Negative Entropy, Zero temperature and stationary Markov chains on the interval.

A. O. Lopes*, J. Mohr* , R. R. Souza* *and Ph. Thieullen†

4th July 2008

Abstract

We analyze some properties of maximizing stationary Markov probabilities on the Bernoulli space $[0,1]^{\mathbb{N}}$, which means we consider stationary Markov chains with state space the interval $S=[0,1]$. More precisely, we consider ergodic optimization for a continuous potential A, where $A:[0,1]^{\mathbb{N}} \to \mathbb{R}$ which depends only on the two first coordinates of $[0,1]^{\mathbb{N}}$. We are interested in finding stationary Markov probabilities μ_∞ on $[0,1]^{\mathbb{N}}$ that maximize the value $\int A d\mu$, among all stationary Markov probabilities μ on $[0,1]^{\mathbb{N}}$. This problem correspond in Statistical Mechanics to the zero temperature case for the interaction described by the potential A. The main purpose of this paper is to show, under the hypothesis of uniqueness of the maximizing probability, a Large Deviation Principle for a family of absolutely continuous Markov probabilities μ_β which weakly converges to μ_∞. The probabilities μ_β are obtained via an information we get from a Perron operator and they satisfy a variational principle similar to the pressure in Thermodynamic Formalism. As the potential A depends only on the first two coordinates, instead of the probability μ on $[0,1]^{\mathbb{N}}$, we can consider its projection ν on $[0,1]^2$. We look at the problem in both ways. If μ_∞ is the maximizing probability on $[0,1]^{\mathbb{N}}$, we also have that its projection ν_∞ is maximizing for A. The hypothesis about stationary on the maximization problem can also be seen as a transhipment problem. Under the hypothesis of A being C^2 and the twist condition, that is, $\frac{\partial^2 A}{\partial_x \partial_y}(x,y) \neq 0$, for all $(x,y) \in [0,1]^2$, we show the graph property of the maximizing probability ν on $[0,1]^2$. Moreover, the graph is monotonous. We also show that, in the sense of Mañé, the maximizing probability is unique. Finally, we exhibit a separating sub-action for A.

*Instituto de Matemática, UFRGS, 91509-900 Porto Alegre, Brasil. Partially supported by CNPq, PRONEX – Sistemas Dinâmicos, Instituto do Milênio, and beneficiary of CAPES financial support.

†Institut de Mathématiques, Université Bordeaux 1, F-33405 Talence, France.

1 Introduction

We are interested here in a problem in ergodic optimization [Jen1] [CG] [CLT] [Mo]: we analyze some properties of maximizing stationary Markov probabilities on the Bernoulli space $[0,1]^{\mathbb{N}}$, which means we consider stationary Markov chains with state space $S = [0,1]$, and search for those which maximize the integral of functions $A : [0,1]^{\mathbb{N}} \to \mathbb{R}$, called potentials or observables.

We denote by $\mathbf{x} = (x_1, x_2, ...)$ a point in $[0,1]^{\mathbb{N}}$. We consider a continuous potential $A : [0,1]^{\mathbb{N}} \to \mathbb{R}$ which depends only on the two first coordinates of $[0,1]^{\mathbb{N}}$. Therefore, we can define $\tilde{A} : [0,1]^2 \to \mathbb{R}$, as $\tilde{A}(x_1, x_2) = A(\mathbf{x})$, where $\mathbf{x}$ is any point in $[0,1]^{\mathbb{N}}$ which has x_1 and x_2 as its two first coordinates. We will drop the symbol $\tilde{}$ and the context will show if we are considering a potential in $[0,1]^2$ or in $[0,1]^{\mathbb{N}}$.

We are interested in finding stationary Markov probabilities μ_∞ on the Borel sets of $[0,1]^{\mathbb{N}}$ that maximize the value

$$\int A(x_1, x_2)\, d\mu(\mathbf{x}),$$

among all stationary Markov probabilities μ on $[0,1]^{\mathbb{N}}$.

By a stationary Markov probability μ we mean any measure obtained from an initial probability θ on $[0,1]$, and a Markovian transition Kernel $P_x(dy) = P(x, dy)$, where θ is invariant for the kernel defined by P. The sigma-algebra we consider in $[0,1]^{\mathbb{N}}$ is the one generated by the cylinders. In the next section we will present precise definitions.

The maximizing probabilities μ_∞, by definition, are invariant for the shift, but most of the time they are not positive in all open sets on $[0,1]^{\mathbb{N}}$.

We present an entropy penalized method (see [GV] for the case of Mather measures) designed to approximate a maximizing probability μ_∞ by (absolutely continuous) stationary Markov probabilities μ_β, $\beta > 0$, obtained from $\theta_\beta(x)$ and $P_\beta(x, y)$ which are continuous functions. The functions θ_β and P_β are obtained from the eigenfunctions and the eigenvalue of a pair of Perron operator (we consider the operators $\varphi \to \mathcal{L}_\beta \varphi(\cdot) = \int e^{\beta\, A(x,\cdot)}\, \varphi(x) dx$ and $\varphi \to \bar{\mathcal{L}}_\beta \varphi(\cdot) = \int e^{\beta\, A(\cdot,y)}\, \varphi(y) dy$ and we use Krein-Ruthman Theorem) in an analogous way as the case described by F. Spitzer in [Sp] for the Bernoulli space $\Omega = \{1, 2, .., d\}^{\mathbb{N}}$ (see also [PP]).

We will show a large deviation principle for the sequence $\{\mu_\beta\}$ which converges to μ_∞ when $\beta \to \infty$. The large deviation principle give us important information on the rate of such convergence.

When the state space is the closed unit interval $[0,1]$, therefore, not countable, strange properties can occur: the natural variational problem of pressure deals with a negative entropy, namely, we have to consider the entropy penalized concept. Negative entropies appear in a natural way when we deal with a continuous state space (see [Ju] for mathematical results and also applications to Information Theory). In physical problems they occur when the spins are in a continuum space (see for instance [Lu] [Cv] [Ni] [RRS] [W]).

In Physics the negative entropies are controversial, but we quote [BBNg]: "The negativity of the entropy change is counterintuitive, but does not violate thermodynamics as one would be inclined to conclude at first."

Our result is similar to [BLT] which considers the states space $S = \{1, 2, .., d\}$ and [GLM] which consider the entropy penalized method for Mather measures [CI] [Fathi].

The main motivation of our paper are the results of [Gom] [GV] [BLT] [GLM] and [GL].

In a certain extent, the problem we consider here can be analyzed just by considering probabilities ν on $[0,1] \times [0,1]$ defined by

$$\nu(\,[a_1, a_2] \times [b_1, b_2]\,) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} \theta(dx_1) P(x_1, dx_2),$$

instead of probabilities μ on $[0,1]^{\mathbb{N}}$ defined by corresponding θ and the markovian kernel $P_x(dy)$. We say that ν is the projection of μ on $[0,1]\times[0,1]$.

From the point of view of Statistical Mechanics we are analyzing a system of neighborhood interactions describe by $A(x, y)$ at temperature zero, where the spin x takes values on $[0, 1]$. This is another point of view for the meaning of the concept of maximizing probability for A. A well known example is when $A(x, y) = x\,y$, and $x, y \in [-1, 1]$ (see [Th] for references), which can be analyzed using the methods described here via change of coordinates. In the so called XY spin model, we have $A(x, y) = \cos(x - y)$, where $x, y \in (0, 2\pi]$ (see [V] [Pe] [Ta]). When there is magnetic term one could consider, for instance, $A(x, y) = \cos(x-y) + l\,\cos(x)$, where l is constant [RRS]. Another set of examples appear in [A] section 9.4.: the continuous Ising model can be described by $A(x, y) = \frac{1}{4}(x-y)^2 + \frac{m^2}{2}\,x^2$. The fist term $\frac{1}{2}(x-y)^2$ describes interaction and the second term $\frac{m^2}{2}\,x^2$ describes the magnetic term. More generally one could consider $A(x, y) = \frac{1}{2}(x - y)^2 + f(x)$, where f describes the magnetic term. We show, among other things, that for such model, given a generic f (in the sense of Mañé [Man]), the maximizing probability for A is unique. In this way, for example, considering fixed the term $\frac{1}{2}(x - y)^2$, then, for a dense set of f, we have that the zero-temperature state for $A(x, y) = \frac{1}{2}(x - y)^2 + f(x)$ is unique (see Remark 2).

The so called spherical model (see [RRS]) is also suitable for application of the concepts we describe here.

We left to the physicists the question: the variational problem of pressure (see section 3) for the potential $\beta\,A$ (where $\beta = \frac{1}{T} > 0$ is large and T is temperature), which deals with a negative entropy, is just a device for getting a nice approximation μ_β of the maximizing probability μ_∞ for A, or if μ_β has the physical meaning of being a Gibbs state for finite temperature. Finally, consider the cost $A : [0,1] \times [0,1] \to \mathbb{R}$, and the problem of maximizing $\int A(x,y)\, d\nu(x,y)$, among probabilities ν over $[0,1] \times [0,1]$ (which can be disintegrated as $\nu(dx,dy) = \theta(dx)P(x,dy)$) with the property of having the same marginals in the x and y coordinates. We refer the reader to [Ra] for a broad description of the Monge-Kantorovich mass transport problem and the Kantorovich-Rubinstein mass transhipment problem. We consider here a special case of such problem. The Kantorovich-Rubinstein mass transhipment problem with variable probability θ on $[0,1]$ can be analyzed with the methods presented here (see remark after lemma 1 in section 2. The probability ν_∞ (on $[0,1]^2$) which is the projection of the maximizing Markov stationary probability μ_∞ for A is the solution of the problem. In this way we obtain a robust method (the LDP is true) to approximate the probability ν_∞, which is solution of the mass transhipment problem, via the entropy penalized method.

Under the twist hypothesis, that is $\frac{\partial^2 A}{\partial_x \partial_y}(x,y) \neq 0$, for all $(x,y) \in [0,1]^2$, we show that the probability ν_∞ on $[0,1]^2$ is a graph. We point out that the existence of a separating sub-action (see section 6) should be of some help in mass transhipment problems.

The twist condition is essential in Aubry Theory for twist maps [Ban] [Go]. It corresponds, in the Mather Theory, to the hypothesis of convexity of the lagrangian [Mat] [CI] [Fathi] [Man]. It is also considered in discrete time for optimization problems as in [Ba] [Mi]. Here, several results can be obtained without it. But, for getting results like the graph property, they are necessary.

In section 2 we present the basic definitions. In section 3 we introduce the Perron operator, the entropy penalized concept and we consider the associated variational problem. In section 4, under the hypothesis of A being C^2 and the twist condition, we show the graph property of the maximizing probability. We also show that for the potential A, in the generic sense of Mañé (see [Man] [BC] [CI] [CLT]), the maximizing probability on $[0,1]^2$ is unique. We get the same results for calibrated sub-actions. In section 5, we present the deviation function I and show the L.D.P.. In section 6, we show the monotonicity of the graph and we exhibit a separating sub-action.

All results presented here can be easily extended to Markov Chains with state space $[0,1]^2$, or, to more general potentials depending on a finite number of coordinates in $[0,1]^{\mathbb{N}}$, that is, to A of the form $A(x_1, x_2, ..., x_n)$, $A : [0,1]^n \to \mathbb{R}$.

We would like to thanks Alexandre Baraviera and Ana Ribeiro-Teixeira for references and interesting conversations on the subject of the paper.

1.1 Main results

Next we will give some definitions in order to state the main results of this work.

$[0,1]^{\mathbb{N}}$ can be endowed with the product topology, and then $[0,1]^{\mathbb{N}}$ becames a compact metrizable topological space. We will define a distance in $[0,1]^{\mathbb{N}}$ by

$$d(\mathbf{x},\mathbf{y}) = \sum_{j\geq 1} \frac{|x_j - y_j|}{2^j}.$$

Definition 1. (a) the shift map in $[0,1]^{\mathbb{N}}$ is defined as $\sigma((x_1, x_2, ...)) = (x_2, x_3, ...)$.

(b) Let $A_1, A_2, ..., A_k$ be non degenerated intervals of $[0,1]$. We call a cylinder of size k the subset of $\mathbb{R}^k$ given by $A_1 \times A_2 \times ... \times A_k$, and we denote it by $A_1....A_k$.

(c) Let $\mathcal{M}_{[0,1]^{\mathbb{N}}}$ be the set of probabilities in the Borel sets of $[0,1]^{\mathbb{N}}$. We define the set of holonomic measures in $\mathcal{M}_{[0,1]^{\mathbb{N}}}$ as

$$\mathcal{M}_0 := \left\{ \mu \in \mathcal{M}_{[0,1]^{\mathbb{N}}} : \int (f(x_1) - f(x_2))\, d\mu(\mathbf{x}) = 0, \quad \forall f \in C([0,1]) \right\} .$$

Remark: (i) A cylinder can also be viewed as a subset of $[0,1]^{\mathbb{N}}$: in this case, we have

$$A_1....A_k = \left\{ \mathbf{x} \in [0,1]^{\mathbb{N}} \ : \ x_i \in A_i, \ \forall\, 1 \leq i \leq i \right\} .$$

(ii) $\mathcal{M}_0$ contain all σ-invariant measures. This is a consequence of the fact that invariant measures for a transformation defined in a compact metric space can be characterized by the measures μ such that $\int f d\mu = \int (f \circ \sigma)\, d\mu$ for all continuous functions defined in $[0,1]^{\mathbb{N}}$ and taking values in $\mathbb{R}$. Note that the set of σ-invariant measures is a proper subset of $\mathcal{M}_0$.

We can also consider the following problem:

$$\max_{\mu \in \mathcal{M}_0} \{ \int A d\mu \}, \tag{1}$$

which is more general then the problem of maximizing $\int A d\mu$ over the stationary probabilities.

We define

$$m := \max_{\mathcal{M}_0} \int A\, d\mu$$

We will see that this two problems are equivalents, as we will construct a stationary Markov measure μ such that $m = \int A\, d\mu$.

Definition 2. (a) A continuous function $u : [0,1] \to \mathbb{R}$ is called a calibrated forward-subaction if, for any y we have

$$u(y) = \max_x [A(x,y) + u(x) - m], \tag{2}$$

(b) A continuous function $u : [0,1] \to \mathbb{R}$ is called a calibrated backward-subaction if, for any x we have

$$u(x) = \max_y [A(x,y) + u(y) - m], \tag{3}$$

Definition 3. A property P is called generic for A, in Mañé sense, if there exists a generic set $\mathcal{O}$ (in the Baire sense) on the set $C^2([0,1])$ such that if f is in $\mathcal{O}$ then $A+f$ has the property P. And we will say that A is generic in the Mañé sense if A has the property P.

The main results of this paper can be summarized by the following theorems (although in the text it will be split in several other results):

Theorem 1. If A is generic in the Mañé sense, then we have

(a) If $\mu, \tilde{\mu} \in \mathcal{M}_0$ are two maximizing measures, i.e., $m = \int A\ d\mu = \int A\ d\tilde{\mu}$. Then

$$\nu = \tilde{\nu},$$

where $\nu(B) = \mu(\Pi^{-1}(B))$, $\tilde{\nu}(B) = \tilde{\mu}(\Pi^{-1}(B))$, for any B Borel set of $[0,1]^2$ and $\Pi : [0,1]^{\mathbb{N}} \to [0,1]^2$ is the projection map in the first two coordinates .

(b) Both the sets of calibrated backward-subactions and calibrated forward-subactions are unitary.

Definition 4. We say that a backward subaction u is separating if

$$\max_y [A(x,y) + u(y) - u(x)] = m \iff y \in \Omega(A),$$

where $\Omega(A)$ will be defined in the section 4.

Theorem 2. If the observable A is Holder continuous, there exists a separating backward-subaction.

Theorem 3. Let $A : [0,1]^{\mathbb{N}} \to \mathbb{R}$ be a continuous potential that depends only in the first two coordinates of $[0,1]^{\mathbb{N}}$. Then

(a) There exist a measure $\mu_\infty \in \mathcal{M}_0$ such that $\int A d\mu_\infty = m$, and a sequence of measures μ_β, $\beta \in \mathbb{R}$ such that

$$\mu_\beta \rightharpoonup \mu_\infty,$$

where μ_β is defined by $\theta_\beta : [0,1] \to \mathbb{R}, K_\beta : [0,1]^2 \to \mathbb{R}$ (see equations (15) and (16)) as

$$\mu_\beta(A_1...A_n) := \int_{A_1...A_n} \theta_\beta(x_1) K_\beta(x_1,x_2)...K_\beta(x_{n-1},x_n) dx_1...dx_n$$

for any cylinder $A_1...A_n$. Also μ_∞ is a stationary Markov measure (see definition bellow).

(b) If A is generic in the Mañé sense, then there exists only one maximizing stationary Markov measure. Supposing that $D = A_1....A_k$ is a cylinder, then the following limit exists

$$\lim_{\beta\to\infty} \frac{1}{\beta} \ln \mu_\beta(D) = -\inf_{\mathbf{x}\in D} I(\mathbf{x})\,.$$

where $I : [0,1]^{\mathbb{N}} \to [0,+\infty]$ is a function defined by

$$I(\mathbf{x}) := \sum_{i\geq 1} V(x_{i+1}) - V(x_i) - (A-m)(x_i, x_{i+1})\,,$$

and V is the unique calibrated forward subaction.

2 Stationary Markov Measures

Definition 5. A function $P : [0,1] \times \mathcal{A} \to [0,1]$ is called a *transition probability function* on $[0,1]$, where $\mathcal{A}$ is the Borel σ-algebra on $[0,1]$, if

(i) for all $x \in [0,1]$, $P(x,\cdot)$ is a probability measure on $([0,1],\mathcal{A})$,

(ii) for all $B \in \mathcal{A}$, $P(\cdot,B)$ is a $\mathcal{A}$-measurable function from $([0,1],\mathcal{A}) \to [0,1]$.

Sometimes we will use the notation $P_x(B)$ for $P(x,B)$.

Any probability ν on $[0,1]^2$ can be disintegrated as $\theta(dx)P(x,dy)$, where θ is a probability on $([0,1],\mathcal{A})$ [Dellach], Pg 78, (70-III).

Definition 6. A probability measure θ on $([0,1],\mathcal{A})$ is called *stationary* for a transition function $P(\cdot,\cdot)$, if

$$\theta(B) = \int P(x,B)d\theta(x) \qquad \text{for all } B \in \mathcal{A}.$$

Given the initial probability θ and the transition P, as above, one can define a Markov process $\{X_n\}_{n\in\mathbb{N}}$ with state space $S = [0,1]$ (see [AL] section 14.2 for general references on the topic). If θ is stationary for P, then, X_n is a stationary stochastic process. The associated probability μ over $[0,1]^{\mathbb{N}}$ is called the Markov stationary probability defined by θ and P.

Definition 7. A probability measure $\mu \in \mathcal{M}_{[0,1]^{\mathbb{N}}}$ will be called a *stationary Markov measure* if there exist θ and P as in the definition 6, such that μ is given by

$$\mu(A_1...A_n) := \int_{A_1...A_n} \theta(dx_1) P(x_1, dx_2)...P(x_{n-1}, dx_n)\,, \tag{4}$$

where $A_1...A_n$ is a cylinder of size n.

Definition 8. $\mathcal{M}_{[0,1]^2}$ will denote the set of probabilities measures in the Borel sets of $[0,1]^2$.

$\mathcal{M}_{[0,1]^2}$ can be endowed with the weak-$\star$ topology, where a sequence $\nu_n \to \nu$, iff, $\int f d\nu_n \to \int f d\nu$, for all continuous functions $f : [0,1]^2 \to \mathbb{R}$. Which is a compact topological space (by Banach-Alaoglu theorem).

Definition 9. (a) A probability measure $\nu \in \mathcal{M}_{[0,1]^2}$ will be called a *induced stationary Markov measure* if there exists a transition function P, and a probability measure θ on $([0,1], \mathcal{A})$, which is stationary for P, such that for each set $(a,b) \times (c,d) \in [0,1]^2$ we have

$$\nu((a,b) \times (c,d)) = \int_{(a,b)} \int_{(c,d)} dP_x(y) d\theta(x)$$

In this case ν can be disintegrated as $\nu = \theta P$.

(b) We will denote by $\mathbf{M}$ the set of induced stationary Markov measures.

Definition 10. (a) A probability measure ν will be called an *induced absolutely continuous stationary Markov measure*, if ν is in $\mathbf{M}$ and can be disintegrated as $\nu = \theta K$, where θ is an absolutely continuous measure given by a continuous density $\theta(x)dx$, and also for each $x \in [0,1]$ the measure $K(x,.)$ is an absolutely continuous measure given by a continuous density $K(x,y)dy$.

(b) We will denote by $\mathbf{M}_{ac}$ the set of induced absolutely continuous stationary Markov measures.

We can see that the above continuous densities $K : [0,1]^2 \to [0,+\infty)$ and $\theta : [0,1] \to [0,+\infty)$ satisfy the following equations:

$$\int K(x,y)\, dy = 1, \qquad \forall\, x \in [0,1], \tag{5}$$

$$\int \theta(x)\, K(x,y)\, dxdy = 1, \tag{6}$$

$$\int \theta(x)\, K(x,y)\, dx = \theta(y), \qquad \forall\, y \in [0,1]. \tag{7}$$

Moreover, any pair of non-negative continuous functions satisfying the three equations above define an induced absolutely continuous stationary Markov measure.

Let $C[0,1]$ denotes the set of continuous functions defined in $[0,1]$ and taking values on $\mathbb{R}$.

Lemma 1.

(a) $\mathbf{M} = \left\{ \nu \in \mathcal{M}_{[0,1]^2} \ : \ \int f(x) - f(y) \ d\nu(x,y) = 0 \ , \ \forall f \in C[0,1] \right\}$,

(b) $\mathbf{M}$ is a closed set in the weak-$\star$ topology.

The above formulation of the set $\mathbf{M}$ is more convenient for the duality of Fenchel-Rockafellar required by proposition 4. It just says that both marginals in the x and y coordinates are the same.

The important information is given by θ and P. Sometimes we consider μ over $[0,1]^{\mathbb{N}}$ and sometimes the corresponding projected ν over $[0,1]^2$. We will forget the word projected from now on, and the context will indicate which one we are working with. Note that, to make the lecture easier, we are using the following notation: ν when we want to refer to a measure in $[0,1]^2$ and μ for the measures in $[0,1]^{\mathbb{N}}$.

Remark : We point out that maximizing $\int A d\nu$ for probabilities on $\nu \in \mathbf{M}$, means a Kantorovich-Rubinstein (mass transhipment) problem where we assume the two marginals are the same (see [Ra] Vol I section 4 for a related problem). The methods presented here can be used to get approximations of the optimal probability by absolutely continuous ones. These probabilities are obtained via the eigenfunctions of a Perron operator.

In the case we are analyzing, where the observable depends only of the two first coordinates, we will establish some connections between the measures in $[0,1]^2$ and the measures in $[0,1]^{\mathbb{N}}$, and we will see that the problem of maximization can be analyzed as a problem of maximization among induced Markov measures in $[0,1]^2$.

Proposition 1. Let $A : [0,1]^{\mathbb{N}} \to \mathbb{R}$ be a potential which depends only in the first two coordinates of $[0,1]^{\mathbb{N}}$. Then the following is true:

(a) There exists a map, not necessarily surjective, of $\mathbf{M}$ in $\mathcal{M}_0$.

(b) There exists a map, not necessarily injective, of $\mathcal{M}_0$ in $\mathbf{M}$.

(c) $\max_{\mu \in \mathcal{M}_0} \int A(x_1, x_2) \ d\mu(\mathbf{x}) = \max_{\nu \in \mathbf{M}} \int A(x,y) \ d\nu(x,y)$

Proof: (a) A measure $\nu \in \mathbf{M}$ can be disintegrated as $\nu = \theta P$, and then can be extended to a measure $\mu \in \mathcal{M}_0$ by

$$\mu(A_1 ... A_n) := \int_{A_1 ... A_n} \theta(dx_1) P(x_1, dx_2) ... P(x_{n-1}, dx_n) \,, \tag{8}$$

Also, we have

$$\int_{[0,1]^{\mathbb{N}}} A(x_1,x_2)d\mu(\mathbf{x}) = \int_{[0,1]^2} A(x,y)d\nu(x,y)\,.$$

(b) A measure $\mu \in \mathcal{M}_0$ can be projected in a measure $\nu \in \mathcal{M}_{[0,1]^2}$, defined by , for each Borel set B of $[0,1]^2$,

$$\nu(B) = \mu(\Pi^{-1}(B))\,,$$

where $\Pi : [0,1]^{\mathbb{N}} \to [0,1]^2$ is the projection in the two first coordinates. Note that, by the lemma 1, $\nu \in \mathbf{M}$. Then we have

$$\int_{[0,1]^{\mathbb{N}}} A(x_1,x_2)d\mu(\mathbf{x}) = \int_{[0,1]^2} A(x,y)d\nu(x,y)\,.$$

(c) It follows easily by (a) and (b). $\square$

Remark: Note that in the item (a), in the particular case where $\nu \in \mathbf{M}_{ac}$, we have that ν can be disintegrated as $\nu = \theta K$, and then the stationary Markov measure μ is given by

$$\mu(A_1...A_n) := \int_{A_1...A_n} \theta(x_1)K(x_1,x_2)...K(x_{n-1},x_n)dx_1...dx_n\,, \qquad (9)$$

where onde $A_1...A_n$ is a cylinder. And μ is a absolutely continuous Markov measure in $[0,1]^{\mathbb{N}}$.

3 The Maximization Problem

We are interested in finding stationary Markov probabilities μ_∞ on $[0,1]^{\mathbb{N}}$ (defined by a initial θ and a transition P) such that maximize the value

$$\int A(x_1,x_2)d\mu(\mathbf{x}),$$

among all stationary (not only Markov) probabilities for the shift $[0,1]^{\mathbb{N}}$. As we just consider potentials of the form $A(x,y)$, in this case, it is not possible to have uniqueness. We just take into account the information of the measure on cylinders of size two. In any case, the Markov stationary probability we will describe below will also solve this maximizing problem.

As we are considering the problem of maximizing $\int A\, d\mu$ over the holonomic measures, which contains all stationary measures in $[0,1]^{\mathbb{N}}$. By the item (c) of the proposition 1: $\max\limits_{\mu\in\mathcal{M}_0} \int A\, d\mu = \max\limits_{\nu\in\mathbf{M}} \int A\, d\nu$.

Hence, the problem we are analyzing is equivalent to the problem of find ν_∞ which is the maximal for $\int A d\nu$, among all $\nu \in \mathbf{M}$. Because once we have ν_∞, by item (a) of the proposition 1, we obtain a maximizing Markov measure μ_∞ among the holonomic measures.

Now we will concentrate on the maximizing problem in $[0,1]^2$.
Let $A : [0,1] \times [0,1] \to \mathbb{R}$ be a continuous function. We will denote by

$$\mathfrak{M}_0 := \left\{ \nu \in \mathbf{M} \ : \ \int A(x,y)\; d\nu(x,y) = m \right\}$$

where

$$m = \max_{\nu \in \ \mathbf{M}} \left\{ \int A(x,y) d\nu(x,y) \right\} .$$

A measure in $\mathfrak{M}_0$ will be called a maximizing measure on $\mathbf{M}$.

Consider now the variational problem

$$\max_{\theta K \in \mathbf{M}_{ac}} \left\{ \int \ \beta\, A(x,y)\ \theta(x) K(x,y) dx dy - \int \theta(x) K(x,y) \log\left(K(x,y)\right) dx dy \right\} . \tag{10}$$

In some sense we are considering above a kind of pressure problem (see [PP]).

Definition 11. We define the *term of entropy* of an absolutely continuous probability measure $\nu \in \mathcal{M}_{[0,1]^2}$, given by a density $\nu(x,y) dx dy$, as

$$S[\nu] = - \int \nu(x,y) \log\left(\frac{\nu(x,y)}{\int \nu(x,z) dz} \right) dx dy\, . \tag{11}$$

It is easy to see that any $\nu = \theta K \in \mathbf{M}_{ac}$ satisfies

$$S[\theta K] = - \int \theta(x) K(x,y) \log\left(K(x,y)\right) dx dy\, . \tag{12}$$

We call $S[\nu] = S[\theta K]$ the entropy penalized of the probability $\nu = \theta K \in \mathbf{M}_{ac}$.

Lemma 2. If $\nu = \theta K \in \mathbf{M}_{ac}$ and K is strictly positive, then $S[\nu] \leq 0$.

Proof: log is a concave function. Hence, by Jensen inequality, we have

$$- \int \theta(x) K(x,y) \log\left(K(x,y)\right) dx dy = \int \theta(x) K(x,y) \log\left(\frac{1}{K(x,y)} \right) dx dy \leq$$

$$\leq \log \int \theta(x) K(x,y) \frac{1}{K(x,y)} dx dy = \log(1) = 0.$$

□

For each β fixed, we will exhibit a measure ν_β in $\mathbf{M}_{ac}$ which maximizes (10). Finally, we will show that such ν_β will approximate in weak convergence the probabilities ν_∞ which are maximizing for A in the set $\mathbf{M}$.

In order to do that, we need to define the following operators, defined in the Banach space of real valued continuous functions defined in $[0,1]$, with the sup norm, which we will denote by $C([0,1])$:

Definition 12. Let $\mathcal{L}_\beta, \bar{\mathcal{L}}_\beta : C([0,1]) \to C([0,1])$ be given by

$$\mathcal{L}_\beta \varphi(y) = \int e^{\beta A(x,y)}\, \varphi(x) dx, \tag{13}$$

$$\bar{\mathcal{L}}_\beta \varphi(x) = \int e^{\beta A(x,y)}\, \varphi(y) dy. \tag{14}$$

We refer the reader to [Ka] and [Sch] chapter IV for a general reference on positive integral operators.

Theorem 4. The operators $\mathcal{L}_\beta$ and $\bar{\mathcal{L}}_\beta$ have the same strictly positive maximal eigenvalue λ_β, which is simple and isolated. The eigenfunctions associated are strictly positive functions.

*Proof :*We can see that $\mathcal{L}_\beta$ is a compact operator, because the image by $\mathcal{L}_\beta$ of the unity closed ball of $C([0,1])$ is a equicontinuous family in $C([0,1])$: we know that $e^{\beta A}$ is a uniformly continuous function, and then, if φ is in the closed unit ball, we have

$$|\mathcal{L}_\beta\varphi(y) - \mathcal{L}_\beta\varphi(z)| \le \int |e^{\beta A(x,y)} - e^{\beta A(x,z)}|\, |\varphi(x)| dx \le |e^{\beta A(x,y)} - e^{\beta A(x,z)}| < \delta\,,$$

if, y and z are close enough. Thus, we can use Arzela-Ascoli Theorem to prove the compacity of $\mathcal{L}_\beta$ (see also [Sch] Chapter IV section 1).

The spectrum of a compact operator is a sequence of eigenvalues that converges to zero, possibly added by zero. This implies that any non-zero eigenvalue of $\mathcal{L}_\beta$ is isolated (i.e. there are no sequence in the spectrum of $\mathcal{L}_\beta$ that converges to some non-zero eigenvalue).

The definition of $\mathcal{L}_\beta$ now shows that it preserves the cone of positive functions in $C([0,1])$, indeed, sending a point in this cone to the interior of the cone. This means that $\mathcal{L}_\beta$ is a strictly positive operator.

The Krein-Ruthman Theorem (Theorem 19.3 in [De]) implies that there is a strictly positive eigenvalue λ_β, which is maximal (i.e. $\lambda_\beta > |\lambda|$, if λ is in the spectrum of $\mathcal{L}_\beta$) and simple (i.e. the eigenspace associated to λ_β is one-dimensional), and is associated to a strictly positive eigenfunction φ_β. As an additional result (that will not be used in this work), we can say that there are no other eigenvalues associated to positive eigenfunctions (see also Theorem 6.3 chapter 2 in [Ka]).

If we proceed in the same way, we get the same conclusions about the operator $\bar{\mathcal{L}}_\beta$, and we get the respective eigenvalue $\bar{\lambda}_\beta$ and eigenfunction $\bar{\varphi}_\beta$.

In order to prove that $\bar{\lambda}_\beta = \lambda_\beta$, we use the positivity of φ_β and $\bar{\varphi}_\beta$ and the fact that $\bar{\mathcal{L}}_\beta$ is the adjoint of $\mathcal{L}_\beta$ (here we see that our operators can be, in fact, defined in the Hilbert space $L^2([0,1])$, which contains $C([0,1])$). We have $< \varphi_\beta, \bar{\varphi}_\beta >= \int \varphi_\beta(x)\bar{\varphi}_\beta(x)dx > 0$, and

$$\lambda_\beta < \varphi_\beta, \bar{\varphi}_\beta >=< \mathcal{L}_\beta \varphi_\beta, \bar{\varphi}_\beta >=< \varphi_\beta, \bar{\mathcal{L}}_\beta \bar{\varphi}_\beta >= \bar{\lambda}_\beta < \varphi_\beta, \bar{\varphi}_\beta > .$$

□

An estimate on the spectral gap for the operator $\mathcal{L}_\beta$, where $\beta > 0$, is given in [Os] [Hop]: suppose

$$\tilde{M} = \sup_{(x,y)\in[0,1]^2} A(x,y), \quad \text{and} \quad \tilde{m} = \inf_{A(x,y)\in[0,1]^2} A(x,y).$$

If λ_β is the main eigenvalue, then any other λ in the spectrum of $\mathcal{L}_\beta$ satisfies

$$\lambda_\beta \, (\frac{\tilde{M}^\beta - \tilde{m}^\beta}{\tilde{M}^\beta + \tilde{m}^\beta}) \; > \; \lambda.$$

With this information one can give an estimate of the decay of correlation for functions evolving under the probability of the Markov Chain associated to such value β (see next proposition). The proof of this claim is similar to the reasoning in chapter 2 page 26 in [PP], which deals with the case where the state space is discrete.

Let us call $\varphi_\beta, \bar{\varphi}_\beta$ the strictly positive eigenfunctions for $\mathcal{L}_\beta$ and $\bar{\mathcal{L}}_\beta$ associated to λ_β, which satisfy $\int \varphi_\beta(x)\,dx = 1$ and $\int \bar{\varphi}_\beta(x)\,dx = 1$. It is easy to see that φ_β and $\bar{\varphi}_\beta$ are continuous functions.

We will define a density $\theta_\beta : [0,1] \to \mathbb{R}$ by

$$\theta_\beta(x) := \frac{\varphi_\beta(x)\ \bar{\varphi}_\beta(x)}{\pi_\beta}, \tag{15}$$

where $\pi_\beta = \int \varphi_\beta(x)\bar{\varphi}_\beta(x)dx$, and a transition $K_\beta : [0,1]^2 \to \mathbb{R}$ by

$$K_\beta(x,y) := \frac{e^{\beta A(x,y)}\ \bar{\varphi}_\beta(y)}{\bar{\varphi}_\beta(x)\,\lambda_\beta}. \tag{16}$$

Let $\nu_\beta \in \mathbf{M}_{ac}$ be defined by

$$d\nu_\beta(x,y) := \theta_\beta(x) K_\beta(x,y) dxdy. \tag{17}$$

Proposition 2. The Markov measure $\nu_\beta = \theta_\beta K_\beta$ defined above maximize

$$\int \beta\, A(x,y)\; \theta(x) K(x,y) dxdy - \int \theta(x) K(x,y) \log\left(K(x,y)\right) dxdy$$

over all absolutely continuous Markov measures. Also

$$\log \lambda_\beta = \int \beta A\; \theta_\beta K_\beta dxdy + S[\theta_\beta K_\beta].$$

Proof: By the definition of the functions θ_β, K_β, we have

$$S[\theta_\beta K_\beta] = -\int (\beta A(x,y) + \log \bar{\varphi}_\beta(y) - \log \bar{\varphi}_\beta(x) - \log \lambda_\beta) d\nu_\beta$$

Then

$$\int \beta A\; \theta_\beta K_\beta dxdy + S[\theta_\beta K_\beta] = \log \lambda_\beta + \int (\log \bar{\varphi}_\beta(x) - \log \bar{\varphi}_\beta(y)) \theta_\beta(x) K_\beta(x,y) dxdy\,,$$

and the last integral is zero because $\nu_\beta = \theta_\beta K_\beta \in \mathbf{M}_{ac}$.

To show that ν_β is maximizing let ν be any measure in $\mathbf{M}_{ac}$ and $0 \leq \tau \leq 1$. We claim that the function

$$I[\tau] := \int \beta A d\nu_\tau + S[\nu_\tau]$$

where $\nu_\tau = (1-\tau)\nu_\beta + \tau\nu$, is concave and $I'(0) = 0$

Indeed, see proof of theorem 33 of [GV]. We point out that the entropy term there has a difference of signal, which implies that the function I in [GV] is convex.

□

Lemma 3. (a) There exists a constant $c > 0$ such that for all $x \in [0,1]$

$$e^{-\beta c} \leq \varphi_\beta(x) \leq e^{\beta c} \qquad e^{-\beta c} \leq \bar{\varphi}_\beta(x) \leq e^{\beta c}.$$

Also the sequences

$$\frac{1}{\beta} \log \pi_\beta \quad \text{and} \quad \frac{1}{\beta} \log \lambda_\beta$$

admit accumulation points when $\beta \to \infty$.

(b) The sets

$$\left\{ \frac{1}{\beta} \log(\varphi_\beta) \;\mid \beta > 1 \right\} \;\text{ and }\; \left\{ \frac{1}{\beta} \log(\bar{\varphi}_\beta) \;\mid \beta > 1 \right\}$$

are equicontinuous, and relatively compact in the supremum norm.

Proof:

(a) We left to the reader.

(b) We just have to prove the equicontinuity of both sets. Once we have that, and considering the fact that both sets are sets of functions defined in the compact set $[0,1]$, we use item (a) and Arzela-Ascoli's Theorem to get the relative compactness of these sets.

To have the equicontinuity for the first set, let y be a point in $[0,1]$, and let $\beta > 1$. Let $\epsilon > 0$. We will use the fact that A is a uniformily continuous map: We know there exists $\delta > 0$, such that $|y-z| < \delta$, implies $|A(x,y) - A(x,z)| < \epsilon$, $\forall x \in [0,1]$. Without any loss of generality, we suppose that $\varphi_\beta(y) \geq \varphi_\beta(z)$. We have:

$$\left|\frac{1}{\beta}\log(\varphi_\beta(y)) - \frac{1}{\beta}\log(\varphi_\beta(z))\right| =$$

$$= \frac{1}{\beta}\left(\log\left(\frac{1}{\lambda_\beta}\int e^{\beta A(x,y)}\varphi_\beta(x)dx\right) - \log\left(\frac{1}{\lambda_\beta}\int e^{\beta A(x,z)}\varphi_\beta(x)dx\right)\right) =$$

$$= \frac{1}{\beta}\log\left(\frac{\int e^{\beta A(x,y)}\varphi_\beta(x)dx}{\int e^{\beta A(x,z)}\varphi_\beta(x)dx}\right) \leq \frac{1}{\beta}\log\left(\frac{\int e^{\beta(A(x,z)+\epsilon)}\varphi_\beta(x)dx}{\int e^{\beta A(x,z)}\varphi_\beta(x)dx}\right) =$$

$$= \frac{1}{\beta}\log\left(e^{\beta\epsilon}\frac{\int e^{\beta A(x,z)}\varphi_\beta(x)dx}{\int e^{\beta A(x,z)}\varphi_\beta(x)dx}\right) = \epsilon\,.$$

We prove the equicontinuity for the second set in the same way.

□

From the above, we can find $\beta_n \to \infty$ which defines convergent subsequences $\frac{1}{\beta_n}\log\varphi_{\beta_n}$.

Let us fix a subsequence β_n s.t. $\beta_n \to \infty$ and s.t. all the three following limits exist:

$$V(x) := \lim_{n\to\infty}\frac{1}{\beta_n}\log\varphi_{\beta_n}(x) \quad , \quad \bar{V}(x) := \lim_{n\to\infty}\frac{1}{\beta_n}\log\bar{\varphi}_{\beta_n}(x)$$

$$\tilde{m} := \lim_{n\to\infty}\frac{1}{\beta_n}\log\lambda_{\beta_n}$$

Note that the limits defining V and $\bar{V}$ converge uniformly. In principle, the function V depends of the sequence β_n we choose.

Proposition 3 (Laplace's Method). Let $f_k : [0,1] \to \mathbb{R}$ be a sequence of functions that converge uniformly, as k goes to ∞, to a function $f : [0,1] \to \mathbb{R}$. Then

$$\lim_k \frac{1}{k}\log\int_0^1 e^{kf_k(x)}dx = \sup_{x\in[0,1]} f(x)$$

Lemma 4.
$$\lim_{n\to\infty}\frac{1}{\beta_n}\log\pi_{\beta_n} = \max_{x\in[0,1]}(V(x)+\bar{V}(x))$$

Proof:

$$\pi_{\beta_n} = \int_0^1 \varphi_{\beta_n}(x)\bar{\varphi}_{\beta_n}(x)dx = \int_0^1 e^{\beta_n(\frac{1}{\beta}\log\varphi_{\beta_n}(x)+\frac{1}{\beta}\log\bar{\varphi}_{\beta_n}(x))}dx$$

And note that $\frac{1}{\beta_n}\log\varphi_{\beta_n}(x)\to V(x)$, $\frac{1}{\beta_n}\log\bar{\varphi}_{\beta_n}(x)\to\bar{V}(x)$ uniformly. Hence it follows by the Laplace's Method. □

Also by the Laplace method we have the following lemma:

Lemma 5.
$$V(y) = \max_{x\in[0,1]}(V(x)+A(x,y)-\tilde{m})$$

and

$$\bar{V}(x) = \max_{y\in[0,1]}(\bar{V}(y)+A(x,y)-\tilde{m}).$$

For some subsequence (of the subsequence $\{\beta_n\}$ fixed after the proof of lemma 3, which we will also denote by $\{\beta_n\}$), the measures ν_{β_n} defined in (17) weakly converge to a measure $\nu_\infty\in\mathbf{M}$. Then

$$\lim_{n\to\infty}\int Ad\nu_{\beta_n} = \int Ad\nu_\infty$$

The measure $\nu_\infty\in\mathbf{M}$, i.e., ν_∞ is a stationary Markov measure.

Theorem 5.
$$\int A(x,y)d\nu_\infty(x,y) = m$$

i.e., ν_∞ is a maximizing measure on $\mathbf{M}$.

In order to prove the theorem we need first some new results.

Proposition 4. Given a potential $A\in C([0,1]^2)$, we have that

$$\sup_{\nu\in\,\mathbf{M}}\int Ad\nu = \inf_{f\in\,C([0,1])}\ \max_{(x,y)}(A(x,y)+f(x)-f(y))$$

This proposition will be a consequence of the Fenchel-Rockafellar duality theorem. Let us fix the setting we consider in order to to apply this theorem.

Let $C([0,1]^2)$ be the set of continuous functions in $[0,1]^2$ and $\mathcal{S}$ the set of signed measures over the Borel $\sigma-$algebra of $[0,1]^2$.

Consider the convex correspondence $H : C([0,1]^2) \to \mathbb{R}$ given by $H(\phi) = \max(A+\phi)$ and

$$\mathcal{C} := \{\phi \in C([0,1]^2) \ : \ \phi(x,y) = f(x) - f(y) \ \text{, for some} \quad f \in C([0,1])\}$$

We define a concave correspondence $G : C([0,1]^2) \to \mathbb{R} \cup \{-\infty\}$ by $G(\phi) = 0$ if $\phi \in \mathcal{C}$ and $G(\phi) = -\infty$ otherwise.

Then the corresponding Fenchel transforms,

$H^* : \mathcal{S} \to \mathbb{R} \cup \{+\infty\}$, $G^* : \mathcal{S} \to \mathbb{R} \cup \{-\infty\}$, are given by

$$H^*(\nu) = \sup_{\phi \in C([0,1]^2)} \left[\int \phi(x,y) d\nu(x,y) - H(\phi) \right]$$

and

$$G^*(\nu) = \inf_{\phi \in C([0,1]^2)} \left[\int \phi(x,y) d\nu(x,y) - G(\phi) \right]$$

Lemma 6. Given H and G as above, then

$$H^*(\nu) = \begin{cases} -\int A(x,y) d\nu(x,y) & \text{if} \quad \nu \in \mathcal{M}_{[0,1]^2} \\ +\infty & \text{otherwise} \end{cases}$$

$$G^*(\nu) = \begin{cases} 0 & \text{if} \quad \nu \in \mathbf{M} \\ -\infty & \text{otherwise} \end{cases}$$

This lemma follows from lemma 2 of [GL].

Proof of proposition 4: The duality theorem of Fenchel-Rockafellar says that

$$\sup_{\phi \in C([0,1]^2)} [G(\phi) - H(\phi)] = \inf_{\nu \in \mathcal{S}} [H^*(\nu) - G^*(\nu)]$$

Hence, by lemma 6

$$\sup_{\phi \in \mathcal{C}} [-\max_{(x,y)} (A+\phi)(x,y)] = \inf_{\nu \in \mathbf{M}} [-\int A(x,y) d\nu(x,y)]$$

Using the definition of $\mathcal{C}$ we have that

$$\sup_{\nu \in \mathbf{M}} \int A d\nu = \inf_{f \in C([0,1])} \max_{(x,y)} (A(x,y) + f(x) - f(y))$$

□

Lemma 7. $\tilde{m} = m$.

Proof: Note that by proposition 4 and lemma 5 we have that $m \leq \tilde{m}$. To show the other inequality remember that

$$\log \lambda_{\beta_n} = \int \beta A \, d\nu_{\beta_n} + S[\nu_{\beta_n}].$$

Then

$$\tilde{m} = \lim_{n\to\infty} \int A \, d\nu_{\beta_n} + \frac{1}{\beta_n} S[\nu_{\beta_n}].$$

Note that $\nu_{\beta_n} \in \mathbf{M}$, which implies $\int A \, d\nu_{\beta_n} \leq m$.
As $S[\nu_{\beta_n}] \leq 0$, we have

$$\int A \, d\nu_{\beta_n} + \frac{1}{\beta_n} S[\nu_{\beta_n}] \leq m \qquad \forall n$$

Then $\tilde{m} \leq m$. □

Proof of Theorem 5: Remember that $\nu_{\beta_n} \rightharpoonup \nu_\infty$, then

$$\lim_{n\to\infty} \int A d\nu_{\beta_n} = \int A d\nu_\infty$$

by lemma 7

$$m = \lim_{n\to\infty} \int A \, d\nu_{\beta_n} + \frac{1}{\beta_n} S[\nu_{\beta_n}] \leq \int A \, d\nu_\infty.$$

□

4 Uniqueness of maximizing measures and Calibrated Sub-Actions

We want to remark here that for the results of this section we were inspired by the works of [Gom], [GLM] and [GL]. Hence, jointing all these ideas, we are able to prove that the sets of calibrated subactions (backward and forward), to be defined in the sequel, are unitary if the potential A is C^2, satisfy the twist property and is also generic in Mañé's sense. The precise definitions will be given in what follows.

We repeat here the important definition of calibrated subactions:

Definition 13. (a) A continuous function $u : [0,1] \to \mathbb{R}$ is called a calibrated forward-subaction if, for any y we have

$$u(y) = \max_x [A(x,y) + u(x) - m], \tag{18}$$

(b) A continuous function $u : [0,1] \to \mathbb{R}$ is called a calibrated backward-subaction if, for any x we have

$$u(x) = \max_y [A(x,y) + u(y) - m], \tag{19}$$

Note that V and $\bar{V}$ defined in lemma 5 are, respectively, forward and backward calibrated subactions (remember that $\tilde{m} = m$ by lemma 7).

Let u be a calibrated backward-subaction, using the fact that $[0,1]$ is compact, there exists $y(x)$ (maybe not unique) such that

$$u(x) = A(x, y(x)) + u(y(x)) - m. \tag{20}$$

Proposition 5. Let $\nu \in \mathfrak{M}_0$ be any maximizing measure, and u be a calibrated backward-subaction. Then ν-almost everywhere

$$u(x) = A(x,y) + u(y) - m.$$

Proof: Note that $u(x) \geq A(x,y) + u(y) - m$ for all $y \in [0,1]$, as $\nu \in \mathfrak{M}_0$ then $\int A d\nu = m$, and as $\int (u(x) - u(y))\, d\nu = 0$, the proposition follows. □

Definition 14.

$$D^+u(x) = \{\, p \in \mathbb{R} \mid \limsup_{|v| \to 0} \frac{u(x+v) - u(x) - p\, v}{|v|} \leq 0 \,\}.$$

The definition of $D^-u(x)$ is similar. For the main properties of $D^-u(x)$ and $D^+u(x)$ see [EvG], [Gom] and [Gom1].

Lemma 8. For any u calibrated backward-subaction , and $(x, y(x))$ satisfying equation (20), we have that

$$D^-u(x) \neq \emptyset \;\; \forall x, \qquad \text{and} \qquad \frac{\partial A}{\partial x}(x, y(x)) \in D^-u(x)$$

and

$$D^+u(y(x)) \neq \emptyset \qquad \text{and} \qquad -\frac{\partial A}{\partial y}(x, y(x)) \in D^+u(y(x)).$$

Proof: Fix $(x, y(x))$ satisfying equation (20), then for any $w \in [0,1]$ we have that

$$u(x+w) \geq A(x+w, y(x)) + u(y(x)) - m$$

using equation 20, we see that

$$u(x+w) - u(x) - A(x+w, y(x)) + A(x, y(x)) \geq 0,$$

then

$$\liminf_{|w| \to 0} \frac{u(x+w) - u(x) - (\frac{\partial A}{\partial x}(x, y(x))\, w + r(w))}{|w|} \geq 0,$$

hence $\frac{\partial A}{\partial x}(x,y(x)) \in D^- u(x)$.

Also for $(x, y(x))$ satisfying (20) and any $w \in [0,1]$ we have

$$u(x) \geq A(x, y(x) + w) + u(y(x) + w) - m,$$

using equation (20), we get that

$$u(y(x) + w) - u(y(x)) + A(x, y(x) + w) - A(x, y(x)) \leq 0$$

then

$$\limsup_{|w|\to 0} \frac{u(y(x)+w) - u(y(x)) - (-\frac{\partial A}{\partial y}(x, y(x))\, w + r(w))}{|w|} \leq 0$$

hence $-\frac{\partial A}{\partial y}(x, y(x)) \in D^+ u(y(x))$.

□

Remember that $D^- u(w) \neq \emptyset$ and $D^+ u(w) \neq \emptyset$ implies that u is differentiable at w.

Lemma 9. For any measure $\nu \in \mathbf{M}$, we have that for almost every point $(x, y) \in \mathrm{supp}(\nu)$, there exists z such that $(z, x) \in \mathrm{supp}(\nu)$.

Proof: Define the set

$$R = \{(x, y) \in \mathrm{supp}(\nu) \ : \ x \neq w, \ \ \forall (z, w) \in \mathrm{supp}(\nu)\}$$

Suppose, by contradiction, that $\nu(R) = \epsilon > 0$.

Let $\pi_j : [0,1]^2 \to [0,1]$ be the projection on the j-th coordinate.

Let ν_2 be the measure on the Borel sets of $[0,1]$ given by $\nu_2(B) = \nu(\pi_2^{-1}(B))$, where B is any Borel set in $[0,1]$.

Consider $R_1 = \pi_1(R)$. We have

$$R_1 = \{x \in \pi_1(\mathrm{supp}(\nu)) \ : \ x \neq w \ \ \forall (z, w) \in \mathrm{supp}(\nu)\}.$$

We claim that

$$\nu_2(R_1) = \int_{\mathrm{supp}(\nu)} \chi_{R_1}(y) d\nu_\infty(x, y) = 0.$$

Indeed, the first equality is immediate. To prove the second equality, take $(x, y) \in \mathrm{supp}(\nu)$. We have two possibilities: If $y \notin \pi_1(\mathrm{supp}(\nu))$, then $y \notin R_1$. And if $y \in \pi_1(\mathrm{supp}(\nu))$ we have $(x, y) \in \mathrm{supp}(\nu)$ and then $y \notin R_1$. This shows the claim.

By the other hand, note that $\pi_1^{-1}(R_1) \cap \mathrm{supp}(\nu) = R$, and thus

$$\int_{\mathrm{supp}(\nu)} \chi_{R_1}(x) d\nu(x, y) = \int_{\mathrm{supp}(\nu)} \chi_{\pi_1^{-1}(R_1)}(x, y) d\nu(x, y) = \nu(R) = \epsilon.$$

Now let U be an open set of $[0,1]$ which contains R_1 and such that $\nu_2(U) < \nu_2(R_1) + \epsilon/2 = \epsilon/2$. Consider a sequence of continuous function f_j such that $f_j \uparrow \chi_U$. Using the monotonous convergence theorem and $\nu \in \mathbf{M}$, we have:

$$\epsilon/2 > \nu_2(U) = \int \chi_U(y) d\nu(x,y) = \lim_j \int f_j(y) d\nu(x,y) =$$

$$= \lim_j \int f_j(x) d\nu(x,y) = \int \chi_U(x) d\nu(x,y) \geq \int \chi_{R_1}(x) d\nu(x,y) = \epsilon$$

which is a contradiction. □

Theorem 6. Let $\nu \in \mathfrak{M}_0$ be any maximizing measure. If the observable A is C^2, and $\frac{\partial^2 A}{\partial x \partial y} > 0$, then the measure ν is supported on a graph.

Proof: Let u be any calibrated backward-subaction. Note that, for any fixed p and x, the equation $p = \frac{\partial A}{\partial x}(x,y)$ has a unique solution $y(x,p)$ because $\frac{\partial^2 A}{\partial x \partial y} > 0$.

Take $(x_0, y_0) \in \text{supp}(\nu)$, then (x_0, y_0) satisfies equation (20). By lemma 9 there exists z_0 such that $(z_0, x_0) \in \text{supp}(\nu)$, it means that $x_0 = y(z_0)$, and thus we can apply lemma in order 8 to obtain that u is differentiable in x_0. Hence

$$Du(x_0) = \frac{\partial A}{\partial x}(x_0, y_0)$$

and then y_0 in the equation above is uniquely defined. □

Lemma 10. If the observable A is C^2, and $\frac{\partial^2 A}{\partial x \partial y} > 0$, then $\cup_{\nu \in \mathfrak{M}_0} \text{supp}(\nu)$ is contained in a graph.

Proof: Let ν_1 and ν_2 be two maximizing measures. Suppose there exists $x \in \pi_1(\text{supp}(\nu_1)) \cap \pi_1(\text{supp}(\nu_2))$. Let y_1 and y_2 be the (unique) points such that $(x, y_1) \in \text{supp}(\nu_1)$ and $(x, y_2) \in \text{supp}(\nu_2)$, and let u be a calibrated backward-subaction. By proposition 5, we have

$$u(x) = A(x, y_1) + u(y_1) - m \quad \text{and} \quad u(x) = A(x, y_2) + u(y_2) - m\,.$$

Following the proof of theorem 6, we get that $y_1 = y_2$. □

Definition 15. Given k and $x, y \in [0,1]$, we will call a k-path beginning in x and ending at y an ordered sequence of points

$$(x_1, ..., x_k) \in [0,1] \times ... \times [0,1]$$

satisfying $x_1 = x$, $x_k = y$.

We will denote by $\mathcal{P}_k(x,y)$ the set of such k-paths.

Remark: 1) Here we shall note that this results can not be a particular case of the results obtain in [Gom], [GLM] for the theory of Aubry-Mather, because in A-M theory a Lagrangian $L:[0,1]\times\mathbb{R}\to\mathbb{R}$, satisfy the hypothesis that $L(x,v)\to+\infty$ when $|v|\to\infty$.

2) A path in A-M theory (see [GLM]) is an orderer sequence of points $(x_0,....,x_k)\in\mathbb{R}^N\times...\times\mathbb{R}^N$, and such that for each x_j we associate a velocity $v_j=x_{j+1}-x_j$, $0\le j<k$, hence with those pair (x_j,v_j) we are able to calculate the action of the path $(x_0,....,x_k)$. While here there is no velocity and only the points of the path are used to calculate the action of the path.

Definition 16. A point $x\in[0,1]$ is called non-wandering with respect to A if, for each $\epsilon>0$, there exists $k\ge 1$ and a k-path $(x_1,....,x_k)$ in $\mathcal{P}_k(x,x)$ such that

$$\left|\sum_{i=1}^{k-1}(A-m)(x_i,x_{i+1})\right|<\epsilon.$$

We will denote by $\Omega(A)$ the set of non-wandering points with respect to A.

Lemma 11. Suppose that the observable A is C^2, and $\frac{\partial^2 A}{\partial x\partial y}>0$. Let $\nu\in\mathfrak{M}_0$ be any maximizing measure, then $\pi_1(\text{supp}(\nu))\subset\Omega(A)$.

Proof: Let u be a backward calibrated subaction and $Y:\text{dom}(Du)\to[0,1]$, where $\text{dom}(Du)$ are the differentiable points of u, the map defined by $Y(x)=y$, where y is the unique point such that (x,y) satisfies (20). As we will see in the proposition 12, this map is monotonous, hence we can define a measurable map $Y:[0,1]\to[0,1]$, as taking the limit in the left. Note that $\nu_\infty\circ\pi_1^{-1}$-a.e. $\pi_1(\text{supp}\nu_\infty)\subset\text{dom}(Du)$.

Let us see that $\nu_\infty\circ\pi_1^{-1}$ is an invariant measure for Y. Indeed, for $f\in C^0(\Omega(A))$, we have that:

$$\int f\circ Y(x)\,d\nu_\infty\circ\pi_1^{-1}(x)=\int f\circ Y(x)\,d\nu_\infty(x,y)=\int f(y)\,d\nu_\infty(x,y)=$$

$$=\int f(x)\,d\nu_\infty(x,y)=\int f(x)\,d\nu_\infty\circ\pi_1^{-1}(x)\,.$$

Take $(x,y)\in\text{supp}\nu_\infty$ and B a ball centered in x, as $\text{supp}\nu_\infty$ is contained in a graph, $\nu_\infty\circ\pi_1^{-1}(B)>0$. Hence, by Poincaré recurrence theorem, there exists $x_1\in B\cap\text{dom}(Du)$ such that $Y^j(x_1)=:x_{j+1}$ return infinitely often to B. Note that the points x_j satisfy the following equation:

$$u(x_j)-u(x_{j+1})=A(x_j,x_{j+1})-m$$

because, by lemma 8, u is differentiable in each x_j and then there exists only one $y(x_j)$ (that coincides with x_{j+1}) which satisfies the equation (20).

We fix $\epsilon > 0$ and $x_j \in B$, we can construct the following path: $(\tilde{x}_1, ..., \tilde{x}_j) = (x, x_2, ..., x_{j-1}, x)$, and we have that

$$\sum_{i=1}^{j-1} A(\tilde{x}_i, \tilde{x}_{i+1}) = u(x_1) - u(x_j) + A(x, x_2) - A(x_1, x_2) + A(x_{j-1}, x) - A(x_{j-1}, x_j) \leq \epsilon$$

if B is small enough, because u is Lipschitz (as A is C^2). □

Definition 17. We call Mañé potential the function $S : [0,1] \times [0,1] \to \mathbb{R}$ defined by

$$S(x, y) = \inf_k S_k(x, y),$$

And Peierls barrier the function $h : [0,1] \times [0,1] \to \mathbb{R} \cup \{+\infty\}$ defined by

$$h(x, y) = \liminf_{k \to \infty} S_k(x, y),$$

where

$$S_k(x, y) = \inf_{(x_1, \ldots, x_k) \in \mathcal{P}_k(x,y)} \left[-\sum_{i=1}^{k-1} (A - m)(x_i, x_{i+1}) \right].$$

It is easy to see that

$$\Omega(A) = \{x \in [0,1] \ : \ S(x, x) = h(x, x) = 0\}$$

The functions S and h have following properties

(a) if $x, y, z \in [0,1]$ then $S(x, z) \leq S(x, y) + S(y, z)$.

(b) $S(\cdot, y)$ is a forward-subaction and $S(x, \cdot)$ is a backward-subaction.

(c) $h(\cdot, y)$ is a calibrated forward-subaction and $h(x, \cdot)$ is a calibrated backward-subaction.

We want to prove, under the condition that A is generic in Mañé sense, that V and $\bar{V}$ are unique (up to a constant).

Definition 18. We will said that A is generic in Mañé sense if $\mathfrak{M}_0(A) = \{\nu\}$ and $\pi_1(supp(\nu)) = \Omega(A)\}$.

We want to prove that, when A is generic in the Mañé sense, then the functions V and $\bar{V}$ are unique. To do that first we show that generically the maximizing measure is unique, as we will se in the following proposition.

Proposition 6. Suppose that the observable A is C^2, and $\frac{\partial^2 A}{\partial x \partial y} > 0$. Then the set

$$G_2 = \{f \in C^2([0,1]) \ | \ \mathfrak{M}_0(A + f) = \{\nu\} \text{ and } \pi_1(supp(\nu)) = \Omega(A + f)\}$$

is generic in $C^2([0,1])$.

We will use a result of [BC] in order to prove the proposition. First we will show that

$$G_1 = \{f \in C^2([0,1]) \mid \mathfrak{M}_0(A+f) = \{\nu\}\} \tag{21}$$

is generic.

Remark : We point out that if one considers above, in the definition of G_2, potentials of the form $A(x,y) + l\,x$, where l is constant, instead of $A(x,y) + f(x)$, the same result is true for a generic $l \in \mathbb{R}$. This new statement is natural (and means something interesting) once it's common to consider a magnetization as a function of this form. In this way, for example, considering fixed the term $\frac{1}{2}(x-y)^2$, for a dense set of $l \in \mathbb{R}$, we have that the zero-temperature state for $A(x,y) = \frac{1}{2}(x-y)^2 + l\,x$ is unique.

Let us fix some notation : C is the set of continuous functions in $[0,1]^2$, $F = C^*$ the vector space of continuous functionals $\nu : C \to \mathbb{R}$, $E = C^2([0,1])$ provided with the C^2 topology, and G is the vector space of finite Borel signed measures on $[0,1]$. $K \subset G$ is the set of Borel probability measure on $[0,1]$, and note that $\mathbf{M} \subset F$. We denote by $F_A : \mathbf{M} \to \mathbb{R}$ the linear functional defined by $F_A(\nu) = \int A\,d\nu$. Note that $\mathfrak{M}_0(A)$ is the set of points of $\nu \in \mathbf{M}$ which maximize $F_A|_{\mathbf{M}}$, finally let $\pi : F \to G$ be the projection induced by $\pi_1 : [0,1]^2 \to [0,1]$.

Lemma 12. There exists a residual subset $\mathcal{O} \subset E$ such that, for all $f \in \mathcal{O}$, we have that

$$\#\pi(\mathfrak{M}_0(A+f)) = 1.$$

Proof: We just note that F_A is a affine subspace of dimension 0 of $\mathbf{M}^*$, then the proposition follows by theorem 5 of [BC]. $\square$

Note that in order to have (21), we need to prove that $\#\mathfrak{M}_0(A+f) = 1$.

Lemma 13. If the observable A is C^2, and $\frac{\partial^2 A}{\partial x \partial y} > 0$, then we have $\#\mathfrak{M}_0(A) = \#\pi(\mathfrak{M}_0(A))$

Proof: By lemma 10 we know that the restriction to $\cup_{\nu \in \mathfrak{M}_0} \mathrm{supp}(\nu)$ of the projection $[0,1]^2 \to [0,1]$ is a injective map. Hence the linear map $\pi : \mathfrak{M}_0(A) \to G$ is injective, and $\#\pi(\mathfrak{M}_0(A)) = \#\mathfrak{M}_0(A)$. $\square$

Proof of proposition 6: Note that by lemmas 12, and 13 we have that the set G_1 given in (21) is generic.

Let $f_0 \in G_1$, and $f_1 \in C^2([0,1])$ such that $f_1 \geq 0$ and $\{x; f_1(x) = 0\} = \pi_1(\mathrm{supp}(\nu))$. Then $\pi_1(\mathrm{supp}(\nu)) \subset \Omega(A + f_0 + f_1)$.

Claim: If $x_1 \notin \pi_1(\mathrm{supp}(\nu))$ then $x_1 \notin \Omega(A + f_0 + f_1)$.

Indeed, $f_1(x_1) > 0$, and

$$h^{(A+f_0+f_1)}(x_1,x_1) = \liminf_{k\to\infty} \left(\inf_{\mathcal{P}_k(x_1,x_1)} \sum_{i=1}^{k-1} (A+f_0+f_1-m)(x_i,x_{i+1}) \right) \geq$$

$$\liminf_{k\to\infty} \left(\inf_{\mathcal{P}_k(x_1,x_1)} \sum_{i=1}^{k-1} (A+f_0+-m)(x_i,x_{i+1}) + f_1(x_1) \right) =$$

$$h^{(A+f_0)}(x_1,x_1) + f_1(x_1) > 0.$$

Hence $\pi_1(\mathrm{supp}(\nu)) = \Omega(A+f_0+f_1)$. $\square$

Remark : We point out that if one consider the corresponding maximization problem among stationary Markov probabilities, that is, probabilities on the paths on $[0,1]^{\mathbb{N}}$, then, under the same hypothesis, we also get the same result. That is, the maximizing probability is unique. For the corresponding problem considering invariant probabilities for the shift, there is no uniqueness.

Proposition 7. If u is a calibrated backward-subaction, then for any x we have

$$u(x) = \sup_{p\in\Omega(A)} \{u(p) - h(p,x)\}.$$

Proof: For $(x_1,...,x_k) \in \mathcal{P}_k(x,\bar{x})$, we have

$$u(x_i) - u(x_{i+1}) \geq A(x_i,x_{i+1}) - m$$

and

$$u(x_k) - u(x_1) \leq -\sum_{i=1}^{k-1} A(x_i,x_{i+1}) - m$$

hence $u(\bar{x}) - u(x) \leq h(x,\bar{x})$, therefore

$$u(x) \geq \sup_{p\in\Omega(A)} \{u(p) - h(x,p)\}.$$

Now we show the another inequality. We denote by $x_1 = x$, as u is a backward calibrated subaction, there exists x_2 such that $u(x_1) = u(x_2) + A(x_1,x_2) - m$, recursively we can construct $(x_1,x_2,...,x_n,...)$ such that $u(x_n) = u(x_{n+1}) + A(x_n,x_{n+1}) - m$.

Let p be an accumulation point of the sequence $\{x_n\}$, we claim that $p \in \Omega(A)$. Indeed, if $x_{n_j} \to p$, fix $j > i$, we construct $(\tilde{x}_1,...,\tilde{x}_{n_j-n_i}) = (p,x_{n_{i+1}},...,x_{n_{j-i}},p)$ and hence

$$\sum_{i=1}^{n_j-n_i-1}(A-m)(\tilde{x}_i,\tilde{x}_{i+1})$$

$$=\sum_{k=n_i}^{n_{j-1}}(A-m)(x_k,x_{k+1})+A(p,x_{n_{i+1}})-A(x_{n_i},x_{n_{i+1}})+A(x_{n_{j-1}},p)-A(x_{n_{j-1}},x_{n_j})$$

$$=u(x_{n_j})-u(x_{n_i})+A(p,x_{n_{i+1}})-A(x_{n_i},x_{n_{i+1}})+A(x_{n_{j-1}},p)-A(x_{n_{j-1}},x_{n_j})$$

Then for $\epsilon>0$ fixed and i large enough we have that

$$|\sum_{i=1}^{n_j-n_i-1}(A-m)(\tilde{x}_i,\tilde{x}_{i+1})|\leq\epsilon$$

therefore $p\in\Omega(A)$.

Now take $(\tilde{x}_1,...,\tilde{x}_{n_j})=(x_1,x_2...,x_{n_j-1},p)$ then

$$-\sum_{i=1}^{n_j-1}(A-m)(\tilde{x}_i,\tilde{x}_{i+1})+u(x)-u(p)$$

$$=-\sum_{i=1}^{n_j-1}(A-m)(x_i,x_{i+1})+A(x_{n_j-1},x_{n_j})-A(x_{n_j-1},p)+u(x)-u(p)$$

$$=u(x_{n_j})-u(p)+A(x_{n_j-1},x_{n_j})-A(x_{n_j-1},p)$$

given $k>0$ there exists n_k such that

$$-\sum_{i=1}^{n_k-1}(A-m)(\tilde{x}_i,\tilde{x}_{i+1})\leq u(p)-u(x)+\frac{1}{k}$$

making $k\to\infty$ we obtain $h(x,p)\leq u(p)-u(x)$ then

$$u(x)=\sup_{p\in\Omega(A)}\{u(p)-h(x,p)\}.$$

□

Proposition 8. There exists a bijective correspondence between the set of calibrated backward-subactions and the set of functions $f\in C^0(\Omega(A))$ satisfying $f(y)-f(x)\leq h(x,y)$, for all points x,y in $\Omega(A)$.

The proof of this proposition is similar to the proof of the theorem 13 in [GL].

Now suppose that A has a unique maximizing measure ν_∞ and also that $\pi_1(\text{supp}(\nu_\infty))=\Omega(A)$. Using the monotonicity argument of proposition 10 bellow, one can define a measurable map $Y : \Omega(A) \to \Omega(A)$. Indeed, when x is such that there is unique y satisfying $(x, y) \in \mathrm{supp}(\nu_\infty)$, then $y = Y(x)$. In the other case, we define Y via the limit coming from the left side. Thus $(x, y) \in \mathrm{supp}(\nu_\infty) \Leftrightarrow y = Y(x)$.

Lemma 14. If A is generic in the Mañé sense, then the measure $\nu_\infty \circ \pi_1^{-1}$ is an invariant ergodic measure for Y.

Proof: First we prove the invariance: Let $f \in C^0(\Omega(A))$. We have:

$$\int f \circ Y(x)\, d\nu_\infty \circ \pi_1^{-1}(x) = \int f \circ Y(x)\, d\nu_\infty(x, y) = \int f(y)\, d\nu_\infty(x, y) =$$

$$= \int f(x)\, d\nu_\infty(x, y) = \int f(x)\, d\nu_\infty \circ \pi_1^{-1}(x)\,.$$

Now we will prove that Y is uniquely ergodic: Let η be a measure in the Borel sets of $\Omega(A)$ which is invariant for Y. If we define, for each Borel set A of $[0,1]^2$, $\nu(A) = \eta(\pi_1(A \cap \mathrm{supp}(\nu_\infty)))$, we have that ν is a measure probability in $[0,1]^2$ such that

(1) $\mathrm{supp}(\nu) \subset \mathrm{supp}(\nu_\infty)$

(2) $\pi_1(\nu) = \eta$.

(3) $\nu \in \mathbf{M}$.

In order to prove (3), consider $f \in C([0,1])$. We have

$$\int f(y) d\nu(x, y) = \int f(Y(x)) d\nu(x, y) = \int f(Y(x)) d\nu(x) =$$

$$= \int f(x) d\nu(x) = \int f(x) d\nu(x, y)\,,$$

where we used, in sequence: (1); (2) ; η is Y-invariant ; (2).

Note that for any calibrated backward-subaction u we have

$$\int A(x, y) d\nu(x, y) = \int (u(x) - u(y) + m)\, d_\nu(x, y) = m\,,$$

where in the second equality we used (1) and proposition 5, and, in the third equality we used (3). Thus we have that ν is a maximizing measure, and by uniqueness $\nu = \nu_\infty$. This implies $\eta = \pi_1(\nu_\infty)$, which shows that there exists an unique invariant measure for Y, which is a ergodic measure. □

Proposition 9. If $\nu \circ \pi_1^{-1}$ is an ergodic measure in $[0,1]$, and u, u' are two calibrated backward-subactions for A, then $u - u'$ is constant in $\pi_1(\mathrm{supp}(\nu))$.

For the proof of this proposition see theorem 17 of [GL].

Theorem 7. If A is generic in the Mañé sense, then the set of calibrated backward-subactions is unitary.

Proof: By the hypothesis ν_∞ is the unique maximizing measure, hence $\nu_\infty \circ \pi_1^{-1}$ is ergodic, and $\pi_1(\text{supp}(\nu_\infty)) = \Omega(A)$.

Let $f, f' : \Omega(A) \to \mathbb{R}$ continuous functions satisfying the hypothesis of proposition 8, then using this proposition we can get two calibrated subactions u_f, $u_{f'}$ such that $f - f' = u_f - u_{f'}$ in $\Omega(A)$, hence by proposition 9 $u_f - u_{f'}$ is constant in $\Omega(A)$, and again, from proposition 8, we show that the set of calibrated backward-subactions is unitary. □

This proves that $\bar{V}$ is unique, to prove that V is unique the arguments are very similar.

5 The Shift in the Bernoulli Space $[0,1]^{\mathbb{N}}$, and a Large Deviation Principle

Let us come back to the maximization problem in $\mathcal{M}_0$

$$\max_{\mu \in \mathcal{M}_0} \{ \int A d\mu \} . \tag{22}$$

The following proposition allow us to conclude that, generically in Mañé's sense, all such maximizing measures, after projection in the first two coordinates, are unique.

Proposition 10. Suppose that ν_∞ is a maximizing measure in $\mathbf{M}$.

(i) If A has an unique maximizing measure in $\mathbf{M}$, then any maximizing measure in $\mathcal{M}_0$ is projected by Π in ν_∞, where $\Pi : [0,1]^{\mathbb{N}} \to [0,1]^2$ is the projection in the first two coordinates.

(ii) ν_∞ can be extended to a maximizing measure $\mu_\infty \in \mathcal{M}_0$ which is an stationary Markov measure.

(iii) If ν_β is the family of measures given by (17), then this measures can be extended to absolutely continuous Markov measures μ_β, this sequence of measures weakly converge to the maximizing measure μ_∞.

Proof: The item (i) follows by items (b) and (c) of the proposition 1 and by proposition 6, the item (ii) follows by the item (a) of the proposition 1, and (iii) is a consequence of the remark after the proof of the proposition 1. □

From now on, until the end of this section, we will suppose that the maximizing measure ν_∞, and the functions V and $\bar{V}$ are unique. This is a generic property in Mañé sense.

Thus, for the maximization problem in the Bernoulli shift, we have shown the existence of a maximizing measure μ_∞ which can be approximated by absolutely continuous stationary Markov measures μ_β, which were explicitly calculated.

Now we will show a Large Deviation Principle for the family of measures $\{\mu_\beta\}$. We also exhibit a Large Deviation Principle for the bidimensional measures ν_β which, by the earlier sections, converge to ν_∞.

Lemma 15. Suppose $k \geq 2$. Let $F_k : [0,1]^k \to \mathbb{R}$ be the function given by

$$F_k(x_1, .., x_k) := \max(V + \bar{V}) - V(x_1) - \bar{V}(x_k) - \sum_{i=1}^{k-1}(A - m)(x_i, x_{i+1}).$$

Let $D_k = A_1....A_k$ be a cylinder of size k. Then, there exists the limit

$$\lim_{\beta\to\infty} \frac{1}{\beta} \log \mu_\beta(D_k) = - \inf_{(x_1,..,x_k)\in D_k} F_k(x_1, .., x_k).$$

Proof: Let us define

$$f_{k,\beta}(x_1, ..., x_k) := \frac{1}{\beta} \log \pi_\beta + \frac{k-1}{\beta} \log \lambda_\beta - \sum_{i=1}^{k-1} A(x_i, x_{i+1}) - \frac{1}{\beta} \log \varphi_\beta(x_1) - \frac{1}{\beta} \log \bar{\varphi}_\beta(x_k).$$

We have that $f_{k,\beta} \to F_k$ uniformly when $\beta \to \infty$. This is a consequence of the uniqueness of V and $\tilde{V}$.

We begin by proving the

Claim: Let $C_k = A_1....A_k$ be a cylinder of size k. We have

$$\limsup_{\beta\to\infty} \frac{1}{\beta} \log \mu_\beta(C_k) \leq - \inf_{(x_1,..,x_k)\in C_k} F_k(x_1, .., x_k)\,.$$

To prove the *Claim*, note that we have

$$\mu_\beta(C_k) = \int_{A_1} ... \int_{A_k} \frac{\varphi_{\beta(x_1)}\bar{\varphi}_{\beta(x_1)}}{\pi_\beta}\ \frac{e^{\beta A(x_1,x_2)}\bar{\varphi}_\beta(x_2)}{\bar{\varphi}_\beta(x_1)\lambda_\beta} ... \frac{e^{\beta A(x_{k-1},x_k)}\bar{\varphi}_\beta(x_k)}{\bar{\varphi}_\beta(x_{k-1})\lambda_\beta} dx_1...dx_k =$$

$$= \int_{A_1} ... \int_{A_k} e^{-\beta f_{k,\beta}(x_1,...x_k)} dx_1...dx_k \leq e^{-\beta \inf_{C_k} f_{k,\beta}(x_1,...x_k)} |C_k|\,, \qquad (23)$$

where $|C_k|$ denotes the Lebesgue measure of C_k. Hence

$$\frac{1}{\beta} \log \mu_\beta(C_k) \leq - \inf_{C_k} f_{k,\beta}(x_1, ...x_k) + \frac{1}{\beta} \log |C_k|,$$

and then, by the uniform convergence, we have:

$$\limsup_{\beta\to\infty} \frac{1}{\beta} \log \mu_\beta(C_k) \leq - \inf_{C_k} F_k(x_1, ...x_k)\,,$$

which finishes the proof of the *Claim*.

Now we will prove the lemma: if we fix $\delta > 0$, using the continuity of F_k we can find a point $(x_1, ..., x_k) \in D_k^0$ (the interior of D_k) such that

$$\inf_{D_k} F_k \leq F_k(x_1, ..., x_k) < \inf_{D_k} F_k + \delta \quad . \tag{24}$$

Now, let D_δ be a cylinder of size k, such that $(x_1, ..., x_k) \in D_\delta \subset D_k^0$, and

$$\inf_{D_k} F_k \leq F_k(y_1, ..., y_k) < \inf_{D_k} F_k + 2\delta \quad \forall (y_1, ..., y_k) \in D_\delta \, . \tag{25}$$

We have that

$$\mu_\beta(D_k) \geq \mu_\beta(D_\delta) \geq e^{-\beta \sup_{D_\delta} f_{k,\beta}(y_1,...,y_k)} |D_\delta| \quad ,$$

where the last inequality cames from (23). Now we use again the uniform convergence of $f_{k,\beta}$ to F_k in order to get

$$\liminf_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta(D_k) \geq -\sup_{D_\delta} F_k \, .$$

By (25), we get

$$\liminf_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta(D_k) \geq -\inf_{D_k} F_k - 2\delta \tag{26}$$

Sending $\delta \to 0$, and using the *Claim*, we finish the proof of the lemma. □

Note that if we set $k = 2$ above, we get a LDP for the family $\nu_\beta \to \nu_\infty$.

Theorem 8. Let $I : [0,1]^{\mathbb{N}} \to [0, +\infty]$ be the function defined by

$$I(\mathbf{x}) := \sum_{i \geq 1} V(x_{i+1}) - V(x_i) - (A - m)(x_i, x_{i+1}) \, .$$

Let $D = A_1....A_k$ be a cylinder of any size k. Then, there exists the limit

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

Note that lemma 5 show us that $I(\mathbf{x})$ is well defined (the sequence of partial sums of the series which define $I(\mathbf{x})$ is a non-decreasing sequence, for each $\mathbf{x}$).

In order to prove Theorem 8 we will need some new results and definitions.

For each $N \geq 2$, let us extend the function F_N to the space $[0,1]^{\mathbb{N}}$:

$$F_N(\mathbf{z}) := F_N(z_1, ..., z_n) = \sup(V + \bar{V}) - V(z_1) - \bar{V}(z_N) - \sum_{i=1}^{N-1} (A - m)(z_i, z_{i+1}) \, .$$

Lemma 16. $\forall\, \mathbf{z} \in [0,1]^{\mathbb{N}}$, we have

$$F_N(\mathbf{z}) \geq \max(V+\bar{V}) - (V(z_1)+\bar{V}(z_1)) \geq 0\,.$$

Proof: By lemma 5

$$\bar{V}(x) - \bar{V}(y) \geq A(x,y) - m \;\;, \forall x, y,$$

then

$$-\sum_{i=1}^{N-1}(A-m)(z_i, z_{i+1}) \geq \bar{V}(z_N) - \bar{V}(z_1)\,.$$

Hence, by definition of F_N

$$F_N(\mathbf{z}) \geq \max(V+\bar{V}) - (V(z_1)+\bar{V}(z_1))\,.$$

□

Lemma 17. (a) for a fixed $\mathbf{x} \in [0,1]^{\mathbb{N}}$, we have that

$$V(x_1) + \sum_{i=1}^{k-1}(A-m)(x_i, x_{i+1}) + \bar{V}(x_k)$$

is decreasing with respect to k.

(b) If $I(\mathbf{x}) < +\infty$, then there exists the limit

$$L(\mathbf{x}) = \lim_{k\to+\infty} V(\sigma^k(\mathbf{x}))) + \bar{V}(\sigma^k(\mathbf{x}))).$$

Proof: (a) This follows from lemma 5 (remember that $\tilde{m} = m$).

(b) We have

$$I(\mathbf{x}) = \sum_{i\geq 1} V(x_{i+1}) - V(x_i) - (A-m)(x_i, x_{i+1}) =$$

$$= \lim_{k\to+\infty} V(x_k) + \bar{V}(x_k) - \lim_{k\to\infty}\left(V(x_1) + \sum_{i=1}^{k-1}(A-m)(x_i, x_{i+1}) + \bar{V}(x_k)\right) \tag{27}$$

Hence, if $I(\mathbf{x}) < +\infty$, it follows, thanks to item (a), that $V(x_k)+\bar{V}(x_k) = V(\sigma^k(\mathbf{x}))) + \bar{V}(\sigma^k(\mathbf{x}))$ must converge.

□

Lemma 18. Suppose $I(\mathbf{x}) < +\infty$. Then, if we define, for each $M \in \mathbb{N}$, the probability measure

$$\mu_M = \frac{1}{M} \sum_{j=1}^{M-1} \delta_{\sigma^j(\mathbf{x})} ,$$

we have that $\Pi(\mu_M) \to \nu_\infty$ in the weak-$\star$ topology (where Π is the projection in the two first coordinates).

Proof: Given $\epsilon > 0$, there exists $N_\epsilon \in \mathbb{N}$ such that , for all $N \geq N_\epsilon$, and all $M > N$,

$$\sum_{i=N}^{M-1} V(x_{i+1}) - V(x_i) - (A-m)(x_i, x_{i+1}) < \epsilon .$$

Thus

$$V(x_M) - V(x_N) + (M-N)m < \sum_{i=N}^{M-1} A(\sigma^i(\mathbf{x})) + \epsilon ,$$

and

$$\frac{1}{M-N} \sum_{i=N}^{M-1} A(\sigma^i(\mathbf{x})) > m + \frac{V(x_M) - V(x_N)}{M-N} - \frac{\epsilon}{M-N} ,$$

and then we get that

$$\liminf_{M \to +\infty} \frac{1}{M} \sum_{i=1}^{M-1} A(\sigma^i(\mathbf{x})) \geq m .$$

Now we remember that

$$\frac{1}{M} \sum_{i=1}^{M-1} A(\sigma^i(\mathbf{x})) = \int A d\mu_M \leq m ,$$

and finally we get

$$\lim_{M \to +\infty} \int A d\mu_M = \lim_{M \to +\infty} \frac{1}{M} \sum_{i=1}^{M-1} A(\sigma^i(\mathbf{x})) = m .$$

If we use the compactness of the closed ball of radius 1 in the weak-$\star$ topology, we get that $\{\mu_M\}$ has convergent subsequences. Any limit of a convergent subsequence is a stationary measure (a σ-invariant measure) and must be a maximizing measure, by the last equality. As any maximizing measure is projected in μ_∞ by Π, we get the lemma. $\square$

Proposition 11. If $I(\mathbf{x}) < +\infty$, then

$$\lim_{k \to +\infty} V(\sigma^k(\mathbf{x})) + \bar{V}(\sigma^k(\mathbf{x})) = \max(V + \bar{V}).$$

Proof: Let $\mathbf{z} = (z_1, z_2, z_3, ...) \in \mathrm{supp}(\mu_\infty)$. We have that $(z_1, z_2) \in \mathrm{supp}(\nu_\infty)$. Thus, by lemma 18 there exists a sub-sequence such that $\Pi(\sigma^{k_l}(\mathbf{x})) \to (z_1, z_2)$.

Fix $\epsilon > 0$. Let

$$B_{k_l,\epsilon}(\mathbf{x}) := \{\mathbf{y} \in [0,1]^{\mathbb{N}} \ : \ |y_j - x_{j+k_l}| \leq \epsilon, \ \forall \, 1 \leq j \leq 2\}$$

be the closed cylinder of size 2 'centered' at $\sigma^{k_l}(\mathbf{x})$.

If l is big enough, we have that

$$B_{k_l,\epsilon}(\mathbf{x}) \subset \{\mathbf{y} \in [0,1]^{\mathbb{N}} \ : \ |y_j - z_j| \leq 2\epsilon, \ \forall \, 1 \leq j \leq 2\} \, .$$

Note that $\mu_\infty(B_{k_l,\epsilon}(\mathbf{x})) = \nu_\infty(B_{k_l,\epsilon}(\mathbf{x})) > 0$, and thus using Lemma 15 with $k = 2$, it follows that there exists a point $(z_{1,\epsilon}, z_{2,\epsilon}, z_{3,\epsilon}, z_{4,\epsilon}, ...) \in B_{k_l,\epsilon}(\mathbf{x})$, such that $F_2((z_{1,\epsilon}, z_{2,\epsilon})) = 0$.

Then, we can use the fact that F_2 depends only on its first 2 coordinates in order to obtain that $F_2(\mathbf{w}_\epsilon) = 0$, where $\mathbf{w}_\epsilon = (z_{1,\epsilon}, z_{2,\epsilon}, z_3, z_4, ...)$ is defined by the point of $[0,1]^{\mathbb{N}}$ whose first 2 coordinates are equal to those of $(z_{1,\epsilon}, z_{2,\epsilon})$, while the other coordinates are equal to those of $\mathbf{z}$.

Now, if we send $\epsilon \to 0$, we have that $\mathbf{w}_\epsilon \to \mathbf{z}$. Thus we can use the continuity of F_N to get that $F_2(\mathbf{z}) = 0$.

Using again the continuity of F_2, we have that $F_2(\sigma^{k_l}(\mathbf{x})) \to 0$.

Lemma 16 shows that

$$\lim_{l \to +\infty} V(\sigma^{k_l}(\mathbf{x})) + \bar{V}(\sigma^{k_l}(\mathbf{x})) = \max(V + \bar{V}),$$

and finally using Lemma 17(b) we prove proposition 11.

□

Proof of theorem 8: First we need to prove the following claim.

Claim:

$$I(\mathbf{x}) = \max(V + \bar{V}) - \lim_{k \to \infty} \left(V(x_1) + \sum_{i=1}^{k-1} (A - m)(x_i, x_{i+1}) + \bar{V}(x_k) \right) .$$

In order to prove the *Claim*, we have to consider two possibilities: if $I(\mathbf{x}) < +\infty$, then (27) can be combined with proposition 11 to give the *Claim*. If $I(\mathbf{x}) = +\infty$, we just have to use the expression

$$I(x) = \lim_{k \to \infty} \left(V(x_k) - V(x_1) - \sum_{i=1}^{k-1} (A - m)(x_i, x_{i+1}) \right) .$$

Thanks to Lemma 15, we just have to show that

$$- \inf_{(x_1, .., x_k) \in D} F_k(x_1, .., x_k) = - \inf_{\mathbf{x} \in D} I(\mathbf{x}) \, .$$

We begin by proving that

$$-\inf_{(x_1,..,x_k)\in D} F_k(x_1,..,x_k) \le -\inf_{\mathbf{x}\in D} I(\mathbf{x})\,.$$

Given $\delta > 0$, there exists a point $(y_1, ..., y_k) \in D$ such that

$$F_k(y_1, ..., y_k) < \inf_{(x_1,..,x_k)\in C} F_k(x_1,..,x_k) + \delta\,.$$

By the definition of F_k,

$$F_k(y_1, ..., y_k) = \max(V + \bar V) - V(y_1) - \bar V(y_k) - \sum_{i=1}^{k-1}(A - m)(y_i, y_{i+1}).$$

For each $j \ge k$ we choose a y_{j+1} that satisfies $\bar V(y_j) = \bar V(y_{j+1}) + A(y_j, y_{j+1}) - m$. Then we define $\mathbf{y} := (y_1, ...y_k, y_{k+1}, ...)$.

Second Claim: $I(\mathbf{y}) = F_k(y_1, ...y_k)$. Indeed,

$$F_k(y_1, ...y_k) = \max(V + \bar V) - \left(V(y_1) + \bar V(y_k) + \sum_{i=1}^{k-1}(A - m)(y_i, y_{i+1})\right) =$$

$$= \max(V + \bar V) - \left(V(y_1) + \bar V(y_j) + \sum_{i=1}^{j-1}(A - m)(y_i, y_{i+1})\right), \quad \forall j \ge k.$$

Then, from the reasoning above and the way we choose $\mathbf{y}$, we get that $F_k(y_1, ...y_k)$ is equal to

$$\max(V + \bar V) - \lim_{j\to\infty}\left(V(y_1) + \bar V(y_j) + \sum_{i=1}^{j-1}(A - m)(y_i, y_{i+1})\right) = I(\mathbf{y})\,.$$

This implies that

$$-\inf_{(x_1,..,x_k)\in D} F_k(x_1,..,x_k) < -I(\mathbf{y}) + \delta \le -\inf_{\mathbf{x}\in D} I(\mathbf{x}) + \delta\,.$$

Making $\delta \to 0$, we have the first inequality.

Now, we will prove the second inequality:

$$-\inf_{\mathbf{x}\in D} I(\mathbf{x}) \le -\inf_{(x_1,..,x_k)\in D} F_k(x_1,..,x_k)\,.$$

In order to prove this, we use Lemma 17(a), and then we get, by the *Claim*,

$$I(\mathbf{x}) = \max(V + \bar V) - \lim_{j\to\infty}\left(V(x_1) + \sum_{i=1}^{j-1}(A - m)(x_i, x_{i+1}) + \bar V(x_j)\right) \ge$$

$$\geq \max(V+\bar{V}) - \left(V(x_1) + \sum_{i=1}^{k-1}(A-m)(x_i,x_{i+1}) + \bar{V}(x_k)\right) = F_k(x_1,..,x_k).$$

□

Here finally we can give the proofs of theorems 1 and 3:

Proof of theorem 1:(a) This is item (i) of proposition 10.

(b) Theorem 7 shows that the set of backward calibrated subactions is unitary , the proof that the set of forward calibrated subactions is unitary is similar.

Proof of theorem 3: (a) It follows by items (ii) and (iii) of the proposition 10 and theorem 5.

(b) Here we use theorem 8, where V is a calibrated subaction, which is unique by theorem 1 item (b).

6 Monotonicity of the graph, and separating subactions

Suppose A is C^2 and satisfies

$$\frac{\partial^2 A}{\partial x \partial y}(x,y) > 0,$$

then, for all $x<x', y<y'$ we have that

$$A(x,y) + A(x',y') > A(x,y') + A(x',y). \tag{28}$$

Let $\bar{V}$ be the calibrated backward-subaction define above.

As A is C^2, then $\bar{V}$ is Lipschitz, hence $\bar{V}$ is differentiable λ-a.e., where λ is the Lebesgue measure. Let $Dom(D\bar{V})$ be the set of points where $\bar{V}$ is differentiable.

Following the proof of theorem 6 we have that for $x \in Dom(D\bar{V})$, there exists only one $y(x)$ such that

$$\bar{V}(x) = A(x,y(x)) + \bar{V}(y(x)) - m. \tag{29}$$

Proposition 12. The function $Y: Dom(D\bar{V}) \to [0,1]$, defined by $Y(x) = y(x)$, $y(x)$ satisfying (29), is monotone nondecreasing.

Proof: Let $x<x'$, let us call $z=Y(x), z'=Y(x')$, suppose that $z>z'$. We know that

$$\bar{V}(x) = A(x,z) + \bar{V}(z) - m \quad , \quad \bar{V}(x') = A(x',z') + \bar{V}(z') - m,$$

and

$$\bar{V}(x) \geq A(x,z') + \bar{V}(z') - m \quad , \quad \bar{V}(x') \geq A(x',z) + \bar{V}(z) - m.$$

Adding the first two equation and comparing with the summation of the last two, we get that

$$A(x,z) + A(x',z') \geq A(x,z') + A(x',z),$$

for $x < x', z' < z$, which is a contradiction with (28). $\square$

If we assume that $\frac{\partial^2 A}{\partial x \partial y}(x,y) < 0$, then a function $Y(x)$ as above can be defined, and it will be monotone non-increasing.

Now we will show the existence of a separating subaction (see [GLT] and [GLM] for related results). The idea is: given a potential A find a subaction u such that in the cohomological equation the equality just happen in points x that are on $\Omega(A)$ (where it has to happen, anyway). In this way, we have a criteria to separate points of $\Omega(A)$ from the other ones. One can then consider a new potential $\tilde{A} = A(x,y) + u(x) - u(y)$ where the maximum of $\tilde{A}$ is exactly attained in $\Omega(\tilde{A})$.

Definition 19. (a) A continuous function $u : [0,1] \to \mathbb{R}$ is called a forward-subaction if, for any $x, y \in [0,1]$ we have

$$u(y) \geq A(x,y) + u(x) - m. \tag{30}$$

(b) A continuous function $u : [0,1] \to \mathbb{R}$ is called a backward-subaction if, for any $x, y \in [0,1]$ we have

$$u(x) \geq A(x,y) + u(y) - m. \tag{31}$$

The proof we presented here is similar to the one in [GLM].

Definition 20. We say that a forward subaction u is separating if

$$\max_x [A(x,y) + u(x) - u(y)] = m \iff x \in \Omega(A),$$

and a backward subaction u is separating if

$$\max_y [A(x,y) + u(y) - u(x)] = m \iff y \in \Omega(A).$$

We will show the existence of a separating backward-subaction.

Lemma 19. If $x \in \Omega(A)$ there exists $\mathbf{x} = (x_1, ..., x_k, ...) \in [0,1]^{\mathbb{N}}$ such that $x_1 = x$ and

$$h(x_k, x_1) \leq \sum_{i=1}^{k-1} (A - m)(x_i, x_{i+1}).$$

Proof: If $x \in \Omega(A)$, then there exists a sequence of paths $\{(x_1^n, ..., x_{j_n}^n)\}_{n\in\mathbb{N}}$ such that $x_1^n = x_{j_n}^n = x$ and $j_n \to \infty$ satisfying

$$\sum_{j=1}^{j_n-1} (A-m)(x_j^n, x_{j+1}^n) \to 0. \tag{32}$$

Because $|x_j^n| \le 1$, there exists a ray $(x_1, ..., x_k, ...)$ which is the limit of the paths above, the convergence being uniform in each compact part.

Fixed $k \in \mathbb{N}$, for $j_n > k$, we have that

$$S^{j_n-k}(x_k, x_1) \le -A(x_k, x_{k+1}^n) + m - \sum_{j=k+1}^{j_n-1} (A-m)(x_j^n, x_{j+1}^n),$$

and

$$S^{j_n-k}(x_k, x_1) + \sum_{j=1}^{j_n-1} (A-m)(x_j^n, x_{j+1}^n) \le$$

$$-A(x_k, x_{k+1}^n) + m + \sum_{j=1}^{k} (A-m)(x_j^n, x_{j+1}^n).$$

Hence taking the $\liminf_{n\to\infty}$ and using (32) we obtain

$$h(x_k, x_1) \le \sum_{j=1}^{k-1} (A-m)(x_j, x_{j+1}).$$

□

Lemma 20. Let u be any backward-subaction, then for all $x \in \Omega(A)$ we have

$$\max_y \{u(y) - u(x) + A(x,y)\} = m.$$

Proof: As u satisfies equation (31), for any $(x_1, ..., x_k) \in \mathcal{P}_k(x,y)$ we have that $u(y) - u(x) \le -\sum_{j=1}^{k-1} (A-m)(x_i, x_{i+1})$, hence $u(y) - u(x) \le h(x,y)$.

Let $x \in \Omega(A)$ and let $\mathbf{x} = (x_1, ..., x_k, ...)$ be the point in $[0,1]^{\mathbb{N}}$ which exists by the proposition (19).

By the lemma (19) we have that

$$u(x_1) - u(x_k) \le h(x_k, x_1) \le \sum_{j=1}^{k-1} (A-m)(x_j, x_{j+1}),$$

and, as it is a backward-subaction,

$$u(x_k) - u(x_1) \le -\sum_{j=0}^{k-1} (A-m)(x_j, x_{j+1}),$$

in particular, for $k = 1$,

$$u(x_2) - u(x_1) = -A(x_1, x_2) + m.$$

This implies

$$\max_y \{u(y) - u(x) + A(x, y)\} = m.$$

$\square$

Lemma 21. If the observable A is Holder continuous, then the function $S_x(\cdot) := S(x, \cdot)$ is uniformly Holder and with the same Holder constant of A.

Proof: Let us fix x, $\epsilon > 0$ and $y, z \in [0, 1]$, then there exists $(x_1, ..., x_k) \in \mathcal{P}_k(x, y)$ such that

$$| - \sum_{i=1}^{k-1} (A - m)(x_i, x_{i+1})| \leq S(x, y) + \epsilon.$$

Consider now the following path: $(\tilde{x}_1, ..., \tilde{x}_k) = (x_1, ..., x_{k-1}, z) \in \mathcal{P}_k(x, z)$, then

$$- \sum_{i=1}^{k-1} (A - m)(\tilde{x}_i, \tilde{x}_{i+1}) = - \sum_{i=1}^{k-1} (A - m)(x_i, x_{i+1}) + A(x_{k-1}, y) - A(x_{k-1}, z),$$

therefore,

$$S(x, z) \leq - \sum_{i=1}^{k-1} (A - m)(\tilde{x}_i, \tilde{x}_{i+1}) \leq S(x, y) + \epsilon + \mathrm{Hol}_\alpha(A)|z - y|^\alpha, \quad \forall \epsilon,$$

i.e., $S(x, y) - S(x, z) \leq \mathrm{Hol}_\alpha(A)|z - y|^\alpha$. Changing the role of y and z we obtain $|S(x, y) - S(x, z)| \leq \mathrm{Hol}_\alpha(A)|z - y|^\alpha$, which give us the Holder continuity of S_x, independently of x. $\square$

Theorem 9. If the observable A is Holder continuous, there exists a separating backward-subaction.

Proof: By definition,

$$S(x, y) \leq -A(x, y) + m \quad \forall\, y \in [0, 1]$$

If $x \notin \Omega(A)$, then $S(x, x) > 0$. Hence

$$S_x(y) - S_x(x) < -A(x, y) + m \quad \forall\, y \in [0, 1]$$

As $\Omega(A)$ is a closed set, then for each $x \notin \Omega(A)$ we can find a neighborhood $V_x \subset [0,1]\backslash\Omega(A)$ of x such that

$$S_x(y) - S_x(z) < -A(z,y) + m, \quad \forall\, y \in [0,1], \forall\, z \in V_x.$$

We can extract, from the family of these neighborhoods $\{V_x\}_{x\notin\Omega(A)}$, a countable family $\{V_{x_j}\}_{j=1}^{\infty}$ which is a covering of $[0,1]\backslash\Omega(A)$.

We define

$$\tilde{S}_{x_j}(z) = S_{x_j}(z) - S_{x_j}(0)$$

as S_{x_j} is uniformly Holder, this implies that $|\tilde{S}_{x_j}(z)| \leq \mathrm{Hol}_\alpha(A) z^\alpha, \quad \forall\, x_j$, therefore the series

$$u(z) = \sum_{j=1}^{\infty} \frac{\tilde{S}_{x_j}(z)}{2^j}$$

is well defined and uniformly convergent, because $[0,1]$ is compact. Note that u is a infinite convex combination of backward-subactions $\tilde{S}_{x_j}$, then u is also a backward-subaction.

Fix $x \in [0,1]\backslash\Omega(A)$, there exists $k \geq 1$ such that $x \in V_{x_k}$. Now, $\forall y \in [0,1]$ we have

$$u(y)-u(x) = \sum_{j=1}^{\infty} \frac{S_{x_j}(y) - S_{x_j}(x)}{2^j} = \frac{S_{x_k}(y) - S_{x_k}(x)}{2^k} + \sum_{j\neq k} \frac{S_{x_j}(y) - S_{x_j}(x)}{2^j} <$$

$$\frac{-A(x,y)+m}{2^k} + \sum_{j\neq k} \frac{-A(x,y)+m}{2^j} < -A(x,y) + m.$$

Hence,

$$\max_y \{u(y) - u(x) + A(x,y)\} < m, \quad \text{if} \quad x \notin \Omega(A),$$

and, as u is a backward-subaction, we have by lemma 20 that

$$\max_y \{u(y) - u(x) + A(x,y)\} = m, \qquad \text{if} \qquad x \in \Omega(A).$$

□

References

[A] S. Adams. *Mathematical Statistical Mechanics.* Max-Plank-Institut fur Math , 2006.

[AL] K. Athreya and S. Lahiri. *Measure Theory and Probability Theory.* Springer Verlag, 2006.

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

[BBNg] I. Brevik, J-M Borven and S. Ng. *Viscous Brane Cohomology with a Brane-Bulk energy interchange term*, General Relativity and Gravitation, V. 38, N. 5, 907-915(9) 2006

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

[Ba] P. Bhattacharya and M. Majumdar. *Random Dynamical Systems*. Cambridge Univ. Press, 2007.

[Cv] M. Cveti, S. Nojiri and S. D. Odintsov, *Black hole thermodynamics and negative entropy in de Sitter and anti-de Sitter EinsteinGaussBonnet gravity*, Nuclear Physics B, V. 628, Issues 1-2, 295-330, , 2002

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

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

[CLT] 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.

[CG] J. P. Conze, Y. Guivarc'h. *Croissance des sommes ergodiques et principe variationnel*, manuscript circa 1993.

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

[Dellach] C. Dellacherie. *Probabilities and potential*. North-Holland, 1978

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

[EvG] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, 1991

[Ev] L. C. Evans. *Weak Convergence Methods for Nonlinear Partial Differential Equations*. Published for the Conference Board of the Mathematical Sciences, Whashington, DC, 1990.

[Fathi] 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.

[FS] A. Fathi and A. Siconolfi. *Existence of C^1 critical subsolutions of the Hamilton-Jacobi equations.* Inv. Math. 155, pp 363-388, 2004.

[GL] 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).

[GLT] E. Garibaldi, A. O. Lopes and P. Thieullen. *On separating sub-actions.* Preprint 2006. To appear.

[GLM] D. Gomes, A. O. Lopes and J. Mohr. *The Mather measure and a Large Deviation Principle for the Entropy Penalized Method.* Preprint 2007. To appear.

[Go] C. Gole. *Sympletic twist maps.* World Sci. Pub Co Inc, 1998.

[Gom] D. A. Gomes. *Viscosity Solution methods and discrete Aubry-Mather problem.* Discrete Contin. Dyn. Syst., 13 (1): 103-116, 2005.

[Gom1] D. A. Gomes. *Calculus of Variations.* 2006

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

[Hop] E. Hopf. *An inequality for Positive Linear Integral Operators*, Journal of Mathematics and Mechanics, V. 12. N.5, 683-692, 1963

[Jen1] O. Jenkinson. Ergodic optimization, *Discrete and Continuous Dynamical Systems, Series A*, 15, 197-224, 2006

[Ju] G. Jumarie. *Relative Information*, Springer Verlag, 1990

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

[Lu] E. Lubkin. *Negative entropy, energy, and heat capacity in connection with surface tension: Artifact of a model or real?*, Inter. Journal of Theoretical Physics, V. 26, N. 5, 1987

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

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

[Mo] I. D. Morris. *A sufficient condition for the subordination principle in ergodic optimization*, Bull. Lond. Math. Soc. 39, no. 2, pp 214-220, 2007

[Mi] I. Mitra. *Introduction to dynamic optimization theory*, Optimization and Chaos, Editors M. Majumdar, T. Mitra and K. Nishimura, Springer Verlag, pp 31-108, 2000

[Ni] R. K. Niven. *Cost of s-fold Decisions in Exact Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac Statistics*, Physica A, Volume 365, Issue 1, pp 142-149, 2006

[Os] A. Ostrowski. *On positive matrices*, Math. Annalen, V. 150, 276-284, 1963

[Pe] M. Pettini. *Geometry and topology in Hamiltonian dynamics and statistical mechanics*, Springer Verlag, 2007.

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

[Ra] S. Rachev and L. Ruschendorf. *Mass transportation problems, Vol I and II.* Springer Verlag, 1998.

[RRS] S. Risau-Gusman, A. C. Ribeiro-Teixeira and D. A. Stariolo. *Topology and Phase Transitions: The Case of the Short Range Spherical Model.* Journ. of Statist. Physics, 2006

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

[Sp] F. Spitzer. *A Variational characterization of finite Markov chains.* The Annals of Mathematical Statistics. (43): N.1 303-307, 1972.

[Ta] M. Takahashi. *Thermodynamics of one-dimensional solvable models*, Cambridge Press, 2005.

[Th] C. Thompson. *Infinite-Spin Ising Model in one dimension.* Journal of Mathematical Physics. (9): N.2 241-245, 1968.

[V] A. van Enter, S. Romano and V. Zagrebnov. *First-order transitions for some generalized XY models.* J. Phys. A 39, no. 26, 439–445, 2006.

[W] W. F. Wrezinski and E. Abdalla. *A precise formulation of the third law of thermodynamics with applications to statistical physics and black holes.* Preprint USP 2007.