Stochastics and Dynamics

A LARGE DEVIATION PRINCIPLE FOR THE EQUILIBRIUM STATES OF HÖLDER POTENTIALS: THE ZERO TEMPERATURE CASE

ALEXANDRE BARAVIERA, ARTUR O. LOPES and PHILIPPE THIEULLEN *
Instituto de Matemática, UFRGS,
Av. Bento Gonçalves 9500, 91500, Porto Alegre, RS, Brazil
** Institut de Mathématiques, LaBAG, CNRS, Université Bordeaux 1,*
341 cours de la libération, F-33405 Talence, France

Received ?
Revised ?

Consider a α-Hölder function $A : \Sigma \to \mathbb{R}$ and assume that it admits a unique maximizing measure $\mu_{\max}$. For each β, we denote μ_β, the unique equilibrium measure associated to βA. We show that (μ_β) satisfies a Large Deviation Principle, that is, for any cylinder C of Σ,

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

where

$$I(x) = \sum_{n\geq 0} \big(V\circ\sigma - V - (A-m)\big)\circ \sigma^n(x), \quad m = \int A\, d\mu_{\max}$$

where $V(x)$ is any strict subaction of A.

Keywords: Equilibrium state, Large Deviation Principle, Zero temperature, Subaction, Maximizing measure, Dual Potential, W Kernel.

1. Introduction

Let $\Sigma = \{x \in \{1, 2, ..., r\}^{\mathbb{N}} \mid M(x_i, x_{i+1}) = 1 \text{ for all } i \geq 0\}$ be a subshift of finite type on r symbols and transition matrix M, and σ the left-shift acting on Σ defined by $\sigma(x_0, x_1, x_2, ...) = (x_1, x_2, x_3, ...)$. Let $A : \Sigma \to \mathbb{R}$ be a fixed $\alpha-$Hölder function which we call observable. The matrix M takes values in $\{0, 1\}$ and indicates whether a transition $i \to j$, $i, j \in \{1, 2, \cdots, r\}$, is allowed :

$$i \to j \text{ is allowed } \iff M(i,j) = 1.$$

A probability σ-invariant measure $\mu_{\max}$ is said to be maximizing if

$$m(A) := \int A\, d\mu_{\max} = \sup\{\int A\, d\mu \mid \mu \text{ invariant for } \sigma\}.$$

*E-mail: baravi@mat.ufrgs.br, alopes@mat.ufrgs.br and Philippe.Thieullen@math.u-bordeaux1.fr

A general α-Hölder function A may admits several maximizing measures. Nevertheless, as it is shown in [10], a "generic"(in a certain space and Hölder topology) function A admits a unique maximizing measure which has a support on a periodic orbit for σ.

In recent years several results about maximizing probabilities for discrete time expanding systems were obtained [4 5 7 16 17 21 24 27]. Some of these results are discrete time versions of similar ones appearing in the so called Aubry-Mather Theory. Analogous concepts like Aubry set, Peierls barrier, viscosity solutions, etc, appear in this setting in a natural way [10]. Subactions (to be defined later) play an important role in the proof of all these results [9].

A connection of these two settings in a particular case is described in [21].

Questions related to maximizing Lyapunov probabilities, that is, the ones maximizing $\int \log T'\, d\mu$, appear in [10]. Questions related to ground states of C^*-Algebras can be found in [12].

The existence of subactions for Anosov Difeos is considered in [19] and for Anosov Flows in [24 20].

The above list of results does not exhaust all the recent works in the subject.

For a fixed potential A, if one considers the equilibrium probability μ_β for the pressure of βA, it is natural to ask what can be said about the weak limits μ_∞ of μ_β, when $\beta \to \infty$. This question, which was addressed in [10 7 17] and by others, is quite related to maximizing probabilities. The main purpose of the present paper is to analyze large deviation properties of μ_β, when $\beta \to \infty$.

We choose in the sequel the following assumption:

Assumption 1. *The Hölder function A admits a unique maximizing measure μ_{max}. (We nevertheless do not make any assumption on the support of that measure.)*

For instance the above assumption is true for observables A constant on a periodic orbit and strictly smaller elsewhere.

We denote by $\mathcal{L}_A$ the Ruelle-Perron operator corresponding to A and $\mathcal{L}_A^*$ its dual operator acting on finite measures:

$$\mathcal{L}_A(\phi)(x_0, x_1, \cdots) = \sum_{y_1 \to x_0} \exp\big(A(y_1, x_0, x_1, \cdots)\big)\phi(y_1, x_0, x_1, \cdots).$$

We refer the reader to [23] for the results we use about Thermodynamic Formalism. We denote also by $\lambda(A)$, ϕ_A and ν_A, respectively, the largest eigenvalue of $\mathcal{L}_A$, the corresponding unique eigenfunction and eigenmeasure of $\mathcal{L}_A$ and $\mathcal{L}_A^*$ normalized by $\int \phi_A\, d\nu_A = 1$ and $\nu_A(\Sigma) = 1$,

$$\mathcal{L}_A(\phi_A) = \lambda(A)\phi_A, \quad \mathcal{L}_A^*(\nu_A) = \lambda(A)\nu_A.$$

The probability measure $\mu_A = \phi_A \nu_A$ is σ-invariant and maximizes the pressure of A. We call ν_A the Gibbs measure of A and μ_A the equilibrium measure of A.

Later we will simplify the notations by introducing λ_β, ϕ_β, ν_β, the largest eigenvalue, the corresponding eigenfunction and eigenmeasure of the function βA. Without assuming uniqueness of the probability measure, it is known [10] that any weak subsequence limit of the (μ_{β_n}) converges, when n goes to infinity, to a maximizing measure.

In fact, here we need less than assumption 1. It is enough that the weak limit of Gibbs states for βA, with $\beta \in \mathbb{R}$, converges to a unique measure, when $\beta \to \infty$. Then, our result would be for this special probability. Reference [7],[17] address the question of such uniqueness in a particular.

According to our assumption 1, we know that (μ_β) converges to the unique maximizing measure $\mu_{\max}$. Our purpose is to show that (μ_β) satisfies a Large Deviation Principle. Let us recall first the definition of this principle:

Definition 2. *We say a one-parameter family of probability measures (μ_β) which converges to some measure ρ when $\beta \to \infty$, satisfies a Large Deviation Principle with deviation function $I : \Sigma \to \mathbb{R}$, if for any cylinder $C \subset \Sigma$*

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

for some non-negative lower semi-continuous function $I(x)$.

A general reference for Large Deviation results in Ergodic Theory is [22].

Considerations about weak limits of equilibrium states μ_β with $\beta \to \infty$ are called the analysis of the temperature zero case [7], [8], [13], [14] [5], [?], [17]. This is so because in Statistical Mechanics β is (up to a physical constant) the inverse of temperature. We also point out the similar results obtained by N. Anantharanam, [1], [2], who also studies a Large Deviation Principle in the Lagrangian case and shows an upper large deviation inequality.

After this paper was accepted we received the information that the present Proposition 7 appears in some form in [15].

2. Main results

In order to write a precise formula for the deviation function $I(x)$, we need the basic notion of subaction and more precisely the notion of strict subaction. A subaction for $A(x)$ is a real continuous function U on Σ such that $A \leq U \circ \sigma - U + m(A)$ everywhere on Σ. A strict subaction possesses a stronger property:

Definition 3. *A continuous function $V : \Sigma \to \mathbb{R}$ is called strict subaction if*

$$V(x) = \max_{y \,:\, \sigma(y) = x} \big(V(y) + A(y) - m(A) \big).$$

(In other terms, V is a subaction and for any $x \in \Sigma$ there exists $y \in \Sigma$ such that $\sigma(y) = x$ and $V(y) + A(x) - m(A) = V(x)$).

Subactions can play an important role in Large Deviation problems as we will see here.

There are several ways to construct subactions. For instance, it is shown in [10], that any accumulation point of $(\frac{1}{\beta} \log \phi_\beta)$, for the uniform convergence topology, is a strict subaction V for A. That is there exists a subsequence of (β_n) such that, uniformly on Σ, the following limit exists:

$$V := \lim_{n \to +\infty} \frac{1}{\beta_n} \log \phi_{\beta_n}.$$

From the proof presented here it follows that the limit of above sequence does not depend of the sequence β_n.

Our main purpose in this paper is to prove that the sequence (μ_β) satisfies a Large Deviation Principle:

Theorem 4. *Let $A : \Sigma \to \mathbb{R}$ be a Hölder observable admitting a unique maximizing measure μ_{max}. Then for any cylinder C of Σ,*

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

where $I(x)$ is given by ($m := m(A)$ to simplify the notation)

$$I(x) = \sum_{n \geq 0} \big(V \circ \sigma - V - (A - m)\big) \circ \sigma^n(x)$$

and $V(x)$ is any strict subaction for A. (Notice that each term in the previous sum is nonnegative.)

The proof of this theorem is very indirect and uses the notion of dual shift and (what we call) the W kernel. We explain in the rest of this section the plan of this proof.

We first explain why $V(x)$ can be any subaction in theorem 4. Let

$$S(p,x) := \lim_{\epsilon \to 0} \sup_{n \geq 1,\ x' \in \Sigma} \Big\{ \sum_{k=0}^{n-1} (A - m) \circ \sigma^k(x') \mid d(x', p) < \epsilon,\ \sigma^n(x') = x \Big\}$$

for any $p, x \in \Sigma$. We showed, see [10] for instance, that $S(p,x)$, as a function of x, is finite and Hölder as soon as p belongs to the support of some maximizing measure. We have

Proposition 5. *Let $A : \Sigma \to \mathbb{R}$ be a Hölder observable admitting a unique maximizing measure μ_{max}. Then for any strict subaction V, $x \in \Sigma$ and $p \in \operatorname{supp}(\mu_{max})$*

$$V(x) = V(p) + S(p,x).$$

In particular, any two strict subactions differ by a constant.

The second notion we need is called W kernel. This function is defined on the natural extension of Σ.

Consider

$$\Sigma^* = \{(y_1, y_2, \cdots) \in \{1 \cdots r\}^{\mathbb{N}} \mid M(y_{i+1}, y_i) = 1,\ \text{for all } i \geq 1\ \},$$

which represents also the Bernoulli space but which indexes increasing in the negative direction of the lattice $\mathbb{Z}$. The reader should consider the pair $(y,x) = (...y_3, y_2, y_1, x_0, x_1, x_2, ..)$ as an element on $\Sigma^* \times \Sigma \subset \{1 \cdots r\}^{\mathbb{Z}}$.

It will convenient later to define a function $M(y,x)$, on $\Sigma^* \times \Sigma$, by $M(y,x) = M(y_1, x_0)$, so that $M(y,x) = 1$ if $(y,x) \in \hat{\Sigma}$ and $M(y,x) = 0$ elsewhere.

We prefer to introduce first the dual Markov chain (Σ^*, σ^*) of transition matrix M^*, the transpose of M, and then defined the natural extension $\hat{\Sigma}$ as a subset of $\Sigma^* \times \Sigma$. We thus introduce

$$\hat{\Sigma} = \{(y,x) \in \Sigma^* \times \Sigma \mid M(y_1, x_0) = 1\}$$

We just have defined the two-sided subshift of finite type and of transition matrix M. It is convenient to write points $y \in \Sigma^*$ and $x \in \Sigma$ in the form

$$y = \langle \cdots, y_3, y_2, y_1|, \quad x = |x_0, x_1, x_2, \cdots\rangle.$$

The left shift has the following definition using these notations

$$\hat{\sigma}(y,x) := (\tau_x^*(y), \sigma(x)), \quad \tau_x^*(y) := \langle \cdots, y_2, y_1, x_0|$$
$$\hat{\sigma}^{-1}(y,x) = (\sigma^*(y), \tau_y(x)), \quad \tau_y(x) := |y_1, x_0, x_1, \cdots\rangle.$$

$\tau_y = \tau_{y_1}$ and $\tau_x^* = \tau_{x_0}^*$ are the inverse branches of σ and σ^* of order 1. We also define inverse branches of order n:

$$\tau_{y,n}(x) = |y_n, \cdots, y_1, x_0, x_1, \cdots\rangle, \quad \tau_{x,n}^*(y) = \langle \cdots, y_2, y_1, x_0, \cdots, x_{n-1}|$$

and the two Birkhoff sums of respectively $B : \Sigma \to \mathbb{R}$ and $B^* : \Sigma^* \to \mathbb{R}$

$$S_n B = \sum_{k=0}^{n-1} B \circ \sigma^k, \quad S_n^* B^* = \sum_{k=0}^{n-1} B^* \circ \sigma^{*k}.$$

We explain in a few lines why we have chosen this nonconventional definition of the natural extension. A natural extension $(\hat{\Sigma}, \hat{\sigma})$ of (Σ, σ), in the sense of Rohlin, it is not unique in the category of measure preserving dynamical systems. It may happen that there exist several topological natural extensions. In the case of the one-sided subshift of finite type it is usual to choose the corresponding two-sided subshift of finite type. It is however important in the present paper to let the past and future variables play the same role. In the considered formalism $\hat{\sigma}$ and $\hat{\sigma^{-1}}$ are natural extension of the two subshift of finite type (Σ, σ) and $(\hat{\Sigma}, \hat{\sigma})$ of transition matrix M and M^* (the transpose of M). The space Σ denotes de Bernoulli space associated to the shift of finite type defined by M.

We are now able to define the W kernel.

Definition 6. *Let $A : \Sigma \to \mathbb{R}$ be a continuous observable (considered as a function on $\hat{\Sigma}$). We call W kernel, $W(y,x)$, a continuous function $W : \hat{\Sigma} \to \mathbb{R}$ such that*

$$A^* := A \circ \hat{\sigma}^{-1} + W \circ \hat{\sigma}^{-1} - W$$

depends only on the variable y. A^* defines thus a continuous function on Σ^*. It is convenient to extend W on the whole product space $\Sigma^* \times \Sigma$ by $W(y,x) = -\infty$, if (y,x) does nor belong to $\hat{\Sigma}$.

Although the fact that any function on $\hat{\Sigma}$ is cohomologous to a function depending only on y (or x) is well known (see [6]), we prefer to give a specific name to the transfer function W because of its importance later in the construction of the deviation function $I(x)$. As we will see soon, $W(y,x)$ is unique up to a function depending only on y. The dual observable $A^*(y)$ thus defined is unique up to a coboundary.

When A depends only of the two first coordinates in Bernoully space, the values $(e^{A(i,j)})$ define a square matrix. From Perron-Frobenius Theorem for this positive operator we obtain a stochastic matrix and finally a stationary Markov Chain probability which defines the Gibbs state for A (see [25] for a proof). This fact was first observed by W. Parry. In this case the adjoint of the matrix $(e^{A(i,j)})$ is the matrix $(e^{A^*(i,j)})$.

We will give a proof of the following lemma

Proposition 7. *Let $A : \Sigma \to \mathbb{R}$ be an Hölder observable.*

(1) *A admits a Hölder W kernel.*

(2) *If W_1 and W_2 are two Hölder W kernels for A, their difference $W_1 - W_2$ depends only on the variable y.*

The W kernel plays a fundamental role in the definition of the deviation function. It has also some independent interest that we describe in the following proposition 8. The Ruelle-Perron $\mathcal{L}_A$ operator gives two important informations: the eigenmeasure ν_A and the eigenfunction ϕ_A. It is usually more difficult to find the eigenfunction than the eigenmeasure and the W kernel can be used instead. (Σ^*, σ^*) is a subshift of finite type and a similar Ruelle-Perron operator can be defined. If A and W are Hölder, the dual observable A^* is also Hölder and we denote by ν_A^* and ϕ_A^* the eigenmeasure and eigenfunction for the largest eigenvalue $\lambda(A^*)$ again normalized by $\int \phi_A^* \, d\nu_A^* = 1$ and $\nu_A^*(\Sigma^*) = 1$. Notice that $\lambda(A) = \lambda(A^*)$ because of the coboundary equation between A and A^*. In particular

$$m(A) = \lim_{\beta \to +\infty} \frac{1}{\beta} \log \lambda(\beta A) = m(A^*).$$

We show that the knowledge of W, ν_A and ν_A^* is enough to find ϕ_A and ϕ_A^*.

Proposition 8. *Let $A : \Sigma \to \mathbb{R}$ be an Hölder observable, W the associated kernel and A^* the corresponding dual observable. Let*

$$c := \log \iint M(y,x) \exp\big(W(y,x)\big) \, d\nu_A^*(y) d\nu_A(x)$$

where $M(y,x) = M(y_1, x_0)$. Then

(1) *The natural extension $\hat{\mu}_A$ of the equilibrium measure μ_A is given by*

$$\hat{\mu}_A(dy, dx) = M(y,x) \exp\big(W(y,x) - c\big) \, \nu_A^*(dy) \times \nu_A(dx).$$

(2) *The normalized eigenfunctions ϕ_A and ϕ_A^* are given by*

$$\phi_A(x) = \int M(y,x) \exp\big(W(y,x) - c\big) \, d\nu_A^*(y),$$

$$\phi_A^*(y) = \int M(y,x) \exp\big(W(y,x) - c\big) \, d\nu_A(x).$$

A similar proposition for Markov expanding transformations F on the interval can be proved. To illustrate this generalization, we show for instance how to construct an explicit W kernel for the Gauss map which enable us to recover the standard invariant measure absolutely continuous with respect to Lebesgue. Related considerations can be found in the very interesting article [3] where results are described without mathematical rigor.

We point out that in a forthcoming paper we will use the W kernel, in the context of Bowen-Series transformations [21], to describe a relation of the Helgason distribution (an eigendistribution for a complex Ruelle operator) of each eigenfunction of the Laplacian in a compact surface of negative constant curvature with an eigenfunction associated to the eigenvalue 1 of the related complex Ruelle operator acting on the boundary of Poincare disk. In this way we will be able to give a mathematical proof, for the compact case (via results of Helgason and Otal), of the main result stated in [3] .

Let us recall first the definition of the Gauss map $T : [0,1[\to [0,1[$, ($\{u\}$ denotes the fractional part of u),

$$a(x) := \left\{\frac{1}{x}\right\} \quad \text{if } x \neq 0, \quad F(0) := 0,$$
$$T(x) := \frac{1}{x} - \left\{\frac{1}{x}\right\} \quad \text{if } x \neq 0, \quad a(0) := +\infty.$$

This dynamical system $([0,1[\backslash\mathbb{Q}, T)$ can be identified to the full shift (Σ,σ), $\Sigma = (\mathbb{N}^*)^{\mathbb{N}}$, on a countable number of symbols using the theory of decomposition into continuous fraction

$$(x_0, x_1, x_2, \cdots) \in \Sigma \longleftrightarrow x = \frac{1}{x_0 + \frac{1}{x_1 + \frac{1}{x_2 + \cdots}}} \in [0,1[\backslash\mathbb{Q}.$$

The dual shift (Σ^*,σ^*) is equal to the original full shift (Σ,σ). If we use the same identification between Σ^* and $[0,1[\backslash\mathbb{Q}$, the natural extension of T is defined on $[0,1[\backslash\mathbb{Q} \times [0,1[\backslash\mathbb{Q}$ and given by

$$\hat{T}(y,x) = (T^{-1}_{a(x)}(y), T(x)) \quad \hat{T}^{-1}(y,x) = (T(y), T^{-1}_{a(y)}(x))$$

where $T_n^{-1}(y) := 1/(n+y)$. We will prove

Proposition 9. *Let* $([0,1[\backslash\mathbb{Q}, T)$ *be the Gauss map and* $A(x) := -\log|T'(x)|$. *Then* $\exp A(x) = x^2$ *and the function* W *defined on* $[0.1[\backslash\mathbb{Q} \times [0,1[\backslash\mathbb{Q}$ *by*

$$W(y,x) := -2\log(1+xy)$$

is a W kernel for A. The corresponding dual observable A^* *satisfies* $\exp A^*(y) = y^2$.

In this case the dual $A^* = A$ and this is very unusual.

From now on we choose a particular Hölder W kernel W. We notice that βW is a W kernel for βA and the corresponding dual observable is equal to βA^*

$$\beta A^* = \beta A \circ \hat{\sigma}^{-1} + \beta W \circ \hat{\sigma}^{-1} - \beta W.$$

For any β, we denote by λ^*_β, ϕ^*_β and ν^*_β, the largest eigenvalue, the corresponding eigenfunction and eigenmeasure of the Ruelle-Perron operator on the dual subshift (Σ^*,σ^*) associated to the observable βA^*. We again normalize by

$$\int \phi^*_\beta \, d\nu^*_\beta = 1, \quad \nu^*_\beta(\Sigma^*) = 1.$$

Note that from the cohomology equation above (by means of W), if $\mu_{\max}$ is a maximizing measure for A, if $\hat{\mu}_{\max}$ denotes its natural extension and $\mu^*_{\max}$ its projection onto Σ^*, then $\mu^*_{\max}$ is also a maximizing measure for A^*. According to our assumption 1, $\mu_{\max}$ and $\mu^*_{\max}$ are unique.

As previously, (A^* is Hölder), the sequence $(\frac{1}{\beta}\log\phi^*_\beta)$ possesses accumulation points V^* for the uniform topology. All these accumulation points are strict subactions and we choose as before a particular subsequence (β_n) such that

$$V^* := \lim_{n\to+\infty}\frac{1}{\beta_n}\log\phi^*_{\beta_n} \quad \text{exists uniformly on } \Sigma^*.$$

The main step in the proof of theorem 4 is given by the following intermediate proposition

Proposition 10. *Let $A:\Sigma\to\mathbb{R}$ be a Hölder observable admitting a unique maximizing measure μ_{max}. Let $W:\hat{\Sigma}\to\mathbb{R}$ be a Hölder W kernel and A^* the corresponding dual Hölder observable. Let $\hat{\mu}_{max}$ be the natural extension of μ_{max}.*

(1) Suppose that for some subsequence (β_n), the following limits exist

$$V := \lim_{n\to+\infty}\frac{1}{\beta_n}\log\phi_{\beta_n}, \quad \textit{and} \quad V^* := \lim_{n\to+\infty}\frac{1}{\beta_n}\log\phi^*_{\beta_n}.$$

Then for any $(p^*,p)\in\operatorname{supp}\hat{\mu}_{max}$

$$\begin{aligned}\gamma(W) &:= W(p^*,p) - V(p) - V^*(p^*)\\ &= \lim_{n\to\infty}\frac{1}{\beta_n}\log\iint M(y,x)\exp\big(\beta_n W(y,x)\big)\,d\nu^*_{\beta_n}(y)d\nu_{\beta_n}(x).\end{aligned}$$

(2) For any $(y,x)\in\hat{\Sigma}$, ($m=m(A)=m(A^*)$)

$$I(x) = W(y,x) - \gamma(W) - V(x) - \lim_{n\to+\infty}\Big(S^*_n(A^*-m)+V^*\Big)\circ\hat{\sigma}^n(y,x).$$

(3) (The Large Deviation Principle). For any cylinder $C\subset\Sigma$

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

and the definition of $I(x)$ is independent of the chosen V (which in fact can be any strict subaction)

In the proof of this theorem we will show that $\lim_{\beta\to+\infty}\frac{1}{\beta}\log\phi_\beta$ exists. Notice that this limit is when $\beta\to\infty$ and not for $\beta_n\to\infty$.

Notice that the above proposition implies Theorem 4.

When $A(x)$ depends only on the first two elements of x, that is, $A(x) = A(x_0,x_1)$, where $x=(x_0,x_1,x_2,\cdots)\in\Sigma$, the proof of our main result can be simplified. We just have to consider an associated Markov Chain and the deviation function $I(x)$ has a simpler formulation. In this case A^* also depends on the first two elements of y, that is $A^*(y)=A^*(y_1,y_2)$, where $y=(y_1,y_2,y_3,\cdots)\in\Sigma^*$. In [7], [17] interesting results about the case $A(x)=A(x_0,x_1)$ are obtained.

Remark 11.

(1) *It is known (see [7], [17] for a proof) that, in the Markov case (A depends on finite number of coordinates), the sequence of equilibrium measures (μ_β) converges to some maximizing measure.*

3. Proof of the Main Results

We analyze now the case of a general Hölder $A : \Sigma \to \mathbb{R}$ depending on an infinite number of coordinates and having a unique maximizing measure.

Given $x = |x_0, x_1, x_2, ...\rangle \in \Sigma$ and $y = \langle \cdots y_2, y_1| \in \Sigma^*$, we will use the notation $\langle y|x\rangle = \langle \cdots, y_3, y_2, y_1 \mid x_0, x_1, x_2, \cdots \rangle \in \hat{\Sigma}$ (for any admissible transition $y_1 \to x_0$). The symbol $|$ is used to say where the coordinate at time 0 is located.

We recall that the left shift σ acting on Σ has inverse branches of all order n given by

$$\tau_{y,n}|x_0, x_1, \cdots\rangle = |y_n, \cdots y_1, x_0, x_1, \cdots\rangle.$$

In the same way $\sigma^* : \Sigma^* \to \Sigma^*$ has inverse branches given by

$$\tau^*_{x,n}\langle \cdots y_2, y_1| = \langle \cdots y_2, y_1, x_0, \cdots, x_{n-1}|.$$

The (bijective) natural extension is given by $\hat{\sigma}(y, x) = (\tau^*_x(y), \sigma(x))$ and its inverse by $\hat{\sigma}^{-1}(y, x) = (\sigma^*(y), \tau_y(x))$. The same formula are true for $\hat{\sigma}^n$ and $\hat{\sigma}^{-n}$ using $\tau_{y,n}$ and $\tau^*_{x,n}$. We also introduce a convenient notation. We call $\mathcal{W}^*(x)$ and $\mathcal{W}(y)$, the local stable and unstable "manifolds", that is for any $(y, x) \in \hat{\Sigma}$

$$\mathcal{W}(y) = \{u \in \Sigma \mid y_1 \to u_0\} \quad \mathcal{W}^*(x) = \{v \in \Sigma^* \mid v_1 \to x_0\}.$$

Our first main objective is to proof the existence of a W kernel.

Proof of Proposition 7-(1). (Sinai's method). We first define a family of cocycles $\Delta(x, x', y)$ given by

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

Note that x and x' are both in $\mathcal{W}(y)$.

The function $\Delta(x, x', y)$ is well defined (and Hölder) because A is Hölder and $\tau_{y,n}$ is contracting. We note that

$$\Delta(x, x', y) = A \circ \tau_y(x) - A \circ \tau_y(x') + \Delta(\tau_y(x), \tau_y(x'), \sigma^*(y)).$$

We are going to prove that

$$W(y, x) := \Delta(x, x', y) - \log \int_{\mathcal{W}(y)} \exp \Delta(u, x', y)\, d\nu_A(u) \quad (\forall\, x' \in \mathcal{W}(y))$$

is a W kernel. Notice that $W(y, x)$ is well defined, that is, does not depend on $x' \in \mathcal{W}(y)$. We recall that ν_A is the normalized eigenmeasure of the Ruelle operator $\mathcal{L}^*_A$ corresponding to the largest eigenvalue $\lambda(A) = \exp P(A)$. Let

$$A^*(y, x) := W \circ \hat{\sigma}^{-1}(y, x) - W(y, x) + A \circ \hat{\sigma}^{-1}(y, x).$$

We just need to prove that A^* depends only on the variable y. We study first the integrand in the definition of W:

$$\mathcal{I} := \log \int_{\mathcal{W}(y)} \exp \Delta(u, x', y) \, d\nu_A(u)$$

Since $\tau_y : \mathcal{W}(y) \to |y_1\rangle$ is injective and has Jacobian $(A - P(A)) \circ \tau_y$ with respect to ν_A

$$\begin{aligned}\mathcal{I} &= \log \int_{\mathcal{W}(y)} \exp(A \circ \tau_y(u) - A \circ \tau_y(x')) \exp \Delta(\tau_y(u), \tau_y(x'), \sigma^*(y)) \, d\nu_A(u) \\ &= (P(A) - A \circ \tau_y(x')) + \log \int_{|y_1\rangle} \exp \Delta(v, \tau_y(x'), \sigma^*(y)) \, d\nu_A(v).\end{aligned}$$

Since

$$\Delta(\tau_y(x), \tau_y(x'), \sigma^*(y)) - \Delta(x, x', y) = A \circ \tau_y(x') - A \circ \tau_y(x).$$

We finally obtain

$$W \circ \hat{\sigma}^{-1} - W + A \circ \hat{\sigma}^{-1} = P(A) + \log \frac{\int_{|y_1\rangle} \exp \Delta(v, \tau_y(x'), \sigma^*(y)) \, d\nu_A(v)}{\int_{\mathcal{W} \circ \sigma^*(y)} \exp \Delta(v, \tau_y(x'), \sigma^*(y)) \, d\nu_A(v)}.$$

The right hand side of the equality is clearly independent of x. □

We notice that Sinai's method gives a dual observable A^* which is normalized in the sense that the function $\phi^*(y) := 1$ is an eigenfunction of the dual Ruelle operator for its largest eigenvalue $\lambda(A) = \exp P(A)$: for all $y \in \Sigma^*$

$$\sum_{y_1 \leftarrow i} \exp(A^* - P(A))\langle \cdots y_2, y_1, i| = 1 \quad \forall\, i = 1 \cdots r.$$

Remember that Σ^* is a subshift of transition matrix M^* and the summation in the above formula is over all transitions i following the symbol y_1. In particular this normalization implies that A^* does not change if a coboundary is added to A. Indeed, if $B = A + c - c \circ \sigma$

$$\begin{aligned}\Delta_B(x, x', y) &= \Delta_A(x, x', y) - [c(x) - c(x')] \\ \nu_B(dx) &= \exp c(x) \nu_A(dx) \\ W_B(y, x) &= W_A(y, x) - c(x) \\ B^*(y) &= W_B \circ \hat{\sigma}^{-1} - W_B + B \circ \hat{\sigma}^{-1} = A^*(y).\end{aligned}$$

We could have chosen another proof using Bowen's ideas. In this case we would lost the normalization of A^* and gained a linear dependence from A to A^*.
Remark 12. *For any W kernel W, for any $(y, x) \in \hat{\Sigma}$, $x' \in \mathcal{W}(y)$, $y' \in \mathcal{W}^*(x)$*

- $W(y, x) - W(y, x') = \Delta(x, x', y)$,
- $W(y, x) - W(y', x) = \Delta^*(y, y', x) := \sum_{k \geq 1} A^* \circ \hat{\sigma}^k(y, x) - A^* \circ \hat{\sigma}^k(y', x)$.

This last equality explains why $I(x)$ in Theorem 4 depends only on x and not on y.

We are now going to prove the second part of Proposition 7.

Proof of Proposition 7-(2). Let W_1 W_2 be two W kernels for the same observable A. Let

$$\begin{aligned} A_1^* &= W_1 \circ \hat\sigma^{-1} - W_1 + A \circ \hat\sigma^{-1}, \\ A_2^* &= W_2 \circ \hat\sigma^{-1} - W_2 + A \circ \hat\sigma^{-1}, \\ A^* &= A_2^* - A_1^* = (W_2 - W_1) \circ \hat\sigma^{-1} - (W_2 - W_1). \end{aligned}$$

The Birkhoff sum of A^*, as a function on $\hat\Sigma$, is equal to zero on any periodic orbit. The same remark is valid when A^* is considered as a function on Σ^*. Since A^* is Hölder, thanks to Livsic theorem [18], A^* is equal to a coboundary:

$$A^*(y) = c(y) - c \circ \sigma^*(y),$$

for some Hölder function $c(y)$. Then

$$(W_2 - W_1 + c) \circ \hat\sigma = (W_2 - W_1 + c)$$

every where on $\hat\Sigma$. Thanks to the transitivity of $\hat\sigma$, $W_2 - W_1 + c$ has to be constant, $W_2 - W_1$ depends only on y. $\square$

Before showing some properties a W kernel possesses, we establish a fundamental lemma which explains in part the disymetry in the definition in W.

Lemma 13. *Let $W : \hat\Sigma \to \mathbb{R}$ be a W kernel for an observable $A(x)$. Let $\mathcal{L}_A$ and $\mathcal{L}_{A^*}$ be the Ruelle operator defined on Σ and Σ^*.*

(1) For any symbol i, any $x \in \Sigma$ and $y \in \Sigma^*$, if $y_1 \to i \to x_0$ are admissible

$$(A^* + W)(\langle y, i | x\rangle) = (A + W)(\langle y | i, x\rangle).$$

(2) In particular, for any $x \in \Sigma$, $y \in \Sigma^*$ and any function $f : \Sigma^* \times \Sigma \to \mathbb{R}$

$$\begin{aligned} &\mathcal{L}_{A^*}\Big(f(\cdot, x) M(\cdot, x) \exp W(\cdot, x)\Big)(y) \\ &\qquad = \mathcal{L}_A\Big(f \circ \hat\sigma(y, \cdot) M(y, \cdot) \exp W(y, \cdot)\Big)(x). \end{aligned}$$

($M(y, x) := M(y_1, x_0)$ where $M(i, j) = 1$ iff $i \to j$ is admissible.)

Proof of Lemma 13. Part(1). Let $y' = \langle y, i|$ and $x' = |i, x\rangle$, then $\hat\sigma^{-1}(y', x) = (y, x')$. By definition of A^*

$$A^*(y', x) = W(y, x') - W(y', x) + A(y, x').$$

Part (2).

$$\begin{aligned} &\mathcal{L}_{A^*}\big(f(\cdot, x) M(\cdot, x) \exp W(\cdot, x)\big)(y) \\ &= \sum_i M(y_1, i) f(\langle y, i | x) M(\langle y, i|, x) \exp\big(A^*(\langle y, i|) + W(\langle y, i | x)\big) \\ &= \sum_i M(y_1, i) f \circ \hat\sigma(y | i, x\rangle) M(i, x_0) \exp\big(A(|i, x\rangle) + W(y | i, x\rangle)\big) \\ &= \sum_i M(i, x_0) f \circ \hat\sigma(y | i, x\rangle) M(y, |i, x\rangle) \exp\big(A(|i, x\rangle) + W(y | i, x\rangle)\big) \\ &= \mathcal{L}_A\big(f \circ \hat\sigma(y, \cdot) M(y, \cdot) \exp W(y, \cdot)\big)(x) \end{aligned}$$

$\square$

We can then prove the following

Proof of Proposition 8. Part (1). For simplicity we note

$$K(y,x) = M(y,x)\exp(W(y,x)-c).$$

For any bounded Borel $f : \Sigma^* \times \Sigma \to \mathbb{R}$ we have

$$\begin{aligned}\iint & f\circ\hat{\sigma}(y,x)K(y,x)\,d\nu_A^*(y)d\nu_A(x) \\ &= \int d\nu_A^*(y)\int \mathcal{L}_{(A-P)}\big(f\circ\hat{\sigma}(y,\cdot)K(y,\cdot)\big)(x)\,d\nu_A(x) \\ &= \int d\nu_A(x)\int \mathcal{L}_{(A^*-P^*)}\big(f(\cdot,x)K(\cdot,x)\big)\,d\nu_A^*(y) \\ &= \iint f(y,x)\,d\nu_A^*(y)d\nu_A(x)\end{aligned}$$

where P (resp. P^*) is the pressure of A (resp. A^*). We already noticed $P=P^*$. The measure $\hat{\mu}_A(dy,dx) = M(y,x)\nu_A^*(dy)\nu_A(dx)$ is invariant and projects onto μ_A and μ_A^*. $\hat{\mu}_A$ is therefore the natural extension of μ_A.

Part (2). Let $\phi_A(x) = \int K(y,x)\,d\nu_A^*(y)$. Then

$$\begin{aligned}\phi_A(x) &= \int \mathcal{L}_{(A^*-P^*)}\big(K(\cdot,x)\big)(y)\,d\nu_A^*(y) \\ &= \int \mathcal{L}_{(A-P)}\big(K(y,\cdot)\big)(x)\,d\nu_A^*(y) \\ &= \mathcal{L}_{(A-P)}(\phi_A)(x).\end{aligned}$$

The proof for ϕ_A^* is similar. $\square$

The proof of Proposition 9 is actually very simple provided we guess the correct W kernel. We explain this fact in the first part. In the second part, we use Sinai's method to construct a W kernel and obtain

$$\tilde{W}(y,x) = \log\frac{1+y}{(1+xy)^2}$$

which is our guess times a function of y as predicted by the general theory.

Proof of Proposition 9. First part: proof using the a priori definition. We first notice that $\exp A(x) = x^2$. Define $\exp A^*(y) := y^2$. We want to show, as in Lemma 13, that for any symbol $n \geq 1$

$$\exp\big(A^*\circ T_n^{-1}(y) + W(T_n^{-1}(y),x)\big) = \exp\big(A\circ T_n^{-1}(x) + W(y,T_n^{-1}(x))\big).$$

Indeed

$$\frac{1}{(n+y)^2}\frac{1}{\big(1+\frac{x}{n+y}\big)^2} = \frac{1}{(n+x)^2}\frac{1}{\big(1+\frac{y}{n+x}\big)^2} = \frac{1}{(n+x+y)^2}.$$

Second part: how guessing the W kernel. We first recall an identity we will use later. Let $(a_0, a_1, \cdots)$ be a sequence of positive integers with reduced quotients (p_n/q_n). Then for any $x \geq 0$, $n \geq 0$

$$\frac{p_n + p_{n-1}x}{q_n + q_{n-1}x} = \cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{a_n + x}}}}$$

and

$$q_n + q_{n-1}x =$$
$$(a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_n + x}}})\ (a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_n + x}})...(a_n + x)$$

We want now to compute

$$\exp \Delta(x, x', y) = \prod_{k \geq 1} \frac{\exp A \circ T_y^{-k}(x)}{\exp A \circ T_y^{-k}(x')}$$

where $(y_1, y_2, \cdots)$ are the continuous fraction expansion of y and

$$y = \cfrac{1}{y_1 + \cfrac{1}{y_2 + \cfrac{1}{y_3 + \cdots}}} \qquad T_y^{-k}(x) = \cfrac{1}{y_k + \cfrac{1}{y_{k-1} + \cdots + \cfrac{1}{y_1 + x}}}.$$

Let

$$\frac{p_1^*}{q_1^*} = \frac{1}{y_k} \qquad \frac{p_2^*}{q_2^*} = \cfrac{1}{y_k + \cfrac{1}{y_{k-1}}} \quad \cdots \quad \frac{p_k^*}{q_k^*} = \cfrac{1}{y_k + \cfrac{1}{y_{k-1} + \cdots + \cfrac{1}{y_1}}}.$$

Then

$$\prod_{k=1}^{n} \exp A \circ T_y^{-k}(x) = \frac{1}{(q_n^* + q_{n-1}^* x)^2}$$
$$\exp \Delta(x, x', y) = \lim_{n \to +\infty} \Big(\frac{q_n^* + q_{n-1}^* x'}{q_n^* + q_{n-1}^* x}\Big)^2 = \Big(\frac{1 + yx'}{1 + yx}\Big)^2,$$

thanks to

$$\frac{q_{n-1}^*}{q_n^*} = \frac{q_{n-1}^*}{y_1 q_{n-1}^* + q_{n-2}^*} = \cfrac{1}{y_1 + \cfrac{q_{n-2}^*}{q_{n-1}^*}} = \cdots = \cfrac{1}{y_1 + \cfrac{1}{y_2 + \cdots + \cfrac{1}{y_k}}}$$

$$\lim_{n \to +\infty} \frac{q_{n-1}^*}{q_n^*} = y.$$

The W kernel is finally given by

$$\int_0^1 \Delta(u, x', y)\, du = \frac{(1 + x'y)^2}{1 + y} \qquad \exp \tilde{W}(y, x) = \frac{1 + y}{(1 + xy)^2}.$$

Notice that in this example,

$$\nu_A(dx) = dx \quad A^*(y) = y^2 \frac{1 + T(y)}{1 + y} \quad \nu_A^*(dy) = \frac{1}{1 + y} dy$$
$$\iint \exp \tilde{W}(y, x)\, d\nu_A^*(y) d\nu_A(x) = \log 2$$

and the density is given by $\phi_A(x) = \frac{1}{\log 2} \frac{1}{1+y}$. □

The definition of the deviation function $I(x)$ in Theorem 4 uses an accumulation point $V(x)$ of $(\frac{1}{\beta} \log \phi_\beta)$. $V(x)$ is actually a strict subaction. In the case of a unique maximizing measure, Proposition 5 tells us that all the strict subactions V are equal up to a constant.

Definition 14. *Let $A : \Sigma \to \mathbb{R}$ be a continuous observable. The A-nonwandering set is the set $\Omega(A, \sigma) = \{x \in \Sigma \mid S(x, x) = 0\}$.*

Proof of Proposition 5 On the one hand, for any subaction (strict or not)

$$S(p, x) \leq V(x) - V(p).$$

On the other hand, since V is a strict subaction, there exists a sequence of points $y = (y_1, y_2, \cdots)$ such that

$$\sigma(y_1) = x, \quad \sigma(y_{n+1}) = y_n, \quad \forall\, n \geq 1$$
$$(A - m)(y_{n+1}) = V(y_n) - V(y_{n+1}) \quad \forall\, n \geq 1.$$

Let $\alpha(y)$ be the set of all accumulation points of the sequence (y_n). This set is compact and σ-invariant; it possesses therefore an invariant measure. Since $\alpha(y) \subset \Omega(A, \sigma)$, see Definition 14, this measure is necessarily maximizing. By assumption, there exists a unique such measure; we thus obtain that $\text{supp}(\mu_{\max}) \subset \alpha(y)$ and that p is an accumulation point of (y_n). The definition of $S(p, x)$ implies

$$S(p, x) \geq V(x) - V(p).$$

□

The rest of this section is now devoted to the proof of Proposition 10 end therefore to Theorem 4.

Proof of Proposition 10 : part (1). Let $\hat{p} = (p^*, p)$ be a point in the support of $\mu_{\max}$. Let $B^* \times B$ be a small cylinder containing $\hat{p}$.

Recall that $K_\beta(y, x)$ is defined on the whole product space $\Sigma^* \times \Sigma$ and is equal to 0 outside $\hat{\Sigma}$.

Let

$$K_{\beta_n} := M(y, x) \exp\big(\beta_n W(y, x) - c_{\beta_n}\big), \quad c_{\beta_n} := \log \iint K_{\beta_n}(y, x)\, d\nu_{\beta_n}^*(y) d\nu_{\beta_n}(x).$$

Then

$$\hat{\mu}_{\beta_n}(B^* \times B)\exp(c_{\beta_n}) = \iint_{B^* \times B} \frac{K_{\beta_n}(y,x)}{\phi^*_{\beta_n}(y)\phi_{\beta_n}(x)}\, d\mu^*_{\beta_n}(y) d\mu_{\beta_n}(x)$$
$$\leq \mu^*_{\beta_n}(B^*)\mu_{\beta_n}(B) \sup_{B^* \times B} \frac{K_{\beta_n}(y,x)}{\phi^*_{\beta_n}(y)\phi_{\beta_n}(x)}.$$

Since $\mu_{\beta_n} \to \mu_{\max}$,

$$\hat{\mu}_{\max}(B^* \times B) \neq 0, \quad \mu^*_{\max}(B^*) \neq 0, \quad \mu_{\max}(B) \neq 0.$$

Letting n go to ∞ and using $\frac{1}{\beta_n}\log\phi_{\beta_n} \to V$, $\frac{1}{\beta_n}\log\phi^*_{\beta_n} \to V^*$, we obtain

$$\limsup_{n\to\infty} \frac{1}{\beta_n} c_{\beta_n} \leq \sup_{(B^* \times B)\cap\hat{\Sigma}} \{W(y,x) - V^*(y) - V(x)\}.$$

Since B^* and B can be chosen as small as we need, we finally get

$$\limsup_{n\to\infty} \frac{1}{\beta_n} c_{\beta_n} \leq W(\hat{p}) - V^*(p^*) - V(p).$$

The lower bound is similar. □

Proof of Proposition 10 : part 2. Let $(\gamma = \gamma(W))$

$$\hat{I}(y,x) = W(y,x) - V^*(y) - V(x) - \gamma$$
$$- \lim_{n\to\infty} \big(V^* - V^* \circ \hat{\sigma}^{-n} + S^*_n(A^* - m)\big) \circ \hat{\sigma}^n.$$

Since by definition $S^*_n(A^* - m) = \sum_{k=0}^{n-1}(A^* - m)\circ\hat{\sigma}^{-k}$, we obtain

$$\hat{I}(y,x) = W(y,x) - V^*(y) - V(x) - \gamma + \hat{R}(y,x),$$

where $\hat{R}(y,x) = \sum_{k\geq 1}\big(V^*\circ\hat{\sigma}^{-1} - V^* - (A^* - m)\big)\circ\hat{\sigma}^k$. We use now the cocycle relation between A^* and A

$$(A^* - m)\circ\hat{\sigma}^k = W\circ\hat{\sigma}^{k-1} - W\circ\hat{\sigma}^k + (A - m)\circ\hat{\sigma}^{k-1}$$
$$\sum_{k=1}^{n}(A^* - m)\circ\hat{\sigma}^k = W - W\circ\hat{\sigma}^n + \sum_{k=0}^{n-1}(A - m)\circ\hat{\sigma}^k$$
$$\hat{I} = \lim_{n\to\infty}(W - V^* - V - \gamma)\circ\hat{\sigma}^n + \sum_{k=0}^{n-1}\big(V\circ\sigma - V - (A - m)\big)\circ\sigma^k.$$

Let $I(x) := \sum_{k\geq 0}(V\circ\sigma - V - (A - m))\circ\sigma^k$.

Either $I(x) = +\infty$ then $\hat{I}(y,x) = +\infty$ too and $I(x) = \hat{I}(y,x)$. Or $I(x) < +\infty$, $\hat{R}(y,x) < +\infty$, the set of accumulation points, $\hat{\omega}(y,x)$, of $(\hat{\sigma}^n(y,x))$ has to be included in $\hat{\Omega}(A,\hat{\sigma})$ and therefore contains the support of the unique maximizing measure $\mu_{\max}$. There exists a subsequence $(\hat{\sigma}^{n_k}(y,x))$ converging to $\hat{p} = (p^*, p)$,

$$\lim_{k\to\infty}(W - V^* - V - \gamma)\circ\hat{\sigma}^{n_k}(y,x) = (W - V^* - V - \gamma)(\hat{p}) = 0$$

and $I(x) = \hat{I}(y,x)$ in this case too. $\Box$

Proof of Proposition 10 : part 3. We choose a subsequence (β_k) such that

$$\frac{1}{\beta_k} \log \phi^*_{\beta_k}(y) \to V^*(y) \quad \text{and} \quad \frac{1}{\beta_k} \log \phi_{\beta_k}(x) \to V(x).$$

Note that under our assumption 1 the probability μ_β converges to the maximizing measure as $\beta \to \infty$.

To simplify the notations, we keep β instead of β_k. We also use the notation

$$K_\beta(y,x) = M(y,x) \exp\big(\beta W(y,x) - c_\beta\big)$$
$$c_\beta = \log \iint M(y,x) \exp \beta W(y,x) \; d\nu^*_\beta(y) d\nu_\beta(x).$$

We recall that $\frac{1}{\beta} c_\beta \to \gamma$. We choose once for all a cylinder $C = |i_0, i_1, \cdots, i_{r-1}\rangle$ of length r. We first show that

$$\bar{l}(C) := \limsup_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta(C) \leq -\inf_C I.$$

We define by induction a decreasing sequence of sets (C_n) in the following way : $C_r = C$, if $C_n = |i_0, \cdots, i_{n-1}\rangle$ has been defined, since C_n is equal to the disjoint sum of $C_{n+1} = |i_0, \cdots, i_n\rangle$ over i_n, there exists at least one i_n such that

$$\bar{l}(C) = \limsup_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta(C_{n+1}).$$

Define $x_C = |i_0, i_1, i_2, \cdots\rangle$. Choose some fixed $y_C = \langle \cdots i_{-2}, i_{-1}| \in \Sigma^*$ so that $(y_C, x_C) \in \hat{\Sigma}$ and call $B_k = \langle i_{-k}, \cdots, i_{-1}|$. On the one hand

$$\begin{aligned} \hat{\mu}_\beta(B_k \times C_n) &= \iint_{B_k \times C_n} \Big(\frac{K_\beta(y,x)}{\phi^*_\beta(y) \phi_\beta(x)} \Big) \, d\mu^*_\beta(y) d\mu_\beta(x) \\ &\geq \inf_{B_k \times C_n} \Big(\frac{K_\beta(y,x)}{\phi^*_\beta(y) \phi_\beta(x)} \Big) \mu^*_\beta(B_k) \mu_\beta(C_n). \end{aligned}$$

On the other hand

$$\begin{aligned} \hat{\mu}_\beta(B_k \times C_n) &= \mu^*_\beta(\langle i_{-k}, \cdots, i_{-1}, i_0, \cdots, i_{n-1}|) \\ &= \int_{B_k} \exp S^*_n(\beta A^* - P_\beta) \circ \tau^*_{x_C,n}(y) \frac{\phi^*_\beta \circ \tau_{x_C,n}(y)}{\phi^*_\beta(y)} \, d\mu^*_\beta(y) \\ &\leq \sup_{y \in B_k} \Big(\exp\big(S^*_n(\beta A^* - P_\beta) \circ \tau^*_{x_C,n}(y)\big) \frac{\phi^*_\beta \circ \tau_{x_C,n}(y)}{\phi^*_\beta(y)} \Big) \mu^*_\beta(B_k). \end{aligned}$$

We first eliminate $\mu^*_\beta(B_k)$ on both sides, then apply $\frac{1}{\beta}\log$ on both sides, take limit when β (or more precisely β_n) goes to ∞, to get

$$\bar{l}(C) + \inf_{B_k \times C_n} (W - \gamma - V^* - V) \leq \sup_{B_k} \Big((S^*_n(A^* - m) + V^*) \circ \hat{\sigma}^n - V^* \Big).$$

Letting first k go to infinity and then n go to infinity, we obtain finally

$$\bar{l}(C) \leq -\hat{I}(y_C, x_C) = -I(x_C) \leq -\inf_{x \in C} I(x).$$

We next show that

$$\underline{l}(C) := \liminf_{\beta \to \infty} \frac{1}{\beta} \log \mu_\beta(C) \geq -\inf_{x \in C} I(x).$$

As before, take any $x = |i_0, i_1, \cdots\rangle \in C$, any $y \in \Sigma^*$ such that $(y, x) \in \hat{\Sigma}$ and define in the same way B_k and C_n. Reversing the two previous inequalities, we obtain

$$\mu_\beta(C_n) \sup_{B_k \times C_n} \left(\frac{K_\beta}{\phi_\beta^* \phi_\beta} \right) \geq \inf_{B_k} \left(\exp\left(S_n^*(\beta A^* - P_\beta) \circ \tau_{x,n}^* \right) \frac{\phi_\beta^* \circ \tau_{x,n}}{\phi_\beta^*} \right)$$
$$\underline{l}(C) + \sup_{B_k \times C_n} (W - \gamma - V^* - V) \geq \inf_{B_k} \Big((S_n^*(A^* - m) + V^*) \circ \hat{\sigma}^n - V^* \Big).$$

and finally

$$\begin{aligned} \underline{l}(C) &\geq -\hat{I}(x, y) = -I(x) \qquad \text{for any } (y, x) \in \hat{\Sigma},\ x \in C \\ &\geq -\inf_{x \in C} I(x). \end{aligned}$$

□

Acknowledgments

A.O. Lopes partially supported by CNPq-Pronex Sistemas Dinamicos, and Instituto do Milenio.

A. Baraviera, partially supported by CNPq - Sistemas Dinamicos and Instituto do Milenio.

References

1. N. Anantharaman, *Counting geodesics which are optimal in homology, Erg Theo and Dyn Syst* **23 Issue 2** (2003) 353-388.
2. N. Anantharaman, *On the zero-temperature or vanishing viscosity limit for Markov processes arising from Lagrangian Dynamics, Journal of European Mathematical Society* **6 (2)** (2004) 207-276.
3. E. Bogomolny and M. Carioli, *Quantum maps of geodesic flows on surfaces of constant negative curvature, IV International Conference on "PATH INTEGRALS FROM meV TO MeV"*, Tutzing May, (1992) 18-21.
4. T. Boush, *Le poisson n'a pas d'arêtes*, *Ann Inst Henry Poincaré, Prob et Stat* **36** (2000) 459-508.
5. T. Bousch and O. Jenkinson, *Cohomology classes of dynamically non-negative C^k functions Invent. Math.* **148 , no. 1** (2002) 207–217
6. R. Bowen, *Equilibrium states and the ergodic theory of Anosov difeomorphisms*, *Lecture Notes in Mathematics* **470** Sringer Verlag (1975).
7. J. Brémont, *On the behaviour of Gibbs measures at temperature zero*, *Nonlinearity* **16** (2003) 419-426

8. Z. N. Coelho, *Entropy and Ergodicity of skew-products over shifts of finite type and central limit asymptotics,* Ph.D. Thesis, Warwick Universit (1990)
9. J.-P. Conze and Y. Guivarc'h, *Croissance des sommes ergodiques,* unpublished preprint (1993)
10. G. Contreras, A. O. Lopes and P. Thieullen, *Lyapunov minimizing measures for expanding maps of the circle, Erg. Theo. and Dyn. Syst.* **21** (2001) 1379-1409
11. A. Dembo and O. Zeitouni, **Large Deviations Techniques**, (Jones and Bartlett Publishers 1993)
12. R. Exel and A. O. Lopes, *C^*-Algebras, approximately proper equivalence relations, and Thermodynamic Formalism, Erg. Theo. and Dyn. Systems.* **24** (2004) 1051-1082
13. A. C. C. van Enter, R. Fernandez and A. D. Sokal, *Regularity properites and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory, Journ. of Stat. Phys.* **72, N 5/6** (1993) 879-1187
14. H.-O. Georgii, **Gibbs measures and phase transitions**, (de Gruyter edit. 1988)
15. Nicolai T. A. Haydn, *Gibbs functionals on subshifts*, *Comm. Math. Phys.* **134, no. 2** (1990) 217-236
16. O. Jenkinson, *Rotation, entropy, and equilibrium states*, *Trans. Amer. Math. Soc.* **353, no. 9** (2001) 3713-3739
17. R. Leplaideur, *A dynamical proof for the convergence of Gibbs measures at temperature* 0*,* preprint (2004)
18. A. Livsic, *Cohomology properties of Dynamical Systems*, *Math. USSR- Izvestia* **6** (1972) 1278-1301
19. A. O. Lopes and P. Thieullen, *Subactions for Anosov Diffeomorphisms*, *Asterisque*, **287**, *Geometric Methods in Dynamics (II)* (2003) 135-146
20. A. O. Lopes and P. Thieullen, *Subactions for Anosov Flows*, *Erg Theo and Dyn Systems*, **25** (2005) 605-628
21. A.O. Lopes and P. Thieullen, *Mather Theory and the Bowen-Series transformation*, to appear in *Annal. Inst. Henry Poincaré, Anal Non-lin.* preprint (2002).
22. S. Orey, *Large deviations in ergodic theory*, In K. L. Chung, E. Cinlar and R. K. Getoor, editor, *Seminar on Stochastic Processes* **12** Birkhauser (1985) 195-249
23. W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, *Asterisque* **187-188** (1990) 1-268
24. M. Pollicott and R. Sharp, *Livsic theorems, maximizing measures and the stable norm*, *Dyn. Sys.* **19** (2004) 75-88
25. F. Spitzer, *A variational characterization of finite Markov chains*, *The Annals of Math Stat.* **43, N 1** (1972) 303-307
26. D. Strook, **An introduction to the theory of large deviations**, (Springer Verlag 1984)
27. G. Yuan and B. R. Hunt, *Optimal orbits of hyperbolic systems Nonlinearity* **12** (1999) 1207-1224