Entropy, Pressure and Duality for Gibbs plans in Ergodic Transport

A. O. Lopes, J. K. Mengue, J. Mohr and R. R. Souza

March 2, 2014

Instituto de Matemática, UFRGS - Porto Alegre, Brasil

Let X be a finite set and $\Omega = \{1, ..., d\}^{\mathbb{N}}$ be the Bernoulli space. Denote by σ the shift map acting on Ω. We consider a fixed Lipschitz cost (or potential) function $c : X \times \Omega \to \mathbb{R}$ and an associated Ruelle operator. We introduce the concept of Gibbs plan for c, which is a probability on $X \times \Omega$ such that the y marginal is invariant. Moreover, we define entropy, pressure and equilibrium plans. The study of equilibrium plans can be seen as a generalization of the equilibrium probability problem where the concept of entropy for plans is introduced. We show that an equilibrium plan is a Gibbs plan.

For a fixed probability μ on X with supp(μ)$= X$, define $\Pi(\mu, \sigma)$ as the set of all Borel probabilities $\pi \in P(X \times \Omega)$ such that the x-marginal of π is μ and the y-marginal of π is σ-invariant. We also investigate the pressure problem over $\Pi(\mu, \sigma)$, that is with constraint μ. Our main result is a Kantorovich duality Theorem on this setting. The pressure without constraint plays an important role in the establishment of the notion of admissible pair. Basically we want to transform a problem of pressure with a constraint μ on X in a problem of pressure without constraint. Finally, given a parameter β, which plays the role of the inverse of temperature, we consider equilibrium plans for βc and its limit π_∞, when $\beta \to \infty$, which is also known as ground state. We compare this with other previous results on Ergodic Transport at temperature zero.

1 Introduction

Kantorovich duality is a general theoretical tool for solving problems. The practical problems where one can get explicit solutions in Classical Transport Theory are in general obtained via duality techniques and the complementary slackness condition. Here we investigate this kind of result in a dynamical setting associated to a generalization of Thermodynamic Formalism which fits well the Transport setting.

We want to show that:

1) the principle of maximizing pressure in Thermodynamic Formalism corresponds in the more general dynamical setting to Kantorovich Duality (section 3).

2) the slackness condition is given by a simple equation which uses a generalized Ruelle operator (presented in section 2).

Let X be a finite set and $\Omega = \{1, ..., d\}^{\mathbb{N}}$ the Bernoulli space with the usual metric[1]. We denote by σ the shift map acting on Ω, by $P(X)$ the set of probabilities over X and by $P(\Omega)$ the set of probabilities acting on the Borel sigma-algebra $\mathcal{B}(\Omega)$. Let $C(X)$ be the set of functions from X to $\mathbb{R}$ and $C(\Omega)$ be the set of continuous functions from Ω to $\mathbb{R}$. We denote by (x, y) the variables on the space $X \times \Omega$.

A Borel probability π on $X \times \Omega$ is called a plan. For a fixed $\mu \in P(X)$ such that supp(μ)$= X$, define $\Pi(\mu, \sigma)$ as the set of all plans satisfying

$$\begin{cases} \int_{X\times\Omega} f(x)\, d\pi(x,y) = \int_X f(x)\, d\mu(x) & \text{for any } f \in C(X), \\ \int_{X\times\Omega} g(y)\, d\pi(x,y) = \int_{X\times\Omega} g(\sigma(y))\, d\pi(x,y) & \text{for any } g \in C(\Omega), \end{cases} \tag{1}$$

which means, the set of probabilities π such that the x-marginal of π is the fixed probability $\mu \in P(X)$ and the y-marginal of π is σ-invariant. Define $\Pi(\cdot, \sigma)$ as the set of plans such that its y-projection is σ-invariant.

We will introduce the entropy $H(\pi)$ of a plan $\pi \in \Pi(\cdot, \sigma)$ and the pressure $P(c)$ of a Lipschitz cost $c : X \times \Omega \to \mathbb{R}$. More precisely

$$P(c) = \sup_{\pi \in \Pi(\cdot,\sigma)} \int_{X\times\Omega} c\, d\pi \,+\, H(\pi).$$

We will show in section 3 the following result which is the **Kantorovich duality** in the (Transport) Thermodynamic Formalism setting:

Variational Principle

$$\inf_{\varphi: P(c-\varphi)=0} \int_X \varphi(x)\, d\mu = \sup_{\pi \in \Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi \,+\, H(\pi).$$

The infimum and supremum will be attained by unique elements $\tilde{\varphi}$ and $\tilde{\pi}$.

If X has only one element we get the classical Thermodynamical Formalism setting. The function $\tilde{\varphi}$ plays in some sense the role of the main

[1] $d(z, y) = \frac{1}{2^n}$ when $z = (z_0, z_1, ...)$, $y = (y_0, y_1, ...)$, $z_n \neq y_n$ and $z_j = y_j$, $j < n$.

eigenvalue of the transfer operator. If X has only one element, $\tilde{\varphi}$ coincides with $\log(\lambda)$, where λ is the main eigenvalue of the classical Ruelle Operator [9].

When X has more than one point the main issue in the theory is to be able to characterize the optimal $\tilde{\pi}$. The main point here is to transform a problem of pressure with a constraint μ, namely

$$\sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi \,+\, H(\pi),$$

in a problem of pressure without constraint

$$P(\tilde{c}) = \sup_{\pi\in\Pi(\cdot,\sigma)} \int_{X\times\Omega} \tilde{c}\, d\pi \,+\, H(\pi).$$

The $\tilde{\varphi}$ helps to do that because $\tilde{c} = c - \tilde{\varphi}$. This can be achieved via the duality of Theorem 16. We will present a worked example (see example 3 on page 18) to illustrate how one can get explicit solutions of the above mentioned problem.

Section 2 generalize to plans what is known in Thermodynamic Formalism. Most of the proofs are a kind of standard generalization of the classical setting. In section 2, μ is not fixed and the results are about plans and not exactly about transport (optimal plans with a fixed μ as marginal). We need this part in section 3 where μ is fixed and our main result is proved.

Now we will present some motivations from results about Ergodic Optimization and transport contained in [7].

Consider a Lipschitz function $c : X \times \Omega \to \mathbb{R}$. We denote by $\pi_{\text{opt}(c)}$ any optimal plan in $\Pi(\mu, \sigma)$, which means that $\pi_{\text{opt}(c)}$ satisfies

$$\sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi = \int_{X\times\Omega} c\, d\pi_{\text{opt}(c)}.$$

It is well known that the optimal plan $\pi_{\text{opt}(c)}$ may not be unique. Associated to the problem of finding an optimal plan $\pi_{\text{opt}(c)}$ and determining $\sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi$ it is natural to consider the dual problem: if c is Lipschitz continuous, there exist Lipschitz functions $V : \Omega \to \mathbb{R}$ and $m : X \to \mathbb{R}$ such that

$$c(x, y) + V(y) - V(\sigma(y)) - m(x) \leq 0\,, \tag{2}$$

and

$$\int_{X\times\Omega} c(x, y)\, d\pi_{\text{opt}(c)} = \int_X m(x)\, d\mu. \tag{3}$$

This result was proved in [7]. Results of such nature are part of what is called Ergodic Transport Theory. We point out that in [7] it is considered a minimization cost problem and here we consider a maximization cost problem. There is no conceptual difference in both settings.

If V satisfies (2) and (3), we say that V is a subaction associated to c and μ. We say that V is a calibrated subaction if for each given $y \in \Omega$ there exist $x \in X$ and a pre-image w of y, such that

$$c(x,w) + V(w) - V(y) - m(x) = 0.$$

The problem above is very much related with the questions which are usually considered in Ergodic Optimization. Indeed, if we consider the particular case where X has a unique point x, then μ will be the Dirac measure δ_x, and any plan in $\Pi(\delta_x, \sigma)$ is a direct product $\delta_x \times \nu$, where ν is a invariant measure. In this case, we can identify $X \times \Omega$ with Ω, $\Pi(\delta_x, \sigma)$ with the set of invariant measures on Ω, $c(x,y)$ with a potential $A(y)$, and in this case m will be constant and equal to the number

$$m(A) = \sup_{\nu\text{-invariant}} \int_\Omega A\, d\nu,$$

which is called sometimes the maximal value of A. In the Ergodic Transport setting we have that m is a function on x and also get the validity of the equations (2) and (3).

When X has two or more points the function $m(x)$ is strongly related with the initial fixed μ in the following sense: If c and m satisfy (2) and (3) for some V, then

$$\sup_{P\in\Pi(\cdot,\sigma)} \int_{X\times\Omega} c(x,y) - m(x)\, dP = \sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c(x,y) - m(x)\, d\pi = 0.$$

It is well known that Ergodic Optimization is related with Thermodynamic Formalism via the zero temperature limit (see [2]). In this way it is natural to investigate the possible generalizations of the transfer operator (also called Ruelle operator), and other properties which appear in Thermodynamic Formalism for the Ergodic Transport setting. This will be done in sections 2 and 3.

We will show in section 4 that in the zero temperature limit the function $\tilde{\varphi}(x)$ will correspond to the function $m(x)$ previously defined. The optimal plan $\pi_{opt(c)}$ will correspond, in the zero temperature, to the limit of the equilibrium plans $\tilde{\pi}$.

The analogous questions in the case where X is not finite and μ is a general probability on X will require a different type of transfer operator.

This will introduce some technical difficulties which are similar to the ones analyzed in [8], where it is considered a general a priory probability. We will not address here this more general problem.

In the appendix we will present some technical results which are needed in our reasoning.

The setting presented here is different from [3] [4] [5] .

2 Thermodynamic formalism over $\Pi(\cdot, \sigma)$

We assume that $c : X \times \Omega \to \mathbb{R}$ is a Lipschitz function and we define the transfer operator L_c, which acts on $C(\Omega)$, in the following way: given $\psi : \Omega \to \mathbb{R}$ we have

$$(L_c\psi)(y) = \sum_{x\in X} \sum_{\sigma(w)=y} e^{c(x,w)}\psi(w) = \sum_{x\in X} \sum_{a\in\{1,\ldots,d\}} e^{c(x,ay)}\psi(ay).$$

Proposition 1. *Let $c : X \times \Omega \to \mathbb{R}$ be a Lipschitz function. There exists a positive Lipschitz function h and a positive number λ, such that,*

$$\sum_{x\in X} \sum_{\sigma(w)=y} e^{c(x,w)}h(w) = \lambda h(y).$$

Moreover, λ is simple and the remainder of spectrum is contained in a disc with radius strictly smaller than λ.

Proof. If $K > 0$ is the Lipschitz constant of c, and we consider the potential $A(y) = \log\left(\sum_x e^{c(x,y)}\right)$, we have that A is a Lipschitz function with constant K. [2] Then, $(L_c\psi)(y) = \sum_{\sigma(w)=y} e^{A(w)}\psi(w)$, and the proposition follows easily from classical arguments, see section 2 in [9]. □

We say that a Lipschitz cost (potential) c is normalized if for any $y \in \Omega$, we have

$$\sum_{x\in X} \sum_{\sigma(w)=y} e^{c(x,w)} = 1.$$

If c is Lipschitz and λ, h are given by the Proposition 1, then $\overline{c}(x, y) = c(x, y) + \log(h(y)) - \log(h(\sigma(y))) - \log(\lambda)$ is normalized.

Let us assume now that $c : X \times \Omega \to \mathbb{R}$ is normalized. In this case we define the dual operator of the transfer operator L_c in the following way: for a given probability ν on Ω, we get a new probability $L_c^*(\nu)$ such that for any continuous function $\psi : \Omega \to \mathbb{R}$ we have

[2] $A(y) = \log\left(\sum_x e^{c(x,y)}\right) \leq \log\left(\sum_x e^{c(x,z)+K|y-z|}\right) = K|y - z| + A(z)$.

$$L_c^*(\nu)(\psi) = \int_\Omega L_c(\psi)\, d\nu = \int_\Omega \sum_{x\in X} \sum_{\sigma(w)=y} e^{c(x,w)}\psi(w)\, d\nu(y). \tag{4}$$

Using results of the classical thermodynamical formalism (see [9]) for the potential $A(y) = \log\left(\sum_x e^{c(x,y)}\right)$ we have that the operator L_c^* has a unique fixed point probability ν_c, i.e. $L_c^*(\nu_c) = \nu_c$. We call ν_c the *Gibbs probability measure* associated to the normalized cost c.

We want to extend the above definitions of transfer operator (and, moreover, of the dual operator) which acts on functions of the variable y to functions which depends on coordinates (x, y). Let c be a normalized cost, then denote

$$\hat{L}_c(u)(y) = \sum_x \sum_{\sigma(w)=y} e^{c(x,w)}u(x,w)$$

for any $u : X \times \Omega \to \mathbb{R}$. Note that $\hat{L}_c$ sends $u : X \times \Omega \to \mathbb{R}$ to the function denoted by $\hat{L}_c(u) : \Omega \to \mathbb{R}$. For such normalized c we denote $\hat{L}_c^*$ the operator on $P(X \times \Omega)$ defined by

$$\hat{L}_c^*(\pi)(u(z,y)) \;=\; \int_{X\times\Omega} \left(\sum_x \sum_{\sigma(w)=y} e^{c(x,w)}u(x,w) \right) d\pi(z,y)\,. \tag{5}$$

Definition 2. *A probability on $X \times \Omega$ which is a fixed point for $\hat{L}_c^*$ is called* **a Gibbs plan** *for the normalized cost (potential) c. It will be denoted by π_c.*

The normalization property implies that $\hat{L}_c^*(\pi)(1) = 1$, i.e. the $\hat{L}_c^*$ preserves the convex and compact set $P(X \times \Omega)$.

Proposition 3. *Given a normalized cost c there exists a unique fixed point π_c for the operator $\hat{L}_c^*$. We have that $\pi_c \in \Pi(\cdot, \sigma)$ and the y-marginal of π_c is the Gibbs measure ν_c. Moreover, the support of π_c is the set $X \times \Omega$.*

We refer the reader to the appendix for the proof of the above result.

Now we define the entropy of plans (Definition 7), the pressure of a Lipschitz cost, and the concept of equilibrium plan for a cost (Definition 8). We will also see some examples and properties of such concepts, and prove a variational principle for the pressure which shows that the equilibrium plan for a cost is the Gibbs plan for the associated normalized cost (Theorem 9). We will follow the main lines of [9], Chapter 3.

We denote by $[x, y_0...y_n] = \{(z, w) \in X \times \Omega : z = x, w_0 = y_0, ..., w_n = y_n\}$ and $[y_1...y_n] = \{w \in \Omega : w_0 = y_1, ..., w_{n-1} = y_n\}$. Consider a fixed plan $\pi \in \Pi(\cdot, \sigma)$ with y-marginal ν and define

$$J^n_\pi(x, y) = \frac{\pi([x, y_0...y_n])}{\nu([y_1...y_n])}$$

if $y = (y_1, y_2, ...) \in$ supp(ν). From the Increasing Martingale Theorem the functions J^n_π converge to a function $J_\pi(x, y)$ in $L^1(X \times \Omega, \mathcal{B}(X \times \Omega), \pi)$ and for π a.e. (x, y). For each plan π this function J_π can be also obtained via the Radon-Nikodyn Theorem. This is carefully explained in the appendix. We have, $J_\pi > 0$ a.e. (π) and $\sum_x \sum_a J_\pi(x, ay) = 1$. For a plan $\pi \in \Pi(\cdot, \sigma)$ we call J_π the **Jacobian of the plan**. For a general $\pi \in \Pi(\cdot, \sigma)$ the Jacobian J_π is not necessarily Lipschitz but just measurable.

Lemma 4. *Suppose π is a plan in $\Pi(\cdot, \sigma)$. Then, for every $w \in C(X, \Omega)$,*

$$\int_{X\times\Omega} \sum_x \sum_a J_\pi(x, ay) w(x, ay)\, d\pi = \int_{X\times\Omega} w(z, y)\, d\pi.$$

This result is proved in the Appendix.

Lemma 5. *If a normalized Lipschitz potential $c(x, y)$ has a Gibbs plan π_c, then $J_{\pi_c} = e^c$.*

Proof. This follows easily from [9] proposition 3.2 and corollary 3.2.2 with simple adaptations in the computations.

□

Example 1:

We consider as an example the case where $X = \{1, 2\}$, $\Omega = \{1, 2\}^{\mathbb{N}}$, and c is such that depends just on two coordinates on y, that is, $c(x, y) = c(x,\, y_1\, y_2) = c^{\,x}(y_1 y_2)$, $x = 1, 2$, let us denote by $a^i_{r,s} = e^{c^i(r,s)}$, $i, r, s = 1, 2$, and

$$A^1 = \begin{pmatrix} a^1_{11} & a^1_{12} \\ a^1_{21} & a^1_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \;,\; A^2 = \begin{pmatrix} a^2_{11} & a^2_{12} \\ a^2_{21} & a^2_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

We want to determine the Gibbs plan π for such c.

The action of L_c over potentials that depends only of one coordinate y_1, $y_1 = 1, 2$, can be written in the form of the action on a vector h

$$(h_1, h_2)[A^1 + A^2] = (h_1, h_2) \begin{pmatrix} a^1_{11} + a^2_{11} & a^1_{12} + a^2_{12} \\ a^1_{21} + a^2_{21} & a^1_{22} + a^2_{22} \end{pmatrix}.$$

In this case $\lambda = \frac{5+\sqrt{17}}{2}$ and $(h_1, h_2) = (3+\sqrt{17}, 5+\sqrt{17})$ are respectively the associated eigenvalue and eigenvector. After normalization, we get the column stochastic matrix

$$\overline{A} = \begin{pmatrix} \frac{a^1_{11}+a^2_{11}}{\lambda} & \frac{a^1_{12}+a^2_{12}}{\lambda}\frac{h_1}{h_2} \\ \\ \frac{a^1_{21}+a^2_{21}}{\lambda}\frac{h_2}{h_1} & \frac{a^1_{22}+a^2_{22}}{\lambda} \end{pmatrix} = \begin{pmatrix} \frac{4}{5+\sqrt{17}} & \frac{4}{5+\sqrt{17}}\frac{3+\sqrt{17}}{5+\sqrt{17}} \\ \\ \frac{4}{3+\sqrt{17}} & \frac{6}{5+\sqrt{17}} \end{pmatrix} = \begin{pmatrix} 0,4384 & 0,3423 \\ \\ 0,5616 & 0,6577 \end{pmatrix}.$$

The stationary initial probability on $\{1,2\}$ of a stochastic matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is given by $\nu = (p_1, p_2)^T = (\frac{b}{b+1-a}, \frac{1-a}{b+1-a})^T$. Finally, we get $p_1 = 0,3786$ and $p_2 = 0,6213$. In order to obtain the Gibbs plan we need to split $\overline{A}$ in the form $\overline{A^1}$, $\overline{A^2}$ where

$$\overline{A^1} = \begin{pmatrix} \frac{2}{5+\sqrt{17}} & \frac{2}{5+\sqrt{17}}\frac{3+\sqrt{17}}{5+\sqrt{17}} \\ \\ \frac{2}{3+\sqrt{17}} & \frac{2}{5+\sqrt{17}} \end{pmatrix} = \begin{pmatrix} 0,2192 & 0,1711 \\ \\ 0,2808 & 0,2192 \end{pmatrix}$$

$$\overline{A^2} = \begin{pmatrix} \frac{2}{5+\sqrt{17}} & \frac{2}{5+\sqrt{17}}\frac{3+\sqrt{17}}{5+\sqrt{17}} \\ \\ \frac{2}{3+\sqrt{17}} & \frac{4}{5+\sqrt{17}} \end{pmatrix} = \begin{pmatrix} 0,2192 & 0,1711 \\ \\ 0,2808 & 0,4385 \end{pmatrix}.$$

This defines the Jacobian of the plan π we are looking for.

Given, $u : X \times \Omega \to \mathbb{R}$, where $u(x,y)$ is such that depend just on the first coordinate of y in the Bernoulli space, we have that $\int u d\pi = u^1_1\pi^1_1 + u^1_2\pi^1_2 + u^2_1\pi^2_1 + u^2_2\pi^2_2$. Using (5) we get

$$\pi^1 = \overline{A^1}\nu = \begin{pmatrix} 0,1893 \\ 0,2425 \end{pmatrix}$$

and

$$\pi^2 = \overline{A^2}\nu = \begin{pmatrix} 0,1893 \\ 0,3787 \end{pmatrix}.$$

In this way we get the values $\pi^k_j = \pi(k, \overline{j})$, $j,k = 1,2$, where $\overline{j}$ is the cylinder of size 1 on $\{1,2\}^{\mathbb{N}}$ with first symbol j.

Note that as we get explicitly the Jacobian of π one can obtain (via the use of (5)) the probability of any cylinder.

In this way we get information on the Gibbs plan π we were looking for. It is easy to see that the above arguments can be applied in the same way for general matrices A^1 and A^2.

Lemma 6. *If b is a normalized Lipchitz potential and $\pi \in \Pi(\cdot, \sigma)$, then*

$$0 \leq -\int_{X\times\Omega} \log(J_\pi)\, d\pi \leq -\int_{X\times\Omega} b\, d\pi,$$

with equality, if and only if, $b = \log(J_\pi)$.

Furthermore,

$$-\int_{X\times\Omega} \log(J_\pi)\, d\pi = \inf_{b\, normalized} -\int_{X\times\Omega} b\, d\pi.$$

The proof of this result appears in the Appendix.

Definition 7. *If $\pi \in \Pi(\cdot, \sigma)$ we define the entropy of π by*

$$H(\pi) = -\int_{X\times\Omega} \log(J_\pi)\, d\pi = \inf_{b\, normalized} -\int_{X\times\Omega} b\, d\pi.$$

The functions $\log(J_\pi^n)$ converge to $\log(J_\pi)$ in $L^1(X \times \Omega, \mathcal{B}(X \times \Omega), \pi)$ and we can compute the entropy from the limit

$$H(\pi) = -\lim_{n\to\infty} \int_{X\times\Omega} \log(J_\pi^n)\, d\pi.$$

In the case X has just one point the above definition matches the usual one for the Kolmogorov entropy (see [6] and [9]). If X has $\#X$ elements and $\Omega = \{1, ..., d\}^{\mathbb{N}}$, then $c(x, y) = -\log(d(\#X))$ is a normalized cost, therefore

$$0 \leq H(\pi) \leq \log(d) + \log(\#X).$$

Definition 8. *The pressure of a Lipschitz continuous cost (potential) c is defined by*

$$P(c) = \sup_{\pi\in\Pi(\cdot,\sigma)} \left(\int_{X\times\Omega} c\, d\pi + H(\pi) \right).$$

A plan $\pi \in \Pi(\cdot, \sigma)$ which realizes the supremum is called **an equilibrium plan for** *c.*

Theorem 9 (Variational Principle over $\Pi(\cdot, \sigma)$)**.** *Let us fix a Lipschitz cost c. Then, $P(c) = \log(\lambda_c)$, where λ_c is the main eigenvalue of L_c. The equilibrium plan for c is unique and given by the Gibbs plan for $\overline{c} := c+\log(h_c)-\log(h_c\circ \sigma) - \log(\lambda_c)$, where h_c is the eigenfunction associated to λ_c.*

We refer the reader to the appendix for a proof.

Now we present some properties of the entropy and pressure, as well as an example.

As X has a finite number of points we can consider the usual (non-dynamical) entropy of a probability measure $\mu \in P(X)$ given by $h(\mu) = -\sum_{x \in X} \mu(x) \log(\mu(x))$. For each σ-invariant probability $\nu \in P(\Omega)$ the Kolmogorov entropy is denoted by $h(\nu)$.

Proposition 10. *Given $\pi \in \Pi(\cdot, \sigma)$, if the x-marginal of π is a probability measure μ and the y-marginal of π is an invariant measure ν, then*

$$H(\pi) \leq h(\mu) + h(\nu).$$

If $\pi = \mu \times \nu$, then $H(\pi) = h(\mu) + h(\nu)$.

For a proof see the Appendix.

Example 2: If $X = \{1, 2\}$ and $\Omega = \{1, 2\}^{\mathbb{N}}$, then:
1. Consider the plan π defined from

$$\pi([x, y_0 y_1 ... y_n]) = \begin{cases} \frac{1}{2^{n+1}} & \text{if } y_0 = x \\ 0 & \text{if } y_0 \neq x \end{cases}$$

The x-marginal of π is $\mu = (\frac{1}{2}, \frac{1}{2})$ and the y-marginal of π is the Bernoulli measure ν with uniform distribution. Then, $h(\mu) = h(\nu) = \log(2)$ and $H(\pi) = \log(2)$. Indeed, we have

$$\begin{aligned} H(\pi) &= \lim_{n\to\infty} -\int_{X\times\Omega} \log\left(\frac{\pi([x, y_0...y_n])}{\nu([y_1...y_n])}\right) d\pi(x, y) \\ &= \lim_{n\to\infty} -\sum_{x} \sum_{y_0,...,y_n} \log\left(\frac{\pi([x, y_0...y_n])}{\nu([y_1...y_n])}\right) \pi([x, y_0...y_n]) \\ &= \lim_{n\to\infty} -\sum_{y_0...y_n} \log\left(\frac{\pi([y_0, y_0...y_n])}{\nu([y_1...y_n])}\right) \pi([y_0, y_0...y_n]) \\ &= \lim_{n\to\infty} -\sum_{y_0...y_n} \log\left(\frac{1}{2}\right) \frac{1}{2^{n+1}} = \log(2) \end{aligned}$$

2. Consider any plan $\pi = \mu \times \nu$ where $\mu = (\frac{1}{2}, \frac{1}{2})$ and ν has support in a periodic orbit. In this case $h(\mu) = \log(2)$ while $h(\nu) = 0$. Once more we get $H(\pi) = \log(2)$ because π is a product plan.

3. Let π be the plan with two atoms $\pi = \frac{1}{2}\delta_{(1,\overline{12})} + \frac{1}{2}\delta_{(2,\overline{21})}$, where $\overline{y_0 y_1} = (y_0, y_1, y_0, y_1, y_0, ...)$. The x-marginal of π is again $\mu = (\frac{1}{2}, \frac{1}{2})$ while the y-marginal is the invariant measure with support equal to the periodic orbit $\overline{12}$. In the present case $H(\pi) = 0$. Indeed,

$$\pi([x, y_0...y_n]) = \begin{cases} \frac{1}{2} & \text{if } x = y_0 = y_2 = y_4 = ... \text{ and } x \neq y_1 = y_3 = ... \\ 0 & \text{else} \end{cases}.$$

Then,

$$\begin{aligned} H(\pi) &= \lim_{n\to\infty} -\sum_x \sum_{y_0...y_n} \log\left(\frac{\pi([x, y_0...y_n])}{\nu([y_1...y_n])}\right) \pi([x, y_0...y_n]) \\ &= \lim_{n\to\infty} -\sum_x \log\left(\frac{\frac{1}{2}}{\frac{1}{2}}\right)\frac{1}{2} = 0. \end{aligned}$$

Proposition 11. *The entropy is a concave and upper semi-continuous function.*

Proof. Both properties are consequences of the definition

$$H(\pi) = \inf_{b\, normalized} - \int_{X\times\Omega} b\, d\pi.$$

The proof follows the same arguments used in [8]. □

Proposition 12. *The pressure has the following properties:*
(a) if $c_1 \geq c_2$, then $P(c_1) \geq P(c_2)$.
(b) $P(c + a) = P(c) + a$ if $a \in \mathbb{R}$.
(c) The pressure is convex .
(d) $|P(c_1) - P(c_2)| \leq \|c_1 - c_2\|$.

The proof is in the Appendix.

3 Kantorovich duality for Thermodynamic Formalism over $\Pi(\mu, \sigma)$

In the last section μ was not fixed. In this section the probability μ on X is fixed.

We define the μ-pressure of c by

$$P_\mu(c) = \sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi + H(\pi). \tag{6}$$

Note that, $P_\mu(c) \leq P(c)$.

By compactness, there exists a plan $\tilde{\pi}_c \in \Pi(\mu, \sigma)$ which attains the supremum

$$\sup_{\pi \in \Pi(\mu,\sigma)} \int_{X\times\Omega} c(x,y) d\pi + H(\pi).$$

Remember that the Lipchitz functions are C^0 dense in the set of continuous functions.

The results that we will prove in this section can be resumed in the following.

Theorem 13 (Variational Principle for $\Pi(\mu, \sigma)$)**.**

$$\inf_{\varphi : P(c-\varphi)=0} \int_X \varphi(x)\, d\mu = \sup_{\pi \in \Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi \,+\, H(\pi). \tag{7}$$

The above infimum and supremum are attained in unique elements $\tilde{\varphi}$ and $\tilde{\pi}$. The maximizer $\tilde{\pi}$ is the Gibbs plan for $c - \tilde{\varphi}$.

The next corollary can be interpreted as the slackness condition on the present setting:

Corollary 14. *Given $\varphi(x)$ such that $P(c - \varphi) = 0$ and a plan $\pi_0 \in \Pi(\mu, \sigma)$, if π_0 is the Gibbs measure of $c - \varphi$, then π_0 attains the supremum and φ attains the infimum in* (7)*.*

Proof. We have

$$0 = P(c-\varphi) = \sup_{\pi \in \Pi(\cdot,\sigma)} \int_{X\times\Omega} (c-\varphi)\, d\pi \,+\, H(\pi) = \int_{X\times\Omega} (c-\varphi)\, d\pi_0 \,+\, H(\pi_0). \tag{8}$$

Therefore

$$\sup_{\pi \in \Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi \,+\, H(\pi) = \int \varphi d\mu$$

and the supremum is attained in π_0. □

Definition 15. *Given a Lipschitz cost (potential) c we define Φ_c as the set of all pairs of continuous functions $(\varphi, \psi) \in C(X) \times C(\Omega)$ which satisfy*

$$\varphi(x) - \psi(y) + (\psi \circ \sigma)(y) \geq c(x,y) - b(x,y), \qquad \forall\, (x,y) \in X \times \Omega \tag{9}$$

for some Lipschitz function b with zero pressure. Following the classical terminology it is natural to call $\varphi(x)$ and $\psi(y) + (\psi \circ \sigma)(y)$ of c-admissible pair.

The theorem stated below is the version for positive temperature of the main theorem in [7] (which in some sense corresponds to zero temperature).

Theorem 16. *Given a Lipschitz cost c we have*

$$\inf_{(\varphi,\psi)\in\Phi_c}\int_X \varphi\, d\mu = \sup_{\pi\in\Pi(\mu,\sigma)}\int_{X\times\Omega} c(x,y)\, d\pi + H(\pi). \tag{10}$$

The supremum in (10) is attained in at least one plan.

In Proposition 18 we will prove that the infimum in (10) is attained in exactly one function φ and that this infimum coincides with the left hand side of the Variation Principle stated above. In order to prove this theorem we follow [14] and we use the next theorem (see also [7]).

Theorem 17 (**Fenchel-Rockafellar duality**)**.** *Suppose E is a normed vector space, Θ and Ξ two convex functions defined on E taking values in $\mathbb{R}\cup\{+\infty\}$. Denote Θ^* and Ξ^*, respectively, the Legendre-Fenchel transform of Θ and Ξ. Suppose there exists $v_0 \in E$, such that $\Theta(v_0) < +\infty$, $\Xi(v_0) < +\infty$ and that Θ is continuous on v_0.*

Then,

$$\inf_{v\in E}[\Theta(v) + \Xi(v)] = \sup_{f\in E^*}[-\Theta^*(-f) - \Xi^*(f)] \tag{11}$$

Moreover, the supremum in (11) is attained in at least one element in E^*.

Proof. (of Theorem 16)

It is enough to consider the case were $P(c) = 0$. Indeed, let us we assume the theorem is proved for costs with zero pressure, if $P(c) \neq 0$, we define $\tilde{c} = c - P(c)$. In this way $(\varphi,\psi) \in \Phi_c$, if and only if, $(\varphi - P(c),\psi) \in \Phi_{\tilde{c}}$. Then,

$$\inf_{(\varphi,\psi)\in\Phi_c}\int_X \varphi\, d\mu = \inf_{(\varphi-P(c),\psi)\in\Phi_{\tilde{c}}}\int_X \varphi\, d\mu = \inf_{(\tilde{\varphi},\psi)\in\Phi_{\tilde{c}}}\int_X \tilde{\varphi}\, d\mu + P(c)$$

$$= \sup_{\pi\in\Pi(\mu,\sigma)}\int_{X\times\Omega} \tilde{c}\, d\pi + H(\pi) + P(c) = \sup_{\pi\in\Pi(\mu,\sigma)}\int_{X\times\Omega} c\, d\pi + H(\pi).$$

Hence, from now on we will assume that $P(c) = 0$.

We want to use, the Fenchel-Rockafellar duality in the proof. For this purpose we define

$$E = C(X\times\Omega),$$

where $C(X\times\Omega)$ is the set of all continuous functions in $X\times\Omega$ taking values in $\mathbb{R}$, with the usual sup norm. Moreover, $E^* = M(X\times\Omega)$ is the set of continuous linear operators in $C(X \times \Omega)$ taking values in $\mathbb{R}$ with the total variation norm. The elements in $M(X \times \Omega)$ are signed measures.

Define $\Theta, \Xi : E \longrightarrow \mathbb{R} \cup \{+\infty\}$ from

$$\Theta(u) = \begin{cases} 0, & \text{if } u(x,y) \geq c(x,y) - b(x,y), \ \forall (x,y), \\ & \ \text{for some } b \text{ with } P(b) = 0, \\ \\ +\infty, & \text{in the other case} \end{cases}$$

and

$$\Xi(u) = \begin{cases} \int_X \varphi \, d\mu, & \text{if } u(x,y) = \varphi(x) - \psi(y) + (\psi \circ \sigma)(y), \\ & \text{where } (\varphi, \psi) \in C(X) \times C(\Omega)\,, \\ \\ +\infty, & \text{in the other case.} \end{cases}$$

Note that Ξ is well defined. Indeed, if $u(x,y) = \varphi_1(x) - \psi_1(y) + (\psi_1 \circ \sigma)(y) = \varphi_2(x) - \psi_2(y) + (\psi_2 \circ \sigma)(y)$, then, integrating over a probability in $\Pi(\mu, \sigma)$, we conclude that $\int_X \varphi_1(x) d\mu = \int_X \varphi_2(x) d\mu$, $\forall\, x \in X$.

Now we will show that the hypothesis in Theorem 17 are satisfied. The convexity of Ξ is immediate. To show the convexity of Θ take u_1 and u_2 such that $\Theta(u_1) = \Theta(u_2) = 0$, then there exist b_1 and b_2 Lipschitz with $P(b_1) = P(b_2) = 0$, such that, $u_1 \geq c - b_1$ and $u_2 \geq c - b_2$. Note that

$$\lambda u_1 + (1-\lambda) u_2 \geq c - (\lambda b_1 + (1-\lambda) b_2),$$

and then using item (c) of Proposition 12, we see

$$a = P(\lambda b_1 + (1-\lambda) b_2) \leq \lambda P(b_1) + (1-\lambda) P(b_2) = 0.$$

Therefore, by item (b) of Proposition 12, $P(\,[\lambda b_1 + (1-\lambda) b_2\,] -\, a) = 0$. In this way

$$\lambda u_1 + (1-\lambda) u_2 \geq c - (\lambda b_1 + (1-\lambda) b_2) \geq c - ([\lambda b_1 + (1-\lambda) b_2] - a).$$

This shows that $\Theta(\lambda u_1 + (1-\lambda) u_2) = 0$ and hence Θ is convex. Finally we exhibit u_0 in the domain of Ξ and Θ: take $u_0 = 1$. Then, $\Xi(1) = 1$ and $\Theta(1) = 0$, because $1 > 0 = c - c$, and $P(c) = 0$. If $w \in C(X \times \Omega)$ such that $\|w - 1\| < 1/2$, then $w > 0 = c - c$, and $P(c) = 0$. This shows that Θ is continuous in u_0.

Let us now compute the Legendre-Fenchel transform of Θ and Ξ. We denote by π an element in $E^* = M(X \times \Omega)$.

For any $\pi \in E^*$, by the definition of Θ we get

$$\begin{aligned}\Theta^*(-\pi) &= \sup_{u\in E}\{\langle -\pi, u\rangle - \Theta(u)\}\\ &= \sup_{u\in E}\{\langle \pi, u\rangle :\ u \le -c+b,\ b \text{ with } P(b)=0\}.\end{aligned}$$

If π is not a positive functional, then there exists a function $v \le 0$, $v \in C(X\times\Omega)$, such that, $\langle \pi, v\rangle > 0$. We can assume that v is Lipschitz. Note that $c+v \le c$, hence using item (a) of Proposition 12 and that $P(c)=0$, we have $P(c+v) \le 0$. Therefore, $v = -c+(c+v) \le -c+(c+v-P(c+v))$, with $P(c+v-P(c+v))=0$. Taking $u=\lambda v$, and considering $\lambda \to +\infty$, we get that

$$\sup_{u\in C(X\times\Omega)} \{\langle \pi, u\rangle :\ u \le -c+b,\quad b \text{ with } P(b)=0\} = +\infty.$$

Therefore, we assume from now on that π is a positive functional. Note that, if $\pi \in M^+(X\times\Omega)$, then

$$\sup_{u\in C(X\times\Omega)} \{\langle \pi, u\rangle :\ u \le -c+b,\quad b \text{ with } P(b)=0\} = \langle \pi, -c\rangle + \sup_{b\,:\,P(b)=0} \langle \pi, b\rangle.$$

Hence, we obtain

$$\Theta^*(-\pi) = \begin{cases} \langle \pi,\, (-c\,)\,\rangle + \sup\limits_{b\,:\,P(b)=0} \langle \pi, b\rangle\,, \text{ if } \pi \in M^+(X\times\Omega) \\ \\ +\infty,\ \text{in the other case.}\end{cases} \tag{12}$$

Analogously, by the definition of Ξ we get that

$$\Xi^*(\pi) = \sup_{u\in E}\{\langle \pi, u\rangle - \Xi(u)\}$$

$$= \sup_{u\in E}\left\{ \begin{array}{l} \langle \pi, u\rangle - \int_X \varphi\, d\mu :\\ u(x,y) = \varphi(x) - \psi(y) + \psi(\sigma(y)) \text{ where } (\varphi,\psi) \in C(X)\times C(\Omega) \end{array}\right\}$$

$$= \sup_{(\varphi,\psi)\in C(X)\times C(\Omega),} \left\{ \langle \pi, \varphi(x) - \psi(y) + \psi(\sigma(y))\rangle - \int_X \varphi\, d\mu \right\}.$$

Note that, if $\langle \pi, \varphi(x)\rangle > \int_X \varphi\, d\mu$, for some φ, choosing $\lambda.\varphi$ with $\lambda \to \infty$, the supremum will be equal to $+\infty$. Also if $\langle \pi, \psi(y) - \psi(\sigma(y))\rangle > 0$, for some ψ, taking $\lambda.\psi$ with $\lambda \to \infty$, the supremum will be $+\infty$. The case where we consider the other inequality is analogous. Then, we can assume that $\langle \pi, \varphi(x)\rangle = \int_X \varphi\, d\mu$ and $\langle \pi, \psi(y) - \psi(\sigma(y))\rangle = 0$.

In order to simplify the notation, we define

$$\Pi^*(\mu) = \left\{ \pi \in M(X \times \Omega) : \begin{array}{l} \langle \pi, \varphi(x) \rangle = \int_X \varphi \, d\mu \text{ and } \langle \pi, \psi(y) - \psi(\sigma(y)) \rangle = 0 \\ \forall (\varphi, \psi) \in C(X) \times C(\Omega) \end{array} \right\}.$$

With this notation we can write

$$\Xi^*(\pi) = \begin{cases} 0, \text{ if } \pi \in \Pi^*(\mu), \\ \\ +\infty, \text{ in the other case.} \end{cases} \tag{13}$$

We observe that if $\pi \in M^+(X \times \Omega) \cap \Pi^*(\mu)$, then $\langle \pi, 1 \rangle = \mu(1) = 1$, $\langle \pi, u \rangle \geq 0$ when $u \geq 0$ and also $\langle \pi, \cdot \rangle$ is linear. From these properties we get that $\pi \in P(X \times \Omega)$. Moreover, by definition of $\Pi^*(\mu)$, the x-marginal of π is μ and the y-marginal of π is σ-invariant. Hence, we conclude $M^+(X \times \Omega) \cap \Pi^*(\mu) = \Pi(\mu, \sigma)$.

By definition 7, if $\pi \in \Pi(\mu, \sigma)$, we get

$$-\sup_{b \,:\, P(b)=0} \pi(b) = -\sup_{b \text{ normalized}} \int_{X \times \Omega} b \, d\pi = H(\pi).$$

The left hand side of (11) is given by

$$\begin{aligned} &\inf_{u \in E} [\Theta(u) + \Xi(u)] \\ &= \inf_{u \in E} \Big\{ \int_X \varphi \, d\mu : \varphi(x) - [\psi - (\psi \circ \sigma)](y) \geq c(x,y) - b(x,y), \\ &\quad \text{for some } b \text{ with } P(b) = 0, (\varphi, \psi) \in C(X) \times C(\Omega) \Big\} \\ &= \inf_{(\varphi,\psi) \in \Phi_c} \int_X \varphi \, d\mu. \end{aligned}$$

The right hand side of (11) is given by

$$\begin{aligned} &\sup_{\pi \in E^*} [-\Theta^*(-\pi) - \Xi^*(\pi)] \\ &= \sup_{\pi \in E^*} \left\{ \begin{array}{ll} \pi(c) + H(\pi) \ , & \text{if } \pi \in \Pi(\mu, \sigma) \\ \\ -\infty, & \text{in the other case} \end{array} \right\} \\ &= \sup_{\pi \in \Pi(\mu,\sigma)} \{\pi(c) + H(\pi)\}. \end{aligned}$$

Therefore, from (11) we get

$$\inf_{(\varphi,\psi)\in\Phi_c}\int_X \varphi\, d\mu = \sup_{\pi\in\Pi(\mu,\sigma)}\int_{X\times\Omega} c\, d\pi \,+\, H(\pi).$$

Theorem 17 claims that

$$\sup_{f\in E^*}[-\Theta^*(-f)-\Xi^*(f)] = \sup_{\pi\in\Pi(\mu,\sigma)}\int_{X\times\Omega} c\, d\pi + H(\pi).$$

is attained, for at least one element (but we already know this by compactness).

□

Example 3:

We consider again as an example the case where $X=\{1,2\}$, $\Omega=\{1,2\}^{\mathbb{N}}$, and c is such that depends just on two coordinates on y, that is, $c(x,y)=c(x,\, y_1\, y_2)=c^{\,x}(y_1y_2)$, $x=1,2$, and

$$A^1=\begin{pmatrix} a^1_{11} & a^1_{12}\\ a^1_{21} & a^1_{22}\end{pmatrix}\ ,\ A^2=\begin{pmatrix} a^2_{11} & a^2_{12}\\ a^2_{21} & a^2_{22}\end{pmatrix},$$

where $a^i_{r,s}=e^{c^i(r,s)}$, $i,r,s=1,2$.

We fix $\mu=(\mu_1,\mu_2)$ and we are going to explain how one can get the solution $\pi\in\Pi(\mu,\sigma)$ of the above transport problem via the equation

$$\inf_{\varphi:P(c-\varphi)=0}\int_X \varphi(x)\, d\mu = \sup_{\pi\in\Pi(\mu,\sigma)}\int_{X\times\Omega} c\, d\pi \,+\, H(\pi).$$

We consider first the left side expression.

The function φ is described by (φ_1,φ_2). The condition $P(c-\varphi)=0$ means that

$$\begin{pmatrix} e^{c^1_{11}}\, e^{-\varphi_1}+e^{c^2_{11}}\, e^{-\varphi_2} & e^{c^1_{12}}\, e^{-\varphi_1}+e^{c^2_{12}}\, e^{-\varphi_2}\\ e^{c^1_{21}}\, e^{-\varphi_1}+e^{c^2_{21}}\, e^{-\varphi_2} & e^{c^1_{22}}\, e^{-\varphi_1}+e^{c^2_{22}}\, e^{-\varphi_2}\end{pmatrix}$$

has a dominant eigenvalue 1.

We are interested in the $z_1=e^{-\varphi_1}$, $z_2=e^{-\varphi_2}$ which are solutions of the equation

$$\det\begin{pmatrix} (e^{c^1_{11}}\, z_1+e^{c^2_{11}}\, z_2)-1 & e^{c^1_{12}}\, z_1+e^{c^2_{12}}\, z_2\\ e^{c^1_{21}}\, z_1+e^{c^2_{21}}\, z_2 & (e^{c^1_{22}}\, z_1+e^{c^2_{22}}\, z_2)-1\end{pmatrix}=0.$$

In this way we get that (z_1,z_2) describes an algebraic curve on $\mathbb{R}^2$. This equation does not discriminate if the eigenvalue 1 is maximal but this is not a big problem. Now we have to find the points (z_1, z_2) of such curve such that its normal vector is colinear with the vector $v(z_1, z_2) = (\mu_1 \frac{1}{z_1}, \mu_2 \frac{1}{z_2})$, which is the gradient of the function $(z_1, z_2) \to \log z_1 \mu_1 + \log z_2 \mu_2$ (Lagrange multipliers). This will determine a finite set (in the generic case) of possible $\varphi = (\varphi_1, \varphi_2)$, which are critical points for $\varphi \to \int \varphi \, d\mu$. We test these possibilities and then we get the minimal φ which we denote by $\tilde{\varphi}$. In this way we determine the left hand side of the last main equality and the value of the μ pressure of c.

Now we consider the potential $\tilde{c} = c - \tilde{\varphi}$. Finally using the same procedure of example 1 one can get the Gibbs plan for $\tilde{c}$. In this way we solve the Ergodic Transport problem for c with a fixed marginal μ.

To show the uniqueness in the next proposition we will use the property that the pressure is an analytical function of the potential (see [13] and [12]).

Proposition 18.

$$\inf_{(\varphi,\psi)\in\Phi_c} \int_X \varphi \, d\mu = \inf_{\varphi : P(c-\varphi)=0} \int_X \varphi \, d\mu \tag{14}$$

The infimum is attained at exactly one function $\tilde{\varphi}$.

Proof. Given $(\varphi, \psi) \in \Phi_c$, there exists b such that $P(b) = 0$ and

$$c(x, y) - \varphi(x) + \psi(y) - (\psi \circ \sigma)(y) \leq b(x, y).$$

Then, by item (a) of proposition 12, $P(c(x, y) - \varphi(x)) \leq 0$. On the other hand, if $P(c(x, y) - \varphi(x)) = a \leq 0$, then we define $b(x, y) = c(x, y) - \varphi(x) - a$. We have that $P(b) = 0$ and $b(x, y) \geq c(x, y) - \varphi(x)$. Hence,

$$\inf_{(\varphi,\psi)\in\Phi_c} \int_X \varphi \, d\mu = \inf_{\varphi : P(c-\varphi)\leq 0} \int_X \varphi \, d\mu.$$

By monotonicity of the pressure, we have

$$\inf_{(\varphi,\psi)\in\Phi_c} \int_X \varphi \, d\mu = \inf_{\varphi : P(c-\varphi)=0} \int_X \varphi \, d\mu.$$

Note also that, if $P(c - \varphi) \neq 0$, we can add the constant $-P(c - \varphi)$ and get $P(c - \varphi - P(c - \varphi)) = 0$. Then,

$$\inf_{\varphi : P(c-\varphi)=0} \int_X \varphi \, d\mu = \inf_{\varphi} \int_X \varphi \, d\mu + P(c - \varphi) = \inf_{\varphi} - \int_X \varphi \, d\mu + P(c + \varphi) \tag{15}$$

Consider the continuous function $F : C(X) \to \mathbb{R}$ given by

$$F(\varphi) = -\int_X \varphi \, d\mu + P(c + \varphi),$$

we see that, if $a \in \mathbb{R}$, then $F(\varphi + a) = F(\varphi)$. This shows we can minimize $F(\varphi)$ among φ such that $\varphi(0) = 0$.

In order to prove the uniqueness of the minimizer of F, we assume that $X = \{0, ..., k\}$ has $k + 1$ elements, $\varphi(0) = 0$ and we identify φ with $v \in \mathbb{R}^k$, in the following way, $\varphi = (0, v_1, ..., v_k)$, i.e., $\varphi(j) = v_j$, $j = 1, 2, ..., k$.

Therefore, $F : \mathbb{R}^k \to \mathbb{R}$ associates to each vector $v \in \mathbb{R}^k$ the number

$$F(v) = -\int_X \varphi \, d\mu + P(c + \varphi).$$

In this way, to finish the proof, we need to show that $F(v)$ has only one minimizer $\tilde{v} = \tilde{v}_c$.

We begin by proving that when $t \to +\infty$, $F(tv) \to +\infty$ uniformily in $\mathbb{S}^{k-1}$, i.e, there exist an $\epsilon > 0$ and $\xi \in \mathbb{R}$, such that, for any $v \in \mathbb{S}^{k-1}$, we have that

$$F(tv) \geq t\varepsilon + \xi. \tag{16}$$

In order to do that, let $K_1 = \{v \in \mathbb{S}^{k-1} : v_i \geq 0 \text{ for some } i\}$ and $K_2 = \{v \in \mathbb{S}^{k-1} : v_i \leq 0 \ \forall i\}$. We have $K_1 \cup K_2 = \mathbb{S}^{k-1}$ and K_1, K_2 are compact sets.

Using the fact that $\text{supp}(\mu) = X$, we have that the functions $-\int_X \varphi \, d\mu + \max_i \varphi(i)$ and $-\int_X \varphi \, d\mu$ are continuous and strictly positive in K_1 and K_2, respectively, where $\varphi = (0, v_1, ..., v_k)$. Then, there exists $\varepsilon > 0$, such that, $-\int_X \varphi \, d\mu + \max_i \varphi(i) \geq \varepsilon$, for all $v \in K_1$, and $-\int_X \varphi \, d\mu \geq \varepsilon$, for all $v \in K_2$.

Let us take $v \in K_1$ and a plan π with x-marginal δ_k, such that $\varphi(k) = \max_i \varphi(i)$, therefore $P(t\varphi) \geq \int_{X\times\Omega} t\varphi \, d\pi + H(\pi) = t\varphi(k) + H(\pi) \geq t\varphi(k) = t \max_i \varphi(i)$. Hence,

$$\begin{aligned} F(tv) &= -\int_X t\varphi \, d\mu + P(c + t\varphi) \geq -\int_X t\varphi \, d\mu + P(t\varphi) + \min c \\ &\geq -\int_X t\varphi \, d\mu + t \max_i \varphi(i) + \min c \geq t\varepsilon + \min c. \end{aligned}$$

Now, take $v \in K_2$ and a plan π with x-marginal δ_0, we have $P(t\varphi) \geq \int_{X\times\Omega} t\varphi \, d\pi + H(\pi) = H(\pi) \geq 0$. Hence,

$$\begin{aligned} F(tv) &= -\int_X t\varphi \, d\mu + P(c + t\varphi) \geq -\int_X t\varphi \, d\mu + P(t\varphi) + \min c \\ &\geq -\int_X t\varphi \, d\mu + \min c \geq t\varepsilon + \min c. \end{aligned}$$

We conclude that (16) holds, and this shows that F assume a minimum $\tilde{v}$ in $\mathbb{R}^k$.

Now we will prove that the minimizer $\tilde{v}$ is unique. Note that F is well defined for any v. We want to show that F is locally analytic. It will be the restriction of a complex analytic function. We use the analyticity of the pressure, which imply that F is analytic on v. Indeed, note that $A_v(y) = A_\varphi(y) = \log\left(\sum_x e^{c(x,y)+\varphi(x)}\right)$ is an analytic function on v (locally can be extended to a complex analytic function) taking values on the Banach space of Holder potentials on the variable y. As the composition of analytic functions is also analytic and the pressure is analytic on the potential (see Theorem 5.26 in [13]) we get our claim. As F is globally defined and locally analytic then it is analytic in the all domain.

We also know that the pressure is convex as a function of c (see Proposition 12). This implies that F is also convex in v.

Suppose that $\tilde{v}$ and $\hat{v}$ are minimizers for F. Using the convexity of F, we know that all convex combinations of $\tilde{v}$ and $\hat{v}$ are minimizers for F.

Now let the function $G : \mathbb{R} \to \mathbb{R}$ be defined by $G(t) = F(\tilde{v} + t(\hat{v} - \tilde{v}))$. G is an analytical function which converges to $+\infty$ when $t \to \pm\infty$.

Note that the second derivative of F can not be 0. Therefore, it can not be constant in a open interval of the real line, and we conclude that $\tilde{v} = \hat{v}$.

□

Corollary 19. *Let $\tilde{\varphi}$ be the unique minimizer for the Fenchel-Rockafellar duality* (10)*, then the Gibbs plan $\pi_{c-\tilde{\varphi}}$, for $c - \tilde{\varphi}$, belongs to $\Pi(\mu,\sigma)$ and is the unique maximizer of* (10).

Proof. Let $\tilde{\varphi}$ be the minimizer of (10) then

$$\int_X \tilde{\varphi}\, d\mu = \sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi \; + \; H(\pi), \tag{17}$$

hence

$$\sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} (c - \tilde{\varphi})\, d\pi \; + \; H(\pi) = 0, \tag{18}$$

which implies that $P_\mu(c - \tilde{\varphi}) = 0$. Also, by Proposition 18 we know that $P(c - \tilde{\varphi}) = 0$. Therefore, $P_\mu(c - \tilde{\varphi}) = P(c - \tilde{\varphi}) = 0$.

Now, let π_μ be a plan that attains the supremum in (18), which exists by Theorem 16, then π_μ also attains the supremum in $P(c - \tilde{\varphi}) = 0$. Finally using Theorem 9, we see that $\pi_\mu = \pi_{c-\tilde{\varphi}}$, as $\pi_{c-\tilde{\varphi}}$ is the unique equilibrium plan for $c - \tilde{\varphi}$, and this implies $\pi_{c-\tilde{\varphi}} \in \Pi(\mu,\sigma)$ and that $\pi_{c-\tilde{\varphi}}$ is the unique maximizer of (18), and hence of (17). □

Proof. (of Theorem 13) It follows by Theorem 16 and Proposition 18 that there exists a unique $\tilde{\varphi}$ such that

$$\int_X \tilde{\varphi}\, d\mu = \inf_{\varphi : P(c-\varphi)=0} \int_X \varphi\, d\mu = \sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c\, d\pi \,+\, H(\pi), \tag{19}$$

and by Corollary 19, we see that $\pi_{c-\tilde{\varphi}}$ is the unique maximizer of (19). □

4 The zero temperature limit

In this section we show that the main result proved in [7] and discussed in the introduction can be obtained from the reasoning of the above section considering the zero temperature limit.

Zero temperature for $\Pi(\cdot,\sigma)$

Given a Lipschitz potential c and a real variable $\beta > 0$, consider the potential $\beta\, c$. The parameter β corresponds to the inverse of the temperature in the Thermodynamic Formalism. We denote by λ_β the main eigenvalue of $L_{\beta c}$ and by h_β the main eigenfunction associate to λ_β (we can suppose that $\min(h_\beta) = 1$ for any β). Denote also by π_β the equilibrium plan for βc.

We note that h_β is the positive eigenfunction of the Ruelle operator with potential

$$A_\beta(y) = \log\left(\sum_x e^{\beta c(x,y)}\right),$$

in the classical Thermodynamic Formalism setting. Hence the Lipschitz constant of $\log(h_\beta)$ increase linearly with β [1]. From the Arzela-Ascolli Theorem $\frac{1}{\beta_n}\log(h_{\beta_n})$ converges for some sequence $\beta_n \to \infty$. Let

$$m = \sup_{\pi\in\Pi(\cdot,\sigma)} \int_{X\times\Omega} c(x,y)\, d\pi.$$

Using Theorem 9 and that $0 \le H(\pi) \le \log(\#X) + \log(d)$, we have

$$\beta m \le \log(\lambda_\beta) \le \beta m + \log(\#X) + \log(d)$$

and then $\lim_{\beta\to\infty} \frac{\log(\lambda_\beta)}{\beta} = m$. By compactness we know that there exist convergent sub-sequences of π_β, $\beta \to \infty$.

Suppose that for some sequence β_n we have $\frac{1}{\beta_n}\log(h_{\beta_n}) \to V$ and $\pi_{\beta_n} \to \pi_\infty$. Applying the Laplace's Method (see [1] and [8]) on the equation

$$\sum_x \sum_a e^{\beta c(x,ay)+\log(h_\beta(ay))-\log(h_\beta(y))-\log(\lambda_\beta)} = 1,$$

we conclude that

$$\sup_x \sup_a [\, c(x,ay) + V(ay) - V(y) - m] = 0, \quad \forall\, y.$$

Let us prove that π_∞ is a maximizing measure for c: analyzing the equation

$$\int_{X\times\Omega} \beta c(x,y)\, d\pi_\beta + H(\pi_\beta) = \sup_{\pi\in\Pi(\cdot,\sigma)} \int_{X\times\Omega} \beta c(x,y)\, d\pi + H(\pi),$$

we conclude that (dividing by β_n, making $\beta_n \to \infty$ and using that H is a bounded function)

$$\int_{X\times\Omega} c(x,y)\, d\pi_\infty = \sup_{\pi\in\Pi(\cdot,\sigma)} \int_{X\times\Omega} c(x,y)\, d\pi = m.$$

Now we prove the duality between the primal and dual problem:

Theorem 20. *m is the smallest real number α, such that, there exists a continuous function $S:\Omega\to\mathbb{R}$, satisfying $c(x,y)+S(y)-S(\sigma(y))-\alpha \le 0$, for any $x\in X$ and $y\in\Omega$.*

In this way, if we define Φ_c as the set of pair (α, S) such that $-\alpha+S(y)-S(\sigma(y)) \le -c(x,y)$ then

$$\inf_{(\alpha,S)\in\Phi_c} \alpha = \sup_{\pi\in\Pi(\cdot,\sigma)} \int_{X\times\Omega} c(x,y)\, d\pi$$

or, equivalently

$$\sup_{(\alpha,S)\in\Phi_c} -\alpha = \inf_{\pi\in\Pi(\cdot,\sigma)} \int_{X\times\Omega} -c(x,y)\, d\pi.$$

Proof. From the above arguments we know that

$$\sup_x \sup_a [\, c(x,ay) + V(ay) - V(y) - m] = 0, \quad \forall\, y$$

proving that m is a possible number. In order to show that m is the smallest possible number, fix α and S such that $c(x,y)+S(y)-S(\sigma(y))-\alpha \le 0$ for any $x\in X, y\in\Omega$. Then

$$\int_{X\times\Omega} c\, d\pi \le \alpha, \quad \forall\, \pi \in \Pi(\cdot,\sigma).$$

In this way

$$m = \sup_{\pi \in \Pi(\cdot,\sigma)} \int_{X\times\Omega} c\, d\pi \leq \alpha.$$

□

Zero temperature for $\Pi(\mu, \sigma)$

Now we consider the analogous problem over $\Pi(\mu, \sigma)$. For each $\beta > 0$, given the potential βc, by Theorem 13, there exists a unique function $\varphi_\beta(x)$ such that $P(\beta c - \varphi_\beta) = 0$ and

$$\int_X \varphi_\beta \, d\mu = \sup_{\pi \in \Pi(\mu,\sigma)} \int_{X\times\Omega} \beta c\, d\pi + H(\pi).$$

Let h_β be the eigenfunction associate to the eigenvalue 1 for $L_{\beta c - \varphi_\beta}$.We suppose $\min(h_\beta) = 1$. Let $\pi_\beta \in \Pi(\mu, \sigma)$ be the equilibrium plan for $\beta c - \varphi_\beta$.

Now we want to prove that the sequences $\frac{\varphi_\beta}{\beta}$ and $\frac{h_\beta}{\beta}$ converge in subsequence, when $\beta \to \infty$. To do this we need show that the Lipschitz constant of $\beta c - \varphi_\beta$ increases linearly with β.

According to (15) we can add a constant to φ_β and take φ^*_β a minimizer of

$$\int_X \varphi \, d\mu + P(\beta c - \varphi),$$

such that $\varphi^*_\beta(0) = 0$, we have that $\varphi_\beta = \varphi^*_\beta + P(\beta c - \varphi^*_\beta)$.

As in the proof of Proposition 18 we suppose $X = \{0, ..., k\}$, we consider for each βc the function $F(\varphi) = -\int_X \varphi\, d\mu + P(\beta c + \varphi)$ and we see, by the same arguments, that $-\varphi^*_\beta$ is a minimizer of F, in particular $F(-\varphi^*_\beta) \leq F(0) = P(\beta c)$. Also, there exists $\epsilon > 0$ such that $F(t\varphi) > t\epsilon + \min(\beta c)$ for any $\varphi \in \mathbb{S}^{k-1}$. Therefore, if $t > \frac{P(\beta c) - \min(\beta c)}{\epsilon}$ then $F(t\varphi) > F(0) \geq F(-\varphi^*_\beta)$, for any $\varphi \in \mathbb{S}^{k-1}$. If we write $-\varphi^*_\beta = \|\varphi^*_\beta\| \tilde{\varphi}^*_\beta$, with $\tilde{\varphi}^*_\beta \in \mathbb{S}^{k-1}$, we see that $\|\varphi^*_\beta\| \leq \frac{P(\beta c) - \min(\beta c)}{\epsilon}$ and then $\frac{\|\varphi^*_\beta\|}{\beta} \leq \frac{P(\beta . c)}{\beta\epsilon} - \frac{\min(c)}{\epsilon}$. From the arguments above we know that $\frac{P(\beta c)}{\beta}$ converges to $\sup_{\pi\in\Pi(\cdot,\sigma)} \int_{X\times\Omega} c\, d\pi$. Then, there exists a constant K such that $\frac{\|\varphi^*_\beta\|}{\beta} \leq K$.

Now we claim that $\frac{\varphi_\beta}{\beta}$ is bounded. Indeed, as $\frac{\varphi_\beta}{\beta} = \frac{\varphi^*_\beta}{\beta} + \frac{P(\beta c - \varphi^*_\beta)}{\beta}$ we have

$$\frac{\|\varphi_\beta\|}{\beta} \leq \frac{2\,\|\varphi^*_\beta\|}{\beta} + \max c + \log(d) + \log(k+1) \leq K_2.$$

Using the estimative above and the fact that X is a finite set we see that the Lipschitz constant of $\frac{\varphi_\beta}{\beta}$ is uniformly bounded, hence the Lipschitz constant of $c - \frac{\varphi_\beta}{\beta}$ is bounded.

It follows that for some subsequence, there exists the limit of $\frac{\varphi_\beta}{\beta}$. In the same way we get a control of the Lipschitz constants of the eigenfunctions $h_\beta(y)$ to $L_{\beta c-\varphi_\beta}$. Applying the Arzela-Ascoli Theorem we obtain the existence of the limit on C^0 norm of $\frac{h_\beta}{\beta}$ for some subsequence $\beta_n \to \infty$.

Suppose that for some sequence β_n we have: $\frac{1}{\beta_n}\log(h_{\beta_n})(y) \to \tilde{V}(y)$, $\frac{\varphi_{\beta_n}(x)}{\beta_n} \to \tilde{m}(x)$ and $\pi_{\beta_n} \to \pi_\infty$. Then, $\pi_\infty \in \Pi(\mu,\sigma)$, and

$$\sup_x \sup_a [\, c(x,ay) + \tilde{V}(ay) - \tilde{V}(y) - \tilde{m}(x)] = 0, \quad \forall\, y. \tag{20}$$

On the other hand, from

$$\int_{X\times\Omega} \beta c(x,y) - \varphi_\beta(x)\, d\pi_\beta + H(\pi_\beta) = \sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} \beta c(x,y) - \varphi_\beta(x)\, d\pi + H(\pi),$$

we conclude that

$$\int_{X\times\Omega} c(x,y)\, d\pi_\infty - \int_X \tilde{m}(x)\, d\mu = \sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c(x,y)\, d\pi - \int_X \tilde{m}(x)\, d\mu.$$

Therefore,

$$\int_{X\times\Omega} c(x,y)\, d\pi_\infty = \sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c(x,y)\, d\pi,$$

which means that π_∞ is an optimal plan for c over $\Pi(\mu,\sigma)$.

Now we prove the duality between the primal and dual problem [7]:

Theorem 21. *Let Φ be the set of functions $\alpha(x)$, such that, there exists a function $S(y)$ satisfying: $c(x,y) + S(y) - S(\sigma(y)) - \alpha(x) \leq 0$, for any $x \in X$ and $y \in \Omega$. Then,*

$$\inf_{\alpha\in\Phi} \int_X \alpha(x)\, d\mu = \sup_{\pi\in\Pi(\mu,\sigma)} \int_{X\times\Omega} c(x,y)\, d\pi.$$

Moreover, suppose that for some sequence β_n we have: $\frac{1}{\beta_n}\log(h_{\beta_n}) \to \tilde{V}$, $\frac{\varphi_{\beta_n}}{\beta_n} \to \tilde{m}$ and $\pi_{\beta_n} \to \pi_\infty$. Then, the infimum on the left hand side is attained for $\alpha(x) = \tilde{m}(x)$ and $S(y) = \tilde{V}(y)$. The supremum on the right hand side is attained in π_∞. The function $\tilde{V}$ is a calibrated subaction.

Proof. Given α and S such that $c(x,y) + S(y) - S(\sigma(y)) - \alpha(x) \leq 0$, for any $x \in X, y \in \Omega$, we get the inequality

$$\int_{X\times\Omega} c(x,y)\, d\pi \leq \int_X \alpha(x)\, d\mu, \quad \forall \pi \in \Pi(\mu,\sigma).$$

Therefore,

$$\sup_{\pi\in\Pi(\mu,\sigma)}\int_{X\times\Omega} c(x,y)\,d\pi \le \int_X \alpha(x)\,d\mu.$$

On the other hand, from the Variational Principle for $\Pi(\mu,\sigma)$ we get the equation

$$\int_X \frac{\varphi_\beta(x)}{\beta}\,d\mu = \int_{X\times\Omega} c(x,y)\,d\pi_\beta + \frac{1}{\beta}H(\pi_\beta).$$

Then, when $\beta_n \to +\infty$, we have that $\tilde{m}(x)\in\Phi$ and

$$\int_X \tilde{m}(x)\,d\mu = \int_{X\times\Omega} c(x,y)\,d\pi_\infty = \sup_{\pi\in\Pi(\mu,\sigma)}\int_{X\times\Omega} c(x,y)\,d\pi.$$

□

5 Appendix

In the appendix we will give the proofs os some technical results.

Proof of Proposition 3. By the Schauder-Tychonov fixed point theorem we can find a plan $\pi_c \in P(X\times\Omega)$ such that $\hat{L}_c^*(\pi_c)=\pi_c$, for such normalized c. We note that $\pi_c\in\Pi(\cdot,\sigma)$ because

$$\begin{aligned}\int_{X\times\Omega}\psi\circ\sigma(y)\,d\pi_c &= \int_{X\times\Omega}\psi\circ\sigma(y)\,d\hat{L}_c^*(\pi_c)\\ = \int_{X\times\Omega}&\left(\sum_x\sum_{\sigma(w)=y} e^{c(x,w)}\psi\circ\sigma(w)\right)d\pi_c = \int_{X\times\Omega}\psi(y)\,d\pi_c\,.\end{aligned}$$

Now we show that the y-marginal of π_c is ν_c. Denote by $\tilde{\nu}_c$ the y-marginal of π_c. Then, for any fixed $\psi\in C(\Omega)$,

$$\begin{aligned}\tilde{\nu}_c(\psi) &= \pi_c(\psi)=\hat{L}_c^*(\pi_c)(\psi(y)) = \int_{X\times\Omega}\left(\sum_x\sum_{\sigma(w)=y} e^{c(x,w)}\psi(w)\right)d\pi_c\\ &= \int_\Omega\left(\sum_x\sum_{\sigma(w)=y} e^{c(x,w)}\psi(w)\right)d\tilde{\nu}_c = L_c^*(\tilde{\nu}_c)(\psi).\end{aligned}$$

Therefore, $\tilde{\nu}_c=\nu_c$, because ν_c is the unique fixed point of L_c^* .

The fixed point π_c for $\hat{L}_c^*$ is unique because it satisfies, for any $u(z,y)$,

$$\begin{aligned}\pi_c(u) &= \hat{L}_c^*(\pi_c)(u) = \int_{X\times\Omega}\Big(\sum_x\sum_{\sigma(w)=y} e^{c(x,w)}u(x,w)\Big)\,d\pi_c(z,y)\\ &= \int_{\Omega}\Big(\sum_x\sum_{\sigma(w)=y} e^{c(x,w)}u(x,w)\Big)\,d\nu_c(y).\end{aligned}$$

Finally, for a fixed (x_0,y_0) and an open set of the form (x_0,A), where A is a cylinder containing y_0 we have

$$\pi_c((x_0,A)) = \int_{\Omega}\Big(\sum_{\sigma(w)=y} e^{c(x_0,w)}\mathbf{I}_A(w)\Big)\,d\nu_c(y) > 0,$$

because ν_c is positive on cylinders and $e^{c(x_0,w)}$ is bounded below. In this way we show that the support of π_c is the full set $X\times\Omega$.

□

Now we will explaining how the Jacobian of a plan π is defined. We will adapt the reasoning of [10] to our setting.

Let $\mathcal{B}$ be the Borel sigma-algebra over $X\times\Omega$. Moreover, let $\sigma^{-1}(\mathcal{B})$ be the sigma algebra generated by cylinders of the form $[\cdot,\cdot\, y_1...y_n], n=1,2...$ where

$$[\cdot,\cdot\, y_1...y_n] = \{(x,(w_0,w_1,...))\in X\times\Omega : w_1=y_1,...,w_n=y_n\}.$$

Remember that for each $(x,a)\in X\times\{1,...,d\}$, $[x,a]=\{(z,(w_0,w_1,...)) : z=x, w_0=a\}$.

Given a plan π with y-marginal ν we define for each (x,a) the measure $\pi^{x,a}$ over $\sigma^{-1}(\mathcal{B})$ by the rule $\pi^{x,a}(A)=\pi([x,a]\cap A)$. Clearly $\pi^{x,a}\ll\pi$ then, from the Radon-Nikodym Theorem, there exists a function

$$E(\mathbb{I}_{[x,a]}\,|\,\sigma^{-1}(\mathcal{B})) := \frac{d\pi^{x,a}}{d\pi}\in L^1(X\times\Omega,\sigma^{-1}(\mathcal{B}),\pi),$$

which is the conditional expectation of $\mathbb{I}_{[x,a]}$ given $\sigma^{-1}(\mathcal{B})$.

In the same way, for each n, we consider $\mathcal{B}_n$ which is the smallest sigma-algebra containing the cylinders of the form $[\cdot,\cdot\, y_1...y_n]$. For each (x,a) let $\pi_n^{x,a}$ over $\mathcal{B}_n$ be defined by $\pi_n^{x,a}(A)=\pi([x,a]\cap A)$. Applying again the Radon-Nikodym Theorem we get a function

$$E(\mathbb{I}_{[x,a]}\,|\,\mathcal{B}_n) := \frac{d\pi_n^{x,a}}{d\pi}\in L^1(X\times\Omega,\mathcal{B}_n,\pi).$$

Note that

$$\int_{X\times\Omega} \mathbb{I}_{[\cdot,\cdot\, y_1...y_n]}\, d\pi_n^{x,a} = \int_{X\times\Omega} \mathbb{I}_{[x,ay_1...y_n]}\, d\pi = \pi([x, ay_1...y_n]),$$

and

$$\int_{X\times\Omega} \mathbb{I}_{[\cdot,\cdot\, y_1...y_n]}\, d\pi_n^{x,a} = \int_{X\times\Omega} \mathbb{I}_{[\cdot,\cdot\, y_1...y_n]} E(I_{[x,a]}\,|\,\mathcal{B}_n)\, d\pi.$$

Then, using the fact that $E(\mathbb{I}_{[x,a]}\,|\,\mathcal{B}_n)$ is constant on the set $[\cdot,\cdot\, y_1...y_n]$, we get

$$E(\mathbb{I}_{[x,a]}\,|\,\mathcal{B}_n)(x_0,(y_0,y_1,...,y_n,...)) = \frac{\pi([x,ay_1...y_n])}{\pi([\cdot,\cdot\, y_1...y_n])} = \frac{\pi([x,ay_1...y_n])}{\nu([y_1...y_n])}.$$

Let

$$J_\pi^n := \sum_{x,a} \mathbb{I}_{[x,a]} E(\mathbb{I}_{[x,a]}\,|\,\mathcal{B}_n),$$

then, for $(x_0,(y_0,y_1,...)) \in X \in \Omega$,

$$J_\pi^n(x_0,y) = \frac{\pi([x_0, y_0y_1...y_n])}{\nu([y_1...y_n])}.$$

From the increasing martingale theorem, when $n \to \infty$,

$$E(\mathbb{I}_{[x,a]}\,|\,\mathcal{B}_n) \to E(\mathbb{I}_{[x,a]}\,|\,\sigma^{-1}(\mathcal{B}))$$

in $L^1(X\times\Omega, \sigma^{-1}(\mathcal{B}), \pi)$ and in a.e. π. Then, by summing over (x,a) we get a function J_π well defined π a.e., such that,

$$J_\pi^n \to J_\pi$$

in $L^1(X\times\Omega, \mathcal{B}, \pi)$ and a.e. π. Note that:

$$J_\pi = \sum_{x,a} \mathbb{I}_{[x,a]} E(\mathbb{I}_{[x,a]}\,|\,\sigma^{-1}(\mathcal{B}))$$

in $L^1(X\times\Omega, \sigma^{-1}(\mathcal{B}), \pi)$.

Following the terminology of [9] and [10] we mention that the information function is defined by

$$I := -\log(J_\pi) = -\sum_{x,a} \mathbb{I}_{[x,a]} \log(E(\mathbb{I}_{[x,a]}\,|\,\sigma^{-1}(\mathcal{B}))).$$

In this case the entropy of π is

$$H(\pi) := \int_{X\times\Omega} -\log(J_\pi)\, d\pi = -\int_{X\times\Omega} \sum_{x,a} \mathbb{I}_{[x,a]} \log(E(\mathbb{I}_{[x,a]} \,|\, \sigma^{-1}(\mathcal{B})))\, d\pi.$$

The number $H(\pi)$ is finite.

This is the end of the basic considerations about the concepts of Jacobian and entropy of a plan.

Proof of Theorem 9. Let λ_c be the main eigenvalue and h_c the positive eigenfunction of L_c, given by Proposition 1, then $\bar{c}(x,y) := c(x,y) + \log(h_c)(y) - \log(h_c \circ \sigma)(y) - \log(\lambda_c)$ is the normalized cost associated to c. As h_c depends only on y, for any $\pi \in \Pi(\cdot,\sigma)$ we have that $-\int_{X\times\Omega} \bar{c}\, d\pi = -\int_{X\times\Omega} c\, d\pi + \log(\lambda_c)$. Hence, by the definition of entropy, we get

$$\begin{aligned} P(c) &= \sup_{\pi\in\Pi(\cdot,\sigma)} \left(\int_{X\times\Omega} c\, d\pi + H(\pi) \right) \\ &\leq \sup_{\pi\in\Pi(\cdot,\sigma)} \left(\int_{X\times\Omega} c\, d\pi - \int_{X\times\Omega} \bar{c}\, d\pi \right) = \log(\lambda_c). \end{aligned}$$

Now we show the other inequality: let $\pi_{\bar{c}}$ be the Gibbs plan to $\bar{c}$. Then, by Lemma 5, $H(\pi_{\bar{c}}) = -\int_{X\times\Omega} \bar{c}\, d\pi_{\bar{c}} = -\int_{X\times\Omega} c\, d\pi_{\bar{c}} + \log(\lambda_c)$.

Therefore,

$$P(c) = \sup_{\pi\in\Pi(\cdot,\sigma)} \left(\int_{X\times\Omega} c\, d\pi + H(\pi) \right) \geq \left(\int_{X\times\Omega} c\, d\pi_{\bar{c}} + H(\pi_{\bar{c}}) \right) = \log(\lambda_c).$$

In order to prove that the equilibrium plan is unique let us suppose that c is normalized. Then $P(c) = 0$ and for all $\pi \in \Pi(\cdot,\sigma)$ we have

$$\int_{X\times\Omega} c\, d\pi - \int_{X\times\Omega} \log(J_\pi)\, d\pi = \int_{X\times\Omega} c\, d\pi + H(\pi) \leq 0,$$

with equality, if and only if, $c = \log(J_\pi)$, by Lemma 6. Suppose now π is such that $\int_{X\times\Omega} c\, d\pi + H(\pi) = 0$. Using Lemma 4, for every $w \in C(X,\Omega)$,

$$\int_{X\times\Omega} \sum_x \sum_a J_\pi(x, ay) w(x, ay)\, d\pi = \int_{X\times\Omega} w\, d\pi,$$

hence, as $J_\pi = e^c$,

$$\int_{X\times\Omega} \sum_x \sum_a e^{c(x,ay)} w(x, ay)\, d\pi = \int_{X\times\Omega} \hat{L}_c(w) d\pi = \int_{X\times\Omega} w\, d\pi.$$

This shows that $\hat{L}_c^*(\pi) = \pi$. Finally, from the uniqueness of the Gibbs plan given by Proposition 3, we get that $\pi = \pi_c$.

□

Now we will prove some other results that we used before.

Proof of Lemma 4. We need to prove that, for every $w \in C(X, \Omega)$, $\pi \in \Pi(\cdot, \sigma)$,

$$\int_{X\times\Omega} \sum_x \sum_a J_\pi(x, ay) w(x, ay)\, d\pi = \int_{X\times\Omega} w\, d\pi.$$

First we show that if w is constant in the cylinders of the form $[x, y_0...y_n]$, then

$$\int_{X\times\Omega} \sum_x \sum_a J_\pi^n(x, ay) w(x, ay)\, d\pi = \int_{X\times\Omega} w\, d\pi.$$

Consider a function $w_n = \mathbb{I}_{[i, j_0 j_1 ... j_n]}$. Then,

$$\begin{aligned}
&\int_{X\times\Omega} \sum_x \sum_a J_\pi^n(x, ay) w_n(x, ay)\, d\pi \\
&= \int_{X\times\Omega} \sum_x \sum_a \frac{\pi([x, ay_0...y_{n-1}])}{\nu([y_0...y_{n-1}])} \mathbb{I}_{[i, j_0 j_1 ... j_n]}(x, ay)\, d\nu(y) \\
&= \pi([i, j_0 j_1 ... j_n]) = \int_{X\times\Omega} w_n\, d\pi.
\end{aligned}$$

From linearity arguments we conclude the first part of the Lemma.

In order to prove the second part of the Lemma we take a function $w_l = \mathbb{I}_{[i, j_0 j_1 ... j_l]}$. Then, using the first part of the Lemma, we obtain

$$\begin{aligned}
\int_{X\times\Omega} \sum_x \sum_a J_\pi(x, ay) w_l(x, ay)\, d\pi &= \lim_{n\to\infty} \int_{X\times\Omega} \sum_x \sum_a J_\pi^n(x, ay) w_l(x, ay)\, d\pi \\
&= \int_{X\times\Omega} w_l\, d\pi,
\end{aligned}$$

where we use that, if $n \geq l$, w_l is also constant in the cylinder of the form $[x, y_0...y_n]$.

From linearity arguments and using the fact that the functions which are constant in cylinders of length $l = 1, 2, 3, ...$ are dense in $C(X, \Omega)$ we conclude the proof.

□

The Proof of Lemma 6 will require the following:

Lemma 22. *If b is a normalized potential which is constant on cylinders of the form $[x, y_0...y_n]$, then*

$$-\int_{X\times\Omega} \log(J_\pi^n)\, d\pi \leq -\int_{X\times\Omega} b\, d\pi.$$

Furthermore, there exists a family of normalized potentials b_ϵ such that

$$-\int_{X\times\Omega} \log(J_\pi^n)\, d\pi = \lim_{\epsilon\to 0} -\int_{X\times\Omega} b_\epsilon\, d\pi.$$

Proof. Let us fix a normalized potential b constant on cylinders of the form $[x, y_0...y_n]$. The functions $u = \frac{e^b}{J_\pi^n}$ and $\log(J_\pi^n)$ are well defined in $\text{supp}(\pi)$. Using Jensen inequality, we have

$$\begin{aligned} 0 &= \int_{X\times\Omega} \log\left(\sum_x \sum_a e^{b(x,ay)}\right) d\pi = \int_{X\times\Omega} \log\left(\sum_x \sum_a J_\pi^n(x,ay)u(x,ay)\right) d\pi \\ &\geq \int_{X\times\Omega} \left(\sum_x \sum_a J_\pi^n(x,ay)\log(u(x,ay))\right) d\pi = \int_{X\times\Omega} \log(u(z,y))\, d\pi \\ &= \int_{X\times\Omega} b(z,y) - \log(J_\pi^n(z,y))\, d\pi. \end{aligned}$$

This shows the first part of the lemma.

In order to show the second part, we consider for each cylinder $[y_1...y_n]$, such that, $\nu([y_1...y_n]) > 0$, the sets
$A = A_{[y_1...y_n]} := \{(x,a) \in X \times \{1,...,d\} \,:\, \pi([x, ay_1...y_n]) = 0\}$ and
$B = B_{[y_1...y_n]} := \{(x,a) \in X \times \{1,...,d\} \,:\, \pi([x, ay_1...y_n]) > 0\}$.

Fixed $\epsilon > 0$ sufficiently small, consider the potential b_ϵ defined by

$$b_\epsilon([x, ay_1...y_n]) = \begin{cases} \log((\#B)\,\epsilon), & if\ (x,a) \in A \\ \log(J_\pi^n([x, ay_1...y_n]) - (\#A)\,\epsilon), & if\ (x,a) \in B \end{cases}$$

If $\nu([y_1, ..., y_n]) = 0$, we define $b_\epsilon([x, ay_1...y_n]) = -\log((\#X)\,d)$, for all $x \in X, a \in \{1,...,d\}$. By construction we see that b_ϵ is a normalized potential. Indeed, take $z = (z_1, z_2, ...) \in \Omega$, there are two cases:
If $\nu([z_1...z_n]) = 0$, then

$$\sum_x \sum_a e^{b_\epsilon(x,az)} = \sum_x \sum_a e^{b_\epsilon([x,az_1...z_n])} = \sum_x \sum_a e^{-\log((\#X)\,d)} = 1.$$

If $\nu([z_1...z_n]) > 0$, then

$$\sum_x \sum_a e^{b_\epsilon(x,az)} = \sum_A e^{\log((\#B)\,\epsilon)} + \sum_B e^{\log(J^n_\pi([x,az_1...z_n])-(\#A)\,\epsilon)}$$
$$= (\#A)(\#B)\,\epsilon + \left(\sum_B J^n_\pi([x,az_1...z_n])\right) - (\#B)(\#A)\,\epsilon$$
$$= \sum_B J^n_\pi([x,az_1...z_n]) = 1.$$

When $\epsilon \to 0$,

$$\begin{aligned}\lim_{\epsilon\to 0}\int_{X\times\Omega} b_\epsilon\,d\pi &= \lim_{\epsilon\to 0}\sum_{\nu([y_1...y_n])>0}\ \sum_{B_{[y_1...y_n]}} b_\epsilon([x,y_0y_1...y_n])\pi([x,y_0y_1...y_n]) \\ &= \int_{X\times\Omega}\log(J^n_\pi)\,d\pi.\end{aligned}$$

□

Proof of Lemma 6. First, we need to prove that, if b is a normalized potential and $\pi \in \Pi(\cdot,\sigma)$, then

$$0 \le -\int_{X\times\Omega}\log(J_\pi)\,d\pi \le -\int_{X\times\Omega} b\,d\pi,$$

with equality, if and only if, $b = \log(J_\pi)$.

We know that $\log(J^n_\pi)$ converges to $\log(J_\pi)$ a.e. (π). Following the Lemma 8.11 and Theorem 8.12 in [10] we conclude that $\log(J^n_\pi)$ converges to $\log(J_\pi)$ in L^1 norm.

We claim that, for all b normalized

$$-\int_{X\times\Omega}\log(J_\pi)\,d\pi \le -\int_{X\times\Omega} b\,d\pi.$$

Indeed, note that functions $u = \frac{e^b}{J_\pi}$ and $\log(J_\pi)$ are well defined a.e π and that log is a strictly concave function, then by Jensen inequality we have

$$\begin{aligned}0 &= \int_{X\times\Omega}\log\left(\sum_x\sum_a e^{b(x,ay)}\right)\,d\pi = \int_{X\times\Omega}\log\left(\sum_x\sum_a J_\pi(x,ay)u(x,ay)\right)\,d\pi \\ &\ge \int_{X\times\Omega}\left(\sum_x\sum_a J_\pi(x,ay)\log(u(x,ay))\right)\,d\pi = \int_{X\times\Omega}\log(u(z,y))\,d\pi \\ &= \int_{X\times\Omega} b(z,y) - \log(J_\pi(z,y))\,d\pi,\end{aligned}$$

proving the claim.

In the case $\int_{X\times\Omega} b(z,y)-\log(J_\pi(z,y))\,d\pi = 0$ we get that for π a.e. y, the Jensen's inequality will be an equality [3]. Then, for π a.e. y we have $u(x,ay)$ is constant equal to 1 . That is, for almost all y we have that for any x and a the equality $\log J_\pi(x,ay) = b(x,ay)$ hold. Using Lemma 4 it follows that π is the Gibbs plan for b, because of the uniqueness assertion of Proposition 3. From Lemma 5 we get that $J_\pi = e^b$, hence $\log J_\pi = b$.

Now, the final claim of Lemma 6,

$$-\int_{X\times\Omega} \log(J_\pi)\,d\pi = \inf_{b\,normalized} -\int_{X\times\Omega} b\,d\pi.$$

is a consequence of the second part from Lemma 22.

□

Proof of Proposition 10. We need to prove that, given $\pi \in \Pi(\cdot,\sigma)$, if the x-marginal of π is a probability measure μ and the y-marginal of π is an invariant measure ν, then

$$H(\pi) \leq h(\mu) + h(\nu).$$

Moreover, if $\pi = \mu\times\nu$, then $H(\pi) = h(\mu) + h(\nu)$.

We remember that the Kolmogorov entropy of ν satisfies

$$h(\nu) = \inf\left\{-\int_\Omega g(y)\,d\nu(y)\ :\ g\,\text{Lipschitz and}\sum_{\sigma(w)=y} e^{g(w)} = 1\right\}.$$

Given $\epsilon > 0$, let g be a Lipschitz function satisfying $\sum_{\sigma(w)=y} e^{g(w)} = 1$ and $-\int_\Omega g(y)\,d\nu(y) < h(\nu)+\epsilon$. The potential $c(x,y) = g(y) + \log(\mu(x))$ is normalized and therefore

$$\begin{aligned} H(\pi) &\leq -\int_{X\times\Omega} g(y) + \log(\mu(x))\,d\pi = -\int_\Omega g(y)\,d\nu(y) - \int_X \log(\mu(x))\,d\mu(x) \\ &\leq h(\nu) + h(\mu) + \epsilon. \end{aligned}$$

Taking $\epsilon \to 0$, we conclude the first part of the proof.

---

[3] $0 = \log\left(\sum_x\sum_a J_\pi(x,ay)u(x,ay)\right) \geq \sum_x\sum_a J_\pi(x,ay)\log(u(x,ay))$ is an equality iff $u(x,ay) = k(y)$ for all $x\in X, a\in\{1,...,d\}$, hence $0=\log(k(y))$ this implies $k(y)=1$.

Now we suppose that $\pi = \mu \times \nu$. Then

$$\begin{aligned}
H(\pi) &= \lim_{n\to\infty} -\int_{X\times\Omega} \log\left(\frac{\pi([x, y_0...y_n])}{\nu([y_1...y_n])}\right) d\pi(x,y) \\
&= \lim_{n\to\infty} -\int_{X\times\Omega} \log(\mu(x)) + \log\left(\frac{\nu([y_0...y_n])}{\nu([y_1...y_n])}\right) d\pi(x,y) \\
&= -\sum_x \log(\mu(x))\mu(x) - \lim_{n\to\infty}\int_\Omega \log\left(\frac{\nu([y_0...y_n])}{\nu([y_1...y_n])}\right) d\nu(y) \\
&= h(\mu) + h(\nu).
\end{aligned}$$

□

Now we will show the last proof that was missing.

Proof of Proposition 12. Itens (a) and (b) are straightforward.

To prove item (c), i.e. the convexity of the pressure map, we suppose that $c = \lambda c_1 + (1-\lambda)c_2$. Given $\epsilon > 0$, there exists π_ϵ, such that,

$$\begin{aligned}
P(c) - \epsilon &\leq \int_{X\times\Omega} c\, d\pi_\epsilon + H(\pi_\epsilon) \\
&= \lambda\left(\int_{X\times\Omega} c_1 d\pi_\epsilon + H(\pi_\epsilon)\right) + (1-\lambda)\left(\int_{X\times\Omega} c_2 d\pi_\epsilon + H(\pi_\epsilon)\right) \\
&\leq \lambda P(c_1) + (1-\lambda)P(c_2).
\end{aligned}$$

In order to prove item (d), for a given $\epsilon > 0$, there exists π_ϵ, such that,

$$\begin{aligned}
P(c_1) - P(c_2) &\leq \int_{X\times\Omega} c_1 d\pi_\epsilon + H(\pi_\epsilon) + \epsilon - P(c_2) \\
&\leq \int_{X\times\Omega} c_1 d\pi_\epsilon + H(\pi_\epsilon) + \epsilon - \left(\int_{X\times\Omega} c_2 d\pi_\epsilon + H(\pi_\epsilon)\right) \\
&\leq \int_{X\times\Omega} |c_1 - c_2| d\pi_\epsilon + \epsilon \\
&\leq \|c_1 - c_2\| + \epsilon.
\end{aligned}$$

□

References

[1] A. T. Baraviera, L. Cioletti, A. O. Lopes, J. Mohr and R. R. Souza, On the general one-dimensional XY model: positive and zero temperature, selection and non-selection. *Rev.Math. Phis.* 23,no. 10, 1063-1113, 82Bxx, 2011.

[2] A. Baraviera, R. Leplaideur and A. O. Lopes, Ergodic Optimization, Zero temperature limits and the Max-Plus Algebra, mini-course in XXIX Colóquio Brasileiro de Matemática - IMPA - Rio de Janeiro (2013)

[3] E. Garibaldi and A. Lopes, Functions for relative maximization, Dynamical Systems, v. 22, 511–528, 2007.

[4] T. Kucherenko and C. Wolf, Geometry and entropy of generalized rotation sets, Arxiv (2012)

[5] S. Lalley, Distribution of periodic orbits of symbolic and Axiom A flows. Adv. in Appl. Math. 8, no. 2, 154–193, 1979.

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

[7] A. Lopes and J. Mengue, Duality Theorems in Ergodic Transport, Journal of Statistical Physics. Vol 149, issue 5, pp 921-942, 2012.

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

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

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

[11] F. Przytycki and M. Urbanski, *Conformal Fractals*, Cambridge Press, 2010.

[12] E. A. da Silva, R. R. da Silva and R. R. Souza The Analyticity of a Generalized Ruelle's Operator, to appear in *Bull. Braz. Math. Soc.*

[13] D. Ruelle, *Thermodynamic Formalism*, Addison Welsey, 1978.

[14] C. Villani, *Topics in optimal transportation*, AMS, Providence, 2003.