Ergod. Th. & Dynam. Sys. (2015), **35**, 1925–1961 © *Cambridge University Press, 2014*
doi:10.1017/etds.2014.15

Entropy and variational principle for one-dimensional lattice systems with a general *a priori* probability: positive and zero temperature

A. O. LOPES, J. K. MENGUE, J. MOHR and R. R. SOUZA

Instituto de Matemática, UFRGS - Porto Alegre, Brazil
(e-mail: arturoscar.lopes@gmail.com)

(*Received* 17 *October* 2012 *and accepted in revised form* 7 *February* 2014)

Abstract. We generalize several results of the classical theory of thermodynamic formalism by considering a compact metric space M as the state space. We analyze the shift acting on $M^{\mathbb{N}}$ and consider a general *a priori* probability for defining the transfer (Ruelle) operator. We study potentials A which can depend on the infinite set of coordinates in $M^{\mathbb{N}}$. We define entropy and by its very nature it is always a non-positive number. The concepts of entropy and transfer operator are linked. If M is not a finite set there exist Gibbs states with arbitrary negative value of entropy. Invariant probabilities with support in a fixed point will have entropy equal to minus infinity. In the case $M = S^1$, and the *a priori* measure is Lebesgue dx, the infinite product of dx on $(S^1)^{\mathbb{N}}$ will have zero entropy. We analyze the Pressure problem for a Hölder potential A and its relation with eigenfunctions and eigenprobabilities of the Ruelle operator. Among other things we analyze the case where temperature goes to zero and we show some selection results. Our general setting can be adapted in order to analyze the thermodynamic formalism for the Bernoulli space with countable infinite symbols. Moreover, the so-called XY model also fits under our setting. In this last case M is the unitary circle S^1. We explore the differentiable structure of $(S^1)^{\mathbb{N}}$ by considering a certain class of smooth potentials and we show some properties of the corresponding main eigenfunctions.

1. *Introduction*

Let (M, d_M) be a compact metric space. We consider the metric in $M^{\mathbb{N}}$ given by

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} d_M(x_n, y_n),$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

where $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$. Note that $\mathcal{B} := M^{\mathbb{N}}$ is compact by Tychonoff's theorem.

We denote by H_α the set of α-Hölder functions $A : \mathcal{B} \to \mathbb{R}$ with the norm

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

where

$$\|A\| = \sup_{x \in \mathcal{B}} |A(x)| \quad \text{and} \quad |A|_\alpha = \sup_{x \neq y} \frac{|A(x) - A(y)|}{d(x, y)^\alpha}.$$

Here $\sigma : \mathcal{B} \to \mathcal{B}$ denotes the shift map which is defined by

$$\sigma(x_1, x_2, x_3, \ldots) = (x_2, x_3, x_4, \ldots).$$

Let $\mathcal{C}$ be the space of continuous functions from $\mathcal{B}$ to $\mathbb{R}$, and we will fix an *a priori* probability measure ν on the Borel sigma algebra over M. We assume that the support of ν is the set M. We stress the crucial point: ν needs to be a *probability measure*, not only a *measure*. Note that from our hypothesis if x_0 is isolated then $\nu(x_0) > 0$.

For a fixed potential $A \in H_\alpha$ we define a transfer operator (also called a Ruelle operator) $\mathcal{L}_A : \mathcal{C} \to \mathcal{C}$ by the rule

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

where $x \in \mathcal{B}$ and $ax = (a, x_1, x_2, \ldots)$ denotes a preimage of x with $a \in M$.

We call this general setting one-dimensional lattice system theory. Rigorous mathematical results on statistical mechanics are presented in [**1**, **6**, **20**, **24**, **27**, **30**, **32**, **50**, **57**, **60**, **61**].

We point out that a Holder potential A defined on $M^{\mathbb{Z}}$ is coboundary with a potential in $M^{\mathbb{N}}$ (same proof as in [**52**]). In this way the statistical mechanics of interactions on $M^{\mathbb{Z}}$ can be understood via the analysis of the similar problem in $M^{\mathbb{N}}$.

In [**3**] the Gibbs measure was investigated at positive and zero temperature for a potential A (which depends on infinite coordinates) in the case $M = S^1$, where the *a priori* measure is Lebesgue measure. This is the so-called XY model (see [**23**, **45**, **59**]) which is considered in several applications to real problems in physics. The spin in each site of the lattice is described by an angle from $[0, 2\pi)$. In the physics literature, as far as we know, the potential A depends on two coordinates. A well-known example in applications is the potential $A(x) = A(x_0, x_1) = \cos(x_1 - x_0 - \alpha) + \gamma \cos(2x_0)$.

The present paper is a generalization of the setting presented in [**3**] for positive and zero temperature. We will also consider here a topic which was not addressed there, namely, the equilibrium (maximizing pressure) measure for the potential A. For a dynamical systems such that any point has an infinite number of preimages it is possible to define entropy via the Ruelle operator as far as one consider an *a priori* probability (and not a measure).

There are several possible points of view for understanding Gibbs states in statistical mechanics (see [**54**, **56**] for interesting discussions). We prefer the transfer (Ruelle) operator method because we believe that the eigenfunctions and eigenprobabilities (which can be derived from the theory) allow a deeper understanding of the problem. For example, the information one can get from the main eigenfunction (defined in the whole lattice) is worthwhile, mainly in the limit when temperature goes to zero.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

We point out that after the present paper was written, in another publication (see [**46**]), the Ruelle operator point of view was also considered for the understanding of continuous-time Gibbs states for the setting of continuous-time Markov chains. The concept of entropy can be defined in the same way as here and it is also negative (due to the fact of the existence of an infinite number of preimages for any point).

Examples. Now we give a brief description of some other examples that fit in our setting. The last example will be explained in detail in §5.

(i) If the alphabet is given by $M = \{1, 2, \ldots, d\}$, and the *a priori* measure is given by $\nu = (1/d) \sum_{i=1}^{d} \delta_i$, then we have the original full shift in a finite set of d symbols and the transfer operator is the classical Ruelle operator associated to a potential $A - \log(d)$ (see, for example, [**35**, **52**]). More precisely,

$$\mathcal{L}_A(\varphi)(x) = \int_M e^{A(ax)} \varphi(ax)\, d\nu(a) = \sum_{a \in \{1,2,\ldots,d\}} e^{A(ax) - \log(d)} \varphi(ax).$$

If we change the *a priori* measure to $\nu = \sum_{i=1}^{d} p_i \cdot \delta_i$, where $p_i > 0$, and $\sum_{i=1}^{d} p_i = 1$, then

$$\mathcal{L}_A(\varphi)(x) = \sum_{a \in \{1,2,\ldots,d\}} e^{A(ax)} \varphi(ax) p_a = \sum_{a \in \{1,2,\ldots,d\}} e^{A(ax) + \log(p_a)} \varphi(ax)$$

is the classical Ruelle operator with potential $A + \log(P)$, where $P(x_1, x_2, \ldots) = p_{x_1}$.

(ii) If $M_0 = \{z_i,\ i \in \mathbb{N}\}$ is a countable infinite subset of S^1, where each point is isolated, and there is only one accumulating point $z_\infty \in S^1 \setminus M_0$, then $M = M_0 \cup \{z_\infty\}$ is a compact set. In this case M can be identified with $\mathbb{N}$, where a special point z_∞ plays the role of infinity (that is, a one-point compactification). We consider here the restricted distance we get from S^1 in M. If $\sum_{i \in \mathbb{N}} p_i = 1$ with $p_i \geq 0$ and $\nu = \sum_{i \in \mathbb{N}} p_i \delta_{z_i}$, then ν is supported on the whole M, but z_∞ is not an atom for ν. The thermodynamic formalism with state space $\mathbb{N}$, or $\mathbb{Z}$, is considered for example in [**19**, **31**, **34**, **51**, **55**, **56**]. We will analyze in §5 some of these results in the present setting.

Our main purpose here is to describe a general theory for the statistical mechanics of one-dimensional spin lattices. We point out that most of the papers on the subject assume that the potential A depends just on two (or a finite number of) coordinates (as, for instance, is the case of [**1**, **6**, **30**]). We consider potentials which can depend on the infinite set of coordinates in $M^{\mathbb{N}}$.

In §3 we consider the entropy, pressure and variational principle and its relations with eigenfunctions and eigenprobabilities of the Ruelle operator. This setting, as far as we know, was not considered before. In this case the entropy, by its very nature, is always a non-positive number. If M is not a finite set, invariant probabilities with support in a fixed point will have entropy equal to minus infinity. The infinite product of $d\nu$ on $M^{\mathbb{N}}$ will have zero entropy. We point out that, although at first glance, the fact that the entropy we define here is negative may look strange, our definition is the natural extension of the concept of Kolomogorov entropy. In the classical case, the entropy is positive because the *a priori* measure is not a probability: it is the counting measure.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Entropy and pressure were considered before in other settings, as for instance in [**32**, §II] or [**60**]. In these works the authors consider a variational principle on boxes of finite length, and then they get the equilibrium as the limit probability on the lattice, when the size of the box goes to infinity. The concept of entropy was considered relative to a certain probability on the box (which in some sense plays the role of the *a priori* probability). Our formalism is derived from the Ruelle operator point of view and is close to the approach described, for instance, in [**52**], where the probabilities are consider directly on $M^{\mathbb{N}}$. As we will see the concepts of entropy and the transfer operator are very much related. When the potential A depends on an infinite number of coordinates in the lattice we believe our approach is more simple to state and to understand.

Other authors in previous works also considered entropy and transfer operators on one-dimensional spin lattices over metric spaces (see, for instance, [**50**, §III] or [**37**, §A3, or Proposition A4.9]), but we believe our approach is different.

Among other things we consider in §4 the case where temperature goes to zero and show some selection results related with the ergodic optimization (see [**17**, **18**, **25**, **33**, **38**, **41**, **42**, **58**]). Using the variational principle we obtain a simple proof of the fact that Gibbs states converge to maximizing measures when the temperature goes to zero (a question not discussed in [**3**]).

An important issue that does not appear in the classical thermodynamic formalism (in the sense of [**52**] and [**35**]) is the differentiable structure. We will show in §6 that for a certain class of smooth potentials A the associated main eigenfunction is also smooth.

2. *Ruelle operator*

Let a^n be an element of M^n having coordinates $a^n = (a_n, a_{n-1}, \ldots, a_2, a_1)$, we denote by $a^n x \in \mathcal{B}$ the concatenation of $a^n \in M^n$ with $x \in \mathcal{B}$, i.e. $a^n x = (a_n, \ldots, a_1, x_1, x_2, \ldots)$. In the case of $n = 1$ we will write $a := a^1 \in M$, and $ax = (a, x_1, x_2, \ldots)$.

The nth iterate of $\mathcal{L}_A$ has the following expression

$$\mathcal{L}_A^n(\varphi)(x) = \int_{M^n} e^{S_n A(a^n x)} \varphi(a^n x)\, dv^n(a^n),$$

where $S_n A(a^n x) = \sum_{k=0}^{n-1} A(\sigma^k(a^n x))$ and $dv^n(a^n) = \prod_{k=1}^{n} dv(a_{n-k+1})$.

Let us show that $\mathcal{L}_A$ preserves the set of Hölder functions.

LEMMA 1. *If $\varphi \in H_\alpha$, then $\mathcal{L}_A(\varphi) \in H_\alpha$.*

Proof. We have

$$\frac{|\mathcal{L}_A(\varphi)(x) - \mathcal{L}_A(\varphi)(y)|}{d(x, y)^\alpha} = \frac{|\int_M e^{A(ax)}\varphi(ax)\, dv(a) - \int_M e^{A(ay)}\varphi(ay)\, dv(a)|}{d(x, y)^\alpha}.$$

Now we use the fact that if $\varphi, A \in H_\alpha$, then $e^A \varphi \in H_\alpha$ and, hence,

$$\frac{\int_M |e^{A(ax)}\varphi(ax) - e^{A(ay)}\varphi(ay)|\, dv(a)}{d(x, y)^\alpha} \le \mathrm{Hol}(e^A\varphi)\nu(M) = \mathrm{Hol}(e^A\varphi). \qquad \square$$

THEOREM 1. *Consider a fixed a priori probability ν. Let us fix $A \in H_\alpha$, then there exists a strictly positive Hölder eigenfunction ψ_A for $\mathcal{L}_A : \mathcal{C} \to \mathcal{C}$ associated to a strictly*

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

positive eigenvalue λ_A. This eigenvalue is simple, which means the eigenfunction is unique (modulo multiplication by constant).

Proof. For each $0 < s < 1$, we define the operator $\mathcal{T}_{s,A}$ on $\mathcal{C}$, given by

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

The introduction of the parameter s in the proof is an adaptation of an argument presented in [**10**] for the present setting.

An easy adaptation of the proof of Proposition 1 in [**3**] shows that $\mathcal{T}_{s,A}$ is an uniform contraction map. Let u_s be the unique fixed point for $\mathcal{T}_{s,A}$, then u_s satisfies

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

By the same arguments used in the proof of Proposition 2 in [**3**], we can prove that the family $\{u_s\}_{0<s<1}$ is an equicontinuous family of functions. It follows from (1) that

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

Hence, $-\|A\| \le (1-s)\min u_s \le (1-s)\max u_s \le \|A\|$, for any $0 < s < 1$.

The family $\{u_s^* = u_s - \max u_s\}_{0<s<1}$ is equicontinuous and uniformly bounded. Let us fix a subsequence $s_n \to 1$ such that $[(1-s_n)\max u_{s_n}] \to k$, and that, using Arzela–Ascoli theorem, $\{u_{s_n}^*\}_{n\ge1}$ has an accumulation point in $\mathcal{C}$, which we will call u.

Observe that for any s

$$\begin{aligned} e^{u_s^*(x)} &= e^{u_s(x)-\max u_s} = e^{-(1-s)\max u_s + u_s(x) - s\max u_s} \\ &= e^{-(1-s)\max u_s} \int_M e^{A(ax)+(su_s(ax)-s\max u_s)}\, d\nu(a). \end{aligned}$$

Taking the limit where n goes to infinity for the sequence s_n, we get that u satisfies

$$e^{u(x)} = e^{-k}\int_M e^{A(ax)+u(ax)}\, d\nu(a) = e^{-k}\mathcal{L}_A(e^u)(x).$$

This shows that $\psi_A := e^u$ is a positive Holder eigenfunction for $\mathcal{L}_A$ associated to the eigenvalue $\lambda_A := e^k$.

The proof of the uniqueness is exactly the same as presented in [**3**] (see the comments after Theorem 3 in that paper). □

The eigenfunction is unique up a multiplicative factor. There are several ways to normalize it. We assume in this moment that the maximum of the eigenfunction is equal to one.

We point out that is possible to generalize the above result for *a priori* probabilities which depend on the point $x \in M$. This will require some mild assumptions on this family of probabilities. We will not address this question here.

We say that a potential B is normalized if $\mathcal{L}_B(1) = 1$, which means it satisfies

$$\int_M e^{B(ax)}\, d\nu(a) = 1 \quad \text{for all } x \in \mathcal{B}.$$

In particular, for all $x \in \mathcal{B}$, $a \to e^{B(ax)}\, d\nu(a)$ is a probability measure on M, and $\mathcal{L}_B u(x)$ can be seen as the expectation of the random variable u with respect to this probability measure defined by the point x.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Let $A \in H_\alpha$, ψ_A and λ_A given by Theorem 1, it is easy to see that

$$\int_M \frac{e^{A(ax)}\psi_A(ax)}{\lambda_A \psi_A(x)}\, dv(a) = 1 \quad \text{for all } x \in \mathcal{B}. \tag{2}$$

Therefore, we define the normalized potential $\bar{A}$ associated to A, as

$$\bar{A} := A + \log \psi_A - \log \psi_A \circ \sigma - \log \lambda_A, \tag{3}$$

where $\sigma : \mathcal{B} \to \mathcal{B}$ is the shift map. As $\psi_A \in H_\alpha$ we have that $\bar{A} \in H_\alpha$.

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

We say a probability measure μ over $\mathcal{F}$ is invariant, if for any Borel set B, we have that $\mu(B) = \mu(\sigma^{-1}(B))$. We denote by $\mathcal{M}_\sigma$ the set of invariant probability measures.

We note that $\mathcal{B}$ is a compact metric space and by the Riesz representation theorem, a probability measure on the Borel σ-algebra is identified with a positive linear functional $L : \mathcal{C} \to \mathbb{R}$ that sends the constant function 1 to the real number 1. We also note that $\mu \in \mathcal{M}_\sigma$ if and only if, for any $\psi \in \mathcal{C}$ we have

$$\int_{\mathcal{B}} \psi \, d\mu = \int_{\mathcal{B}} \psi \circ \sigma \, d\mu.$$

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

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

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

The next theorem is a generalization of Propositions 4 and 5 of [**3**]. Here we consider $\mathcal{L}_A : H_\alpha \to H_\alpha$.

THEOREM 2. *Let A be a Hölder continuous potential, not necessarily normalized, ψ_A and λ_A the eigenfunction and eigenvalue given by Theorem 1. We associate to A the normalized potential $\bar{A} = A + \log \psi_A - \log \psi_A \circ \sigma - \log \lambda_A$. Then:*

(i) *there exists a unique fixed point μ_A for $\mathcal{L}_{\bar{A}}^*$, which is a σ-invariant probability measure;*

(ii) *the measure*

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

satisfies $\mathcal{L}_A^*(\rho_A) = \lambda_A \rho_A$; therefore, ρ_A is an eigenmeasure for $\mathcal{L}_A^*$;

(iii) *for any Hölder continuous function $w : \mathcal{B} \to \mathbb{R}$, we have that, in the uniform convergence topology,*

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

and

$$\frac{\mathcal{L}_A^n(w)}{(\lambda_A)^n} \to \psi_A \int_{\mathcal{B}} w \, d\rho_A,$$

where $\mathcal{L}_A^n$ denotes the nth iterate of the operator $\mathcal{L}_A : H_\alpha \to H_\alpha$.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Proof. (i) We begin by observing that the normalization property implies that the convex and compact set of Borel probability measures on $\mathcal{B}$ is preserved by the operator $\mathcal{L}^*_{\bar{A}}$. Therefore, using the Tychonoff–Schauder theorem we conclude the existence of a fixed point μ_A for the operator $\mathcal{L}^*_{\bar{A}}$. Now we prove that μ_A is σ-invariant: if $\psi \in \mathcal{C}$, we have

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

where in the last equality we used the normalization hypothesis for $\bar{A}$. The uniqueness of the fixed point will be obtained in the proof of item (iii).

(ii) For any $\psi \in \mathcal{C}$, $\mathcal{L}^*_{\bar{A}}(\mu_A) = \mu_A$ implies that

$$\begin{aligned}\int_{\mathcal{B}} \psi \, d\mu_A &= \int_{\mathcal{B}} \psi \, d\mathcal{L}^*_{\bar{A}}(\mu_A) = \int_{\mathcal{B}} \mathcal{L}_{\bar{A}}(\psi) \, d\mu_A \\ &= \int_{\mathcal{B}} \left(\int_M \psi(ax) \frac{e^{A(ax)} \psi_A(ax)}{\lambda_A \psi_A(x)} \, d\nu(a) \right) d\mu_A(x).\end{aligned}$$

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

$$\int_{\mathcal{B}} \frac{\varphi}{\psi_A} \, d\mu_A = \frac{1}{\lambda_A} \int_{\mathcal{B}} \left(\int_M \varphi(ax) \frac{e^{A(ax)}}{\psi_A(x)} \, d\nu(a) \right) d\mu_A(x),$$

which is equivalent to

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

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

(iii) In order to prove item (iii) we will need two claims. The first claim can be proved by induction.

First claim: for any normalized Holder potential B, if $\|w\|$ denotes the uniform norm of the Holder function $w : \mathcal{B} \to \mathbb{R}$, we have

$$|\mathcal{L}^n_B(w)(x) - \mathcal{L}^n_B(w)(y)| \le \left[C_{e^B} \|w\| \left(\frac{1}{2^\alpha} + \cdots + \frac{1}{2^{n\alpha}} \right) + \frac{C_w}{2^{n\alpha}} \right] d(x, y)^\alpha,$$

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

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

Therefore, by the Arzela–Ascoli theorem there exists an accumulation point, $\bar{\omega}$, for $\{\mathcal{L}^n_{\bar{A}}\omega\}_{n\ge 0}$, i.e. there exists a subsequence $\{n_k\}_{k\ge 0}$ such that

$$\bar{\omega}(x) = \lim_{k\ge 0} \mathcal{L}^{n_k}_{\bar{A}}\omega(x). \tag{5}$$

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

To prove this claim, we begin by observing that

$$\sup \bar{\omega} \ge \sup \mathcal{L}_{\bar{A}}\bar{\omega} \tag{6}$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

(in fact, this inequality holds for any function w). Now, equation (5) implies

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

(possibly by a different subsequence) and this shows that what we have in (6) is indeed an equality: in fact, we have

$$\sup \bar{\omega} = \sup \mathcal{L}_{\bar{A}}^{n} \bar{\omega} \quad \text{for all } n \geq 0.$$

Now, let x_M^n be a maximum point of $\mathcal{L}_{\bar{A}}^{n} \bar{\omega}$, for any $n \geq 0$. We have

$$\bar{\omega}(x_M^0) = \mathcal{L}_{\bar{A}}^{n} \bar{\omega}(x_M^n),$$

and this proves the second claim, because the normalization property implies that $\mathcal{L}_{\bar{A}}^{n} \bar{\omega}(x_M^n)$ is a convex combination of $\bar{\omega}$ in the preimages of x_M^n (here we also use the fact that the support of the *a priori* probability is the *all* space M).

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

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

which shows that $\bar{\omega}$ does not depend on the subsequence chosen. Therefore, for any $x \in \mathcal{B}$ we have

$$\mathcal{L}_{\bar{A}}^{n} \omega(x) \to \bar{\omega} = \int_{\mathcal{B}} \omega\, d\mu_A.$$

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

To finish the proof of item (iii), as $A = \bar{A} - \log \psi_A + \log \psi_A \circ \sigma + \log \lambda_A$, we have

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

and therefore

$$\frac{\mathcal{L}_A^n(w)(x)}{\lambda_A^n} = \frac{1}{\lambda_A^n} \int_{M^n} e^{S_n A(a^n x)} w(a^n x)\, d\nu^n(a^n) = \psi_A(x) \int_{M^n} \frac{e^{S_n \bar{A}(a^n x)}}{\psi_A(a^n x)} w(a^n x)\, d\nu^n(a^n)$$

$$= \psi_A(x) \mathcal{L}_{\bar{A}}^{n}\left(\frac{w}{\psi_A}\right) \to \psi_A(x) \int_{\mathcal{B}} \frac{w}{\psi_A}\, d\mu_A = \psi_A(x) \int_{\mathcal{B}} w\, d\rho_A. \qquad \square$$

We call μ_A the *Gibbs probability* (or Gibbs state) for A. We will leave the term *equilibrium probability* (or equilibrium state) for the one which maximizes pressure. As we will see, this invariant probability measure over $\mathcal{B}$ describes the statistics in equilibrium for the interaction described by the potential A. The assumption that the potential is Hölder implies that the decay of iteration is fast.

We normalize ψ_A by assuming that $\max \psi_A = 1$. There are other possible normalizations. Therefore, ρ_A is not a probability measure. From the item (iii) above we conclude that

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

PROPOSITION 1. *The only Holder continuous eigenfunction ψ of $\mathcal{L}_A$ which is totally positive is ψ_A.*

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Proof. Suppose $\psi : \mathcal{B} \to \mathbb{R}$ is a Holder continuous eigenfunction of $\mathcal{L}_A$ associated to some eigenvalue λ. It follows from item (iii) of Theorem 2 that

$$\frac{\mathcal{L}_A^n(\psi)}{\lambda_A^n} \to \psi_A \int_{\mathcal{B}} \psi \, d\rho_A \quad \text{when } n \to \infty.$$

Therefore, if $\psi > c > 0$, then $\int_{\mathcal{B}} \psi \, d\rho_A > 0$. Moreover, $\mathcal{L}_A^n(\psi) = \lambda^n \psi$. This is only possible if $\lambda = \lambda_A$ and $\psi = \psi_A$. □

The next result follows from Proposition 7 of [**3**].

PROPOSITION 2. *Suppose $\bar{A}$ is normalized, then the eigenvalue $\lambda_{\bar{A}} = 1$ is maximal. Moreover, the remainder of the spectrum of $\mathcal{L}_{\bar{A}} : H_\alpha \to H_\alpha$ is contained in a disk centered at zero with radius strictly smaller than one.*

The proof is the same as presented in [**52**, Theorem 2.2 (ii)].

We denote $\lambda_{\bar{A}}^1 < \lambda_{\bar{A}} = 1$ the spectral radius of $\mathcal{L}_{\bar{A}}$ when restricted to the set

$$\left\{ w \in H_\alpha \;\middle|\; \int_{\mathcal{B}} w \, d\mu_A = 0 \right\},$$

(which is the orthogonal complement of the space of constant functions, i.e. the orthogonal complement of the eigenspace associated to the maximal eigenvalue). One can also show the exponential decay of correlation for Hölder functions [**3**], which implies mixing and ergodic properties for μ_A.

PROPOSITION 3. *If $v, w \in \mathcal{L}^2(\mu_A)$ are such that w is Hölder and $\int_{\mathcal{B}} w \, d\mu_A = 0$, then, there exists $C > 0$ such that for all n*

$$\int_{\mathcal{B}} (v \circ \sigma^n) w \, d\mu_A \leq C(\lambda_{\bar{A}}^1)^n.$$

In particular, μ_A is mixing and therefore ergodic.

For the proof one can adapt [**52**, Theorem 2.3].

3. *Entropy and variational principle*

In this section we will introduce a notion of entropy. Initially, this will be done only for Gibbs probabilities, and then we will extend this definition to invariant probabilities. After that we prove that the Gibbs probability obtained in the general setting above satisfies a variational principle. We will also study some general properties of this notion of entropy and compare it with the classical Kolmogorov entropy when $M = \{1, \ldots, d\}^{\mathbb{N}}$. Finally, we will show that this definition is an extension of the notion of entropy for Markov measures (which are the Gibbs measures when the potential depends only on the first two coordinates), as introduced in [**45**].

Remember that $\mathcal{M}_\sigma$ is the set of invariant probability measures.

Definition 1. Let ν be a fixed *a priori* probability on M. We denote by $\mathcal{G} = \mathcal{G}_\nu$ the set of Gibbs measures, which means the set of $\mu \in \mathcal{M}_\sigma$, such that, $\mathcal{L}_B^*(\mu) = \mu$, for some normalized potential $B \in H_\alpha$. We define the entropy of $\mu \in \mathcal{G}$ as

$$h(\mu) = h^\nu(\mu) = -\int_{\mathcal{B}} B(x) \, d\mu(x).$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

We will see below that $-\int_{\mathcal{B}} B\, d\mu$ is the infimum of

$$\left\{ -\int_{\mathcal{B}} A\, d\mu + \log(\lambda_A) \;\middle|\; A \in H_\alpha \right\}.$$

The above definition is different from that briefly mentioned in [**3**, §3].

Remark. This concept of entropy depends on the choice of the *a priori* measure ν, which we choose to be a probability. In the classical case, when $M = \{1, \dots, d\}$, the entropy $H(\mu)$ is computed with the *a priori* measure ν given by $\sum_{j=1}^{d} \delta_j$ (which is not a probability). A comparison of the value of the above entropy $h(\mu)$, when $M = \{1, \dots, d\}$, with the classical Kolmogorov entropy $H(\mu)$ (for the full shift) is discussed below, after Proposition 7. For example, if the *a priori* probability is $\nu = (1/d) \sum_{j=1}^{d} \delta_j$, to get the entropy $h(\mu)$ you just have to add $-\log d$ to the classical one $H(\mu)$. Therefore, in this particular case, the above definition results in a number between $-\log(d)$ and zero. We point out that in the case M has infinite cardinality the above definition $h(\mu)$ makes sense, is well defined, and it is the natural generalization of the previous concept.

Remark. Let μ be a Gibbs measure and B the normalized potential associated to μ, if we call $J = e^{-B}$, we have an equivalent definition of entropy given by

$$h(\mu) = \int_{\mathcal{B}} \log(J(x))\, d\mu(x).$$

We point out that $J = e^{-B}$ does not correspond to the usual concept of Jacobian of the measure μ. For example, consider a finite alphabet $M = \{1, \dots, d\}$ and ν the *a priori* probability given by $\nu(i) = p_i$, where $p_i \geq 0$ and $\sum_{i=1}^{d} p_i = 1$. In this setting, the Ruelle operator is given by

$$\mathcal{L}_B w(x) = \sum_{i=1}^{d} e^{B(ix)} w(ix)\, p_i,$$

which can be rewritten as

$$\mathcal{L}_B w(x) = \sum_{i=1}^{d} e^{B(ix)+\log(p_i)} w(ix).$$

The last formulation fits in the classical thermodynamical formalism setting (see [**52**]), for a potential $\tilde{B}(ix) = B(ix) + \log(p_i)$, where we know that the Jacobian (defined as $\lim_{n\to\infty}(\mu[x_1 x_2 \cdots x_n]/\mu[x_2 x_3 \cdots x_n])$, when $[x_1 x_2 \cdots x_n]$ is the usual cylinder set) is given by $e^{-\tilde{B}} = e^{-B-\log(p_i)}$.

PROPOSITION 4. *If $\mu \in \mathcal{G}$, then we have $h(\mu) \leq 0$.*

Proof. Let μ be a probability on $\mathcal{G}$ with associated normalized potential B. We have

$$h(\mu) = -\int_{\mathcal{B}} B(x)\, d\mu(x) = \int_{\mathcal{B}} \log e^{-B(x)}\, d\mu(x) \leq \log \int_{\mathcal{B}} e^{-B(x)}\, d\mu(x)$$
$$= \log \int_{\mathcal{B}} e^{-B(x)}\, d\mathcal{L}_B^*(\mu)(x) = \log \int_{\mathcal{B}} \mathcal{L}_B e^{-B(x)}\, d\mu(x) = 0,$$

where we have used Jensen's inequality and also $\mathcal{L}_B e^{-B(x)} = 1$. □

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

This negative entropy property will be useful in the next section to give an easy proof that the Gibbs measures of βA select maximizing measures for A, when $\beta \to +\infty$. This result is obtained in the classical thermodynamic formalism theory because the entropy is bounded. It is also not difficult to get this in the present setting because the current notion of entropy is bounded above (by zero).

Now we state a lemma that will be used to prove the main result of this section, namely, the variational principle of Theorem 3. This lemma was shown to be true in the case M is finite (and the classical Kolmogorov entropy) in [**43**].

LEMMA 2. *Let us fix a Hölder continuous potential A and a measure $\mu \in \mathcal{G}$ with associated normalized potential B. We call $\mathcal{C}^+$ the space of continuous positive functions on $\mathcal{B}$. We have*

$$h(\mu) + \int_{\mathcal{B}} A(x)\, d\mu(x) = \inf_{u \in \mathcal{C}^+} \left\{ \int_{\mathcal{B}} \log\left(\frac{\mathcal{L}_A u(x)}{u(x)} \right) d\mu(x) \right\}.$$

Proof. If we take $\tilde{u}(x) = e^{-A(x)+B(x)}$, then

$$\log\left(\frac{\mathcal{L}_A \tilde{u}(x)}{\tilde{u}(x)} \right) = \log\left(\frac{\int_M e^{B(ax)}\, d\nu(a)}{e^{-A(x)+B(x)}} \right) = A(x) - B(x).$$

Integrating, we get

$$\begin{aligned} \int_{\mathcal{B}} \log\left(\frac{\mathcal{L}_A \tilde{u}(x)}{\tilde{u}(x)} \right) d\mu(x) &= \int_{\mathcal{B}} A(x)\, d\mu(x) - \int_{\mathcal{B}} B(x)\, d\mu(x) \\ &= h(\mu) + \int_{\mathcal{B}} A(x)\, d\mu(x). \end{aligned}$$

Now, let us consider a general $\bar{u} \in \mathcal{C}^+$. Using the fact that e^{-A+B} is a positive function, we can write $\bar{u}(x) = u(x)e^{-A(x)+B(x)}$. Note that, in this case,

$$\mathcal{L}_A \bar{u}(x) = \int_M e^{B(ax)} u(ax)\, d\nu(a) = \mathcal{L}_B u(x).$$

Hence,

$$\log\left(\frac{\mathcal{L}_A \bar{u}(x)}{\bar{u}(x)} \right) = \log(\mathcal{L}_B u(x)) - \log u(x) + A(x) - B(x),$$

and therefore, by integration, we get

$$\begin{aligned} \int_{\mathcal{B}} \log\left(\frac{\mathcal{L}_A \bar{u}(x)}{\bar{u}(x)} \right) d\mu(x) &= \int_{\mathcal{B}} \log(\mathcal{L}_B u(x))\, d\mu(x) - \int_{\mathcal{B}} \log u(x)\, d\mu(x) \\ &\quad + \int_{\mathcal{B}} A(x)\, d\mu(x) - \int_{\mathcal{B}} B(x)\, d\mu(x). \end{aligned}$$

Now, all we need to prove is that

$$\int_{\mathcal{B}} \log(\mathcal{L}_B u(x))\, d\mu(x) - \int_{\mathcal{B}} \log u(x)\, d\mu(x) \geq 0.$$

In order to do that, we use Jensen inequality, and we get $\log(\mathcal{L}_B u(x)) \geq \mathcal{L}_B \log u(x)$, which implies

$$\int_{\mathcal{B}} \log(\mathcal{L}_B u(x))\, d\mu(x) \geq \int_{\mathcal{B}} \mathcal{L}_B \log u(x)\, d\mu(x) = \int_{\mathcal{B}} \log u(x)\, d\mu(x),$$

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

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Let $\mu \in \mathcal{G}$, with associated normalized potential B, $A \in H_\alpha$ and ψ_A and λ_A, respectively, the positive eigenfunction and the maximal eigenvalue of $\mathcal{L}_A$, given by Theorem 1. From Lemma 2 we have

$$h(\mu) + \int_{\mathcal{B}} A \, d\mu = \inf_{u \in \mathcal{C}^+} \left\{ \int_{\mathcal{B}} \log \left(\frac{\mathcal{L}_A u(x)}{u(x)} \right) d\mu(x) \right\}$$
$$\leq \left\{ \int_{\mathcal{B}} \log \left(\frac{\mathcal{L}_A \psi_A}{\psi_A} \right) d\mu(x) \right\} = \log \lambda_A.$$

This implies

$$h(\mu) = -\int_{\mathcal{B}} B \, d\mu \leq -\int_{\mathcal{B}} A \, d\mu + \log \lambda_A \quad \text{for all } A \in H_\alpha,$$

with equality if $A = B$ (as $\lambda_B = 1$). Therefore,

$$h(\mu) = \inf_{A \in H_\alpha} \left\{ -\int_{\mathcal{B}} A \, d\mu + \log \lambda_A \right\},$$

with the minimum attained at B. Now, based on the last equation, we can extend the definition of entropy for all invariant measures.

Definition 2. Let μ be an invariant measure. We define the entropy of μ as

$$h^\nu(\mu) = h(\mu) = \inf_{A \in H_\alpha} \left\{ -\int_{\mathcal{B}} A \, d\mu + \log \lambda_A \right\},$$

where λ_A is the maximal eigenvalue of $\mathcal{L}_A$, given by Theorem 1.

This value is non-positive and can be $-\infty$ as we will see later.

Definition 3. Given a Hölder potential A we call the pressure of A the value

$$P(A) = \sup_{\mu \in \mathcal{M}_\sigma} \left\{ h(\mu) + \int_{\mathcal{B}} A(x) \, d\mu(x) \right\}.$$

A probability which attains such a maximum value is called equilibrium state for A.

In the literature sometimes this value is called the free energy (see [**20**] for instance).

Now we will show the variational principle of pressure which characterizes the equilibrium state.

THEOREM 3. (Variational principle) *Let $A \in H_\alpha$ be a Hölder continuous potential and λ_A be the maximal eigenvalue of $\mathcal{L}_A$, then*

$$\log \lambda_A = P(A) = \sup_{\mu \in \mathcal{M}_\sigma} \left\{ h(\mu) + \int_{\mathcal{B}} A(x) \, d\mu(x) \right\}.$$

Moreover, the supremum is attained on the Gibbs measure, i.e. the measure μ_A that satisfies $\mathcal{L}^*_{\bar{A}}(\mu_A) = \mu_A$.

Therefore, the Gibbs state and the equilibrium state for A are given by the same measure μ_A, which is the unique fixed point for the dual Ruelle operator associated to the normalized potential $\bar{A}$.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Proof. Consider a fixed $A \in H_\alpha$, by the definition of entropy, we have

$$\sup_{\mu\in\mathcal{M}_\sigma}\left\{h(\mu)+\int_{\mathcal{B}} A(x)\,d\mu(x)\right\}$$
$$=\sup_{\mu\in\mathcal{M}_\sigma}\left\{\inf_{B\in H_\alpha}\left\{-\int_{\mathcal{B}} B\,d\mu+\log\lambda_B\right\}+\int_{\mathcal{B}} A(x)\,d\mu(x)\right\}$$
$$\le\sup_{\mu\in\mathcal{M}_\sigma}\left\{-\int_{\mathcal{B}} A\,d\mu+\log\lambda_A+\int_{\mathcal{B}} A(x)\,d\mu(x)\right\}=\log\lambda_A.$$

Hence,

$$\log\lambda_A\ge\sup_{\mu\in\mathcal{M}_\sigma}\left\{h(\mu)+\int_{\mathcal{B}} A(x)\,d\mu(x)\right\}.$$

On the order hand, as $A\in H_\alpha$, from Theorem 1 we know that there exists λ_A and φ_A, such that $\mathcal{L}_A(\varphi_A)=\lambda_A\varphi_A$. Now, if we define $\bar{A}=A+\log\varphi_A-\log\varphi_A\circ\sigma-\log\lambda_A$, then, by Theorem 2, there exists a measure μ_A such that $\mathcal{L}^*_{\bar{A}}(\mu_A)=\mu_A$. This implies $\mu_A\in\mathcal{G}$, and

$$h(\mu_A)=-\int_{\mathcal{B}}\bar{A}(x)\,d\mu_A(x)=-\int_{\mathcal{B}} A\,d\mu_A+\log\lambda_A.$$

Therefore,

$$\log\lambda_A=h(\mu_A)+\int_{\mathcal{B}} A\,d\mu_A\le\sup_{\mu\in\mathcal{M}_\sigma}\left\{h(\mu)+\int_{\mathcal{B}} A(x)\,d\mu(x)\right\}.\qquad\square$$

In [**45**] a variational principle of pressure was considered. Other variational principles of pressure were described in [**32**, **50**]. Our approach and also the kind of probabilities we consider are different from those in this last reference.

THEOREM 4. (Pressure as minimax) *Given a Hölder potential A*

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

Proof. This follows at once from Lemma 2 (see also [**47**]). $\square$

It is known that periodic orbits can be used to gain information about the pressure in the classical thermodynamic formalism setting, and also to approximate the equilibrium measure (see [**52**, Ch. 5] and [**44**]). Therefore, the next corollary can be useful.

COROLLARY 1. *For each $a^n\in M^n$, $a^n=(a_n,\ldots,a_1)$ let $a^\infty\in\mathcal{B}$ be the periodic orbit of period n obtained by the successive concatenation of a^n, i.e. $a^\infty=(a_n,\ldots,a_1,a_n,\ldots,a_1,\ldots)$. Then,*

$$P(A)=\lim_{n\to\infty}\frac{1}{n}\log\left(\int_{M^n} e^{S_nA(a^\infty)}\,d\nu^n(a^n)\right).$$

Proof. Using Theorem 2(iii) and then the variational principle we conclude that, for any fixed $x\in\mathcal{B}$,

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\int_{M^n} e^{S_nA(a^nx)}\,d\nu^n(a^n)\right)=\lim_{n\to\infty}\frac{1}{n}\log(\mathcal{L}^n_A(1)(x))=\log(\lambda_A)=P(A).$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Using the fact that A is Hölder continuous, there exists a constant $C > 0$ such that

$$|S_n A(a^n x) - S_n A(a^\infty)| \le C\left(\frac{1}{2^\alpha} + \frac{1}{2^{2\alpha}} + \cdots + \frac{1}{2^{n\alpha}}\right) d(x, a^\infty)^\alpha.$$

Therefore, using that $\mathcal{B}$ is a compact set with finite diameter, there exists a constant $C > 0$, such that,

$$|S_n A(a^n x) - S_n A(a^\infty)| \le C,$$

for any $n \in \mathbb{N}$, $x \in \mathcal{B}$ and $a \in M$. Then,

$$\int_{M^n} e^{S_n A(a^n x) - C}\, d\nu^n(a^n) \le \int_{M^n} e^{S_n A(a^\infty)}\, d\nu^n(a^n) \le \int_{M^n} e^{S_n A(a^n x) + C}\, d\nu^n(a^n),$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\int_{M^n} e^{S_n A(a^\infty)}\, d\nu^n(a^n)\right) = \lim_{n\to\infty} \frac{1}{n} \log\left(\int_{M^n} e^{S_n A(a^n x)}\, d\nu^n(a^n)\right) = P(A).$$

□

Now we present a few properties of entropy.

PROPOSITION 5. *The entropy has the following properties:*

(i) $h(\mu) \le 0$ *for any invariant measure* μ*;*

(ii) $\nu^\infty = \nu \times \nu \times \nu \times \cdots$ *has zero entropy;*

(iii) *the entropy is upper semi-continuous;*

(iv) *the entropy is a concave function in the space of invariant probabilities.*

Proof. (i) We point out that $A = 0$ is a normalized function, hence for any invariant measure μ we have $h(\mu) \le 0$.

(ii) We are going to show that ν^∞ is the equilibrium measure for $A = 0$. Indeed,

$$\begin{aligned}\int_{\mathcal{B}} g\, d\mathcal{L}_A^*(\nu^\infty) &= \int_{\mathcal{B}} \mathcal{L}_A g\, d\nu^\infty = \int_{\mathcal{B}} \int_M e^0 g(ax)\, d\nu(a)\, d\nu^\infty(x) \\ &\overset{z=ax}{=} \int_{\mathcal{B}} g(z)\, d\nu^\infty(z).\end{aligned}$$

which shows that $\mathcal{L}_A^*(\nu^\infty) = \nu^\infty$.

(iii) Fix an $\varepsilon > 0$ and suppose μ_n converges to μ. By definition of $h(\mu)$, we can choose $A \in H_\alpha$ such that

$$-\int_{\mathcal{B}} A\, d\mu + \log \lambda_A \le h(\mu) + \varepsilon.$$

If n is large enough, we have $|\int_{\mathcal{B}} A\, d\mu_n - \int_{\mathcal{B}} A\, d\mu| < \varepsilon$. Then,

$$h(\mu_n) \le -\int_{\mathcal{B}} A\, d\mu_n + \log \lambda_A \le -\int_{\mathcal{B}} A\, d\mu + \log \lambda_A + \varepsilon \le h(\mu) + 2\varepsilon,$$

therefore $\limsup_{n\to+\infty} h(\mu_n) \le h(\mu) + 2\varepsilon$.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

(iv) Let μ_1 and μ_2 be σ-invariant probabilities, $\varepsilon \in (0, 1)$ and $\mu = \varepsilon\mu_1 + (1-\varepsilon)\mu_2$. Then

$$\begin{aligned} h(\varepsilon\mu_1 + (1-\varepsilon)\mu_2) = h(\mu) &= \inf_A \left(-\int_{\mathcal{B}} A \, d\mu + \log \lambda_A \right) \\ &= \inf_A \left(-\varepsilon \int_{\mathcal{B}} A \, d\mu_1 + (1-\varepsilon) \int_{\mathcal{B}} A \, d\mu_2 + \log \lambda_A \right) \\ &\geq \inf_A \left(-\varepsilon \int_{\mathcal{B}} A \, d\mu_1 + \varepsilon \cdot \log \lambda_A \right) \\ &\quad + \inf_A \left(-(1-\varepsilon) \int_{\mathcal{B}} A \, d\mu_2 + (1-\varepsilon) \cdot \log \lambda_A \right) \\ &= \varepsilon h(\mu_1) + (1-\varepsilon) h(\mu_2). \end{aligned}$$

□

Remark. The entropy of a probability measure supported on periodic orbit can be $-\infty$. Indeed, suppose $M = [0, 1]$, and $A_c : M^{\mathbb{N}} \to \mathbb{R}$ given by $A_c(x) = \log((c/(1-e^{-c}))e^{-cx_1})$. Suppose the *a priori* measure ν is the Lebesgue measure. We have that for each $c > 0$, the function A_c is a C^1 normalized potential (therefore belongs to H_α), which depends only on the first coordinate of x. Note that $\mathcal{L}_{A_c}(1) = 1$. Let μ be the Dirac measure on 0^∞. We have

$$h(\mu) \leq -\int_{\mathcal{B}} A_c \, d\mu = -A_c(0^\infty) = -\log\left(\frac{c}{1-e^{-c}}\right) \to -\infty$$

when $c \to \infty$. This shows that $h(\mu) = -\infty$. An easy adaptation of the arguments can be done to prove that, in this setting, invariant measures supported on periodic orbits have entropy $-\infty$.

Note the subtle point that the entropy depends on the *a priori* probability and moreover all subsequent concepts we introduced, such as for example the Ruelle operator,

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

assume conditions on the preimages of σ. Therefore, given an iterate σ^n, if one wants to consider the entropy of a σ^n-invariant probability, then we need to specify a certain *a priori* probability. We will address this question now.

3.1. *Entropy of iterates.* Suppose M^n is the compact set given by

$$\{(x_1, x_2, \ldots, x_n) \mid x_i \in M, \ \forall 1 \leq i \leq n\},$$

with the sum or the maximum norm. Let σ^n be the shift map defined on the Bernoulli space given by $\mathcal{B}^n \equiv (M^n)^{\mathbb{N}}$. We know that σ^n is the nth iterate of σ in the original Bernoulli space, but we prefer to see σ^n as a new map defined on a new Bernoulli space. If we do that, all of the theory developed above applies to σ^n, we have a Ruelle operator with an *a priori* measure given by ν^n, and therefore the entropy of a Gibbs measure to the new map σ^n is well defined. Note that the new Bernoulli set $\mathcal{B}^n$ can be identified with the original $\mathcal{B}$ and an invariant measure for σ is also an invariant measure for σ^n.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

PROPOSITION 6. *If $\mathcal{G}^n$ denotes the set of Gibbs measures on $\mathcal{B}^n$, then $\mu \in \mathcal{G}$ implies $\mu \in \mathcal{G}^n$ and*

$$h_\mu^{\nu^n}(\sigma^n) = h_\mu(\sigma^n) = nh_\mu(\sigma) = nh_\mu^\nu(\sigma). \tag{7}$$

Proof. Note that if μ is a σ-invariant measure, then μ is invariant for σ^n. Also, if $B \in H_\alpha$ is a normalized potential for σ, then the Birkhoff sum $B^n \equiv \sum_{j=0}^{n-1} B \circ \sigma^j$ is a normalized potential for the map σ^n. Let us first prove that if μ is the Gibbs measure for the Ruelle operator associated to B, then μ (which is indeed a measure on $\mathcal{B}^n$) is also Gibbs for B^n. In order to do that, note that

$$\begin{aligned}\mathcal{L}_{B^n}^m(\varphi)(x) &= \int_{(M^n)^m} e^{S_m B^n(a^{nm}x)} \varphi(a^{nm}x)\, d(\nu^n)^m(a^{nm}) \\ &= \int_{(M^n)^m} e^{S_{mn} B(a^{nm}x)} \varphi(a^{nm}x) d\nu^{nm}(a^{nm}) = \mathcal{L}_B^{mn}(\varphi)(x) \to \int_{\mathcal{B}} \varphi \, d\mu.\end{aligned}$$

Now the σ^n-entropy of μ, given by the integral of the B^n, equals n times the σ-entropy of μ, because, using the fact that μ is σ-invariant, we have

$$\begin{aligned}h_\mu(\sigma^n) &= -\int_{\mathcal{B}^n} B^n \, d\mu = -\int_{\mathcal{B}^n} \sum_{j=0}^{n-1} B \circ \sigma^j \, d\mu \\ &= -\sum_{j=0}^{n-1} \int_{\mathcal{B}} B \circ \sigma^j \, d\mu = -n \int_{\mathcal{B}} B \, d\mu = nh_\mu(\sigma). \qquad \square\end{aligned}$$

3.2. *Relations with Kolmogorov entropy.* Let us consider the construction of the entropy by partitions method, in the case M is finite. We begin by remembering that, by the Kolmogorov–Sinai theorem, the classical entropy of μ, which we will denote by $H(\mu)$, is given by

$$H(\mu) = \lim_{n\to\infty} -\frac{1}{n} \sum_{i_1,\ldots,i_n} \mu([i_1 \cdots i_n]) \log(\mu([i_1 \cdots i_n])). \tag{8}$$

PROPOSITION 7. *Let $M = \{1, \ldots, d\}$ and $\nu = \sum_{i=1}^d p_i \delta_i$ be the a priori probability on M. For any Gibbs measure μ:*

(i)

$$H(\mu) = h^\nu(\mu) - \sum_{i=1}^d \log(p_i) \cdot \mu([i]);$$

(ii)

$$h^\nu(\mu) = -\lim_{n\to\infty} \frac{1}{n} \sum_{i_1,\ldots,i_n} \mu([i_1 \cdots i_n]) \log\left(\frac{\mu([i_1 \cdots i_n])}{p_{i_1} \cdots p_{i_n}}\right)$$

where

$$[i_1 \cdots i_n] = \{x \in M^{\mathbb{N}} \mid x_1 = i_1, \ldots, x_n = i_n\}.$$

Proof. If μ is a Gibbs measure, there exists a normalized potential A associated to μ, which implies

$$\int_M e^{A(ax)} \, d\nu(a) = \sum_{i=1}^d e^{A(ix)} p_i = 1 \quad \text{for all } x \in M^{\mathbb{N}},$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

which is equivalent to

$$\sum_{i=1}^{d} e^{A(ix)+\log(p_i)} = 1 \quad \text{for all } x \in M^{\mathbb{N}}.$$

Moreover, $\mathcal{L}_A^*(\mu) = \mu$ implies that μ is a fixed point for the classical Ruelle operator with the normalized potential $A + \log(P)$, where $P(y_1, y_2, \ldots) = p_{y_1}$. Therefore,

$$H(\mu) = -\int_{\mathcal{B}} A + \log(P)\, d\mu = h^{\nu}(\mu) - \int_{\mathcal{B}} \log(P)\, d\mu = h^{\nu}(\mu) - \sum_{i=1}^{d} \log(p_i) \cdot \mu([i]).$$

which ends the proof of item (i).

In order to prove item (ii), we note that, from the last equation, and using that μ is an σ-invariant measure, we have, for any $n \geq 1$,

$$\begin{aligned} H(\mu) &= h^{\nu}(\mu) - \frac{1}{n}\int_{\mathcal{B}} \log(P) + \cdots + \log(P \circ \sigma^{n-1})\, d\mu \\ &= h^{\nu}(\mu) - \frac{1}{n}\int_{\mathcal{B}} \log(p_{x_1} \cdots p_{x_n})\, d\mu(x) \\ &= h^{\nu}(\mu) - \frac{1}{n}\sum_{i_1,\ldots,i_n} \mu([i_1 \cdots i_n]) \log(p_{i_1} \cdots p_{i_n}). \end{aligned}$$

Then,

$$\begin{aligned} h^{\nu}(\mu) &= H(\mu) + \frac{1}{n}\sum_{i_1,\ldots,i_n} \mu([i_1 \cdots i_n]) \log(p_{i_1} \cdots p_{i_n}) \\ &= H(\mu) + \lim_{n\to\infty} \frac{1}{n}\sum_{i_1,\ldots,i_n} \mu([i_1 \cdots i_n]) \log(p_{i_1} \cdots p_{i_n}) \\ &= -\lim_{n\to\infty} \frac{1}{n}\sum_{i_1,\ldots,i_n} \mu([i_1 \cdots i_n]) \log\left(\frac{\mu([i_1 \cdots i_n])}{p_{i_1} \cdots p_{i_n}}\right), \end{aligned}$$

where in the last equation we used (8). □

In particular, it follows from item (i) above that, when $p_i = 1/d$, for all i, we have

$$h^{\nu}(\mu) = H(\mu) - \log(d).$$

The above proposition can be interpreted in the following way: in the classical definition of Kolmogorov entropy it is considered the *a priori* measure $\nu = \sum_{i=1}^{\infty} \delta_i$ on M, which is not a probability.

3.3. *Markov chains with values on S^1.* Now we recall the concept of Markov measures and show that the entropy defined above is an extension of the concept of entropy for Markov measures, as introduced in [**45**].

Let $K : M^2 \to \mathbb{R}$, $\theta : M \to \mathbb{R}$, satisfying

$$\begin{aligned} &\int_M K(x_1, x_2)\, d\nu(x_2) = 1 \quad \text{for all } x_1 \\ \text{and} \quad &\int_M \theta(x_1) K(x_1, x_2)\, d\nu(x_1) = \theta(x_2) \quad \text{for all } x_2. \end{aligned} \tag{9}$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

We call K a transition kernel and θ the stationary measure for K. As in [**45**], we define the absolutely continuous Markov measure associated to K and θ, as

$$\mu(A_1 \cdots A_n \times M^{\mathbb{N}}) := \int_{A_1 \cdots A_n} \theta(x_1)\, K(x_1, x_2) \cdots K(x_{n-1}, x_n)\, d\nu(x_n) \cdots d\nu(x_1), \tag{10}$$

for any cylinder $A_1 \cdots A_n \times M^{\mathbb{N}}$.

The next proposition show us the importance of absolutely continuous Markov measures.

PROPOSITION 8. *We will show that:*

(i) *given a Hölder continuous potential $A(x_1, x_2)$ (not necessarily normalized) depending on two coordinates, there exists a Markov measure that is Gibbs for A;*

(ii) *the converse is also true: given an absolutely continuous Markov measure defined by K and θ, there exists a certain Hölder continuous normalized potential $A(x_1, x_2)$, such that the Markov measure defined by θ and K is the Gibbs measure for A.*

Therefore, any absolutely continuous Markov measure is Gibbs for a potential depending on two variables and, conversely, any potential depending on two variables has a Gibbs measure which is an absolutely continuous Markov Measure.

In other words, if we restrict our analysis to potentials that depend just on the first two coordinates, we have that the set of absolutely continuous Markov Measures coincides with the set of Gibbs measures.

Proof. (i) Given a potential $A(x_1, x_2)$, non-normalized, as in [**45**] define $\theta_A : M \to \mathbb{R}$ by

$$\theta_A(x_1) := \frac{\psi_A(x_1)\ \bar{\psi}_A(x_1)}{\pi_A}, \tag{11}$$

and a transition $K_A : M^2 \to \mathbb{R}$ by

$$K_A(x_1, x_2) := \frac{e^{A(x_1,x_2)}\ \bar{\psi}_A(x_2)}{\bar{\psi}_A(x_1)\,\lambda_A}, \tag{12}$$

where ψ_A and $\bar{\psi}_A$ are the eigenfunctions associated to the maximal eigenvalue λ_A of the operators

$$L_A\psi(x_2) = \int_M e^{A(x_1,x_2)}\psi(x_1)\, d\nu(x_1) \quad \text{and} \quad \bar{L}_A\psi(x_1) = \int_M e^{A(x_1,x_2)}\psi(x_2)\, d\nu(x_2) \tag{13}$$

and $\pi_A = \int_M \psi_A(x_1)\bar{\psi}_A(x_1)\, d\nu(x_1)$.

Then, by the same arguments used to prove Theorem 16 of [**3**], we obtain that the Markov measure μ_A defined by (10) (considering K_A and θ_A) is Gibbs for A, i.e. a fixed point for the dual Ruelle operator $\mathcal{L}^*_{\bar{A}}$, where $\bar{A} = A + \log\psi_A(x_1) - \log\psi_A(x_2) - \log\lambda_A$.

(ii) Let K and θ satisfy (9), and define $A = \log K$, we have $\bar{L}_A(1) = 1$ which implies $\lambda_A = 1$ and $\bar{\psi}_A = 1$. Let ψ_A be maximal eigenfunction for L_A.

Using (12), we get $K_A(x_1, x_2) = e^{A(x_1,x_2)} = K(x_1, x_2)$. Define $\theta_A = \psi_A/\pi_A$. We have that θ_A is an invariant density for K, therefore $\theta_A = \theta$. Then, also by Theorem 16 of [**3**], we have that the Markov measure defined by K and θ is Gibbs for A. □

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

The next proposition shows us that the concept of entropy introduced in Definition 2 is a generalization of the concept of entropy defined in [**45**], which could only be applied to absolutely continuous Markov measures.

PROPOSITION 9. *Let μ be the Markov measure defined by a transition kernel K and a stationary measure θ, given in (10). The definition of entropy given in* [**45**]*:*

$$S(\theta K) = -\int_{M^2} \theta(x_1)K(x_1, x_2)\log(K(x_1, x_2))\, d\nu(x_1)\, d\nu(x_2) \leq 0$$

coincides with the present Definition 2.

Proof. As in the proof of Proposition 8, note that the normalized potential associated to $A(x, y) = \log K(x, y)$ is

$$\bar{A}(x_1, x_2) = \log K(x_1, x_2) + \log \psi_A(x_1) - \log \psi_A(x_2),$$

where ψ_A is the maximal eigenfunction of the operator L_A. Note also that $\bar{A}$ depends only on the first two coordinates.

Let μ be the Gibbs measure associated to $\bar{A}$, hence by Definition 2 we have

$$\begin{aligned} h(\mu) &= -\int_{\mathcal{B}} \bar{A}(x_1, x_2)\, d\mu(x) = -\int_{\mathcal{B}} \log K(x_1, x_2)\, d\mu(x) \\ &= \int_{M^2} \log K(x_1, x_2)\theta(x_1)K(x_1, x_2)\, d\nu(x_1)\, d\nu(x_2) = S(\theta K). \end{aligned}$$ □

4. *Zero temperature*

Consider a fixed Hölder potential A and a real variable $\beta > 0$. We denote, respectively, by $\psi_{\beta A}$ and $\mu_{\beta A}$, the eigenfunction for the Ruelle operator associated to βA and the equilibrium measure (Gibbs) for βA. We would like to investigate general properties of the limits of $\mu_{\beta_n A}$ and of $(1/\beta_n)\log \psi_{\beta_n A}$ when $\beta_n \to \infty$. Some results of this section are generalizations of those in [**3**]. It is well known that the parameter β represents the inverse of the temperature.

It is fair to call 'Gibbs state at zero temperature for the potential A' any of the weak limits of convergent subsequences $\mu_{\beta_n A}$. Even when the potential A is Hölder, Gibbs state at zero temperature do not have to be unique. In the case there exist the weak limit $\mu_{\beta A} \rightharpoonup \mu$, $\beta \to \infty$, we say that there exists a selection of Gibbs states for A at temperature zero.

Remark. Given β and A, the Hölder constant of $u_{\beta A} = \log(\psi_{\beta A})$, depends on the Hölder constant for βA, and is given by $\beta(2^\alpha/(2^\alpha - 1))\mathrm{Hol}_A$ (see [**3**]). As we normalize $\psi_{\beta A}$ assuming that $\max \psi_{\beta A} = 1$, the family of functions $(1/\beta)\log(\psi_{\beta A})$, $\beta > 0$, is uniformly bounded. Note that when we normalize $\psi_{\beta A}$ the Hölder constant of $\log(\psi_{\beta A})$ remains unchanged, which assures the family $(1/\beta)\log(\psi_{\beta A})$, $\beta > 0$, is equicontinuous.

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

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

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

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

The limits of $\mu_{\beta A}$ are related (see below) with the following problem: given $A : \mathcal{B} \to \mathbb{R}$ Holder, we want to find probabilities that maximize, over $\mathcal{M}_\sigma$, the value

$$\int_{\mathcal{B}} A(x)\, d\mu(\mathbf{x}).$$

We define

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

Any of the probability measures which attains the maximal value will be called a maximizing probability measure, which will be sometimes denoted generically by μ_∞. As $\mathcal{M}_\sigma$ is compact, there exist always at least one maximizing probability measure. It is also true that there exists ergodic maximizing probability measures. Indeed, the set of maximizing probability measures is convex, compact and the extreme probability measures of this convex set are ergodic (cannot be expressed as a convex combination of others [**35**]). Any maximizing probability measure is a convex combination of ergodic ones [**53**]. Results obtained in this setting belong to what is called ergodic optimization theory [**33**].

The possible limits of $(1/\beta_n) \log \psi_{\beta_n A}$ are related (see below) with the following concept.

Definition 4. A continuous function $u : \mathcal{B} \to \mathbb{R}$ is called a *calibrated subaction* for $A : \mathcal{B} \to \mathbb{R}$, if, for any $y \in \mathcal{B}$, we have

$$u(y) = \max_{\sigma(x)=y} [A(x) + u(x) - m(A)]. \tag{14}$$

This can also be expressed as

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

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

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

The above equation for u can be seen as a kind of discrete version of the concept of subsolution of the Hamilton–Jacobi equation [**5**, **15**, **21**]. It can also be seen as a kind of dynamic additive eigenvalue problem [**13**, **14**, **26**].

We note that $m(A)$ can be characterized by

$$m(A) = \inf\{\gamma \mid \exists u \in C,\ \gamma + u \circ \sigma - u - A \geq 0\},$$

where $\mathcal{C}$ denotes the set of continuous real-valued functions. In some sense this corresponds to the dual problem in transport theory [**39**]. Any invariant measure help us to estimate $m(A)$ from below. The continuous functions on the dual problem help us to estimate $m(A)$ from above.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

If u is a calibrated subaction, then $u + c$, where c is a constant, is also a calibrated subaction. An interesting question is when such calibrated subaction u is unique up to an additive constant (see [**45**] and [**25**]).

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

$$\int_{\mathcal{B}} [u(\sigma(x)) - u(x)]\, d\mu = 0.$$

Therefore, if μ_∞ is a maximizing probability measure for A and u is a calibrated subaction for A, then (see, for instance, [**17**, **33**, **58**] for a similar result) for any x in the support of μ_∞, we have

$$u(\sigma(x)) - u(x) - A(x) + m(A) = 0. \tag{15}$$

In this way if we know the value $m(A)$, then a calibrated subaction u for A can help us to identify the support of maximizing probabilities. The above equation can be eventually true outside the union of the supports of the maximizing probabilities (see an interesting example due to Leplaideur [**16**]).

We show below that if there exists a subsequence $\beta_n \to \infty$, such that on the uniform convergence

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

then such V is a calibrated subaction for A. When there exists a V which is the limit

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

(not just via a subsequence) we say we have selection of subactions at temperature zero. Positive results in this direction are presented in [**4**, **40**, **45**].

There exists here a subtle point. Subaction is a concept in ergodic optimization and does not depend on the existence of an *a priori* probability ν in M. On the other hand, the eigenfunction $\psi_{\beta_n A}$ is associated to a Ruelle operator, which depends on the *a priori* measure. In any case, for any given *a priori* probability ν, if the associated family of eigenfunctions $\psi_{\beta_n A}$ converges, it will converge to a subaction for A.

LEMMA 3. *For any β, we have $-\|A\| < (1/\beta) \log \lambda_\beta < \|A\|$.*

Proof. Fix $\beta > 0$. Let $\bar{x}$ be the maximum of $\psi_{\beta A}$ in $\mathcal{B}$ and $\tilde{x}$ be the minimum of $\psi_{\beta A}$ in $\mathcal{B}$. If $\|A\|$ is the uniform norm of A, we have

$$\lambda_\beta = \frac{1}{\psi_{\beta A}(\bar{x})} \int_M e^{\beta A(a\,\bar{x})} \psi_{\beta A}(a\,\bar{x})\, d\nu(a) \le \int_M e^{\beta A(a\,\bar{x})}\, d\nu(a) \le e^{\beta\|A\|},$$

and

$$\lambda_\beta = \frac{1}{\psi_{\beta A}(\tilde{x})} \int_M e^{\beta A(a\,\tilde{x})} \psi_{\beta A}(a\,\tilde{x})\, d\nu(a) \ge \int_M e^{\beta A(a\,\bar{x})}\, d\nu(a) \ge e^{-\beta\|A\|}. \qquad \square$$

The next result can be seen as a kind of measure-theoretical version of the Laplace's method.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

LEMMA 4. *Suppose $W_t : M \to \mathbb{R}$ converges uniformly to $W : M \to \mathbb{R}$, when $t \to +\infty$. Then*

$$\lim_{t\to\infty} \frac{1}{t} \log \int_M e^{tW_t(a)}\, d\nu(a) = \max_{a\in M} W(a).$$

Proof. Let $m = \max\{W(a) : a \in M\}$. Let $\bar{a} \in M$, such that $W(\bar{a}) = m$. Given $\varepsilon > 0$, there exist t_0 and δ, such that $W_t(a) > m - \varepsilon$, for any $a \in B(\bar{a}, \delta) \equiv \{a : d(a, \bar{a}) < \delta\}$ and $t > t_0$.

Therefore, if $t > t_0$, we have that

$$\int_M e^{tW_t(a)}\, d\nu(a) \geq \int_{B(\bar{a},\delta)} e^{tW_t(a)}\, d\nu(a) > \nu(B(\bar{a}, \delta))e^{t(m-\varepsilon)},$$

thus, if $t > t_0$,

$$\frac{1}{t} \log \int_M e^{tW_t(a)}\, d\nu(a) > \frac{1}{t} \log(\nu(B(\bar{a}, \delta))) + m - \varepsilon.$$

Hence,

$$\liminf_{t\to+\infty} \frac{1}{t} \log \int_M e^{tW_t(a)}\, d\nu(a) \geq m.$$

The other inequality is analogous, using the fact that: given ε, there exists t_0 such that, $W_t(a) < m + \varepsilon$, for any $t > t_0$ and $a \in M$. □

PROPOSITION 10. *Given a potential A Hölder continuous:*

(i) *we have*

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

(ii) *any limit, in the uniform topology,*

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

is a calibrated subaction for A.

Proof. Let β_n be a subsequence such that the following limit exists: $(1/\beta_n) \log \lambda_{\beta_n} \to k$, when $n \to \infty$. By taking a subsequence of β_n we can assume that also there exists V Holder, such that $V := \lim_{n\to\infty}(1/\beta_n) \log(\psi_{\beta_n A})$.

Given $x \in \mathcal{B}$, consider the equation

$$\lambda_{\beta_n} = \frac{1}{\psi_{\beta_n A}(x)} \int_M e^{\beta_n A(a\, x)} \psi_{\beta_n A}(a\, x)\, d\nu(a).$$

It follows from Lemma 4 that, when $n \to \infty$,

$$k = \max_{a\in M}\{A(ax) + V(ax) - V(x)\}.$$

First we show that $k \geq m(A)$. From the above it follows that

$$-V(\sigma(x)) + V(x) + A(x) \leq k.$$

Let μ be a σ-invariant probability measure, then

$$\int_{\mathcal{B}} A(x)\, d\mu(x) = \int_{\mathcal{B}} [-V(\sigma(x)) + V(x) + A(x)]\, d\mu(x) \leq k.$$

This implies $m(A) \leq k$.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Now we show that $m(A) \geq k$. For any $x \in \mathcal{B}$ there exist $y = a_x\, x$ such that $\sigma(y) = x$, and

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

Therefore, the compact set $K = \{y \mid -V(\sigma(y)) + V(y) + A(y) = k\}$ is such that, $K' = \cap_n \sigma^{-n}(K)$ is non-empty, compact and σ-invariant. If we consider a σ-invariant probability measure μ with support on K', we have that $\int_{\mathcal{B}} A(y)\, d\mu(y) = k$. From this it follows that $m(A) \geq k$.

From the above arguments $k = m(A)$ is the unique possible accumulation point of the bounded function $\beta \to (1/\beta) \log \lambda_{\beta\, A}$, then

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

Moreover, from the above expressions, we can say that any limit of convergent subsequence $\lim_{n\to\infty}(1/\beta_n) \log(\psi_{\beta_n A})$ is a calibrated subaction. □

Now we return to study the Gibbs measures at zero temperature. In the case $\mu_{\beta A} \rightharpoonup \mu_\infty$, when $\beta \to \infty$ (not just a subsequence), as we said before, we have selection of probability at temperature zero (see [**38**, **40**, **45**] for general positive results and [**11**, **12**, **61**] for negative results). The next result uses the variational principle proved in the previous section and the property that the entropy of an invariant probability is not positive.

THEOREM 5. *Consider a Hölder potential A. Suppose that for some subsequence we have $\mu_{\beta_n A} \rightharpoonup \mu_\infty$. Then μ_∞ is a maximizing probability, i.e.*

$$\int_{\mathcal{B}} A(x)\, d\mu_\infty(x) = m(A).$$

In the case the maximizing probability for A is unique, we have selection of Gibbs probability at temperature zero.

Proof. By definition, $\mu_{\beta_n A} \rightharpoonup \mu_\infty$, if and only if,

$$\lim_{n\to\infty} \int_{\mathcal{B}} w\, d\mu_{\beta_n A} = \int_{\mathcal{B}} w d\mu_\infty \quad \text{for all } w \in \mathcal{C}.$$

Now using Theorem 3 and the fact that $h(\mu) \leq 0$, we obtain

$$m(A) = \lim_{\beta\to\infty} \frac{\log \lambda_{\beta A}}{\beta} = \lim_{n\to\infty} \left(\int_{\mathcal{B}} A\, d\mu_{\beta_n A} + \frac{1}{\beta_n} h(\mu_{\beta_n A}) \right)$$
$$\leq \lim_{n\to\infty} \int_{\mathcal{B}} A\, d\mu_{\beta_n A} = \int_{\mathcal{B}} A\, d\mu_\infty$$

Hence, $m(A) \leq \int_{\mathcal{B}} A\, d\mu_\infty$. Also, as μ_∞ is a σ-invariant measure, we have that $m(A) \geq \int_{\mathcal{B}} A\, d\mu_\infty$. This implies that $m(A) = \int_{\mathcal{B}} A\, d\mu_\infty$. □

Questions related to the large deviation property on the XY model, when $\beta \to \infty$, are considered in [**40**]. The existence of a calibrated subaction plays an important role in this kind of result.

We consider now a different kind of question. From an easy adaptation of Theorem 2.1 in [**18**] one can show the following result.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

PROPOSITION 11. *We have*

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

We consider from now on a potential A which depends on two coordinates $A : S^1 \times S^1 \to \mathbb{R}$ of class C^∞. A smooth real-valued function on a manifold M is a Morse function if it has no degenerate critical points.

In [**1**] it is shown that in the C^∞ topology it is generic the set of potentials A such that for any n the function $\sum_{j=0}^{n-1} A \circ \sigma^j : (S^1)^{n+1} \to \mathbb{R}$ is a Morse function. In this case for each n there exist a finite number of points where the values of $(1/n) \sum_{j=0}^{n-1} A \circ \sigma^j$ are maximal. Moreover, there exists a positive number D such that for all n the number of critical points of $\sum_{j=0}^{n-1} A \circ \sigma^j$ is smaller than D^n.

One can consider for the function $\sum_{j=0}^{n-1} A \circ \sigma^j : (S^1)^{n+1} \to \mathbb{R}$ periodic boundary conditions on $(S^1)^n$. By adapting the proof of the above result we get that the Morse property, in this case, is also true. It follows from the above proposition that, generically on the potential A, the maximizing probability can be approximated by probabilities with support on periodic orbits (which are isolated in $(S^1)^{n+1}$).

Other references related to the topic are [**6**, **9**, **22**]. In a future work we will analyze questions related to zeta functions for generic potentials A (see [**52**] and [**44**]).

5. *An application to the non-compact case*

An interesting example of application of the above theory is the following: consider $M_0 = \{z_i, i \in \mathbb{N}\}$ an increasing infinite sequence of points in $[0, 1)$ and suppose that $z_\infty := 1 = \lim_{i\to\infty} z_i$. We will also suppose $z_1 = 0$. Therefore, each point of M_0 is isolated, and there is only one accumulating point $z_\infty = 1$. We consider the induced euclidean metric then $M = M_0 \cup \{1\}$ is a compact set. The state space M_0 can be identified with $\mathbb{N}$, and M has a special point $z_\infty = 1$ playing the role of the infinity. Let $\mathcal{B}_0 = M_0^{\mathbb{N}}$ and $\mathcal{B} = M^{\mathbb{N}}$. Note that $\mathcal{B}_0$ is not compact.

Some results in §5.1 for the shift with countable symbols (see [**19**, **55**]) can be recovered from our previous results as we will see. We will also study the limit, when the temperature goes to zero, of Gibbs states, and some results in Ergodic Optimization will be obtained. In particular we will show the existence of subactions, under some suitable hypothesis. We will get this last result via limit at temperature zero of eigenfunctions at positive temperature. Results in §5.2 for this setting appear in [**8**, **9**, **31**, **34**, **36**, **51**].

5.1. *Thermodynamic formalism.*

LEMMA 5. *Suppose that $A : \mathcal{B}_0 \to \mathbb{R}$ is a Hölder continuous potential. Then it can be extended as a Hölder continuous function $A : \mathcal{B} \to \mathbb{R}$.*

Proof. The extension is a consequence of the fact that any uniformly continuous function can be extended as a uniformly continuous function to the closure of its domain. It is easy to see that this extension is also Holder continuous. □

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Now let us fix an *a priori* measure $\nu := \sum_{i\in\mathbb{N}} p_i \delta_{z_i}$ on M (or M_0), where $p_i > 0$ and $\sum_{i\in\mathbb{N}} p_i = 1$. In fact, we have that $z_\infty = 1$ belongs to the support of μ, but is not an atom of μ. All other points of M (i.e. the points of M_0) are atoms for ν. On this way for each Hölder continuous potential $A : \mathcal{B}_0 \to \mathbb{R}$ we can consider the following transfer operator on $\mathcal{C}(\mathcal{B}_0)$:

$$\mathcal{L}_A(w)(x) := \int_M e^{A(ax)} w(ax)\, d\nu(a) = \sum_{i\in\mathbb{N}} e^{A(z_i x)} w(z_i x) p_i.$$

PROPOSITION 12. *Let $A : \mathcal{B}_0 \to \mathbb{R}$ be a Hölder potential. Then:*

(i) *there exists a positive number λ_A and a positive Hölder function $\psi_A : \mathcal{B}_0 \to \mathbb{R}$, such that, $\mathcal{L}_A \psi_A = \lambda_A \psi_A$.*

If we consider the normalized potential $\bar{A} = A + \log \psi_A - \log \psi_A \circ \sigma - \log \lambda_A$, then:

(ii) *there exists a unique fixed point μ_A for $\mathcal{L}^*_{\bar{A}}$, which is a σ-invariant probability measure on $\mathcal{B}_0$;*

(iii) *the measure*

$$\rho_A = \frac{1}{\psi_A} \mu_A$$

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

(iv) *for any Hölder function $w : \mathcal{B}_0 \to \mathbb{R}$, we have that, in the uniform convergence topology,*

$$\frac{\mathcal{L}^n_A(w)}{(\lambda_A)^n} \to \psi_A \int_{\mathcal{B}_0} w\, d\rho_A,$$

and

$$\mathcal{L}^n_{\bar{A}} \omega \to \int_{\mathcal{B}_0} \omega\, d\mu_A.$$

Proof. Using Lemma 5 we can extend $\mathcal{L}_A$ to $\mathcal{C}(\mathcal{B})$. From Theorem 1 we obtain $\lambda_A > 0$ and $\psi_A > 0$ Hölder continuous such that

$$\sum_i e^{A(z_i x)} \psi_A(z_i x) p_i = \lambda_A . \psi_A(x) \quad \text{for all } x \in \mathcal{B}.$$

In particular, the restriction $\psi_A : \mathcal{B}_0 \to \mathbb{R}$ will also satisfy the expression

$$\sum_i e^{A(z_i x)} \psi_A(z_i x) p_i = \lambda_A . \psi_A(x) \quad \text{for all } x \in \mathcal{B}_0.$$

This proves item (i).

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

In order to prove item (ii) we observe that from Theorem 2 there exists μ_A on $\mathcal{B}$ satisfying item (ii). We want to prove that $\mu_A(\mathcal{B} - \mathcal{B}_0) = 0$ or, equivalently, $\mu_A(\mathcal{B}_0) = 1$. On this way we only need to show that $\mu_A(\{x \in \mathcal{B} \mid x_1 = 1\}) = 0$, because for all $n \geq 1$, we have

$$\mu_A(\{x \in \mathcal{B} \mid x_n = 1\}) = \mu_A(\sigma^{-n+1}(\{x \in \mathcal{B} \mid x_1 = 1\})),$$

therefore, we will have

$$\mu_A(\mathcal{B} - \mathcal{B}_0) \leq \sum_{n=1}^{\infty} \mu_A(\{x \in \mathcal{B} \mid x_n = 1\}) = 0.$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

To prove that $\mu_A(\{x \in \mathcal{B} \mid x_1 = 1\}) = 0$, we fix $\varepsilon > 0$, and we consider a Hölder function w_ε, such that $\chi_{\{x\in\mathcal{B}|x_1=1\}} \le w_\varepsilon \le 1$ and $w_\varepsilon(x) = 0$, if $x_1 < 1 - \varepsilon$. Then, using Theorem 2(iii),

$$\begin{aligned}
&\mu_A(\{x \in \mathcal{B} \mid x_1 = 1\}) \\
&\le \int_{\mathcal{B}} w_\varepsilon \, d\mu_A = \lim_{n\to+\infty} \mathcal{L}_{\bar{A}}^n w_\varepsilon(0^\infty) \\
&= \lim_{n\to+\infty} \sum_{i_1 : z_{i_1} > 1-\varepsilon} p_{i_1} \sum_{i_2,\ldots,i_n} e^{S_{\bar{A}}^n(z_{i_1} z_{i_2}\cdots z_{i_n} 0^\infty)} w_\varepsilon(z_{i_1} z_{i_2} \cdots z_{i_n} 0^\infty) p_{i_2} \cdots p_{i_n} \\
&\le \lim_{n\to+\infty} \sum_{i_1 : z_{i_1} > 1-\varepsilon} p_{i_1} e^{\|\bar{A}\|} \sum_{i_2,\ldots,i_n} e^{S_{\bar{A}}^{n-1}(z_{i_2}\cdots z_{i_n} 0^\infty)} p_{i_2} \cdots p_{i_n} = \sum_{i_1 : z_{i_1} > 1-\varepsilon} p_{i_1} e^{\|\bar{A}\|},
\end{aligned}$$

where we used in the last equation the normalization property.

Now the claim follows easily when we use the fact that the *a priori* measure is supported on $[0, 1)$, which makes

$$\sum_{i_1 : z_{i_1} > 1-\varepsilon} p_{i_1} \to 0 \quad \text{when } \varepsilon \to 0.$$

The items (iii) and (iv) follow when we restrict to $\mathcal{B}_0$ the result of Theorem 2. □

Now let us compare this setting with some results contained in [**55**]. The operator $\mathcal{L}_A$ can be written as

$$\mathcal{L}_A(w)(x) = \sum_i e^{A(z_i x)} w(z_i x) p_i = \sum_i e^{A(z_i x) + \log(p_i)} w(z_i x),$$

that is, the classical Ruelle operator with potential $B := A + \log(P)$, where $P(y_1, y_2, y_3, \ldots) = P(y_1) = p_i$, if, $y_1 = z_i$. We denote this operator by L_B or $L_{A+\log(P)}$.

Clearly $(A + \log(P))(z_i, y_2, y_3, \ldots) \to -\infty$, when $i \to +\infty$, because $p_i \to 0$, when $i \to +\infty$. Furthermore, if we define

$$\mathrm{Var}_n(B) = \sup\{|B(x) - B(y)| \mid x_1 = y_1, \ldots, x_n = y_n\},$$

then, there exists $C > 0$, such that $\mathrm{Var}_n(B) \le C(1/2^{n\alpha})$, for any $n \ge 1$. This means that B is locally Hölder continuous (see [**55**]).

Define

$$Z_n(B, a) := \sum_{\substack{\sigma^n(y) = y \\ y_1 = a}} e^{S_n B(y)}.$$

PROPOSITION 13. *Fix $a \in M_0$, then there exists a constant M_a and an integer N_a, such that, for any $n > N_a$,*

$$\frac{Z_n(B, a)}{(\lambda_A)^n} \in [M_a^{-1}, M_a].$$

Proof. Let $x = a^\infty = (a, a, a, a, \ldots)$. When we apply Proposition 12(iv) for $w \equiv 1$ we get

$$\frac{\mathcal{L}_A^n(w)(x)}{(\lambda_A)^n} \to C > 0.$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Remember that we use the notation by $a^n = (a_n, \ldots, a_1) \in M_0^n$, then there exist $M_1 > 0$ and $N_a > 0$ such that for $n \geq N_a$:

$$\frac{1}{(\lambda_A)^{n+1}} \sum_{a^n} e^{S_n B(a^n x)} = \frac{1}{(\lambda_A)^{n+1}} \sum_{a^n} e^{S_n A(a^n x) + \log(P(a_n)\cdots P(a_1))} \in [M_1^{-1}, M_1].$$

Let y be the periodic point with period $n+1$ obtained by the successive concatenation of $(a, a_n, \ldots, a_1)$ and let $z = \sigma(y)$. We have

$$\begin{aligned}
&|S_n B(a^n x) - S_n B(z)| \\
&\quad = |S_n B(a_n, \ldots, a_1, a, a, a, \ldots) - S_n B(a_n, \ldots, a_1, a, a_n, a_{n-1}, \ldots)| \\
&\quad = |S_n A(a_n, \ldots, a_1, a, a, a, \ldots) - S_n A(a_n, \ldots, a_1, a, a_n, a_{n-1}, \ldots)| \\
&\quad \leq C\left(\frac{1}{2^\alpha} + \frac{1}{2^{2\alpha}} + \cdots + \frac{1}{2^{n\alpha}}\right) \cdot d(x, y)^\alpha.
\end{aligned}$$

Using the fact that $\mathcal{B}_0$ has finite diameter we obtain a constant $M_2 > 0$, such that

$$|S_n B(a^n x) - S_n B(z)| \leq M_2 \quad \text{for all } n \in \mathbb{N},\ a^n \in M_0^n.$$

Furthermore, using the property $\sigma^n(z) = y$, we conclude that $B(\sigma^n z) = A(\sigma^n(z)) + \log(P(a))$ is bounded independently of z. Therefore, there exists a constant $M_3 > 0$, such that

$$|S_n B(a^n x) - S_{(n+1)} B(z)| \leq M_3 \quad \text{for all } n \in \mathbb{N},\ a^n \in M_0^n.$$

Note that $S_{(n+1)} B(z) = S_{(n+1)} B(y)$. Then,

$$\sum_{a^n} e^{S_n B(a^n x) - M_3} \leq \sum_{\substack{\sigma^{n+1}(y) = y \\ y_1 = a}} e^{S_{(n+1)} B(y)} \leq \sum_{a^n} e^{S_n B(a^n x) + M_3}.$$

Therefore, when $n \geq N_a$, we get

$$\left(\frac{1}{(\lambda_A)^{n+1}} \sum_{\substack{\sigma^{n+1}(y) = y \\ y_1 = a}} e^{S_{(n+1)} B(y)}\right) \in [(M_1 e^{M_3})^{-1}, (M_1 e^{M_3})].$$

Choosing $M_a = M_1 e^{M_3}$ we conclude the proof. □

In this way, we can say that $B = A + \log(P)$ is positive recurrent (see [**55**, Definition 2]). Following [**55**, Theorem 4] we get a Ruelle–Perron–Frobenius Theorem (as in Theorem 12). It follows from the above proposition that λ_A is the Gurevic pressure of B (see [**55**, Definition 1]).

We would like to point out some differences on the topology considered in our setting with the classical one used in the theory of thermodynamic formalism with state space $\mathbb{N}$. The set $M_0^{\mathbb{N}}$ can be identified with $\mathbb{N}^{\mathbb{N}}$, but the metric space $M_0^{\mathbb{N}}$ is different from the metric space $\mathbb{N}^{\mathbb{N}}$ with the discrete product topology. Here, we consider a distance (induced in the subset $M_0 \cup \{z_\infty\} \subset [0, 1]$), such that, for any two points $x = (x_1, x_2, \ldots)$, $y = (y_1, y_2, \ldots) \in M_0^{\mathbb{N}}$

$$d(x, y) = \sum_{n \in \mathbb{N}} \frac{1}{2^n} d_{[0,1]}(x_n, y_n).$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

On the other hand, the metric considered in [**55**] is of the form: for two points $x,\ y \in \mathbb{N}^{\mathbb{N}}$

$$\widetilde{d}(x,\ y) = \frac{1}{2^n} \quad \text{if } x_1 = y_1,\ \ldots,\ x_{n-1} = y_{n-1},\ x_n \neq y_n.$$

Using that the diameter of [0, 1] is one, it follows that $d(x,\ y) \leq \widetilde{d}(x,\ y)$. In particular, any convergent sequence on the metric $\widetilde{d}$ is a convergent sequence on the metric d, and any continuous/Hölder function A for the metric d is a continuous/Hölder function for the metric $\widetilde{d}$. But the same is not true in the opposite direction. This is a subtle question. Results in [**55**] and here are obtained under slight different hypothesis. Anyway, in physical applications this is probably not very important.

Considering the dual space, it follows from the relation $d(x,\ y) \leq \widetilde{d}(x,\ y)$ that any open set for the metric d is an open set for the metric $\widetilde{d}$. Then, the Borel σ-algebra generated by d is contained in the Borel σ-algebra generated by $\widetilde{d}$; on the other hand, the cylinder sets [**55**] are closed sets for the metric d, therefore, they belong to the σ-algebra generated by d. In this way, the Borel σ-algebra generated by d or by $\widetilde{d}$ is the same.

5.2. *Ergodic optimization.* Once more we point out that the concepts of subaction and maximizing measures do not involve an *a priori* measure. In this way the statement of the next theorem does not use any *a priori* probability. On the other hand, some condition on A must be assumed in order to obtain positive results in ergodic optimization. For example, the potential $A : \mathcal{B}_0 \to \mathbb{R}$, given by

$$A(x) = -\sum_{i=1}^{\infty} \frac{1}{2^i}\, d(x_i,\ 1),$$

does not have maximizing measures on $\mathcal{B}_0$.

We need some hypothesis on A in such way it prevents the 'the mass from going to infinity'. Under some appropriate and natural conditions on A, we will obtain below the existence of calibrated subactions and maximizing measures. In the proof, we consider a limit involving a subsequence of eigenfunctions of $\mathcal{L}_{\beta A}$, when the temperature $1/\beta$ goes to zero.

THEOREM 6. *Suppose that $A : \mathcal{B}_0 \to \mathbb{R}$ is a Hölder continuous potential. Consider the Hölder continuous extension $A : \mathcal{B} \to \mathbb{R}$. If the extension satisfies*

$$A(x_1,\ \ldots,\ x_{n-1},\ 1,\ x_{n+1},\ x_{n+2},\ \ldots) < A(x_1,\ \ldots,\ x_{n-1},\ 0,\ x_{n+1},\ x_{n+2},\ \ldots) \tag{16}$$

for any $n \in \mathbb{N}$ and $x_i \in M$, then:

(i) *a has a calibrated subaction V on $\mathcal{B}_0$, that means for any $x \in \mathcal{B}_0$,*

$$0 = \max_{a \in M_0} (A(ax) + V(ax) - V(x) - m(A));$$

(ii) *any maximizing measure for A has support on $\mathcal{B}_0$.*

Proof. We define an *a priori* probability measure on M_0 by the expression

$$\nu = \sum_{i \in \mathbb{N}} p_i \cdot \delta_{z_i},$$

where $\sum_{i \in \mathbb{N}} p_i = 1$ and $p_i > 0$.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

CLAIM. *Denote by $\psi_{\beta A}$ the maximal eigenfunction for the Ruelle operator given by the potential βA and the a priori measure ν. Let V be the limit of some subsequence $(1/\beta_n)\log\psi_{\beta_n A}$. Then, we have:*

(i) $\psi_{\beta A}(1, x_1, x_2, x_3, \ldots) < \psi_{\beta A}(0, x_1, x_2, x_3, \ldots)$ *for all* $x \in \mathcal{B}$;

(ii) $V(1, x_1, x_2, x_3, \ldots) \leq V(0, x_1, x_2, x_3, \ldots)$ *for all* $x \in \mathcal{B}$.

Proof of (i). We know by item (iii) of Theorem 2 that

$$\frac{\mathcal{L}^n_{\beta A}(1)(x)}{(\lambda_{\beta A})^n} \to \psi_{\beta A}(x)\int_{\mathcal{B}} 1\, d\rho_{\beta A} \quad \text{for all } x \in \mathcal{B}.$$

Hence,

$$\rho_{\beta A}(\mathcal{B})\psi_{\beta A}(x) = \lim_{n\to\infty}\int_{M^n} e^{S_n\beta A(a^n x)-n\log\lambda_{\beta A}}\, d\nu^n(a^n) \quad \text{for all } x \in \mathcal{B}.$$

Suppose now that $z = (1, x_1, x_2, x_3, \ldots)$ and $y = (0, x_1, x_2, x_3, \ldots)$, with $x \in \mathcal{B}$, then, using (16), we get

$$\int_{M^n} e^{S_n\beta A(a^n z)-n\log\lambda_{\beta A}}\, d\nu^n(a^n) < \int_{M^n} e^{S_n\beta A(a^n y)-n\log\lambda_{\beta A}}\, d\nu^n(a^n),$$

which implies $\psi_{\beta A}(z) < \psi_{\beta A}(y)$.

Proof of item (ii). By hypothesis V satisfies

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

as the log function is monotone, we get $V(z) \leq V(y)$, which finishes the proof of the claim. □

Now let $R_- = A + V - V\circ\sigma - m(A)$, where V was defined above. We know, by the subaction equation, that $R_- \leq 0$. From (16) and the above claim, we get that for any $x \in \mathcal{B}_0$:

$$\begin{aligned} R_-(1x) &= A(1x) + V(1x) - V(x) - m(A) \\ &< A(0x) + V(0x) - V(x) - m(A) = R_-(0x) \leq 0. \end{aligned}$$

Using now the fact that V is a calibrated subaction (on $\mathcal{B}$) we conclude the proof of (i): the last inequality shows that

$$\max_{a\in M_0}(A(ax) + V(ax) - V(x) - m(A)) = \max_{a\in M}(A(ax) + V(ax) - V(x) - m(A)).$$

We point out that the fact that the extension has a maximizing measure is a consequence of the compactness of $\mathcal{B}$. In order to prove (ii) we will fix a $x \in \mathcal{B}\backslash\mathcal{B}_0$ and prove that x does not belong to the support of any maximizing measure.

Note that if $R_-(\sigma^{k-1}(x)) < 0$, then x does not belong to the support of the maximizing probability μ. Indeed,

$$\int_{\mathcal{B}} R_-\circ\sigma^{k-1}\, d\mu = \int_{\mathcal{B}} R_-\, d\mu = \int_{\mathcal{B}} A\, d\mu - m(A) = 0,$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

which, combined with the continuity of $R_- \leq 0$, proves that $R_- \circ \sigma^{k-1}$ vanishes at the support of μ. Now we will prove that $R_-(\sigma^{k-1}(x)) < 0$. So, let $k \in \mathbb{N}$ be such that $x_k = 1$ and $x_l < 1$, for all $1 \leq l < k$. Let $y \in \mathcal{B}$ be given by $y_i = x_i$, if $i \neq k$, and $y_k = 0$.

We have that $\sigma^{k-1}(x) = (1, x_{k+1}, x_{k+2}, \ldots)$ and $\sigma^{k-1}(y) = (0, x_{k+1}, x_{k+2}, \ldots)$, therefore

$$\begin{aligned} R_-(\sigma^{k-1}(x)) &= A(\sigma^{k-1}(x)) + V(\sigma^{k-1}(x)) - V(\sigma^k(x)) - m(A) \\ &< A(\sigma^{k-1}(y)) + V(\sigma^{k-1}(y)) - V(\sigma^k(y)) - m(A) = R_-(\sigma^{k-1}(y)) \leq 0, \end{aligned}$$

where we used above the hypothesis (16), item (ii) of the claim, and also $\sigma^k(x) = \sigma^k(y)$. □

In the proof of the above theorem we show the following result.

COROLLARY 2. *Given an a priori probability measure $\nu = \sum_{i=1}^{\infty} p_i \delta_{z_i}$, $p_i > 0$, under the hypothesis of the above theorem, then, there exists a subsequence $\{\beta_n\}$ and a Hölder continuous function $V : \mathcal{B}_0 \to \mathbb{R}$, such that*

$$\frac{1}{\beta_n} \log(\psi_{\beta_n A}) \to V$$

uniformly on $\mathcal{B}_0$. Furthermore, any possible limit V is a calibrated subaction for A on $\mathcal{B}_0$.

An example of potential satisfying the hypothesis of the above theorem is given by

$$A(x) = -d(x, 0^\infty)$$

where $0^\infty = (0, 0, 0, 0, \ldots)$. Note that in the claim of hypothesis (16) we can change 0 by any fixed z_i, and the result we get will be the same.

It is important to remark that when the temperature changes, then the operator varies in a different way of what happens in the classical sense: for a fixed $\beta > 0$, we have

$$\begin{aligned} \int_M e^{\beta\, A(ax)} w(ax)\, d\nu(a) &= \sum_i e^{\beta\, A(z_i x)} w(z_i x).p_i \\ &= \sum_i e^{\beta\, A(z_i x) + \log(P(z_i))} w(z_i x). \end{aligned}$$

Then, (in this work) the main eigenvalue and eigenfunction, respectively, $\lambda_\beta > 0$ and $\psi_\beta > 0$, are associated to $\beta(A + \log(P)/\beta)$ (in the setting of the classical Ruelle operator).

In this way we can think of the function $z_i \to \log(P(z_i))/\beta$ as a perturbation of the potential A, that goes to zero when $\beta \to \infty$ (but not uniformly).

6. *The differentiable structure and the involution kernel*

We consider in this section the XY model. This is the case where $M = S^1$, and the *a priori* measure is the Lebesgue measure on the circle. Here $(S^1)^{\mathbb{N}}$ has a differentiable structure.

We know that, in the case where the potential A is Holder, the eigenfunction ψ_A is also Holder and belongs to the same Holder class. The main result of this section is Theorem 7, where we prove that, under mild assumptions concerning the differentiability of A (see Definition 6), the associated eigenfunction ψ_A is differentiable in each coordinate x_j of x.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

In this setting, we point out that in [**45**] several questions were analyzed which involve differentiability for potentials which depend just on two coordinates. Here we consider more general potentials.

Let $\mathcal{B}^* = \{(\ldots, y_2, y_1) \in (S^1)^{\mathbb{N}}\}$, and we denote by the pair

$$(y|x) = (\ldots, y_2, y_1|x_1, x_2, \ldots),$$

the general element of $\hat{\mathcal{B}} := \mathcal{B}^* \times \mathcal{B} = (S^1)^{\mathbb{Z}}$, the natural extension of $\mathcal{B}$. Here we will follow the ideas of [**2**].

We denote by $\hat{\sigma}$ the shift on $\hat{\mathcal{B}}$, i.e.

$$\hat{\sigma}(\ldots, y_2, y_1|x_1, x_2, \ldots) = (\ldots, y_2, y_1, x_1|x_2, x_3, \ldots).$$

Definition 5. Let $A : \mathcal{B} \to \mathbb{R}$ be a continuous potential (considered as a function on $\hat{\mathcal{B}}$). A continuous function $W : \hat{\mathcal{B}} \to \mathbb{R}$ is called an involution kernel, if

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

depends only on the variable y.

The involution kernel is not unique.

Let us fix $x' \in \mathcal{B}$ and A a Hölder continuous potential, then we define

$$W(y|x) = \sum_{n \geq 1} A(y_n, \ldots, y_1, x_1, x_2, \ldots) - A(y_n, \ldots, y_1, x'_1, x'_2, \ldots). \tag{17}$$

An easy calculation shows that $W(y|x)$ is a involution kernel (see [**2**]).

The Ruelle–Perron operator $\mathcal{L}_A$ gives two important pieces of information: the eingenmeasure ρ_A and the eingenfunction ψ_A. As in [**2**], we can use the involution Kernel in order to obtain ψ_A, if we know the eigenmeasure of the Ruelle–Perron operator associated to $A^*(y)$ (see Proposition 14). We will show that the involution kernel allows one to differentiate ψ_A with respect to each coordinate x_j of x, using the expression of ψ_A given in Proposition 14 (see Theorem 7).

Let $A : \mathcal{B} \to \mathbb{R}$ be a Hölder continuous potential and $W : \hat{\mathcal{B}} \to \mathbb{R}$ an involution kernel, then for any $a \in S^1$, $x \in \mathcal{B}$ and $y \in \mathcal{B}^*$, we have

$$(A^* + W)(ya|x) = (A + W)(y|ax). \tag{18}$$

Questions related to ergodic transport theory and the involution kernel are analyzed in [**16**, **48**] and [**39**].

Remember that

$$m(A) = \sup_{\mu \text{ is } \sigma\text{-invariant}} \int_{\mathcal{B}} A d\mu,$$

and define

$$m(A^*) = \sup_{\mu \text{ is } \sigma^*\text{-invariant}} \int_{\mathcal{B}^*} A^* d\mu.$$

The next result is an adaptation to the present setting of a result in [**2**].

LEMMA 6. *Let $\mathcal{L}_A$ and $\mathcal{L}_{A^*}$ be the Ruelle operators defined on $\mathcal{B}$ and $\mathcal{B}^*$, and $W(y|x)$ an involution kernel.*

Then, for any $x \in \mathcal{B}$, $y \in \mathcal{B}^*$, and any function $f : \hat{\mathcal{B}} \to \mathbb{R}$

$$\mathcal{L}_{A^*}(f(\cdot|x)\, e^{W(\cdot|x)})(y) = \mathcal{L}_A(f \circ \hat{\sigma}(y|\cdot)\, e^{W(y|\cdot)})(x). \tag{19}$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Proof. Under our notation we write $A(y|x) = A(x)$ and $A^*(y|x) = A^*(y)$. Consider $x \in \mathcal{B}$, $y \in \mathcal{B}^*$ fixed, then by the definition of $\mathcal{L}_A$ and $\mathcal{L}_{A^*}$, and equation (18), we obtain

$$\begin{aligned}\mathcal{L}_{A^*}(f(\cdot|x)\ e^{W(\cdot|x)})(y) &= \int_{S^1} f(ya|x)\, e^{(A^*(ya)+W(ya|x))}\, d\nu(a)\\ &= \int_{S^1} f\circ\hat{\sigma}(y|ax)\, e^{(A(ax)+W(y|ax))}\, d\nu(a)\\ &= \mathcal{L}_A(f\circ\hat{\sigma}(y|\cdot)\, e^{W(y|\cdot)})(x). \qquad \square\end{aligned}$$

Let ρ_A and ρ_{A^*} the eigenmeasures for $\mathcal{L}_A^*$ and $\mathcal{L}_{A^*}^*$, given in Theorem 2. Suppose c is such that $\int\int_{\mathcal{B}\times\mathcal{B}^*} e^{W(y|x)-c}\, d\rho_{A^*}(y)\, d\rho_A(x) = 1$.

PROPOSITION 14. *Suppose* $K(y|x) = e^{W(y|x)-c}$. *Then,*

$$d\hat{\mu}_A = K(y|x)\, d\rho_{A^*}(y)\, d\rho_A(x)$$

is invariant for $\hat{\sigma}$ *and is the natural extension of the Gibbs measure* μ_A.

The function $\psi_A(x) = \int_{\mathcal{B}^*} K(y|x)\, d\rho_{A^*}(y)$ *is the main eigenfunction for* $\mathcal{L}_A$, *and the function* $\psi_{A^*}(y) = \int_{\mathcal{B}} K(y|x)\, d\rho_A(x)$ *is the main eigenfunction for* $\mathcal{L}_{A^*}$. *Furthermore,* $\lambda_A = \lambda_{A^*}$.

Proof. We denote by $K(y|x) = e^{W(y|x)-c}$, and we define a positive function ψ, by the expression $\psi(x) = \int_{\mathcal{B}^*} K(y|x)\, d\rho_{A^*}(y)$. In order to prove that ψ is an eigenfunction for $\mathcal{L}_A$, we remember that $\mathcal{L}_{A^*}^*(\rho_{A^*}) = \lambda_{A^*}\rho_{A^*}$, hence

$$\begin{aligned}\psi(x) &= \int_{\mathcal{B}^*} K(y|x)\, d\left(\frac{1}{\lambda_{A^*}}\mathcal{L}_A^*(\rho_{A^*})\right)(y) = \int_{\mathcal{B}^*} \frac{1}{\lambda_{A^*}}\mathcal{L}_{A^*}(K(\cdot|x))(y)\, d\rho_{A^*}(y)\\ &= \int_{\mathcal{B}^*} \frac{1}{\lambda_{A^*}}\mathcal{L}_A(K(y|\cdot))(x)\, d\rho_{A^*}(y) = \frac{1}{\lambda_{A^*}}\mathcal{L}_A(\psi)(x),\end{aligned}$$

where in the third equality we have used (19) with $f = 1$. This means that ψ is a positive eigenfunction for $\mathcal{L}_A$, now using Proposition 1 we get that $\psi = \psi_A$ and $\lambda_{A^*} = \lambda_A$. The proof for the case of ψ_A^* is similar.

By the same arguments used above, for any bounded Borel $f : \mathcal{B}^* \times \mathcal{B} \to \mathbb{R}$, we have

$$\begin{aligned}&\int_{\mathcal{B}}\int_{\mathcal{B}^*} f\circ\hat{\sigma}(y|x)K(y|x)\, d\rho_{A^*}(y)\, d\rho_A(x)\\ &\quad= \int_{\mathcal{B}^*} d\rho_{A^*}(y)\int_{\mathcal{B}} \frac{1}{\lambda_A}\mathcal{L}_A(f\circ\hat{\sigma}(y|\cdot)K(y|\cdot))(x)\, d\rho_A(x)\\ &\quad= \int_{\mathcal{B}} d\rho_A(x)\int_{\mathcal{B}^*} \frac{1}{\lambda_{A^*}}\mathcal{L}_{A^*}(f(\cdot|x)K(\cdot|x))\, d\rho_{A^*}(y)\\ &\quad= \int_{\mathcal{B}}\int_{\mathcal{B}^*} f(y|x)\, K(y|x)\, d\rho_{A^*}(y)\, d\rho_A(x),\end{aligned}$$

hence $d\hat{\mu}_A = K(y|x)\, d\rho_{A^*}(y)\, d\rho_A(x)$ is invariant for $\hat{\sigma}$.

Finally, let us prove that $\hat{\mu}_A$ is the natural extension of μ_A. Given a function $f(x)$ we get that

$$\begin{aligned}\int_{\mathcal{B}}\int_{\mathcal{B}^*} f(x)\, d\hat{\mu}_A(y,x) &= \int_{\mathcal{B}}\int_{\mathcal{B}^*} f(x)\, K(y|x)\, d\rho_{A^*}(y)\, d\rho_A(x)\\ &= \int_{\mathcal{B}} f(x)\psi_A(x)\, d\rho_A(x) = \int_{\mathcal{B}} f(x)\, d\mu_A(x).\end{aligned}$$

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Therefore, the measure $d\hat{\mu}_A(y, x) = K(y|x)\, d\rho_{A^*}(y)\, d\rho_A(x)$ projects onto μ_A and μ_{A^*} (by the same arguments). The probability $\hat{\mu}_A$ is therefore the natural extension of μ_A. □

Remark. Note that, as $\lambda_{\beta A} = \lambda_{\beta A^*}$, we have that $m(A) = \lim_{\beta\to\infty}(1/\beta)\log\lambda_{\beta A} = m(A^*)$.

Definition 6. Suppose that A is Lipschitz continuous and satisfies both of the following conditions:

(a) A is differentiable in each coordinate x_j of $x \in \mathcal{B}$;

(b) given $\varepsilon > 0$, there exists $H_\varepsilon > 0$, such that, for all x, if $|h| < H_\varepsilon$, then

$$\left|\frac{A(x + he_j) - A(x)}{h} - D_jA(x)\right| \le \frac{\varepsilon}{2^j} \quad \text{for all } j \in \mathbb{N}, \tag{20}$$

where $D_jA(x)$ denote the derivative of A with respect to the jth coordinate.

We will denote the class of such potentials by $\mathcal{D}$.

The potential $A(x) = \sum_{n=1}^{\infty}(1/2^n)\sin(x_n + 1/2^n)$ belongs to the class $\mathcal{D}$. Moreover, any potential which depends on finite coordinates and is of class C^2 belongs to $\mathcal{D}$.

PROPOSITION 15. *Suppose that A belongs to the class $\mathcal{D}$. Given an involution kernel W we have that, for any j,*

$$\frac{\partial}{\partial x_j}W(y|x) = \sum_{n\ge1} D_{n+j}A(y_n, \ldots, y_1, x_1, x_2, \ldots).$$

Proof. Let us first prove that the sum in the right-hand side is convergent. Indeed, using that A is Lipschitz, there exist $K > 0$ such that

$$|A(x) - A(\tilde{x})| < Kd(x, \tilde{x}).$$

If we denote $x_h^j = x + he_j$ and $y^nx = (y_n, \ldots, y_1, x_1, x_2, \ldots)$, then we have

$$\left|\frac{A(y^nx + he_{j+n}) - A(y^nx)}{h}\right| \le \frac{K}{h}\frac{d_{S^1}(x_j + h, x_j)}{2^{n+j}} = \frac{K}{2^{n+j}}.$$

Using that A belongs to $\mathcal{D}$, we get that, given $\varepsilon > 0$, there exists $H_\varepsilon > 0$ such that, for each $|h| < H_\varepsilon$, we have

$$|D_{n+j}A(y^nx)| \le \left|\frac{A(y^nx_h^j) - A(y^nx)}{h}\right| + \frac{\varepsilon}{2^{n+j}} \le \frac{K}{2^{n+j}} + \frac{\varepsilon}{2^{n+j}},$$

which implies

$$\sum_{n\ge1}|D_{n+j}A(y^nx)| < \frac{K+\varepsilon}{2^j} < \infty.$$

Now, we will prove the proposition:

$$\begin{aligned}
&\left|\frac{W(y|(x + he_j)) - W(y|x)}{h} - \sum_{n\ge1} D_{n+j}A(y^nx)\right| \\
&= \left|\frac{1}{h}\left(\sum_{n\ge1}A(y^nx + he_{j+n}) - A(y^nx)\right) - \sum_{n\ge1}D_{n+j}A(y^nx)\right| \\
&= \left|\sum_{n\ge1}\left(\frac{A(y^nx + he_{j+n}) - A(y^nx)}{h} - D_{n+j}A(y^nx)\right)\right| \le \sum_{n\ge1}\frac{\varepsilon}{2^{n+j}} = \frac{\varepsilon}{2^j},
\end{aligned}$$

for all $|h| \le H_\varepsilon$, as A belongs to $\mathcal{D}$. □

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

From the final part of the last proof we have that, for all $|h| \le H_\varepsilon$, and for all x and y,

$$\left| \frac{W(y|(x+he_j)) - W(y|x)}{h} - \frac{\partial W(y|x)}{\partial x_j} \right| \le \frac{\varepsilon}{2^j}. \tag{21}$$

THEOREM 7. *Let $\psi_A(x) = \int_{\mathcal{B}^*} e^{W(y|x)-c}\, d\rho_{A^*}(y)$, and suppose A belongs to the class $\mathcal{D}$. Then, the eigenfunction ψ_A is differentiable in each coordinate x_j. Moreover,*

$$\frac{\partial}{\partial x_j}\psi_A(x) = \int_{\mathcal{B}^*} e^{W(y|x)-c} \sum_{n\ge 1} D_{n+j} A(y_n, \ldots, y_1, x_1, x_2, \ldots)\, d\rho_{A^*}(y).$$

Proof. Consider $j \in \mathbb{N}$. We have

$$\begin{aligned}
&\left| \frac{\psi_A(x+he_j) - \psi_A(x)}{h} - \int_{\mathcal{B}^*} e^{W(y|x)-c} \frac{\partial W(y|x)}{\partial x_j}\, d\rho_{A^*}(y) \right| \\
&= \left| \int_{\mathcal{B}^*} \left(\frac{e^{W(y|x+he_j)-c} - e^{W(y|x)-c}}{h} - e^{W(y|x)-c} \frac{\partial W(y|x)}{\partial x_j} \right) d\rho_{A^*}(y) \right| \\
&= \left| \int_{\mathcal{B}^*} e^{W(y|x)-c} \left(\frac{e^{W(y|x+he_j)-W(y|x)} - 1}{h} - \frac{\partial W(y|x)}{\partial x_j} \right) d\rho_{A^*}(y) \right| \\
&\le \left| \int_{\mathcal{B}^*} e^{W(y|x)-c} \left(\frac{e^{W(y|x+he_j)-W(y|x)} - 1}{h} - \frac{W(y|x+he_j) - W(y|x)}{h} \right) d\rho_{A^*}(y) \right| \\
&\quad + \left| \int_{\mathcal{B}^*} e^{W(y|x)-c} \left(\frac{W(y|x+he_j) - W(y|x)}{h} - \frac{\partial W(y|x)}{\partial x_j} \right) d\rho_{A^*}(y) \right|.
\end{aligned}$$

Now, observe that the second integral above goes to zero when $h \to 0$, as a consequence of (21). The first integral also goes to zero when $h \to 0$ because, using the fact that

$$\frac{e^{ah} - 1}{h} - a = \sum_{k\ge 2} \frac{a^k h^{k-1}}{k!},$$

with $a = (W(y|x+he_j) - W(y|x))/h$, we have

$$\begin{aligned}
&\left| \frac{e^{W(y|x+he_j)-W(y|x)} - 1}{h} - \frac{W(y|x+he_j) - W(y|x)}{h} \right| \\
&= \left| \sum_{k\ge 2} \frac{(W(y|x+he_j) - W(y|x))^k}{k!h} \right| \\
&= \left| \sum_{k\ge 2} \frac{(W(y|x+he_j) - W(y|x))^{k-1}}{k!} \frac{(W(y|x+he_j) - W(y|x))}{h} \right| \\
&\to 0 \quad \text{when } h \to 0.
\end{aligned}$$

In the last expression we used the uniform continuity of W and also (21). □

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

Remark. In the case where A depends only on the two first coordinates, we have that (see (13))

$$\psi_A(x_1) = \frac{1}{\lambda_A} \int_{S^1} e^{A(y_1,x_1)} \psi_A(y_1)\, d\nu(y_1).$$

Hence, ψ_A satisfies the equation

$$\frac{\partial}{\partial x_1}\psi_A(x_1) = \frac{1}{\lambda_A} \int_{S^1} e^{A(y_1,x_1)} D_2A(y_1, x_1)\, \psi_A(y_1)\, d\nu(y_1).$$

References

[**1**] M. Asaoka, T. Fukaya, K. Mitsui and M. Tsukamoto. Growth of critical points in one-dimensional lattice systems. *Preprint*, 2012.

[**2**] A. Baraviera, A. O. Lopes and P. Thieullen. A large deviation principle for equilibrium states of Hölder potentials: the zero temperature case. *Stoch. Dynam.* **6** (2006), 77–96.

[**3**] A. T. Baraviera, L. Cioletti, A. O. Lopes, J. Mohr and R. R. Souza. On the general one-dimensional XY model: positive and zero temperature, selection and non-selection. *Rev. Math. Phys.* **23**(10) (2011), 1063–1113, 82Bxx.

[**4**] A. T. Baraviera, A. O. Lopes and J. K. Mengue. On the selection of subaction and measure for a subclass of potentials defined by P. Walters. *Ergod. Th. & Dynam. Sys.* **33**(5) (2013), 1338–1362.

[**5**] P. Bernard and G. Contreras. A generic property of families of Lagrangian systems. *Ann. of Math.* **167**(3) (2008), 1099–1108.

[**6**] M. Bertelson. Topological invariant for discrete group actions. *Lett. Math. Phys.* **62** (2004), 147–156.

[**7**] M. Bertelson and M. Gromov. Dynamical Morse entropy. *Modern Dynamical Systems and Applications*. Cambridge University Press, Cambridge, 2004, pp. 27–44.

[**8**] R. Bissacot and R. Freire Jr. On the existence of maximizing measures for irreducible countable Markov shifts: a dynamical proof. *Ergod. Th. & Dynam. Sys.* to appear.

[**9**] R. Bissacot and E. Garibaldi. Weak KAM methods and ergodic optimal problems for countable Markov shifts. *Bull. Braz. Math. Soc.* **41**(3) (2010), 321–338.

[**10**] T. Bousch. La condition de Walters. *Ann. Sci. Éc. Norm. Super. (4)* **34**(2) (2001), 287–311.

[**11**] J.R. Chazottes, J.M. Gambaudo and E. Ulgade. Zero-temperature limit of one dimensional Gibbs states via renormalization: the case of locally constant potentials. *Ergod. Th. & Dynam. Sys.* **31**(4) (2011), 1109–1161.

[**12**] J. R. Chazottes and M. Hochman. On the zero-temperature limit of Gibbs states. *Commun. Math. Phys.* **297**(N1) (2010), 265–281.

[**13**] W. Chou and R. J. Duffin. An additive eigenvalue problem of physics related to linear programming. *Adv. Appl. Math.* **8** (1987), 486–498.

[**14**] W. Chou and R. Griffiths. Ground states of one-dimensional systems using effective potentials. *Phys. Rev. B* **34**(9) (1986), 6219–6234.

[**15**] G. Contreras and R. Iturriaga. *Global Minimizers of Autonomous Lagrangians (22° Colóquio Brasileiro de Matemática)*. IMPA, Rio de Janeiro, 1999.

[**16**] G. Contreras, A. Lopes and E. Oliveira. Ergodic transport theory, periodic maximizing probabilities and the twist condition. *Modeling, Optimization, Dynamics and Bioeconomy (Proceedings in Mathematics)*. Eds. D. Zilberman and A. Pinto. Springer, to appear.

[**17**] G. Contreras, A. O. Lopes and Ph. Thieullen. Lyapunov minimizing measures for expanding maps of the circle. *Ergod. Th. & Dynam. Sys.* **21** (2001), 1379–1409.

[**18**] J. P. Conze and Y. Guivarc'h. Croissance des sommes ergodiques et principe variationnel. Manuscript circa 1993.

[**19**] Y. Daon. Bernoullicity of equilibrium measures on countable Markov shifts. *Discrete Contin. Dyn. Syst.* **33**(9) (2013), 4003–4015.

[**20**] R. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer, Berlin, 2005.

[**21**] A. Fathi. Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens. *C. R. Acad. Sci. Sér. I Math.* **324** (1997), 1043–1046.

[**22**] T. Fukaya and M. Tsukamoto. Asymptotic distribution of critical values. *Geom. Dedicata* **143** (2009), 63–67.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

[**23**] Y. Fukui and M. Horiguchi. One-dimensional Chiral XY Model at finite temperature. *Interdisciplinary Inf. Sci.* **1**(2) (1995), 133–149.
[**24**] G. Gallavotti. *Statistical Mechanics: A Short Treatise*. Springer, Berlin, 2010.
[**25**] E. Garibaldi and A. O. Lopes. On Aubry–Mather theory for symbolic dynamics. *Ergod. Th. & Dynam. Sys.* **28**(3) (2008), 791–815.
[**26**] E. Garibaldi and A. O. Lopes. The effective potential and transshipment in thermodynamic formalism at temperature zero. *Stoch. Dyn.* **13**(1) (2013),1250009.
[**27**] H.-O. Georgii. *Gibbs Measures and Phase Transitions*. de Gruyter, Berlin, 1988.
[**28**] D. A. Gomes, A. O. Lopes and J. Mohr. The Mather measure and a large deviation principle for the entropy penalized method. *Commun. Contemp. Math.* **13**(2) (2011), 235–268.
[**29**] D. A. Gomes and E. Valdinoci. Entropy penalization methods for Hamilton–Jacobi equations. *Adv. Math.* **215**(1) (2007), 94–152.
[**30**] M. Gromov. Singularities, expanders and topology of maps. Part 2: From combinatorics to topology via algebraic isoperimetry. *Geom. Funct. Anal.* **20**(2) (2010), 416–526.
[**31**] G. Iommi. Ergodic optimization for renewal type shifts. *Monatsh. Math.* **150**(2) (2007), 91–95.
[**32**] R. B. Israel. *Convexity in the Theory of Lattice Gases*. Princeton University Press, Princeton, NJ, 1979.
[**33**] O. Jenkinson. Ergodic optimization. *Discrete Contin. Dyn. Syst. Ser. A* **15** (2006), 197–224.
[**34**] O. Jenkinson, R. D. Mauldin and M. Urbanski. Zero temperature limits of Gibbs-equilibrium states for countable alphabet subshifts of finite type. *J. Stat. Phys.* **119**(3–4) (2005), 765–776.
[**35**] G. Keller. *Gibbs States in Ergodic Theory*. Cambridge University Press, Cambridge, 1998.
[**36**] T. Kempton. Zero temperature limits of Gibbs equilibrium states for countable markov shifts. *J. Stat. Phys.* **143** (2011), 795–806.
[**37**] O. Lanford. Entropy and equilibrium states in classical statistical mechanics. *Statistical Mechanics and Mathematical Problems (Battelle Rencontres, Seattle, WA, 1971) (Lecture Notes in Physics, 20)*. Springer, Berlin, 1973.
[**38**] R. Leplaideur. A dynamical proof for convergence of Gibbs measures at temperature zero. *Nonlinearity* **18**(6) (2005), 2847–2880.
[**39**] A. Lopes and J. Mengue. Duality theorems in ergodic transport. *J. Stat. Phys.* **149**(5) (2012), 921–942.
[**40**] A. Lopes and J. Mengue. Selection of measure and a large deviation principle for the general one-dimensional XY model. *Dyn. Syst.* **29**(1) (2014), 24–39.
[**41**] A. Lopes and P. Thieullen. Subactions for Anosov flows. *Ergod. Th. & Dynam. Sys.* **25**(2) (2005), 605–628.
[**42**] A. Lopes and P. Thieullen. Subactions for Anosov diffeomorphisms. *Asterisque Geom. Dynam. (II)* **287** (2013), 135–146.
[**43**] A. O. Lopes. An analogy of charge distribution on Julia sets with the Brownian motion. *J. Math. Phys.* **30**(9) (1989), 2120–2124.
[**44**] A. O. Lopes and J. Mengue. Zeta measures and thermodynamic formalism for temperature zero. *Bull. Braz. Math. Soc.* **41**(3) (2010), 449–480.
[**45**] A. O. Lopes, J. Mohr, R. Souza and Ph. Thieullen. Negative entropy, zero temperature and stationary Markov chains on the interval. *Bull. Braz. Math. Soc.* **40** (2009), 1–52.
[**46**] A. O. Lopes, A. Neumann and Ph. Thieullen. A thermodynamic formalism for continuous time Markov chains with values on the Bernoulli space: entropy, pressure and large deviations. *J. Stat. Phys.* **152**(5) (2013), 894–933.
[**47**] A. O. Lopes and E. Oliveira. Entropy and variational principles for holonomic probabilities of IFS. *Discrete Cont. Dyn. Syst. A* **23**(3) (2009), 937–955.
[**48**] A. O. Lopes, E. R. Oliveira and Ph. Thieullen. The dual potential, the involution kernel and transport in ergodic optimization. *Preprint*, 2008.
[**49**] R. Mañé. *The Hausdorff Dimension of Invariant Probabilities of Rational Maps (Lecture Notes in Mathematics, 1331)*. Springer, Berlin, 1988, pp. 86–117.
[**50**] D. H. Mayer. *The Ruelle–Araki Transfer Operator in Classical Statistical Mechanics (Lecture Notes in Physics, 123)*. Springer, Berlin, 1980.
[**51**] I. D. Morris. Entropy for zero-temperature limits of Gibbs-equilibrium states for countable-alphabet subshifts of finite type. *J. Stat. Phys.* **126**(2) (2007), 315–324.
[**52**] W. Parry and M. Pollicott. *Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics (Astérisque, 187–188)*. Societé Mathématique de France, Paris, 1990.
[**53**] M. Pollicott and M. Yuri. *Dynamical Systems and Ergodic Theory*. Cambridge University Press, Cambridge, 1998.
[**54**] D. Ruelle. *Thermodynamic Formalism*, 2nd edn. Cambridge University Press, Cambridge, 2004.
[**55**] O. Sarig. Thermodynamic formalism for countable Markov shifts. *Ergod. Th. & Dynam. Sys.* **19** (1999), 1565–1593.
[**56**] O. Sarig. Thermodynamic formalism for topological Markov shifts. *Lecture Notes*. Penn State, 2009.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press

[**57**] B. Simon. *The Statistical Mechanics of Lattice Gases*. Princeton University Press, Princeton, NJ, 1993.
[**58**] R. R. Souza. Sub-actions for weakly hyperbolic one-dimensional systems. *Dynam. Syst.* **2**(18) (2003), 165–179.
[**59**] C. Thompson. Infinite-spin Ising model in one dimension. *J. Math. Phys.* **9**(2) (1968), 241–245.
[**60**] A. C. D. van Enter, R. Fernandez and A. D. Sokal. Regularity properties and pathologies of position-space renormalization-group transformations: scope and limitations of Gibbsian theory. *J. Stat. Phys.* **72**(5/6) (1993), 879–1187.
[**61**] A. C. D. van Enter and W. M. Ruszel. Chaotic temperature dependence at zero temperature. *J. Stat. Phys.* **127**(3) (2007), 567–573.

https://doi.org/10.1017/etds.2014.15 Published online by Cambridge University Press