The Dual Potential, the Involution Kernel and Transport in Ergodic Optimization

A.O. Lopes, E.R. Oliveira, and Ph. Thieullen

Abstract Consider the shift σ acting on the Bernoulli space $\Sigma = \{1, 2, \ldots, n\}^{\mathbb{N}}$. We denote $\hat{\Sigma} = \{1, 2, \ldots, n\}^{\mathbb{Z}} = \Sigma \times \Sigma$. We analyze several properties of the maximizing probability $\mu_{\infty,A}$ of a Hölder potential $A : \Sigma \to \mathbb{R}$. Associated to $A(x)$, via the involution kernel, $W(x, y)$, $W : \hat{\Sigma} \to \mathbb{R}$, one can get the dual potential $A^*(y)$, where $(x, y) \in \hat{\Sigma}$. We denote μ_{∞,A^*} the maximizing probability for A^*. We would like to consider the transport problem from $\mu_{\infty,A}$ to μ_{∞,A^*}. In this case, it is natural to consider the cost function $c(x, y) = I(x) - W(x, y) + \gamma$, where I is the deviation function for $\mu_{\infty,A}$, as the limit of Gibbs probabilities $\mu_{\beta A}$ for the potential βA when $\beta \to \infty$. The value γ is a constant which depends on A. We could also take $c = -W$ above. We denote by $\mathcal{K} = \mathcal{K}(\mu_{\infty,A}, \mu_{\infty,A^*})$ the set of probabilities $\hat{\eta}(x, y)$ on $\hat{\Sigma}$, such that $\pi_x^*(\hat{\eta}) = \mu_{\infty,A}$, and $\pi_y^*(\hat{\eta}) = \mu_{\infty,A^*}$. We describe the minimal solution $\hat{\mu}$ (which is invariant by the shift on $\hat{\Sigma}$) of the Transport Problem, that is, the solution of

$$\inf_{\hat{\eta} \in \mathcal{K}} \int \int c(x, y)\, d\hat{\eta} = -\max_{\hat{\eta} \in \mathcal{K}} \int \int (W(x, y) - \gamma)\, d\hat{\eta}.$$

The optimal pair of functions for the Kantorovich Transport dual Problem is $(-V, -V^*)$, where we denote the two calibrated sub-actions by V and V^*, respectively, for A and A^*. We show that the involution kernel W is cyclically monotone. In other words, satisfies a twist condition in the support of $\hat{\mu}$. We analyze the question: is the support of $\hat{\mu}$ a graph? We also investigate the question of finding an explicit expression for the function $f : \Sigma \to \mathbb{R}$ whose c-subderivative determines the graph. We also analyze the same kind of problem for expanding transformations on the circle.

A.O. Lopes • E.R. Oliveira (✉)
Instituto de Matemática-UFRGS, Avenida Bento Gonçalves, 9500 Porto Alegre, RS, Brazil
e-mail: alopes@mat.ufrgs.br; oliveira.elismar@gmail.com

Ph. Thieullen
Institut de Mathématiques, Université Bordeaux, 33405 Talence, France
e-mail: philippe.thieullen@math.u-bordeaux1.fr

J.-P. Bourguignon et al. (eds.), *Dynamics, Games and Science*, CIM Series in Mathematical Sciences 1, DOI 10.1007/978-3-319-16118-1_20

1 Introduction

It seems natural to try to investigate the connections of Transport Theory with Ergodic Theory. Some results on this direction appear in [18, 32–34, 51]. Here we follow a different path.

Given a continuous function $A : \Sigma = \{1, 2, 3, .., d\}^{\mathbb{N}} \to \mathbb{R}$, we call $\mu_{\infty,A}$ a maximizing probability for A, if $\int A d\nu$ attains the maximal value in $\mu_{\infty,A}$, when the probabilities ν range among the set of invariant for the shift acting on the Bernoulli space Σ. We denote by $m(A)$ this maximal value.

Such maximizing probabilities $\mu_{\infty,A}$ can be seen as the equilibrium states at zero temperature for a system on the one dimensional lattice $\mathbb{N}$ with d spins in each site and under the influence of an interacting potential A (see [5, 8, 12, 14, 27, 35, 42, 46]).

A main conjecture on the area claims that for a generic Hölder potential A the maximizing probability has support in a unique periodic orbit for the shift (for a partial result see [12]). This conjecture was recently proved by G. Contreras (see [10]).

We address the question of finding the optimal transport plan from a certain maximizing probability to another. More precisely, we would like to consider the transport problem from $\mu_{\infty,A}$ to μ_{∞,A^*}, where $A : \Sigma = \{1, 2, 3, .., d\}^{\mathbb{N}} \to \mathbb{R}$ is a Hölder potential and A^* its dual (see [2]).

We consider here that A acts on the variable x and A^* in the variable y. A function $W(x, y)$ called the involution kernel will play an important role in the theory. The twist condition for W is a kind of convexity assumption. We will describe bellow with all details the setting we are going to consider in the present paper. We will also provide several examples to illustrate the theory.

We assume here in most (but not all) of the results that the maximizing probability $\mu_{\infty,A}$ (on Σ) for A is unique.

We denote by $\hat{\mu}$ the minimizing probability over $\hat{\Sigma} = \{1, 2, 3, .., d\}^{\mathbb{Z}} = \Sigma \times \Sigma$, for the natural Kantorovich Transport Problem associated to the $-W$, where $W(x, y)$, for $(x, y) \in \Sigma \times \Sigma$, is the involution kernel associated to A (see [2]).

We will denote by $\hat{\sigma}$ the shift on $\hat{\Sigma}$. The probability $\hat{\mu}_{max}$ denotes the natural extension of $\mu_{\infty,A}$ as described in [2].

We point out that by its very nature the Classical Transport Theory is not a Dynamical Theory (in the sense of considering invariant probabilities) [48, 53, 54]. One has to consider a cost which is obtained from dynamical properties in order to get optimal plans which are invariant for $\hat{\sigma}$.

Recent results in Ergodic Transport are [13, 22, 36, 37, 41, 44].

We will consider a cost which is the involution kernel W. First we show that:

Theorem 1 *The minimizing Kantorovich probability $\hat{\mu}$ on $\hat{\Sigma}$ associated to $-W$, where W is the involution kernel for A, is $\hat{\mu}_{max}$. Same property is true for c instead of W*

One of our main results is Theorem 5 which claims that the support of $\hat{\mu}_{max}$ is W-cyclically monotone. We do not assume the twist condition in the above result.

The calibrated subactions V play an important role in Ergodic Optimization. They can help to find the support of the maximizing probability (see [5, 27] or [12] for instance). Moreover, if we denote $R(x) = V(\sigma(x)) - V(x) - A(x) + m(A)$, then $I(x) = \sum_{n\geq 0} R(\sigma^n(x))$ defines a nonnegative lower semicontinuous function (can be infinite at several points) which is the deviation function for the family of Gibbs states associated to A when the temperature converges to zero [2] (see [3, 36] for the case of the XY model). For a class of explicit nontrivial examples of subactions V see [4].

Theorem 2 *If V is the calibrated subaction for A, and V^* is the calibrated subaction for A^*, then, the pair $(-V, -V^*)$ is the dual $(-W + I)$-Kantorovich pair of $(\mu_{\infty,A}, \mu_{\infty,A^*})$, when I is the deviation function for A.*

Finding the optimal transport measure between two probabilities is the solution of the so called relaxed problem [53]. If we want to find a measurable transformation (the Monge problem) which transfers one probability to another we need to show that the graph property is true in the support of such probability (which does not always happen if one considers a general cost function) [53].

Finally, we analyze later here the graph property for the support of the $\hat{\mu}_{max}$ (over $\hat{\Sigma} = \{1, 2, 3, .., d\}^{\mathbb{Z}}$) which is the minimizing probability for the cost function $-W$.

One can consider in the Bernoulli space $\Sigma = \{0, 1\}^{\mathbb{N}}$ the lexicographic order. In this way, $x < z$, if and only if, the first element i such that, $x_j = z_j$ for all $j < i$, and $x_i \neq z_i$, satisfies the property $x_i < z_i$. Moreover, $(0, x_1, x_2, \ldots) < (1, x_1, x_2, \ldots)$.

One can also consider the more general case $\Sigma = \{0, 1, \ldots, d-1\}^{\mathbb{N}}$, but in order to simplify the notation and to avoid technicalities, we consider only the case $\Sigma = \{0, 1\}^{\mathbb{N}}$.

Definition 1 We say a continuous $G : \hat{\Sigma} = \Sigma \times \Sigma \to \mathbb{R}$ satisfies the twist condition on $\hat{\Sigma}$, if for any $(a, b) \in \hat{\Sigma} = \Sigma \times \Sigma$ and $(a', b') \in \Sigma \times \Sigma$, with $a' > a$, $b' > b$, we have

$$G(a, b) + G(a', b') < G(a, b') + G(a', b). \tag{1}$$

The twist condition is inspired in the Aubry-Mather Theory [1, 11, 23–25]. It is a quite natural concept in Classical Optimization and Transport Theory [6, 13, 15, 40, 45, 48, 53, 54] (see [37] for dynamical examples).

The twist condition is also described by the concept of **global** cyclically monotonicity (see [53])

We point out that in Mather Theory in order to have the graph property (see [11, 43]) for the minimal action measure it is necessary to assume that Lagrangian is convex in the velocity. We need in our setting some technical assumptions to replace this important property. We believe that the twist condition is the natural one.

Definition 2 We say a continuous $A : \Sigma \to \mathbb{R}$ satisfies the twist condition, if its involution kernel W satisfies the twist condition.

The involution kernel of A is not unique (see [2]), but if the above property is true for some W, then it will also be true for any other one.

Our final result is:

Theorem 3 *Suppose the involution kernel W satisfies the twist condition on $\hat{\Sigma}$, then, the support of $\hat{\mu}_{max} = \hat{\mu}$ on $\hat{\Sigma}$ is a graph. Moreover, if $d = 2$, then there exists at most one point in the support of $\hat{\mu}$ which has two points in the support of $\hat{\mu}$ in its vertical fiber. The σ orbit of this point is a zero measure set.*

There are examples where the existence of this exceptional point occurs and this is associated to the concept of turning point (see [13, 37, 40]).

Similar results occur for the case of a general d. A similar definition can be considered for an expanding transformation on [0, 1], and we are also able to get the analogous graph property result. This also includes the case of $T(x) = -2x$ (mod 1).

We present in the Appendix at the end of the paper several examples (and computations) where one can write the involution kernel W explicitly and the twist condition is satisfied. First we will explain all the preliminaries we will need later.

Consider X a compact metric space. Given a continuous transformation $f : X \to X$, we denote by $\mathcal{M}_f$ the convex set of f-invariant Borel probability measures. As usual, we consider in $\mathcal{M}_f$ the weak* topology. The standard model used in ergodic optimization is the triple $(X, f, \mathcal{M}_f)$. Given a potential $A \in C^0(X)$, we denote

$$m(A) = \max_{\nu \in \mathcal{M}_f} \int_X A(x)\, d\nu(x). \tag{2}$$

We are interested here in the characterization and main properties of A-maximizing probabilities, that is, the probabilities belonging to the set

$$\{\ \mu \in \mathcal{M}_f : \int_X A(x)\, d\mu(x) = m(A)\ \}. \tag{3}$$

We will assume here that A is Hölder.

In the following we will also assume that the maximizing probability $\mu_{\infty,A} = \mu_\infty$ is unique.

Under reasonable hypothesis (expanding, hyperbolic, etc.) several results were obtained related to this maximizing question, among them [2, 5, 7–9, 12, 14, 23, 24, 26–28, 35, 38, 46, 50, 52]. For maximization with constraints see [20, 39]. Questions related to the dynamics on the boundary of the fat attractor appear in [37]. Naturally, if we change the maximizing notion for the minimizing one, the analogous properties will also be true.

Our focus here will be mainly on symbolic dynamics and on expanding transformations on S^1 or the interval $[0, 1]$. We recall some basic definitions (see [5] or [12] for example).

Let $\sigma : \Sigma \to \Sigma$ be a subshift of finite type defined by a matrix C of 0 and 1, where $\sigma(x_0, x_1, x_2, ..) = (x_1, x_2, x_3, ..)$. In this case we are considering $X = \Sigma =$

$\{1, 2, 3, .., d\}_C^{\mathbb{N}}$ and $f = \sigma$. Remind that, for a fixed $\lambda \in (0, 1)$, we consider for Σ the metric $d(\mathbf{x}, \bar{\mathbf{x}}) = \lambda^k$, where $\mathbf{x} = (x_0, x_1, \ldots), \bar{\mathbf{x}} = (\bar{x}_0, \bar{x}_1, \ldots) \in \Sigma$ and $k = \min\{j : x_j \neq \bar{x}_j\}$. In this situation, given a Hölder potential $A : \{1, 2, 3, .., d\}^{\mathbb{N}} \to \mathbb{R}$, one should be interested in A-maximizing probabilities for the triple $(\Sigma, \sigma, \mathscr{M}_\sigma)$, where the probabilities are consider over $\mathscr{B}$, the σ-algebra of Borel of Σ. In order to simplify the notation here we will consider the full Bernoulli space (all entries of C are equal to 1).

Given an $C^{1+\alpha}$ expanding transformation T of fixed degree on S^1 and $A : S^1 \to \mathbb{R}$ we will be interested in A- maximizing probabilities on $(S^1, T, \mathscr{M}_T)$, where the probabilities are consider over $\mathscr{B}$, the σ-algebra of Borel of S^1.

One can consider the analogous setting for $C^{1+\alpha}$ expanding transformations of fixed degree over $[0, 1]$.

Convex potentials $A : [0, 1] \to \mathbb{R}$ and the transformation $T : [0, 1] \to [0, 1]$, given by $T(x) = 2\,x$ (mod 1), were considered in [29] where it was shown that the maximizing probabilities in this case are Sturmian measures. For $T(x)$ equal to $-2\,x$ (mod 1) however, the situation is completely different (see [31]).

Definition 3 A function $u \in C^0(\Sigma)$ is a sub-action for the potential A if, for any $\mathbf{x} \in \Sigma = \{1, 2, 3, \ldots, d\}_C^{\mathbb{N}}$, we have

$$u(\mathbf{x}) \leqslant u(\sigma(\mathbf{x})) - A(\mathbf{x}) + \beta_A. \tag{4}$$

Let (Σ^*, σ^*) be the dual subshift.

In the case of the full Bernoulli space (all entries of C equal 1) then $\Sigma^* = \{1, 2, 3, .., d\}^{\mathbb{N}}$ and $\sigma^*(y_0, y_1, y_2, \ldots) = (y_1, y_2, \ldots)$.

We consider the space of the dynamics $(\hat{\Sigma}, \hat{\sigma})$, the natural extension of (Σ, σ), as subset of $\Sigma^* \times \Sigma$. In fact, if $\mathbf{y} = (\ldots, y_1, y_0) \in \Sigma^*$ and $\mathbf{x} = (x_0, x_1, \ldots) \in \Sigma$, then $\hat{\Sigma}$ will be the set of points

$$< y, x >= (\ldots, y_1, y_0 | x_0, x_1, \ldots) \in \Sigma^* \times \Sigma,$$

such that (y_0, x_0) is an allowed word (no restrictions when we consider the full Bernoulli space). In this case

$$\hat{\sigma}\,(\ldots, y_1, y_0 | x_0, x_1, \ldots) \;=\; (\ldots, y_1, y_0, x_0 | x_1, x_2, \ldots).$$

We point out that we use here the notation $< y, x >= (x, y)$. For functions $b : \hat{\Sigma} \to \mathbb{R}$, we denote its value on $< y, x >$ by $b(x, y)$. We define the map $\tau : \hat{\Sigma} \to \Sigma$ by $\tau(x, y) = \tau_{\mathbf{y}}(\mathbf{x}) = (y_0, x_0, x_1, \ldots)$. Note that, if $\pi_x : \hat{\Sigma} \to \Sigma$ is the projection in the x coordinate, then, $\tau_y(x) = \pi_x \circ \hat{\sigma}^{-1}\,(x, y)$. We denote by $\pi_y(x, y) = y$ the projection on the second coordinate. Note that $\hat{\sigma}^{-1}(x, y) = (\tau_y(x), \sigma^*(y))$.

Definition 4 A continuous function $V : \Sigma \to \mathbb{R}$ is called calibrated subaction for A, if

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

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

Note that for all z we have $V(\sigma(z)) - V(z) - A(z) + m(A) \geq 0$. We show bellow some explicit expressions for calibrated subactions for a class of potentials A.

We point out that we will also consider here analogous results for an expanding transformation $T : S^1 \to S^1$ (or, $T : [0,1] \to [0,1]$) of class $C^{1+\alpha}$, and a Hölder potential $A : S^1 \to \mathbb{R}$ (or, $A : [0,1] \to \mathbb{R}$) as in [12]. The case $T(x) = -2x$ (mod 1) is one of the examples we have on mind.

In this case one could consider analogous problems in $S^1 \times S^1$, or, $S^1 \times \Sigma$, if one consider the symbols i which index the inverse branches τ_i of T [37, 40]. The existence of involution kernel, L.D.P. properties, etc., are also true.

The calibrated sub-action is unique (up to an additive constant) if the maximizing probability is unique (see [2, 12, 21]). We point out that we called strict in [2] what we denote here by calibrated. We will use from now on the notation of [2].

Definition 5 Given $A : \Sigma \to \mathbb{R}$ Lipchitz potential, consider $A^*(y)$ (the dual potential), where $A : \Sigma^* \to \mathbb{R}$, and $W(x,y) = W_A(x,y)$ its involution kernel.

This means, by definition that for all $< y, x >= (x,y) \in \hat{\Sigma}$

$$A^*(y) = A(\tau_y(x)) + W(\tau_y(x), \sigma^*(y)) - W(x,y). \tag{5}$$

This expression can be also written in the form

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

If A depends on just two coordinates we can take A^* as the transpose of A. Therefore, the above definition extends this concept in the case A depends on infinite coordinates on the Bernoulli space. We say A is involutive if $A = A^*$.

We address the question of regularity of the involution kernel W (is bi-Hölder) in the item (d) in the Appendix.

We denote by M the Bernoulli space or the unitary circle. Suppose T is an expanding transformation on M (T can be the shift σ or the transformation T defined above).

For a Lipchitz potential $A : M \to \mathbb{R}$ the pressure of A is the value

$$P(A) = \sup_{\mu \text{ invariant for } T} \{h(\mu) + \int A \, d\mu\},$$

where $h(\mu)$ is the Kolmogorov entropy of the invariant probability μ.

The equilibrium state for A is the probability μ which realizes the above supremum.

Given a Hölder function $A : M \to \mathbb{R}$, by definition the Ruelle operator $\mathscr{L}_A : C(M) \to C(M)$ acts on continuous functions $\phi : M \to \mathbb{R}$, in such way that, $\mathscr{L}_A(\phi) = \varphi$, where

$$\varphi(x) = \mathscr{L}_A(\phi)(x) = \sum_{T(y)=x} e^{A(y)}\, \phi(y).$$

This operator (sometimes called transfer operator) helps to understand equilibrium states in Thermodynamic Formalism. This corresponds to the analysis of the Statistical Mechanics of the one-dimensional lattice at positive temperature (see [47]). Maximizing probabilities correspond to the limit of equilibrium states when temperature goes to zero (ground states) as one can see for instance in [5].

When A is such that $\mathscr{L}_A(1) = 1$ we say that A is normalized.

The dual operator $\mathscr{L}_A^*$ acts on the space of probabilities measures on M. Given a probability μ, then, $\mathscr{L}_A^*(\mu) = \nu$ where the probability measure ν is the unique one satisfying

$$\int \phi \; d\, \mathscr{L}_A^*\,(\mu) = \int \phi\, d\nu = \int \mathscr{L}_A\,(\phi)\, d\mu$$

for any continuous function ϕ.

An important result claims that there exists a positive value λ which is simultaneous an eigenvalue for $\mathscr{L}_A$ and $\mathscr{L}_A^*$ (see [47]). This λ is the spectral radius of $\mathscr{L}_A$. This defines a main eigenfunction for $\mathscr{L}_A$ and a main eigenprobability for $\mathscr{L}_A^*$.

In [33] it is shown that the dual of the Ruelle operator $\mathscr{L}_A^*$ is a contraction for the 1-Wasserstein distance when A is normalized. The fixed point probability is the main eigenprobability for $\mathscr{L}_A^*$.

We suppose that c is a normalization constant for W in the sense that

$$\int \int e^{W(x,y)-c}\, d\nu_{A^*}(y)\, d\nu_A(x) \;=\; 1, \tag{6}$$

where ν_A and ν_{A^*} are respectively the eigen-probability for the dual Ruelle operator of A and A^* [12]. We also denote by ϕ_A and ϕ_{A^*} the corresponding eigenfunctions for $\mathscr{L}_A$. Finally, $\mu_A = \nu_A\, \phi_A =$ and $\mu_{A^*} = \nu_{A^*}\, \phi_{A^*}$ are the invariant probabilities which are the solutions of the respective pressure problems for A and A^*. For a fixed A we consider a real parameter β, and the corresponding potentials βA, and the eigenfunctions $\phi_{\beta A}$, and so on.

In Statistical Mechanics β is the inverse of temperature. In this way asymptotic results when $\beta \to \infty$ can be consider as the ones which describes the system in equilibrium at temperature zero. Note that βW is an involution kernel for βA, and its dual is βA^*.

It is known (see for instance [12]) that a sub-action V can obtained as the limit

$$V(x) = \lim_{\beta\to\infty} \frac{1}{\beta} \log \phi_{\beta A}(x). \tag{7}$$

This V is a calibrated sub-action for A (see [2, 12, 20]). We can also get a calibrated sub-action V^* for A^* using the limit

$$V^*(y) = \lim_{\beta\to\infty} \frac{1}{\beta} \log \phi_{\beta A^*}(y) \ . \tag{8}$$

From [2] (see also [42]) we have

$$\phi_{A^*}(y) = \int e^{\,W_A(x,y)-c}\, d\nu_A(x).$$

Finally, we define for each $x \in \Sigma$,

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

where V is a (any) calibrated sub-action.

The function I, where $I : \Sigma \to \mathbb{R} \cup \{\infty\}$, can have infinite values, but it is lower semi-continuous. In [2] it is shown that for any cylinder set $C \subset \Sigma$,

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

In this way we get a Large Deviation principle for $\mu_{\beta A} \to \mu_\infty$.

Remember that we denote by μ_∞^* the unique maximizing probability for A^* (it is unique because μ_∞ is unique for A, and, moreover, A and A^* are cohomologous in $\hat{\Sigma}$).

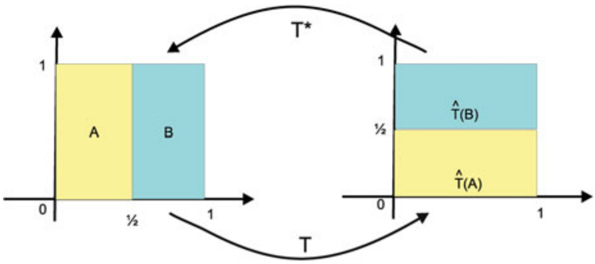
All the results described above are true for expanding transformations T of class $C^{1+\alpha}$ on the circle S^1. In this case we have to consider the natural extension $\hat{T}$ of T. This also includes the case of $T(x) = -2x$ (mod 1).

In the case $T : S^1 \to S^1$, given by $T(x) = 2x$ (mod 1), we define $\hat{T}$ in the following way: the Baker transformation associated to T, denoted by $\hat{T}(x_1, x_2)$, where $\hat{T} : [0,1]^2 \to [0,1]^2$, is such that satisfies for all $(x_1, x_2) \in [0,1]^2$, $\hat{T}(x_1, T^*(x_2)) = (T(x_1), x_2)$ (see picture bellow) . In this case $T^* : S^1 \to S^1$, with $T^*(y) = 2y$ (mod 1), $\hat{T}$ plays the role of $\hat{\sigma}$, and T^* plays the role of σ^*, on the definitions and results above.

All the above apply for an expanding transformation $T : S^1 \to S^1$, or $T : [0,1] \to [0,1]$.

The transformation $\hat{T}$ on $S^1 \times S^1$, contract vertical fibers by forward iteration and expand (and cut) vertical fibers by backward iteration.

Characterization of S



Remember that we said that $W : \hat{\Sigma} = \Sigma \times \Sigma \to \mathbb{R}$ satisfies the twist condition on $\hat{\Sigma}$, if for any $(a, b) \in \hat{\Sigma} = \Sigma \times \Sigma$ and $(a', b') \in \Sigma \times \Sigma$, with $a' > a$, $b' > b$, we have

$$W(a, b) + W(a', b') < W(a, b') + W(a', b). \tag{9}$$

We have the analogous definition for expanding transformations on the interval:

Definition 6 We say $W : [0, 1]^2 \to \mathbb{R}$ continuous satisfies the twist condition on $[0, 1]^2$, if for any $(a, b) \in [0, 1]^2$ and $(a', b') \in [0, 1]^2$, with $a' > a$, $b' > b$, we have

$$W(a, b) + W(a', b') < W(a, b') + W(a', b). \tag{10}$$

Same definition for W on $S^1 \times S^1$.

When $x, y \in [0, 1]$ (or, on S^1), the condition

$$\frac{\partial^2 W}{\partial x\, \partial y} < 0,$$

implies the twist condition for W. The twist condition can be seen as a kind of transversality condition (see [37])

Example 1 Consider the transformation $T : S^1 \to S^1$, given by $T(x) = -2x$ (mod 1) and $A(x) = a + bx + cx^2$, where a, b, c are constants and $c > 0$. In item (b) in the Appendix we show an explicit expression for the W-kernel and we prove that W satisfies the twist condition. From this, we can get an explicit expression for the calibrated subaction for a certain potential (see Remark 6 in the Appendix).

We point out that for considering the system above in S^1 we have to assume above that $A(0) = A(1)$. If we are interested in the case of $[0, 1]$ the same result can be obtained but we do not have to assume $A(0) = A(1)$.

Moreover, we also show in item (c) in the Appendix that a certain class of analytic perturbations of $A(x) = a + bx + cx^2$ produces W-kernels which are twist.

Example 2 In item (b) in the Appendix we show an example of a W-kernel for a continuous potential A, and for the action of the shift σ on the Bernoulli space $\{0, 1\}^{\mathbb{N}}$, which is twist.

Example 3 Consider the Gauss map $T(x) = \frac{1}{x} - [\frac{1}{x}]$ on $[0, 1]$.

We can define the Baker transformation associated to T, denoted by $\hat{T}(x_1, x_2)$, where $\hat{T} : [0, 1]^2 \to [0, 1]^2$. The involution kernel W for $A(x_1) = -\log T'(x_1)$ is $W(x_1, x_2) = -2 \log(1 + x_1 x_2)$ (see [2]).

It is known that the dual of $A = -\log T'$ is $A^* = -\log T'$ (see Proposition 4 in [2]).

The maximizing probability for such potential $-\log T'(x) = 2\log(x)$ is the δ-Dirac in the fixed point b, where b is the golden mean $b = \frac{\sqrt{5}-1}{2}$ (see for instance [14]). In this case $m(A) = 2\log(b)$.

Note that W is differentiable on any point $(x_1, x_2) \in [0, 1]^2$.

One can easily see that an explicit calibrated sub-action u (unique up to an additive constant because the maximizing probability is unique [20]) satisfying

$$u(x) \leqslant u(T(x)) - A(x) + m(A), \tag{11}$$

is $u(x) = W(x, b) = -2 \log(1 + x b)$.

Note that

$$\frac{\partial^2 W}{\partial x\, \partial y} < 0,$$

and, therefore, W is twist.

Example 4 Suppose $T(x) = -2x$ (mod 1), $T : [0, 1] \to [0, 1]$ and $A : [0, 1] \to \mathbb{R}$ is Hölder and monotonous. Under some assumptions on A one can get cases where the maximizing probability is unique and with support on the right fixed point p (see [31]). In the same way as in last example one can show that $V(x) = W(x, p)$ is a calibrated subaction.

If one considers on the interval $[0, 1]$ the potential $A(x) = x^2$ is under such assumptions. One can show that $A^*(y) = y^2$, and $W(x, y) = (1/3)(x^2 + y^2) - (4/3)xy$ (see Remark 6 in item (b) in the Appendix). In the same way $\frac{\partial^2 W(x,y)}{\partial x\, \partial y} < 0$.

Example 5 Consider the transformation $T : S^1 \to S^1$, given by $T(x) = -2x$ (mod 1) and $A(x) = -(x - \frac{1}{2})^2$ (a continuous potential on S^1) for which all results in [2] apply (see also [37] where it is shown in this case the graph property). The maximizing probability has support in the periodic orbit of period 2 (see [29, 30]).

One can define the continuous Baker transformation associated to T, denoted by $\hat{T}(x_1, x_2)$, where $\hat{T} : [0, 1]^2 \to [0, 1]^2$ is such that satisfies for all $(x_1, x_2) \in [0, 1]^2$, $\hat{T}(x_1, T(x_2)) = (T(x_1), x_2)$.

In this case, we show in Remark 6 in the Appendix that a smooth W-kernel is:

$$W(x, y) = -(1/3)x^2 - (1/3)y^2 + (4/3)xy - (2/3)x - (1/3)y.$$

The dual potential A^* is equal to A.

This W-kernel is **not** twist because $\frac{\partial^2 W(x,y)}{\partial x\, \partial y} > 0$.

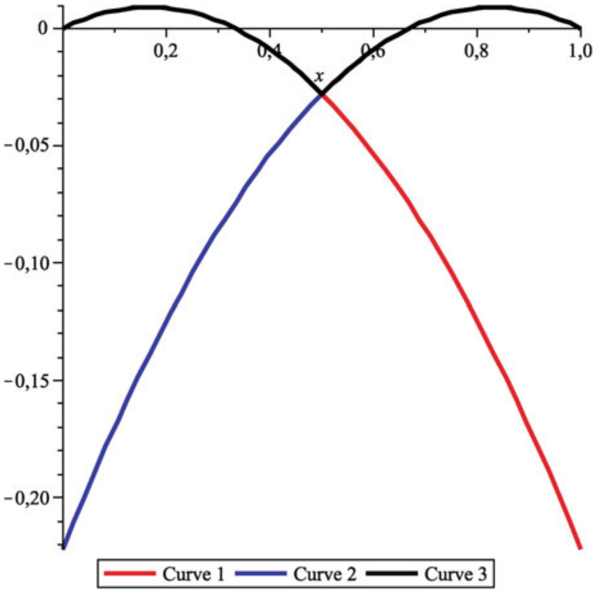
It follows from a general result presented in [31] that any maximizing measure for this potential is $\mu_\infty = (1-t)\delta_{1/3} + t\delta_{2/3}$, where $t \in [0, 1]$, so the critical value is $m = A(1/3) = A(2/3)$.

It is easy to verify that,

$$\begin{aligned} V(x) &= (W(x, 1/3) - W(1/3, 1/3))\chi_{[(0,1/2)]}(x) + W(x, 2/3) - W(2/3, 2/3)\chi_{[1/2,1]}(x) \\ &= \max\{W(x, 1/3) - W(1/3, 1/3), W(x, 2/3) - W(2/3, 2/3)\} \end{aligned}$$

is a calibrated subaction for A.

$W(x, 1/3) - W(1/3, 1/3)$=*red*, $W(x, 2/3) - W(2/3, 2/3)$ = *blue* and ϕ=*black*—The calibrated subaction is the supremum of the two functions described in the picture



This calibrated subaction is not analytic but piecewise analytic (see [40] for more general results).

Example 6 Consider the transformation $T : S^1 \to S^1$, given by $T(x) = -2x$ (mod 1) and $A(x) = (x - \frac{1}{2})^2$ (a continuous potential on S^1) for which all results in [2] apply.

In this case we show in item (b) in the Appendix that a smooth W-kernel is:

$$W(x, y) = (1/3)x^2 + (1/3)y^2 - (4/3)xy + (2/3)x + (1/3)y,$$

the dual potential A^* is equal to A and this involution kernel W is twist.

Similar results can be obtained for $T : S^1 \to S^1$, given by $T(x) = 2x$ (mod 1) and $A(x) = -(x - \frac{1}{2})^2$ (a continuous potential on S^1)

Definition 7 Given $G : \hat{\Sigma} \to \mathbb{R}$ upper semi-continuous, and $f(x)$ continuous, where $f : \Sigma \to \mathbb{R}$, we define the G-transform of f, denoted by $f^{\#}(y)$, where $f^{\#} : \Sigma^* \to \mathbb{R}$, the function such that

$$f^{\#}(y) = \max_{x\in\Sigma} \{-f(x) + G(x, y)\}. \tag{12}$$

We can use also the notation $f^{\#}_G$, instead of $f^{\#}$, if we want to stress the dependence on G.

In this case we say that $f^{\#}$ is the G-conjugate of f [53, 54]. We use the notation of [49, p. 268]. Note that, if we add a constant to f, then new $f^{\#}$ will be obtained from the old one by subtracting the same constant. Therefore, in this case the sum $f(x) + f^{\#}(y)$ will be the same. We are interested, for example, when $G = -W$ or $G = -W + I$. A similar definition and properties can be consider for expanding transformations on [0, 1].

Proposition 1 *If V is a subaction for A, then $V^{\#} = V^{\#}_W$ is a subaction for A^*.*

Proof Given y there exist z^0 such that

$$\begin{gathered}
V^{\#}(\sigma^*(y)) - V^{\#}(y) = \max_{x\in\Sigma} \{-V(x) + W(x, \sigma^*(y))\}- \\
\max_{z\in\Sigma} \{-V(z) + W(z, y)\} = \\
\max_{x\in\Sigma} \{-V(x) + W(x, \sigma^*(y))\} - (-V(z_0) + W(z_0, y)) \geq \\
-V(\tau_y(z_0)) + W(\tau_y(z_0), \sigma^*(y)) + V(z_0) - W(z_0, y) \geq \\
A(\tau_y(z_0)) - m(A) + W(\tau_y(z_0), \sigma^*(y)) - W(z_0, y) = \\
A^*(y) - m(A) = A^*(y) - m(A^*).
\end{gathered}$$

The subaction you get by $-W$-transform is not necessarily calibrated.

Note that if we add a constant to W (the new W will be also a W-Kernel), then all of the above will be also true.

In a similar way like in the reasoning of last proposition one can get:

Proposition 2 *If V^* is a sub-action for A^*, then*

$$(V^*)^{\#}_W(x) = \max_{z\in\Sigma^*} \{-V^*(z) + W(x, z)\}$$

is a subaction for A.

Analogous definitions can be consider for an expanding transformation $T : S^1 \to S^1$. This also includes the case of $T(x) = -2x$ (mod 1).

2 The Transport Problem

We assume that the maximizing probability μ_∞ for A is unique. We denote by μ_∞^* a fixed maximizing probability for A^*. We denote by $\mathcal{K}(\mu_\infty, \mu_\infty^*)$ the set of probabilities $\hat{\eta}(x, y)$ on $\hat{\Sigma}$, such that

$$\pi_x^*(\hat{\eta}) = \mu_\infty, \text{ and } \pi_y^*(\hat{\eta}) = \mu_\infty^*.$$

We are going to consider bellow the cost function $c(x, y) = I(x) - W(x, y) + \gamma$, which is defined for x such that $I(x) \neq \infty$.

The Kantorovich Transport Problem Given A (and all the probabilities described above) we are interested in the minimization problem

$$C(\mu_\infty, \mu_\infty^*) = \inf_{\hat{\eta} \in \mathcal{K}(\mu_\infty, \mu_\infty^*)} \int \int (I(x) - W(x, y) + \gamma)\, d\,\hat{\eta} =$$

$$\inf_{\hat{\eta} \in \mathcal{K}(\mu_\infty, \mu_\infty^*)} \int \int c(x, y)\, d\,\hat{\eta} =$$

$$\max_{\hat{\eta} \in \mathcal{K}(\mu_\infty, \mu_\infty^*)} \int \int (W(x, y) - \gamma - I(x))\, d\,\hat{\eta} \tag{13}$$

where, I is the deviation function for $\mu_\infty = \lim_{\beta \to \infty} \mu_{\beta A}$ (see [2]),

$$c_\beta = \int \int e^{\beta\, W(y,x)}\, d\nu_{\beta A}(x)\, d\nu_{\beta A^*}(y), \tag{14}$$

and

$$\gamma = \lim_{\beta \to \infty} \frac{1}{\beta} \log c_\beta\,, \tag{15}$$

as in proposition 5 in [2]. We call $c(x, y) = -W(x, y) + \gamma + I(x)$ the cost function. Therefore, c is lower semi-continuous. A probability $\hat{\eta}$ on $\hat{\Sigma}$ which attains such minimum is called an optimal transport probability. We denote it by $\hat{\mu}$. We will show later that $\hat{\mu}_{max}$, the natural extension of μ_∞, will be the optimal transport probability $\hat{\mu}$.

One of our main results is Theorem 5 which claims that: The support of $\hat{\mu}_{max}$ is c-cyclically monotone. In other words, the twist condition for c is true when restricted to the support of the maximizing probability $\hat{\mu}_{max}$.

Remark 1 Note that if we subtract the deviation function $I(x)$ of the cost function, that is, if we consider a new cost $c(x, y) = -W(x, y) + \gamma$, the problem above will not change, because I is constant zero in the support of μ_∞. In other words

$$C(\mu_\infty, \mu_\infty^*) = \inf_{\hat\eta \in \mathscr{K}(\mu_\infty, \mu_\infty^*)} \int\int (-W(x,y) + \gamma)\, d\hat\eta,$$

and, the optimal transport probability will be the same. In some sense this setting is nicer because the cost c is a continuous function on $\hat\Sigma$.

Definition 8 A pair of functions $f(x)$ and $f^\#(y)$ will be called c-admissible (or, just admissible for short) if

$$f^\#(y) = \min_{x\in\Sigma} \{-f(x) + c(x,y)\}. \tag{16}$$

In other words $-f^\#$ is the $-c$-conjugate of $-f$. Note that in this case, $\forall x \in \Sigma$, $y \in \Sigma^*$, we have that $f(x) + f^\#(y) \le c(x,y)$. We denote by $\mathscr{F}$ the set of all admissible pairs $(f(x), f^\#(y))$.

The Kantorovich Dual Problem Given A and the corresponding c (W and all the probabilities described above) we are interested in the maximization problem

$$D(\mu_\infty, \mu_\infty^*) = \max_{(f,f^\#)\in\mathscr{F}} \left(\int f d\mu_\infty + \int f^\# d\mu_\infty^* \right). \tag{17}$$

A pair of admissible $(f, f^\#) \in \mathscr{F}$ which attains the maximum value will be called an optimal pair.

The Kantorovich duality theorem (see [53]) claims that under general conditions $D(\mu_\infty, \mu_\infty^*) = C(\mu_\infty, \mu_\infty^*)$. The main tool to prove this result is the Fenchel-Rockafellar duality Theorem.

Theorem 4 (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,

$$\inf_{v\in E} [\Theta(v) + \Xi(v)] = \max_{f\in E^*} [-\Theta^*(-f) - \Xi^*(f)] \tag{18}$$

We will not present the proof of this general theorem but we will present a nice geometric proof in a simple case (one-dimensional) in item (e) in the Appendix. We suppose, from now on, that the maximizing probability for A, denoted by μ_∞ is unique. We denote, as in [12] the calibrated sub-actions V and V^* by

$$V(x) = \lim_{\beta\to\infty} \frac{1}{\beta} \log \phi_{\beta A}(x) \text{ and } V^*(y) = \lim_{\beta\to\infty} \frac{1}{\beta} \log \phi_{\beta A^*}(y)\,. \tag{19}$$

The above convergence is uniform and V is (up to constant) the unique calibrated sub-action for A (see [2, 12, 20]). We will show later that $(f, f^{\#})$ such that $f(x) = -V(x)$ and $f^{\#}(y) = -V^*(y)$ is the optimal pair.

Important Property If $\hat{\mu}$ is an optimal transport probability and if $(f, f^{\#})$ is an optimal pair in $\mathscr{F}$, then the support of $\hat{\mu}$ is contained in the set

$$\{ < y, x > \in \hat{\Sigma} \mid \text{such that } (f(x) + f^{\#}(y)) = c(x, y) \}. \tag{20}$$

It follows from the prime and dual linear programming problem formulation. The condition above is the complementary slackness condition (see [17, 19, 48]).

The reciprocal of this result is also true (see [54, Remark 5.13, p. 59]).

If x and y are such that $(f(x) + f^{\#}(y)) = c(x, y)$ we say that they are realizers for the cost c. In [13] it is shown that the set of realizers for $I - W$ is an invariant set for the dynamics of $\hat{\sigma}$. In this section we are mainly concerned with the support and not with all realizers.

If one finds $\hat{\mu}$ an admissible pair $(f, f^{\#})$ satisfying the above claim (for the support), then, one solves the Kantorovich problem, that is, one finds the optimal transport probability $\hat{\mu}$.

No we will prove Theorem 1.

Proposition 3 *The minimizing Kantorovich probability $\hat{\mu}$ on $\hat{\Sigma}$ associated to $-W$ is $\hat{\mu}_{max}$.*

Proof Proposition 10 (1) in [2] claims that if $\hat{\mu}_{max}$ is the natural extension of the maximizing probability μ_∞, then for all $< p^*|p >$ in the support of $\hat{\mu}_{max}$ we have

$$-V(p) - V^*(p^*) = -W(p, p^*) + \gamma.$$

This is the same as saying that in the support of $\hat{\mu}_{max}$

$$-V(p) - V^*(p^*) = -W(p, p^*) + \gamma + I(p) = c(p, p^*),$$

because I is zero in the support of μ_∞. Then if $-V(x)$ and $-V^*(y)$ is an admissible pair, then $\hat{\mu}_{max}$ is the optimal transport probability for such $c(x, y)$. This will be shown in the next proposition. We will show bellow that the $-c$-transform of V is V^*.

Note that if W is a W-Kernel for A, for all β, we have that βW is a W-Kernel for βA. We denote by c_β the normalizing constant for βW, as in [2]. It is known that $\frac{1}{\beta} \log c_\beta = \gamma$.

Now we will show Theorem 2.

Proposition 4 *The pair $(-V, -V^*)$ is admissible.*

Proof For a fixed y we have to show that

$$-V^*(y) = (-V)^{\#}_c = \inf_{x \in \Sigma} \{-(-V(x)) + c(x, y)\} .$$

This is the same as

$$V^*(y) = \sup_{x\in\Sigma} \{ (-V(x)) - c(x,y) \} = \sup_{x\in\Sigma} \{ -V(x) - (\gamma - W(x,y) + I(x)) \} ,$$

or, for all x

$$-V^*(y) \leq V(x) + c(x,y)\,. \tag{21}$$

From Proposition 3 in [2] (we just write here $W(x,y)$, instead of $W(y,x)$ there) we have

$$\phi_{\beta A^*}(y) = \int e^{\beta\, W_A(x,y) - c_\beta} \frac{1}{\phi_{\beta A}(x)}\, d\mu_{\beta A}(x) = \int e^{\beta\, W_A(x,y) - c_\beta - \log \phi_{\beta A}(x)}\, d\mu_{\beta A}(x).$$

Consider now the limit

$$V^*(y) \;=\; \lim_{\beta\to\infty} \frac{1}{\beta}\, \log(\phi_{\beta A^*}(y)) =$$

$$\lim_{\beta\to\infty} \frac{1}{\beta}\, \log \int e^{\beta\, W_A(x,y) - c_\beta - \log \phi_{\beta A}(x)}\, d\mu_{\beta A}(x).$$

From [12] the function $\frac{1}{\beta}\, \log(\phi_{\beta A}(x))$ converges uniformly with β to $V(x)$. Therefore, one can write

$$\lim_{\beta\to\infty} \frac{1}{\beta}\, \log \int e^{\beta\, W_A(x,y) - c_\beta - \log \phi_{\beta A}(x)}\, d\mu_{\beta A}(x) =$$

$$\lim_{\beta\to\infty} \frac{1}{\beta}\, \log \int e^{\beta\, (\, W_A(x,y) - \gamma - V(x)\,)}\, d\mu_{\beta A}(x)$$

Now, by Varadhan's Integral Lemma [16] we obtain

$$V^*(y) \;= \sup_x \{W_A(x,y) - \gamma - V(x) - I(x)\} = \sup_x \{-V(x) + W(x,y) - \gamma - I(x)\},$$

where I is the deviation function.

Finally, we get that $\hat{\mu}_{max}$ is the optimal transport probability for such $c(x,y)$. From now on we will use either the notation $\hat{\mu}$ or $\hat{\mu}_{max}$ for the optimal transport probability. In [40] Transport Theory is used as a tool to show that in some cases the calibrated subaction is piecewise analytic. In [13] some generic properties of the potential A is considered and special results about the realizers of the $W - I$ are obtained.

The last theorem says: for any $y \in \Sigma^*$ we have

$$V^*(y) \;= \sup_{x\in\Sigma} \{-V(x) - c(x,y)\}. \tag{22}$$

Note that when $y = p^*$, for p^* in the support of μ^*_∞, the supremum

$$V^*(p^*) = \sup_x\{-V(x) + W(x, p^*) - \gamma - I(x)\} = \sup_x\{-V(x) - c(x, p^*)\},$$

is realized at $x = p$, for p in the support of μ_∞ (with $< p^*, p >$ in the support of $\hat{\mu}$).

Remark 2 Remember that, if the maximizing probability for A^* is unique, then there is a unique calibrated sub-action for A^* (up to additive constant) [2, 20].

Analogous definitions and properties can be obtained for $T : S^1 \to S^1$. This also includes the case of $T(x) = -2x$ (mod 1). We could likewise consider the analogous problem for A^*: given A^* (obtained from A) fixed, denote $I^* : \Sigma^* \to \mathbb{R}$, the non-negative deviation function for $\mu_{\beta A^*} \to \mu^*_\infty$. Denote $c^*(x, y) = (I^*(y) - W(x, y) + \gamma)$.

Then, consider the problem

$$C(\mu_\infty, \mu^*_\infty) = \inf_{\hat{\eta} \in \mathcal{K}(\mu_\infty, \mu^*_\infty)} \int\int (I^*(y) - W(x, y) + \gamma)\, d\,\hat{\eta} =$$

$$\inf_{\hat{\eta} \in \mathcal{K}(\mu_\infty, \mu^*_\infty)} c^*(x, y)\, d\,\hat{\eta} = \inf_{\hat{\eta} \in \mathcal{K}(\mu_\infty, \mu^*_\infty)} \int\int (-W(x, y) + \gamma)\, d\,\hat{\eta},$$

which have the same minimizing measures, as for the minimization for $c(x, y) = (I(x) - W(x, y) + \gamma)$ among probabilities on $\mathcal{K}(\mu_\infty, \mu^*_\infty)$.

Note also that from Proposition 3 in [2] we have

$$\phi_{\beta A}(x) = \int e^{\beta\, W_A(x,y) - c_\beta} \frac{1}{\phi_{\beta A^*}(y)}\, d\mu_{\beta A^*}(y) =$$

$$\int e^{\beta^*\, W_A(x,y) - c_\beta - \log \phi_{\beta A^*}(y)}\, d\mu_{\beta A^*}(y).$$

In the same way as before one can show that for any $x \in \Sigma$, we have

$$V(x) = (-V^*)^{\#}_{c^*} = \sup_{y \in \Sigma^*} \{-V^*(y) - c^*(x, y)\}. \tag{23}$$

Note that $c(x, y) = c^*(x, y)$ in the support of the minimizing $\hat{\mu}_{max}$ for c (or for c^*).

Remark 3 It is not necessarily true that $(\,(-V^*)^{\#}_{c^*}\,)^{\#}_{c^*} = -V^*$. However, the expression is true when restricted to the support of the optimal transport probability $\hat{\mu}_{max}$. In the same way $(\,(-V)^{\#}_{c}\,)^{\#}_{c} = -V$ in the support of $\hat{\mu}_{max}$.

3 Graph Properties and the Twist Condition

Consider a lower semi-continuous cost function $c(x, y)$ on $\hat{\Sigma}$ (or, a continuous cost function $-W(x, y)$ on $\hat{\Sigma}$). We refer the reader to [48, 53, 54] and [19] for general references on optimal mass transportation problems.

Definition 9 A set $S \subset \hat{\Sigma}$ is called c-cyclically monotone, if for any finite number of points (x_j, y_j) in S, $j \in \{1, 2, \ldots, n\}$, and any permutation σ of the n letters, we have

$$\sum_{j=1}^{n} c(x_j, y_j) \leq \sum_{j=1}^{n} c(x_{\sigma(j)}, y_j). \tag{24}$$

Proposition 5 (See Theorem 2.3 [19]) *For a continuous function $c(x, y) \geq 0$, where $\hat{\Sigma}$, if $\rho \in \mathscr{K}(\mu_\infty, \mu_\infty^*)$ is optimal for c, then, ρ has a c-cyclically monotone support.*

Corollary 1 *The support of $\hat{\mu}_{max}$, the natural extension of μ_∞ is c-cyclically monotone.*

We will present bellow in the next theorem a direct proof of this fact.

Definition 10 A function $f : \Sigma \to \mathbb{R} \cup \{\infty\}$ is c-concave, if there exist a set $A \subset \Sigma \times \mathbb{R}$ such that

$$f(y) = \sup_{(x,\lambda)\in A} \{c(x, y) + \lambda\}$$

Definition 11 A function $f : X \to \mathbb{R} \cup \{\infty\}$ is c-convex, if $(-f)$ is c-concave.

Definition 12 Given $x \in \Sigma$, the set $\hat{\partial}_c f(x)$ is the set of $y \in \hat{\Sigma}$ such that, for all $z \in \Sigma$ we have

$$f(z) - f(x) \leq c(z, y) - c(x, y)$$

In this case we say y is a c-sub-derivative for f in x.

An important problem is to know, for a certain given x, if the $\hat{\partial}_c f(x)$ has cardinality 1.

Proposition 6 (See Theorem 2.7 in [19], Lemma 2.1 in [49] and Section 4 in [48]) *For $S \subset \hat{\Sigma}$ to be c-cyclically monotone, it is necessary and sufficient that $S \subset \hat{\partial}_c(f)(x) = \{(x, y) \,|\, f(z) - f(x) \leq c(z, y) - c(x, y)\,,\ \forall z \in X\}$, for some c concave f, where $f : \Sigma \to \mathbb{R} \cup \{\infty\}$.*

Moreover: f is defined in the following way: choose $(x_0, y_0) \in S$*, then*

$$f(x) = \inf_{n\in\mathbb{N},\,(x_j,y_j)\in S,\,1\le j\le n} [\,(\,c(x,y_n) - c(x_n,y_n)\,)+ \\ (\,c(x_n,y_{n-1}) - c(x_{n-1},y_{n-1})\,) + \ldots \\ +(\,c(x_2,y_1) - c(x_1,y_1)\,) + (\,c(x_1,y_0) - c(x_0,y_0)\,)\,].$$

Note that if $S \subset \hat{\Sigma}$ is a graph, then for each $x \in \Sigma$ in the x-projection of S, we have that $\hat{\partial}_c(f)(x)$ has cardinality 1. Consider fixed $(x_0, y_0), (x_1, y_1)$ in the support of $\hat{\mu}_{max}$ and $(x_0, y_1), (x_1, y_0) \in \hat{\Sigma}$. Given a function $f(x, y)$ we denote

$$\Delta_f\,((x_0,y_1),(x_1,y_0)) \;=\; (f(x_0,y_0) + f(x_1,y_1)) - (f(x_0,y_1) + f(x_1,y_0)\,), \tag{25}$$

and

$$b(x, y) = I(x) + \gamma - W(x, y) + V(x) + V^*(y). \tag{26}$$

The c-cyclically monotone condition for the support of $\hat{\mu}_{max}$ will follow from the claim

$$\Delta_c\,((x_0,y_1),(x_1,y_0)) \;=\; (\,c(x_0,y_0) + c(x_1,y_1)) - (\,c(x_0,y_1) + c(x_1,y_0)\,) \le 0. \tag{27}$$

This is so because any permutation of letters can be obtained by a series of composition of transformations that exchange just two letters. It will follow from the proof bellow that $\Delta_c \circ \sigma = \Delta_c$.

The next result does not assume a global assumption on twist condition for c.

Theorem 5 *Given* $A : \Sigma \to \mathbb{R}$ *Hölder, then* $c(x, y) = I(x) - W(x, y) + \gamma \ge 0$*, for all* $(x, y) \in \Sigma$*. Moreover, for* $(x_0, y_0), (x_1, y_1)$ *in the support of* $\hat{\mu}_{max}$*, we have* $\Delta_c \le 0$*. Therefore, the support of* $\hat{\mu}_{max}$ *is c-cyclically monotone. In other words, the twist condition for c (or, for W) is true when restricted to the support of the maximizing probability* $\hat{\mu}_{max}$*.*

Proof First we point out that $\Delta_c = \Delta_b$. We will show that under our hypothesis is true that $\Delta_b \le 0$. First note that

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

Remember (see [2]) that

$$I(x) = \sum_{n=0}^{\infty} [V \,\circ\, \sigma - V - A]\,\hat{\sigma}^n\,(x,y)$$

We denote

$$I_n(x,y) = \sum_{k=0}^{n-1} [V \circ \sigma - V - A] \circ \hat{\sigma}^k \, (x,y) = I_n(x),$$

and

$$R_n\,(x,y) = I_n(x,y) + [\,\gamma + V(x) + V^*(y) - W(x,y)\,] - \\ [\,\gamma + V + V^* - W\,]\,\hat{\sigma}^n\,(x,y).$$

We claim that if (x,y) is in the support of $\hat{\mu}_{max}$, then $b(x,y) = 0$. Moreover, for all $(x,y) \in \Sigma$, we have $b(x,y) \geq 0$. One can prove this result by means of Varadhan's Integral Lemma [16] with the same reasoning as in the last proposition of the previous section. We will give bellow a direct proof of the claim.

Either $I(x) = \infty$, and the claim is trivially true or $I(x)$ is finite. In this case, any accumulation point of $\hat{\sigma}^n(x,y)$ will be in the support of $\hat{\mu}_{max}$.

Moreover, $b(x,y) = R(x,y) = \lim_{n\to\infty} R_n(x,y) \geq 0$. As in the support of $\hat{\mu}_{max}$, we have that $R(x,y) = 0$, then, $b(x,y) = 0$. In any case $R(x,y) \geq 0$. This shows the claim. We point out that $\Delta_c = \Delta_b = \Delta_W$ in the case $I(x)$ is finite.

We also remark that if (x_0, y_0) is in support of $\hat{\mu}_{max}$, then as $R(x_0, y_0)$ is zero, it follows that $R(x_0, y)$ is finite. This is so because (x_0, y) is in the stable manifold of (x_0, y_0) and

$$R_n(x_0,y) - R_n(x_0,y_0) =$$

$$\sum_{k=1}^{n} \{\, [V^* \circ \hat{\sigma}^{-1} - V^* - A^*]\hat{\sigma}^k(x_0,y) - [V^* \circ \hat{\sigma}^{-1} - V^* - A^*]\hat{\sigma}^k(x_0,y_0)\,\}.$$

Finally, if (x_0, y_0) and (x_1, y_1) are both in the support of $\hat{\mu}_{max}$, then $R(x_0, y_1) < \infty$, $R(x_1, y_0) < \infty$ and $I(x_0) = 0 = I(x_1)$. In this case, for any (x,y) of the form (x_0, y_0), (x_1, y_1), (x_1, y_0), or (x_0, y_1)

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

As we know that R is non-negative, then

$$[b(x_0,y_0) + b(x_1,y_1)] - [b(x_1,y_0) + b(x_0,y_1)] = 0 - [b(x_1,y_0) + b(x_0,y_1)] \leq 0.$$

This shows that $\Delta_b \leq 0$.

We did not use the twist condition above. Note that we could alternatively consider the function $g : \Sigma \to \mathbb{R}$ defined in the following way: choose $(x_0, y_0) \in S$, then

$$g(x) = \inf_{n\in\mathbb{N},\,(x_j,y_j)\in S,\,1\le j\le n} [\,(\,W(x,y_n) - W(x_n,y_n)\,) +$$
$$(\,W(x_n,y_{n-1}) - W(x_{n-1},y_{n-1})\,) + \dots$$
$$+(\,W(x_2,y_1) - W(x_1,y_1)\,) + (\,W(x_1,y_0) - W(x_0,y_0)\,)\,],$$

which has the advantage of just taking into account a continuous function W. The graph property for $S =$ support of $\hat{\mu}$, and all kinds of different considerations can be obtained from such g. We want to show now that if W satisfies the twist condition and the maximizing probability for A is unique, then the support of $\hat{\mu}$ on $\hat{\Sigma}$ is a graph. Our proof works for the Venously space $\{0, 1, 2, .., d\}^{\mathbb{N}}$ as well for the interval $[0, 1]$ [considering T either conjugated to $2x$ (mod 1) or to $-2x$ (mod 1)].

Consider the cost $c(x, y) = I(x)-W(x, y)-\gamma$, and a subset $S \subset X\times Y$ c-cyclically monotone.

Lemma 1 *Suppose the c satisfies the twist condition and let S be a c-cyclically monotone subset, if* $(a, b), (a', b') \in S$ *and* $a \neq a'$ *and* $b \neq b'$*, then* $a < a'$ *and* $b > b'$*, or* $a > a'$ *and* $b < b'$*.*

Proof Indeed, suppose $a < a'$ then, if $b < b'$, the twist condition on W implies that

$$c(a, b) + c(a', b') > c(a, b') + c(a', b).$$

On the other hand, S is c-cyclically monotone subset, so

$$c(a, b) + c(a', b') \le c(a, b') + c(a', b),$$

that is an absurd.

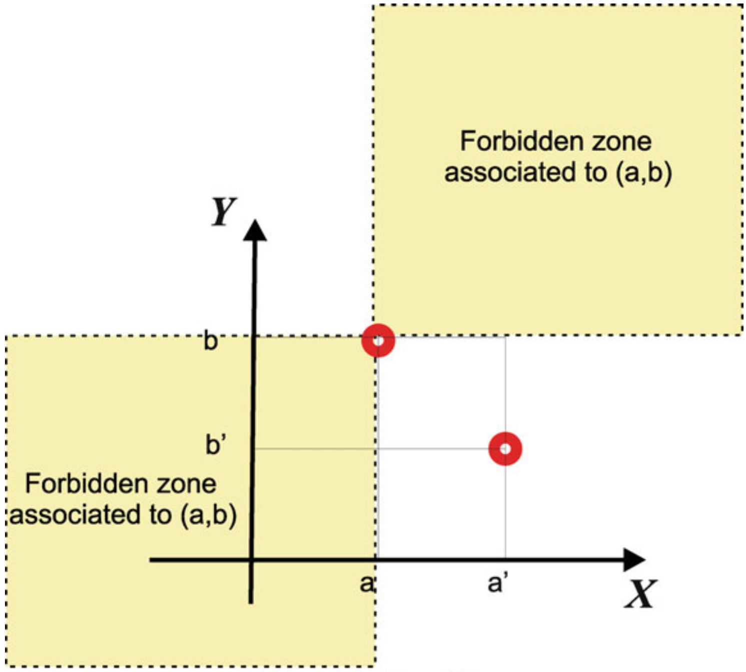
A similar property is true for W. This Lemma means that the correct figure associated to a pair of points in S is given by:

We point out that, in principle, could exist points z of S in the vertical fiber passing by a or in the horizontal fiber passing by b.

Now we will show Theorem 3.

Theorem 6 (Graph Theorem) *Suppose the involution kernel W satisfies the twist condition and let* $\hat{\mu}$ *be the c-minimizing measure of probability to the transport problem, then* $S = supp\ \hat{\mu}$ *is a graph in x (up to an orbit of measure zero), moreover this graph is monotone not increasing.*

Proof In fact we will just use the twist condition for W on the support of the optimal transport probability. In order to get advantage of the geometrical and combinatorial arguments we will present pictures for the case of a transformation $T : [0, 1] \to$



Characterization of S

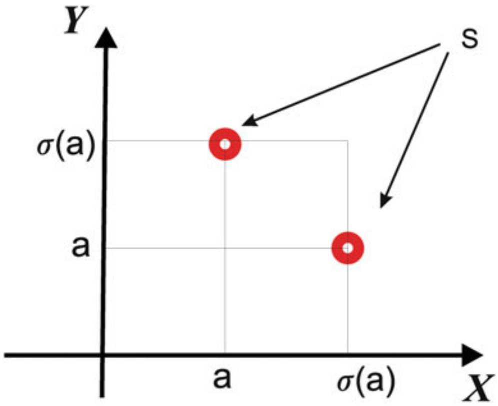
Characterization of S

$[0, 1]$, given by $T(x) = 2x$ (mod 1). Define $v^+(x) = \max\{y|(x, y) \in S\}$ and $v^-(x) = \min\{y|(x, y) \in S\}$. In order to prove that supp $\hat{\mu}$ is a graph we need to prove that $v^-(x) = v^+(x)$ for any x in the support of μ_∞. We say that a point (x, y) in the support of $\hat{\mu}$ is non-graph, if there exist another point of the form (x, z), in the support of $\hat{\mu}$, and such that $z \neq y$. Note that the image of two points in the support of $\hat{\mu}$ on the fiber over x will go on two different points in the support of $\hat{\mu}$ on the fiber over $\sigma(x)$. That is, the forward image by $\hat{\sigma}^n$ of non-graph points will go on non-graph points. This maybe can not be true for backward images by $\hat{\sigma}^n$.

Suppose the support of the maximizing probability μ_∞ (unique) is a periodic orbit. If S is not a graph, then $v^-(x) < v^+(x)$ for some x. As the transformation $\hat{\sigma}$ contracts each fiber by forward iteration, we have that, the image of the interval fiber from $(x, v^-(x))$ to $(x, v^+(x))$, by a finite iterate of $\hat{\sigma}$, goes inside the fiber $(x, v^-(x))$ to $(x, v^+(x))$. Therefore, σ^* has a periodic point in the support of μ^*_∞. If the maximizing probability μ_∞ is unique for A, then μ^*_∞ is unique for the maximization problem for A^*. In this case the support of μ^*_∞ is this periodic orbit. Therefore, there is a minimal distance (in vertical fiber) between non-graph points and this is in contradiction with the contraction on vertical fibers. The conclusion is that S is a graph if the support of the maximizing probability μ_∞ is a periodic orbit.

Remark 4 In the case of the shift, if suppμ_∞ is a periodic orbit, one can easily show that if suppμ_∞ = the orbit by σ of $(a_0, a_1, \ldots, a_{(n-1)}, a_0, \ldots)$ then suppμ^*_∞ = orbit by σ^* of $(a_{(n-1)}, \ldots, a_2, a_1, a_0, a_{(n-1)}, \ldots)$.

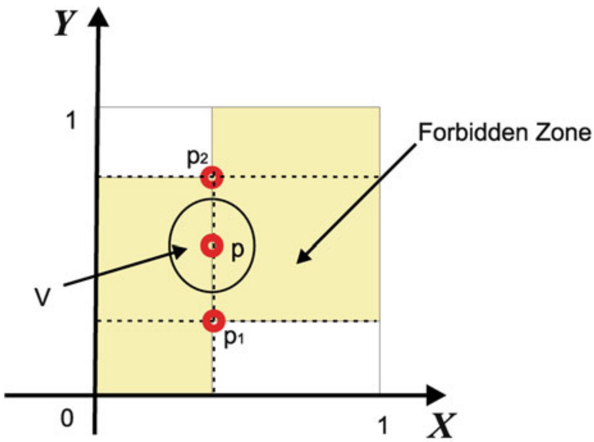
Support of $\hat{\mu}$ in the periodic case



Support of $\hat{\mu}$ in the periodic case.

We suppose from now on that the support of the maximizing probability μ_∞ is not a periodic orbit.

Characterization of S



Characterization of S

Suppose, that $v^-(x) < v^+(x)$ for some x, then we claim that there is no other point in support of $\hat{\mu}$ in the fiber by x between $p_1 = v^-(x)$ and $p_2 = v^+(x)$. Indeed, from the above picture we see that if there exists a point (x, p) in the support of $\hat{\mu}$ such that $v^-(x) = p_1 < p < p_2 = v^+(x)$, then, as $\hat{\mu}$ is ergodic, should exist a point

(q_1, q_2) in a small neighborhood V of (x, p) such that returns by a forward n-iterate by $\hat{\sigma}$ to V.

This iterate has to return to the fiber, and this contradicts the fact that the support of the maximizing probability μ_∞ is not a periodic orbit.

If the support of μ_∞ is not a periodic orbit, then we claim that does not exist two pairs $(x_1, y_1), (x_1, z_1)$ and $(x_2, y_2), (x_2, z_2)$, in the support of $\hat{\mu}$, such that, the orbits by σ of x_1 and x_2 are different.

In order to simplify the argument and the notation we consider bellow $T^*(x) = 2x \pmod 1$, but we point out the reasoning apply to any expanding transformation of degree d. Given y_n and z_n, $n = 1, 2$, there exists a rational point of the form $s_n = \frac{q}{2^k}$, with $0 < q < 2^k$, $q, k \in \mathbb{N}$, such that $y_n < s_n < z_n$, $n = 1, 2$. Consider the s_n determined by the smallest possible value k.

The pair of points $\hat{T}^{-r}(x_n, y_n)$ and $\hat{T}^{-r}(x_n, z_n)$, $r \geq 0$, determine non-graph points in the same fiber, for any $r > 0$, until time $r = k$. In time $r = k-1$, it happens for the first time that the horizontal fiber through $1/2$ cuts the vertical segment connecting $\hat{T}^{-(k-1)}(x_n, y_n)$ and $\hat{T}^{-(k-1)}(x_n, z_n)$.

In this way, for each n, we get a horizontal forbidden region A_n (a horizontal strip from one vertical side to the other vertical side of $[0, 1] \times [0, 1]$) determined by such pair $\hat{T}^{k-1}(x_n, y_n)$ and $\hat{T}^{-(k-1)}(x_n, z_n)$, $n = 1, 2$, which contains the horizontal fiber through $1/2$.

If we apply the argument for $n = 1$, then the next forbidden region A_2 for $n = 2$ will contain the previous one A_1. Moreover, considering the full forbidden region determined by these two pair of points we reach a contradiction.

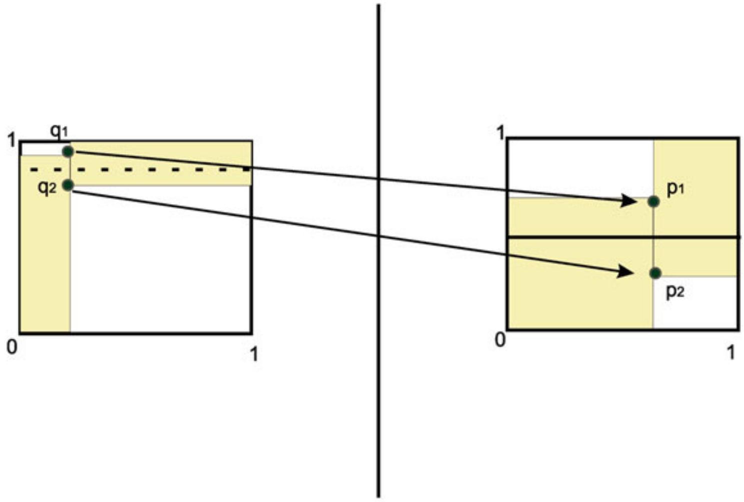
In the picture bellow we show the final pair of points q_1 and q_2 in a $\hat{\sigma}$-orbit (in the same vertical fiber) which has the property that its images p_1 and p_2 are on different sides of the upper and down rectangles. The images of p_1 and p_2 by $\hat{\sigma}$ are not anymore in the same vertical fiber (neither their future iterates). There is no room for getting a different pair of p_1 and p_2 like this (because of the forbidden region).

In this way, from above, we get that could exist just one orbit of x by σ such that over the fiber over x there is two points in the support. That is, the projection $K \subset \Sigma$ on the x-axis of the non-graph points have to be the orbit of a single point x. Therefore, $\mu_\infty(K) = \sum_k \mu_\infty(\{\sigma^k(x)\})$.

We assume first that the set of non-graph points have probability 1 and we will reach a contradiction. Indeed, $\mu_\infty(\{\sigma^k(x)\}) \geq \mu_\infty(\{\sigma^j(x)\})$, for $k \geq j$, and the μ_∞ probability of the set $\{x\}$ is zero or is positive.

Remember that the support of $\hat{\mu}$ is invariant by $\hat{\sigma}$. Now we will show that, indeed, if there exists non-graph points, this set has probability 1.

Note that if the vertical fiber by $x \in \Sigma$ is such that $v^-(x) < v^+(x)$, then $\sigma(x)$ also has this property. If the transformation $\hat{\sigma}$ we consider preserves orientation in the vertical fiber then the iterates are in the same order. Otherwise they exchange order. That is, the set of points (x, y) which are not graph point are invariant by forward iteration by $\hat{\sigma}$. Moreover, $\hat{\sigma}$ is a forward contraction in vertical fibers. Denote by $B = \{\, (x, v^+(x))\}$ in the support of $\hat{\mu}$ such that $\{v^-(x) < v^+(x)\,\}$. The set B is the upper part of the non-graph part of the set S.



The dynamics on the support

The dynamics on the support

We will show that $\hat{\mu}(B) = 0$ or $\hat{\mu}(B) = 1$. We suppose first that $\hat{\sigma}$ preserves order in the fiber by forward iteration. Consider $\tilde{B}$ the set $\{ (x, y)\}$ in the support of $\hat{\mu}$ such that for some $n \geq 0$ we have $\{\hat{\sigma}^n (x, y) \in B\}$. Note that as B is forward invariant, once $\hat{\sigma}^n (x, y) \in B$, for some fixed n, then $\hat{\sigma}^m (x, y) \in B$, for any $m \geq n$.

We will show that $\hat{\sigma}^{-1}\tilde{B} = \tilde{B}$. The fact that $\hat{\sigma}^{-1}\tilde{B} \subset \tilde{B}$ follows easily from the definition of $\tilde{B}$. Given $x \in \tilde{B}$, there exists $n \geq 0$ such that $\hat{\sigma}^n (x, y) \in B$. If $n \geq 1$, then $\hat{\sigma}^{n-1} (\hat{\sigma}(x, y)) \in B$ and, therefore, $(x, y) \in \hat{\sigma}^{-1}\tilde{B}$. In the other case $(x, y) \in B$, but then $(\hat{\sigma}(x, y)) \in B$, because $\hat{\sigma}$ preserves order in the fiber, and does not exist more than two points in the vertical fiber over $\sigma(x)$ which are in S. Therefore, $(x, y) \in \hat{\sigma}^{-1}\tilde{B}$.

As $\hat{\mu}$ is ergodic, then $\hat{\mu}(\tilde{B}) = 0$ or $\hat{\mu}(\tilde{B}) = 1$.

If $\hat{\mu}(\tilde{B}) = 1$, then take a Birkhoff point $z \in \tilde{B}$ for the ergodic probability $\hat{\mu}$. Therefore, we get that the asymptotic frequency of visit to the set $C = \{ (x, v^-(x))\}$ in the support of $\hat{\mu}$ such that $\{v^-(x) < v^+(x)\}$ (the bellow part of the non-graph part of set S) is zero. Finally, we get that $\hat{\mu}(C) = 0$. In the same way $\hat{\mu}(B) = 1$.

If $\hat{\mu}(\tilde{B}) = 0$, we get that $\hat{\mu}(B) = 0$. Now, using a similar argument for the lower part of the non-graph part we get that $\hat{\mu}(C) = 1$.

This shows that the π_1 projection of the non-graph points has probability one and this proves the theorem.

The above reasoning also applies to $T(x) = -2x$ (mod 1) and to the shift in the Bernoulli space.

4 Selection of Minimizing Sequences

In this section we want to exhibit a nice expression for the function f (defined before) such that, the set $\{(x, \hat{\partial}_c f(x)) \,|\, x \in \}$ support $\{\mu_\infty\}$ = support of $\hat{\mu}_{max}$, in the case the support of $\hat{\mu}_{max}$ is a periodic orbit. In the end of the section we address briefly the general case.

Definition 13 We say that $c : \hat{\Sigma} = \Sigma \times \Sigma \to \mathbb{R}$, upper semicontinuous, satisfies the twist condition on $\hat{\Sigma}$, if (bellow we just consider values of c which are finite) for any $(a, b) \in \hat{\Sigma} = \Sigma \times \Sigma$ and $(a', b') \in \Sigma \times \Sigma$, with $a' > a$, $b' > b$, we have

$$c(a, b) + c(a', b') > c(a, b') + c(a', b). \tag{28}$$

If W is twist and $c(x, y) = I(x) - W(x, y) + \gamma$, then c is twist. We assume from now on this property.

Theorem 7 *Suppose the support of $\hat{\mu}_{max}$ is a periodic orbit. Choose (x_0, y_0) in such way that $x_0 \in \Sigma$ is the smaller point in the projection and $y_0 \in \hat{\Sigma}$ the smaller on the fiber over x_0. From the above, in this case for any given $z \in \Sigma$, the f defined before is such that*

$$\begin{gathered} f(z) = [\,(\,c(z, y_n) - c(x_n, y_n)\,)+ \\ (\,c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1})\,)\ + \ldots \\ + \ldots + \{(\,c(x_3, y_2) - c(x_2, y_2)\,)\,\} + \\ (\,c(x_2, y_1) - c(x_1, y_1)\,) + (\,c(x_1, y_0) - c(x_0, y_0)\,)\ \ ], \end{gathered}$$

where we use all the possible x_i which are in the support of the maximizing probability for A on the left of z, and for each x_i we choose the corresponding y_i. In the notation of f above, the last one $(x_n, y_n) = (x_n(z), y_n(z))$ is such that $(x_n(z), y_n(z)) = (x_{k-1}, y_{k-1})$. Which means $n = k - 1$.

Moreover, $x_0 < x_1 < x_2 < \ldots < x_n$.

If $z = x_k$ for some element x_k in the support of μ_A, then, in the notation of f above, if $x_{k-1} < z < x_k$, then $(x_n, y_n) = (x_n(z), y_n(z))$ is such that $(x_n(x_k), y_n(x_k)) = (x_{k-1}, y_{k-1})$. The case $z = x_k$ is include in the expression above for f. In this case $x_k = x_{n+1}$ following the above notation. The index of the x_i has no dynamical meaning.

Proof Consider the cost $c(x, y) = I(x) - W(x, y) - \gamma$, and a subset $S \subset X \times Y$ c-cyclically monotone. Also, assume that c verifies the twist condition: If $a < a'$ and $b < b'$ then $c(a, b) + c(a', b') > c(a, b') + c(a', b)$.

In this way, the definition of c implies that: $W(a, b) + W(a', b') < W(a, b') + W(a', b)$.

Define $\Delta(x, x', y) = W(x, y) - W(x', y)$, so the twist condition can be restated as: if $a < a'$, and $b < b'$, then, $\Delta(a, a', b) < \Delta(a, a', b')$.

Therefore, if we define the map $y \to \Delta(a, a', y)$ we get a increasing map. Observe that:

(i) $\Delta(x, x', y) = -\Delta(x', x, y)$
(ii) $\Delta(x, x, y) = 0$
(iii) $\Delta(x, x', y) + \Delta(x', x'', y) = \Delta(x, x'', y)$

In particular the map, $y \to \Delta(a', a, y)$ is decreasing if $a' > a$.
Using the fact that $c(x, y) = I(x) - W(x, y) - \gamma$ we get,

$$\partial_c f(x) = \{y \in Y | f(z) - f(x) \leq I(z) - I(x) - [W(z, y) - W(x, y)], \forall z \in X\}.$$

We know that S is c-cyclically monotone, if and only if, $S \subset \hat{\partial}_c f(x_0)$ where f is a c-convex function given by:

$$f(z) = \min_{(x_i, y_i) \subset S, i=1..n} \sum_{i=0}^{n} c(x_{i+1}, y_i) - c(x_i, y_i),$$

where $(x_0, y_0) \in S$ is as fixed point and $x_{n+1} = z$. Using $c(x, y) = I(x) - W(x, y) - \gamma$ we get,

$$f(z) = min_{(x_i, y_i) \subset S, i=1..n} \sum_{i=0}^{n} I(x_{i+1}) - I(x_i) - [W(x_{i+1}, y_i) - W(x_i, y_i)] =$$

$$= min_{(x_i, y_i) \subset S, i=1..n} \sum_{i=0}^{n} I(x_{i+1}) - I(x_i) + [\Delta(x_i, x_{i+1}, y_i)] =$$

$$= min_{(x_i, y_i) \subset S, i=1..n} I(z) - I(x_0) + \sum_{i=0}^{n} \Delta(x_i, x_{i+1}, y_i).$$

Lemma 2 *If, $(x_i, y_i) \subset S, i = 0, 1, 2$ is such that $x_0 < x_1 < x_2 < z$ and $y_2 < y_1 < y_0$ then, $\Delta(x_0, x_1, y_0) + \Delta(x_1, z, y_1) > \Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \Delta(x_2, z, y_2)$ (Figs. 1 and 2).*

Proof Observe that, $\Delta(x_1, z, y_1) = \Delta(x_1, x_2, y_1) + \Delta(x_2, z, y_1) > \Delta(x_1, x_2, y_1) + \Delta(x_2, z, y_2)$, because $\Delta(x_2, z, \cdot)$ is increasing and $y_1 > y_2$.

Lemma 3 *If, $(x_i, y_i) \subset S, i = 0, 1, 2$ is such that $x_0 < x_1 < z < x_2$ and $y_2 < y_1 < y_0$ then, $\Delta(x_0, x_1, y_0) + \Delta(x_1, z, y_1) < \Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \Delta(x_2, z, y_2)$.*

In particular, $\Delta(x_0, x_1, y_0) + \Delta(x_1, z, y_1) < \Delta(x_0, x_2, y_0) + \Delta(x_2, z, y_2)$ (Figs. 3 and 4).

Proof Observe that, $\Delta(x_1, z, y_1) = \Delta(x_1, x_2, y_1) + \Delta(x_2, z, y_1) < \Delta(x_1, x_2, y_1) + \Delta(x_2, z, y_2)$, because $\Delta(x_2, z, \cdot)$ is decreasing and $y_1 > y_2$.

alopes@mat.ufrgs.br

Fig. 1 Bad

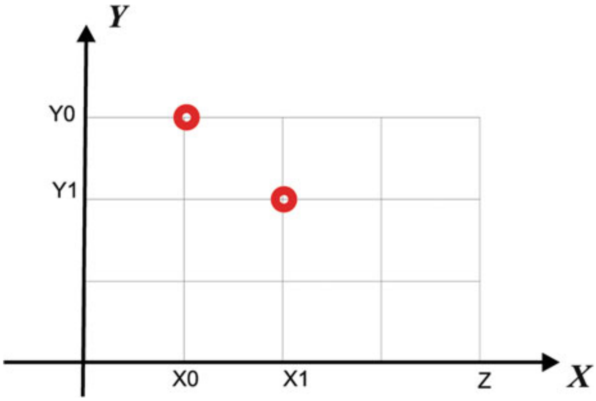


Fig. 2 Good

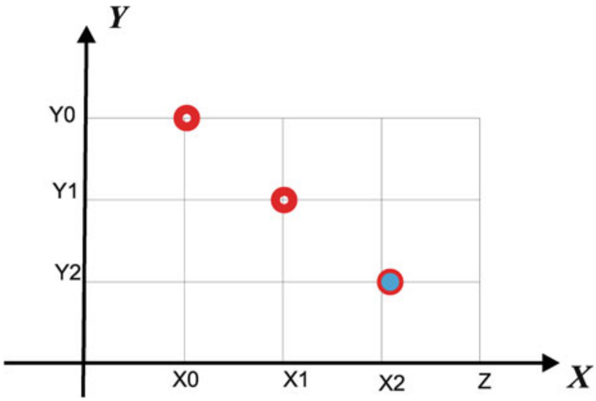


Fig. 3 Bad

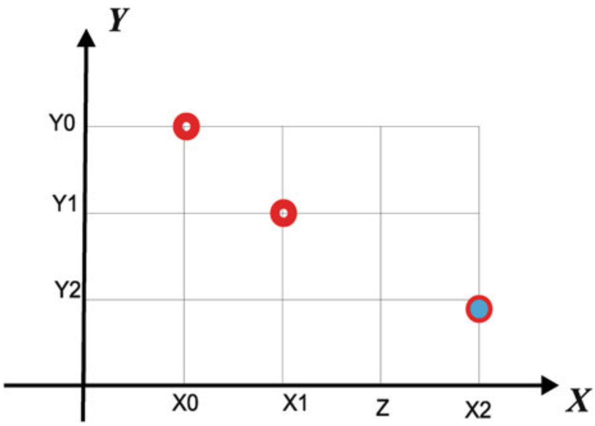
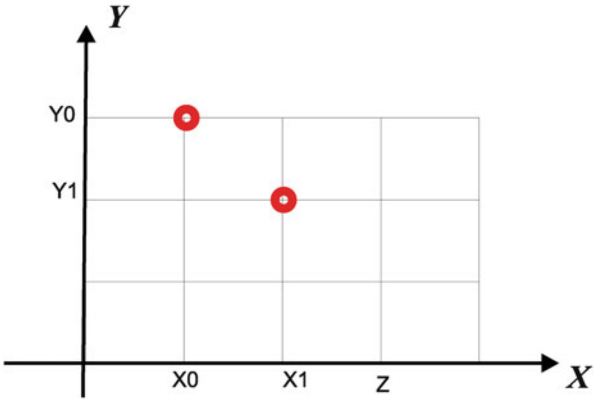


Fig. 4 Good



Now observe that,
$\Delta(x_0, x_2, y_0) + \Delta(x_2, z, y_2) = \Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_0) + \Delta(x_2, z, y_2) >$
$\Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \Delta(x_2, z, y_2) > \Delta(x_0, x_1, y_0) + \Delta(x_1, z, y_1)$.

Now one can generalize the idea above: Suppose that, $(x_i, y_i) \subset S, i = 0, 1, 2, \ldots, n$ is such that $x_0 < x_1 < \ldots < x_k < z < x_{k+1} < \ldots < x_n$ and $y_n < \ldots < y_2 < y_1 < y_0$, then, $\Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \ldots + \Delta(x_k, z, y_k) < \Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \ldots + \Delta(x_n, z, y_n)$.

In order to see this, we proceed by induction in the right side of the inequality above:
$\Delta(x_{n-1}, x_n, y_{n-1}) + \Delta(x_n, z, y_n) > \Delta(x_{n-1}, x_n, y_{n-1}) + \Delta(x_n, z, y_{n-1}) = \Delta(x_{n-1}, z, y_{n-1})$.

In this step we discard the pair (x_n, y_n). We must to repeat this process while $n - j > k$, discarding all points in the right side of z. So the conclusion is, that we can discard all points in the right side of z decreasing the sum, and we can introduce a point between the last point in the left size of z, and z, decreasing the sum.

We discard $(x_2, y_2), (x_3, y_3), (x_4, y_4)$, from right size and insert (A, B) between (x_1, y_1) and z (Figs. 5 and 6).

The case in which $z < x_0$ must be analyzed now:

Observe that:
$\Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \Delta(x_2, x_3, y_2) + \Delta(x_3, x_4, y_3) + \Delta(x_4, x_5, y_4) + \Delta(x_5, z, y_5) >$
$\Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \Delta(x_2, x_3, y_2) + \Delta(x_3, x_4, y_3) + [\Delta(x_4, x_5, y_4) + \Delta(x_5, z, y_4)] =$
$\Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \Delta(x_2, x_3, y_2) + \Delta(x_3, x_4, y_3) + \Delta(x_4, z, y_4)$,

and successively to eliminate 4 and 3.

Now we check what happen with permutations of the order in the projected points.

Note that the sum $\sum_{i=0}^{n} c(x_{i+1}, y_i) - c(x_i, y_i)$ can change by sorting the sequence of points $(x_i, y_i) \subset S, i = 1..n$. So we need to consider the natural question about the better way to rename this points.

Fig. 5 Bad

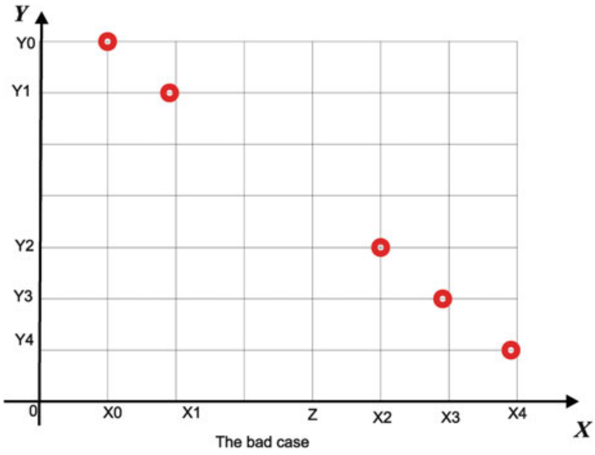
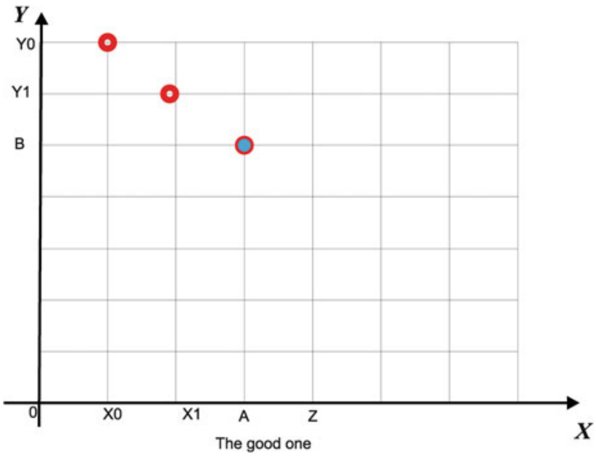


Fig. 6 Good



Please, check the below figure (Figs. 7 and 8):

We claim that it is possible discard all the points at the right side of z and also all the points between x_0 and z that are no ordered in order to minimize the sum above.

In fact:

$$\Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) +$$

$$\Delta(x_2, x_3, y_2) + \Delta(x_3, x_4, y_3) + [\Delta(x_4, x_5, y_4) + \Delta(x_5, z, y_5)] >$$

$$\Delta(x_0, x_1, y_0) + \Delta(x_1, x_2, y_1) + \Delta(x_2, x_3, y_2) + [\Delta(x_3, x_4, y_3) + \Delta(x_4, z, y_4)] >$$

$$\Delta(x_0, x_1, y_0) + [\Delta(x_1, x_2, y_1) + \Delta(x_2, x_3, y_2)] + [\Delta(x_3, z, y_3)] >$$

$$\Delta(x_0, x_1, y_0) + [\Delta(x_1, x_3, y_1) + \Delta(x_3, z, y_3)] >$$

$$\Delta(x_0, x_1, y_0) + \Delta(x_1, z, y_1).$$

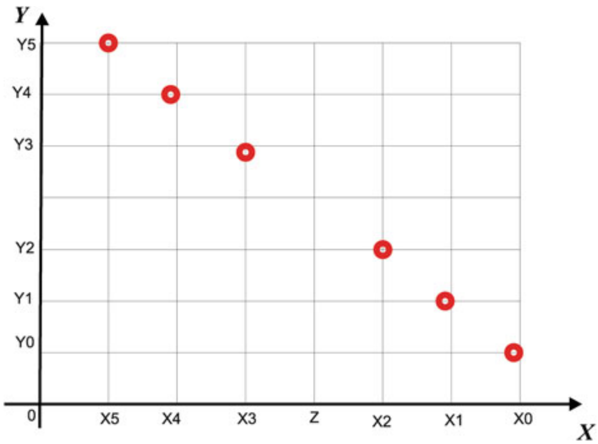


Fig. 7 Bad

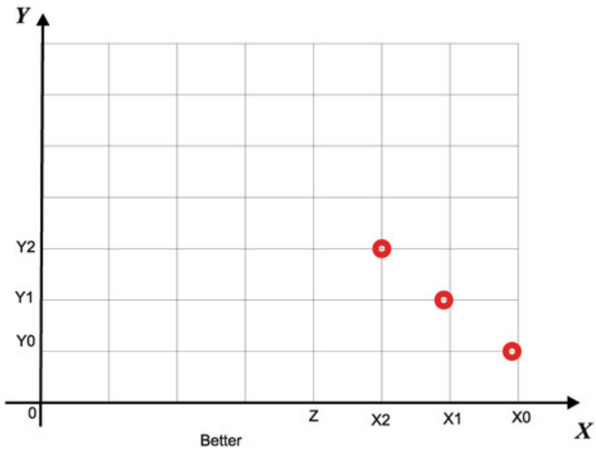


Fig. 8 Good

So the sequence $(x_0, y_0), (x_1, y_1)$ in this order minimize this sum. We know that the graph property is true. But suppose we have a more general case where $\Delta(x, z, y)$ can be consider and we do not have the graph property.

Consider the sequence $(x_0, y_0), (x_1, y_1)$ and suppose $z > x_1 > x_0$. Additionally suppose that $(x_1, .) \cap S \neq \{y_1\}$, so we can compares the sum $\Delta(x_0, x_1, y_0) + \Delta(x_1, z, y_1)$ with $\Delta(x_0, x_1, y_0) + \Delta(x_1, z, y)$ for any $y \in (x_1, .) \cap S \neq \{y_1\}$.

We claim that this function is monotone increasing in y.

In fact suppose that $y' < y_1 < y'' < y_0$, as in Figs. 9 and 10. Observe that, $\Delta(x_1, z, y_1) < \Delta(x_1, z, y'')$ and $\Delta(x_1, z, y_1) > \Delta(x_1, z, y')$ because $x_1 < z$. The conclusion is that if the support of $\hat{\mu}_{max}$ is a periodic orbit, then, we choose

alopes@mat.ufrgs.br

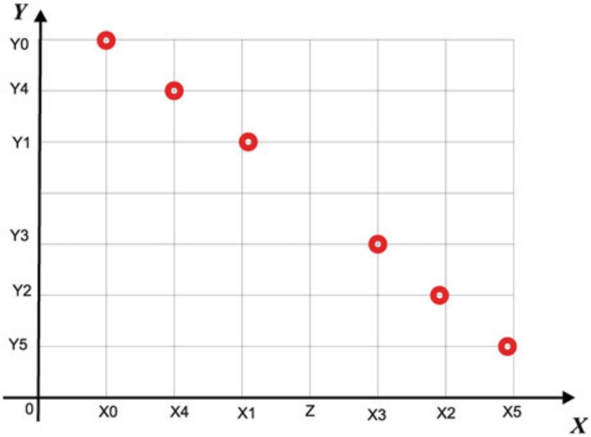


Fig. 9 Too bad

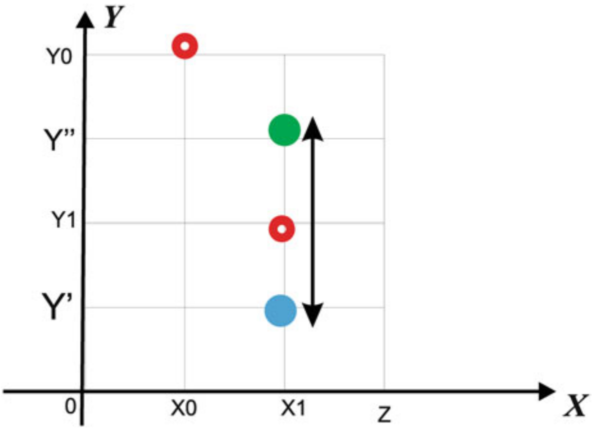


Fig. 10 Going down is better

(x_0, y_0) in the support of $\hat{\mu}_{max}$. From the above, in this case given $z \in \Sigma$, then

$$f(z) = [\,(\,c(z, y_n) - c(x_n, y_n)\,)]+$$
$$(\,c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1})\,) \, + \ldots$$
$$+ \ldots + \{(\,c(x_3, y_2) - c(x_2, y_2)\,)\,\} +$$
$$(\,c(x_2, y_1) - c(x_1, y_1)\,) + [(\,c(x_1, y_0) - c(x_0, y_0)\,) \;].,$$

where we use all the possible x_i, $i = 1, 2, .., n$, on the left of z, and for each x_i we choose the corresponding y_i such that (x_i, y_i) is in the support of $\hat{\mu}_{max}$. Moreover, $x_0 < x_1 < x_2 < \ldots < x_n$.

Finally, we can say that $\hat{\partial}_c f(x_k) = y_k$, for any k.

One can get similar results for the function g (obtained just from the kernel W) defined before.

From the reasoning above (for the case of W satisfying the twist condition), in the case μ_∞ is not a periodic orbit, then in definition of f, the infimum is not attained in a finite sequence of x_n in the support of μ_∞.

Acknowledgements The first author is partially supported by CNPq, CAPES and FAPERGS.

Appendix

Here we consider first the shift $\Sigma = \{0, 1\}^{\mathbb{N}}$, and Σ as a metric space with the usual distance:

$$d(x,y) = \begin{cases} 0, \text{ if } & x = y \\ (1/2)^n, \text{ if } & n = \min\{i \mid x_i \neq y_i\}. \end{cases}$$

Additionally, we suppose that Σ is ordered by $x < y$, if $x_i = y_i$ for $i = 1..n-1$, and $x_n = 0$ and $y_n = 1$.

As the usual, we consider the dynamical system (Σ, σ) where $\sigma : \Sigma \to \Sigma$ is given by $\sigma(x) = \sigma(x_1, x_2, x_3, \ldots) = (x_2, x_3, x_4, \ldots)$.

(a) Potentials and the Involution Kernel

As usual we denote

$$\tau_x^*(y) = (x_1, y_1, y_2, y_3, \ldots) \text{ and } \tau_y(x) = (y_1, x_1, x_2, x_3, \ldots),$$

and

$$\hat{\sigma}(x,y) = (\sigma(x), \tau_x^*(y)) \text{ and } \hat{\sigma}^{-1}(x,y) = (\tau_y x, \sigma^*(y)),$$

the skew product map, where $\sigma^*(y = (y_1, y_2, y_3, \ldots)) = (y_2, y_3, y_4, \ldots)$.

We also define $\tau_{k,y} x = (y_k, y_{k-1}, \ldots y_2, y_1, x_0, x_1, x_2, \ldots)$, where $x = (x_0, x_1, x_2, \ldots)$, $y = (y_1, y_2, y_3, \ldots)$. In a similar way we define $\tau_{k,y}^* x$.

Given a continuous function $A : \Sigma \to \mathbb{R}$, remember that a continuous function $W : \Sigma \times \Sigma \to \mathbb{R}$ is an involution kernel for A if $(W \circ \hat{\sigma}^{-1} - W + A \circ \hat{\sigma}^{-1})(x,y)$ does not depends on x; In this case the continuous function $A^*(y) = (W \circ \hat{\sigma}^{-1} - W + A \circ \hat{\sigma}^{-1})(x,y)$ is called the W-dual potential of A.

As in [2] we define the cocycle $\Delta_A(x, x', y)$, where

$$\Delta_A(x,x',y) = \sum_{n \geq 1} A \circ \hat{\sigma}^{-n}(x,y) - A \circ \hat{\sigma}^{-n}(x',y) = \sum_{n \geq 1} A \circ \tau_{n,y}(x) - A \circ \tau_{n,y}(x'),$$

and its dual version $\Delta_{A^*}(x, y, y')$, where

$$\Delta_{A^*}(x,y,y') = \sum_{n \geq 1} A^* \circ \hat{\sigma}^n(x,y) - A^* \circ \hat{\sigma}^n(x,y') = \sum_{n \geq 1} A^* \circ \tau_{n,x}^*(y) - A^* \circ \tau_{n,x}^*(y').$$

Note that:

(i) $\Delta_A(x,x',y) = -\Delta_A(x',x,y)$, in particular $\Delta_A(x,x,y) = 0$,
(ii) $\Delta_A(x,x',y) + \Delta_A(x',x'',y) = \Delta_A(x,x'',y)$,
(iii) $\Delta_A(x,x',y) = \Delta_A(\tau_y x, \tau_y x', \sigma^*(y)) + [A \circ \tau_y x - A \circ \tau_y x']$,

and the same relations are true for $\Delta_{A^*}(x,y,y')$.

Using this properties one can prove that, for any involution kernel we have $W(x,y) - W(x',y) = \Delta_A(x,x',y)$ and $W(x,y) - W(x,y') = \Delta_{A^*}(x,y,y')$.

From this fact, we get that the difference between two involution kernels for A is a continuous function of y: {Involution kernels for $A\}/C^0(\Sigma) = W^0$, where $W^0(x,y) = \Delta_A(x,x',y)$ for a fix $x' \in \Sigma$ is called a fundamental involution kernel of A. Indeed, the property (iii) shows that W^0 is an involution kernel for A.

On the other hand, given another involution kernel, W we have $W(x,y) - W(x',y) = \Delta_A(x,x',y)$, thus

$$W(x,y) = W(x',y) + \Delta_A(x,x',y) = W(x',y) + W^0(x,y) = g(y) + W^0(x,y),$$

where $g(y) = W(x',y) \in C^0(\Sigma)$.

As an example we compute the general dual potential. First for $W^0(x,y) = \Delta_A(x,x',y)$ we get:

$$\begin{aligned} A_0^*(y) = &\quad (W^0(\tau_y x, \sigma^*(y))) - W^0(x,y) + A(\tau_y x) \\ = &\ \Delta_A(\tau_y x, x', \sigma^*(y)) - \Delta_A(x,x',y) + A(\tau_y x) \\ = &\qquad A(\tau_y x') + \Delta_A(\tau_y x', x', \sigma^*(y)). \end{aligned}$$

Given another involution kernel, W we have $W(x,y) = W(x',y) + W^0(x,y)$ thus

$$A^*(y) = (W \circ \hat{\sigma}^{-1} - W + A \circ \hat{\sigma}^{-1})(x,y) = W(x',\sigma^*(y)) - W(x',y) + A_0^*(y).$$

(b) The Twist Property of an Involution Kernel

If $A : \Sigma \to \mathbb{R}$ is a potential and W an arbitrary involution kernel for A, as we said before, W has the twist property, if for any, $a, b, a', b' \in \Sigma$

$$W(a,b) + W(a',b') < W(a,b') + W(a',b),$$

provided that $a < a'$ and $b < b'$.

If we rewrite this inequality as,

$$\begin{aligned} W(a,b) + W(a',b') &< W(a,b') + W(a',b) \\ W(a,b) - W(a',b) &< W(a,b') - W(a',b') \\ \Delta_A(a,a',b) &< \Delta_A(a,a',b'), \end{aligned}$$

we get an alternative criteria for the twist property, that is, W has the twist property, if for any, $a, a' \in \Sigma$ the function $y \to \Delta_A(a,a',y)$, is strictly increasing, provided that $a < a'$.

Remark 5 This characterization shows a very important fact. The twist property is a property of A, so we can said that A is a twist potential or equivalently A has a twist involution kernel (as, obviously other involution kernel is also twist).

Remark 6 As an initial approximation we can consider a different setting of dynamics. Let $T(x) = -2x \pmod 1$, and

$$\tau_0 x = -\frac{1}{2}x + \frac{1}{2}, \text{ and } \tau_1 x = -\frac{1}{2}x + 1,$$

the inverse branches that defines the skew maps (that are not the actual natural extension of T):

$$\hat{T}(x, y) = (T(x), \tau_x^*(y)) \text{ and } \hat{T}^{-1}(x, y) = (\tau_y x, T^*(y)).$$

So, one can compute an involutive (that is, $A^*(y) = A(y)$) smooth kernel for $A_1(x) = x$ and $A_2(x) = x^2$ given by

$$W_1(x, y) = -\frac{1}{3}(x + y) \text{ and } W_2(x, y) = \frac{1}{3}(x^2 + y^2) - \frac{4}{3}xy.$$

As a corollary we get that any potential $A(x) = a + bx + cx^2$ has a smooth involution kernel given by $W(x, y) = a + bW_1(x, y) + cW_2(x, y)$.

Here and in the next paragraphs, we will denote

$$W_A(x, y) := a + bW_1(x, y) + cW_2(x, y),$$

where $A(x) = a + bx + cx^2$ is a polynomial of degree 2.

We observe that the twist property can be derived from the positivity of the second mix derivative of the involution kernel when it is smooth. Note that,

$$\frac{\partial^2 W_1}{\partial x \partial y} = 0, \text{ and } \frac{\partial^2 W_2}{\partial x \partial y} = -\frac{4}{3},$$

thus W_1 is not twist and W_2 is. Actually any potential $A(x) = a + bx + cx^2$ where $c > 0$ is twist.

Remark 7 In this remark we are going to consider the case of $A(x) = a + bx + cx^2$ where $c < 0$ (not twist). In this case we will be able to compute the calibrated subaction explicitly, which, we believe, it is interesting in itself.

As a first example consider $A(x) = -(x - 1)^2$ which is a convex potential.

From [30, 31] we get that the unique maximizing measure for this potential is $\mu_\infty = \delta_{2/3}$, so the critical value is $m = A(2/3)$. Using the fact that $m = A(2/3)$ one can show that there is a unique (up to constants) calibrated subaction ϕ given by:

$$\phi(x) = W(x, 2/3) - W(2/3, 2/3) = -\frac{1}{3}x^2 + \frac{2}{9}x$$

where the kernel is given by

$$W(x,y) = -(1/3)x^2 - (1/3)y^2 + (4/3)xy - (2/3)x - (2/3)y.$$

As a second example consider $A(x) = -(x-\frac{1}{2})^2$ which it is also a concave potential.

The general arguments in [31] shown that any maximizing measure for this potential is $\mu_\infty = (1-t)\delta_{1/3} + t\delta_{2/3}$, where $t \in [0,1]$, so the critical value is $m = A(1/3) = A(2/3)$. In this case the involutive smooth involution kernel is:

$$W(x,y) = -(1/3)x^2 - (1/3)y^2 + (4/3)xy - (2/3)x - (1/3)y.$$

It is easy to verify that,

$$\phi(x) = V_1(x)\chi_{[(0,1/2)]}(x) + V_2(x)\chi_{[1/2,1]}(x) = \max\{V_1(x), V_2(x)\},$$

is indeed a calibrated subaction for A, where
$V_1(x) = W(x,1/3) - W(1/3,1/3) = \Delta(x,1/3,1/3) = -(1/3)x^2 + (1/9)x$,
$V_2(x) = W(x,2/3) - W(2/3,2/3) = \Delta(x,2/3,2/3) = -(1/3)x^2 + (5/9)x - 2/9$,
Note that,

$$\begin{aligned}
\phi(\tau_0 x) &= V_1(\tau_0 x)\chi_{[(0,1/2)]}(\tau_0 x) + V_2(\tau_0 x)\chi_{[1/2,1]}(\tau_0 x) \\
&= V_1(\tau_0 x) = \Delta(\tau_0 x, 1/3, 1/3) \\
&= \Delta(\tau_{1/3} x, \tau_{1/3} 1/3, T^* 1/3) \\
&= \Delta(x,1/3,1/3) - [A(\tau_{1/3} x) - A(\tau_{1/3} 1/3)] \\
&= V_1(x) - [A(\tau_0 x) - m].
\end{aligned}$$

Thus $\phi(\tau_0 x) + A(\tau_0 x) - m = V_1(x)$. Analogously, $\phi(\tau_1 x) + A(\tau_1 x) - m = V_2(x)$ so

$$\begin{aligned}
\phi(x) &= \max\{V_1(x), V_2(x)\} \\
&= \max\{\phi(\tau_0 x) + A(\tau_0 x) - m, \phi(\tau_1 x) + A(\tau_1 x) - m\} \\
&= \max_{y\in\Sigma}\{\phi(\tau_y x) + A(\tau_y x) - m\}.
\end{aligned}$$

(c) Twist Criteria

Is natural to consider a criteria for the twist property for a class of functions that has a small dependence on the cubic (or higher order) terms. Let $P_2^+ = \{p(x) = a + bx + cx^2 \mid c > 0\}$ be the set of strictly convex polynomial. Consider $p \in P_2^+$, and define

$$\mathscr{C}_\varepsilon(p) = \{A \in C^3([0,1]) | A(x) = p(x) + \varepsilon R(x), \text{ where } \frac{\partial R}{\partial x} \in C^3([0,1])\}$$

Theorem 8 *For any $p \in P_2^+$, there exists $\varepsilon > 0$ such that all $A \in \mathscr{C}_\varepsilon(p)$ is twist.*

Proof Consider $p \in P_2^+$ fixed. So, p has a smooth and involutive involution kernel given by

$$W_p(x,y) = (a + bW_1 + cW_2)(x,y),$$

that is, $p^*(y) = p(y)$, where $W_1(x,y) = -\frac{1}{3}(x+y)$ and $W_2(x,y) = \frac{1}{3}(x^2+y^2) - \frac{4}{3}xy$, are the involution kernel associated to x and x^2 respectively. Let, $A = p + \varepsilon R \in \mathscr{C}_\varepsilon(p)$, and W_R be the involution kernel for R. Since R is C^3 we get that, its corresponding involution kernel W_R is C^2 in the variable x. Using the linearity of the cohomological equation, we get $W_A(x,y) = p(W)(x,y) + \varepsilon W_R(x,y)$, and differentiating with respect to x, we have

$$\begin{aligned}\tfrac{\partial}{\partial x} W_A(x,y) = (b\tfrac{\partial}{\partial x}W_1 + c\tfrac{\partial}{\partial x}W_2)(x,y) + \varepsilon\tfrac{\partial}{\partial x}W_R(x,y) =\\ -\tfrac{1}{3}b + \tfrac{2}{3}cx - \tfrac{4}{3}cy + \varepsilon\tfrac{\partial}{\partial x}W_R(x,y)\end{aligned}$$

Since $-\frac{4}{3}c < 0$, and $\frac{\partial}{\partial x}W_R(x,y) \in C^0([0,1]^2)$ the compactness of $[0,1]^2$ implies that $\frac{\partial}{\partial x}W_A(x,\cdot)$ is a strictly decreasing function for any ε small enough, which is sufficient to ensure the twist property.

Remark 8 If, $A \in C^\infty([0,1])$ is strongly convex, we can consider a perturbation of A of order 2 given by

$$B_\varepsilon(x) = A(0) - A'(0)x + \frac{A''(0)}{2}x^2 + \varepsilon\sum_{n\geq 3}\frac{A^{(n)}(0)}{n!}x^n \in \mathscr{C}_\varepsilon(p_A),$$

where $p_A = A(0) - A'(0)x + \frac{A''(0)}{2}x^2 \in P_2^+$. Thus, we can find $\varepsilon_0 > 0$ such that B_ε is twist for any $0 < \varepsilon < \varepsilon_0$.

(d) The Involution Kernel is Bi-Hölder

We consider now $T(x) = 2x \pmod 1$ on the interval $[0,1]$ and the shift σ on $\Omega = \{0,1\}^{\mathbb{N}}$. A natural question is the regularity of the involution kernel W. We denote τ_j, $j = 0, 1$ the two inverse branches of T. Given $w = (w_1, w_2, \ldots) \in \{0,1\}^{\mathbb{N}}$ we denote by $\tau_{k,w}$ the transformation in $[0,1]$ given by $\tau_{k,w}(x) = (\tau_{w_k} \circ \tau_{w_{k-1}} \circ \ldots \circ \tau_{w_1})(x)$. We have that, for a fixed x_0

$$\Delta(x, x_0, w) = \sum_{k=1}^{\infty} A(\tau_{k,w}(x)) - A(\tau_{k,w}(x_0))$$

and, the involution kernel W can be described as: for any (x,w) we have $W(x,w) = \Delta(x, x_0, w)$. It is easy to see that W is Hölder on the variable x. Consider $a, b \in \Omega$ and suppose that $d(a,b) = 2^{-n}$. In this way $a_j = b_j, j = 1, 2\ldots, n-1, n$. We denote $\bar{a} = \sigma^n(a)$ and $\bar{b} = \sigma^n(b)$.

Proposition 7 *Suppose A is α−Hölder. Consider* $a, b \in \Omega$ *such that* $d(a, b) = 2^{-n}$. *For a fixed* $x \in [0, 1]$ *we have* $|\, W(x, a) - W(x, b)\, | \leq C\,(2^{-n})^{\alpha}$.

Proof Note that for $z = \tau_{n,a}(x) = \tau_{n,b}(x)$ and $z_0 = \tau_{n,a}(x_0) = \tau_{n,b}(x_0)$ we have

$$W(x, a) - W(x, b) = \sum_{k=1}^{\infty} A(\tau_{k,a}(x)) - A(\tau_{k,a}(x_0)) - A(\tau_{k,b}(x)) + A(\tau_{k,b}(x_0)) =$$

$$\sum_{k=1}^{\infty} [\, A(\tau_{k,a}(x)) - A(\tau_{k,b}(x))\,] - [\, A(\tau_{k,a}(x_0)) - A(\tau_{k,b}(x_0))\,] =$$

$$\sum_{k=1}^{\infty} [\, A(\tau_{k,\bar{a}}(z)) - A(\tau_{k,\bar{b}}(z))\,] - [\, A(\tau_{k,\bar{a}}(z_0)) - A(\tau_{k,\bar{b}}(z_0))\,].$$

Note also that $|z - z_0| \leq d(a, b) = 2^{-n}$. Consider $z = z_0 + h$, then

$$A(\tau_{k,\bar{a}}(z_0 + h)) - A(\tau_{k,\bar{a}}(z_0)) \leq C_A\, d(\tau_{k,\bar{a}}(z_0 + h), \tau_{k,\bar{a}}(z_0))^{\alpha} \leq$$

$$C_A\, (\, 2^{-k}\, h\,)^{\alpha} = C_A (\, 2^{-k}\,)^{\alpha}\, h^{\alpha}.$$

Then,

$$\sum_{k=1}^{\infty} [\, A(\tau_{k,\bar{a}}(z)) - A(\tau_{k,\bar{a}}(z_0))\,] - [\, A(\tau_{k,\bar{b}}(z)) - A(\tau_{k,\bar{b}}(z_0))\,]$$

$$\leq C_A \sum_{k=1}^{\infty} 2\, (\, 2^{-k}\,)^{\alpha}\, h^{\alpha} \leq C_A \sum_{k=1}^{\infty} 2\, (2^{\alpha})^{-k}\, h^{\alpha} \leq C\, d(a, b)^{\alpha}.$$

From the above we get:

Theorem 9 *If* $A : S^1 \to \mathbb{R}$ *is Hölder then* $W : S^1 \times \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ *is bi-Hölder.*

(e) The Fenchel-Rockafellar Theorem Given $f : \mathbb{R} \to \mathbb{R}$ defined on the variable x, the Legendre transform of f, denoted by f^*, is the function on the variable p defined by

$$f^*(p) = \sup_{x \in \mathbb{R}} \{p\, x - f(x)\}.$$

Theorem 10 (Fenchel-Rockafellar) *Suppose* $f(x)$ *is smooth strictly convex,* $f : \mathbb{R} \to \mathbb{R}$, *and,* $g(x)$ *is smooth strictly concave,* $g : \mathbb{R} \to \mathbb{R}$. *Denote by* f^* *and* g^* *the corresponding Legendre transforms on the variable p. Then,*

$$\inf_{x \in \mathbb{R}} \{\, f(x) \, - \, g(x)\,\} \, = \, \sup_{p \in \mathbb{R}} \{\, g^*(p) \, - f^*(p)\,\}$$

Fig. 11 The infimum

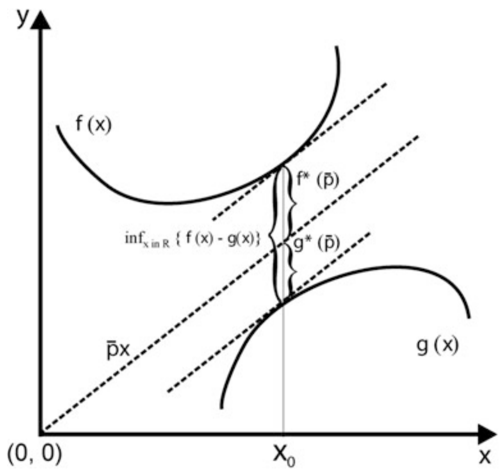
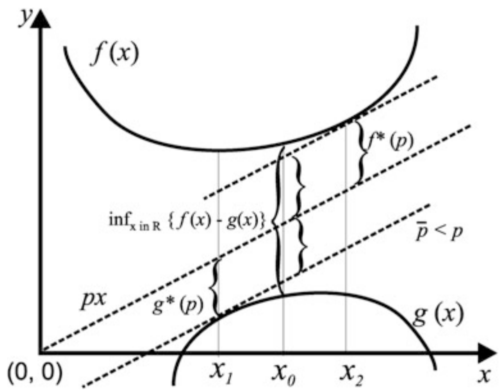


Fig. 12 The supremum



Proof By convexity and concavity properties we have that there exists x_0 such that

$$\inf_{x\in\mathbb{R}} \{ f(x) - g(x) \} = f(x_0) - g(x_0).$$

It is also true that $f'(x_0) - g'(x_0) = 0$. Denote by $\overline{p}$ that value $\overline{p} = f'(x_0)$. We illustrate the proof via two pictures in a certain particular case. Figure 11 shows a geometric picture of the position and values of $f(x_0) - g(x_0)$, $g^*(\overline{p})$ and $f^*(\overline{p})$. Note that in this picture we have that $f(x_0) - g(x_0) > 0$. This picture also shows the graph of $\overline{p}\,x$ as a function of x. We observe that the Legendre transform is not linear on the function. Let's consider different values of p and estimate $f^*(p)$ and $g^*(p)$. Suppose first $p > \overline{p}$. In Fig. 12 we show the graph of $p\,x$, and the values of $f^*(p)$ and $g^*(p)$. We denote by x_2 the value such that

$$f^*(p) = \sup_{x\in\mathbb{R}} \{p\,x - f(x)\} = p\,x_2 - f(x_2).$$

Note that $x_2 > x_0$. We denote by x_1 the value such that

$$0 < g^*(p) = \sup_{x\in\mathbb{R}}\{p\,x - g(x)\} = p\,x_1 - g(x_1).$$

Note that $x_1 < x_0$.
Note also that $f^*(p)$ and $g^*(p)$ have different signs. From this picture one can see that $g^*(p) - f^*(p) < f(x_0) - g(x_0)$. In the case $p < \overline{p}$ a similar reasoning can be done.

References

1. Bangert, V.: Mather sets for twist maps and geodesics on tori. In: Dynamics Reported, vol. 1, pp. 1–56. Wiley, Chichester (1988)
2. Baraviera, A., Lopes, A.O., Thieullen, Ph.: A large deviation principle for equilibrium states of holder potentials: the zero temperature case. Stoch. Dyn. **6**, 77–96 (2006)
3. Baraviera, A.T., Cioletti, L.M., Lopes, A.O., Mohr, J., Souza, R.R.: On the general *XY* Model: positive and zero temperature, selection and non-selection. Rev. Math. Phys. **23**(10), 1063–1113 (2011)
4. Baraviera, A., Lopes, A.O., Mengue, J.: On the selection of subaction and measure for a subclass of potentials defined by P. Walters. Ergodic Theory Dyn. Syst. **33**(5), 1338–1362 (2013)
5. Baraviera, A., Leplaideur, R., Lopes, A.O.: Ergodic optimization, zero temperature limits and the max-plus algebra. In: Mini-course in XXIX Coloquio Brasileiro de Matemática, Rio de Janeiro (2013)
6. Bhattacharya, P., Majumdar, M.: Random Dynamical Systems. Cambridge University Press, Cambridge (2007)
7. Bissacot, R., Garibaldi, E.: Weak KAM methods and ergodic optimal problems for countable Markov shifts. Bull. Braz. Math. Soc. **41**(3), 321–338 (2010)
8. Bousch, T.: Le poisson n'a pas d'arêtes. Ann. Inst. Henri Poincaré Probab. Stat. **36**, 489–508 (2000)
9. Bousch, T.: La condition de Walters. Ann. Sci. l'École Normale Supérieure **34**, 287–311 (2001)
10. Contreras, G.: Ground states are generically a periodic orbit, Arxiv (2013)
11. Contreras, G., Iturriaga, R.: Global minimizers of autonomous Lagrangians. 22° Colóquio Brasileiro de Matemática, IMPA (1999)
12. Contreras,G., Lopes, A.O., Thieullen, Ph.: Lyapunov minimizing measures for expanding maps of the circle. Ergodic Theory Dyn. Syst. **21**, 1379–1409 (2001)
13. Contreras, G., Lopes, A.O., Oliveira, E.: Ergodic Transport Theory, periodic maximizing probabilities and the twist condition. In: D. Zilberman, A. Pinto (eds.) Modeling, Optimization, Dynamics and Bioeconomy. Springer Proceedings in Mathematics, pp. 183–219. Springer, Cham (2014)
14. Conze, J.P., Guivarc'h, Y.: Croissance des sommes ergodiques et principe variationnel, manuscript circa (1993)
15. Delon, J., Salomon, J., Sobolevski, A.: Fast transport optimization for Monge costs on the circle. SIAM J. Appl. Math. **7**, 2239–2258 (2010)
16. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications. Springer, New York (1998)
17. Evans, L., Gomes, D.: Linear programming interpretation of Mather's variational principle. ESAIM Control Optim. Cal. Var. **8**, 693–702 (2002)

18. Galatolo, S., Pacifico, M.: Lorenz-like flows: exponential decay of correlations for the Poincaré map, logarithm law, quantitative recurrence. Ergodic Theory Dyn. Syst. **30**(6), 1703–1737 (2010)
19. Gangbo, W., McCann, R.J.: The geometry of optimal transportation. Acta Math. **177**, 113–161 (1996)
20. Garibaldi, E., Lopes, A.O.: Functions for relative maximization. Dyn. Syst. **22**, 511–528 (2007)
21. Garibaldi, E., Lopes, A.O., Thieullen, Ph.: On calibrated and separating sub-actions. Bull. Braz. Math. Soc. **40**(4), 577–602 (2009)
22. Garibaldi, E., Lopes, A.O.: The effective potential and transshipment in thermodynamic formalism at temperature zero. Stoch. Dyn. **13**(1), 1250009 (13 p) (2013)
23. Garibaldi, E., Thieullen, Ph.: Minimizing orbits in the discrete Aubry-Mather model. Nonlinearity **24**(2), 563–611 (2011)
24. Garibaldi, E., Thieullen, Ph.: Description of some ground states by Puiseux technics. J. Stat. **146**(1), 125–180 (2012)
25. Gole, C.: Sympletic Super-Twist Maps. World Scientific, Singapore (1998)
26. Hunt, B.R., Yuan, G.C.: Optimal orbits of hyperbolic systems. Nonlinearity **12**, 1207–1224 (1999)
27. Jenkinson, O.: Ergodic optimization. Discrete Continuous Dyn. Syst. Ser. A **15**, 197–224 (2006)
28. Jenkinson, O.: Every ergodic measure is uniquely maximizing. Discrete Continuous Dyn. Syst. Ser. A **16**, 383–392 (2006)
29. Jenkinson, O.: A partial order on x2 -invariant measures. Math. Res. Lett. **15**(5), 893–900 (2008)
30. Jenkinson, O.: Optimization and majorization of invariant measures. Electron. Res. Announc. Am. Math. Soc. **13**, 1–12 (2007)
31. Jenkinson, O., Steel, J.: Majorization of invariant measures for orientation-reversing maps. Ergodic Theory Dyn. Syst. **30**(5), 1471–1483 (2010)
32. Kloeckner, B.: Optimal Transport and dynamics of circle expanding maps acting on measures. Ergodic Theory Dyn. Syst. **33**(2), 529–548 (2013)
33. Kloeckner, B., Lopes, A.O., Stadlbauer, M.: Contraction in the Wasserstein metric for some Markov chains, and applications to the dynamics of expanding maps, preprint (2014)
34. Kloeckner, B., Giulietti, P., Lopes, A.O., Marcon, D.: On the Geometry of Thermodynamical Formalism, preprint (2014)
35. Leplaideur, R.: A dynamical proof for the convergence of Gibbs measures at temperature zero. Nonlinearity **18**(6), 2847–2880 (2005)
36. Lopes, A., Mengue, J.: Duality theorems in ergodic transport. J. Stat. Phys. **149**(5), 921–942 (2012)
37. Lopes, A.O., Oliveira, E.R.: On the thin boundary of the fat attractor, preprint UFRGS (2011)
38. Lopes, A.O., Thieullen, Ph.: Sub-actions for Anosov diffeomorphisms. Astérisque **287**, 135–146 (2003)
39. Lopes, A.O., Thieullen, P.: Mather measures and the Bowen-Series transformation. Ann. Inst. Henri Poincaré Analyse non Linéaire **23**, 663–682 (2006)
40. Lopes, A.O., Oliveira, E.R., Smania, D.: Ergodic transport theory and piecewise analytic subactions for analytic dynamics. Bull. Braz. Math. Soc. **43**(3), 467–512 (2012)
41. Lopes, A.O., Mengue, J.K., Mohr, J., Souza, R.R.: Entropy, pressure and duality for Gibbs plans in ergodic transport. Bull. Braz. Math. Soc. (to appear)
42. Lopes, A., Mengue, J.K., Mohr, J., Souza, R.R.: Entropy and variational principle for one-dimensional lattice systems with a general a-priori probability: positive and zero temperature. Ergodic Theory Dyn. Syst. (to appear)
43. Mather, J.: Action minimizing invariant measures for positive definite Lagrangian Systems. Math. Z. **207**(2), 169–207 (1991)
44. Mengue, J.K., Oliveira, E.R.: Duality results for Iterated Function Systems with a general family of branches, preprint Arxiv (2014)

45. Mitra, T.: Introduction to dynamic optimization theory. In: Majumdar, M., Mitra, T., Nishimura, K. (eds.) Optimization and Chaos. Studies in Economic Theory. Springer, Heidelberg (2000)
46. Morris, I.D.: A sufficient condition for the subordination principle in ergodic optimization. Bull. Lond. Math. Soc. **39**(2), 214–220 (2007)
47. Parry, W., Pollicott, M.: Zeta functions and the periodic orbit structure of hyperbolic dynamics. In: Astérisque vol. 187–188. Société mathématique de France, Paris (1990)
48. Rachev, S., Ruschendorf, L.: Mass Transportation Problems, vol. 1 and 2. Springer, New York (1998)
49. Ruscheendorf, L.: On *c*-optimal random variables. Stat. Probab. Lett. **27**, 267–270 (1996)
50. Savchenko, S.V.: Cohomological inequalities for finite Markov chains. Funct. Anal. Appl. **33**, 236–238 (1999)
51. Souza, R.R.: Ergodic and Thermodynamic Games, preprint (to appear in Stochastics and Dynamics, 2014)
52. Tal, F.A., Zanata, S.A.: Maximizing measures for endomorphisms of the circle. Nonlinearity **21**, 2347–2359 (2008)
53. Villani, C.: Topics in Optimal Transportation. AMS, Providence (2003)
54. Villani, C.: Optimal Transport: Old and New. Springer, Berlin (2009)