Optimal transportation and pressure at zero temperature

Jairo K. Mengue

Universidade Federal do Rio Grande do Sul

February 3, 2025

Abstract

Given two compact metric spaces X and Y, a Lipschitz continuous cost function c on $X \times Y$ and two probabilities $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, we propose to study the Monge-Kantorovich problem and its duality from a zero temperature limit of a convex pressure function. We consider the entropy defined by $H(\pi) = -D_{KL}(\pi | \mu \times \nu)$, where D_{KL} is the Kullback-Leibler divergence, and then the pressure defined by the variational principle

$$P(\beta A) = \sup_{\pi \in \Pi(\mu,\nu)} \left[\int \beta A \, d\pi + H(\pi) \right],$$

where $\beta > 0$ and $A = -c$. We will show that it admits a dual formulation and when $\beta \to +\infty$ we recover the solution for the usual Monge-Kantorovich problem and its Kantorovich duality. Such approach is similar to one which is well known in Thermodynamic Formalism and Ergodic Optimization, where β is interpreted as the inverse of the temperature ($\beta = \frac{1}{T}$) and $\beta \to +\infty$ is interpreted as a zero temperature limit.

1 Introduction

We consider two compact metric spaces (X, d_X), (Y, d_Y) and the product space $X \times Y$ with the metric $d_X + d_Y$. We consider also the Borel sigma algebra over each space and two probabilities, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ such that $\mathrm{supp}(\mu) = X$ and $\mathrm{supp}(\nu) = Y$. A probability $\pi \in \mathcal{P}(X \times Y)$ will be called a *transference plan* if it has $x-$marginal μ and $y-$marginal ν, that is,

$$\int f(x) \, d\pi(x,y) = \int f(x) \, d\mu(x) \ \text{ and } \ \int g(y) \, d\pi(x,y) = \int g(y) \, d\nu(y)$$

for any $f \in C(X)$ and $g \in C(Y)$. We denote by $\Pi(\mu, \nu)$ the set of transference plans. Let $c : X \times Y \to \mathbb{R}$ be a Lipschitz continuous function (a cost function) and consider the problem of to find the number

$$\alpha(c) = \min_{\pi \in \Pi(\mu,\nu)} \int c(x,y)\, d\pi$$

and an optimal transference plan (Monge-Kantorovich problem).

A classical result concerning such variational principle is Kantorovich Duality: Let $\Phi_c := \{(\varphi, \psi) \in C(X) \times C(Y) \,|\, c(x,y) \geq \varphi(x) + \psi(y) \,\forall x, y\}$, then

$$\min_{\pi \in \Pi(\mu,\nu)} \int c(x,y)\, d\pi = \sup_{(\varphi,\psi) \in \Phi_c} \int \varphi(x)\, d\mu(x) + \int \psi(y)\, d\nu(y).$$

Such duality can be formulated in a broader sense (instead compact spaces and Lipschitz functions) and in [14, 15] can be found a general exposition of the optimal transportation theory. Anyway we remark that for any $p \geq 1$, if $X = Y$ and the cost function is of the form $c(x,y) = [d(x,y)]^p$ then c is Lipschitz continuos.

As the set $\Pi(\mu, \nu)$ is convex and the map $\pi \mapsto \int c(x,y)\, d\pi$ is affine, the Monge-Kantorovich problem can be interpreted as a variational principle in convex analysis. In the present work we propose to consider the Kullback-Leibler divergence and relative entropy and then a different variational principle with introduction of such entropy (which we call the positive temperature approach). Furthermore, we will recover a solution of the Monge-Kantorovich problem and Kantorovich dual problem as a zero temperature limit, similarly as occur in Thermodynamic Formalism and Ergodic Optimization (see [1]). We believe that section 2 of [11] also helps to clarify our strategy.

If η and ρ are probabilities on a same measurable space, the Kullback-Leibler divergence is known as

$$D_{KL}(\eta|\rho) := \begin{cases} \int \log(\frac{d\eta}{d\rho})\, d\eta & \text{if } \eta \ll \rho \\[2ex] +\infty & \begin{array}{l} \text{if } \eta \text{ is not absolutely continuous} \\ \text{with respect to } \rho \end{array} \end{cases},$$

where $\frac{d\eta}{d\rho}$ denotes the Radon-Nikodym derivative.

Following [8] we define the entropy of a probability $\eta \in \mathcal{P}(X \times Y)$ relative to $\mu \in \mathcal{P}(X)$ by $H^\mu(\eta) = -D_{KL}(\eta \,|\, \mu \times \rho)$, where ρ is the $y-$marginal of η. We refer [8] for additional results and characterization of H^μ. In [8, 9, 10, 11] can be found such entropy in the context of Ruelle Operator and a variational principle of the pressure, which is defined from a dynamical system or an iterated function system (IFS).

Supposing $\pi \in \Pi(\mu, \nu)$ we have $H(\pi) := H^\mu(\pi) = -D_{KL}(\pi \,|\, \mu \times \nu)$. In the present work, using such entropy $H(\pi)$, we introduce the following definition of pressure:

$$P(A) = \sup_{\pi \in \Pi(\mu,\nu)} [\int A\, d\pi + H(\pi)],$$

where $A : X \times Y \to \mathbb{R}$ is continuous. Let us consider the problem of to find a transference plan which attains the supremum for P (an equilibrium). In [7,

9, 10] appear some duality results for variational principles in different settings which consider a mixing of transport theory and ergodic theory. In such works it is considered a dynamic or IFS, but the pressure $P(A)$ above defined does not consider any dynamic. Anyway, in [2] and its correction [3] there is a study of a general concept of pressure (without a dynamic) in convex analysis, being possible to check that it contains $P(A)$ as an example.

Our goal is to prove the following theorem.

Theorem 1.1. *Under above setting, writing $A = -c$ and considering $\beta > 0$:*
1- there is a pair of functions $\varphi_\beta \in C(X)$ and $\psi_\beta \in C(Y)$ such that

$$\int e^{\beta A(x,y)+\varphi_\beta(x)+\psi_\beta(y)} d\mu(x) = 1 \,\forall y \in Y \ \ and \ \ \int e^{A(x,y)+\varphi_\beta(x)+\psi_\beta(y)} d\nu(y) = 1 \,\forall x \in X.$$

Such functions are unique in the following sense: if $(\tilde{\phi}_\beta, \tilde{\psi}_\beta)$ is another pair, then there is a constant d_β such that $\tilde{\phi}_\beta = \phi_\beta + d_\beta$ and $\tilde{\psi}_\beta = \psi_\beta - d_\beta$. Furthermore, these functions are Lipschitz continuous and satisfies $\mathrm{Lip}(\varphi_\beta) \le \beta \, \mathrm{Lip}(A)$ and $\mathrm{Lip}(\psi_\beta) \le \beta \, \mathrm{Lip}(A)$.
2 - The probability $d\pi_\beta := e^{\beta A + \varphi_\beta + \psi_\beta} d\mu \, d\nu$ belongs to $\Pi(\mu, \nu)$. Furthermore,

$$P(\beta A) = \int \beta A \, d\pi_\beta + H(\pi_\beta) = -\int \varphi_\beta \, d\mu_\beta - \int \psi_\beta \, d\nu_\beta.$$

3 - The family of functions $(\frac{\varphi_\beta}{\beta})_{\beta>0}$ and $(\frac{\psi_\beta}{\beta})_{\beta>0}$ are equicontinuous and we can suppose uniformly bounded. Any uniform limit (φ, ψ) of $(\frac{\varphi_\beta}{\beta}, \frac{\psi_\beta}{\beta})$ as $\beta \to +\infty$ is a solution of the Kantorovich dual problem, that is, it belongs to Φ_c, also verifying $\alpha(c) = \int \varphi(x) \, d\mu(x) + \int \psi(y) \, d\nu(y)$. Any weak* limit π of π_β as $\beta \to +\infty$ is a solution of the Monge-Kantorovich problem, that is, it belongs to $\Pi(\mu, \nu)$ and satisfies $\int c \, d\pi = \alpha(c)$.

For compact metric spaces, the Kantorovich-Rubinstein Theorem claims (see Thm. 1.14 in [15]) that under the case $X = Y$ and $c(x, y) = d(x, y)$ (distance function) we have

$$\alpha(c) = \sup_{\mathrm{Lip}(\varphi) \le 1} \int \varphi \, d(\mu - \nu).$$

As a corollary of above theorem we get a solution on the Kantorovich-Rubinstein dual problem.

Corollary 1.2. *Considering the hypotheses and notations of Theorem 1.1 and supposing $X = Y$ and $c(x, y) = d(x, y)$, any uniform limit φ of $\frac{\varphi_\beta}{\beta}$ as $\beta \to +\infty$ is a solution of the Kantorovich-Rubinstein dual problem, that is, $\mathrm{Lip}(\varphi) \le 1$ and $\alpha(c) = \int \varphi \, d(\mu - \nu)$.*

Concerning the velocity of convergence of π_β to the optimal transference plan π, we study if it satisfies a Large Deviation Principle (see [6]).

Proposition 1.3. *Under the setting of Theorem 1.1, supposing there exist the uniform limits $\varphi = \lim_{\beta \to +\infty} \frac{\varphi_\beta}{\beta}$, $\psi = \lim_{\beta \to +\infty} \frac{\psi_\beta}{\beta}$ and the weak* limit $\pi = \lim_{\beta \to +\infty} \pi_\beta$ we have that (π_β) satisfies a large deviation principle with rate function $I(x, y) = c(x, y) - \varphi(x) - \psi(y)$.*

Naturally we can ask if known results of Thermodynamic Formalism and Ergodic Optimization could be translated to Transport Theory. Let us denote, for $A = -c$,

$$m(A) := \sup_{\pi \in \Pi(\mu, \nu)} \int A \, d\pi = -\alpha(c);$$

$$\mathcal{M}_{max}(A) := \{\pi \in \Pi(\mu, \nu) \mid \int A \, d\pi = m(A)\};$$

$$H_{\max} := \sup_{\pi \in \mathcal{M}_{max}(A)} H(\pi)$$

(we can have $H_{\max} = -\infty$). $\mathcal{M}_{max}(A)$ is the set of optimal plans to c and $H_{\max}$ is the biggest of the entropies of optimal plans. Next proposition is similar to a well know result in Thermodynamic Formalism (see [5]). It presents a characterization of the possible limits of π_β using the entropy H.

Proposition 1.4. *Under above setting, with $A = -c$, the function*

$$\beta \mapsto [P(\beta A) - \beta m(A)], \ \beta > 0$$

is non-increasing and $\lim_{\beta \to +\infty} P(\beta A) - \beta m(A) = H_{\max}$. Furthermore, any accumulation point for π_β as $\beta \to +\infty$ has a maximal entropy H over optimal plans, that is, if π_∞ is an accumulation point of π_β as $\beta \to +\infty$, then $\pi_\infty \in \mathcal{M}_{max}(A)$ and $H_{\max} = H(\pi_\infty)$ (which can be $-\infty$).

2 Basic properties of entropy and pressure

We remember that the entropy of a probability $\eta \in \mathcal{P}(X \times Y)$ relative to $\mu \in \mathcal{P}(X)$ was defined as $H^\mu(\eta) = -D_{KL}(\eta \mid \mu \times \rho)$, where ρ is the $y-$marginal of π. From Theorems 4.4 and 5.1 of [8] we get

$$H^\mu(\eta) = -\sup\{\int u(x,y) d\eta \mid \int e^{u(x,y)} d\mu(x) = 1 \, \forall y, \ u \, \text{Lipschitz}\}.$$

If $\pi \in \Pi(\mu, \nu)$ then ν is the $y-$marginal of π and we get

$$H(\pi) := -D_{KL}(\pi \mid \mu \times \nu) = -\sup\{\int u \, d\pi \mid \int e^{u(x,y)} d\mu(x) = 1 \, \forall y, \ u \, \text{Lipschitz}\}. \quad (1)$$

Proposition 2.1. *Properties of the entropy H on $\Pi(\mu, \nu)$:*
1. $H \leq 0$ and $H(\mu \times \nu) = 0$;
2. H is a concave function;
3. H is upper semi-continuous, that is, if π_n converges to π in the weak* topology, then $\limsup_n H(\pi_n) \leq H(\pi)$.

Proof. Let us denote by $N = \{u : X \times Y \to \mathbb{R} \mid \int e^{u(x,y)} d\mu(x) = 1 \, \forall y, \ u \, \text{Lipschitz}\}$. We have, $H(\pi) = \inf_{u \in N}[-\int u \, d\pi]$.

1. If we consider $u = 0$ we get $u \in N$ and then $H(\pi) \leq 0$. On the other hand, $H(\mu \times \nu) = -D_{KL}(\mu \times \nu \mid \mu \times \nu) = 0$.

2. For $\pi_1, \pi_2 \in \Pi(\mu, \nu)$ and $\lambda \in [0,1]$, if $\pi = \lambda \pi_1 + (1 - \lambda)\pi_2$ we have

$$H(\pi) = \inf_{u \in N}[-\lambda \int u d\pi_1 - (1 - \lambda) \int u d\pi_2]$$

$$\geq \inf_{u_1 \in N}[-\lambda \int u_1 d\pi_1] + \inf_{u_2 \in N}[-(1 - \lambda) \int u_2 d\pi_2]$$

$$= \lambda H(\pi_1) + (1 - \lambda)H(\pi_2).$$

3. It suppose that $\pi_n \to \pi$ in the weak* topology.

- If $H(\pi) = -\infty$ then for each $k < 0$ there is $u \in N$ such that $-\int u \, d\pi \leq k$. As $\pi_n \to \pi$, for sufficiently large n we get $-\int u \, d\pi_n \leq k+1$, that is, $H(\pi_n) \leq k+1$. Consequently $\limsup_n H(\pi_n) \leq k+1$. As k is arbitrary, $\limsup_n H(\pi_n) = -\infty$.

- If $H(\pi) > -\infty$, for each $\epsilon > 0$ there is $u \in N$ such that $-\int u \, d\pi \leq H(\pi) + \epsilon$. For sufficiently large n we get $-\int u \, d\pi_n \leq -\int u \, d\pi + \epsilon$, that is, $H(\pi_n) \leq H(\pi) + 2\epsilon$. As ϵ is arbitrary we get $\limsup_n H(\pi_n) \leq H(\pi)$. ∎

Example 2.2. *The entropy H is not affine (see also Example 2.4 in [10]).*

Suppose $X = \{1, 2\}$, $Y = \{1, 2\}$ $\mu = (1/2, 1/2)$, $\nu = (1/2, 1/2)$ and consider the following probabilities in $\{1, 2\} \times \{1, 2\}$:

$$
\pi_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad \pi_2 = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \quad \pi = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}.
$$

We have that $\pi, \pi_1, \pi_2 \in \Pi(\mu, \nu)$, $\pi = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2$, but $H(\pi) = 0$ while $H(\pi_1) = H(\pi_2) = -\log(2)$ (consequently $H(\pi) \neq \frac{1}{2}H(\pi_1) + \frac{1}{2}H(\pi_2)$). Indeed, $\pi = \mu \times \nu$ and then $H(\pi) = 0$. On the other hand, as $X \times Y$ is a finite set, $\pi_1 \ll \mu \times \nu$ and its Radon-Nikodym derivative satisfies (a.e.)

$$
\frac{d(\pi_1)}{d(\mu \times \nu)} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} =: J_1.
$$

We can conclude this, observing that, for any $f : \{1, 2\} \times \{1, 2\} \to \mathbb{R}$, we have

$$
\int f \, d\pi_1 = f(1,1) \cdot \frac{1}{2} + f(2,2) \cdot \frac{1}{2} = \sum_{i,j} f(i,j) \cdot J_1(i,j) \cdot \frac{1}{4} = \int f \cdot J_1 \, d(\mu \times \nu).
$$

Consequently $H(\pi_1) = -\int \log(\frac{d(\pi_1)}{d(\mu \times \nu)}) d\pi_1 = -\log(2)$. Similarly $H(\pi_2) = -\log(2)$.

Notation 2.3. *If Z is a compact metric space, let us denote the supremum norm on $C(Z)$ by $|f|_\infty := \sup_{z \in Z} |f(z)|$.*

Next proposition particularly shows that the pressure in the present work is contained in the definition of pressure in [2] and its correction [3].

Proposition 2.4. *Properties of the pressure*
For any $A, B \in C(X \times Y)$, $c \in \mathbb{R}$ and $\lambda \in [0, 1]$ we have:
1. Image: $P(A) \in [\min(A), \max(A)]$;
2. monotonicity: $A \leq B \Rightarrow P(A) \leq P(B)$;
3. translation invariance: $P(A + c) = P(A) + c$;
4. convexity: $P(\lambda A + (1 - \lambda)B) \leq \lambda P(A) + (1 - \lambda)P(B)$;
5. continuity: $|P(A) - P(B)| \leq |A - B|_\infty$.

Proof. 1. Considering $\pi = \mu \times \nu$ we get

$$
P(A) \geq \iint A \, d\mu d\nu + H(\mu \times \nu) = \iint A \, d\mu d\nu \geq \min(A).
$$

on the other hand, as $H \leq 0$, for any $\pi \in \Pi(\mu, \nu)$ we have

$$\int A \, d\pi + H(\pi) \leq \int A \, d\pi \leq \max(A)$$

and consequently $P(A) \leq \max(A)$.

2. and 3.: it is an immediate consequence of the definition.

4.

$$P(\lambda A + (1 - \lambda)B) = \sup_{\pi}[\lambda \int A \, d\pi + (1 - \lambda) \int B \, d\pi + \lambda H(\pi) + (1 - \lambda)H(\pi)]$$

$$\leq \sup_{\pi_1}[\lambda(\int A \, d\pi_1 + H(\pi_1))] + \sup_{\pi_2}[(1 - \lambda)(\int B \, d\pi_2 + H(\pi_2))]$$

$$= \lambda P(A) + (1 - \lambda)P(B).$$

5. Supposing $P(A) \geq P(B)$ we get

$$P(A) - P(B) = [\sup_{\pi_1} \int A \, d\pi_1 + H(\pi_1)] - [\sup_{\pi_2} \int B \, d\pi_2 + H(\pi_2)]$$

$$\leq [\sup_{\pi_1} \int B \, d\pi_1 + H(\pi_1) + |A - B|_\infty] - [\sup_{\pi_2} \int B \, d\pi_2 + H(\pi_2)]$$

$$= |A - B|_\infty.$$

■

The set $\Pi(\mu, \nu)$ is convex and we say that $\pi \in \Pi(\mu, \nu)$ is a vertex of $\Pi(\mu, \nu)$ if there is not $\pi_1, \pi_2 \in \Pi(\mu, \nu)$ and $\lambda \in (0, 1)$ such that $\pi = \lambda \pi_1 + (1 - \lambda)\pi_2$.

Proposition 2.5. *For each $A \in C(X \times Y)$ there is at least one probability $\pi \in \Pi(\mu, \nu)$ such that $P(A) = \int A \, d\pi + H(\pi)$. It is possible that none of the vertices of $\Pi(\mu, \nu)$ reaches the supremum defining $P(A)$.*

Proof. By definition of supremum, for each n there is a probability $\pi_n \in \Pi(\mu, \nu)$ such that

$$\int A \, d\pi_n + H(\pi_n) \geq P(A) - \frac{1}{n}.$$

As $X \times Y$ is compact there exists a probability π and a subsequence (π_{n_j}) of (π_n) such that $\pi_{n_j} \to \pi$ in the weak* topology. We have $\pi \in \Pi(\mu, \nu)$ and as H is upper semi-continuous we get

$$\int A \, d\pi + H(\pi) \geq \limsup_{n_j}[\int A \, d\pi_{n_j} + H(\pi_{n_j})] = P(A).$$

As the reverse inequality is consequence of the definition of P we get the equality.

Let us present an example where the optimal π is unique but it is not a vertex. Consider $X = \{1, 2\}$, $Y = \{1, 2\}$ $\mu = (1/2, 1/2)$, $\nu = (1/2, 1/2)$ and the probabilities

$$\pi_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad \pi_2 = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \quad \pi_3 = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}.$$

The vertices of $\Pi(\mu, \nu)$ are just the probabilities π_1 and π_2. Following Example 2.2, for $A = 0$ we get that $P(A) = \sup_{\pi \in \Pi(\mu, \nu)} H(\pi) = 0 = \int A \, d\pi_3 + H(\pi_3)$. On the other hand, $\int A \, d\pi_1 + H(\pi_1) = -\log(2) = \int A \, d\pi_2 + H(\pi_2)$. More generally if A is approximately zero the vertices π_1 and π_2 are not optimal for the pressure.

■

3   Proof of Theorem 1.1

3.1   Proof of item 1

We can drop β in the proof of item 1.

Proposition 3.1. *There is a pair of functions $\varphi \in C(X)$ and $\psi \in C(Y)$ such that*

$$\int e^{A(x,y)+\varphi(x)+\psi(y)} d\mu(x) = 1 \,\forall y \in Y \ \ \text{and} \ \ \int e^{A(x,y)+\varphi(x)+\psi(y)} d\nu(y) = 1 \,\forall x \in X.$$

Remark 3.2. *If $X = Y$ is a finite set and $\mu = \nu$ is the probability with uniform distribution, we are constructing a double stochastic matrix and such proposition can be found in [12, 13].*

The proof of this proposition follows some ideas of [4] (see also [11]).

Proof. For each $0 < s < 1$, we define $T^s_\mu : C(X) \to C(Y)$ by

$$T^s_\mu(f)(y) = -\log \int_X e^{A(x,y)+sf(x)} d\mu(x)$$

and $T^s_\nu : C(Y) \to C(X)$ by

$$T^s_\nu(g)(x) = -\log \int_Y e^{A(x,y)+sg(y)} d\nu(y).$$

We have

$$|T^s_\mu(f_1) - T^s_\mu(f_2)|_\infty = \sup_y |\log \int_X e^{A(x,y)+sf_1(x)} d\mu(x) - \log \int_X e^{A(x,y)+sf_2(x)} d\mu(x)|$$
$$\leq s|f_1 - f_2|_\infty$$

and analogously $|T^s_\nu(g_1) - T^s_\nu(g_2)|_\infty \leq s|g_1 - g_2|_\infty$. It follows that $T^s_\nu \circ T^s_\mu$ is a contraction on $C(X)$ and $T^s_\mu \circ T^s_\nu$ is a contraction on $C(Y)$ for the supremum norm. Consequently they have unique fixed points φ^s and ψ^s, respectively.

We claim that $T^s_\mu(\varphi^s) = \psi^s$ and $T^s_\nu(\psi^s) = \varphi^s$. Indeed, denoting $g^s := T^s_\mu(\varphi^s)$ we get $T^s_\mu \circ T^s_\nu(g^s) = T^s_\mu \circ T^s_\nu \circ T^s_\mu(\varphi^s) = T^s_\mu(\varphi^s) = g^s$. As ψ^s is the unique fixed point of $T^s_\mu \circ T^s_\nu$ we conclude that $g^s = \psi^s$ and then $\psi^s = T^s_\mu(\varphi^s)$. Furthermore $T^s_\nu(\psi^s) = T^s_\nu(T^s_\mu(\varphi^s)) = \varphi^s$.

We claim that for any $s \in (0,1)$, the functions φ^s and ψ^s are Lipschitz continuous, with $\text{Lip}(\varphi^s) \leq \text{Lip}(A)$ and $\text{Lip}(\psi^s) \leq \text{Lip}(A)$. Indeed, for any $x_1, x_2 \in X$ we have

$$\varphi^s(x_1) = -\log \int e^{A(x_1,y)+s\psi^s(y)} d\nu(y) = -\log \int e^{A(x_2,y)-A(x_2,y)+A(x_1,y)+s\psi^s(y)} d\nu(y)$$

$$\leq -\log \int e^{A(x_2,y)-\text{Lip}(A)\cdot d(x_1,x_2)+s\psi^s(y)} d\nu(y)$$

$$= -\log \int e^{A(x_2,y)+s\psi^s(y)} d\nu(y) + \text{Lip}(A) \cdot d(x_1, x_2)$$

$$= \varphi^s(x_2) + \text{Lip}(A) \cdot d(x_1, x_2)$$

It follows that $\frac{\varphi^s(x_1)-\varphi^s(x_2)}{d(x_1,x_2)} \leq \text{Lip}(A)$ for any $x_1, x_2 \in X$. Analogously we can obtain that $\text{Lip}(\psi^s) \leq \text{Lip}(A)$.

We have that $-\mathrm{Lip}(A) \cdot diam(X) \leq \varphi^s - \max(\varphi^s) \leq 0$ and then the family $(\varphi^s - \max(\varphi^s))_{0<s<1}$ is equicontinuous and uniformly bounded and similarly to $(\psi^s - \max(\psi^s))_{0<s<1}$. Before to apply Arzelà-Ascoli theorem, let us to consider the numbers $\max(\psi^s) + s \max(\varphi^s)$.

We claim that

$$-\max(A) \leq \max(\psi^s) + s \max(\varphi^s) \leq lip(A) \cdot diam(X) - min(A).$$

Indeed, for any $y \in Y$ we have $e^{-\psi^s(y)} = \int e^{A(x,y)+s\varphi^s(x)} \, d\mu(x)$ and then $-\psi^s(y) \leq \max(A) + s \max(\varphi^s)$. Consequently $-\max(A) \leq \max(\psi^s) + s \max(\varphi^s)$. On the other hand,

$$-\psi^s(y) \geq \min(A) + s \min(\varphi^s) \geq \min(A) + s \max(\varphi^s) - s \, \mathrm{Lip}(A) \cdot diam(X)$$

and then $-\min(A) + \mathrm{Lip}(A) \cdot diam(X) \geq \max(\psi^s) + s \max(\varphi^s)$. This concludes the proof of the claim.

With similar arguments we conclude that $\max(\varphi^s) + s \max(\psi^s)$ is bounded. Therefore, there exist an increasing sequence of positive numbers s_n which converges to 1, real numbers l_1, l_2 and functions $\varphi_\beta \in C(X)$ and $\psi_\beta \in C(Y)$ such that $\max(\psi^{s_n}) + s_n \max(\varphi^{s_n}) \to l_1$, $\max(\varphi^{s_n}) + s_n \max(\psi^{s_n}) \to l_2$, $\varphi^{s_n} - \max(\varphi^{s_n}) \to \varphi$ (uniformly) and $\psi^{s_n} - \max(\psi^{s_n}) \to \psi$ (uniformly). Particularly we get $\max(\varphi_\beta) = \max(\psi_\beta) = 0$, $\mathrm{Lip}(\varphi) \leq \mathrm{Lip}(A)$ and $\mathrm{Lip}(\psi) \leq \mathrm{Lip}(A)$.

We have,

$$
\begin{aligned}
e^{-\psi(y)} &= \lim_n e^{-\psi^{s_n}(y)+\max(\psi^{s_n})} = \lim_n \int e^{A(x,y)+s_n\varphi^{s_n}(x)} \, d\mu(x) e^{\max(\psi^{s_n})} \\
&= \lim_n \int e^{A(x,y)+s_n(\varphi^{s_n}(x)-\max(\varphi^{s_n}))} \, d\mu(x) e^{\max(\psi^{s_n})+s_n\max(\varphi^{s_n})} \\
&= \int e^{A(x,y)+\varphi(x)} \, d\mu(x) \cdot e^{l_1}.
\end{aligned}
$$

Consequently

$$\int e^{A(x,y)+\varphi(x)+\psi(y)+l_1} \, d\mu(x) = 1 \,\, \forall y.$$

With similar computations we get

$$\int e^{A(x,y)+\varphi(x)+\psi(y)+l_2} \, d\nu(y) = 1 \,\, \forall x.$$

Applying Fubini's Theorem we can conclude that

$$\iint e^{A(x,y)+\varphi(x)+\psi(y)+l_1} \, d\nu(y)d\mu(x) = 1 = \iint e^{A(x,y)+\varphi(x)+\psi(y)+l_2} \, d\nu(y)d\mu(x).$$

Then $l_1 = l_2$. Finally, replacing ψ by $\psi+l$ where $l := l_1 = l_2$ we complete the proof. As a final remark, observe that by construction, the number l satisfies

$$-\max(A) \leq l \leq lip(A) \cdot diam(X) - min(A) \tag{2}$$

$\blacksquare$

The proof of next proposition follows ideas of [12, 13].

Proposition 3.3. *Suppose there are two pair of continuous functions* (φ_1, ψ_1), (φ_2, ψ_2) *such that*

$$\int e^{A(x,y)+\varphi_i(x)+\psi_i(y)} d\mu(x) = 1 \,\forall y \in Y \quad and \quad \int e^{A(x,y)+\varphi_i(x)+\psi_i(y)} d\nu(y) = 1 \,\forall x \in Y. \quad (3)$$

Then there is a constant d such that $\varphi_2 = \varphi_1 + d$ *and* $\psi_2 = \psi_1 - d$.

Proof. Let $p(x) = \varphi_2(x) - \varphi_1(x)$ and $q(y) = \psi_2(y) - \psi_1(y)$. We have

$$\int e^{A(x,y)+\varphi_1(x)+p(x)+\psi_1(y)+q(y)} d\mu(x) = 1 \,\forall y \in Y \quad (4)$$

and

$$\int e^{A(x,y)+\varphi_1(x)+p(x)+\psi_1(y)+q(y)} d\nu(y) = 1 \,\forall x \in X. \quad (5)$$

Let us denote by x_0 and y_0 two points satisfying $p_0 := p(x_0) = \min_{x \in X} p(x)$ and $q_0 := q(y_0) = \max_{y \in Y} q(y)$. For any $y \in Y$ we have

$$e^{q(y)} \overset{(4)}{=} \frac{1}{\int e^{A(x,y)+\varphi_1(x)+p(x)+\psi_1(y)} d\mu(x)} \leq \frac{1}{e^{p_0} \int e^{A(x,y)+\varphi_1(x)+\psi_1(y)} d\mu(x)} \overset{(3)}{=} e^{-p_0}$$

and for any $x \in X$ we have

$$e^{p(x)} \overset{(5)}{=} \frac{1}{\int e^{A(x,y)+\varphi_1(x)+\psi_1(y)+q(y)} d\nu(y)} \geq \frac{1}{e^{q_0} \int e^{A(x,y)+\varphi_1(x)+\psi_1(y)} d\nu(y)} \overset{(3)}{=} e^{-q_0}.$$

It follows that $q(y) \leq -p_0$ and $p(x) \geq -q_0$ and then $-p_0 = q_0$.

Now we will prove that p is constant. Indeed, if $p(\tilde{x}) > p_0$ for some $\tilde{x}$, then, denoting by $\epsilon = \frac{p(\tilde{x})-p_0}{2} > 0$, there exists $\delta > 0$ such that $p(x) - p_0 > \epsilon$ for all x in the ball $B(\tilde{x}, \delta)$. It follows that

$$1 \overset{(4)}{=} \int e^{A(x,y)+\varphi_1(x)+p(x)+\psi_1(y_0)+q_0} d\mu(x) \overset{q_0=-p_0}{=} \int e^{A(x,y)+\varphi_1(x)+\psi_1(y_0)} e^{p(x)-p_0} d\mu(x)$$

$$\geq \int e^{A(x,y)+\varphi_1(x)+\psi_1(y_0)} d\mu(x) + \int_{B(\tilde{x},\delta)} e^{A(x,y)+\varphi_1(x)+\psi_1(y_0)} (e^\epsilon - 1) d\mu(x) \overset{(3)}{>} 1$$

This is a contradiction. Consequently $p(x) = p_0$ for any $x \in X$. Analogously we can show that q is constant. Finally as $q_0 = -p_0$ we get $q = -p$ and the proof is complete.

∎

3.2 Proof of item 2

Consider φ_β and ψ_β as defined in item 1. Then let us define the probability $\pi_\beta \in \mathcal{P}(X \times Y)$ by

$$\int u(x,y) \, d\pi_\beta(x,y) := \int e^{\beta A(x,y)+\varphi_\beta(x)+\psi_\beta(y)} u(x,y) d\mu(x) \, d\nu(y),$$

where $u \in C(X \times Y)$. We have then $d\pi_\beta := e^{\beta A+\varphi_\beta+\psi_\beta} d\mu \, d\nu$ and consequently $\pi_\beta \ll \mu \times \nu$ with Radon-Nikodym derivative $\frac{d\pi_\beta}{d(\mu \times \nu)} = e^{\beta A+\varphi_\beta+\psi_\beta}$. Observe that for any $f \in C(X)$, applying Fubini's Theorem, we have

$$\int f(x) \, d\pi(x,y) = \int e^{\beta A(x,y)+\varphi_\beta(x)+\psi_\beta(y)} f(x) d\mu(x) \, d\nu(y)$$
$$= \int e^{\beta A(x,y)+\varphi_\beta(x)+\psi_\beta(y)} d\nu(y) f(x) \, d\mu(x)$$
$$= \int 1 \cdot f(x) \, d\mu(x) = \int f(x) \, d\mu(x).$$

Analogously we have $\int g(y)\,d\pi(x,y) = \int g(y)\,d\nu(y)\ \forall g \in C(Y)$ and consequently $\pi_\beta \in \Pi(\mu,\nu)$.

By definition of entropy H, as $\frac{d\pi_\beta}{d(\mu\times\nu)} = e^{\beta A + \varphi_\beta + \psi_\beta}$ we have

$$H(\pi_\beta) = -\int \beta A + \varphi_\beta + \psi_\beta\,d\pi_\beta = -\int \beta A d\pi_\beta - \int \varphi_\beta\,d\mu - \int \psi_\beta\,d\nu$$

and then

$$\int \beta A\,d\pi_\beta + H(\pi_\beta) = -\int \varphi_\beta\,d\mu - \int \psi_\beta\,d\nu. \tag{6}$$

Let us show now that

$$P(\beta A) = \sup_{\pi \in \Pi(\mu,\nu)} \int \beta A\,d\pi + H(\pi) = -\int \varphi_\beta\,d\mu - \int \psi_\beta\,d\nu. \tag{7}$$

Let $\pi \in \Pi(\mu,\nu)$ be any transference plan. From equation (1) we have

$$H(\pi) = \inf\{-\int u(x,y)d\pi \mid \int e^{u(x,y)}d\mu(x) = 1\,\forall y,\ u\,\text{Lipschitz}\}.$$

In this way $H(\pi) \leq -\int \beta A + \varphi_\beta + \psi_\beta\,d\pi$ and then, as $\pi \in \Pi(\mu,\nu)$,

$$\int \beta A\,d\pi + H(\pi) \leq -\int \varphi_\beta\,d\mu - \int \psi_\beta\,d\nu.$$

This inequality combined with (6) concludes the proof of equation (7).

3.3   Proof of item 3

From item 1. we have $\text{Lip}(\varphi_\beta) \leq \text{Lip}(\beta A)$ and then $\text{Lip}(\frac{\varphi_\beta}{\beta}) \leq \text{Lip}(A)$. Analogously $\text{Lip}(\frac{\psi_\beta}{\beta}) \leq \text{Lip}(A)$. Furthermore, (following the proof of Proposition 3.1) we can suppose that $-\text{Lip}(A)diam(X) \leq \varphi_\beta/\beta \leq 0$ and $-\text{Lip}(A)diam(Y) + \frac{l_\beta}{\beta} \leq \psi_\beta/\beta \leq \frac{l_\beta}{\beta}$, where l_β satisfies (see equation (2))

$$-\beta \max(A) \leq l_\beta \leq \beta\,\text{Lip}(A)diam(X) - \beta\min(A).$$

From Arzelà-Ascoli Theorem, there exists an increasing sequence $\beta_n \to +\infty$ and Lipschitz functions φ and ψ such that $\frac{\varphi_{\beta_n}}{\beta_n}$ converges uniformly to φ and $\frac{\psi_{\beta_n}}{\beta_n}$ converges uniformly to ψ.

A proof of the following result can be found in [11].

Lemma 3.4. *Let U, Z be compact metric spaces. Let $W_\beta : U \times Z \to \mathbb{R}$ be a family of measurable functions converging uniformly to a continuous function $W : U \times Z \to \mathbb{R}$, as $\beta \to +\infty$, and let ρ be a finite measure on U with $\text{supp}(\rho) = U$. Then*

$$\frac{1}{\beta} \log \int_U e^{\beta W_\beta(u,z)}d\rho(u) \to \sup_{u \in U} W(u,z)$$

uniformly on Z, as $\beta \to +\infty$. The same is true if we replace β by a sequence β_n which converges to $+\infty$.

Using this lemma and taking $\lim_{\beta_n \to +\infty} \frac{1}{\beta_n} \log$ in both sides of the equations

$$\int e^{\beta A(x,y)+\varphi_\beta(x)+\psi_\beta(y)} d\mu(x) = 1 \,\forall y \in Y \ \text{ and } \ \int e^{\beta A(x,y)+\varphi_\beta(x)+\psi_\beta(y)} d\nu(y) = 1 \,\forall x \in X$$

we get

$$\sup_{x \in X} A(x,y)+\varphi(x)+\psi(y) = 0 \,\forall y \in Y \text{ and } \sup_{y \in Y} A(x,y)+\varphi(x)+\psi(y) = 0 \,\forall x \in X. \quad (8)$$

Particularly $\sup_{x,y} A(x,y)+\varphi(x)+\psi(y) = 0$ and then, as $A = -c$ we get $(\varphi, \psi) \in \Phi_c$. Furthermore, (φ, ψ) is a pair of conjugate $c-$concave functions (see [15] p. 33), that is,

$$\psi(y) = \inf_{x \in X}[c(x,y) - \varphi(x)] \,\forall y \in Y \text{ and } \varphi(x) = \inf_{y \in Y}[c(x,y) - \psi(y)] \,\forall x \in X. \quad (9)$$

If π_∞ is any limit of π_β in the weak* topology (as $\beta \to +\infty$), then clearly $\pi_\infty \in \Pi(\mu, \nu)$. Let us suppose that π_∞ is a limit from a subsequence of β_n, which we also denote by β_n. Taking $\lim_{\beta_n \to +\infty} \frac{1}{\beta_n}(\cdot)$ in both sides of equation (6) and using that H is non-positive we get

$$\int A \, d\pi_\infty \geq - \int \varphi \, d\mu - \int \psi \, d\nu,$$

and then (as $A = -c$)

$$\int c \, d\pi_\infty \leq \int \varphi \, d\mu + \int \psi \, d\nu.$$

Consequently

$$\alpha(c) \leq \int c \, d\pi_\infty \leq \int \varphi \, d\mu + \int \psi \, d\nu \leq \alpha(c)$$

(for the last inequality we use that for any $\pi \in \Pi(\mu, \nu)$ we have $\int \varphi \, d\mu + \int \psi \, d\nu \leq \int c \, d\pi$, because $\varphi + \psi \leq c$). Therefore $\alpha(c) = \int \varphi \, d\mu + \int \psi \, d\nu = \int c \, d\pi_\infty$. Particularly we conclude that any limit (φ, ψ) is an optimal dual pair.

On the other hand, if we start by considering π_∞ as any limit of π_β, for example taking an increasing sequence β_m, then from such sequence we can take a subsequence (also denoted by β_m) and Lipschitz functions φ and ψ such that $\frac{\varphi_{\beta_m}}{\beta_m}$ converges uniformly to φ and $\frac{\psi_{\beta_m}}{\beta_m}$ converges uniformly to ψ. Finally, we can reply above computations to reach the same conclusion and then π_∞ is an optimal transference plan.

4   Other proofs

4.1   Proof of Corollary 1.2

We follow the proof of item 3. of Theorem 1.1 and consider an increasing sequence $\beta_n \to +\infty$ and a function φ such that $\frac{\varphi_{\beta_n}}{\beta_n}$ converges uniformly to φ. We can take a subsequence of (β_n) also denoted by (β_n) such that $(\frac{\psi_{\beta_n}}{\beta_n})$ converges uniformly to a function ψ. We remark that $\text{Lip}(\varphi) \leq \text{Lip}(c)$ and $\text{Lip}(\psi) \leq \text{Lip}(c)$.

Supposing $X = Y$ and $c(x,y) = d(x,y)$ (distance function) we obtain from equation (9)

$$\psi(y) = \inf_{x \in X}[d(x,y) - \varphi(x)] \,\forall y \in Y \text{ and } \varphi(x) = \inf_{y \in Y}[d(x,y) - \psi(y)] \,\forall x \in X. \quad (10)$$

We remark that $\mathrm{Lip}(d) = 1$. Indeed, for any $(x_1, y_1) \neq (x_2, y_2) \in X \times X$ we have

$$\frac{d(x_1, y_1) - d(x_2, y_2)}{d_{X \times Y}((x_1, y_1), (x_2, y_2))} = \frac{d(x_1, y_1) - d(x_2, y_2)}{d(x_1, x_2) + d(y_1, y_2)} \leq \frac{d(x_1, x_2) + d(x_2, y_1) - d(x_2, y_2)}{d(x_1, x_2) + d(y_1, y_2)}$$

$$\leq \frac{d(x_1, x_2) + d(y_1, y_2)}{d(x_1, x_2) + d(y_1, y_2)} = 1$$

and for $x_1 \neq x_2 = y_1 = y_2$ we reach the equality.

Let us show that $\psi = -\varphi$. Analyzing just the function ψ, as $\mathrm{Lip}(\psi) \leq 1$ we get $\psi(y) - \psi(x) \leq 1 \cdot d(x, y)$ for any x, y and consequently $-\psi(x) \leq \inf_y[d(x, y) - \psi(y)]$. As (considering $y = x$ in the infimum) $-\psi(x) \geq \inf_y[d(x, y) - \psi(y)]$ we conclude that $-\psi(x) = \inf_y[d(x, y) - \psi(y)] \overset{(10)}{=} \varphi(x)$.

Applying item 3. of Theorem 1.1 we get

$$\alpha(d) = \alpha(c) = \int \varphi \, d\mu + \int \psi \, d\nu = \int \varphi \, d\mu - \int \varphi \, d\nu = \int \varphi \, d(\mu - \nu).$$

4.2 Proof of Proposition 1.3

We suppose that there exists the uniform limits $\varphi = \lim_{\beta \to +\infty} \frac{\varphi_\beta}{\beta}$, $\psi = \lim_{\beta \to +\infty} \frac{\psi_\beta}{\beta}$ and the weak* limit $\pi = \lim_{\beta \to +\infty} \pi_\beta$. We will show that π_β satisfies a large deviation principle with rate function $I(x, y) = c(x, y) - \varphi(x) - \psi(y)$.

Definition 4.1. *Let $(\mu_\beta)_{\beta > 0}$ be a family of probabilities on a metric space Ω. We say that (μ_β) satisfy a large deviation principle (LDP) if there exists a lower semi-continuous rate function $I : \Omega \to [0, +\infty]$ such that*
1. $\limsup_{\beta \to +\infty} \frac{1}{\beta} \log \mu_\beta(C) \leq -\inf_{\omega \in C} I(\omega)$, for any closed set $C \subset \Omega$;
2. $\liminf_{\beta \to +\infty} \frac{1}{\beta} \log \mu_\beta(U) \geq -\inf_{\omega \in U} I(\omega)$, for any open set $U \subset \Omega$.

We refer [6] for general results concerning large deviations. We will apply the following result in this section:

Lemma 4.2. *Consider the family $(\pi_\beta)_{\beta > 0}$ of probabilities on $X \times Y$. Suppose that for any function $f \in C(X \times Y)$ there exists the limit*

$$\Gamma(f) := \lim_{\beta \to \infty} \frac{1}{\beta} \log(\int e^{\beta f} \, d\pi_\beta).$$

Then (π_β) satisfies a LDP with rate function $I(x, y) = \sup_{x,y}[f(x, y) - \Gamma(f)]$. Furthermore,

$$\Gamma(f) = \sup_{x,y}[f(x, y) - I(x, y)], \, \forall \, f \in C(X \times Y). \tag{11}$$

Proof. See Theorem 4.4.2 in [6] (Bric's inverse Varadhan's Lemma). ∎

Remark 4.3. *Using continuous functions of the form*

$$\delta^n_{(x_0, y_0)}(x, y) = \begin{cases} -n^2 \left[d(x, x_0) + d(y, y_0) \right] & \text{if } 0 \leq d(x, x_0) + d(y, y_0) < 1/n \\ -n & \text{if } 1/n \leq d(x, x_0) + d(y, y_0) \end{cases}$$

we can conclude that there is a unique lower semi-continuous function I satisfying equation (11).

Now we complete the proof of Proposition 1.3. We have, by definition of π_β (see item 2 of Theorem 1.1),

$$\Gamma(f) = \lim_{\beta \to \infty} \frac{1}{\beta} \log(\int e^{\beta A(x,y) + \varphi_\beta(x) + \psi_\beta(y) + \beta f(x,y)} \, d\mu(x) d\nu(y)).$$

From lemma 3.4 we conclude that $\Gamma(f) = \sup_{x,y}[A(x,y) + \varphi(x) + \psi(y) + f(x,y)]$. Consequently (π_β) satisfies a LDP with (a continuous) rate function $I(x,y) = c(x,y) - \varphi(x) - \psi(y)$.

4.3   Proof of Proposition 1.4

The proof below follows ideas present in [5].

Proof. Let π_∞ be an accumulation probability measure of the family (π_β) in the weak* topology, as $\beta \to +\infty$. We know from Theorem 1.1 that $\pi_\infty \in M_{max}(A)$. We have, for $\epsilon > 0$,

$$P(\beta A) \geq \beta \int A \, d\pi_{\beta+\epsilon} + H(\pi_{\beta+\epsilon}) = (\beta + \epsilon) \int A \, d\pi_{\beta+\epsilon} + H(\pi_{\beta+\epsilon}) - \epsilon \int A \, d\pi_{\beta+\epsilon}$$
$$= P((\beta + \epsilon)A) - \epsilon \int A \, d\pi_{\beta+\epsilon}.$$

Then

$$P((\beta + \epsilon)A) \leq P(\beta A) + \epsilon \int A \, d\pi_{\beta+\epsilon}$$

and therefore

$$P((\beta + \epsilon)A) - (\beta + \epsilon)m(A) \leq P(\beta A) + \epsilon \int A \, d\pi_{\beta+\epsilon} - (\beta + \epsilon)m(A)$$
$$= P(\beta A) - \beta m(A) + \epsilon(\int A \, d\pi_{\beta+\epsilon} - m(A)) \leq P(\beta A) - \beta m(A).$$

This shows that $\beta \mapsto [P(\beta A) - \beta m(A)]$ is not increasing.

We have

$$[P(\beta A) - \beta m(A)] = [H(\pi_\beta) + \beta(\int A \, d\pi_\beta - m(A))] \leq H(\pi_\beta).$$

As $[P(\beta A) - \beta m(A)]$ is not increasing and the entropy H is upper semi-continuous, if $\mu_{\beta_i} \to \mu_\infty$ we have

$$\lim_{\beta \to +\infty} [P(\beta A) - \beta m(A)] = \lim_{\beta_i \to +\infty} P(\beta_i A) - \beta_i m(A) \leq \limsup_{\beta_i \to +\infty} H(\pi_{\beta_i}) \leq H(\pi_\infty).$$

On the other hand, for any $\pi \in \mathcal{M}_{max}(A)$ we have $\beta m(A) = \beta \int A \, d\pi$ and

$$\lim_{\beta \to +\infty} [P(\beta A) - \beta m(A)] \geq \lim_{\beta \to +\infty} [[\beta \int A \, d\pi + H(\pi)] - \beta \int A \, d\pi] = H(\pi).$$

This shows that $H(\mu_\infty) = H_{\max}$ and that $\lim_{\beta \to +\infty} [P(\beta A) - \beta m(A)] = H_{\max}$.

∎

References

[1] A. Baraviera, R. Leplaideur, and A. Lopes. *Ergodic optimization, zero temperature limits and the max-plus algebra. Paper from the 29th Brazilian mathematics colloquium – 29º Colóquio Brasileiro de Matemática, Rio de Janeiro, Brazil, July 22 – August 2, 2013.* Publ. Mat. IMPA. Rio de Janeiro: Instituto Nacional de Matemática Pura e Aplicada (IMPA), 2013.

[2] A. Bis, M. Carvalho, M. Mendes, and P. Varandas. A convex analysis approach to entropy functions, variational principles and equilibrium states. *Commun. Math. Phys.* 394(1):215–256, 2022.

[3] A. Bis, M. Carvalho, M. Mendes, P. Varandas, and X. Zhong. Correction to: A convex analysis approach to entropy functions, variational principles and equilibrium states *Commun. Math. Phys.* 401(3):3335–3342, 2023.

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

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

[6] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.*, volume 38 of *Stoch. Model. Appl. Probab.* Berlin: Springer, 2nd ed., corrected 2nd printing edition, 2010.

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

[8] A. Lopes and J. Mengue. On information gain, Kullback-Leibler divergence, entropy production and the involution kernel. *Discrete Contin. Dyn. Syst.*, 42(7):3593–3627, 2022.

[9] A. Lopes, J. Mengue, J. Mohr and R. Souza. Entropy, Pressure and Duality for Gibbs plans in Ergodic transport. *Bull Braz Math Soc* 46, 353–389, 2015.

[10] J. Mengue and E. Oliveira. Duality results for iterated function systems with a general family of branches. *Stoch. Dyn.*, 17(3):23, 2017. Id/No 1750021.

[11] J. Mengue and E. Oliveira. Large Deviation for Gibbs Probabilities at Zero Temperature and Invariant Idempotent Probabilities for Iterated Function Systems. *Journal of Statistical Physics*, 192, 18, 2025.

[12] R. Sinkhorn. A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices. *The Annals of Mathematical Statistics*, vol. 35, no. 2, 876–879, 1964.

[13] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics* Vol. 21, No. 2, 343–348, 1967.

[14] C. Villani. Optimal transport : old and new Springer Berlin, Heidelberg, 2009.

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