expansive Measures – Carlos A. Morales e Víctor F. Sirvent
- Funções de Operador e o Estudo do Espectro – Augusto Armando de Castro Júnior
- Introdução à Geometria Finsler – Umberto L. Hryniewicz e Pedro A. S. Salomão
- Introdução aos Métodos de Crivos em Teoria dos Números – Júlio Andrade
- Otimização de Médias sobre Grafos Orientados – Eduardo Garibaldi e João Tiago Assunção Gomes

Abstract

The purpose of this book is to present for math students some of the main ideas and issues which are considered in Ergodic Optimization. This text is also helpful for a mathematician with no experience in the area.

Our focus here are the questions which are associated to selection of probabilities when temperature goes to zero.

We believe that concrete examples can be very useful for any person which is reading this topic for the first time. We care about this point. The reader will realize that the use of the so called Max-Plus Algebra resulted in a very helpful tool for computing explicit solutions for the kind of problems we are interested here.

We are not concerned in our text in presenting results in their more general form nor point out who were the person that did this or that first. By the other hand, in the bibliography, we try to mention all results which appear in the literature. The reader can look there, and find other references which complement our short exposition. We are sorry for the case we eventually do not mention some paper in the area. Simply we were not aware of the work.

We believe there is a need for a kind of text like ours. The purpose is to present some of the basic ideas of this beautiful theory for a more broad audience.

We would like to thank several colleagues who read drafts of this book and made several suggestions for improvement.

A. T. Baraviera
R. Leplaideur
A. O. Lopes

Rio de Janeiro, may 7, 2013.

Contents

Chapter 1

Preliminaires

1.1 The configuration's space

1.1.1 Topological properties

We consider the space $\Omega = \{1, 2, ..., k\}^{\mathbb{N}}$ where the elements are sequences $x = (x_0, x_1, x_2, x_3...)$, in which $x_i \in \{1, 2, ..., k\}, i \in \mathbb{N}$. An element in Ω will also be called an *infinite word* over the alphabet $\{1, \ldots, k\}$, and x_i will be called a *digit* or a *symbol*.

The distance between two points $x = x_0, x_1, \ldots$ and $y = y_0, y_1, \ldots$ is given by

$$d(x, y) = \frac{1}{2^{\min\{n, \ x_n \neq y_n\}}}.$$

Example. In the case $k = 4$, $d(1, 2, 1, 3, 4..), (1, 2, 1, 2, 3, ..)) = \frac{1}{2^3}$.

We can represent this distance graphically as shown in figure 1.1.
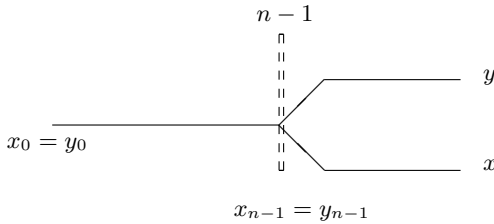
Figure 1.1: The sequence x and y coincide from the digit 0 up to the digit $n-1$, and then split.

A finite string of symbols $x_0 \ldots x_{n-1}$ is also called a *word*, of length n. For a word w, its length is denoted by $|w|$. A *cylinder* (of length n) is denoted by $[x_0 \ldots x_{n-1}] \subset \Omega$. It is the set of points y such that $y_i = x_i$ for $i = 0, \ldots n - 1$. For example, the cylinder $[2, 1, 1]$ is the set of $y = (y_0, y_1, y_2, y_3, y_4, \ldots)$ such that $y_0 = 2, y_1 = 1, y_2 = 1$.

Note that cylinders of length n form a partition of Ω. Given $x \in \Omega$, then $C_n(x)$ will denote the unique element of this partition which contains x. That is, $C_n(x) = [x_0 \ldots x_n]$, when $x = x_0 x_1 x_2 x_3 \ldots$. The cylinder $C_n(x)$ also coincides with the ball $B(x, \frac{1}{2^n})$. The set of cylinders of length n will be denoted by $\mathcal{C}_n(\Omega)$.

If $\omega = \omega_0 \ldots \omega_{n-1}$ is a finite word of length n and $\omega' = \omega'_0 \omega'_1 \ldots$ is a word (of any length possibly infinite), then $\omega\omega'$ is the word

$$\omega_0 \ldots \omega_{n-1}\omega'_0\omega'_1 \ldots$$

It is called the *concatenation* of ω and ω'.

The set (Ω, d) is a compact metric space. Compactness also follows from the fact that Ω is a product of compact spaces. Note that the topology induced by the distance d coincides with the product topology. The cylinders are clopen (closed and open) sets and they generate the topology.

The general subshifts of finite type

We will need to consider later the general subshift of finite type. A good reference for symbolic dynamics is [95].

Definition 1.1. A *transition matrix* is a $d \times d$ matrix with entries in $\{0, 1\}$.

If $T = (T_{ij})$ is a $d \times d$ transition matrix, the subshift of finite type Σ_T associated to T is the set of sequences $x = x_0 x_1 x_2 \ldots x_n \ldots$ such that for every j,

$$T_{x_j x_{j+1}} = 1.$$

It is exactly the set of infinite words such that a subword ij appears only if $T_{ij} = 1$. Equivalently, it is the set of words such that the subword ij never appears if $T_{ij} = 0$.

Example. The full shift $\Omega = \{1, \ldots d\}^{\mathbb{N}}$ is the subshift of finite type associated to the $d \times d$ matrix with all entries equal to 1.

The best way to understand what is the subshift of finite associated to a matrix T is to consider paths: for a given i, the set of j such that $T_{ij} = 1$, is the set of letters j authorized to follow i. Then, Σ_T is the set of infinite words we can write respecting theses rules, or equivalently, the set of infinite paths that we can do.

Example. The transition matrix is $T = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$.

When there are more 1's than 0's in T it is simpler to describe Σ_T (there are less restrictions). In the present case we get the set of infinite words x with letters in the alphabet 1, 2, 3 and 4, such that, 12, 14, 23, 41 and 44, never appear in each sequence $x = x_0 x_1 x_2 x_3 \ldots$

Definition 1.2. Let T be a $d \times d$ transition matrix, and Σ_T be the associated subshift of finite type. Two digits i and j are said to be *associated*, if there exists a path from i to j, and, another from j to i. We set $i \sim j$.

Equivalently, $i \sim j$ means that there exists a word in Σ_T of the form

$$i \ldots j \ldots i.$$

Obviously $\sim$ is an equivalence relation, and this defines classes of equivalence. It can happen that there are equivalence classes strictly contained in $\{1, \ldots d\}$.

Exercise 1

Find examples of transition matrices such that one digit is associated only to itself and not to any other digit.

Definition 1.3. The subsets of words in Σ_T such that all their digits belong to the same equivalence classes of $\sim$ are called the irreducible components of Σ_T.

1.1.2   Dynamics

Definition 1.4. The shift $\sigma : \Omega \to \Omega$, is defined by

$$\sigma(x_0 x_1 x_2 x_3 \ldots) = x_1 x_2 x_3 \ldots .$$

The shift expands distance by a factor 2:

$$d(\sigma(x), \sigma(y)) = 2 \, d(x, y).$$

Hence, it is Lipschitz and thus continuous.

Definition 1.5. Given $x \in \Omega$ the set $\{\sigma^n(x), n \geq 0\}$ is called the orbit of x. It is denoted by $\mathcal{O}(x)$.

The main goal in Dynamical Systems is to describe orbits and their behaviors. Let us first present the simplest of all.

Definition 1.6. A point $x \in \Omega$ is said to be periodic if there exists $k > 0$, such that, $\sigma^k(x) = x$. In that case the period of x is the smaller positive integer k with this property.

A periodic point of period 1 is called a fixed point.

Examples

$111\ldots$ is a fixed point.

In $\{1, 2, 3, 4\}^{\mathbb{N}}$, $x := 1323132313231323\ldots$ has period 4 and the four points

$$13231323\ldots, \quad 32313231\ldots, \quad 23132313\ldots, \quad 31323132\ldots$$

form the orbit of x.

A n-periodic point is entirely determined by its first n-digits; actually it is the infinite concatenation of these first digits:

$$x = \underbrace{x_0 \ldots x_{n-1}} \ \underbrace{x_0 \ldots x_{n-1}} \ \underbrace{x_0 \ldots x_{n-1}} \ldots$$

As a notation we shall set $x = (x_0 \ldots x_{n-1})^\infty$.

Definition 1.7. Given a point x in Ω, a point $y \in \Omega$, such that, $\sigma(y) = x$ is called a *preimage of x*.

A point y such that $\sigma^n(y) = x$ is called a *n-preimage of x*. The set $\{y \mid$ there exists an n such that $\sigma^n(y) = x \}$ is called the *preimage set of x*.

For the case of the full shift in $\Omega = \{1, \ldots k\}^{\mathbb{N}}$, each point x has exactly k-preimages. They are obtained by the concatenation process ix, with $i = 1, \ldots, k$. The set of 1-preimage is $\sigma^{-1}(\{x\})$. The set of n-preimages is $(\sigma^n)^{-1}(\{x\})$ which is simply denoted by $\sigma^{-n}(\{x\})$.

Definition 1.8. A Borel set A is said to be σ-invariant if it satisfies one of the following equivalent properties:

1. For any $x \in A$, $\sigma(x)$ belongs to A.

2. $\sigma^{-1}(A) \supset A$.

Example. If x is periodic, $\mathcal{O}(x)$ is σ-invariant.

Exercise 2
Show equivalence of both properties mentioned in Definition 1.8.

Exercise 3
If x is periodic, do we have $\sigma^{-1}(\mathcal{O}(x)) = \mathcal{O}(x)$?

Back to the general subshifts of finite type

We have defined above the general subshift of finite type. For such a subshift, we have also defined the irreducible components. Here, we give a better description of these components with respect to the dynamics.

Definition 1.9 (and proposition[1]). A σ-invariant compact set $\mathbb{K}$ is called transitive if it satisfies one of the two equivalent properties:

(i) For every pair of open sets of $\mathbb{K}$, U and V, there exists $n > 0$, such that, $\sigma^{-n}(U) \cap V \neq \emptyset$.

(ii) There exists a dense orbit.

Exercise 4
Show that if $\mathbb{K}$ is transitive, then the set of points in $\mathbb{K}$ with dense orbit is a G_δ-dense set.

We claim that irreducible components are also transitive components. Indeed, any open set contains a cylinder and it is thus sufficient to prove (i) with cylinders. Now, considering two cylinders of the form $[x_0 \ldots x_k]$ and $[y_0 \ldots y_n]$, the relation defining irreducible components shows that there exists a connection

$$x_0 \ldots x_k z_1 \ldots z_m y_0 \ldots y_n,$$

remaining in $\mathbb{K}$.

1.1.3 Measures

We refer the reader to [18] and [57] for general results in measure theory.

We denote by $\mathcal{B}$ the Borel σ-algebra over Ω, that is, the one generated by the open sets.

We will only consider signed measures (probabilities) μ on Ω over this sigma-algebra $\mathcal{B}$, which we call Borel signed measures (probabilities). Due to the fact that cylinders are open sets and generate the topology, they also generate the σ-algebra $\mathcal{B}$. Therefore, the values $\mu(C_n)$, where C_n runs over all the cylinders of length n, and n runs over $\mathbb{N}$, determine uniquely μ.

We remind the relation between Borel measures and continuous functions:

---

[1]We do not prove the proposition part.

Theorem 1.10 (Riesz). *The set of Borel signed measures is the dual of the set $\mathcal{C}^0(\Omega)$.*

In other words, this means that any linear bounded transformation $G : \mathcal{C}^0(\Omega) \to \mathbb{R}$, is of the form

$$G : f \mapsto G(f) = \int f \, d\nu,$$

where ν is a fixed signed measure. The map $G \to \nu$ is a bijection.

Probabilities ν are characterized by the two properties: for any $f \geq 0$, we have that $G(f) \geq 0$, and $G(1) = 1$.

A subset K of a linear space is convex, if, for any given $x, y \in K$, and any λ, such that $0 \leq \lambda \leq 1$, we have that $\lambda \, x + (1 - \lambda) \, y$ is in K.

Corollary 1.11. *The set of all probabilities on a compact space is compact and convex for the weak*-topology.*

We remind that $\mu_n \xrightarrow{\text{w}^*} \mu$ means that for every continuous function $f : \Omega \to \mathbb{R}$,

$$\int f \, d\mu_n \to_{n \to \infty} \int f \, d\mu.$$

We point out that any indicator function of a cylinder, $\mathbb{1}_{C_n}$, is continuous. We recall that the support of a (probability) measure μ is the set of points x such that

$$\forall \varepsilon, \ \mu(B(x, \varepsilon)) > 0.$$

In our case, x belongs to the support, if and only, if $\mu(C_n(x)) > 0$, for every n. The support is denoted by supp (μ).

Exercise 5
Show that supp (μ) is compact.

We recall that a probability over Ω is a measure μ such that $\mu(\Omega) = 1$. In this book we shall only consider probabilities. Moreover, most of the time, we shall be interested in properties holding *almost everywhere*, that is, properties which are true for all points $x \in \Omega$, up to a set of probability zero.

Definition 1.12. We say that a probability μ is *invariant for the shift* σ, if for any $A \in \mathcal{B}$,

$$\mu(\sigma^{-1}(A)) = \mu(A).$$

We will also say that μ is σ-invariant, or simply invariant (as σ is the unique dynamics we shall consider).

To consider invariant measure means the following thing: if we see the action of $\mathbb{N}$ as a temporal action on the system, the systems is closed, in the sense that along the time, there is neither creation nor disappearance of mass in the system. In other words, the mass of a certain given set is constant along the time evolution.

Using the vocabulary from Probability Theory, the study of invariant probabilities correspond to the study of Stationary Processes (see [86] [112]).

Exercise 6
Show that if μ is invariant, supp μ is invariant. Is it still the case if μ is not invariant ?

In Ergodic Theory one is mainly interested in invariant probabilities, and the properties that are true for points x which are in a set which have mass equal to one.

1.2 Invariant measures

In this section we present some particular invariant measures in our setting and we will also present some more general results.

1.2.1 Examples of invariant measures

Periodic measures

If x is a point in Ω, δ_x is the *Dirac* measure at x, that is

$$\delta_x(A) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ otherwise.} \end{cases}$$

Then, if x is n-periodic,

$$\mu := \frac{1}{n} \sum_{j=0}^{n-1} \delta_{\sigma^j(x)}$$

is σ-invariant.

The Bernoulli product measure

Let consider $\Omega = \{1, 2\}^{\mathbb{N}}$, pick two positive numbers p and q such that $p + q = 1$. Consider the measure $\mathbb{P}$ on $\{1, 2\}$ defined by

$$\mathbb{P}(\{1\}) = p, \ \mathbb{P}(\{2\}) = q.$$

Then, consider the measure $\mu := \otimes \mathbb{P}$ on the product space $\{1, 2\}^{\mathbb{N}}$. We remind that such μ is defined in such way that

$$\mu([x_0 \ldots x_{n-1}]) = p^{\# \text{ of 1's in the word } x} \ q^{\# \text{ of 2's in the word } x},$$

where x is the finite word $x_0 \ldots x_{n-1}$.

We claim that such μ is an invariant measure. Indeed, we have for any cylinder $[x_0 \ldots x_{n-1}]$,

$$\sigma^{-1}([x_0 \ldots x_n]) = [1x_0 \ldots x_n] \sqcup [2x_0 \ldots x_n].$$

Then,

$$\mu([1x_0 \ldots x_n]) = p\mu([1x_0 \ldots x_n])$$

and

$$\mu([2x_0 \ldots x_n]) = q\mu([1x_0 \ldots x_n]).$$

In this way $\mu(\sigma^{-1}[x_0 \ldots x_n]) = \mu([x_0 \ldots x_n])$.

This example corresponds to the model of tossing a coin (head identified with 1 and tail identified with 2) in an independent way a certain number of times. We are assuming that each time we toss the coin the probability of head is p and the probability of tail is q.

Therefore, $\mu([211])$ describes the probability of getting tail in the first time and head in the two subsequent times we toss the coin, when we toss the coin three times.

Remark 1.13. The previous example shows that there are uncountably many σ-invariant probabilities on $\{1, 2\}^{\mathbb{N}}$. ∎

Markov chain

Let us start with some example. Again, we consider the case $\Omega = \{1,2\}^{\mathbb{N}}$. Pick p and q two positive numbers in $]0,1[$, and set

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} = \begin{pmatrix} P(1,1) & P(1,2) \\ P(2,1) & P(2,2) \end{pmatrix}.$$

The first writing of P shows that 1 is an eigenvalue. If we solve the equation

$$(x,y).P = (x,y),$$

we find a one-dimensional eigenspace (directed by a *left eigenvector*) with $y = \dfrac{1-p}{1-q}x$. Therefore, there exists a unique left eigenvector (π_1, π_2) such that

$$(\pi_1, \pi_2).P = (\pi_1, \pi_2) \text{ and } \pi_1 + \pi_2 = 1.$$

Note that π_1 and π_2 are both positive.

The measure μ is then defined by

$$\mu([x_0 \ldots x_n]) = \pi_{x_0} P(x_0, x_1) P(x_1, x_2) \ldots P(x_{n-2}, x_{n-1}).$$

A simple way to see the measure μ is the following: a word $\omega = \omega_0 \ldots \omega_{n-1}$ has to be seen as a path of length n, starting at state $\omega_0 \in \{1,2\}$ and finishing at state ω_{n-1}. The measure $\mu([\omega])$ is the probability of this space among all the paths of length n. This probability is then given by the initial probability of being in state ω_0 (given by π_{ω_0}) and then probabilities of transitions from the state ω_j to ω_{j+1} (equal to $P(\omega_j, \omega_{j+1})$), these events being independent.

A probability of this form is called the *Markov measure obtained from the 2×2 line stochastic matrix P and the initial vector of probability π*.

Exercise 7
Show that the Bernoulli measure constructed above is also a Markov measure.

More generally we have:

Definition 1.14. A $d \times d$ matrix P such that all entries are positive and the sum of the elements in each line is equal to 1 is called a **line stochastic matrix**.

One can show[2] that for a line stochastic matrix (with all entries strictly positive) there exist only one vector $\pi = (\pi_1, \pi_2, ..., \pi_d)$, such that all $\pi_j > 0$, $j \in \{1, 2, .., d\}$, $\sum_{j=1}^{d} \pi_j = 1$, and

$$\pi = \pi P.$$

π is called the **left invariant probability vector for the Markov Chain defined by** P.

Definition 1.15 (and proposition)**.** Given a $d \times d$ line stochastic matrix P, and its left invariant probability vector $\pi = (\pi_1, \pi_2, ..., \pi_d)$, we define, μ on $\Omega = \{1, 2, \ldots, d\}^{\mathbb{N}}$, in the following way: for any cylinder $[x_0 x_1 \ldots x_k]$

$$\mu([x_0, x_1 \cdots x_k]) = \pi_{x_0} P(x_0, x_1) P(x_1, x_2) P(x_2, x_3) \ldots P(x_{k-1}, x_k).$$

This measure μ **is invariant for the shift** and it is called the *Markov measure associated to P and* π. For a fixed d we denote the set of Markov measures over $\Omega = \{1, 2, .., d\}^{\mathbb{N}}$ by $\overline{\mathcal{G}}$

Note that for a Markov probability any cylinder set has positive measure

1.2.2   General results on invariant measures

The definition of invariant measure involves Borel sets. The next result gives another characterization of invariant measures (see [127]):

Proposition 1.16. *The measure* μ *is invariant, if and only if, for any continuous function* f

$$\int f(x) \, d\mu(x) = \int f(\sigma(x)) \, d\mu(x).$$

---

[2]Actually we will give a proof of that result in Theorem 2.14.

It is easy to see from the above that in the case $\mu_n \xrightarrow{\text{w*}} \mu$, and each μ_n is invariant, then, μ is invariant.

We denote by $\mathcal{M}_\sigma$ the set of invariant probabilities on Ω. It is a closed subset of probabilities for the weak*-topology, hence it is compact and convex.

We say that a point z in a convex compact set C is extremal, if it is not a trivial convex combination of other elements in the set C. That is, we can not write z as $z = \lambda x + (1 - \lambda)y$, where $0 < \lambda < 1$, and x, y are in the convex set C.

Definition 1.17. An extremal measure in $\mathcal{M}_\sigma$ is called ergodic.

This definition is however not useful and clearly not easy to check. The next proposition gives other criteria for a measure to be ergodic.

Proposition 1.18. *A probability μ is ergodic, if and only if, it satisfies one of the following properties:*

1. *Any invariant Borel set has full measure or zero measure.*

2. *For any continuous $f : \Omega \to \mathbb{R}$, if $f = f \circ \sigma$ μ-a.e., then, f is constant.*

Exercise 8
Show that a Markov measure is ergodic.

We can now state the main theorem in Ergodic Theory:

Theorem 1.19. (Birkhoff Ergodic Theorem) *Let μ be σ-invariant and ergodic. Then, for every continuous function $f : \Omega \to \mathbb{R}$ there exist a Borel set K, such that $\mu(K) = 1$, and for every x in K*

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} f(\sigma^{k-1}(x)) = \int f \, d\mu.$$

The Birkhoff Theorem says that, under the assumption of ergodicity, a time average is equal to a spatial average. Here is an example of application:

"Average cost of car ownership rises to \$8,946 per year."

What does the term average mean ? One can imagine that we count how much a certain person spends every year for its car, and then do the average cost. This is a time-average. The main problem of this average is to know if it representative of the cost of anybody.

On the contrary, one can pick some region, then count how many people spend in 1 year for their car, and take the average amount. This is a spatial average. The main problem is to know if it represents how much each person is going to spend along the years (at beginning the car is new, and then get older !).

The ergodic assumption means that the repartition of old and new cars in the space is "well" distributed and/or that the chosen person in the first way to compute the average cost is "typical". Then, the Birkhoff theorem says that both averages are equal.

Notation. We set $S_n(f)(x) := f(x) + \ldots + f \circ \sigma^{n-1}(x)$.

Example. In a previous example we considered a Bernoulli measure modeling the tossing of a coin with probabilities p of head and q of tail. Consider the indicator function $\mathbb{1}_{[2]}$ of the cylinder $[2]$. For a fixed n and for a fixed $x = x_0x_1x_2\ldots$ the value $\sum_{k=1}^{n} \mathbb{1}_{[2]}(\sigma^{k-1}(x))$, counts the number of times we get tail (or, the value 2) in the finite string $x_0x_1x_2\ldots x_{n-1}$.

Note that $\int \mathbb{1}_{[2]} \, d\mu$ by definition is equal to $\mu([2]) = q$. One can show that the measure μ we get is ergodic.

Therefore, from Birkhoff Theorem, we can say that for μ-almost every x, we have that

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{[2]}(\sigma^{k-1}(x)) = \int \mathbb{1}_{[2]} \, d\mu = q.$$

The value $\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{[2]}(\sigma^{k-1}(x))$ is the empirical mean value of number of times we get tail if the sequence of events is obtained from flipping the coin n times, which is described by $x_0x_1x_2\ldots x_{n-1}$, where $x = (x_0, x_1, x_2, \ldots)$.

We finish this section with another application of ergodicity:

Proposition 1.20. *Let μ be an invariant ergodic probability. Let x be a "generic" point in $\operatorname{supp}\mu$. Then, x returns infinitely many times, as closed as wanted, to itself.*

Proof. Pick $\varepsilon > 0$, and consider the ball $B(x,\varepsilon)$. It is a clopen set, hence $\mathbb{1}_{B(x,\varepsilon)}$ is continuous. The point x is generic for μ, then $\lim_{n\to\infty} \frac{1}{n} S_n(\mathbb{1}_{B(x,\varepsilon)})(x) = \mu(B(x,\varepsilon))$, and this last term is positive. Therefore, there are infinitely many n such that $\mathbb{1}_{B(x,\varepsilon)}(\sigma^n(x)) = 1$. $\qquad\square$

Remark 1.21. Actually, one can get a stronger result than the one claimed by the above Birkhoff Theorem: for every μ ergodic, there exists a set G_μ of full μ-measure, such that for every $x \in G_\mu$ and for every continuous function f,

$$\lim_{n\to\infty} \frac{1}{n} S_n(f)(x) = \int f\, d\mu.$$

A very important property is that for two distinct ergodic probabilities μ and ν, $G_\mu \cap G_\nu = \emptyset$ (see [127]). ∎

1.3 Ergodic optimization and temperature zero

The set $\mathcal{M}_\sigma$ of invariant measures is quite large. It is thus natural to ask about measures with special properties. In this direction, a well known class is the one which can be obtained from *Thermodynamic Formalism*. The first results on this topic are from the 70's. We will briefly describe some basic results on this setting below and we will present some more details in Chapter 2. Anyway, our goal here is to focus on *Ergodic Optimization*.

Definition 1.22. Let $A : \Omega \to \mathbb{R}$ be a continuous function. An invariant measure μ is said to be A-maximizing if

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

Note that this maximum is well defined because A is continuous and $\mathcal{M}_\sigma$ is compact for the weak*-topology.

Examples

• Consider $\Omega = \{1,2\}^{\mathbb{N}}$ and $A = \dfrac{1}{2}\mathbb{I}_{[1,2]} + \dfrac{1}{2}\mathbb{I}_{[2,1]}$. In this case the maximizing probability is unique and has support in the periodic orbit of period two $12121212\ldots$. This maximizing measure is $\mu := \dfrac{1}{2}\delta_{(12)^\infty} + \dfrac{1}{2}\delta_{(21)^\infty\ldots}$.

• There can be several maximizing measures. Consider $A(x) = -\min(d(x,1^\infty), d(x,2^\infty)) \leqslant 0$. The two measures $\mu_1 := \delta_{1^\infty}$ and $\mu_{2^\infty} := \delta_{2^\infty}$ are A-maximizing. Consequently, any convex combination

$$\mu_t := t\mu_{1^\infty} + (1-t)\mu_{2^\infty}, \ t \in [0,1],$$

is also A-maximizing. It can be shown, in this particular case, that any maximizing probability is of this form. Note that μ_t is not ergodic if $t \neq 0, 1$.

• Suppose $\Omega = \{1,2,3,4\}^{\mathbb{N}}$ and consider $A = \mathbb{I}_{[1]\cup[2]}$. any measure with support in $\{1,2\}^{\mathbb{N}}$ is A-maximizing. This example shows that there can be uncountably many ergodic maximizing measures.

Chapters 3 and 4 are devoted to more general results and tools for the study of maximizing measures. Chapter 7 is devoted to the study of an explicit example.

As we mentioned above, another important class of invariant measures appears in Thermodynamic Formalism. This will be the topic of Chapter 2.

Consider a fixed function $A : \Omega \to \mathbb{R}$, which is called a potential. Without entering too much into the theory, for a real parameter β we shall associate to $\beta.A$ a functional $\mathcal{P}(\beta)$, and for each β some measure μ_β called *equilibrium state for the potential $\beta.A$*. In Statistical Mechanics, β represents the inverse of the temperature. Then, $\beta \to +\infty$ means that the temperature goes to 0.

We want to point out the relations between the two different ways to singularize measures:

1. $\beta \mapsto \mathcal{P}(\beta)$ is convex and admits an asymptote for $\beta \to +\infty$. The slope is given by $m(A) = \max \int A \, d\mu$.

2. Any accumulation point for μ_β, as $\beta \to +\infty$, is A-maximizing. Then, the main question is to know if there is *convergence*, and, in the affirmative case, how does μ_β select the limit ?

These are the mains points which we consider in Chapter 3 (general results) and Chapter 7 (an specific example).

Chapter 2

Thermodynamic Formalism and Equilibrium states

2.1 General results

2.1.1 Motivation and definitions

A dynamical system, in general, admits a large number of different ergodic measures. A natural problem would be to find a way to singularize some special ones among the others.

Given $A : \Omega \to \mathbb{R}$, the thermodynamic formalism aims to singularize measures via the following principle:

$$\mathcal{P}(A) := \sup_{\mu \in \mathcal{M}_\sigma} \left\{ h_\mu + \int A \, d\mu \right\}. \qquad (2.1)$$

The quantity h_μ is called the Kolmogorov entropy. It is a non-negative real number, bounded by $\log d$ if $\Omega = \{1, \ldots d\}^{\mathbb{N}}$. Roughly speaking, it measures the chaos seen by the measure μ. It is a concept defined for invariant probabilities. If some measure has positive and large entropy, this means that the system ruled by this measure

is very complex. If the invariant measure has support in a periodic orbit its entropy is zero.

Definition 2.1. Any measure which realizes the maximum in (2.1) is called an *equilibrium state* for A. The function A is called the *potential* and $\mathcal{P}(A)$ is *the pressure* of the potential.

For a given σ-invariant measure μ, the quantity $h_\mu + \int A \, d\mu$ is called the *free-energy* of the measure (with respect to the potential A).

This study is inspired by Statistical Mechanics: the set $\Omega := \{1, \ldots k\}^{\mathbb{N}}$ is the one-dimensional lattice with k possible values for the spin in each site. The potential A measures the interaction between sites. At a macroscopic level, the lattice has a stable behavior, at the microcosmic scale, due to internal agitation, the configuration of the material changes. Therefore, a *state of the system* is a probability on Ω, the set of possible configurations; it is obtained by a principle of the kind of the law of large numbers. The equilibrium at the macroscopic scale is exactly given by the states which maximize the free energy.

In Statistical Mechanics people usually consider the influence of the temperature T by introducing a term $\beta := \frac{1}{T}$ and studying the equilibrium for $\beta.A$. For some fixed $A : \Omega \to \mathbb{R}$ and $\beta \in \mathbb{R}$, we shall write $\mathcal{P}(\beta)$ instead of $\mathcal{P}(\beta.A)$. It is an easy exercise to check that $\beta \mapsto \mathcal{P}(\beta)$ is convex and continuous. Therefore, the slope is non-decreasing and a simple computation shows that it converges to $m(A) := \sup_\mu \int A \, d\mu$ as $\beta \to \infty$.

It seems clear that a motivation for people who introduced the thermodynamic formalism theory into the Dynamical systems (Bowen, Ruelle and Sinai) was to study the functional $\mathcal{P}(\beta)$, $\beta \in \mathbb{R}$. Given that, one natural question is to understand the limit of $\mathcal{P}(\beta)$, when β goes to $+\infty$, that is the zero temperature case. This is one of the purposes of this text.

2.1.2 Entropy and existence of equilibrium states

We refer the reader to [29][124] [11] [87] [97] for the some of the results we use from Thermodynamic Formalism. We present here some results which are related to optimization in ergodic theory.

We emphasize that the results we present here are stated for Ω but they holds for any general irreducible subshift of finite type.

Entropy

For general results on entropy see also [13] and [110]. The complete description of entropy is somehow complicated and not relevant for the purpose of this course. A simple definition for the general case could be the following:

Theorem 2.2 (and definition). *Let μ be a σ-invariant ergodic probability. Then, for μ-a.e. $x = x_0 x_1 x_2 \ldots$,*

$$\lim_{n \to \infty} -\frac{1}{n} \log \mu([x_0 x_1 \ldots x_n])$$

exists and is independent of x. It is equal to h_μ.

If μ_0 and μ_1 are both invariant and ergodic probabilities, for every $\alpha \in [0,1]$ set $\mu_\alpha := \alpha.\mu_1 + (1-\alpha).\mu_0$. Then,

$$h_{\mu_\alpha} = \alpha h_{\mu_1} + (1-\alpha) h_{\mu_0}.$$

Roughly speaking, for μ ergodic and for μ-typical $x = x_0 x_1 \ldots$, h_μ is the exponential growth for $\mu([x_0 \ldots x_{n-1}])$, with $n \to \infty$.

Examples

• Let us consider $\Omega = \{1,2\}^{\mathbb{N}}$ and μ the Bernoulli measure given by the line stochastic matrix $P = \begin{pmatrix} p & q \\ p & q \end{pmatrix}$. We have seen before that

$$\mu([x_0 \ldots x_{n-1}]) = p^{\text{\# of 1's in the word}} q^{\text{\# of 2's in the word}}.$$

We also have seen before that $\frac{1}{n} \text{\# of 1's in the word} \to_{n \to \infty} \mu([1]) = p$, and, $\frac{1}{n} \text{\# of 2's in the word} \to_{n \to \infty} \mu([2]) = q$. Therefore,

$$h_\mu = -p \log p - q \log q.$$

• More generally, if μ is the Markov measure associated to a line stochastic matrix P, and, $\pi = (\pi_1, \pi_2, , , , \pi_d)$ is the stationary vector, then,

$$h(\mu) = -\sum_{i,j=1}^{n} \pi_i \ P(i,j) \ \log P(i,j).$$

Proposition 2.3. *The (metric) entropy is upper-semicontinuous :
If μ_n converges to μ in the weak*-topology, then*

$$h_\mu \geqslant \limsup_{n \to \infty} h_{\mu_n}.$$

Existence of equilibrium states

This is an immediate consequence of Proposition 2.3.

Theorem 2.4. *If A is continuous, then there exists at least one equilibrium state for A.*

Proof. The function $\mu \to h_\mu + \int A \, d\mu$ is upper-semicontinuous in the compact set of the invariant probabilities. Therefore attains the maximal value in at least one invariant probability. $\square$

At the end of this chapter we will study the special case where A depend just on the two first coordinates. In particular, we will be able to exhibit an explicit expression for the corresponding equilibrium state.

2.2   Uniqueness of the Equilibrium State

Uniqueness does not always hold. Nevertheless a key result is the following:

Theorem 2.5. *If $A : \Omega \to \mathbb{R}$ is Hölder continuous, then there is a unique equilibrium state for A. Moreover, it is a* Gibbs measure *and $\beta \to \mathcal{P}(\beta)$ is analytic.*

The concept of Gibbs measure will be explained below. We recall that $A : \Omega \to \mathbb{R}$ is said to be α-**Hölder**, $0 < \alpha < 1$, if there exists $C > 0$, such that, for all x, y we have $| \, A(x) - A(y) \, | \leq C \, d(x,y)^\alpha$.

For a fixed value α, we denote by $\mathcal{H}_\alpha$ the set of α-Hölder functions $A : \Omega \to \mathbb{R}$. $\mathcal{H}_\alpha$ is a vector space.

For a fixed α, the norm we consider in the set $\mathcal{H}_\alpha$ of α-Hölder potentials A is

$$||A||_\alpha = \sup_{x \neq y} \frac{|A(x) - A(y)|}{d(x,y)^\alpha} + \sup_{x \in \Omega} |A(x)|.$$

For a fixed α, the vector space $\mathcal{H}_\alpha$ is complete for the above norm.

We want to emphasize Theorem 2.5. If existence of equilibrium state is done via a general result of maximization of a semi-continuous function, uniqueness, by the other hand, is obtained for Hölder continuous potential via a completely different way: the key tool is an operator acting on continuous and Hölder continuous functions. Then, the pressure and the equilibrium state are related to the spectral properties of this operator which had its origin in Statistical Mechanics where is called the transfer operator.

The hypothesis of A being Hölder continuous means (in the Statistical Mechanics setting) that the interactions described by it decay very fast (in an exponential way) for spins which are located more and more distant in the lattice $\mathbb{N}$. This decay is not so fast for a potential which is continuous but not Hölder.

In the general case, for a fixed Hölder potential A, given $\beta_1 \neq \beta_2$, we have that $\mu_{\beta_1} \neq \mu_{\beta_2}$. In this case the generic sets of Remark 1.21 are disjoint. Moreover, for any β, the probability μ_β has support on the all set Ω.

Remark 2.6. The non-uniqueness of the equilibrium state for a potential A is associated to the phenomena of phase transition (see [69] [65] [108] [58] [16]). ∎

2.2.1 The Transfer operator

In this section we consider a fixed α-Hölder potential $A : \Omega \to \mathbb{R}$.

Definition 2.7. We denote by $\mathcal{L}_A : \mathcal{C}^0(\Omega) \to \mathcal{C}^0(\Omega)$ the Transfer operator corresponding to the potential A, which is given in the following way: for a given ϕ we will get another function $\mathcal{L}_A(\phi) = \varphi$,

such that,

$$\varphi(x) = \sum_{a \in \{1,2,...,d\}} e^{A(ax)} \phi(ax).$$

In another form

$$\varphi(x) = \varphi(x_0 x_1 \cdots) = \sum_{a \in \{1,2,...,d\}} e^{A(ax_0 x_1 x_2,...)} \phi(ax_0 x_1 x_2...).$$

The transfer operator is also called the Ruelle-Perron-Frobenius operator.

It is immediate to check that $\mathcal{L}_A$ acts on continuous functions. It also acts on α-Hölder functions if A is α-Hölder.

If μ is a probability, $f \mapsto \int \mathcal{L}_A(f) \, d\mu$ is a bounded linear form on $\mathcal{C}^0(\Omega)$. Therefore, by Riesz Theorem (see Theorem 1.10), there exists ν such that for every $f \in \mathcal{C}^0(\Omega)$, $\int f d\nu = \int \mathcal{L}_A(f) \, d\mu$. We set $\mathcal{L}_A^*(\mu) = \nu$, and we call $\mathcal{L}_A^* : \mu \mapsto \nu$ the dual operator of $\mathcal{L}_A$.

Note that if $\mathcal{L}_A(1) = 1$, then, $\mathcal{L}_A^*(\mu)$ is a probability, in the case μ is a probability.

Theorem 2.8 (see [123]). *Let λ_A be the spectral radius of $\mathcal{L}_A$. Then, λ_A is an eigenvalue for $\mathcal{L}_A^*$: there exists a probability measure ν_A such that*

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

This probability is called the eigenmeasure and/or the conformal measure.

We remind that the spectral radius is given by

$$\lambda_A := \limsup_{n \to \infty} \frac{1}{n} \log |||\mathcal{L}_A^n|||, \text{ and } |||\mathcal{L}_A^n||| = \sup_{||\psi||=1} ||\mathcal{L}^n(\psi)||_\infty.$$

Note that the Schauder-Tychonov Theorem shows that there exists an eigenmeasure. Indeed, consider the function acting on the convex and compact set of probabilities $\mu \to \frac{\mathcal{L}_A^*(\mu)}{\int \mathcal{L}_A(1) \, d\mu}$. It is however not clear that the associated eigenvalue is the spectral radius of $\mathcal{L}_A$.

The main ingredient to prove uniqueness of the equilibrium state is:

Theorem 2.9. *The operator $\mathcal{L}_A$ is quasi-compact on $\mathcal{H}_\alpha$: λ_A is simple isolated and the unique eigenvalue with maximal radius. The rest of the spectrum is contained in a disk $\mathbb{D}(0, \rho\lambda_A)$ with $0 < \rho < 1$.*

>From Theorem 2.9 we get a unique H_A, up to the normalization $\int H_A \, d\nu_A = 1$, such that

$$\mathcal{L}_A(H_A) = \lambda_A H_A.$$

Exercise 9
Show that the measure defined by $\mu_A = H_A \nu_A$ is σ-invariant.

It can be proved that this measure is actually a *Gibbs* measure: which means, there exists $C_A > 0$ such that for every $x = x_0 x_1 \ldots$ and for every n,

$$e^{-C_A} \leqslant \frac{\mu_A([x_0 \ldots x_{n-1}])}{e^{S_n(A)(x) - n\log\lambda_A}} \leqslant e^{C_A}. \tag{2.2}$$

Actually, these two inequalities yields that the free energy for μ_A is $\log\lambda_A$. Moreover, the left-side inequality yields that for any other ergodic measure ν,

$$h_\nu + \int A \, d\nu < \log\lambda_A.$$

In particular we get $\mathcal{P}(A) = \lambda_A$ and μ_A is the unique equilibrium state for A.

The same kind of results can be get for $\beta.A$ instead of A. Now, the spectral gap obtained in Theorem 2.9 and general results for perturbations of spectrum of operators yield that $\beta \mapsto \mathcal{P}(\beta)$ is locally analytic. A simple argument of connectness shows that it is globally analytic.

Remark 2.10. The corresponding theory when A is just continuous (not Hölder) is quite different (see [137]) ∎

We point out that different kinds of Transfer operators has been used in other areas and other settings: Differential and Complex Dynamics, Differential Geometry, Number Theory, Eigenvalues of the Laplacian, Zeta functions, C^*-Algebras, Computer Science, Quantum Computing, Economics, Optimization, etc... (see [124], [29], etc...)

2.2.2   Some more results

The theory described above can be generalized to other cases. In the special case of Ω (the full shift Bernoulli space) we can actually prove that λ_A is the unique dominating eigenvalue. Moreover, we get for every ψ Hölder continuous,

$$\mathcal{L}_A^n(\psi) = e^{n\mathcal{P}(A)} \int \psi \, d\nu_\beta.\phi + e^{n(\mathcal{P}(A)-\varepsilon)}\psi_n, \tag{2.3}$$

where ε is a positive real number (depending on A), ψ_n is continuous and $||\psi_n||_\infty \leqslant C.||\psi||_\infty$, for every n and C is a constant (depending on A). >From this one gets

$$\mathcal{P}(A) = \lim_{n\to\infty} \frac{1}{n} \log \mathcal{L}_A^n(\mathbb{1}),$$

which yields $\dfrac{d\mathcal{P}}{d\beta}(\beta) = \displaystyle\int A \, d\mu_\beta$.

We have already mentioned in the beginning of this chapter that for convexity reason, if β goes to $+\infty$, the slope of the graph of $\mathcal{P}(\beta)$ goes to $m(A) = \sup \displaystyle\int A \, d\mu$. We can now give a more precise result:

Proposition 2.11. *The graph of $\mathcal{P}(\beta)$ admits an asymptote if β goes to $+\infty$. The slope is given by $m(A) = \sup \displaystyle\int A \, d\mu$. Any accumulation point for μ_β is a A-maximizing measure.*

Proof. Convexity is a consequence of the definition of the pressure via a supremum. Let μ_∞ be any A-maximizing measure. Since the entropy is bounded we immediately get

$$m(A) \leqslant \frac{h_{\mu_\infty}}{\beta} + \int A \, d\mu_\infty \leqslant \frac{\mathcal{P}(\beta)}{\beta} \leqslant \frac{\log d}{\beta} + \int A \, d\mu_\beta \leqslant \frac{\log d}{\beta} + m(A).$$

This proves that the asymptotic slope for $\mathcal{P}(\beta)$ is $m(A)$:

$$\lim_{\beta\to+\infty} \frac{\mathcal{P}(\beta)}{\beta} = m(A). \tag{2.4}$$
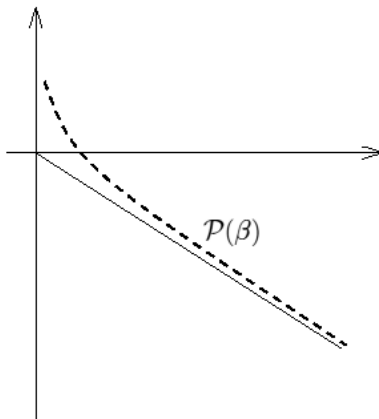
Figure 2.1:

Now, $\beta \mapsto \mathcal{P}(\beta) - \beta m(A)$ has non-positive derivative, and, it is thus a decreasing function. It is non-negative too (as the previous series of inequalities shows), and then it admits a non-negative limit as $\beta \to +\infty$.

We have mentioned above $\dfrac{d\mathcal{P}}{d\beta}(\beta) = \displaystyle\int A \, d\mu_\beta$. If we consider any accumulation point μ_∞ for μ_β, $\displaystyle\int A \, d\mu_\beta$ converges to the slope of the asymptote, that is, to $m(A)$, and then, μ_∞ is A-maximizing. $\qquad\square$

We finish this subsection with an important remark. We have seen the double-inequality (2.2)

$$e^{-C_A} \leqslant \frac{\mu_A([x_0 \ldots x_{n-1}])}{e^{S_n(A)(x) - n \log \lambda_A}} \leqslant e^{C_A}.$$

We emphasize here that C_A is proportional to $||A||_\infty$. Therefore, replacing A by $\beta.A$, and letting $\beta \to +\infty$, this implies that $C_{\beta.A} \to +\infty$. In other words, we will get that the constant explodes when $\beta \to +\infty,$.

Nevertheless we have:

Proposition 2.12. *Assume that A is α-Hölder. There exists a universal constant C such that $\dfrac{1}{\beta} \log H_{\beta.A}$ is α-Hölder with norm bounded by $C.||A||_\alpha$.*

Consequently, "at the scale " $\dfrac{1}{\beta}$ log we recover bounded quantities.
We will elaborate more on this point in the future.

2.2.3   An explicit computation for a particular case

Let us now assume that A depends on two coordinates, that is

$$A(x_0 x_1 x_2 \ldots) = A(x_0, x_1).$$

The exact form of the equilibrium state

Here we show how we get an explicit expression for the equilibrium
state using the Transfer Operator. Our computation does not prove
that the measure we construct is an equilibrium state, and neither
proves uniqueness of the equilibrium state. However, the example
gives a good intuition of the main issues on this class of problems.

We denote by $A(i, j)$ the value of A in the cylinder $[ij]$, $i, j \in \{1, 2, .., d\}$. In this case, the Transfer operator takes a simple form:

$$\mathcal{L}_A(\phi)(x_0 x_1 x_2 \ldots) = \sum_{a \in \{1,2,\ldots,d\}} e^{A(ax_0)} \, \phi(a x_0 x_1 x_2 \ldots).$$

Let M be the matrix will all positive entries given by $M_{i,j} = e^{A(i,j)}$.

Lemma 2.13. *The spectral radius of $\mathcal{L}_A$ is also the spectral radius of M.*

Proof. Assume that ϕ is a function depending only on one coordinate,
i.e.,

$$\phi(x_0 x_1 x_2 \ldots) = \phi(x_0).$$

Then, by abuse of notation, the function ϕ is described by the vector
$(\phi(1), \phi(2), ..., \phi(d))$. For every j

$$\mathcal{L}_A(\phi)(j) = \sum_{j=1}^{s} M_{ij}.\phi(i),$$

which can be written as the matrix M acting on the line vector ϕ, that is, $\mathcal{L}_A(\phi) = \phi M$. In an alternative way we can also write $\mathcal{L}_A(\phi) = M^*.\phi$, where M^* denotes de transpose of the matrix M. This yields that the spectral radius λ_M of M is lower or equal to λ_A.

We remind that the spectral radius is given by

$$\lambda_A := \limsup_{n \to \infty} \frac{1}{n} \log |||\mathcal{L}_A^n|||, \text{ and } |||\mathcal{L}_A^n||| = \sup_{||\psi||=1} ||\mathcal{L}^n(\psi)||_\infty.$$

The operator $\mathcal{L}_A$ is positive and this shows that for every n, $|||\mathcal{L}_A^n||| = ||\mathcal{L}_A^n(\mathbb{1})||_\infty$. Now, $\mathbb{1}$ depends only on 1 coordinate, which then shows $\mathcal{L}_A^n(\mathbb{1}) = M^n(\mathbb{1})$. This yields $\lambda_A \leqslant \lambda_M$. □

Theorem 2.14. (Perron-Frobenius) *Suppose $B = (b_{ij})$ is a $d \times d$ matrix with all entries strictly positive, $1 \leq i, j \leq d$. Then, there exists $\lambda > 0$, and, vectors $\mathbf{l} = (l_1, \cdots, l_d)$ and $\mathbf{r} = (r_1, \cdots, r_d)$ such that*

1. *for all i $l_i > 0$, and $r_i > 0$.*

2. *For all i $\sum_{j=1}^d b_{ij} r_j = \lambda r_i$ and for all j $\sum_{i=1}^d l_i b_{ij} = \lambda l_j$.*

(i.e., $\mathbf{r}$ is a right eigenvector of B and $\mathbf{l}$ is a left eigenvector for B).

Proof. We first show that there exists at least one vector $\mathbf{r}$ with all coordinates positive, and $\lambda \geq 0$, such that $\sum_{j=1}^d b_{ij} r_j = \lambda r_i$.

Consider the convex set $\mathcal{H}$ of vectors $h = (h_1, \cdots, h_d)$ such that $h_i \geq 0$, $1 \leq i \leq d$ and $\sum_{i=1}^d h_i = 1$. The matrix B determines a continuous transformation $G : \mathcal{H} \to \mathcal{H}$, given by $Gh = h'$, where

$$h_i' = \frac{\sum_{j=1}^d b_{ij} h_j}{\sum_{i=1}^d \sum_{j=1}^n b_{ij} h_j}$$

Note that h' has all entries strictly positive. In this way the image of $\mathcal{H}$ by G is strictly inside $\mathcal{H}$.

The Brouwer fixed point (see cap VII [113]) assures that there exists at least on fixed point.

As the image of $\mathcal{H}$ by G is strictly inside $\mathcal{H}$ we have that the fixed point has all entries positive.

If $\mathbf{r}$ is such fixed point, that is, $G(\mathbf{r}) = \mathbf{r}$, which means,

$$r_i = \frac{\sum_{j=1}^d b_{ij} r_j}{\sum_{i=1}^n \sum_{j=1}^d b_{ij} r_j}$$

Taking $\lambda = \sum_{i=1}^d \sum_{j=1}^d b_{ij} r_j$, we get the right eigenvector.

Considering B^* (the transpose of B) instead of B, we get another eigenvector $\mathbf{l}$. Note that $\langle \mathbf{r}, \mathbf{l} \rangle$ is a positive number. Then,

$$\lambda \langle \mathbf{r}, \mathbf{l} \rangle = \langle B\mathbf{r}, \mathbf{l} \rangle = \langle \mathbf{r}, B^* \mathbf{l} \rangle = \lambda^* \langle \mathbf{r}, \mathbf{l} \rangle$$

and $\lambda = \lambda^*$. □

Remark 2.15. 1. Actually, it is possible to prove more. The eigenvalue λ is the spectral radius of the matrix B and $\mathbf{r}$ and $\mathbf{l}$ are simple eigenvectors.

2. We have seen that in the case A depends only on 2 coordinates, the Transfer operator acts as a matrix. In the general case, Theorem 2.9 extends Theorem 2.14.

3. A line stochastic matrix (see Subsec. 1.2.1) has spectral radius equal to 1 and satisfies hypotheses of Theorem 2.14. It thus admits a left-eigenvector associated to the eigenvalue 1.

■

Let us consider the matrix $d \times d$ matrix M such that for each entry $M_{i,j} = e^{A(i,j)}$. By Theorem 2.14, there exist left and right eigenvectors $\mathbf{l} = (l_1, \ldots l_d)$ and $\mathbf{r} = (r_1, \ldots r_d)$, both associated to the spectral radius of $\mathcal{L}_A$ (due to Lemma 2.13), say λ_A.

Let us define the 2×2 matrix $P_A = P_A(i, j)$ with

$$P_A(i, j) = \frac{e^{A(i,j)} r_j}{\lambda_A r_i}. \tag{2.5}$$

Note that P_A is a line stochastic matrix:

$$\sum_j P_A(i, j) = \frac{\sum_j e^{A(i,j)} r_j}{\lambda_A r_i} = 1.$$

Let us fix the normalization in the following way: $\sum_j r_j = 1$ and $\sum_i l_i r_i = 1$. Then, the vector $\pi = (\pi_1, \ldots \pi_d)$ defined by $\pi_i = l_i r_i$ satisfies

$$\pi.P_A = \pi.$$

That is, such π is the stationary vector for P_A.

The right-eigenvector $\mathbf{r}$ has to be seen as the eigenmeasure ν_A from Theorem 2.8, and actually $r_j = \nu_A([j])$. The normalized left-eigenvector l has to be seen as the eigenfunction H_A from Theorem 2.9, with normalization $\int H_A \, d\nu_A = 1$.

The associated invariant Markov measure μ_A is defined by

$$\mu_A([x_0 \ldots x_{n-1}]) = \pi_{x_0} P_A(x_0, x_1) \ldots P_A(x_{n-2}, x_{n-1}).$$

The exact computation yields

$$\mu_A([x_0 \ldots x_{n-1}]) = \phi_A(x_0) e^{S_n(A)(x) - n \log \lambda_A} . \nu_A([x_{n-1}]).$$

Since ν and ϕ_A have positive entries, this shows that μ_A is a Gibbs measure.

Things can be summarized as follows:

Let $M = (M_{ij})$ be the matrix with entries $e^{A(i,j)}$. Let $\mathbf{r} = (r_1, \ldots r_d)$ be the right-eigenvector associated to λ with normalization $\sum r_i = 1$. Let $\mathbf{l} = (l_1, \ldots, l_d)$ be the left-eigenvector for λ with renormalization $\sum l_i r_i = 1$. Then, $\mathbf{r}$ gives the eigenmeasure ν_A, and $\mathbf{l}$ gives the density H_A.
The Gibbs measure of the cylinder $[x_0 \ldots x_{n-1}]$ is $\mu_A([x_0 \ldots x_{n-1}]) = l_{x_0} e^{S_n(A)(x) - n \log \lambda} r_{x_{n-1}}$

Remark 2.16. Note that in the case $A \equiv 0$, the measure μ_A is the Markov probability associated to the line stochastic matrix with all entries equal to $\frac{1}{d}$. We denote by the probability μ_{top} the measure of maximal entropy. It is the unique invariant probability with entropy $\log d$. ∎

About uniqueness

This subsection can be avoid in a first lecture. The reasoning bellow can be seen as the description of a simple computable way to get an equilibrium state, and, moreover, to show that it is the unique one.

First, we remind

$$\mu_A([i_0 \ldots i_{n-1}]) = l_{i_0} e^{S_n(A)(x) - n \log \lambda_A} r_{i_{n-1}}.$$

Taking the logarithm of this equality and dividing by n, we get

$$\frac{1}{n} \log \mu_A([x_0 \ldots x_{n-1}]) = \frac{1}{n} S_n(A)(x) - \log \lambda_A + \frac{1}{n}(\log l_{x_0} + \log r_{x_{n-1}}).$$

If $n \to +\infty$, the left hand side term goes to $-h_{\mu_A}$ (for μ generic x). The first summand in the right hand side goes to $\int A \, d\mu_A$, by Birkhoff theorem. The last summand goes to zero because l_{x_0} is a constant and $\mathbf{r}$ has all its entries bound away from 0. This shows that

$$h_{\mu_A} + \int A \, d\mu_A = \log \lambda_A.$$

We will now show that for every Markov measure μ

$$h(\mu) + \int A d\mu \leq \lambda(A). \tag{2.6}$$

This will show that the probability μ_A realizes the supremum of the free energy among the Markov measures (see [134]). This is a restrictive result (with respect to the initial problem of maximization of the free energy among "all" invariant probabilities) but this gives a hint of the main issues and ideas we are interested here. Actually, the key point is the relative entropy of two measures (see [89]).
We need the next lemma whose proof can be obtained in [127].

Lemma 2.17. *If $p_1, \cdots, p_d$ and $q_1, \cdots, q_d$, satisfy[1]*

$$\sum_{j=1}^{d} p_i = \sum_{j=1}^{d} q_j = 1,$$

with $p_i \geq 0$, $q_i \geq 0$, $i = 1, \cdots, d$, then,

$$\sum_{i=1}^{d} q_i \log q_i - \sum_{i=1}^{d} q_i \log p_i = \sum_{i=1}^{d} q_i \log \frac{q_i}{p_i} \geq 0,$$

and the equality holds, if and only if, $p_i = q_i$, $i = 1 \cdots, d$.

---

[1]By convention $0 \log 0 = 0$.

Consider μ a Markov measure. We fix the vector $\pi = (\pi_1, ..., \pi_d)$, and the matrix P, satisfying $\pi P = \pi$ (which defines this Markov measure μ).

Equation (2.5) yields $A(i,j) = \log P_A(i,j) + \log \lambda_A + \log r_i - \log r_j$. By abuse of notation we set $r_i = r(x)$ if $x \in [i]$. Then, we get

$$
\begin{aligned}
h_\mu + \int A \, d\mu & = -\sum_i \pi_i P_{ij} \log P_{ij} + \int \log P_A(x_0, x_1) \\
& \quad + \log \lambda_A + \int \log r - \log r \circ \sigma \, d\mu \\
& = -\sum_{i,j} \pi_i P_{ij} \log P_{ij} + \sum_{i,j} \pi_i P_{ij} \log P_A(i,j) + \log \lambda_A \\
& = -\sum_i \pi_i \sum_j P_{ij} \log P_{ij} - P_{ij} \log P_A(i,j) + \log \lambda_A \\
& \leqslant \log \lambda_A,
\end{aligned}
$$

by Lemma 2.17. Then, (2.6) holds.

Chapter 3

Maximizing measures and Ground states

3.1 Selection at temperature zero

3.1.1 The main questions

We remind Definition 1.22 which address to the concept of A-maximizing measure: it is a σ-invariant probability μ such that

$$\int A \, d\mu = \max_{\nu \in \mathcal{M}_\sigma} \int A \, d\nu.$$

Existence of maximizing probabilities follows from the compactness of $\mathcal{M}_\sigma$ (and the continuity of the map $\nu \mapsto \int A d\nu$).

The first kind of problems we are interested in are related to maximizing measures. We can for instance address the questions:

1. For a given potential A, how large is the set of maximizing measures ?

2. How can we construct/get maximizing measures ?

3. For a given maximizing measure, what we can say about its support ?

We have already mentioned above the relation between maximizing measures and equilibrium states (see Proposition 2.11). Assume that A is Hölder continuous, then, any accumulation point for μ_β is a A-maximizing measure. We point out that, for simplicity, from now on, we replace all the subscribed $\beta.A$ by β.

The relation between equilibrium states and maximizing measures motivates a new definition:

Definition 3.1. Let A be Hölder continuous. A σ-invariant probability measure μ is called a *ground state* (for A) if it is an accumulation point for μ_β, as β goes to $+\infty$.

Remark 3.2. The Hölder continuity of A is only required to get uniqueness for the equilibrium state for $\beta.A$. ∎

After last chapter, clearly a ground state is a A-maximizing measure, but *a priori* a maximizing measure is not necessarily a ground state. The natural questions we can address are then:

1. Is there convergence of μ_β, as β goes to $+\infty$?

2. If there is convergence, how does the family of measures μ_β select the limit ?

If the limit for μ_β, as β goes to $+\infty$, exists, we say that $\mu = \lim_{\beta \to +\infty} \mu_\beta$ is **selected when temperature goes to zero**.

We remind that if $\mu_{\infty,1}$ and $\mu_{\infty,2}$ are two invariant maximizing probabilities, then any convex combinations of these two measures is also a maximizing measure. This gives a hint of how difficult the problem of selection is. There are too many possibilities.

In case of convergence, one natural question is then to study the speed of convergence for μ_β. More precisely, if C is a cylinder such that $C \cap \mathcal{A} = \emptyset$, then $\lim_{\beta \to +\infty} \mu_\beta(C) = 0$ (because any possible accumulation point gives 0-measure to C). It is thus natural to study the possible limit:

$$\lim_{\beta \to +\infty} \frac{1}{\beta} \log(\mu_\beta(C)).$$

We will see that this question is related to the behavior of $\frac{1}{\beta} \log H_\beta$. Then we address the question:

1. Is there convergence of the family $\frac{1}{\beta} \log H_\beta$, as β goes to $+\infty$?

2. If there is convergence, how does the family of functions $\frac{1}{\beta} \log H_\beta$ select the limit ?

In the case $V = \lim_{\beta \to \infty} \frac{1}{\beta} \log H_\beta$ we say that V **was selected when temperature goes to zero**.

3.1.2 A naive computation. The role of subactions

We explore here a different but related question: consider $A : \Omega \to \mathbb{R}$ Lipschitz and $\beta > 0$. The transfer operator yields for every x,

$$e^{\mathcal{P}(\beta)} H_\beta(x) = \sum_{i=1}^{d} e^{\beta . A(ix)} H_\beta(ix), \qquad (3.1)$$

where all the subscribed $\beta.A$ have been replaced by β for simplicity. H_β denotes the main eigenfunction (properly normalized) for βA.

We have seen that, if β goes to $+\infty$, there is control on the Lipschitz (or, Hölder) constant, more precisely, this control exists at the $\frac{1}{\beta}$ log-scale.

Actually, Proposition 2.12 shows that the $\frac{1}{\beta} \log H_\beta$ form an equicontinuous family. There are different choices for normalizing the family of eigenfunction of the Ruelle operator in order to get that the family is bounded (independent of β), but we will not elaborate on this in the moment. We can consider an accumulation point of this sequence, say V, and we can show it is Lipschitz continuous. For simplicity we keep writing $\beta \to +\infty$, even if we actually consider subsequences. Then, taking $\frac{1}{\beta} \log$ of (3.1), considering $\beta \to +\infty$, equality (2.4) (and some more effort) implies

$$m(A) + V(x) = \max_i \{V(ix) + A(ix)\}. \qquad (3.2)$$

We will show in a moment a proof of this result in a particular case

Definition 3.3. A continuous function $u : \Omega \to \mathbb{R}$ is called a *calibrated subaction* for $A : \Omega \to \mathbb{R}$, if for any $x \in \Omega$, we have

$$u(x) = \max_{\sigma(y)=x} [A(y) + u(y) - m(A)]. \tag{3.3}$$

We will show the proof of a particular case of the general result we mentioned above.

Theorem 3.4. *Assume A is a potential which depends on two coordinates. Suppose β_n is such that for each j*

$$\lim_{\beta_n \to \infty} \frac{1}{\beta_n} \log l^{\beta_n A}(j) = V(j).$$

Then, V is a calibrated subaction.

To get this result we consider the eigenvalue equation: for each β and for each $1 \le j \le d$, we have

$$\sum_{i=1}^{d} \frac{l_i^{\beta A} e^{\beta A(i,j)}}{\lambda(\beta A) l_j^{\beta A}} = 1.$$

For each j and β_n, there exist a $i_n = i_n^j$, such that, $\frac{l_{i_n}^{\beta A} e^{\beta A(i_n,j)}}{\lambda(\beta A) l_j^{\beta A}}$ attains the maximal possible value among the $i \in \{1, 2., , .d\}$.

For a fixed j there exists a $i_j \in \{1, 2., , .d\}$ such that the value i_n^j is attained an infinite number of times.

Remember that $\lim_{\beta \to \infty} \frac{P(\beta A)}{\beta} = \lim_{\beta \to \infty} \frac{\log(\lambda(\beta A))}{\beta} = m(A)$.

Therefore,

$$0 = \frac{1}{\beta_n} \log\left(\sum_{i=1}^{d} \frac{l_i^{\beta_n A} e^{\beta_n A(i,j)}}{\lambda(\beta_n A) l_j^{\beta_n A}}\right) \le$$

$$\frac{1}{\beta_n} \log\left(d \, \frac{l_{i_j}^{\beta_n A} e^{\beta_n A(i_j,j)}}{\lambda(\beta_n A) l_j^{\beta_n A}}\right) \le$$

$$A(i_j, j) + \frac{1}{\beta_n} \log l^{\beta_n A}(i_j) - \frac{1}{\beta_n} \log l^{\beta_n A}(j) - \frac{1}{\beta_n} \log \lambda(\beta_n A).$$

Now, taking limit in the above expression, when $n \to \infty$, we get for each fixed j

$$0 \le A(i_j, j) + V(i_j) - V(j) - m(A).$$

In this way, for each j we get

$$V(j) + m(A) \le \max_{i \in \{1,2,\dots,d\}} \{A(i,j) + V(i)\}.$$

In the case, for a given j there exists i and $\epsilon > 0$, such that,

$$V(j) + m(A) + \epsilon < A(i,j) + V(i),$$

then,

$$1 < \frac{l_i^{\beta_n A} e^{\beta_n A(i,j)}}{\lambda(\beta_n A) l_j^{\beta_n A}},$$

for some large n. But, this is a contradiction.

$\square$

In the case the maximizing probability for A is unique, then, from [48] [59], there exists the limit

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log l^{\beta A} = V.$$

Consequently, any accumulation point for $\dfrac{1}{\beta} \log H_\beta$ is a calibrated subaction. Now, Equality (3.2) yields for every i and every x,

$$A(ix) \le m(A) + V(x) - V(ix),$$

which can be rewritten as $A(y) = m(A) + V \circ \sigma(y) - V(y) + g(y)$, where g is a non-positive function. By Proposition 2.12 it is also Lipschitz continuous.

Note that in the case u is a calibrated subaction, then, $u + c$, where c is a constant, is also a calibrated subation. We say that the calibrated subaction is unique, if it is unique up to an additive constant. One can show that in the case there is more than one maximizing probability, the calibrated subaction is not unique.

A calibrated subaction u satisfies

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

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

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

Suppose μ is maximizing for A and u is a calibrated subaction for A. It follows that for any x in the support of μ_∞ we have

$$u(\sigma(x)) - u(x) - A(x) + m(A) = 0. \tag{3.4}$$

Indeed, $g(x) = u(\sigma(x)) - u(x) - A(x) + m(A) \geq 0$, and the integral $\int g(x) d\mu(x) = 0$.

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. The equality to zero in the above equation can be true outside the union of the supports of the maximizing probabilities μ (see example of R. Leplaideur in [111]). It is known that generically on the Holder class on A the equality is true just on the support of the maximizing probability.

The study of the ergodic properties of maximizing probabilities is the purpose of Ergodic Optimization.

Now we recall some questions already stated in the introduction of this chapter:

1. Is there convergence for $\dfrac{1}{\beta} \log H_\beta$ as β goes to $+\infty$?

2. If there is convergence, how does the family of functions $\dfrac{1}{\beta} \log H_\beta$ select the limit ?

We also remind the reader that in the case $V = \lim_{\beta \to \infty} \dfrac{1}{\beta} \log H_\beta$ we say that V **was selected when temperature goes to zero**.

Given a general Hölder potential A can be exist calibrated subactions which are not selected. The ones that are selected are special among the calibrated subactions.

Example. For $\Omega = \{1,2\}^{\mathbb{N}}$ we study the convergence and the selection at temperature zero.

Suppose A is a two by two matrix.

If $A(1,1)$ is strictly bigger than the other $A(1,j)$ there is a unique maximizing probability with support on δ_{1^∞}. There is only one calibrated subaction up to an additive constant, therefore we get selection of subaction and probability. A similar result happen when $A(2,2)$ is strictly bigger than all the other $A(i,j)$.

Suppose now that $A(1,1) = 0 = A(2,2)$, and, moreover that $A(1,2), A(2,1) < 0$.

Using the notation of previous sections we have that $H_i^\beta = l_i^\beta$, $i = 1, 2$. The equation for the right eigenvalue r_i^β of A is the equation of the left eigenvalue for the symmetric matrix A^t.

Then, we get the system

$$(\lambda_\beta - 1)H_1^\beta = e^{\beta A(2,1)} H_2^\beta, \tag{3.5}$$

$$(\lambda_\beta - 1)H_2^\beta = e^{\beta A(1,2)} H_1^\beta \tag{3.6}$$

The trace of the matrix with entries $e^{\beta A(i,j)}$ is 2 and the determinant is $1 - e^{\beta (A(1,2)+A(2,1))}$. Solving a polynomial of degree two we get the maximal eigenvalue

$$\lambda_\beta = 1 + \sqrt{e^{\beta[A(1,2)+A(2,1)]}}.$$

We can take $H_2^\beta = 1$ and $H_1^\beta = e^{\beta \frac{1}{2}[A(2,1)-A(1,2)]}$.

In this case $V(2) = \lim_{\beta \to \infty} \frac{1}{\beta} \log H_2^\beta = 0$ and, moreover, we have $V(1) = \lim_{\beta \to \infty} \frac{1}{\beta} \log H_1^\beta = \frac{1}{2}[A(2,1) - A(1,2)]$.

In this case we have selection of subaction assuming the normalization $H_2^\beta = 1$ for all β.

An easy computation shows that in this case $l_1^\beta = r_2^\beta$ and $l_2^\beta = r_1^\beta$. We can assume the normalization $r_1^\beta + r_2^\beta = 1$ and $l_1^\beta r_1^\beta + l_2^\beta r_2^\beta = 1$.

Therefore, $\mu_\beta([1]) = l_1^\beta r_1^\beta = l_2^\beta r_2^\beta = \mu_\beta([2])$. In this case we have selection of the probability $\frac{1}{2}\delta_{1^\infty} + \frac{1}{2}\delta_{2^\infty}$ in the zero temperature limit.

Concerning the first question which concerns measures (question 1), there is one known example of non-convergence (see [46]), when

$\beta \to \infty$. Let us now gives some more details in the last question. Obviously, if there exists a unique maximizing measure, we have convergence because there is a unique possible accumulation point. In this case the question is not so interesting.

Let us then assume that there are at least two different maximizing measures $\mu_{max,1}$ and $\mu_{max,2}$. Just by linearity, any convex combination of both measure $\mu_t = t\mu_{max,1} + (1-t)\mu_{\max,2}$, $t \in [0,1]$, is also a A-maximizing measure. The question of selection is then to determine why the family (or, even a subfamily) choose to converge to some specific limit, if there are so many possible choices.

Among maximizing measures a special subset is the one whose elements are *ground states*, a notion (and a terminology) that comes from Statistical Mechanics, as we said before. One of the most famous examples is the so called spontaneous magnetization: consider a one-dimensional lattice with spins which can be up (say $+1$) or down (say -1). At high temperature there is a lot of randomness in the spins. But, if the temperature decreases, then, the behavior of the spins change in the lattice, and, at very low temperature, a given arrangement of spins in the lattice, is such that they are all up or they are all down. This procedure modifies the magnetic properties of the material. In Statistical Mechanics β is the inverse of the temperature, thus, $\beta \to +\infty$ means to approximate zero temperature. The main goal is thus to develop mathematical tools for understanding why materials have a strong tendency to be highly ordered at low temperature. They reach a crystal or quasi crystal configuration. This is also called the spontaneous magnetization when temperature decreases. An invariant probability with support in a union of periodic orbits plays the role of a "magnetic state".

A main conjecture in Ergodic Optimization claims that for a generic potential in the Holder topology the maximizing probability is a unique periodic orbit (for a partial result see [48]).

Note that Gibbs states for Hölder potentials give positive mass to any cylinder set. In this case, for any positive temperature, there is no magnetization. In the limit when temperature goes to zero the Gibbs state may, or may not, split in one or more ground states (which can have, or not, support in periodic orbits). All these questions are important issues here.

We point out that when the potential is not of Hölder class it is possible to get "Gibbs states" which have support in periodic orbits even at positive temperature.

Questions about selection of subactions naturally appears in this problem.

3.1.3 Large deviations for the family of Gibbs measures

In the study of Large Deviations when temperature goes to zero one is interested in the limits of the following form:

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta(C), \tag{3.7}$$

where C is a fixed cylinder set on Ω.

In principle, the limit may not exist. We remind that general references in Large deviations are [52] [53]).

Definition 3.5. We say there exists a Large Deviation Principle for the one parameter family μ_β, $\beta > 0$, if there exists a non-negative function I, where $I : \Sigma \to \mathbb{R} \cup \{\infty\}$ (which can have value equal to infinity in some points), which it is lower semi-continuous, and satisfies the property that for any cylinder set $C \subset \Sigma$,

$$\lim_{\beta \to +\infty} \frac{1}{\beta} \log \mu_{\beta\,A}(C) = - \inf_{x \in C} I(x).$$

In the affirmative case an important point is to be able to identify such function I.

Theorem 3.6 (see [12, 109]). *Assume that A admits a unique maximizing measure. Let V be a calibrated subaction. Then, for any cylinder $[i_0 i_1 ... i_n]$, we have*

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta([i_0 i_1 ... i_n]) = - \inf_{x \in [i_0 i_1 ... i_n]} \{I(x)\},$$

where

$$I(x) = \sum_{n=0}^{\infty} \left[V \circ \sigma - V - (A - m(A)) \right] \sigma^n (x).$$

In the proof of the above theorem in [12] it is used the involution kernel which will be briefly described in a future section. For a proof of this result without using the involution kernel see [109].

In the case the potential A depends on two coordinates and the maximizing probability is unique we get

$$I(x) = I(x_0, x_1..., x_k, ..) = \sum_{j=0}^{\infty} [V(x_{j+1}) - V(x_j) - A(x_j, x_{j+1}) + m(A)].$$

Example. Assume that the potential A depends only on two coordinates, say $A(x) = A(x_1, x_2)$.
We have seen that in that case

$$\mu_\beta = ([i_0 \ldots i_{n-1}]) = l_{i_0}^\beta e^{\beta.S_n(A)(x) - n \log \lambda_{\beta.A}} r_{i_{n-1}}^\beta,$$

where $\mathbf{l}^\beta = (l_1^\beta, \ldots, l_d^\beta)$ and $\mathbf{r}^\beta = (r_1^\beta, \ldots, r_d^\beta)$ are respectively the left eigenvector and the right eigenvector for the matrix with entries $e^{\beta.A(i,j)}$ and $\lambda_{\beta.A}$ is the associated eigenvalue (and the spectral radius of the matrix).
Studying the Large Deviation, we are naturally led to set

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta([i_0 i_1...i_{n-1}]) = [A(i_0, i_1) + ... + A(i_{n-2}, i_{n-1})] +$$

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log l_{(i_0)i_0}^\beta - \lim_{\beta \to \infty} \frac{1}{\beta} \log r_{i_{n-1}}^\beta - n.m(A). \qquad (3.8)$$

The main problem in the above discussion is the existence of the limits, when $\beta \to \infty$.
In the case there exists the limits

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log l^\beta = V$$

and

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log r^\beta = V^*,$$

we get the explicit expression

$$\lim_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta([i_0 i_1 ... i_{n-1}]) =$$

$$[A(i_0, i_1) + ... + A(i_{n-2}, i_{n-1})] + V(i_0) - V^*(i_{n-1}) - n.m(A). \quad (3.9)$$

There is a need for understanding better the relation of l^β and r^β. This is sometimes not so easy as we will see in an example which will be presented in the last chapter.

3.1.4   Uniqueness and lock-up on periodic orbits

One can ask questions about the behavior of typical trajectories of maximizing probabilities. From the classical result given by Birkhoff's we know that in the case μ is a maximizing measure, then, for μ-almost every x,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} A \circ \sigma^j(x) = m(A) = \int A d\mu.$$

In this case we say that this orbit beginning in x is A-maximizing. It is natural to analyze the problem under this viewpoint:

1. How can we detect that an certain given orbit is A-maximizing?

2. What is the relation of maximizing measures with maximizing orbits?

3. What are the main properties of the A-maximizing orbits?

We will see that this point of view will produce methods which will help to solve the cohomological inequality:

$$A \geqslant m(A) + V \circ T - V. \quad (3.10)$$

Definition 3.7. A coboundary is a function of the form $\psi \circ \sigma - \psi$.

Exercise 10
Show that a coboundary has zero integral for any σ−invariant measure.

We mentioned above the notion of crystals or quasi-crystals. In our settings crystals means periodic orbits; we will not deal with the question of quasi-crystal in this text. Consequently, an important point we want to emphasize here is the role of the periodic orbits. Nevertheless, the regularity of the potential is of prime importance in that study. The analogous results for the class of Continuous potentials and the class of Hölder Continuous potentials can be very different.

Theorem 3.8 (see [27]). *Generically for the $\mathcal{C}^0$-norm the potential A has a unique maximizing measure. This measure is not-supported in a periodic orbit.*

We shall give the proof of this theorem inspired from [27]. We emphasize that this proofs can be extended (concerning the uniqueness of the maximizing measure) to any separable space. A proof of generic uniqueness for Hölder continuous functions can be found in [48].

Proof. The set $\mathcal{C}^0(\Omega)$ is separable. Let (ψ_n) be a dense sequence in $\mathcal{C}^0(\Omega)$. Two different measures, say μ and ν, must give different values for some ψ_n. This means

$$
\begin{aligned}
\{A, \ \#\mathcal{M}_{max} > 1\} \ &= \ \left\{A, \ \exists n \ \exists \mu,\nu \in \mathcal{M}_{max} \ \int \psi_n \, d\nu \neq \int \psi_n \, d\mu\right\} \\
&= \ \bigcup_n \left\{A, \ \exists \mu,\nu \in \mathcal{M}_{max} \ \int \psi_n \, d\nu \neq \int \psi_n \, d\mu\right\} \\
&= \ \bigcup_n \bigcup_m \left\{A, \ \exists \mu,\nu \in \mathcal{M}_{max} \ \left|\int \psi_n \, d\nu - \int \psi_n \, d\mu\right| \geqslant \frac{1}{m}\right\}.
\end{aligned}
$$

Set $F_{n,m} := \left\{A, \ \exists \mu,\nu \in \mathcal{M}_{max} \ \left|\int \psi_n \, d\nu - \int \psi_n \, d\mu\right| \geqslant \frac{1}{m}\right\}$. We claim that theses sets are closed. For this we need a lemma:

Lemma 3.9. *Let (A_k) be a sequence of continuous potentials converging to A. Let μ_k be any maximizing measure for A_k and μ be an accumulation point for μ_k.*
Then, $\displaystyle\lim_{k \to +\infty} m(A_k) = m(A)$ and μ is a A-maximizing measure.

Proof of Lemma 3.9. For any $\varepsilon > 0$, and for k sufficiently large,

$$A - \varepsilon \leqslant A_k \leqslant A + \varepsilon.$$

This shows $m(A) - \varepsilon \leqslant m(A_k) \leqslant m(A) + \varepsilon$. Furthermore, we have $m(A_k) = \int A_k \, d\mu_k$, $m(A) = \int A \, d\mu$ and (up to a subsequence), $\lim_{k \to +\infty} \int A_k \, d\mu_k = \int A \, d\mu$, because μ_k converges to μ on the weak* topology and A_k goes to A on the strong topology.   □

We can thus show that the set $F_{n,m}$ is closed in $\mathcal{C}^0(\Omega)$. Indeed, considering a sequence A_k converging to A (for the strong topology), we get two sequences (μ_k) and (ν_k) of A_k-maximizing measures such that

$$\left| \int \psi_n \, d\mu_k - \int \psi_n \, d\nu_k \right| \geqslant \frac{1}{m}.$$

We pick a subsequence such that μ_k and ν_k converge for this subsequence. Lemma 3.9 shows that the two limits, say μ and ν, are A-maximizing and they satisfy

$$\left| \int \psi_n \, d\mu - \int \psi_n \, d\nu \right| \geqslant \frac{1}{m}.$$

To complete the proof concerning generic uniqueness, we need to prove that the sets $F_{n,m}$ have empty interior.

For such A, we claim that the function $\varepsilon \mapsto m(A + \varepsilon.\psi_n)$ is convex but not differentiable at $\varepsilon = 0$. Indeed, assume μ and ν are A-maximizing and

$$\int \psi_n \, d\mu \geqslant \int \psi_n \, d\nu + \frac{1}{m},$$

then the right derivative is bigger than $\int \psi_n \, d\mu$, and, the left derivative is lower than $\int \psi_n \, d\nu$.

It is known that a convex function is differentiable Lebesgue everywhere (actually everywhere except on a countable set), which proves that there are infinitely many ε accumulating on 0 such that $A + \varepsilon.\psi_n$ cannot be in $F_{n,m}$.

Let us now prove that, generically, the unique maximizing measure is not supported on a periodic orbit.

Let us consider some periodic orbit $\mathcal{O}$ and $\mu_{\mathcal{O}}$ the associated invariant measure. If A is such that $\mu_{\mathcal{O}}$ is not A-maximizing, then for every A_ε closed, $\mu_{\mathcal{O}}$ is still not A_ε-maximizing (see the proof of Lemma 3.9). This proves that the set of A such that $\mu_{\mathcal{O}}$ is A-maximizing is a closed set in $\mathcal{C}^0$.

In order to prove that it has empty interior, let consider A such that $\mu_{\mathcal{O}}$ is A-maximizing, and some measure μ closed to $\mu_{\mathcal{O}}$ in the weak* topology. This can be chosen in such way that $\int A \, d\mu \geqslant m(A) - \varepsilon$, with $\varepsilon > 0$ as small as wanted.

There exists a set K_ε such that $\mu_{\mathcal{O}}(K_\varepsilon) < \varepsilon$ and $\mu(K_\varepsilon) > 1 - \varepsilon$, and, then, we can find a continuous function $0 \leqslant \psi_\varepsilon \leqslant 1$, null on the periodic orbit $\mathcal{O}$, and, such that, $\int \psi_\varepsilon \, d\mu > 1 - 2\varepsilon$.

Then,

$$\int A + 2\varepsilon \psi_\varepsilon \, d\mu_{\mathcal{O}} = m(A) < m(A) - \varepsilon + 2\varepsilon - 4\varepsilon^2 \leqslant \int A + 2\varepsilon.\psi_\varepsilon \, d\mu_\varepsilon.$$

This shows that $\mu_{\mathcal{O}}$ is not $A + 2\varepsilon \psi_\varepsilon$-maximizing, thus the set of potentials such that $\mu_{\mathcal{O}}$ is maximizing has empty interior.

As there are only countably many periodic orbits, this proves that generically, a periodic orbit is not maximizing. $\qquad\square$

It is known that generically on the Holder class the maximizing probability for A is unique (see [48]).

Let us thus mention an important conjecture:

Conjecture. *Generically for the Lipschitz norm, the potential A has a unique maximizing measure and it is supported by a periodic orbit.*

The main results in that direction are:

Theorem 3.10 (see [138]). *Let μ be a maximizing measure for the Lipchitz potential A. Assume μ is not supported by a periodic orbit. Then, for any $\varepsilon > 0$, there exists A_ε ε-closed to A for the Lipschitz norm such that μ is not A_ε-maximizing.*

Theorem 3.11 (see [48])**.** *Let A be Lipschitz which has a unique A-maximizing measure μ. Assume that μ has periodic support. Then, there exists ε such that for every A_ε ε-closed to A (for the Lipschitz norm), μ is the unique A_ε-maximizing measure.*

Theorems 3.11 and 3.10 show that the unique possibility for a maximizing measure to keep being maximizing for small perturbation in the Lipschitz norm is that it is supported by a periodic orbit. This is the lock-up on periodic orbits. The extended version of Theorem 3.8 for the Lipschitz norm would imply that generically there is a unique maximizing measure. Nevertheless, this conjecture is not yet proved on the full extent. A partial result in this direction is [48].

Questions about ground states which have support on periodic orbits are questions about magnetization at temperature zero.

On the other hand, it is extremely simple to get an example with non-uniqueness for the maximizing measure. Let $\mathbb{K}$ be any σ-invariant compact set such that it contains the support of at least two different invariant measure (in other words $\mathbb{K}$ is not uniquely ergodic). Then set

$$A := -d(., \mathbb{K}).$$

Then, A is Lipschitz and any measure with support in the set $\mathbb{K}$ is A-maximizing. It is so simple to get examples where uniqueness fails that generic uniqueness cannot be seen as sufficient to consider the problem as solved.

The problem of selection is so fascinating, that, in our opinion, is interesting in itself.

3.1.5   First selection: entropy criterion

In this section we consider $A : \Omega \to \mathbb{R}$ a Lipschitz continuous potential. Note that in that specific case, results also hold for Hölder continuous potentials. We denote by $\mathcal{M}_{max}$ the set of maximizing measures.

Definition 3.12. The Mather set of A is the union of the support of all the A-maximizing measures.

Theorem 3.13. *Any ground state has maximal entropy among the set of maximizing measures. In other words, any accumulation point μ_∞ for μ_β satisfies*

$$h_{\mu_\infty} = \max\left\{h_\nu, \nu \in \mathcal{M}_{max}\right\}.$$

Proof. First, note that $\mathcal{M}_{max}$ is closed, thus compact in $\mathcal{M}_\sigma$. The entropy is upper-semi-continuous, then, there exists measures in $\mathcal{M}_{max}$ with maximal entropy.

Let μ_∞ be such a measure, and set $h_{max} = h_{\mu_\infty}$. Then, we get

$$h_{\max} + \beta.m(A) = h_{\mu_\infty} + \beta.\int A\,d\mu_\infty \leqslant \mathcal{P}(\beta). \qquad (3.11)$$

We remind that $\beta \mapsto \mathcal{P}(\beta)$ admits an asymptote as $\beta \to +\infty$, which means that there exists some h such that

$$\mathcal{P}(\beta) = h + \beta.m(A) + o(\beta),$$

with $\lim\limits_{\beta \to +\infty} o(\beta) = 0$. Then, Inequality (3.11) shows that $h_{max} \leqslant h$. Now consider μ_∞ any accumulation point for μ_β. Theorem 2.11 says that μ_∞ is A-maximizing. On the other hand,

$$h + \beta.m(A) + o(\beta) = \mathcal{P}(\beta) = h_{\mu_\beta} + \beta.\int A\,d\mu_\beta \leqslant h_{\mu_\beta} + \beta.m(A)$$

yields $h_{max} \geqslant h_{\mu_\infty} \geqslant \limsup\limits_{\beta \to +\infty} h_{\mu_\beta} \geqslant h \geqslant h_{max}$. $\qquad\square$

Remark 3.14. In Statistical Mechanics, h_{max} is called the residual entropy: it is the entropy of the system at temperature zero, when it reaches its ground state. ∎

Consequently, if the Mather set admits a unique measure of maximal entropy, μ_β converges to this measure, as $\beta \to +\infty$.

We have seen in Proposition 2.11 that the pressure function $\beta \mapsto \mathcal{P}(\beta)$ admits an asymptote as $\beta \to +\infty$. Actually this asymptote is given by

$$h_{max} + \beta.m(A).$$

Theorem 3.13 justifies the study of the set of maximizing orbits. This is the goal of the next section.

3.2   The Aubry set and the subcohomological inequality

3.2.1   The Aubry set and the Mañé potential

Theorem 3.15. *There exists an invariant set $\mathcal{A}$, called the* Aubry set, *which satisfies the following properties:*

1. *$\mathcal{A}$ contains the Mather set, or, equivalently, any A-maximizing measure has its support in $\mathcal{A}$.*

2. *Restricted to $\mathcal{A}$, $A - m(A)$ is equal to a Lipschitz coboundary.*

This theorem helps to answer the question: how can one detect maximizing orbits ? The terminology is borrowed from the Aubry-Mather Theory (see [115] [47] [56] [114] [66]).

Definition 3.16. Given A, the Mañé potential is:

$$S_A(x,y) :=$$

$$\lim_{\epsilon \to 0} \left[\sup \left\{ \sum_{i=0}^{n-1} \left[A(\sigma^i(z)) - m_A \right] \;\middle|\; n \in \mathbb{N}, \; \sigma^n(z) = y, \; d(z,x) < \epsilon \right\} \right].$$

The Mañé potential $S_A(x,y)$ describes in some sense the "optimal way" to go from x to y, following the dynamics, and also maximizing the "cost" given by A. For a fixed ϵ, the supremum may be not attained by a finite piece of orbit.

We point out that for a fixed x the function $y \mapsto S(x,y)$ is Holder. Nevertheless, for a fixed y the function $x \mapsto S(x,y)$ is not necessarily continuous.

Proof of Theorem 3.15 . The proof has several steps. We set $\mathcal{A} := \{ x \in \Omega \mid S_A(x,x) = 0 \}$. Let us check that $\mathcal{A}$ satisfies the required conditions.

   ● some useful equations.
Let us pick any calibrated subaction V. We remind that this means

$$V(\sigma(x)) \geqslant A(x) - m(A) + V(x),$$

which can be written on the form

$$A(x) = m(A) + V \circ \sigma(x) - V(x) + g(x), \tag{3.12}$$

where g is a non-positive Lipschitz-continuous function.

Let z be in Ω. Equality (3.12) yields

$$S_n(A - m(A))(z) = S_n(g)(z) + V \circ \sigma^n(z) - V(z). \tag{3.13}$$

Now, pick y and consider z such that $\sigma^n(z) = y$. This yields

$$S_n(A - m(A))(z) = S_n(g)(z) + V(y) - V(z) \leq$$

$$g(z) + V(y) - V(z) \leq V(y) - V(z). \tag{3.14}$$

Therefore, for every x and y, the continuity of V shows

$$S_A(x, y) \leqslant g(x) + V(y) - V(x) \leqslant V(y) - V(x). \tag{3.15}$$

• The Aubry set contains the Mather set.
We show that if μ is a A-maximizing ergodic measure, then, supp $\mu \subset \mathcal{A}$. This will imply, in particular, that $\mathcal{A}$ is not empty. Consider x a generic point for μ (and, also in the support of μ). As V is Lipschitz, then, Equality (3.13) used for $z := x$ yields $\lim_{n \to \infty} \frac{1}{n} S_n(A - m(A))(x) = 0$. Consequently, g is a non-positive function satisfying $\int g \, d\mu = 0$. Since μ is ergodic and g is continuous, we have that $g_{|\text{supp } \mu} \equiv 0$. This implies, as we have shown before,

$$A(x) = m(A) + V \circ \sigma(x) - V(x),$$

for every $x \in \text{supp } \mu$.

Now, x returns infinitely many times as closed as wanted to itself (see Proposition 1.20). This yields that for a given $\varepsilon = \frac{1}{2^N}$, there are infinitely many n_i such that

$$x = x_0 x_1 \ldots x_{n_i - 1} x_0 x_1 \ldots x_{N-1} \ldots.$$

Equivalently, this means that the word $x_0 \ldots x_{N-1}$ appears infinitely many times into $x = x_0 x_1 \ldots$.

This yields that for such n_i, the word

$$z = x_0 x_1 \ldots x_{n_i-1} x,$$

coincides with x for at least $n_i + N$ digits. Lipschitz regularity for A and g yields

$$|S_{n_i}(A)(z) - S_{n_i}(A)(x)| \leqslant C. \frac{1}{2^N}, \text{ and } |S_{n_i}(g)(z) - S_{n_i}(g)(x)| \leqslant C. \frac{1}{2^N}.$$

Remind that $g \circ \sigma^k(x) = 0$ for every k because x belongs to $\operatorname{supp} \mu$. Remember that V is Lipschitz continuous. Therefore, we get

$$|S_{n_i}(A - m(A))(z)| \leqslant C \frac{1}{2^N}, \sigma^{n_i}(z) = x \text{ and } d(z, x) \leqslant \frac{1}{2^{n_i+N}}.$$

This yields that $S_A(x, x) \geqslant 0$ and Inequality (3.15) yields $S_A(x, x) \leqslant 0$. Therefore, x belongs to $\mathcal{A}$.

 • We prove that $A - m(A)$ restricted to $\mathcal{A}$ is a coboundary.
Let x be in $\mathcal{A}$. Note that Inequality (3.15) yields

$$0 \leqslant S_A(x, x) \leqslant g(x) \leqslant 0,$$

which shows that g is equal to 0 on $\mathcal{A}$. Therefore, $A - m(A)$ is a coboundary on $\mathcal{A}$.

The last point to check is that $\mathcal{A}$ is σ-invariant. Let x be in $\mathcal{A}$. Take a fixed $\varepsilon_0 > 0$ and suppose that z satisfies $\sigma^n(z) = x$ and $d(x, z) < \varepsilon$. Then,

$$
\begin{aligned}
S_n(A)(z) &= A(z) + A \circ \sigma(z) + \ldots + A \circ \sigma^{n-1}(z) \\
&= A(\sigma(z)) + \ldots + A \circ \sigma^{n-2}(\sigma(z)) + A(z) \\
&= A(\sigma(z)) + \ldots + A \circ \sigma^{n-2}(\sigma(z)) + A(x) + A(z) - A(x) \\
&= S_n(A)(\sigma(z)) + A(z) - A(x).
\end{aligned}
$$

Note that $d(\sigma(z), \sigma(x)) = 2d(z, x) < 2\varepsilon$ and $|A(z) - A(x)| \leqslant C.d(z, x) \leqslant 2C\varepsilon$. Taking the supremum over all the possible n, for fixed ε, and then letting $\varepsilon \to 0$ we get

$$S_A(x, x) = S_A(\sigma(x), \sigma(x)).$$

□

Proposition 3.17. *The Aubry set is compact.*

Proof. The definition of S_A yields

$$S_A(x, x) = \lim_{\varepsilon \to 0} \sup_n \{S_n(g)(z), \ \sigma^n(z) = x \ d(x, z) < \varepsilon\}.$$

The function g is non-positive, thus $S_n(g)(z)$ is always non-positive. We claim that $S_A(x, x) = 0$ holds, if and only if, for every n, there exists z, such that $\sigma^n(z) = x$ and $S_n(g)(z) = 0$. Now, Lipschitz continuity of g shows that this condition is closed: if x does not satisfies this condition, there exists n such that for every z, satisfying $\sigma^n(z) = x$, we get $S_n(g)(z) < 0$. This is obviously true for every x' sufficiently close to x. Therefore, the set of x such that $S(x, x) = 0$ is a closed set.

We now prove the claim. If the later condition is not satisfied, then pick n_0 such that for every n_0-preimage z of x, $S_{n_0}(g)(z) < 0$. The set of n_0-preimages is finite, thus, there exists ε such that for every z

$$\sigma^{n_0}(z) = x \implies S_{n_0}(g)(z) < -\varepsilon.$$

Now, for $n > n_0$ and z such that $\sigma^n(z) = x$,

$$S_n(g)(z) = S_{n-n_0}(g)(z) + S_{n_0}(g)(\sigma^{n-n_0}(z)) \leqslant S_{n_0}(g)(\sigma^{n-n_0}(z)) < -\varepsilon.$$

This shows $S_A(x, x) \leqslant -\varepsilon < 0$.

Conversely, assume that $S_A(x, x) = 0$. Pick n_0 and then consider a sequence of n-preimages z_n of x converging to x such that

$$S_n(g)(z_n) \to_{n \to \infty} 0.$$

Consider the new sequence of points $\xi_n := \sigma^{n-n_0}(z_n)$. Each ξ_n is a n_0-preimage of x, then, there exists z, such that, $\sigma^{n_0}(z) = x$, and for infinitely many n, we have $\xi_n = z$. Therefore, considering only these n's we get

$$0 \geqslant S_{n_0}(g)(z) \geqslant S_n(g)(z_n),$$

and the right hand side term goes to 0, if $n \to +\infty$. This yields $S_{n_0}(g)(z) = 0$, and this holds for every n.

$\square$

3.2.2 The Aubry set for locally constant potentials

Here we consider a potential A depending on 2 coordinates, that is: $A(x_0 x_1 x_2 \ldots) = A(x_0, x_1)$. We set $A(i, j)$ for $A(x)$ with $x = ij \ldots$.

Note that in that case H_β is constant in each 1-cylinder set $[i]$. This also holds for $\dfrac{1}{\beta} \log H_\beta$, and, then, for any accumulation point V. Indeed, recall that $LH_\beta = \lambda_\beta H_\beta$. We can get the function H_β by the expression

$$H_\beta(x) = \lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \frac{L^k 1(x)}{\lambda_\beta^k}.$$

Note that

$$L1(x_0 x_1 \ldots) = \sum_{i \in \{1, \ldots, d\}} e^{\beta A(i, x_0)} 1(x_0 x_1 \ldots) = \sum_{i \in \{1, \ldots, d\}} e^{\beta A(i, x_0)}.$$

Iterating k times the action of the operator L, we get

$$L^k 1(x_0 x_1 \ldots) = \sum_{i_k, i_{k-1}, \ldots, i_1 \in \{1, \ldots, d\}} e^{\beta(A(i_k, i_{k-1}) + \cdots + A(i_1, x_0))}$$

Then, the function $H_\beta(x_0 x_1 \ldots)$ depends only on x_0 and so it is constant in each cylinder of size one $[i]$, as claimed.

If x is in Ω and z is such that $\sigma^n(z) = x$ and, moreover, $d(x, z) < 1$, then, $V(x) = V(z)$. Hence,

$$S_n(A - m(A))(z) = S_n(g)(z) \leqslant 0.$$

If y is in $\mathcal{A}$, $g_{|\mathcal{A}} \equiv 0$, then $S_n(A - m(A))(y) = V \circ \sigma^n(y) - V(y)$. If $\sigma^n(y)$ and y are in the same 1-cylinder, then $S_n(A - m(A))(y) = 0$. Furthermore, if z is the periodic orbit given by the concatenation of $y_0 y_1 \ldots y_{n-1}$, then, we get $S_n(A - m(A))(z) = S_n(A - m(A))(y)$, because all the transitions $z_i \to z_{i+1}$ are the same as for y (for $i \leqslant n - 1$).

This shows that $m(A)$ is reached by periodic orbits.

Definition 3.18. A periodic orbit obtained as the concatenation of $z_0 \ldots z_{n-1}$ is said to be *simple* if all the digits z_i are different.

Example. $123123123\ldots$ is a simple periodic orbit of length 3. By the other hand $121412141214\ldots$ is not a simple periodic orbit.

One simple periodic orbit of length n furnishes n *bricks*, which are the words producing the n points of the orbit :

Example. The bricks of $123123123\ldots$ are 123, 231 and 312.

Then, the Aubry set $\mathcal{A}$ is constructed as follows :

1. List all the simple periodic orbits. This is a finite set.

2. Pick the ones such that their Birkhoff means are maximal. This maximal value is $m(A)$. Such a simple periodic orbit is also called maximizing.

3. Consider the associated bricks for all these simple maximizing periodic orbits.

4. The set $\mathcal{A}$ is the subshift of finite type constructed from these bricks:

   (a) Two bricks can be combined if they have a common digit. On one of the simple loops one glues the other simple loop. The bricks $x_0 x_1 \ldots x_n$ and $x_n y_1 \ldots y_k$ produce the periodic orbit $x_0 x_1 \ldots x_n y_1 y_2 \ldots y_k x_n x_0 x_1 \ldots x_n y_1 y_2 \ldots y_k x_n \ldots$

   Example. It is easy to see that 123 and 345 produce the new orbit $123453123453\ldots$.

   (b) $\mathcal{A}$ is the closure of the set of all the periodic orbits obtained by this process.

We can also define $\mathcal{A}$ from its transition matrix. If i does not appear in any of the bricks, we set $T_{ij} = 0$, for every j. If i appears in a brick, set $T_{ij} = 1$, if ij appears in a brick (for $j \neq i$). If i is also a brick (that means that the fixed point $iiiiii\ldots$ is a maximizing orbit) set also $T_{ii} = 1$. Set $T_{ij} = 0$ otherwise. Then, $\mathcal{A} = \Sigma_T$.

Example. If the bricks are (up to permutations) abc, cde, fgh, gi

and fj, the transition matrix restricted to these letters is

$$T = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$
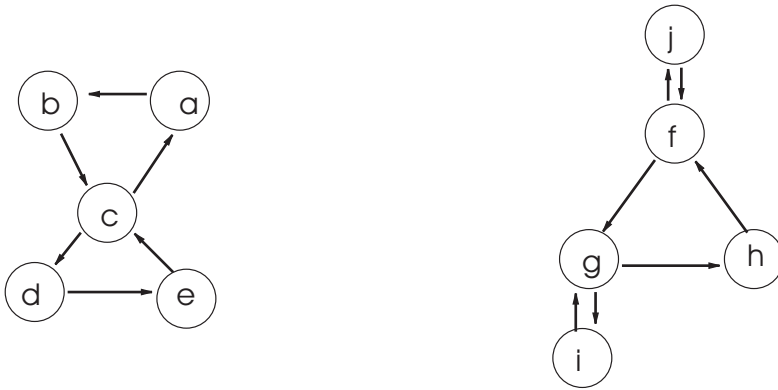

Figure 3.1: The graph for T

The set $\mathcal{A}$ is a subshift of finite type. It can thus be decomposed in irreducible components, say $\mathcal{A}_1 \ldots, \mathcal{A}_r$. Each component admits a unique measure of maximal entropy, say $\mu_1, \ldots, \mu_r$. Let h_i be the associated entropies. We assume that the order has been chosen in such way that

$$h_1 \geqslant h_2 \geqslant \ldots \geqslant h_r.$$

In that case, the topological entropy for $\mathcal{A}$ is h_1. More precisely, assume that j_0 is such that

$$h_1 = h_2 = \ldots = h_{j_0} > h_{j_0+1} \geqslant h_{j_0+2} \ldots.$$

Then, $\mathcal{A}$ admits exactly j_0 ergodic measures of maximal entropy h_1. Any ground state is a convex combination of these j_0 ergodic measures.

In that special case, it is proved that there is only one ground state:

Theorem 3.19 (see [33, 91, 45])**.** *If A depends only on two coordinates, then μ_β converges, as $\beta \to +\infty$.*

Example. In the previous example, $\mathcal{A}$ has two irreducible components. The first one has entropy $\frac{1}{3} \log 2 \sim 0.23$. The second one has entropy ~ 0.398. In that case the ground state is the unique measure of maximal entropy, which has support in the second irreducible component.

3.2.3 Some consequences of Theorem 3.19

A dual viewpoint for the selection problem is the following: for every β the probability μ_β has full support. One can say that a measure, in our case the probability μ_β, can be represented by its set of generic points (see Remark 1.21). This set of points is dense in Ω.

For each $\beta > 0$, this set remains dense in Ω, but in the limit, when $\beta \to +\infty$, this set accumulates on the Aubry set $\mathcal{A}$. More precisely, it is going to accumulate on the irreducible components which have positive weight for some ground state.

The selection problem is thus to determine what are the components of $\mathcal{A}$ where this set of generic points for μ_β accumulates.

It may happen that an irreducible component of $\mathcal{A}$ has maximal entropy but has no weight at temperature zero. In [91], the author introduced the notion of isolation rate between the irreducible components and showed that only the most isolated component have weight at temperature zero.

Chapter 4

The Peierl's barrier

4.1 Irreducible components of the Aubry set

4.1.1 Definition of the Peierl's barrier

We have seen above that if A depends on two coordinates the Aubry set $\mathcal{A}$ is a subshift of finite type. It thus has well-defined irreducible components, each one being the support of a unique measure of (relative) maximal entropy. In the case of a more general potential A, there are no reasons why $\mathcal{A}$ should be a subshift of finite type. Actually it can be any invariant subset as it was shown above : pick any compact set $\mathcal{A}$ and consider $A := -d(., \mathcal{A})$.

Given that, it is not obvious how define the irreducible components of $\mathcal{A}$ and how to determine the measures of maximal entropy.

Definition 4.1. The Peierl's barrier between x and y is defined by

$$h(x, y) := \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \{S_n(A - m(A))(z), \ \sigma^n(z) = y \ d(x, z) < \varepsilon\}.$$

We remind that we got $A = m(A) + V \circ \sigma - V + g$, where V is a calibrated subaction (obtained via a converging subsequence for $\frac{1}{\beta} \log H_\beta$) and g is a non-positive Lipschitz function. Replacing this

expression for A into the definition of the Peierl's barrier we get

$$h(x,y) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \{S_n(g)(z) + V(y) - V(z), \ \sigma^n(z) = y \ d(x,z) < \varepsilon\}$$

$$= \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \{S_n(g)(z), \ \sigma^n(z) = y \ d(x,z) < \varepsilon\} + V(y) - V(x).$$

This shows that to compute $h(x,y)$, we morally have to find a sequence of pre-images for y which converges as fast as possible to x. We shall see later that it is of prime importance the equality $h(x,y) + h(y,x) = h(x,x)$.

Theorem 4.2. *For any x, the Peierl's barrier $y \mapsto h(x,y)$ is a Lipschitz calibrated subaction. Moreover, $h(x,x) = 0$, if and only if, x belongs to $\mathcal{A}$.*

Proof. Pick x and y. Consider z such that $\sigma^n(z) = y$, $d(x,z) < \varepsilon$. Note that in Ω, z is just the concatenation $z_0 \ldots z_{n-1}y$. For y' close to y (namely $y_0 = y_0'$), we consider $z' := z_0 \ldots z_{n-1}y'$; the Lipschitz regularity for g yields

$$|S_n(g)(z) - S_n(g)(z')| \leqslant C.d(y,y'),$$

for some constant C. If we consider a sequence of z realizing the lim sup and then take $\varepsilon \to 0$, we get

$$h(x,y) \leqslant h(x,y') + C.d(y,y').$$

The same argument shows $h(x,y') \leqslant h(x,y) + C.d(y,y')$, and, then $y \mapsto h(x,y)$ is Lipschitz continuous.

Let us show that this function is a calibrated subaction. In this way, consider y, n and ε, such that, $\sigma^n(z) = y$ and $d(x,z) < \varepsilon$. Then, $d(x,z) < \varepsilon$ and $\sigma^{n+1}(z) = \sigma(y)$. Moreover,

$$S_{n+1}(A - m(A))(z) = S_n(A - m(A))(z) + A(y) - m(A). \quad (4.1)$$

Consider a sequence of z realizing the lim sup for $h(x,y)$, now taking the limit along the subsequence, and, considering $\varepsilon \to 0$, we get

$$h(x,y) + A(y) - m(A) \leqslant h(x,\sigma(y)).$$

This shows that $y \mapsto h(x,y)$ is a subaction. It remains to show that for a fixed value for $y' = \sigma(y)$, the equality is achieved by one of

the preimages of y'. This follows from taking the z's and the n's in Equality (4.1) which realize the $\limsup$ for the left hand-side of the equality. Then, we get

$$h(x, y') \leqslant h(x, y) + A(y) - m(A),$$

with $\sigma(y) = y'$. As the reverse inequality holds, the global equality follows.

Both definitions of the Mañé potential and the Peierl's barrier are very similar, except that in the first case we consider the supremum, and in the other case we consider the $\limsup$. This immediately shows that

$$h(x, y) \leqslant S_A(x, y),$$

and then, $h(x, x) = 0$ yields $S_A(x, x) = 0$ (due to Inequality 3.15), thus x belongs to $\mathcal{A}$.

Let us prove the converse. Take x in $\mathcal{A}$. Consider $\rho > 0$ small and ε_0, such that, for every $\varepsilon < \varepsilon_0$,

$$\sup_n \{S_n(A - m(A))(z), \ \sigma^n(z) = x \ d(z, x) < \varepsilon\} \geqslant -\rho. \qquad (4.2)$$

In the following, we assume for simplicity that x is not periodic, but the proof can be easily extended to that case.

Inequality(4.2) holds for every ε. We will construct a subsequence (n_k) by induction. We pick any ε and consider n_0 realizing the supremum up to $-\rho$:

$$S_{n_0}(A - m(A))(z_0) > -2\rho \text{ with } \sigma^{n_0}(z_0) = x \text{ and } d(x, z_0) < \varepsilon.$$

Now we use Inequality(4.2) but with[1]

$$\varepsilon_1 < \min\{d(x, z), \sigma^n(z) = y, \ n \leqslant n_0\}.$$

We get $n_1 > n_0$ and z_1, such that,

$$S_{n_1}(A - m(A))(z_1) > -2\rho \text{ with } \sigma^{n_1}(z_1) = x \text{ and } d(x, z_1) < \varepsilon_1.$$

We then proceed by induction with

$$\varepsilon_{k+1} < \min\{d(x, z), \sigma^n(z) = y \ n \leqslant n_k\}.$$

---

[1]Here we use that this distance is positive, that is that x is not periodic.

Then, we have

$$-2\rho \quad < \quad S_{n_k}(A - m(A))(z_k), \text{ with } \sigma^{n_k}(z_k) = y \text{ and } d(x, z_k) < \varepsilon$$

then

$$-2\rho \quad \leqslant \quad \limsup_{n\to\infty} \{S_n(A - m(A))(z), \ \sigma^n(z) = x \ d(x, z) < \varepsilon\}.$$

This holds for every $0 < \varepsilon < \varepsilon_0$, and then $h(x, x) \geqslant -2\rho$. Now, we take $\rho \to 0$, and we get $h(x, x) \geqslant 0$. The reverse equality is always true, so $h(x, x) = 0$. $\qquad\square$

4.1.2 The irreducible components of the Aubry set

Now, we show that the Peierl's barrier allows to define "irreducible" components of the Aubry set.

Lemma 4.3. *For any x, y and z*

$$h(x, y) \geqslant h(x, z) + h(z, y). \tag{4.3}$$

Proof. Let $\varepsilon > 0$ be fixed. Consider a preimage y' of y close to z and a preimage z' of z close to x. For small ε, z' satisfies $z' = z'_0 \ldots z'_{n-1} z$ and then $y'' := z'_0 \ldots z'_{n-1} y'$ is also a preimage of y. The cocycle relation yields, if $\sigma^m(y') = y$,

$$S_{n+m}(A - m(A))(y'') = S_n(A - m(A))(y'') + S_m(A - m(A))(y').$$

The Lipschitz regularity shows that $S_n(A - m(A))(y'')$ differs from $S_n(A - m(A))(z')$ by the term $\pm C.\varepsilon$. If we assume that the n's and the $m's$ are chosen to realize the respective lim sup, the term on the left-hand side is lower than the lim sup. Then, the Lemma is proved. $\qquad\square$

Lemma 4.4. *For any x in $\mathcal{A}$, $h(x, \sigma(x)) + h(\sigma(x), x) = 0$.*

Proof. Lemma 4.3 and Theorem 4.2 show that

$$h(x, \sigma(x)) + h(\sigma(x), x) \leqslant h(x, x) = 0.$$

It remains to prove it is non-negative. The Peierl's barrier is a sub-action thus

$$h(x, \sigma(x)) \geqslant A - m(A) + h(x, x) = A - m(A).$$

Pick $\varepsilon > 0$ and consider y, a preimage of x ε-close to $\sigma(x)$. Suppose $y = y_0 \dots y_{n-1}x$. Then, $x' := x_0 y_0 \dots y_{n-1}x$ is a preimage of x ε-close[2] to x. In particular $A(x') = A(x) \pm C.\varepsilon$. We emphasize that it is equivalent to have $x' \to x$ or $y \to \sigma(x)$.

We assume that these quantities are chosen in such way that $S_{n+1}(A - m(A))(x')$ converges to the lim sup, if n goes to $+\infty$. Now,

$$S_{n+1}(A - m(A))(x') = A(x') - m(A) + S_n(A - m(A))(y).$$

Taking $n \to +\infty$ and then $\varepsilon \to 0$, the right-hand side term is lower than $h(\sigma(x), x) + A(x) - m(A)$ and the left-hand side term goes to $h(x, x) = 0$. This yields

$$0 \leqslant h(\sigma(x), x) + A(x) - m(A) \leqslant h(\sigma(x), x) + h(x, \sigma(x)).$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Lemma 4.5. *Let x y and z be in $\mathcal{A}$. If $h(x, y) + h(y, x) = 0$ and $h(y, z) + h(z, y) = 0$, then $h(x, z) + h(z, x) = 0$.*

Proof. Inequality (4.3) shows

$$h(x, z) + h(z, x) \geqslant h(x, y) + h(y, z) + h(z, y) + h(y, x) = 0 + 0 = 0.$$

It also yields $h(x, z) + h(z, x) \leqslant h(x, x) = 0$. $\qquad\qquad\qquad\qquad\qquad$ □

Lemma 4.5 proves that $h(x, y) + h(y, x) = 0$ is a transitive relation. Since it is obviously symmetric and reflexive, then it is an equivalence relation on $\mathcal{A}$.

Definition 4.6. The equivalence classes for the relation

$$h(x, y) + h(y, x) = 0,$$

are called irreducible components of $\mathcal{A}$.

Note that x and $\sigma(x)$ belong to the same class, which shows that the classes are invariant. The continuity for the Peierl's barrier has been proved with respect to the second variable for a fixed first variable. It is thus not clear that an irreducible component is closed.

---

[2] Actually $\frac{\varepsilon}{2}$-close to x.

4.1.3 The locally constant case

If the potential is locally constant, we have seen in Subsection 3.2.2 that the Aubry set $\mathcal{A}$ is a subshift of finite type, for which the notion of irreducible component has already been defined (see 1.3). We have to check that the two notions coincide.

We have seen that the irreducible components of a subshift of finite type are exactly the transitive components. We shall use this description to show that irreducible components of the Aubry (in the sense of the Peierl's barrier) set are the irreducible components (with respect to subshifts).

Lemma 4.7. *Assume A depends only on two coordinates. Let x and y be in Ω. Assume that $d(z, x) \leqslant \frac{1}{4}$, $\sigma^n(z) = y$, and, moreover, that there exists $1 \leqslant k < n$, such that, $d(\sigma^k(z), x) \leqslant \frac{1}{4}$. Then,*

$$S_n(A - m_A)(z) \leqslant S_{n-k}(A - m_A)(\sigma^k(z)).$$

Proof. We consider a calibrated subaction V. We have seen that it depends only on one coordinate. We remind that for every ξ,

$$A(\xi) = m_A + V \circ \sigma(\xi) - V(\xi) + g(\xi) \tag{4.4}$$

holds, where g is a non-positive function (depending only on 2 coordinates). Now we have

$$S_k(A - m_A)(z) = S_k(g)(z) + V(\sigma^k(z)) - V(z) = S_k(g)(z) \leqslant 0.$$

$\square$

From Lemma 4.7 we claim that for every x and y,

$$h(x, y) = S_A(x, y) = \max \left\{ S_n(A - m_A)(z) \, \sigma^n(z) = y \, d(z, x) \leqslant \frac{1}{4} \right\}. \tag{4.5}$$

Lemma 4.8. *Consider some fixed y in Ω. The map $x \mapsto h(x, y)$ is continuous*

Proof. Consider a sequence (x_n) converging to x. Assume that all these x_n and x coincide for at least 2 digits. Then,

$$d(z, x_n) \leqslant \frac{1}{4} \iff d(z, x) \leqslant \frac{1}{4}.$$

For y and n, consider any z realizing the maximum into the definition of $S_A(x_n, y)$. It also realizes the maximum for $S_A(x, y)$ and more generally for every $S_A(x_k, y)$.   □

Lemma 4.8 shows that any irreducible components of the Aubry in the sense of definition 4.6 is closed. It is also invariant. We thus have to show it is transitive.

Let us consider some components and pick two open sets U and V (for the components). We still assume that Equality (4.4) holds. Consider $x \in U \cap \mathcal{A}$ and $y \in V \cap \mathcal{A}$. By definition

$$h(x, y) + h(y, x) = 0.$$

By Equality (4.5), $h(x, y)$ is realized by some $S_n(A - m_A)(z)$ and $h(y, x)$ is realized by some $S_m(A - m_A)(z')$. Moreover, we can also assume that z belongs to U and z' belongs to V.

Indeed, if it is not the case, we can always follow preimages of x in the component $\mathcal{A}$ which are exactly on the set $g^{-1}(\{0\})$.

>From the two pieces of orbits which are $z, \sigma(z), \ldots, \sigma^n(z)$ and $z', \sigma(z'), \ldots, \sigma^m(z')$ we can construct a periodic orbit. Denote by ξ the point of this periodic orbit in U. Then, we claim that

$$h(y, \xi) = h(y, x) = S_m(A - m_A)(z')$$

and,

$$h(\xi, y) = h(x, y) = S_n(A - m_A)(z).$$

This shows that ξ belongs to $\mathcal{A}$ and to the same component than y. Therefore $\sigma^{-m}(U) \cap V \neq \emptyset$ and the component is transitive.

4.2   On the road to solve the subcohomological inequality

4.2.1   Peierl's barrier and calibrated subactions

From properties of the Peierl's barrier we get the following result: a calibrated subaction is entirely determined by its values on the Aubry set.

Theorem 4.9 (see [59]Th. 10). *Any calibrated subaction u satisfies for any y*

$$u(y) = \sup_{\mathbf{x} \in \mathcal{A}} [h(\mathbf{x}, y) + u(\mathbf{x})], \qquad (4.6)$$

Proof. We show that any calibrated subaction u is entirely determined by its values on the Aubry set $\mathcal{A}$.

Let us thus consider some calibrated subaction u. Let y be in Ω. Let y_{-1} be any preimage of y such that

$$u(y) = A(y_{-1}) - m(A) + u(y_{-1}).$$

More generally we consider a sequence y_{-n} such that $\sigma(y_{-n}) = y_{-n+1}$ and

$$u(y_{-n+1}) = A(y_{-n}) - m(A) + u(y_{-n}).$$

We claim that any accumulation point for (y_{-n}) belongs to $\mathcal{A}$. Indeed, let us consider some converging sequence y_{-n_k}, converging to x.

For $k' > k$ we get $\sigma^{n_{k'} - n_k}(y_{-n_{k'}}) = y_{-n_k}$. Therefore, we have

$$S_{n_{k'} - n_k}(A - m(A))(y_{-n_{k'}}) = u(y_{-n_k}) - u(y_{-n_{k'}}),$$

and, $y_{n_k} \to x$ yields

$$S_A(x, x) \geqslant 0.$$

As $S_A(x, x) \leqslant 0$ always holds (see Inequality (3.15)) we get that the limit point x belongs to $\mathcal{A}$.

Now we have that $S_{n_k}(A - m(A))(y_{n_k}) = u(y) - u(y_{n_k})$, which yields

$$h(x, y) \geqslant u(y) - u(x).$$

In particular $u(y) \leqslant \sup_{x' \in \mathcal{A}} \{h(x', y) + u(x')\}$ holds.

Actually, the reasoning we have just done allows to get a stronger result. Consider z in Ω, then we can write

$$A(z) = m(A) + u \circ \sigma(z) - u(z) + g(z),$$

where g is a non-positive Lipschitz function. Now, consider z_n such that $\sigma^n(z) = y$. Therefore,

$$S_n(A - m(A))(z_n) = u(y) - u(z) + S_n(g)(z) \leqslant u(y) - u(z).$$

This shows that

$$h(x', y) \leqslant u(y) - u(x')$$

always holds (consider any x' and take a subsequence of z_n converging to it). Then, Equality (4.6) holds. $\qquad\square$

Moreover, the irreducible component get a special importance :

Theorem 4.10. *If x and z are in the same irreducible component of $\mathcal{A}$, then for any y,*

$$h(x, y) + u(x) = h(z, y) + u(z).$$

Proof. We remind two inequalities and an important equality:

$$
\begin{aligned}
h(x, y) &\geqslant h(x, z) + h(z, y), \\
u(x) &\geqslant h(z, x) + u(z), \\
0 &= h(x, z) + h(z, x).
\end{aligned}
$$

Then,

$$
\begin{aligned}
u(x) + h(x, y) &\geqslant u(x) + h(x, z) + h(z, y) \\
&\geqslant u(x) - h(z, x) + h(z, y) + u(z) - u(z) \\
&\geqslant u(x) - u(z) - h(z, x) + h(z, y) + u(z) \\
&\geqslant h(z, y) + u(z).
\end{aligned}
$$

Exchanging the roles of x and z we get that the reverse inequality also holds. $\qquad\square$

4.2.2 Selection of calibrated subactions

We remind that for a given $\beta > 0$, the equilibrium state μ_β is also a Gibbs measure obtained by the product of the *eigenfunction H_β* and the *eigenprobability*[3] ν_β. We also remind that any accumulation point for the family $\dfrac{1}{\beta} \log H_\beta$ is a calibrated subaction.

Uniqueness of the maximizing measure gives a partial answer to these questions:

---

[3]which is also the conformal measure.

Theorem 4.11. *Assume that there is a unique A-maximizing measure, then all the calibrated subactions are equal up to an additive constant.*

Proof. In that case $\mathcal{A}$ is uniquely ergodic and has thus a single irreducible component. If x_0 is any point of $\mathcal{A}$, Theorems 4.10 and 4.9 show that any calibrated subaction is entirely determined by its value on x_0. □

We point out that even in that simple case, the convergence for $\frac{1}{\beta} \log H_\beta$ is not clear.

In that direction we mention one of the results in [93]. For simplicity we state it using the setting of [15].

Theorem 4.12 (see [93]). *Assume that $\Omega = \{1, 2, 3\}^{\mathbb{N}}$ and A satisfies*

$$A(x) := \begin{cases} -d(x, 1^\infty) \text{ if } x = 1 \ldots, \\ -3d(x, 2^\infty) \text{ if } x = 2 \ldots, \\ -\alpha < 0 \text{ if } x = 3 \ldots. \end{cases}$$

Then, $\nu_\beta \to \delta_{1^\infty}$, as $\beta \to +\infty$, and, moreover, $\frac{1}{\beta} \log H_\beta$ converges.

This theorem shows that flatness is a criterion for selection: the Aubry set in that case is reduced to $\{1^\infty\} \cup \{2^\infty\}$ and the two unique ergodic maximizing measures are the Dirac measures δ_{1^∞} and δ_{2^∞}. The potential is "more flat" in 1^∞ than in 2^∞. Therefore, this Theorem says that the locus where the potential is flatter gets all the mass in the limit of the eigenmeasure, when $\beta \to \infty$. In that case this is sufficient to determine all the calibrated subactions.

More generally if the Aubry set $\mathcal{A}$ is not a subshift of finite type, the problem concerning selection is that

1. there is no satisfactory theory for the analysis of the measure of maximal entropy for general subshifts.

2. we do not know about the existence or uniqueness of conformal measures (one of the key points in Theorem 4.12 in the procedure of selection of calibrated subactions).

We shall also see that the problem of selection of subaction is related to the multiplicity of an eigenvector in the Max-Plus formalism.

Chapter 5

The Involution Kernel

5.1 Introduction and main definitions

In this section we will present the involution kernel which is a concept that is sometimes useful for understanding problems of different areas: large deviations ([12]), issues of differentiability of the main eigenfunction and piecewise differentiability of the subaction [111] [110], optimal transport (see [49] [106],[99],[100]), etc... It is also related to the Gromov distance on hyperbolic spaces (see [101]).

The main issue here is that is sometimes helpful for understanding a problem to look for the dual problem.

We recall that the Bernoulli space is the set $\{1, 2, ..., d\}^{\mathbb{N}} = \Omega$. A general element x in the Bernoulli space Ω will be denoted in this section by $x = (x_0, x_1, .., x_n, ..)$. The function σ denotes the shift acting on Ω.

We will consider another copy of Ω which will be denoted by $\tilde{\Omega}$. Points in this set are denoted by $w = (w_0, w_1, w_2, .., w_n, ..)$.

We denote $\{1, 2, ..., d\}^{\mathbb{Z}} = \hat{\Omega} = \tilde{\Omega} \times \Omega = \{1, 2, ..., d\}^{\mathbb{N}} \times \{1, 2, ..., d\}^{\mathbb{N}}$. Points in this set are denoted by

$$(w \,|\, x) = (w, x) = (...w_n...w_3, w_2, w_1, w_0 \,|\, x_0, x_1, .., x_n, ..).$$

For a fixed $i \in \{1, 2, .., d\}$, the function $\psi_i : \Omega \to \Omega$ indicates the i-th inverse branch of σ, $i \in \{1, 2, .., d\}$. This means $\psi_i(x) =$

$(i, x_0, x_1, .., x_n, ..)$. We can also use the notation

$$\psi_w(x) = (w_0, x_0, x_1, .., x_n, ..),$$

$w \in \Omega$, in the case, $w = (w_0, w_1, .., w_n, ..)$. In other words $\psi_w = \psi_{w_0}$.

We also denote by $\tilde{\sigma}$ the shift on $\tilde{\Omega}$. Finally, $\mathbb{T}^{-1}$ is the backward shift on $\hat{\Omega}$ given by

$$\mathbb{T}^{-1}(w, x) = (\tilde{\sigma}(w), \psi_{w_0}(x)).$$

We will present some general results for a Hölder function A which does not necessarily depends on two coordinates. Later, in this section, we will assume that A depends on two coordinates.

It is known [12] [110] that given $A : \Omega \to \mathbb{R}$ Hölder, in the variable $A(x)$, there exists a dual function $A^* : \tilde{\Omega} \to \mathbb{R}$, in the variable $A^*(w)$, and $W : \tilde{\Omega} \times \Omega \to \mathbb{R}$, such that

$$A^*(w) = A \circ \mathbb{T}^{-1}(w, x) + W \circ \mathbb{T}^{-1}(w, x) - W(w, x).$$

The functions $A^* : \tilde{\Omega} \to \mathbb{R}$ and $W : \tilde{\Omega} \times \Omega \to \mathbb{R}$ are both Hölder. We say that W is **the involution kernel** and A^* is **the dual potential** for A.

The A^* and W are not uniquely defined.

The expression for A^* can be also written as

$$A^*(w) = A(w_0, x_0, x_1, ..)+$$

$$W(...w_2, w_1 \,|\, w_0, x_0, x_1, ..) - W(.., w_1, w_0 \,|\, x_0, x_1, ..).$$

We say that A is symmetric if $A = A^*$.

Suppose A is fixed, and W and A^* are also fixed. For a given real parameter β we have

$$\beta \, A^*(w) = \beta A \circ \mathbb{T}^{-1}(w, x) + \beta W \circ \mathbb{T}^{-1}(w, x) - \beta W(w, x).$$

It follows that for any real β we have that βW is the involution kernel and βA^* is the dual potential for βA

Given A, we denote

$$\Delta(x, x', w) = \sum_{n \geq 1} A \circ \tau_{w,n}(x) - A \circ \tau_{w,n}(x').$$

The involution kernel W can be explicitly computed in the following way: for any (w, x) we define W by $W(w, x) = \Delta_A(x, x', w)$, where we choose a point x' for good [12].

It is known the following relation: for any $x, x', w \in \Omega$, we have that $W(w, x) - W(w, x') = \Delta(x, x', w)$.

We denote $\phi_{\beta A} = \phi_\beta : \Omega \to \mathbb{R}$ the main eigenfunction of the Ruelle operator for βA and $\lambda(\beta A) = \lambda(\beta)$ the main eigenvalue. In the same way we denote $\phi_{\beta A^*} = \phi_\beta^*$, where, $\phi_{\beta A^*} : \tilde{\Omega} \to \mathbb{R}$, the main eigenfunction of the Ruelle operator for βA^* and $\lambda(\beta A)^*$, which is the corresponding main eigenvalue.

One can show (see [110]) that $\lambda(\beta A) = \lambda(\beta A)^*$.

$\nu_\beta = \nu_{\beta A}$ denotes the normalized eigenprobability for the Ruelle operator for βA, and, $\nu_\beta^* = \nu_{\beta A^*}$ denotes the normalized eigenprobability for the Ruelle operator for βA^*.

Finally, the probabilities $\mu_\beta = \mu_{\beta A} = \phi_{\beta A} \nu_{\beta A}$ and $\mu_\beta^* = \mu_{\beta A^*} = \phi_A^* \nu_A^*$, are, respectively, the equilibrium states for βA and βA^*.

In the case A is symmetric we have $\nu_\beta = \nu_\beta^*$, $\mu_\beta = \mu_\beta^*$, etc...

We denote the β normalizing constant by

$$c(\beta) = \log \int \int e^{\beta W(w,x)} \, d\nu_{\beta A^*}(w) \, d\nu_{\beta A}(x),$$

and

$$\gamma = \lim_{\beta \to \infty} \frac{1}{\beta} c(\beta).$$

Therefore,

$$\gamma = \lim_{\beta \to \infty} \frac{1}{\beta} c(\beta) = sup_{w,x} \{W(w,x) - V(x) - V^*(w) - I(x) - I^*(w)\}.$$

The probability $e^{\beta W(w,x)} \, d\nu_{\beta A^*}(w) \, d\nu_{\beta A}(x)$ (after normalization) is invariant for the shift $\hat{\sigma}$ acting on $\{1, 2, .., d\}^{\mathbb{Z}}$.

It is known (see [12]) that

$$\phi_{\beta A}(x) = \int e^{\beta W(w,x) - c(\beta)} \, d\nu_{\beta A^*}(w), \qquad (5.1)$$

is the normalized eigenfunction for the Ruelle operator of βA.

Moreover,

$$\phi_{\beta A^*}(w) = \int e^{\beta W(w,x) - c(\beta)} \, d\nu_{\beta A}(x), \qquad (5.2)$$

Note from above that we also get

$$\phi_{\beta A}(x) = \int e^{\beta W(w,x) - c(\beta)} \, (\phi_{\beta A^*}(w))^{-1} \, d\mu_{\beta A^*}(w), \qquad (5.3)$$

and this is a relation between $\phi_{\beta A^*}$ and $\phi_{\beta A}$.

The main point above is that, via an integral Kernel e^W, one can get information of the eigenfunction via the eigenmeasure of the dual problem.

We denote respectively by V and V^* any calibrated subaction for A and A^*.

If the maximizing probability for A is unique, the maximizing probability for A^* is also unique (see [110]). We can also get a large deviation principle for $\mu_{\beta A^*}$.

We denote by I^* the deviation function for A^* (see [12]). We presented on the previous section 3.1.3 some basic properties of the deviation function I for A. The only difference now is that we consider the same for A^*.

Suppose V is the limit of a subsequence $\frac{1}{\beta_n} \log \phi_{\beta_n}$, where ϕ_{β_n} is an eigenfunction of the Ruelle operator for $\beta_n A$. Suppose V^* is obtained in an analogous way for A^* (using a common subsequence). Then, there exists γ such that

$$\gamma + V(x) = \sup_{w \in \Omega} \left[W(w,x) - V^*(w) - I^*(w) \right]. \qquad (5.4)$$

This follows from Varadhan's Integral Lemma (see [52]) and the fact that I is the Large deviation function for the family $\mu_{\beta A}$ (see [100] [111] [48]).

Then, V is the $W(w,x) - I^*(w) - \gamma$ transform of V^* (see [136] for definitons)

There is a dual expression

$$\gamma + V^*(w) = \sup_{x \in \Omega} \left[\, W(w,x) - V(x) - I(x)\,\right]. \qquad (5.5)$$

Note that the above equation is not for a general pair of calibrated subactions V and V^*, it is an equation for selected subactions.

We point out that when $A = A^*$ any calibrated subaction V obtained via the limit of eigenfunction of the Ruelle operator when β to ∞ will satisfy the equation

$$\gamma + V(w) = \sup_{x \in \Omega} \left[\, W(w,x) - V(x) - I(x)\,\right] \qquad (5.6)$$

for some constant γ.

In the general case $(A \neq A^*)$ we have that the limit calibrated subaction V satisfies the equation

$$V(x) = \max_{y \in \Omega} \left\{ W(y,x) - \max_{z \in \tilde\Omega} \left\{ W(y,z) - V(z) \right\} \right\}.$$

Note that there is no γ in the above equation.

Under the hypothesis of twist condition for A (to be defined bellow), and in the case the maximizing probability has support in a periodic orbit, the above equation can help to obtain explicit expression for subactions (see [111] [49]).

We consider now on $\Omega = \{1, ..., d\}^{\mathbb{N}}$ the lexicographic order. This order is obtained from the order $1 \leq 2 \leq 3 \leq ... \leq d$.

Following [100] we define:

Definition 5.1. We say a continuous $G : \hat\Omega = \tilde\Omega \times \Omega \to \mathbb{R}$ satisfies the **twist** condition on $\hat\Omega$, if for any $(a,b) \in \hat\Omega = \tilde\Omega \times \Omega$ and $(a',b') \in \tilde\Omega \times \Omega$, with $a' > a$, $b' > b$, we have

$$G(a,b) + G(a',b') < G(a,b') + G(a',b). \qquad (5.7)$$

Definition 5.2. We say a continuous $A : \Omega \to \mathbb{R}$ satisfies the twist condition, if some of its involution kernels W satisfies the twist condition.

The twist condition plays in Ergodic Optimization, in some sense, the same role as the "convexity hypothesis of the Lagrangian" which is of fundamental importance in Aubry-Mather theory (see [115], [47], [56] or [63]).

An alternative definition for the twist condition would be to claim that for any $a' > a$, $b' > b$ we have

$$G(a, b) + G(a', b') > G(a, b') + G(a', b). \qquad (5.8)$$

This will make no difference in the nature of the results we will get.

Several examples of potentials satisfying the twist condition appear in [100].

Linear combination with positive coefficients of potentials A which satisfy the twist condition also satisfies the twist condition.

Given x, we denote $w_x \in \tilde{\Omega}$ any point such that

$$\gamma + V(x) = [\, W(w_x, x) - V^*(w_x) - I^*(w_x) \,].$$

The proof of the above appears in [111].

Proposition 5.3. *If A is twist, then $x \to w_x$ is monotonous non-increasing, where w_x was chosen to be optimal.*

For a proof see [111].

The results above were used to show how to get subactions in an explicit form for potentials A which satisfy some restrictions (see [111] [100].

5.2 Examples

We point out that, for some kind of questions, it is more easy to ma-
nipulate expressions with the involution kernel than with the Peierl's
barrier.

For example, note that if A_1 and A_2 are two potentials and

$$A_1^*(w) = A_1 \circ \mathbb{T}^{-1}(w,x) + W_1 \circ \mathbb{T}^{-1}(w,x) - W_1(w,x),$$

$$A_2^*(w) = A_2 \circ \mathbb{T}^{-1}(w,x) + W_2 \circ \mathbb{T}^{-1}(w,x) - W_2(w,x),$$

then the involution kernel for $(A_1 + A_2)$ is $(W_1 + W_2)$ and its dual
potential is $(A_1^* + A_2^*)$.

We use the notation $W(w,x) = W(...w_2, w_1, w_0 \,|\, x_0, x_1, x_2, ..)$ and

$$\hat{\chi}_{[z_k,..z_2,z_1,z_0 \,|\, x_0,x_1,...x_n]}(w,x),$$

denotes the indicator function in $\hat{\Omega}$ of the cylinder set

$$[z_k, ..z_2, z_1, z_0 \,|\, x_0, x_1, ...x_n].$$

All the above was for case of general potentials. Now, we analyze
briefly the case where the potential depends on two coordinates.

The real function on $\tilde{\Omega}$ denoted by $\tilde{\chi}_{[a_1,a_0]}$ is such that is 1, if
$w = (...w_2, w_1, w_0) = (...w_2, a_0, a_1)$, and, zero otherwise.

Example.

Given $A = \chi_{[a_0,a_1]}$, the indicator function of the cylinder $[a_0, a_1]$,
defined on Ω, then, for any

$$(w,x) = (...w_2, w_1, w_0 \,|\, x_0, x_1, x_2, ...),$$

we have

$$\tilde{\chi}_{[a_1,a_0]}(...w_2, w_1, w_0) = A(w_0, x_0, ...)+$$

$$\hat{\chi}_{[a_0 \mid a_1]}(...w_2, w_1 \mid w_0, x_0, x_1, ...) - \hat{\chi}_{[a_0 \mid a_1]}(...w_2, w_1, w_0 \mid x_0, x_1..).$$

Therefore the dual potential of $A = \chi_{[a_0, a_1]}$ is $A^* = \tilde{\chi}_{[a_1, a_0]}$ and its involution kernel is $W = \hat{\chi}_{[a_0 \mid a_1]}$.

We point out that in the case the potential depends on infinite coordinates there is no simple expression for the dual potential (and the involution kernel)

Any potential A which depends on two coordinates on the full shift $\Omega = \{1, ..., d\}^{\mathbb{N}}$ can be written in the form

$$A = \sum_{i,j \in \{1,...,d\}} \alpha_{i,j} \, \chi_{[i,j]}$$

and therefore, has involution kernel

$$W = \sum_{i,j \in \{1,...,d-\}} \alpha_{i,j} \, \hat{\chi}_{[i,j]},$$

and dual potential

$$A^* = \sum_{i,j \in \{1,...,d\}} \alpha_{i,j} \, \chi_{[j,i]}.$$

In this way if we consider A as a matrix, **the dual A^* is the transpose of A.**

Example. If $A : \{1, 2\}^{\mathbb{N}} \to \mathbb{R}$, which depends on two coordinates, satisfies: $A(1,1) = 2$, $A(2,2) = 5$, $A(1,2) = 7$, and $A(2,1) = 6$, then, the involution kernel W is

$$W = 2 \, \hat{\chi}_{[1 \mid 1]} + 5 \, \hat{\chi}_{[2 \mid 2]} + 7 \, \hat{\chi}_{[1 \mid 2]} + 6 \, \hat{\chi}_{[2,1]}.$$

This involution kernel W satisfies the twist condition.

The dual A^* of A (via the involution kernel W^*) increase the scope of the concept of transpose for the case A depends on an infinite number of coordinates.

Example.

When $\Omega = \{0,1\}^{\mathbb{N}}$ we we denote M_k the cylinder $[\underbrace{1111...111}_{k} \, 0]$,

$k \in \mathbb{N}$, $k \geq 1$.

A well known class of potentials (see [69] [108] [58] [16]) which are not Hölder is the following: suppose γ is a real positive parameter, then take $A = A_\gamma$ such that

$$A_\gamma(\underline{x}) = -\gamma \log\left(\frac{k+1}{k}\right) \quad \text{if } \underline{x} \in M_k, \ k \neq 0$$

$$A_\gamma(\underline{x}) = -\log(\zeta(\gamma)) \quad \text{if } \underline{x} \in [0]$$

$$A_\gamma(11...)) = 0,$$

where $\zeta(\gamma) = 1^{-\gamma} + 2^{-\gamma} + \cdots$.

We denote bellow by $W(k,j)$ the value of the involution kernel in the set of points $(w|x) = (...w_1 w_0 \,|\, x_0 x_1 x_2 ...)$ such that $w \in M_k$ and $x \in M_j$, $k, j \in \mathbb{N}$.

The involution kernel W for these potentials satisfy

$$e^{W(k,j)} = \frac{(j+1+k)^{-\gamma}}{(j+1)^{-\gamma}(k+1)^{-\gamma}} \zeta(\gamma),$$

$k, j \in \mathbb{N}$. Note that

$$W(k,j) = W(j,k).$$

The dual potential A^* satisfies $A^* = A$. This involution kernel satisfies the twist condition.

These class of potentials A_γ present the phenomena of phase transition on the one dimensional lattice $\Omega = \{0,1\}^{\mathbb{N}}$.

The above example came from discussions with A. Araujo.

Example. Suppose the maximizing probability is unique for A.

Suppose V is the limit of $\frac{1}{\beta} \log \phi_{\beta_n}$, where ϕ_β is the eigenfunction of the Ruelle operator for βA. Suppose V^* is the subaction for A^*.

If $\hat{\mu}_{max}$ is the natural extension of the maximizing probability $\mu_{\infty A}$, then for all (p^*, p) in the support of $\hat{\mu}_{max}$ we have the following expression taken from Proposition 10 in [12]

$$V(p) \, + \, V^*(p^*) \, = \, W(p^*, p) - \gamma \,.$$

The above expression appears in a natural way in problems in transport theory (see [136]). It is called the complementary slackness condition. In this case the involution kernel is the natural dynamically defined cost to be considered in the problem (see [100] [111] [49] [107] [61]). The subactions V and V^* define the dual Kantorovich pair associated to the problem.

We denote by I^* the deviation function for the family of Gibbs states $\mu_{\beta A^*}$, $\beta \to \infty$, according to what was described in section 3.1.3. Remember that I^* is zero over the support of the limit measure $\mu^*_{\infty \, A}$.

If (p^*, p) in the support of $\hat{\mu}_{max}$ (then, $p \in [0, 1]$ is in the support of $\mu_{\infty \, A}$ and $p^* \in \Omega$ is in the support of $\mu^*_{\infty \, A}$), then

$$V(p) = \sup_{w \in \Omega} \, W(w, p) - V^*(w) - I^*(w) - \gamma \, =$$

$$W(p^*, p) - V^*(p^*) - I^*(p^*) \, - \gamma \, = \, W(p^*, p) - V^*(p^*) \, - \gamma.$$

Remember that V^* is the $W - I^* - \gamma$ transform of V.

We point out that in problems on Ergodic Transport the following question is important. It is known that if V is calibrated for A, then, it is true that, for any z in the support of the maximizing probability $\mu_{\infty \, A}$, we have $V(z) + A(z) - m(A) = V(T(z))$. Generically (on the Holder class) on A, one can show that this equality it true just on the support of the maximizing probability. An important issue is to know if this last property is also true for the maximizing probability of A^* (see [49] for a generic result).

Chapter 6

Max Plus algebra

In this chapter we will not consider dynamical systems and invariant measures for a while in order to introduce the so called Max Plus algebra. The results of this section will be used in the next one.

6.1 Motivation

The Max-plus algebra is essentially the algebra of the real numbers with two binary operations, $a \oplus b = \max(a, b)$ and $a \otimes b = a + b$; there are many distinct motivations to introduce this mathematical object, some of them being, for example, from problems in operational research. Let us imagine, for example, a factory where some worker i needs to wait for some of his colleagues j and k to finish their tasks, which takes the times, respectively, T_{ij} and T_{ik}, in order to do his job (of the worker i) which takes the time a_i. Hence, in the factory, the total time of this work is given by $a_i + \max\{T_{ij}, T_{ik}\}$; this value can be rewritten in the max-plus notation as $a_i \otimes (T_{ij} \oplus T_{ik})$. This shows that in a larger system involving a larger number of workers and distinct steps, the max-plus algebra is a convenient and elegant way to formulate the problem of distribution of tasks in order. One can use the techniques of this theory in order to make the whole process more efficient.

From a pure mathematical point of view we can also see this algebra, for example, as the structure of the exponential growth of real functions. Given a function $h \colon \mathbb{R} \to \mathbb{R}$ we define the exponential growth of h as the limit (if, of course, the limit exists)

$$e(h) = \lim_{x \to +\infty} \frac{1}{x} \log h(x).$$

To fix ideas, consider the simple case

$$f(x) = e^{ax} \qquad \text{and} \qquad g(x) = e^{bx}.$$

From the definition it is easy to see that $e(f) = a$ and $e(g) = b$. Now, what is the exponential growth of fg or $f + g$? For $fg = e^{ax}e^{bx} = e^{a+b}$ we have that the function fg grows with a rate of $a + b$, say $e(f) + e(g)$. Hence $e(fg) = e(f) + e(g) = e(f) \otimes e(g)$. For $f + g$, we have that $f + g = e^{ax} + e^{bx} = e^{\max\{a,b\}x}(1 + o(1))$. Hence, $e(f+g) = \max\{e(f), e(g)\} = e(f) \oplus e(g)$. For this reason, it is not a surprise that this technique appears also in the setting of zero temperature limits, where we are exactly talking about comparing certain exponential growth rates.

General references on Max-Plus Algebra are [4], [5], [41], [42], [10], [43], [55] and [44].

6.2 Notation and basic properties

In this text we use
$$\bar{\mathbb{R}} = \mathbb{R} \cup -\infty$$
with the convention that $x + (-\infty) = \infty$ for any $x \in \bar{\mathbb{R}}$.

We endow this set with two operations:

$$a \oplus b = \max(a, b)$$

$$a \otimes b = a + b.$$

With this notation, the convention above is rewritten as $a \otimes -\infty = -\infty$ and we also have $a \oplus -\infty = a$, showing that $-\infty$ is the neutral element for the binary operation $\oplus$.

For the operation $\otimes$ we have that $a \otimes 0 = a + 0 = a$ for any $a \in \mathbb{R}$, showing that 0 is its neutral element.

The so called max-plus algebra is then the semi-ring over $\bar{\mathbb{R}}$ defined by the operations $\oplus$ and $\otimes$.

Some of the main properties of this algebra are listed below:

Lemma 6.1. *given a, b and c in $\bar{\mathbb{R}}$,we have*

1- *Associativity:* $a \oplus (b \oplus c) = (a \oplus b) \oplus c$ *and* $a \otimes (b \otimes c) = (a \otimes b) \otimes c$

2- *Commutativity:* $a \oplus b = b \oplus a$ *and* $a \otimes b = b \otimes a$

3- *Distributivity:* $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$

4- *Additive identity:* $a \oplus (-\infty) = (-\infty) \oplus a = a$

5- *Multiplicative identity:* $a \otimes 0 = 0 \otimes a = a$

6- *Multiplicative inverse: if $a \neq -\infty$ then there exists a unique b such that $a \otimes b = 0$*

7- *Absorbing element:* $a \otimes -\infty = -\infty \otimes a = -\infty$

8- *Idempotency of addition:* $a \oplus a = a$.

Proof. We will show just some of the above properties, leaving the others to the reader.

Distributivity, for example, follows from

$$a \otimes (b \oplus c) = a + \max(b, c) = \max(a + b, a + c) = (a \otimes b) \oplus (a \otimes c).$$

For the multiplicative inverse, just notice that for any a we can take $b = -a$ and so

$$a \otimes b = a + (-a) = 0.$$

$\square$

6.3   Linear algebra

6.3.1   Vectors

A d-dimensional vector v is an element of $\bar{\mathbb{R}}^d$, which is denoted by $v = (v_1, v_2, \ldots, v_d)$, or, as is usual, represented as a column vector

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}.$$

Given two vectors u and v in $\bar{\mathbb{R}}^d$ and $\lambda \in \bar{\mathbb{R}}^d$, we can define the sum of two vectors as

$$u \oplus v := (u_1 \oplus v_1, u_2 \oplus v_2, \ldots, u_d \oplus v_d),$$

and, the product by an scalar as

$$\lambda \otimes u := (\lambda \otimes u_1, \lambda \otimes u_2, \ldots, \lambda \otimes u_d).$$

6.3.2   Matrices

An $m \times n$ matrix A is defined as in the usual case. Given two $m \times n$ matrices A and B, we define $A \oplus B$ as the matrix whose entries are

$$(A \oplus B)_{ij} := A_{ij} \oplus B_{ij} = \max \{A_{ij}, B_{ij}\}.$$

Given $\lambda \in \bar{\mathbb{R}}$, we also define the matrix $\lambda \otimes A$ as

$$(\lambda \otimes A)_{ij} = \lambda \otimes A_{ij} = \lambda + A_{ij}.$$

>From the basic properties of the operations $\oplus$ and $\otimes$, it is not hard to see that the matrix operations above satisfies the following properties:

Lemma 6.2. *Given $m \times n$ matrices A, B and C, and some $\lambda \in \bar{\mathbb{R}}$ we have:*
There exists a matrix $[-\infty]$ such that

$$A \oplus [-\infty] = A,$$

$$A \oplus B = B \oplus A,$$

$$A \oplus (B \oplus C) = (A \oplus B) \oplus C,$$

$$\lambda \otimes A = A \otimes \lambda,$$

$$\lambda \otimes (A \oplus B) = (\lambda \otimes A) \oplus (\lambda \otimes B).$$

Given an $m \times n$ matrix A and an $n \times l$ matrix B we can define the matrix product AB as

$$(AB)_{ij} = \bigoplus_k (A_{ik} \otimes B_{kj}) = \max_k \left(A_{ik} + B_{kj} \right).$$

Lemma 6.3. *Moreover,*

$$(AB)C = A(BC),$$

$$\lambda \otimes AB = A(\lambda \otimes B) = AB \otimes \lambda.$$

If $m = n$ we say that the matrix A is a square matrix of order n. Consider the matrix

$$I_n = \begin{bmatrix} 0 & -\infty & -\infty & \dots & -\infty \\ -\infty & 0 & -\infty & \dots & -\infty \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\infty & \dots & \dots & -\infty & 0 \end{bmatrix}.$$

Then, we can show that

$$AI_n = I_n A = A,$$

for any order n matrix A.

6.3.3   Eigenvectors and eigenvalues

Now consider a $n \times n$ matrix A whose entries are elements of $\bar{\mathbb{R}}$ and a column vector v.

We define the product Av such that

$$(Ax)_i = \bigoplus_j (A_{ij} \otimes v_j) = \max_j \left(A_{ij} + v_j \right).$$

For a given $\lambda \in \bar{\mathbb{R}}$ we also define

$$\lambda v = (\lambda \otimes v_1, \ldots, \lambda \otimes v_n) = (\lambda + v_1, \ldots, \lambda + v_n).$$

In this setting it is a very natural question to look for max plus eigenvectors and eigenvalues for A, in the sense that, $Av = \lambda v$; this notation can be translated in terms of our usual operations as

$$\max_j (A_{ij} + v_j) = \lambda + v_i \quad \text{for any } i = 1, \ldots, n.$$

Example.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} \max(1, -1) \\ \max(0, 0) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

We also have

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} \max(0, 0) \\ \max(-1, 1) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

In this case we see that 1 is an eigenvalue associated with two distinct eigenvectors.

Example. Consider

$$\begin{bmatrix} -\infty & a \\ b & \infty \end{bmatrix},$$

which has eigenvalue $\lambda = (a + b)/2$ and eigenvector

$$\begin{bmatrix} x \\ x + (b - a)/2 \end{bmatrix} = x \otimes \begin{bmatrix} 0 \\ (b - a)/2 \end{bmatrix}$$

(where any choice of x is allowed).

The most important result, with respect to our purposes, is that matrices with real entries have a unique eigenvalue.

Theorem 6.4. *Let A be a $d \times d$ matrix with all entries $a_{ij} \in \mathbb{R}$; then, there exists a real number λ and a vector v, such that, $Av = \lambda v$; moreover, the eigenvalue λ is unique.*

Proof. First of all, notice that if $Mu = \mu u$, then $\lambda Mv = \lambda \otimes \mu v = (\lambda + \mu)v$. Note that

$$\lambda M = \begin{bmatrix} \lambda + m_{11} & \lambda + m_{12} & \cdots & \lambda + m_{1d} \\ \lambda + m_{21} & \lambda + m_{22} & \cdots & \lambda + m_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda + m_{d1} & \cdots & \cdots & \lambda + m_{dd} \end{bmatrix}$$

Hence, there is no loss of generality if we assume that the entries of A are all non-negative. So, we get

$$0 \leq a_{ij} \leq L.$$

Now, let us define the map $T \colon \mathbb{R}^n \to \mathbb{R}^n$ as

$$(Tx)_i = \max_j \left(A_{ij} + x_j \right) - \min_k \max_j \left(A_{kj} + x_j \right).$$

It is easy to see that the expression depends continuously on the vector x. It is also clear from the definition that $(Tx)_i \geq 0$. On the other hand, we have

$$(Tx)_i \leq \max_j \left(L + x_j \right) - \min_k \max_j \left(0 + x_j \right) =$$

$$\max_j \left(L + x_j \right) - \max_j \left(x_j \right) = L.$$

In particular, this shows that the region $\{x_j : 0 \leq x_j \leq L\}$ is mapped inside itself by T; since T is continuous, this implies (by means of Brouwer fixed point theorem) that T has at least one fixed point v. Hence,

$$v = T(v) \Rightarrow v_i = (Tv)_i = \max_j \left(A_{ij} + v_j \right) - \min_k \max_j \left(A_{kj} + v_j \right).$$

Denoting

$$\lambda = \min_k \max_j \left(A_{kj} + v_j \right),$$

then, the expression above implies

$$v = Av - \lambda \Rightarrow \lambda + v = Av \Rightarrow \lambda v = Av,$$

in the max-plus sense, as claimed.

For the uniqueness, let us assume, by contradiction, that we have two distinct eigenvalues λ and μ. In other words, there exists vectors v and u, such that,

$$Av = \lambda v \qquad \text{and} \qquad Au = \mu u.$$

Without loss of generality we can assume $\lambda < \mu$. It is possible to take a large t, such that, $tv \geq u$ (in the sense that $tv_i \geq u_i$, for each $i \in \{1, \ldots, d\}$). Then,

$$tv \oplus u = tv.$$

Hence,

$$A^n(tv \oplus u) = A^n(tv) \oplus A^n(u) = A^n(tv) \Rightarrow$$
$$tA^n(v) \oplus A^n(u) = tA^n(v) \Rightarrow t\lambda^n v \oplus \mu^n = t\lambda^n v,$$

which is equivalent to say that for any n, we have $t\lambda^n v \geq \mu^n u$, and this is a contradiction, since $\lambda < \mu$. Then, we get $\lambda = \mu$ and the max-plus eigenvalue is unique. $\qquad\square$

If we drop the hypothesis of real entries the situation is quite different. Consider, for example,

$$A = \begin{bmatrix} 1 & 1 & -\infty & -\infty \\ 1 & 1 & -\infty & -\infty \\ -\infty & -\infty & 2 & 2 \\ -\infty & -\infty & 2 & 2 \end{bmatrix}.$$

Then, it is not very hard to see that

$$A \begin{bmatrix} -\infty \\ -\infty \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\infty \\ -\infty \\ 3 \\ 3 \end{bmatrix} = 2 \otimes \begin{bmatrix} -\infty \\ -\infty \\ 1 \\ 1 \end{bmatrix},$$

and, that

$$A \begin{bmatrix} 1 \\ 1 \\ -\infty \\ -\infty \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ -\infty \\ -\infty \end{bmatrix} = 1 \otimes \begin{bmatrix} 1 \\ 1 \\ -\infty \\ -\infty \end{bmatrix}.$$

Hence, 1 and 2 are max-plus eigenvalues of A showing that the uniqueness of the eigenvalue does not hold for matrices with $-\infty$ entries.

6.4 Final remarks

There exists some variations of the max-plus algebra, one of then being the min-plus algebra, where the binary operation $\oplus$ is replaced by $a \oplus b = \min(a, b)$; this algebra is also known as tropical algebra. In this context its interesting, for example, to study the behavior of polynomials like

$$p(x) = a_0 \oplus (a_1 \otimes x) \oplus (a_2 \otimes x \otimes x) = \min\{a_0, a_1 + x, a_2 + 2x\}.$$

Its graph, for instance, is a union of the segments (some of them of infinite length) and the reader is invited to sketch this picture. The geometrical investigation of those objects is known as tropical geometry.

Chapter 7

An explicit example with all the computations

We consider the following particular case: $\Omega := \{1, 2, 3\}^{\mathbb{N}}$ and A is a non-positive potential depending only on two coordinates. For each pair (i, j) we set $A(i, j) = -\varepsilon_{ij}$. We assume that

$$\varepsilon_{11} = \varepsilon_{22} = 0,$$

and for every other pair $\varepsilon_{ij} > 0$.

Note that there are only two ergodic A-maximizing measures, namely, $\delta_{1\infty}$ and $\delta_{2\infty}$, which are the Dirac measures at $1^{\infty} := 111\ldots$ and at $2^{\infty} := 222\ldots$. The Aubry set is exactly the union of these two fixed points and each one is an irreducible component.

We remind that for each β, the unique equilibrium state is given by

$$\mu_{\beta} = H_{\beta}\nu_{\beta},$$

where H_{β} and ν_{β} are the eigenvectors for the transfer operator. Its spectrum is $e^{\mathcal{P}(\beta)}$.

We have seen in Subsection 2.2.2 that $\mathcal{P}'(\beta) = \displaystyle\int A \, d\mu_{\beta}$. This quantity is negative because A is non-positive and negative on a set of positive measure (for μ_{β} and for any β). We have seen (see the

comments after Remark 3.14) that the asymptote of the pressure is of the form

$$h_{max} + \beta.m(A),$$

where h_{max} is the residual entropy and is the entropy of the Aubry set. In our case, we have $h_{max} = 0$ and $m(A) = 0$. Then, $\lim_{\beta \to +\infty} \mathcal{P}(\beta) = 0$. The first role of the max-plus algebra seems to determine how the pressure goes to 0 as β goes to $+\infty$.

Proposition 7.1. *There exists a positive sub exponential function g and a positive real number ρ such that $\mathcal{P}(\beta) = g(\beta)e^{-\rho.\beta}$.*

Proof. First we consider any accumulation point $-\rho$ for $\frac{1}{\beta} \log \mathcal{P}(\beta)$. We shall prove that this $-\rho$ is actually unique, namely, it does not depend on the chosen subsequence.

Considering the subsequence which realizes the expression for $-\rho$, we can always extract another subsequence such that $\frac{1}{\beta} \log H_\beta$ converges. We denote by V this limit, and we have seen above that V is indeed a calibrated subaction.

For simplicity we shall write $\beta \to +\infty$ even if we consider a restricted subsequence.

Moreover, H_β and V only depend on one coordinate, and we shall write $H_\beta(i)$ or $V(i)$ for $H_\beta(x)$ and $V(x)$ with $x = i \ldots$.

The eigenfunction H_β is an eigenvector for $\mathcal{L}_\beta$ and this yields

$$\begin{aligned}
e^{\mathcal{P}(\beta)}H_\beta(1) &= e^{\beta.A(1,1)}H_\beta(1) + e^{\beta.A(2,1)}H_\beta(2) + e^{\beta.A(3,1)}H_\beta(3) \\
e^{\mathcal{P}(\beta)}H_\beta(2) &= e^{\beta.A(1,2)}H_\beta(1) + e^{\beta.A(2,2)}H_\beta(2) + e^{\beta.A(3,2)}H_\beta(3).
\end{aligned}$$

Replacing with the values for A we get the two following equations:

$$\begin{aligned}
(e^{\mathcal{P}(\beta)} - 1)H_\beta(1) &= e^{-\beta\varepsilon_{21}}H_\beta(2) + e^{-\beta\varepsilon_{31}}H_\beta(3), \quad (7.1a) \\
(e^{\mathcal{P}(\beta)} - 1)H_\beta(2) &= e^{-\beta\varepsilon_{12}}H_\beta(1) + e^{-\beta\varepsilon_{32}}H_\beta(3). \quad (7.1b)
\end{aligned}$$

Now, $\frac{\mathcal{P}(\beta)}{\beta} \to_{\beta \to +\infty} 0$, which yields $\lim_{\beta \to +\infty} \frac{e^{\mathcal{P}(\beta)} - 1}{\mathcal{P}(\beta)} = 1$ and, finally, $\lim_{\beta \to +\infty} \frac{1}{\beta} \log(e^{\mathcal{P}(\beta)} - 1) = -\rho.$

Taking $\dfrac{1}{\beta}$ log and doing $\beta \to +\infty$ in (7.1a) and (7.1b) we get

$$
\begin{aligned}
-\rho + V(1) &= \max(-\varepsilon_{21}V(2), -\varepsilon_{31}V(3)), & \text{(7.2a)} \\
-\rho + V(2) &= \max(-\varepsilon_{12}V(1), -\varepsilon_{32}V(3)), & \text{(7.2b)}
\end{aligned}
$$

which can be written under the form

$$
-\rho \otimes \begin{pmatrix} V(1) \\ V(2) \end{pmatrix} = \begin{pmatrix} -\infty & -\varepsilon_{21} & -\varepsilon_{31} \\ -\varepsilon_{12} & -\infty & -\varepsilon_{32} \end{pmatrix} \otimes \begin{pmatrix} V(1) \\ V(2) \\ V(3) \end{pmatrix}. \quad \text{(7.3)}
$$

Now, we use Theorem 4.9 to get an expression of $V(3)$ in terms of $V(1)$ and $V(2)$. Indeed, we have

$$
V(3) = \max(V(1) + h(1^\infty, 3), V(2) + h(2^\infty, 3)),
$$

where h is the Peierl's barrier and 3 means any point starting with 3. Copying the work done to get Equality (4.5), we claim that

$$
h(1^\infty, 3) = -\varepsilon_{13} \text{ and } h(2^\infty, 3) = -\varepsilon_{23}.
$$

This yields,

$$
\begin{pmatrix} V(1) \\ V(2) \\ V(3) \end{pmatrix} = \begin{pmatrix} 0 & -\infty \\ -\infty & 0 \\ -\varepsilon_{13} & -\varepsilon_{23} \end{pmatrix} \otimes \begin{pmatrix} V(1) \\ V(2) \end{pmatrix}. \quad \text{(7.4)}
$$

Merging (7.3) and (7.4) we finally get

$$
-\rho \otimes \begin{pmatrix} V(1) \\ V(2) \end{pmatrix} = \begin{pmatrix} -\varepsilon_{13} - \varepsilon_{31} & -\varepsilon_{21} \oplus (-\varepsilon_{31} - \varepsilon_{23}) \\ -\varepsilon_{12} \oplus (-\varepsilon_{32} - \varepsilon_{13}) & -\varepsilon_{23} - \varepsilon_{32} \end{pmatrix} \otimes \begin{pmatrix} V(1) \\ V(2) \end{pmatrix}.
$$

$$\text{(7.5)}$$

This last equation shows that $-\rho$ is an eigenvalue for the above matrix. We have seen (see Theorem 6.4) that such matrix admits an unique eigenvalue (but not necessarily an unique eigenvector). This shows that $\dfrac{1}{\beta} \log \mathcal{P}(\beta)$ admits a unique accumulation point as $\beta \to +\infty$, thus converges.

Setting $g(\beta) := \mathcal{P}(\beta).e^{\rho\beta}$ we get the proof of the proposition. $\qquad \square$

Remark 7.2. We point out that the max-plus algebra allows to determine the value for $-\rho$, but the function $g(\beta)$ remains unknown. For the convergence or, not, of $\mu_{\beta}, \beta \to \infty$, it is necessary a better understanding of the behavior of $g(\beta)$.
∎

Following the reasoning we did before, we get

$$-\rho = \max \begin{cases} A(1,3) + A(3,1) = -\varepsilon_{13} - \varepsilon_{31}, \\[2mm] A(3,2) + A(2,3) = -\varepsilon_{32} - \varepsilon_{23}, \\[2mm] \dfrac{A(2,1) + A(1,2)}{2} = -\dfrac{\varepsilon_{12} + \varepsilon_{21}}{2}, \\[2mm] \dfrac{A(2,1) + A(1,3) + A(3,2)}{2} = -\dfrac{\varepsilon_{21} + \varepsilon_{13} + \varepsilon_{32}}{2}, \\[2mm] \dfrac{A(1,2) + A(2,3) + A(3,1)}{2} = -\dfrac{\varepsilon_{12} + \varepsilon_{23} + \varepsilon_{31}}{2}, \\[2mm] \dfrac{A(2,3) + A(3,1) + A(1,3) + A(3,2)}{2}. \end{cases}$$

$$(7.6)$$

We emphasize that the last quantity is actually the mean value of the two first ones, and then $-\rho \geqslant \frac{A(2,3)+A(3,1)+A(1,3)+A(3,2)}{2}$ always holds, as soon as, $-\rho \geqslant A(1,3) + A(3,1)$ and $-\rho \geqslant A(3,2) + A(2,3)$ hold.

We can now finish the proof of the convergence of μ_{β}. We recall that any accumulation point must be of the form

$$\alpha.\delta_{1\infty} + (1 - \alpha)\delta_{2\infty},$$

with $\alpha \in [0, 1]$. It is thus sufficient to show that $\dfrac{\mu_{\beta}([1])}{\mu_{\beta}([2])}$ converges as $\beta \to +\infty$ to prove the convergence of μ_{β}. However, we emphasize that this is very particular to our case (two ergodic A-maximizing measures). This reasoning may not work for a more general case. Nevertheless, one of the by-product results of our proof is that for

the general case, it seems possible to determine the convergence we get here, in a similar (but more complex) way. The complexity is an issue which is due, essentially, to the large amount of possible combinatorics.

First, we get some equations for the measure ν_β. We remind that this measure is $\beta.A$-conformal. This yields,

$$
\begin{aligned}
\nu_\beta([1]) &= \nu_\beta\left(\bigsqcup_{n=1}^{\infty}[1^n2] \sqcup [1^n3]\right) \\
&= \sum_{n=1}^{+\infty}\nu_\beta([1^n2]) + \nu_\beta([1^n3]) \\
&= \sum_{n=1}^{+\infty}e^{\beta.S_n(A)(1^n2)-n\mathcal{P}(\beta)}\nu_\beta([2]) + \sum_{n=1}^{+\infty}e^{\beta.S_n(A)(1^n3)-n\mathcal{P}(\beta)}\nu_\beta([3]) \\
&= \frac{e^{-\beta\varepsilon_{12}-\mathcal{P}(\beta)}}{1-e^{-\mathcal{P}(\beta)}}\nu_\beta([2]) + \frac{e^{-\beta\varepsilon_{13}-\mathcal{P}(\beta)}}{1-e^{-\mathcal{P}(\beta)}}\nu_\beta([3]).
\end{aligned}
$$

We remind that $\nu_\beta([1]) + \nu_\beta([2]) + \nu_\beta([3]) = 1$, then, we get a linear equation between $\nu_\beta([1])$ and $\nu_\beta([2])$. Doing the same work with the cylinder $[2] = \bigsqcup_{n=1}^{+\infty}[2^n1] \sqcup [2^n3]$ we get the following system:

$$
\begin{cases}
(e^{\mathcal{P}(\beta)} - 1 + e^{-\beta.\varepsilon_{13}})\nu_\beta([1]) + (e^{-\beta.\varepsilon_{13}} - e^{-\beta.\varepsilon_{12}})\nu_\beta([2]) = e^{-\beta.\varepsilon_{13}}, \\
(e^{-\beta.\varepsilon_{23}} - e^{-\beta.\varepsilon_{21}})\nu_\beta([1]) + (e^{\mathcal{P}(\beta)} - 1 + e^{-\beta.\varepsilon_{23}})\nu_\beta([2]) = e^{-\beta.\varepsilon_{23}}.
\end{cases}
\tag{7.7}
$$

The determinant of the system is

$$
\begin{aligned}
\Delta(\beta) := (e^{\mathcal{P}(\beta)} - 1)^2 + (e^{\mathcal{P}(\beta)} - 1)(e^{-\beta.\varepsilon_{13}} + e^{-\beta.\varepsilon_{23}}) + \\
e^{-\beta(\varepsilon_{12}+\varepsilon_{23})} + e^{-\beta.(\varepsilon_{21}+\varepsilon_{13})} - e^{\beta.(\varepsilon_{12}+\varepsilon_{21})},
\end{aligned}
$$

and, we get

$$
\frac{\nu_\beta([1])}{\nu_\beta([2])} = \frac{(e^{\mathcal{P}(\beta)} - 1)e^{-\beta.\varepsilon_{13}} + e^{-\beta.(\varepsilon_{12}+\varepsilon_{23})}}{(e^{\mathcal{P}(\beta)} - 1)e^{-\beta.\varepsilon_{23}} + e^{-\beta.(\varepsilon_{21}+\varepsilon_{13})}}.
\tag{7.8}
$$

On the other hand, Equations (7.2a) and (7.2a) yield the following formula:

$$\frac{H_\beta(1)}{H_\beta(2)} = \frac{(e^{\mathcal{P}(\beta)} - 1)e^{-\beta.\varepsilon_{31}} + e^{-\beta.(\varepsilon_{21}+\varepsilon_{32})}}{(e^{\mathcal{P}(\beta)} - 1)e^{-\beta.\varepsilon_{32}} + e^{-\beta.(\varepsilon_{12}+\varepsilon_{31})}}. \tag{7.9}$$

Therefore, from Equations (7.8) and (7.9) we get

$$\frac{\mu_\beta([1])}{\mu_\beta([2])} = \frac{H_\beta(1)}{H_\beta(2)} \frac{\nu_\beta([1])}{\nu_\beta([2])}$$

$$= \frac{(e^{\mathcal{P}(\beta)} - 1)e^{-\beta.\varepsilon_{13}} + e^{-\beta.(\varepsilon_{12}+\varepsilon_{23})}}{(e^{\mathcal{P}(\beta)} - 1)e^{-\beta.\varepsilon_{23}} + e^{-\beta.(\varepsilon_{21}+\varepsilon_{13})}} \frac{(e^{\mathcal{P}(\beta)} - 1)e^{-\beta.\varepsilon_{31}} + e^{-\beta.(\varepsilon_{21}+\varepsilon_{32})}}{(e^{\mathcal{P}(\beta)} - 1)e^{-\beta.\varepsilon_{32}} + e^{-\beta.(\varepsilon_{12}+\varepsilon_{31})}}$$

$$= \frac{\left((e^{\mathcal{P}(\beta)} - 1)e^{\beta.(\varepsilon_{12}+\varepsilon_{21}-\varepsilon_{13})} + e^{\beta.(\varepsilon_{21}-\varepsilon_{23})}\right)}{\left((e^{\mathcal{P}(\beta)} - 1)e^{\beta.(\varepsilon_{12}+\varepsilon_{21}-\varepsilon_{23})} + e^{\beta.(\varepsilon_{12}-\varepsilon_{13})}\right)} \times$$

$$\frac{\left((e^{\mathcal{P}(\beta)} - 1)e^{\beta.(\varepsilon_{12}+\varepsilon_{21}-\varepsilon_{31})} + e^{\beta.(\varepsilon_{12}-\varepsilon_{32})}\right)}{\left((e^{\mathcal{P}(\beta)} - 1)e^{\beta.(\varepsilon_{12}+\varepsilon_{21}-\varepsilon_{32})} + e^{\beta.(\varepsilon_{21}-\varepsilon_{31})}\right)}. \tag{7.10}$$

Convergence will follow from the next proposition.

Proposition 7.3. *The function g admits a limit as β goes to $+\infty$.*

Proof. We remind that ν_β is the eigenmeasure for the dual transfer operator. This yields:

$$e^{\mathcal{P}(\beta)} = \int \mathcal{L}_\beta(\mathbb{1}) \, d\nu_\beta$$

$$= (1 + e^{-\beta.\varepsilon_{21}} + e^{-\beta.\varepsilon_{31}})\nu_\beta([1]) + (1 + e^{-\beta.\varepsilon_{12}} + e^{-\beta.\varepsilon_{32}})\nu_\beta([2])$$

$$+ (e^{-\beta.\varepsilon_{13}} + e^{-\beta.\varepsilon_{23}} + e^{-\beta.\varepsilon_{33}})\nu_\beta([3]). \tag{7.11}$$

Let us set

$$\begin{cases} X := e^{\mathcal{P}(\beta)}, \\ A := e^{-\beta.\varepsilon_{13}}, \\ A' := e^{-\beta.\varepsilon_{23}}, \\ B := e^{-\beta.(\varepsilon_{12}+\varepsilon_{23})}, \\ B' := e^{-\beta.(\varepsilon_{21}+\varepsilon_{13})}, \\ C := e^{-\beta.(\varepsilon_{12}+\varepsilon_{21})}, \\ a := e^{-\beta.\varepsilon_{21}} + e^{-\beta.\varepsilon_{31}}, \\ b := e^{-\beta.\varepsilon_{12}} + e^{-\beta.\varepsilon_{32}}, \\ c := e^{-\beta.\varepsilon_{13}} + e^{-\beta.\varepsilon_{23}} + e^{-\beta.\varepsilon_{33}}. \end{cases}$$

From the system (7.7), we get exact values for $\nu_\beta([1])$, $\nu_\beta([2])$ and $\nu_\beta([3])$. Replacing these values in (7.11), this yields

$$X = \frac{(1+a)(A(X-1)+B)+(1+b)(A'(X-1)+B')+c((X-1)^2-C)}{(X-1)^2+(A+A')(X-1)+B+B'-C}.$$
(7.12)

This can also be written

$$X-1 = \frac{a(A(X-1)+B)+b(A'(X-1)+B')+(c-1)((X-1)^2-C)}{(X-1)^2+(A+A')(X-1)+B+B'-C},$$

and this yields

$$(X-1)^3 + (A+A'+1-c)(X-1)^2$$
$$+(B+B'-C-a.A-b.A')(X-1)+C(c-1)-a.B-b.B' = 0.$$
(7.13)

Remember that all the terms $A, A', B, \ldots$ go exponentially fast to 0 as $\beta \to +\infty$, $X-1$ behaves like $g(\beta)e^{-\rho.\beta}$, and, moreover g is sub-exponential. We can thus use Taylor development to replace $X-1$ by $g(\beta)e^{-\rho.\beta}$, and, keep in each summand of (7.13) the largest term; larger here means that we are comparing it to other terms, if there is a sum, but also to terms of the other summands.

It is for instance clear that $(A+A'+1-c)(X-1)^2$ has as dominating term $g^2(\beta)e^{-2\rho.\beta}$, whereas $(X-1)^3$ has as dominating term $g^3(\beta)e^{-3\rho.\beta}$, which is exponentially smaller than $g^2(\beta)e^{-2\rho.\beta}$.

In the same direction, note that the term B is killed by part of $b.A'$ and similarly B' is killed by part of $a.A$. The term in $X - 1$ is actually equal to

$$e^{-\beta.(\varepsilon_{12}+\varepsilon_{23})} + e^{-\beta.(\varepsilon_{21}+\varepsilon_{13})} - e^{-\beta(\varepsilon_{12}+\varepsilon_{21})}$$
$$-e^{-\beta.(\varepsilon_{21}+\varepsilon_{13})} - e^{-\beta(\varepsilon_{13}+\varepsilon_{31})} - e^{-\beta.(\varepsilon_{12}+\varepsilon_{23})} - e^{-\beta(\varepsilon_{23}+\varepsilon_{32})}$$
$$= -e^{-\beta(\varepsilon_{12}+\varepsilon_{21})} - e^{-\beta(\varepsilon_{13}+\varepsilon_{31})} - e^{-\beta(\varepsilon_{23}+\varepsilon_{32})}.$$

Note that this term will be multiplied by $g(\beta)e^{-\rho.\beta}$ and compared to $g^2(\beta)e^{-2\rho.\beta}$. The fact that $\rho \leqslant \dfrac{\varepsilon_{12} + \varepsilon_{21}}{2}$ shows that the term $-e^{-\beta(\varepsilon_{12}+\varepsilon_{21})}$ can also be forgotten because it will furnish a quantity exponentially smaller than $g^2(\beta)e^{-2\rho.\beta}$.

The term without $(X - 1)$ is (with change of sign)

$$e^{-\beta.(\varepsilon_{12}+\varepsilon_{21})} \oplus e^{-\beta(\varepsilon_{12}+\varepsilon_{23}+\varepsilon_{21})} \oplus e^{-\beta(\varepsilon_{12}+\varepsilon_{23}+\varepsilon_{31})}$$
$$\oplus e^{-\beta(\varepsilon_{21}+\varepsilon_{12}+\varepsilon_{13})} \oplus e^{-\beta(\varepsilon_{21}+\varepsilon_{13}+\varepsilon_{32})}$$
$$= \quad e^{-\beta.(\varepsilon_{12}+\varepsilon_{21})} \oplus e^{-\beta(\varepsilon_{12}+\varepsilon_{23}+\varepsilon_{31})} \oplus e^{-\beta(\varepsilon_{21}+\varepsilon_{12}+\varepsilon_{13})}$$

from the max-plus formalism. Comparing all these terms with the term in $g^2(\beta)e^{-2\rho.\beta}$, we actually recover 7.6: comparing the terms in 2ρ and the terms in ρ, leads to compare $-\rho$ with $-(\varepsilon_{13} + \varepsilon_{31}) \oplus (-\varepsilon_{23}-\varepsilon_{32})$, and comparing the term with 2ρ, with the terms without ρ, leads to compare -2ρ with $-(\varepsilon_{12}+\varepsilon_{21})\oplus(-\varepsilon_{12}-\varepsilon_{23}-\varepsilon_{31})\oplus(-\varepsilon_{21}-\varepsilon_{31}-\varepsilon_{32})$.

Now, considering the dominating term at exponential scale in (7.13) yields an equality of the form

$$\widetilde{a}g^2(\beta) - \widetilde{b}g(\beta) - \widetilde{c} = \text{ term exponentially small,} \qquad (7.14)$$

where $\widetilde{a}$ is either 0 or 1, $\widetilde{b} \in \{0, 1, 2, 3\}$ and $\widetilde{c} \in \{0, 1, 2, 3\}$ and not all the coefficients $\widetilde{a}$, $\widetilde{b}$ and $\widetilde{c}$ are zero[1]. As g is positive and all the coefficient are not zero, we get that necessarily $\widetilde{a} = 1$. Now, considering any accumulation point G for $g(\beta)$, as β goes to $+\infty$, we get

$$G^2 - \widetilde{b}G - \widetilde{c} = 0. \qquad (7.15)$$

---

[1]because we exactly consider the dominating exponential scale.

Such equation admits all its roots in $\mathbb{R}$, and, at least one of them is non-negative. But, the key point here is that the roots form a finite set, and this set contains the set of accumulation points for $g(\beta)$ as $\beta \to +\infty$. On the other hand, g is a continuous function, thus the set of accumulation points for g is an interval. This shows that it is reduced to a single point, and then $g(\beta)$ converges as $\beta \to +\infty$. $\qquad \square$

We remind that $\mathcal{P}(\beta)$ goes to 0 as β goes to $+\infty$, and then $e^{\mathcal{P}(\beta)} - 1$ behaves like $g(\beta)e^{-\beta \cdot \rho}$. We replace this in Equation (7.10). The final expression is thus:

$$
\frac{\mu_\beta([1])}{\mu_\beta([2])} = \frac{\left(g(\beta)e^{\beta \cdot (\varepsilon_{12} + \varepsilon_{21} - \varepsilon_{13} - \rho)} + e^{\beta \cdot (\varepsilon_{21} - \varepsilon_{23})} \right)}{\left(g(\beta)e^{\beta \cdot (\varepsilon_{12} + \varepsilon_{21} - \varepsilon_{23} - \rho)} + e^{\beta \cdot (\varepsilon_{12} - \varepsilon_{13})} \right)}
$$

$$
\times \frac{\left(g(\beta)e^{\beta \cdot (\varepsilon_{12} + \varepsilon_{21} - \varepsilon_{31} - \rho)} + e^{\beta \cdot (\varepsilon_{12} - \varepsilon_{32})} \right)}{\left(g(\beta)e^{\beta \cdot (\varepsilon_{12} + \varepsilon_{21} - \varepsilon_{32} - \rho)} + e^{\beta \cdot (\varepsilon_{21} - \varepsilon_{31})} \right)}.
$$

$$(7.16)$$

We know by Proposition 7.3 that $g(\beta)$ converges to a nonnegative limit, then Equality (7.16) has a limit if $\beta \to +\infty$ (in $[0, +\infty]$). If the limit is 0, this means that μ_β goes to $\delta_{2\infty}$, if the limit is $+\infty$ this means that μ_β goes to $\delta_{1\infty}$, if it is equal to $\alpha \in]0, +\infty[$, then μ_β goes to $\dfrac{\alpha}{\alpha + 1}\delta_1 + \dfrac{1}{\alpha + 1}\delta_{2\infty}$.

Bibliography

[1] S. Addas-Zanata and F. Tal, Support of maximizing measures for typical C^0 dynamics on compact manifolds. Discrete Contin. Dyn. Syst. 26 (2010), no. 3

[2] S. Addas-Zanata and F. Tal, Maximizing measures for endomorphisms of the circle. Nonlinearity 21 (2008), no. 10, 2347–2359.

[3] S. Addas-Zanata and F. Tal, On maximizing measures of homeomorphisms on compact manifolds. Fund. Math. 200 (2008), no. 2, 145–159

[4] M. Akian, R. Babat and A. Guterman, Handbook of Linear Algebra, Leslie Hogben (Editor), Section 25, 25–1 to 25–17, Chapmann and Hall, (2006)

[5] M. Akian, S. Gaubert, and A. Guterman, Linear independence over tropical semirings and beyond. (English summary) Tropical and idempotent Mathematics, Contemp. Math., 495, Amer. Math. Soc., Providence, RI, (2009)

[6] V. Anagnostopoulou, and O. Jenkinson, Which beta-shifts have a largest invariant measure? J. Lond. Math. Soc. (2) 79 (2009), no. 2, 445–464

[7] V. Anagnostopoulou, K. Diaz-Ordazand, O. Jenkinson and C. Richard, Entrance time functions for flat spot maps. Nonlinearity 23 (2010), no. 6, 147–1494

[8] V. Anagnostopoulou, K. Diaz-Ordazand, O. Jenkinson and C. Richard, Anagnostopoulou, V.; Diaz-Ordaz, K.; Jenkinson, O.; Richard, C. The flat spot standard family: variation of the entrance time median. Dyn. Syst. 27 (2012), no. 1, 29–43.

[9] V. Anagnostopoulou, K. Diaz-Ordazand, O. Jenkinson and C. Richard, Sturmian maximizing measures for the piecewise-linear cosine family. Bull. Braz. Math. Soc. (N.S.) 43 (2012), no. 2, 285–302.

[10] N. Bacaer, Convergence of numerical methods and parameter dependence of min-plus eigenvalue problems, Frenkel-Kontorova models and homogenization of Hamilton- Jacobi equations, ESAIM: Math. Model. Numer. Anal. 35 (2001) 1185–1195.

[11] V. Baladi, Positive Transfer Operators and decay of correlations, World Scientific

[12] A. T. Baraviera, A. O. Lopes and P. Thieullen, A large deviation principle for equilibrium states of Hölder potenciais: the zero temperature case, *Stochastics and Dynamics* **6** (2006), 77–96.

[13] A. T. Baraviera, L. Cioletti, A. O. Lopes, J. Mohr and R. R. Souza, On the general one-dimensional XY model: positive and zero temperature, selection and non-selection. *Rev.Math. Phis.* 23 (2011), no. 10, 1063–1113, 82Bxx.

[14] R. Baman, J. Kiwi, J. Rivera-Letelier and R. Urzão, On the topology of solenoidal attractors of the cylinder. Ann. Inst. H. Poincaré Anal. Non Linéaire 23 (2006), no. 2, 209–236

[15] 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)

[16] A. Baraviera, R. Leplaideur and A. O. Lopes, The potential point of view for Renormalization, Stoch. and Dynamics, Vol 12. N 4, (2012) 1250005(1-34)

[17] A. T. Baraviera, A. O. Lopes and J. K. Mengue, On the selection of subaction and measure for a subclass of potentials defined by P. Walters. *Ergodic Theory and Dynamical Systems*, Available on CJO2012.

[18] R. Bartle, The elements of integration, John Wiley (1966)

[19] R. Bissacot and E. Garibaldi, Weak KAM methods and ergodic optimal problems for countable Markov shifts, *Bull. Braz. Math. Soc.* 41, N 3, 321-338, 210.

[20] R. Bissacot and R. Freire Jr., On the existence of maximizing measures for irreducible countable Markov shifts: a dynamical proof, to appear in Erg. Theo. and Dyn. Systems

[21] J. Bochi and A. Navas, A geometric path from zero Lyapunov exponents to rotation cocycles, Arxiv (2011)

[22] T. Bousch Le poisson n'a pas d'arantes, Ann. Inst. H. Poincaré Probab. Statist. 36 (2000), no. 4, 489–508.

[23] T. Bousch, Le poisson n'a pas d'arêtes, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, **36** (2000), 489–508.

[24] T. Bousch, Un lemme de Mão bilateral, C. R. Math. Acad. Sci. Paris 335 (2002), no. 6, 533–536

[25] T. Bousch and O. Jenkinson, Cohomology classes of dynamically non-negative C^k functions. Invent. Math. 148 (2002), no. 1, 207–217.

[26] T. Bousch and J. Mairesse Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture. J. Amer. Math. Soc. 15 (2002), no. 1, 77–111

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

[28] T. Bousch, Le lemme de Mão-Conze-Guivarc'h pour les systémes amphidynamiques rectifiables, Ann. Fac. Sci. Toulouse Math. (6) 20 (2011), no. 1, 1–14

[29] R. Bowen, Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms, Lecture Notes in Math., vol. 470. Springer, Berlin (1975)

[30] X. Bressaud e A. Quas, Rate of approximation of minimizing measures, *Nonlinearity* **20** no. 4, (2007), 845–853.

[31] F. Branco, Sub-actions and maximizing measures for one-dimensional transformations with a critical point. Discrete Contin. Dyn. Syst. 17 (2007), no. 2, 271–280.

[32] S. Branton, Sub-actions for Young towers. Discrete Contin. Dyn. Syst. 22 (2008), no. 3, 541–556.

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

[34] J. Brémont, Finite flowers and maximizing measures for generic Lipschitz functions on the circle. Nonlinearity 19 (2006), no. 4, 813–828.

[35] J. Brémont, Entropy and maximizing measures of generic continuous functions. C. R. Math. Acad. Sci. Paris 346 (2008), no. 3-4, 199–201.

[36] J. Brémont and S. Seuret, The singularity spectrum of the fish's boundary. Ergodic Theory Dynam. Systems 28 (2008), no. 1, 49–66.

[37] J. Brémont, Dynamics of injective quasi-contractions. Ergodic Theory Dynam. Systems 26 (2006), no. 1, 1944.

[38] X. Bressaud and A. Quas, Rate of approximation of minimizing measures. Nonlinearity 20 (2007), no. 4, 845–853.

[39] P. Butkovic, Max-linear Systems: Theory and Algorithms, Springer Monographs in Mathematics, Springer-Verlag

[40] D. Collier and I. D. Morris, Approximating the maximum ergodic average via periodic orbits. Ergodic Theory Dynam. Systems 28 (2008), no. 4, 1081–1090

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

[42] W. Chou and R. Grifiths, Ground states of one-dimensional systems using effective potentials, Phys. Rev. B 34 (1986) 6219–6234.

[43] M. Chung, Eigenvalues and eigenvectors in the Max-Plus Algebra, Master's Thesis, University of Colorado, Denver (1995)

[44] M. C. Concordel, Periodic homogenization of Hamilton-Jacobi equations: Additive eigenvalues and variational formula, Indiana Univ. Math. J. 45 (1996) 1095–1118.

[45] 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.* 31 (2011), no. 4, 1109–1161

[46] J. R. Chazottes and M. Hochman, On the zero-temperature limit of Gibbs states, *Commun. Math. Phys.* vol 297, N1, 2010.

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

[48] 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.

[49] G. Contreras, A. Lopes and E. Oliveira, Ergodic Transport Theory, periodic maximizing probabilities and the twist condition, to appear in *Modeling, Optimization, Dynamics and Bioeconomy*, Springer Proceedings in Mathematics, Edit. David Zilberman and Alberto Pinto.

[50] J. P. Conze and Y. Guivarc'h, Croissance des sommes ergodiques et principe variationnel, manuscript circa (1993).

[51] A. Davie, M. Urbanski annd A. Zdunik, Maximizing measures of metrizable non-compact spaces. Proc. Edinb. Math. Soc. (2) 50 (2007), no. 1, 123–151

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

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

[54] R. Exel and A. Lopes, C^*-Algebras, approximately proper equivalence relations and Thermodynamic Formalism, Erg Theo and Dyn Sys, 1051–1082, Vol 24, t (2004).

[55] K. G. Farlow, Max-Plus Algebra, Master's Thesis, Virginia Polytechnic Institute and State University (2009).

[56] 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).

[57] P. Fernandez, Medida e Integração, (1976) IMPA

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

[59] 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)

[60] E. Garibaldi and A. Lopes, Functions for relative maximization, Dynamical Systems, v. 22, 511–528, (2007)

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

[62] E. Garibaldi, A. Lopes and P. Thieullen, On calibrated and separating subactions, Bull. of the Bras. Math. Soc. Vol 40, 577–602, (4)(2009)

[63] E. Garibaldi and Ph. Thieullen, Minimizing orbits in the discrete Aubry-Mather model, Nonlinearity, 24 (2011), no. 2, 563–611

[64] E. Garibaldi and Ph. Thieullen, Description of some ground states by Puiseux technics, *Journ. of Statis. Phys*, 146, no. 1, 125–180, (2012)

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

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

[67] D. A. Gomes, A. O. Lopes and J. Mohr, The Mather measure and a large deviation principle for the entropy penalized method. Commun. Contemp. Math. 13 (2011), no.2, 235–268.

[68] A. Grant, Finding Optimal Orbits of Chaotic Systems, Phd thesis, Univ. of Maryland (2005)

[69] F. Hofbauer, Examples for the nonuniqueness of the equilibrium state. Trans. Amer. Math. Soc. 228, no. 223–241 (1977)

[70] B. R. Hunt, Maximum local Lyapunov dimension bounds the box dimension of chaotic attractors, Nonlinearity 9 (1996), no. 4, 845–852.

[71] B. R. Hunt and Y. Guocheng, Optimal orbits of hyperbolic systems. *Nonlinearity* **12**, (1999), 1207–1224.

[72] B. Hunt B and E. Ott, Optimal periodic orbits of chaotic systems occur at low period, Phys. Rev. E 54 32837 (1996)

[73] E. Harriss and O. Jenkinson, Flattening functions on flowers. Ergodic Theory Dynam. Systems 27 (2007), no. 6, 1865–1886

[74] Godofredo Iommi, Ergodic Optimization for Renewal Type Shifts, Monatshefte first Mathematik, Volume 150, Number 2 (2007), 91–95,

[75] O. Jenkinson. Ergodic optimization, Discrete and Continuous Dynamical Systems, Series A, V. 15, 197–224, (2006).

[76] O. Jenkinson and J. Steel, Majorization of invariant measures for orientation-reversing maps. Ergodic Theory Dynam. Systems 30 (2010), no. 5, 1471–1483.

[77] O. Jenkinson and I. D. Morris, Lyapunov optimizing measures for C 1 expanding maps of the circle. Ergodic Theory Dynam. Systems 28 (2008), no. 6, 1849–1860

[78] O. Jenkinson, A partial order on ×2 -invariant measures. Math. Res. Lett. 15 (2008), no. 5, 893–900.

[79] O. Jenkinson, Optimization and majorization of invariant measures. Electron. Res. Announc. Amer. Math. Soc. 13 (2007), 112

[80] O. Jenkinson, Every ergodic measure is uniquely maximizing. Discrete Contin. Dyn. Syst. 16 (2006), no. 2, 383–392.

[81] O. Jenkinson, Maximum hitting frequency and fastest mean return time. Nonlinearity 18 (2005), no. 5, 2305–2321

[82] O. Jenkinson, Directional entropy of rotation sets. C. R. Acad. Sci. Paris S. I Math. 332 (2001), no. 10, 921–926.

[83] O. Jenkinson, Geometric barycentres of invariant measures for circle maps. Ergodic Theory Dynam. Systems 21 (2001), no. 2, 511532.

[84] O. Jenkinson, R. D. Mauldin and M. Urbanski, Zero Temperature Limits of Gibbs-Equilibrium States for Countable Alphabet Subshifts of Finite Type, Journ. of Statis. Physics, Volume 119, Numbers 3–4 (2005)

[85] T. Jordan and M. Rams, Multifractal analysis of weak Gibbs measures for non-uniformly expanding C 1 maps. Ergodic Theory Dynam. Systems 31 (2011), no. 1, 143–164.

[86] S. Karlin and H. Taylor, A First Course in Stochastic Processes, Academic Press.

[87] G. Keller, Gibbs States in Ergodic Theory, Cambrige Press, (1998).

[88] T. Kempton, Zero Temperature Limits of Gibbs Equilibrium States for Countable Markov Shifts, *J. Stat Phys* 143, 795–806, (2011).

[89] U. Krengel, *Ergodic theorems*, (1985), De Gruyter Publ.

[90] O. Lanford, Entropy and Equilibrium States in Classical Statistical Mechanics. Statistical mechanics and mathematical problems. Battelle Rencontres, Seattle, Wash., 1971. Lecture Notes in Physics, 20. Springer-Verlag, Berlin-New York, 1973.

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

[92] R. Leplaideur, Totally dissipative measures for the shift and conformal s - finite measures for the stable holonomies. Bull. Braz. Math. Soc. 41 (2010), no. 1, 136.

[93] R. Leplaideur, Flatness is a criterion for selection of maximizing measures. J. Stat. Phys. 147 (2012), no. 4, 728–757.

[94] R. Leplaideur, Thermodynamic formalism for a family of non-uniformly hyperbolic horseshoes and the unstable Jacobian. Ergodic Theory Dynam. Systems 31 no. 2, 423–447, (2011),

[95] D. Lind and B. Marcus. *An introduction to Symbolic Dynamics and Coding.* Cambridge University Press, 1995.

[96] 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).

[97] A. O. Lopes. Thermodynamic Formalism, Maximizing Probabilities and Large Deviations. Work in progress. Lecture Notes - Dynamique en Cornouaille (2012).

[98] 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).

[99] A. Lopes and J. Mengue, Duality Theorems in Ergodic Transport, Journal of Statistical Physics. Vol 149, issue 5, pp 921–942 (2012).

[100] A. O. Lopes, E. R. Oliveira and Ph. Thieullen, The dual potential, the involution kernel and transport in ergodic optimization, *preprint* Arxiv.

[101] A. Lopes and P. Thieullen, Eigenfunctions of the Laplacian and associated Ruelle operator, Nonlinearity, Volume 21, Number 10, pp-2239–2254 (2008)

[102] A. Lopes, V. Rosas and R. Ruggiero Cohomology and Subcohomology for expansive geodesic flows, Discrete and Continous Dynamical Systems Volume: 17, N. 2, pp 403–422 (2007).

[103] A. Lopes and P. Thieullen, Subactions for Anosov Diffeomorphisms, Asterisque, volume 287, *G*eometric Methods in Dynamics (II) - pp 135–146 (2003).

[104] A. Lopes and P. Thieullen, Subactions for Anosov Flows, Erg Theo and Dyn Syst vol 25 Issue 2, pp 605–628 (2005)

[105] A. Lopes and P. Thieullen, Mather measures and the Bowen-Series transformation, Annal Inst Henry Poincare - Anal non-linear, 663–682, vol 5, (2006)

[106] A. O. Lopes and E. Oliveira, Entropy and variational principles for holonomic probabilities of IFS, *Disc. and Cont. Dyn. Systems* series A, vol 23, N 3, 937–955, (2009).

[107] A. O. Lopes and E. R. Oliveira, On the thin boundary of the fat attractor, preprint (2012).

[108] 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).

[109] A. Lopes and J. Mengue, Selection of measure and a Large Deviation Principle for the general one-dimensional XY model, preprint, UFRGS, (2011).

[110] A. Lopes, J. K. Mengue, J. Mohr and R. R. Souza Entropy and Variational Principle for one-dimensional Lattice Systems with a general a-priori measure: positive and zero temperature, preprint (2012).

[111] A. Lopes, E. R. Oliveira and D. Smania, Ergodic Transport Theory and Piecewise Analytic Subactions for Analytic Dynamics, Bull. of the Braz. Math Soc. Vol 43 (3) 467–512 (2012)

[112] A. Lopes and S. Lopes, Uma introdução aos Processos Estocásticos para estudantes de Matemática, a ser publicado.

[113] E. L. Lima, Curso de Análise , Vol II, IMPA (1981).

[114] R. Mañé. Generic properties and problems of minimizing measures of Lagrangian systems, *Nonlinearity*, Vol 9, 273–310, (1996).

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

[116] I. D. Morris, Entropy for Zero-Temperature Limits of Gibbs-Equilibrium States for Countable-Alphabet Subshifts of Finite Type, Journ. of Statis. Physics, Volume 126, Number 2 (2007), 315–324.

[117] I. D. Morris, A rapidly-converging lower bound for the joint spectral radius via multiplicative ergodic theory. Adv. Math. 225 (2010), no. 6, 3425–3445.

[118] I. D. Morris, Ergodic optimization for generic continuous functions. Discrete Contin. Dyn. Syst. 27 (2010), no. 1, 383–388.

[119] I. D. Morris, The Mão-Conze-Guivarc'h lemma for intermittent maps of the circle. Ergodic Theory Dynam. Systems 29 (2009), no. 5, 1603–1611.

[120] I. D. Morris, Maximizing measures of generic Hölder functions have zero entropy. Nonlinearity 21 (2008), no. 5, 993–1000.

[121] I. D. Morris, A sufficient condition for the subordination principle in ergodic optimization. Bull. Lond. Math. Soc. 39 (2007), no. 2, 214–220.

[122] I. D. Morris, Entropy for zero-temperature limits of Gibbs-equilibrium states for countable-alphabet subshifts of finite type. J. Stat. Phys. 126 (2007), no. 2, 315–324.

[123] K. Oliveira and M. Viana, Thermodynamical formalism for robust classes of potentials and non-uniformly hyperbolic maps. Ergodic Theory and Dynamical Systems 28-02 (2008), pp 501–533.

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

[125] V. Pit, Invariant relations for the Bowen-Series transform, Conform. Geom. Dyn. 16 (2012), 103–123.

[126] M. Pollicott and M. Yuri, Dynamical systems and Ergodic Theory, Cambrige Press, (1998).

[127] M. Pollicott and R. Sharp, Livsic theorems, maximizing measures and the stable norm. Dyn. Syst. 19 (2004), no. 1, 75–88.

[128] A. Quas and J. Siefken, Ergodic optimization of supercontinuous functions on shift spaces, Ergodic Theory and Dynamical Systems / Volume 32 / Issue 06 (2012), 2071–2082.

[129] Y. Sinai, Probabilty Theory, an introductory course, Springer Veralag.

[130] O. Sarig, Thermodynamic formalism for countable Markov shifts, *Ergodic Theory and Dynamical Systems* 19, 1565–1593, (1999).

[131] O. Sarig, Lecture notes on thermodynamic formalism for topological Markov shifts, *Penn State*, (2009).

[132] E. A. da Silva, R. R. da Silva and R. R. Souza The Analyticity of a Generalized Ruelle's Operator, to appear Bull. Braz. Math. Soc.

[133] R. R. Souza, Sub-actions for weakly hyperbolic one-dimensional systems, *Dynamical Systems* 18 **(2)**, 165–179, (2003).

[134] F. Spitzer. A Variational characterization of finite Markov chains. *The An-nals of Mathematical Statistics.* (43): N.1 303–307, (1972).

[135] J. Steel, Concave unimodal maps have no majorisation relations between their ergodic measures. Proc. Amer. Math. Soc. 139 (2011), no. 7, 2553–2558

[136] C. Villani, Topics in optimal transportation, AMS, Providence, (2003).

[137] P. Walters, Ergodic Theory, Springer Verlag.

[138] G. Yuan and B. Hunt, Optimal orbits of hyperbolic systems, Nonlinearity 12 (1999), no. 4, 1207–1224.