Paris Saclay University
LMO

Master's thesis

Multi-Marginal Optimal Partial Transport

Author: Hamza Ennaji

Advisor: Prof. Quentin Mérigot

Contents

Acknowledgement

I would like to thank my supervisor, Prof. Quentin Mérigot, for the patient guidance, encouragement and advice he has provided throughout this period. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly. I would also like to thank all the professors of the LMO for the teachings they gave this year.

I must express my gratitude to my parents for their continued support and encouragement. I was continually amazed by their patience while experiencing all of the ups and downs of my stay in France.

Abstract

The purpose of this work is to present the Multi-Marginal Optimal Partial Transport problem, by modelling a problem arising from engineering. In the first section, we give a short overview on the Optimal Transport, Partial Transport problems and we introduce the Multi-Marginal problem with some existence and duality results. While in the second section we present the Multi-Marginal Partial problem, we give few existence and duality results and we prove a modulus of continuity result for the cost function. The last section is devoted to the notion of entropic regularization of optimal transport problems. We prove a $\Gamma-$convergence result for the regularized problem for a fixed regularization parameter and a discretization parameter tending to zero. Finally, we present some algorithms that are used for solving OT problems with some numerical results.

1 Notation

- The spaces in use X, Y, X_i are either compact spaces or the Euclidean space $\mathbb{R}^d$.
- $C(X), C_b(X)$ are the spaces of continuous and bounded continuous functions on X.
- $\|.\|$ stands for the Euclidean norm on $\mathbb{R}^d$, $i.e$, $\|x\|^2 = \sum_{i=1}^d |x_i|^2$ for $x \in \mathbb{R}^d$.
- $\mathcal{M}(X)$ and $\mathcal{P}(X)$ denote respectively the space of measures and probability measures on X.
- For $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ we have

$$\Gamma(\mu, \nu) = \{\gamma \in \mathcal{P}(X \times Y), (\pi_X)_\sharp \gamma = \mu,\ (\pi_Y)_\sharp \gamma = \nu\}$$

$$\Gamma_{\leq m}(\mu, \nu) = \{\gamma \in \mathcal{M}_+(X \times Y), (\pi_X)_\sharp \gamma \leq \mu,\ (\pi_Y)_\sharp \gamma \leq \nu,\ \gamma(X \times Y) = m.\}$$

  where $m \in (0, 1)$ and π_X, π_Y are the canonical projections.
- For $\phi \in C_b(X), \psi \in C_b(Y)$, $\phi \oplus \psi$ denotes $\phi(x) + \psi(y)$.
- $\mathbb{1} = (1, \cdots, 1)^\dagger$ is the unit vector of $\mathbb{R}^d$.
- δ_x is the Dirac measure at a point x.
- For $\phi \in C_b(X)$ and $\mu \in \mathcal{P}(X)$, $< \phi \mid \mu >= \int_X \phi(x) d\mu(x)$.
- $\mu \ll \nu$: μ is absolutely continuous with respect to ν.
- $\mu_n \rightharpoonup \mu$ means that the sequence of probability measures μ_n converges to μ in duality with C_b.

2 Introduction

Multimarginal optimal transport has began to attract considerable attention duo to its applications in a variety of domains such as economics, with for example, the matching team problem [11], and in physics, with density functional theory [8, 13]. This problem, which is considered as a generalization of the Monge-Kantorovich problem, was generalized in turn to the Multimarginal Optimal Partial Transport Problem by linking it with the so-called Barycentre problem proposed in [1]. Our work on the MMOPT is motivated by modelling a real life problem arising from engineering, namely form the operating process of nuclear plants.

A nuclear power plant produces heat energy form atoms rather than burning coal, or other fuel. The produced heat is used to make steam turbines connected to generators which produce electricity for example. Uranium fuel is loaded up into the reactor. In the core of the reactor, the atoms split apart and release heat energy producing neutrons. Control roads made of materials such as cadmium & boron can be raised of lowered into the reactor to soak up neutrons and slow down or speed up the chain reaction. Water is pumped through the reactor to collect the heat produced by the reaction. It circulates in a closed loop linking the reactor with a heat exchanger, which is a device allowing heat to move form a fluid to another avoiding any mixture or contact between the two fluids. More precisely, the water from the reactor gives its energy to cooler water circulating in another closed loop, transforming it into steam. Since the two loops are unconnected, the contaminated water with radioactivity is kept safely contained in one place and well away from most of the equipment. The temperature of the water which is released in the environment is higher than the temperature of the incoming water. Electricity companies have developed software which are able to estimate this temperature as a function of a few parameters of the plant: the temperature of the cool water $y_1 \in \mathbb{R}$, its debit $y_2 \in \mathbb{R}$, the radioactivity level in the primary circuit $y_3 \in \mathbb{R}$, the steam pressure $y_4 \in \mathbb{R}$, the turbine inlet temperature $y_5 \in \mathbb{R}$ and so on. Given these parameters, the software outputs an estimation of the temperature of the released water $b(y_1, \ldots, y_N)$. In order to avoid harming the ecosystem, one wishes to guarantee that these output temperature will never (or very rarely) become too high.

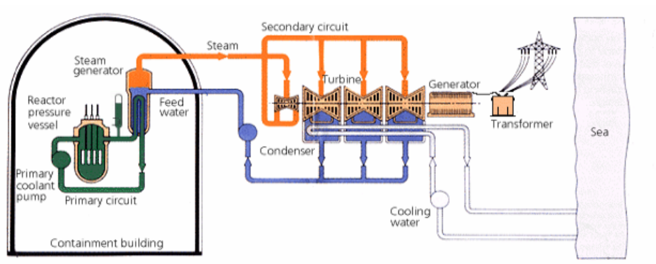

Figure 1: Nuclear power station using pressurised water reactor

If we know the law of the parameters $y_1, \ldots, y_N$, described by a probability distribution $\gamma \in \mathcal{P}(\mathbb{R}^N)$, we can compute the law of the output temperatures by taking the pushforward under b, denoted $b_\sharp\gamma \in \mathcal{P}(\mathbb{R})$. We denote by F_γ the cumulative distribution function of $b_\sharp\gamma$ defined by $F_\gamma(x) = b_\sharp\gamma((-\infty, x])$. If we are given a percentage $m \in (0,1)$, the generalized inverse of F_γ at level m is the minimum temperature x such that $F_\gamma(x) \geq m$, or equivalently, such that $\gamma(\{y \in \mathbb{R}^N \mid b(y_1, \ldots, y_N) \leq x) \geq m$. More formally

$$F_\gamma^{-1}(m) = \inf\{x \in \mathbb{R} : F_\gamma(x) \geq m\}$$

In plain language, $F_\gamma^{-1}(0,99) \leq 20°$ says that 99% of the time, the temperature of the released water remains below 20°. This gives a good way to estimate the potential harm to the environment caused by the plant, but it is difficult to treat mathematically (leading to a degenerate optimal transport problem). We will therefore use the integrated expression

$$R_m(b_\sharp\gamma) = \int_{F_\gamma^{-1}(m)}^{+\infty} x db_\sharp\gamma(x)$$

In practice, γ is very difficult to estimate (because it is a probability measure on a high dimensional space $\mathbb{R}^N$), but it is quite easy to estimate the law of each parameters independently. In other words, we do not know γ, but we know its marginals $\mu_1, \ldots, \mu_N \in \mathcal{P}(\mathbb{R})$. The probability measure μ_i is the law of the parameter $y_i \in \mathbb{R}$. So to compute the risk, we are led to make an assumption

- We could suppose naively that $\gamma = \otimes_{i=1}^N \mu_i$. In this case the parameters are independent, $i.e$, we are in an "optimistic" scenario when rare or "bad" events have little probability of happening at the same time. As we will discuss in the the remark (4.9) this approach is not likely to work only is the case where b is separable in its variables

- We suppose that γ is given by a copula function, $i.e$, $\gamma = \phi(r_1, \cdots, r_N)$ where $\phi : A \subset \mathbb{R}^N \longrightarrow \Gamma(\mu_1, \cdots, \mu_n)$. The problem reads

$$\max_{r_1,\cdots,r_N \in A} R_m(b_\sharp(\phi(r_1, \cdots, r_N)).$$

  But it turns out that this problem is often non-convex(unless, for example, if the embedding is affine). In addition $\phi(A) \subsetneq \Gamma(\mu_1, \cdots, \mu_N)$, which means that we are underestimating the risk.

- Finally, we assume only that γ satisfies the marginal constraints, $i.e$ $\gamma \in \Gamma(\mu_1, \cdots, \mu_N)$. Hence the problem reads

$$\begin{aligned}
\max_{\gamma \in \Gamma(\mu_1,\cdots,\mu_N)} R_m(b_\sharp\gamma) &= \max_\gamma \max_{\substack{0 \leq \sigma \leq b_\sharp\gamma \\ \sigma(\mathbb{R})=m}} \int x d\sigma(x) \\
&= \max_\gamma \max_{\substack{0 \leq \tau \leq \gamma \\ \tau(\mathbb{R}^N)=m}} \int b(y_1, \cdots, y_N) d\tau(y_1, \cdots, y_N) \\
&= \max_{\tau \in \Gamma_{\leq m}(\mu_1,\cdots,\mu_N)} \int b d\tau \quad (2.1)
\end{aligned}$$

where

$$\Gamma_{\leq m}(\mu_1, \cdots, \mu_N) = \{\tau \in \Gamma_{\leq}(\mu_1, \cdots, \mu_N), \tau(\mathbb{R}^N) = m\}$$

If we denote by $c(y_1, \cdots, y_N) = -b(y_1, \cdots, y_N)$, the problem becomes

$$\min_{\tau \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_N)} \int c d\tau$$

which is exactly a multi-marginal optimal partial transport problem as we will see in details in the following chapters.

Remark 2.1 The risk can be written also as $R_m(b_\sharp(\gamma)) = \int_m^1 F_\gamma^{-1}(m) d\mathcal{L}(x)$, with $\mathcal{L}$ is the Lebesgue measure on $\mathbb{R}^d$. Indeed, since $F_\gamma^{-1}{}_\sharp \mathcal{L} = b_\sharp \gamma$, we immediately find:

$$R_m(\gamma) = \int_{F_\gamma^{-1}(m)}^{+\infty} x db_\sharp \gamma(x) = \int_m^1 F_\gamma^{-1}(m) d\mathcal{L}(x)$$

by a simple change of variables.

3 Background on OT and variants

We start with an overview on the "classical" Monge-Kantorovich problem. Imagine that we have a certain amount of soil which should be used to till some holes in the ground. This task should be done while minimizing a certain cost, for example, the traveled distance, or the effort that one does transporting this amount of soil with a wheelbarrow.
The modern formulation of Monge problem as it was introduced by Gaspard Monge in 1776 in his celebrated paper "Mémoire sur la théorie des déblais et des remblais", consists in minimizing the quantity $\int \|x-T(x)\|^2 d\mu(x)$, where μ and ν are probability measures on two compact spaces X and Y, representing the initial and the target distribution of soil respectively, and T is a Borel map from X to Y that pushes μ onto ν, *i.e*, ν coincides with the measure obtained by picking every atom at x an putting it at $T(x)$. More precisely

$$(\mathcal{MP}) : \min \int_X \|x - T(x)\|^2 d\mu(x), \quad \mu(T^{-1}(B)) = \nu(B) \tag{3.1}$$

for any Borelian B. Unfortunately, this problem is difficult to solve due to the nonlinear constraint $T_\sharp\mu = \nu$. Moreover, such maps T, usually called a transport plan, may not exist. One may think about a Dirac measure $\mu = \delta_x$, and another measure ν with no atoms. So if a transport map $T : X \to Y$ exists, then $1 = \mu(T^{-1}(\{T(x)\})) > \nu(\{T(x)\})$, so the constraint $T_\sharp\mu = \nu$ is violated. To overcome theses difficulties, Leonid Kantooivich proposed a "relaxation" of Monge problem in his paper "On the translocation of masses", by allowing mass splitting. More formally, he considered the problem

$$(\mathcal{KP}) : \min \int_{X\times Y} c d\gamma, \quad \gamma \in \Gamma(\mu, \nu) \tag{3.2}$$

where $\Gamma(\mu,\nu)$ is the set of probability measures on $X \times Y$ with marginals μ and ν respectively, *i.e*, $(\pi_X)_\sharp\gamma = \mu$ and $(\pi_Y)_\sharp\gamma = \nu$, where π_X, π_Y are the projections form $X \times Y$ onto X and Y respectively. It turns out that this problem enjoys several interesting properties that make the analysis easier. Namely, the set $\Gamma(\mu,\nu)$ is always non-empty since it contains $\mu \otimes \nu$. Moreover, it's richer than the set of transport maps, in the sense that a transport map T induces a transport plan $\gamma_T = (Id \times T)_\sharp\mu$ (but the converse is false). In addition to this, $\Gamma(\mu,\nu)$ is a convex and compact subset of $\mathcal{P}(X \times Y)$ endowed with narrow topology, which helps to get existence results under weak assumptions on the cost function c. One can prove easily the following

Theorem 3.1 *[22] Let X and Y be compact metric spaces, $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ and $c : X \times Y \to \mathbb{R}(\mathbb{R} \cup \{\infty\})$ a continuous function (respectively, lower semi-continuous and bounded from bellow). Then $(\mathcal{KP})$ admits a solution.*

This kind of results can be extender to Polish spaces. This type of existence results can be found, with different proofs, in [22–24].
We notice that $(\mathcal{KP})$ is a linear optimization problem under convex constraints (actually affine). So it is natural to derive its dual formulation. It consists basically

in a min − max interchange.
Let $\gamma \in \mathcal{M}_+(X \times Y)$ and $\phi \in C_b(X), \psi \in C_b(Y)$. We have

$$\sup_{\phi,\psi} \int_X \phi d\mu + \int_Y \psi d\nu - \int_{X\times Y} \phi \oplus \psi d\gamma = \begin{cases} 0 \quad if \quad \gamma \in \Gamma(\mu,\nu) \\ \infty \quad if \quad not \end{cases} . \tag{3.3}$$

So we can get ride of the constraint $\gamma \in \Gamma(\mu, \nu)$ if we add the previous sup. This leads to consider the following problem

$$\min_{\gamma\in\Gamma(\mu,\nu)} \int c d\gamma + \sup_{\phi,\psi} \int_X \phi d\mu + \int_Y \psi d\nu - \int_{X\times Y} \phi \oplus \psi d\gamma$$

By interchanging the inf and sup we get

$$\sup_{\phi,\psi} \int_X \phi d\mu + \int_Y \psi d\nu + \inf_{\gamma\in\Gamma(\mu,\nu)} \int (c - \phi \oplus \psi) d\gamma$$

Moreover, we can write the inf on γ as a constraint on the potentials ϕ, ψ:

$$\inf_{\gamma\in\Gamma(\mu,\nu)} \int (c - \phi \oplus \psi) d\gamma = \begin{cases} 0 \quad if \quad \phi \oplus \psi \leq c \text{ on } X \times Y. \\ -\infty \quad otherwise \end{cases} . \tag{3.4}$$

We define the dual problem $(\mathcal{KD})$ as follows

$$(\mathcal{KD}) : \ \sup \Big\{ \int_X \phi d\mu + \int_Y \psi d\nu, \ \phi \in C_b(X), \ \psi \in C_b(Y) \text{ and } \phi \oplus \psi \leq c \Big\} \tag{3.5}$$

Proposition 3.2 *Suppose that X and Y are compact and c is continuous. Then there exists a solution (ϕ, ψ) to $(\mathcal{KD})$.*

3.1 Multi-marginal Optimal Transport

In this section we present the multimarginal problem, which consists in studying the same Monge-Kantorovich problem as in the previous section, but with considering more than two marginals. This kind of problems appear in many fields such as physics and economics [8, 11, 13].

3.1.1 Existence results

Let $X_1, \cdots, X_n$ be metric spaces and $c : X_1 \times \cdots \times X_n \mapsto [0, \infty]$ be the cost function which will be assumed to be continuous or lower semi-continuous. Given n probability measures $\mu_i \in \mathcal{P}(X_i), i = 1, \cdots, n$, Kantorovich problem can be formulated as follows:

$$(\mathcal{KP}_n) : \inf_{\gamma\in\Gamma(\mu_1,\cdots,\mu_n)} \Big\{ K(\gamma) = \int_{X_1\times\cdots\times X_n} c d\gamma \Big\} \tag{3.6}$$

where $\Gamma(\mu_1, \cdots, \mu_n)$ is the set of transport plans between the μ_i, *i.e*,

$$\Gamma(\mu_1, \cdots, \mu_n) = \Big\{ \gamma \in \mathcal{P}(X_1 \times \cdots \times X_n), (\pi_i)_{\sharp\gamma} = \mu_i, \ for \ i = 1, \cdots, n \Big\}. \tag{3.7}$$

The first existence result of a minimizer to (3.6) is the following.

Theorem 3.3 *Let $(X_i)_{i=1,\cdots,n}$ be compact metric spaces, $\mu_i \in \mathcal{P}(X_i)$, $i = 1, \cdots, n$ and $c \in C^0(X_1 \times \cdots \times X_n; \mathbb{R})$. Then $(\mathcal{KP}_n)$ admits a solution.*

Proof. Since c is continuous then $\gamma \mapsto K(\gamma)$ is also continuous. It remains to prove that $\Gamma(\mu_1, \cdots, \mu_n)$ is compact to conclude via Weirestrass criterion. Let $\gamma_j \in \Gamma(\mu_1, \cdots, \mu_n)$ be a minimizing sequence, since $\gamma_j(X_1 \times \cdots \times X_n) = 1$, then $(\gamma_j)_{j\in\mathbb{N}}$ is bounded in $C^*(X_1 \times \cdots \times X_n)$. By Banach-Alaoglu theorem, there exist $\gamma_{j_k} \rightharpoonup \gamma \in \mathcal{P}(X_1 \times \cdots \times X_n)$. It remains to check that $\gamma \in \Gamma(\mu_1, \cdots, \mu_n)$. To do so, we fix, for all $i = 1, \cdots, n$, a function $\phi \in C(X_i)$ and we write that:

$$\int_{X_1\times\cdots\times X_n} \phi(x) d\gamma_{j_k}(x) = \int_{X_i} \phi(x_i) d\mu_i(x_i).$$

by passing to the limit we get

$$\int_{X_1\times\cdots\times X_n} \phi(x) d\gamma(x) = \int_{X_i} \phi(x_i) d\mu_i(x_i).$$

as desired. □

The following lemma allows us to obtain the existence of a minimizer under weaker assumptions on c.

Lemma 3.4 *If $c : X_1 \times \cdots \times X_n \mapsto \mathbb{R} \cup \infty$ is lower semi-continuous and bounded from below, then $\gamma \in \mathcal{M}_+(X_1 \times \cdots \times X_n) \mapsto K(\gamma) = \int c d\gamma \in \mathbb{R} \cup \{\infty\}$ is lsc for the weak convergence of measures.*

Proof. Since c is lsc then one can find a sequence of continuous functions c_k converging increasingly towards c. Then $K(\gamma) = \lim_{k\to\infty} \int c_k d\gamma = \sup_k K_k(\gamma)$. Since the c_k are continuous then K_k are also continuous. Consequently, K is lsc as a sup of continuous functions. □

We get easily the following theorem.

Theorem 3.5 *Let $(X_i)_{i=1,\cdots,n}$ be compact metric spaces, $\mu_i \in \mathcal{P}(X_i)$, $i = 1, \cdots, n$ and $c : X_1 \times \cdots \times X_n \mapsto \mathbb{R} \cup \{\infty\}$ be lsc and bounded from below. Then $(\mathcal{KP}_n)$ admits a solution.*

We can prove more a more general result (see the appendix).

3.2 Duality

As in the two marginal setting, we derive the dual problem $(\mathcal{KD}_n)$ for $(\mathcal{KP}_n)$. Let $\gamma \in \mathcal{M}_+(X_1 \times \cdots \times X_n)$ and $\phi_i \in C_b(X_i)$ with $i = 1, \cdots, n$. We have

$$\sup_{(\phi_i)_{i=1}^n} \sum_{i=1}^n \int_{X_i} \phi_i d\mu_i - \int_{X_1\times\cdots\times X_n} \bigoplus_{i=1}^n \phi_i(x_i) d\gamma = \begin{cases} 0 & if \quad \gamma \in \Gamma(\mu_1, \cdots, \mu_n) \\ \infty & if \quad not \end{cases} . \tag{3.8}$$

So we can get ride of the constraint $\gamma \in \Gamma(\mu_1, \cdots, \mu_n)$ if we add the previous sup. This leads to consider the following problem

$$\min_{\gamma\in\Gamma(\mu_1,\cdots,\mu_n)} \int c d\gamma + \sup_{(\phi_i)_{i=1}^n} \sum_{i=1}^n \int_{X_i} \phi_i d\mu_i - \int_{X_1\times\cdots\times X_n} \bigoplus_{i=1}^n \phi_i(x_i) d\gamma$$

By interchanging the inf and sup we get

$$\sup_{(\phi_i)_{i=1}^n} \sum_{i=1}^n \int_{X_i} \phi_i d\mu_i + \inf_{\gamma \in \Gamma(\mu_1,\cdots,\mu_n)} \int (c - \bigoplus_{i=1}^n \phi_i(x_i)) d\gamma$$

Moreover, we can write the $\inf_\gamma$ as a constraint on the potentials $(\phi_1, \cdots, \phi_n)$:

$$\inf_{\gamma \in \Gamma(\mu_1,\cdots,\mu_n)} \int (c - \bigoplus_{i=1}^n \phi_i(x_i)) d\gamma = \begin{cases} 0 \ \ if \quad \bigoplus_{i=1}^n \phi_i \le c \ on \ X_1 \times \cdots \times X_n. \\ -\infty \ \ otherwise \end{cases} . \tag{3.9}$$

We define the dual problem $(\mathcal{KD}_n)$ as follows

$$(\mathcal{KD}_n) : \ \sup \Big\{ \sum_{i=1}^n \int_{X_i} \phi_i d\mu_i, \ \phi_i \in C_b(X_i) \ and \ \bigoplus_{i=1}^n \phi_i \le c \Big\} \tag{3.10}$$

Proposition 3.6 *Suppose that $X_1, \cdots, X_n$ are compact and c is continuous. Then there exists a solution $(\phi_1, \cdots, \phi_n)$ to $(\mathcal{KD}_n)$.*

Proof. Let $(\phi_1^k, \cdots, \phi_n^k)_{k\in\mathbb{N}}$ be a maximizing sequence. Without loss of generality, we can assume that this sequence of n-tuple of functions is c-conjugate with respect to c, *i.e*: for all $i = 1, \cdots, n$

$$\phi_i^k(x_i) = \inf c(x_1, \cdots, x_n) - \sum_{j=1, j\neq i}^n \phi_j^k(x_j), \ \ for \ x_j \in X_j, j \neq i.$$

Since c is continuous on compact sets, and hence uniformly continuous, one has

$$|c(x_1, \cdots, x_n) - c(y_1, \cdots, y_n)| \le \omega_c\Big(d_1(x_1, y_1) + \cdots + d_n(x_n, y_n)\Big)$$

where ω_c is the modulus of continuity of c and d_i is a given metric on X_i for $i = 1, \cdots, n$. So for two elements $x_i, y_i \in X_i$ we have

$$\begin{aligned} &\Big(c(x_1, \cdots, x_i, \cdots, x_n) - \sum_{j=1,j\neq i}^n \phi_j^k(x_j)\Big) - \Big(c(x_1, \cdots, y_i, \cdots, x_n) - \sum_{j=1,j\neq i}^n \phi_j^k(x_j)\Big) \\ &\qquad = c(x_1, \cdots, x_i, \cdots, x_n) - c(x_1, \cdots, y_i, \cdots, x_n) \le \omega_c\Big(d_i(x_i, y_i)\Big). \end{aligned} \tag{3.11}$$

which gives

$$|\phi_i^k(x_i) - \phi_i^k(y_i)| \le \omega_c\Big(d_i(x_i, y_i)\Big).$$

All the ϕ_i^k are bounded since they are continuous on compact sets X_i. Since $\bigoplus_{i=1}^n \phi_i^k \le c$, we also have for $x_i, x_i^* \in X_i, 1 \le i \le n-1$:

$$\Big(\phi_1^k(x_1) - \phi_1^k(x_1^*)\Big) + \cdots + \Big(\phi_{n-1}^k(x_{n-1}) - \phi_{n-1}^k(x_{n-1}^*)\Big) + \phi_n^k(x_n) + \sum_{i\le n-1} \phi_i^k(x_i^*) \le c$$

For all $1 \le i \le n-1$ we set $\tilde{\phi}_i^k = \phi_i^k - \phi_i^k(x_i^*)$. It is clear that theses functions share the same modulus of continuity as the ϕ_i^k, and they satisfy $\tilde{\phi}_i^k(x_i^*) = 0$. Moreover, we have for $x_n \in X_n$ and $i \le n-1$:

$$\tilde{\phi}_i^k(x_i) = \inf c - \sum_{j\neq i} \tilde{\phi}_j^k \in [\min c - \sum_{j\neq i} \omega_c(diam(X_j)), \max c]$$

On the other hand, we have

$$\phi_n^k(x_n) = \inf c - \sum_{j\leq n-1} \phi_j^k \in [\min c - \sum_{j\leq n-1} \omega_c(diam(X_j)), \max c]$$

and this gives uniform bounds on the potentials $(\phi_i^k)_{1\leq i\leq n}$ for all $k \in \mathbb{N}$. So the family $(\phi_1^k, \cdots, \phi_n^k)_{k\in\mathbb{N}}$ is equibounded, and by Ascoli-Arzelà's theorem, up to an extraction, we get that $\phi_i^k \underset{k\to\infty}{\to} \phi_i$ for all $i = 1, \cdots, n$. By uniform convergence we have

$$\int_{X_i} \phi_i^k d\mu_i \underset{k\to\infty}{\to} \int_{X_i} \phi_i d\mu_i$$

and

$$\bigoplus_{i=1}^n \phi_i^k(x_i) \leq c(x_1, \cdots, x_n) \Rightarrow \bigoplus_{i=1}^n \phi_i(x_i) \leq c(x_1, \cdots, x_n)$$

So that $(\phi_1, \cdots, \phi_n)$ is a solution to $(\mathcal{KD}_n)$. □

We give a proof to the duality result $(\mathcal{KP}_n) = (\mathcal{KD}_n)$ based on some properties of Legendre-Fenchel transform, namely, a function f is lsc and convex if and only if $f^{**} = f$. It is adapted from [22].

Lemma 3.7 *Let $p \in C(X_1 \times \cdots \times X_n)$ and define*

$$\Theta(p) = -\max\left\{\sum_{i=1}^n \int_{X_i} \phi_i d\mu_i, \quad \bigoplus_{i=1}^n \phi_i(x_i) \leq c(x_1, \cdots, x_n) - p(x_1, \cdots, x_n)\right\}$$

Then Θ is convex and lsc for uniform convergence on $X_1 \times \cdots \times X_n$, and it's Legendre-Fenchel transform is given by

$$\Theta^*(\gamma) = \begin{cases} K(\gamma) & if \quad \gamma \in \Gamma(\mu_1, \cdots, \mu_n) \\ +\infty & otherwise \end{cases} . \tag{3.12}$$

for $\gamma \in \mathcal{M}(X_1, \cdots, X_n)$

Proof. Let $p, q \in C(X_1 \times \cdots \times X_n)$ and $(\phi_1, \cdots, \phi_n)$, $(\psi_1, \cdots, \psi_n)$ be respectively the corresponding optimal potentials. For $i = 1, \cdots, n$ and $t \in [0, 1]$, we define the following convex combinations $r = (1-t)p + tq$, $\chi_i = (1-t)\phi_i + t\psi_i$. We then have

$$\Theta(r) \leq -\Big(\sum_{i=1}^n \int_{X_i} \chi_i d\mu_i\Big) = -\Big(\sum_{i=1}^n \int_{X_i} ((1-t)\phi_i + t\psi_i) d\mu_i\Big)$$
$$\leq (1-t)\Theta(p) + t\Theta(q). \tag{3.13}$$

which proves convexity. Let $p_j \to p$ in $C(X_1 \times \cdots \times X_n)$. There exists a subsequence p_{j_k} such that $\Theta(p_{j_k}) = \liminf \Theta(p_j)$. Ascoli-Arzelà theorem ensures equicontinuity and equiboundedness of the p_{j_k} and so are the corresponding potentials $(\phi_1^{j_k}, \cdots, \phi_n^{j_k})$. Consequently, we can assume that $\phi_i^{j_k} \underset{k\to+\infty}{\to} \phi_i$ uniformly for all $i = 1, \cdots, n$. Moreover, $\sum_{i=1}^n \phi_i^{j_k}(x_i) \leq c(x_1, \cdots, x_n) - p_{j_k}(x_1, \cdots, x_n)$, and we may pass to the limit to obtain that $\sum_{i=1}^n \phi_i(x_i) \leq c(x_1, \cdots, x_n) - p(x_1, \cdots, x_n)$. Finally

$$\Theta(p) \leq -\Big(\sum_{i=1}^n \int_{X_i} \phi_i(x_i) d\mu_i\Big) = \lim_{k\to+\infty} \Theta(p_{j_k}) = \liminf \Theta(p_j)$$

and this proves lower semi-continuity. Now take $\gamma \in \Gamma(\mu_1, \cdots, \mu_n)$ and write

$$\Theta^*(\gamma) = \sup_p \int_{X_1\times\cdots\times X_n} p d\gamma + \sup_{\phi_1,\cdots,\phi_n} \Big\{ \sum_{i=1}^n \int_{X_i} \phi_i(x_i) d\mu_i, \ \ \bigoplus_{i=1}^n \phi_i \leq c - p \Big\}.$$

We observe that if $\gamma \notin \mathcal{M}_+(X_1, \cdots, X_n)$, there exists $p_0 \leq 0$ such that $\int p_0 d\gamma \geq 0$. So if we take all the potentials ϕ_i to be zero and $p = c + np_0$, we get that $\Theta^*(\gamma) = +\infty$. On the other hand, if we take $p = c - \bigoplus_{i=1}^n \phi_i$, we obtaint that

$$\Theta^*(\gamma) = \sup_{\phi_1,\cdots,\phi_n} \int_{X_1\times\cdots\times X_n} \Big(c(x_1, \cdots, x_n) - \sum_{i=1}^n \phi_i(x_1, \cdots, x_n) \Big) d\gamma + \sum_{i=1}^n \int_{X_i} \phi_i d\mu_i = K(\gamma)$$

□

Theorem 3.8 *If $X_1, \cdots, X_n$ are compact spaces and c is continuous, then the duality formula $\min(\mathcal{KP}_n) = \sup(\mathcal{KD}_n)$ holds.*

Proof. Since Θ is convexe and lsc, we have $\Theta(0) = -\sup(\mathcal{KD}_n) = \Theta^{**}(0)$. But, $\Theta^{**}(0) = \sup_{\gamma\in\mathcal{M}(X_1,\cdots,X_n)} < \gamma, 0 >_{C^*,C} -\Theta^*(\gamma) = -inf\Theta^* = -\min(\mathcal{KP}_n)$. □

3.3 Optimal Partial Transport

We give a brief interpretation of the partial transport problem as in [9]. Consider two measures μ and ν with densities f and g respectively. The density f represent the distribution of bakeries and g the distribution of coffee shops. The classical Monge-Kantorovich problem consists to decide which bakery should supply bread to each coffee shop in order to minimize a certain given cost. In general, this problem is studied in the case where produced bread is totally consumed, *i.e*,

$$\int f(x)dx = \int g(y)dy < \infty.$$

A natural question that can arise is the following: what happens when the supply and demand are not equal, *i.e*, if we transport only a quantity $0 < m < \min\{\|f\|_{L^1}, \|g\|_{L^1}\}$, then what are the bakeries that will continue producing bread, and which coffee shops should they supply so that we keep minimizing the transport cost ?
This problem has been studied extensively in the last few years, and we refer the reader to [9, 17] and the references therein for more details and extensions.

In this section we present the formulation of the partial transport problem, some existence results, and its dual formulation.
Let $X = Y = \mathbb{R}^d$, and assume that c is lower semincontinuous and bounded from below. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and fix $m \in]0, 1[$. A natural extension of Kantorovich problem is the following

$$(\mathcal{KP}_{\leq}) : \inf_{\gamma\in\Gamma_{\leq m}} \Big\{ K(\gamma) = \int_{X\times Y} c d\gamma \Big\} \tag{3.14}$$

where $\Gamma_{\leq m}(\mu, \nu) = \{\gamma \in \mathcal{M}_+(X \times Y),\ (\pi_x)_\sharp\gamma \leq \mu,\ (\pi_y)_\sharp\gamma \leq \nu,\ and\ \gamma(X \times Y) = m\}$. As usual the proof of the existence of minimizers for (3.14) is bases on lower-semicontinuity and compactness arguments. Recall this general version of Prokhorov's theorem, and accept it as true for the moment.

Theorem 3.9 *Let X be a compact space and $\mathcal{F}$ be a subset of the space of sub-probability measures, i.e, $\mathcal{F} \in \{\mu \in \mathcal{M}_+(X), \mu(X) < 1\}$. Then $\mathcal{F}$ is tight if and only if $\mathcal{F}$ is weakly relatively sequentially compact.*

We can state the following theorem.

Theorem 3.10 *Let X and Y be compact spaces. Assume that c is lower-semicontinuous and bounded from below. Then $(\mathcal{KP}_{\leq})$ admits a minimizer.*

Proof. By Prokhorov's theorem Λ is relatively compact. To get narrow compactness, take $(\gamma_n) \in \Lambda$ and assume that $\gamma_n \rightharpoonup \gamma$. We have to prove that $\gamma \in \Lambda$. To do so, take an arbitrary function $\phi \in C_b(X)$ and notice that $\phi \in C_b(X \times Y)$ as well. Hence we have

$$\int \phi d(\pi_x)_\sharp \gamma = \int \phi(x) d\gamma(x,y) = \lim_{n \to +\infty} \int \phi(x) d\gamma_n = \lim_{n \to +\infty} \int \phi d(\pi_x)_\sharp \gamma_n \leq \int \phi d\mu$$

which means that $(\pi_x)_\sharp \gamma \leq \mu$. Similarly we prove that $(\pi_y)_\sharp \gamma \leq \nu$. On the other hand we have that $\int_{X \times Y} d\gamma = \lim_n \int_{X \times Y} d\gamma_n = m$ and hence, $\gamma(X \times Y) = m$. This proves that $\gamma \in \Lambda$. On the other hand, since c is lsc and bounded from below, there exists an increasing sequence $c_n \in C_n(X \times Y; \mathbb{R})$ such that $c(x,y) = \sup_n c_n(x,y)$ so that $\int c d\gamma = \sup_n \int c_n d\gamma$. Since $\lim_{j \to +\infty} \int c_n d\gamma_j = \int c_n d\gamma$ we have

$$\int c d\gamma = \sup_n \int c_n d\gamma \leq \liminf \int c d\gamma_n$$

Consequently, $\gamma \mapsto \int c d\gamma$ is lsc with respect to narrow convergence. □

Now we focus on how to derive the dual problem of $(\mathcal{KP}_{\leq})$. First we introduce a Lagrange multiplier λ conjugate to the constraint $\gamma(X \times Y) = m$. Consider then the problem

$$(\mathcal{KP}^\lambda_{\leq}) : \inf_{\gamma \in \Gamma_{\leq}(\mu,\nu)} \int (c(x,y) - \lambda) d\gamma(x,y)$$

where $\Gamma_{\leq}(\mu,\nu)$ is the set of all measures whose marginals are dominated by μ and ν. We write that

$$\begin{aligned}
(\mathcal{KP}^\lambda_{\leq}) = \inf \Big\{ &< c - \lambda, \gamma >|\ \gamma \geq 0, < \phi, \mu - (\pi_x)_\sharp \gamma > \geq 0, \\
&< \psi, \nu - (\pi_y)_\sharp \gamma > \geq 0,\ \forall \phi \in C_b(X), \forall \psi \in C_b(Y) \Big\} \\
= \inf_{\gamma \geq 0} \sup_{\phi,\psi \leq 0} &< c - \lambda, \gamma > + < \phi, \mu - (\pi_x)_\sharp \gamma > + < \psi, \nu - (\pi_y)_\sharp \gamma > \\
= \inf_{\gamma \geq 0} \sup_{\phi,\psi \leq 0} &< (c - \lambda) - (\phi \otimes 1 + 1 \otimes \psi), \gamma > + < \phi, \mu > + < \psi, \nu > \quad (3.15)
\end{aligned}$$

By interchanging the inf and sup and noticing that

$$\inf_{\gamma \geq 0} < (c - \lambda) - (\phi \otimes 1 + 1 \otimes \psi), \gamma >= \begin{cases} 0, & if\ \phi \oplus \psi \leq c - \lambda \\ -\infty, & if\ not. \end{cases}$$

hence the dual problem can be written as

$$(\mathcal{KD}_{\leq}^{\lambda}): \sup_{\substack{\phi,\psi\leq 0\\ \phi\oplus\psi\leq c-\lambda}} \int_X \phi d\mu + \int_Y \psi d\nu \tag{3.16}$$

We note that, if the optimizer of $(\mathcal{KP}_{\leq}^{\lambda})$ is unique, we denote it γ_λ and its total mass $m(\lambda) = \mathbb{R}^d \times \mathbb{R}^d$. Then $m(\lambda) = -\partial(\mathcal{KP}_{\leq}^{\lambda})/\partial\lambda$ increases from 0 to m as λ increases [9]. Thus, for a properly chosen $\lambda \geq 0$, we are given a mass m. Finally, we show that only one measure in $\Gamma_{\leq m}$ with mass m is optimal. We note also that (3.16) can be obtained by Kantorovich duality for the cost $c - \lambda$.

Remark 3.11 In the particular case of a quadratic cost $c(x, y) = |x - y|^2/2$, the problem

$$\inf_{\tilde\phi+\tilde\psi\geq <x,y>} \left\{ \int_X \tilde\phi d\mu + \int_Y \tilde\psi d\nu,\ \tilde\phi(x) \geq (|x|^2 - \lambda)/2,\ \tilde\psi(y) \geq (|y|^2 - \lambda)/2 \right\} \tag{3.17}$$

is equivalent to $(\mathcal{KP}_{\leq}^{\lambda})$ in the sense that $(\tilde\phi = (|x|^2 - \lambda)/2 - \phi, \tilde\psi = (|y|^2 - \lambda)/2 - \psi)$ minimize (3.17) when (ϕ, ψ) maximize $(\mathcal{KP}_{\leq}^{\lambda})$.

Indeed, if (ϕ, ψ) is optimal, we easily check that

$$(|x|^2 - \lambda)/2 - \phi + (|y|^2 - \lambda)/2 - \psi = (|x|^2 + |y|^2)/2 - \lambda - (\phi + \psi) \geq \lambda - c$$

this shows that $\tilde\phi + \tilde\psi \geq < x, y >$. Since $\phi \leq 0$ and $\psi \leq 0$ we see that $\tilde\phi(x) \geq (|x|^2 - \lambda)/2$ and $\tilde\psi(y) \geq (|y|^2 - \lambda)/2$, *i.e*, $(\tilde\phi, \psi)$ is admissible for (3.17). The converse is obviously true. Moreover, the difference between $(\mathcal{KP}_{\leq}^{\lambda})$ and (3.17) is determined by the second order moments and total masses of μ and ν. More precisely, we have

$$\begin{aligned} \sup\left\{ \int_X \tilde\phi d\mu + \int_Y \tilde\psi d\nu \right\} - \inf\left\{ \int_X \tilde\phi d\mu + \int_Y \tilde\psi d\nu \right\} \\ = \sup\left\{ \int_X (\tilde\phi + \phi) d\mu + \int_Y (\tilde\psi + \psi) d\nu \right\} \\ = \int_X |x|^2 d\mu + \int_Y |y|^2 d\nu - (\mu(X) + \nu(Y)) = W_2^2(\mu, \delta_0) + W_2^2(\nu, \delta_0) - (\mu(X) + \nu(Y)). \end{aligned} \tag{3.18}$$

Using Brenier's result [6], we can always assume that $\tilde\phi, \tilde\psi$ are convex functions. Hence, the optimal solutions of $(\mathcal{KP}_{\leq}^{\lambda})$ and (3.17) are linked by

$$\gamma\Big(\{(x, \nabla\tilde\psi), x \in U\}\Big) = \gamma(\mathbb{R}^d \times \mathbb{R}^d) = \int_U f(x)dx.$$

where $U = \{\tilde\psi(x) \geq (|x|^2 - \lambda)/2\}$.

4 Multi-marginal Optimal Partial Transport

4.1 Formulation

The multi-marginal partial transport problem was studied by Brendan Pass and Jun Kitagawa in [18] as an extension of the partial barycentre problem. In this section we give the formulation of the multimarginal partial problem, we recall for completeness, the partial barycentre problem and we state the main results of [18].
As before, consider N compact spaces, $X_1, \cdots, X_N$ and N measures, $\mu_i \in \mathcal{M}(X_i)$ for $i = 1, \cdots, N$ and fix $0 \leq m \leq \min_{1 \leq i \leq N} \mu_i(X_i)$. We denote by $\Gamma_{\leq m}(\mu_1, \cdots, \mu_N)$ the set

$$\Gamma_{\leq m}(\mu_1, \cdots, \mu_N) = \{\gamma \in \mathcal{M}_+(X_1 \times \cdots \times X_N),\ (\pi_{x_i})_\sharp \gamma \leq \mu_i,\ \gamma(X_1 \times \cdots \times X_N) = m\}$$

We call a solution of the multi-marginal partial problem a measure $\tilde{\gamma} \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_N)$ achieving the minimum value in

$$(MMP_m) : \min_{\gamma \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_N)} \left\{ K(\gamma) = \int_{X_1 \times \cdots \times X_N} c d\gamma \right\} \tag{4.1}$$

This is problem is equivalent to the partial barycentre problem in the case $c(x_1, \cdots, x_N) = \sum_{ij=1}^N |x_i - x_j|^2$. Before stating the result making the relation between the two problems, we recall "standard" barycentre problem. Given N $\mu_1, \cdots, \mu_N$ measures on $\mathbb{R}^d$, with equal mass m, the goal is to minimize the quantity

$$(BC) : \min_{\nu \in \mathcal{M}_m^2} \sum_{i=1}^N \min_{\gamma \in \Gamma(\mu_i, \nu)} \int |x - y|^2 d\gamma \tag{4.2}$$

where $\mathcal{M}_m^2$ is the set of measures with total mass m and of finite second order moment, $i.e$,

$$\mathcal{M}_m^2 = \{\nu \in \mathcal{M}_+(\mathbb{R}^d), \nu(\mathbb{R}^d) = m,\ \int |x|^2 d\nu < \infty\}$$

This problem was studied by Agueh and Carlier in [1] were they show the equivalence with the standard multimarginal problem. Jun Kitagawa and Brendan Pass proposed a similar problem, which consists in finding a minimizer of

$$(PBC_m) : \sum_{i=1}^N \min_{\gamma \in \Gamma_{\leq m}(\mu_i, \nu)} \int |x - y|^2 d\gamma. \tag{4.3}$$

and they show that (MMP_m) and (PBC_m) are equivalent in the following sense:

Proposition 4.1 • *Let μ_i for $i = 1, \cdots, N$ be absolutely continuous measures and $0 \leq m \leq \min_{i=1,\cdots,N} \mu_i(\mathbb{R}^d)$. Define the map F by*

$$F(x_1, \cdots, x_N) = \frac{1}{N} \sum_{i=1}^N x_i$$

Then if γ is an optimal plan in (MMP_m), $F_\sharp \gamma$ is optimal in (PBC_m).

- *If μ is a minimizer in (PBC_m) then the measure $(T_1^\nu, \cdots, T_N^\nu)_\sharp \nu$ is optimal in (MMP_m), where T_i^ν is the optimal mapping such that $\left(T_i^\nu \times Id\right)_\sharp \nu$ solves $\min_{\gamma \in \Gamma_{\leq m}(\mu_i, \nu)} \int |x-y|^2 d\gamma$*

Furthermore, uniqueness in of the minimizer in one problem implies uniqueness in the other one.

Moreover, the main result of [18] consists in establishing uniqueness of the minimizer in (PBC_m) under assumptions on the measures and the transported mass. More precisely,

Theorem 4.2 *Assume that the absolutely continuous measures μ_i have finite mass with densities g_i. We define the measure $\tilde{\mu} = \mu_1 \wedge \cdots \wedge \mu_N$, which is the absolutely continuous measure with density $\tilde{g} = \min_{1 \leq i \leq N} g_i$. If the mass m is such that*

$$\tilde{\mu}(\mathbb{R}^d) \leq m \leq \min_{1 \leq i \leq N} \mu_i(\mathbb{R}^d)$$

then (PBC_m) admits a unique minimizer. Consequently, the multimarginal partial problem (MMP_m) admits a unique minimizer under these assumptions.

4.2 Modulus of continuity of the transport cost

In this section we give an estimate of the modulus of continuity of the transport cost. This is motivated by our modelling problem presented in the first chapter. More precisely, we have seen that the marginals $\mu_1, \cdots, \mu_N$ representing the parameters of work of a nuclear plant are only known up to an error, *i.e*, they represent only estimations of the real parameters. We would therefore like the risk to be continuously dependent on this estimate. We add to this that such continuous dependency will play a major role in the discretization of the problem as we will see later.
Our goal is to prove the following

Theorem 4.3 *Let X, Y be compact metric spaces endowed with the Euclidean metric, and $\mu, \tilde{\mu} \in \mathcal{P}(X)$, $\nu, \tilde{\nu} \in \mathcal{P}(Y)$. We denote by γ_{Opt} the optimal transport plan between μ, ν and $\bar{\gamma}_{Opt}$ an the optimal plan between $\tilde{\mu}$ and $\tilde{\nu}$. Assume that the cost function c is of class $\mathcal{C}^{1,1}$, and that $W_2(\mu, \tilde{\mu}), W_2(\nu, \tilde{\nu}) < \varepsilon$, for some $\varepsilon > 0$. Then*

$$\left| \int c d\gamma_{Opt} - \int c d\bar{\gamma}_{Opt} \right| \lesssim \mathcal{O}(\varepsilon).$$

and the constants in $\mathcal{O}(\varepsilon)$ depend only on c.

Proof. Let $\Lambda \in \Gamma(\gamma_{Opt}, \bar{\gamma}_{Opt})$ such that

$$\int_{(X \times Y)^2} |\alpha - \beta|^2 d\Lambda = \min_{\Upsilon \in \Gamma(\gamma_{Opt}, \bar{\gamma}_{Opt})} \int |\alpha - \beta|^2 d\Upsilon$$

where $\alpha = (x, y) \in X \times Y$ and $\beta = (\tilde{x}, \tilde{y}) \in X \times Y$. We have

$$\begin{aligned} \left| \int c d\gamma_{Opt} - \int c d\bar{\gamma}_{Opt} \right| &= \left| \int c(\beta) d\Lambda(\alpha, \beta) - \int c(\alpha) d\Lambda(\alpha, \beta) \right| \\ = \left| \int (c(\beta) - c(\alpha)) d\Lambda(\alpha, \beta) \right| &= \left| \int \int_0^1 \nabla c(\alpha + \lambda(\beta - \alpha)) \cdot (\beta - \alpha) d\lambda d\Lambda(\alpha, \beta) \right| \\ &\leq \left| \int \nabla c(\alpha) \cdot (\beta - \alpha) d\Lambda \right| + \mathcal{R} \quad (4.4) \end{aligned}$$

Where $\mathcal{R}$ is a remainder term that can be bounded by

$$|\mathcal{R}| \leq \frac{Lip(\nabla c)}{2} \int |\alpha - \beta|^2 d\Lambda = \frac{Lip(\nabla c)}{2} W_2^2(\gamma_{Opt}, \bar{\gamma}_{Opt})$$

On the other hand, applying Cauchy-Schwartz inequality nor we obtain

$$\left| \int \nabla c(\alpha) \cdot (\beta - \alpha) d\Lambda \right| \leq \|\nabla c\|_{L^2(\Lambda)} W_2(\gamma_{Opt}, \bar{\gamma}_{Opt}).$$

By assumption $W_2(\mu, \tilde{\mu}), W_2(\nu, \tilde{\nu}) < \varepsilon$, we then have

$$W_2^2(\gamma_{Opt}, \bar{\gamma}_{Opt}) \leq W_2^2(\mu, \tilde{\mu}) + W_2^2(\nu, \tilde{\nu}) < 2\varepsilon.$$

By an approximation argument, we may assume that mu,nu are absolutely continuous. Then, by Brenier's theorem, there exists T, $\tilde{T}$ transport maps such that $T_\sharp \mu = \tilde{\mu}$, and $\tilde{T}_\sharp \nu = \tilde{\nu}$. We then have

$$W_2^2(\mu, \tilde{\mu}) = \int |x - T(x)|^2 d\mu(x), \quad W_2^2(\nu, \tilde{\nu}) = \int |x - \tilde{T}(x)|^2 d\nu(x)$$

and $\tilde{\gamma} = (T, \tilde{T})_\sharp \gamma_{Opt} \in \Gamma(\tilde{\mu}, \tilde{\nu})$. We need to check that it's marginals are $\tilde{\mu}$ and $\tilde{\nu}$. We have

$$\int \phi(x) d\tilde{\gamma}(x, y) = \int \phi(T(x)) d\gamma_{Opt}(x, y) = \int \phi \circ T d\gamma$$
$$= \int \phi \circ T d\mu = \int \phi d\tilde{\mu}. \quad (4.5)$$

and similarly, $\int \phi(y) d\tilde{\gamma}(x, y) = \int \phi d\tilde{\nu}$. We have

$$\begin{aligned} W_2^2(\gamma_{Opt}, \tilde{\gamma}) &\leq \int |(x, y) - (T(x), \tilde{T}(y))|^2 d\gamma_{Opt}(x, y) \\ &= \int |x - T(x)|^2 d\gamma_{Opt}(x, y) + \int |y - \tilde{T}(y)|^2 d\gamma_{Opt}(x, y) \\ &= \int |x - T(x)|^2 d\mu(x) + \int |y - \tilde{T}(y)|^2 d\nu(y) \\ &= W_2^2(\mu, \tilde{\mu}) + W_2^2(\nu, \tilde{\nu}) < 2\varepsilon^2. \quad (4.6) \end{aligned}$$

as desired. Combining this with the previous inequalities, we obtain

$$\left| \int c d\gamma_{Opt} - \int c d\bar{\gamma} \right| \leq f(\varepsilon) \triangleq 2\varepsilon \|\nabla c\|_{L^2(\Lambda)} + \frac{\sqrt{2}}{2} Lip(\nabla c) \varepsilon.$$

□

We immediately have the following

Corollary 4.4 *Given probability measures $(\mu_i^\varepsilon)_{1 \leq i \leq n}$ and $(\mu_i)_{1 \leq i \leq n}$ on some compact space X such that $\mu_i^\varepsilon \underset{\varepsilon \to 0}{\rightharpoonup} \mu_i$. If the optimal transport plan γ between the $(\mu_i)_{1 \leq i \leq n}$ is unique and γ_*^ε is the optimal transport plan between the $(\mu_i^\varepsilon)_{1 \leq i \leq n}$, then*

$$\gamma_*^\varepsilon \rightharpoonup \gamma \text{ as } \varepsilon \to 0.$$

4.3 Duality

Now, we turn our attention to the corresponding dual problem. As usual, the constraint $\gamma \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_n)$ will become, in some sense, the functional to maximize and the functional $\int c d\gamma$ will become the constraint in the dual problem. We start by noticing that

$$\inf_{\gamma \in \Gamma_{\leq m}(\mu_1,\cdots,\mu_n)} \int c d\gamma = \inf_{\gamma \in \mathcal{M}_+(X_1 \times \cdots X_n)} \int c d\gamma + \Upsilon(\gamma)$$

where $\Upsilon(\gamma) = 0$ if $\gamma \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_n)$ and $+\infty$ if not. We claim that Υ can be written as

$$\Upsilon(\gamma) = \sup\left\{ \sum_{i=1}^{n} \int_{X_i} \phi_i d\mu_i - \sum_{i=1} < \phi \otimes 1 \mid \gamma > + \kappa\Big(m - \int d\gamma\Big),\ \phi \in C_b(X_i), \phi_i \leq 0, \kappa \geq 0 \right\}$$

In fact,

- If $\gamma \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_n)$, we see that $\kappa(m - \int d\gamma) \leq 0$, since $\kappa \geq 0$ and $\int d\gamma \geq 0$. Hence $\sup_{\kappa \geq 0} \kappa(m - \int d\gamma) = 0$. On the other hand, since $(\pi_x)_\sharp \gamma \leq \mu_i$, $\sum_{i=1}^{n} < \phi_i \mid \mu_i - (\pi_x)_\sharp > \leq 0$ provided $\phi_i \leq 0$ for all $i = 1, \cdots, n$. Hence $\Upsilon(\gamma) = 0$ if $\gamma \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_n)$.
- If $\gamma \notin \Gamma_{\leq m}(\mu_1, \cdots, \mu_n)$, we can take the potentials ϕ_i to be zero and note that, in this case, $\int d\gamma \leq m$ implies that $\sup_{\kappa \geq 0} \kappa(m - \int d\gamma) = +\infty$. Thus, we can write

$$\begin{aligned}
(\mathcal{KP}^n_{\leq}) : & \inf_{\gamma \in \Gamma_{\leq m}(\mu_1,\cdots,\mu_n)} \int_{X_1 \times \cdots \times X_n} c d\gamma \\
&= \inf_{\gamma \geq 0} \sup_{\kappa \geq 0, \phi_i \leq 0} \left\{ < c \mid \gamma > + \sum_{i=1}^{n} < \phi_i \mid (\mu_i - (\pi_{x_i})_\sharp \gamma) > + \kappa\Big(m - \int d\gamma\Big) \right\} \\
&= \inf_{\gamma \geq 0} \sup_{\kappa, \phi_i \leq 0} \left\{ < c - \kappa - \sum_{i=1}^{n} \phi_i \oplus 1 \mid \gamma > + \sum_{i=1}^{n} \int_{X_i} \phi_i d\mu_i + \kappa m \right\} \quad (4.7)
\end{aligned}$$

we interchange the inf and sup to get

$$\begin{aligned}
& \inf_{\gamma \in \Gamma_{\leq m}(\mu_1,\cdots,\mu_n)} \int_{X_1 \times \cdots \times X_n} c d\gamma \\
&= \sup_{\kappa \geq 0, \phi_i \leq 0} \inf_{\gamma \geq 0} \left\{ < c - \kappa - \sum_{i=1}^{n} \phi_i \oplus 1 \mid \gamma > + \sum_{i=1}^{n} \int_{X_i} \phi_i d\mu_i + \kappa m \right\} \\
&= \sup_{\kappa \geq 0, \phi_i \leq 0} \left\{ \sum_{i=1}^{n} \int_{X_i} \phi_i d\mu_i + \kappa m + \inf_{\gamma \geq 0} < c - \kappa - \sum_{i=1}^{n} \phi_i \oplus 1 \mid \gamma > \right\} \quad (4.8)
\end{aligned}$$

We observe that

$$\inf_{\gamma \geq 0} < c - \kappa - \sum_{i=1}^{n} \phi_i \oplus 1 \mid \gamma > = \begin{cases} 0, \ if \ \ \oplus_{i=1}^{n} \phi_i \leq c - \kappa \\ -\infty, \ ifnot. \end{cases}$$

hence the dual problem can be written as

$$(\mathcal{DP}^n_{\leq}) : \ \sup\left\{ \sum_{i=1}^{n} \int_{X_i} \phi_i d\mu_i + \kappa m,\ \kappa \geq 0, \phi_i \leq 0,\ \oplus_{i=1}^{n} \phi_i \leq c - \kappa \right\} \quad (4.9)$$

Proposition 4.5 *Suppose that $X_1, \cdots, X_n$ are compact and c is continuous. Then there exists a solution $(\kappa, \phi_1, \cdots, \phi_n)$ to $(\mathcal{DP}^n_{\leq})$.*

Proof. Let $(\phi_1^k, \cdots, \phi_n^k)_{k\in\mathbb{N}}$, $(\kappa_k)_{k\in\mathbb{N}}$ be a maximizing sequence. As in the proof of Proposition 2.2.1, the sequence of potentials ϕ_i can be always improved by taking c-transforms so that all the ϕ_i^k share the same modulus of continuity as c and are equibounded. So the family $(\phi_1^k, \cdots, \phi_n^k)_{k\in\mathbb{N}}$ is equicontinuous and equibounded, and by Ascoli-Arzelà's theorem, up to an extraction, we get that $\phi_i^k \underset{k\to\infty}{\to} \phi_i$ for all $i = 1, \cdots, n$. On the other hand, we have that $\kappa_k \leq c - \oplus_{i=1}^n \phi_i^k \leq C$, where C is a positive constant. This show that, up to an extraction if necessary, $\kappa_k \underset{k\to\infty}{\to} \kappa \geq 0$. By uniform convergence we have

$$\int_{X_i} \phi_i^k d\mu_i + m\kappa_k \underset{k\to\infty}{\to} \int_{X_i} \phi_i d\mu_i + m\kappa$$

and

$$\bigoplus_{i=1}^n \phi_i^k(x_i) \leq c(x_1, \cdots, x_n) - \kappa_k \Rightarrow \bigoplus_{i=1}^n \phi_i(x_i) \leq c(x_1, \cdots, x_n) - \kappa$$

So that $(\kappa, \phi_1, \cdots, \phi_n)$ is a solution to $(\mathcal{KD}^n_{\leq})$. □

To prove the duality formula $(\mathcal{KP}^n_{\leq}) = (\mathcal{DP}^n_{\leq})$ we introduce as before the functional Θ defined by

Lemma 4.6 *Let $p \in C(X_1 \times \cdots \times X_n)$ and define*

$$\Theta(p) = -\max\Big\{\sum_{i=1}^n \int_{X_i} \phi_i d\mu_i + \kappa m, \quad \oplus_{i=1}^n \phi_i \leq (c-\kappa) - p, \ \kappa \geq 0, \phi_i \leq 0\Big\}$$

Then Θ is convex and lsc for uniform convergence on $X_1 \times \cdots \times X_n$, and it's Legendre-Fenchel transform is given by

$$\Theta^*(\gamma) = \begin{cases} K(\gamma) & if \quad \gamma \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_n) \\ +\infty & otherwise \end{cases}. \tag{4.10}$$

for $\gamma \in \mathcal{M}(X_1, \cdots, X_n)$

Proof. Let $p, q \in C(X_1 \times \cdots \times X_n)$ and $(\kappa, \phi_1, \cdots, \phi_n)$, $(\iota, \psi_1, \cdots, \psi_n)$ be respectively the corresponding optimal solutions. For $i = 1, \cdots, n$ and $t \in [0, 1]$, we define the following convex combinations $r = (1-t)p + tq$, $\chi_i = (1-t)\phi_i + t\psi_i$, $\theta = (1-t)\kappa + t\iota$. Theses combinations are admissible since

$$\begin{aligned} \oplus_{i=1}^n \chi_i = \oplus_i^n (1-t)\phi_i + t\psi_i = (1-t) \oplus_{i=1}^n \phi_i + t \oplus_{i=1}^n \chi_i \\ = (1-t)(c - \kappa - p) + t(c - \iota - q) = (c - \theta) - r. \end{aligned} \tag{4.11}$$

Moreover,

$$\begin{aligned} \Theta(r) \leq -\Big(\sum_{i=1}^n \int_{X_i} \chi_i d\mu_i + \theta m\Big) = -\Big(\sum_{i=1}^n \int_{X_i} ((1-t)\phi_i + t\psi_i) d\mu_i + (1-t)\kappa + t\iota\Big) \\ \leq (1-t)\Theta(p) + t\Theta(q). \end{aligned} \tag{4.12}$$

which proves convexity. Let's prove lsc. Let $p_j \to p$ in $C(X_1 \times \cdots \times X_n)$. There exists a subsequence p_{j_k} such that $\Theta(p_{j_k}) = \liminf \Theta(p_j)$. Ascoli-Arzelà theorem ensures equicontinuity and equiboundedness of the p_{j_k} and so are the corresponding potentials $(\phi_1^{j_k}, \cdots, \phi_n^{j_k})$, and particularly, (κ_{j_k}) is bounded. Consequently, we can assume that $\phi_i^{j_k} \underset{k\to+\infty}{\to} \phi_i$ uniformly for all $i = 1, \cdots, n$, and $\kappa_{j_k} \underset{k\to+\infty}{\to} \kappa$. Moreover, $\sum_{i=1}^n \phi_i^{j_k}(x_i) \leq (c(x_1, \cdots, x_n) - \kappa_{j_k}) - p_{j_k}(x_1, \cdots, x_n)$, and we may pass to the limit to obtain that $\sum_{i=1}^n \phi_i(x_i) \leq (c(x_1, \cdots, x_n) - \kappa) - p(x_1, \cdots, x_n)$. Finally

$$\Theta(p) \leq -\Big(\sum_{i=1}^n \int_{X_i} \phi_i(x_i) d\mu_i + m\kappa\Big) = \lim_{k\to+\infty} \Theta(p_{j_k}) = \liminf \Theta(p_j)$$

and this proves lower semi-continuity. Now take $\gamma \in \Gamma(\mu_1, \cdots, \mu_n)$ and write

$$\Theta^*(\gamma) = \sup_p \int_{X_1\times\cdots\times X_n} p d\gamma + \sup_{\phi_i \leq 0, \kappa \geq 0} \Big\{ \sum_{i=1}^n \int_{X_i} \phi_i(x_i) d\mu_i + \kappa m,\ \bigoplus_{i=1}^n \phi_i \leq (c-\kappa) - p \Big\}.$$

We observe that if $\gamma \notin \mathcal{M}_+(X_1, \cdots, X_n)$, there exists $p_0 \leq 0$ such that $\int p_0 d\gamma > 0$. So if we take all the potentials ϕ_i to be zero and $p = c + Np_0$, for $N \to \infty$, we get that $\Theta^*(\gamma) = +\infty$. On the other hand, if we take $p = c - \kappa - \bigoplus_{i=1}^n \phi_i$, we obtaint that

$$\Theta^*(\gamma) = \sup_{\phi_1, \cdots, \phi_n} \int_{X_1\times\cdots\times X_n} \Big(c(x_1, \cdots, x_n) - \kappa - \sum_{i=1}^n \phi_i(x_1, \cdots, x_n)\Big) d\gamma + \sum_{i=1}^n \int_{X_i} \phi_i d\mu_i + \kappa m = K(\gamma)$$

$\square$

Theorem 4.7 *Suppose that $X_1, \cdots, X_n$ are compact spaces and c is continuous. The the duality formula* $\min(\mathcal{KP}^n_\leq) = \max(\mathcal{KD}^n_\leq)$ *holds.*

Proof. Since Θ is convex and lsc, we have $\Theta(0) = -\sup(\mathcal{KD}^n_\leq) = \Theta^{**}(0)$. But, $\Theta^{**}(0) = \sup_{\gamma\in\mathcal{M}} < \gamma \,|> -\Theta^*(0) = -\inf_\gamma \Theta^* = -\min(\mathcal{KP}^n_\leq)$. Hence, $\min(\mathcal{KP}^n_\leq) = \max(\mathcal{DP}^n_\leq)$. $\square$

Remark 4.8 One natural question that can arise is the following: Is the application $(\mu_1, \cdots, \mu_n) \mapsto C_m(\mu_1, \cdots, \mu_n) := \min_{\gamma\in\Gamma_{\leq m}(\mu_1,\cdots,\mu_n)} \int c d\gamma$ continuous ? We give a negative answer when the cost function can take infinite values, by adapting a counter-example of [7]. Assume that $\mu_i = \mu$ for all $i = 1, \cdots, n$. So the problem rewrites

$$(\mathcal{KP}^n_\leq):\ C_m(\mu) = \min_{\gamma\in\Gamma_\leq(\mu)} \int c d\gamma$$

where $\Gamma_{\leq m}(\mu) = \{\gamma \in \mathcal{P}(X^n),\ (\pi_{x_i})_\sharp\gamma = \mu,\ \forall i = 1, \cdots, n,\ \gamma(X^n) \geq m\}$. Since the cost c is by assumption lsc, then $\mu \mapsto C_m(\mu)$ is lsc. Indeed, if $\mu_n \rightharpoonup \mu$, we have for all $\gamma \in \Gamma_{\leq m}(\mu_n)$

$$(\pi_{x_i})_\sharp\gamma \leq \mu_n \underset{n\to\infty}{\Rightarrow} (\pi_{x_i})_\sharp\gamma \leq \mu$$

so obviously γ is also admissible for μ. And we have

$$\liminf_{n\to\infty} C_m(\mu_n) = \liminf_{n\to\infty} \min_{\gamma\in\Gamma_{\leq m}(\mu_n)} \int c d\gamma \geq \min_{\gamma\in\Gamma_{\leq m}(\mu)} \int c d\gamma \geq C_m(\mu)$$

We shall prove that this application cannot be upper semicontinuous. We consider the two marginal setting, *i.e*, $n = 2$, and consider $c(x, y) = 1/|x - y|$. Take $\mu = 1/2\delta_x + 1/2\delta_y$ for $x \neq y$, and $\mu_k = (1/2 + 1/k)\delta_x + (1/2 - 1/k)\delta_y$. For $\phi \in C_b(X \times X)$ we have

$$< \mu_k \mid \phi >= (1/2+1/k)\phi(x)+(1/2-1/k)\phi(y) \underset{k\to\infty}{\to} 1/2\phi(x)+1/2\phi(y) =< \mu \mid \phi > \tag{4.13}$$

which means that $\mu_k \rightharpoonup \mu$. But $C_m(\mu) = 1/|x - y|$, and $C_m(\mu_k) = +\infty$ for $k \geq 1$. This shows that $\limsup_{\to\infty} C_m(\mu_k) \leq C_m(\mu)$ is violated. Hence the application cannot be continuous. In [7] the authors studies some conditions implying continuity of $C(\mu)$ with respect to narrow convergence. More particularly they show that $C(\mu)$ is Lipschitz-continuous on every bounded set of $L^p(\mathbb{R}^d), p > 1$ for Coulomb-type costs.

Remark 4.9 Take $b = -c$ and assume that b is non-negative and decreasing in the x_i variables, *i.e*, $x_i \mapsto b(x_1, \cdots, x_i, \cdots, x_n)$ is decreasing for all $i = 1, \cdots, n$. Suppose that the measures μ_i are supported in $\mathbb{R}^+$. We write

$$\min_{\gamma\in\Gamma_{\leq m}(\mu_1,\cdots,\mu_n)} \int cd\gamma = -\max_{\gamma\in\Gamma_{\leq m}(\mu_1,\cdots,\mu_n)} \int bd\gamma.$$

We are looking for conditions ensuring triviality of the MMP solution. In other words, when does the equality

$$(1) := \max_{\gamma\in\Gamma_{\leq m}(\mu_1,\cdots,\mu_n)} \int cd\gamma = (2) := \max_{\substack{\nu_i(\mathbb{R})\geq m\\ \nu_i\leq\mu_i}} \int bd\nu_1 \otimes \cdots \otimes \nu_n$$

holds ? The inequality $(1) \geq (2)$ always holds since the constraints $\nu_i \leq \mu_i$ and $\nu_i(\mathbb{R}) \geq m$, imply that $d\nu_i \otimes \cdots \otimes d\nu_n \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_n)$. We assume that the cost c is C^1. Let γ_{Opt} be the solution of MMP, we can write

$$\begin{aligned}\int_{\mathbb{R}^n} bd\gamma_{Opt} &= \int_{\mathbb{R}^n} \Big(b(0,\cdots,0) + \sum_{i=1}^n x_i \int_0^1 \partial_{x_i} b(sx_1,\cdots,sx_n)ds\Big)d\gamma_{Opt}\\ &= b(0,\cdots,0)\gamma_{Opt}(\mathbb{R}^n) + \sum_{i=1}^n \int_{\mathbb{R}} (x_i \int_0^1 \partial_{x_i} b(sx_1,\cdots,sx_n)ds)d\gamma_{Opt}\end{aligned} \tag{4.14}$$

We have

$$\begin{aligned}\int b\otimes_{i=1}^n d\nu_i - \int bd\gamma_{Opt} &= b(0,\cdots,0)\Pi_{i=1}^n \nu_i(\mathbb{R}) + \sum_{i=1}^n < x_i \int_0^1 \partial_{x_i} b(sx_1,\cdots,sx_n)ds|\nu_i >\\ &- b(0,\cdots,0)\gamma_{Opt}(\mathbb{R}^n) - \sum_{i=1}^n < x_i \int_0^1 \partial_{x_i} b(sx_1,\cdots,sx_n)ds|\mu_i >\\ &= b(0,\cdots,0)(\Pi_{i=1}^n \nu_i(\mathbb{R}) - \gamma_{Opt}(\mathbb{R}^n)) + \sum_{i=1}^n < x_i \int_0^1 \partial_{x_i} b(sx_1,\cdots,sx_n)ds|\nu_i - \mu_i >\end{aligned} \tag{4.15}$$

since by assumption, $\nu_i \leq \mu_i$, $x_i \geq 0$ and $\partial_{x_i} b \leq 0$ we deduce that

$$\int b \otimes_{i=1}^{n} d\nu_i \geq \int b d\gamma_{Opt} = \max_{\gamma \in \Gamma_{\leq m}(\mu_1, \cdots, \mu_n)} \int b d\gamma$$

and consequently (2) $\geq$ (1). We note that the computations namely in the expansions are justified as follows. Let γ_{Opt} is the optimal plan between the μ_i, and $\phi_i, i = 1, \cdots, n$ are the Kantorovich potentials, if we assume that $\mu_i \ll \mathcal{L}$ (this implies that $\nu \ll \mathcal{L}$), we have the optimality condition

$$\phi_1(x_1) + \cdots + \phi_n(x_n) + \kappa = -b(x_1, \cdots, x_n) \ on \ spt(\gamma_{Opt}).$$

So at least on the support of γ_{Opt}, b is separated in the variables x_i.

5 Entropic Regularization

5.1 General setting

In this section, all probability measures are absolutely continuous with the Lebesgue measure, and we will often conflate the probability measure with the density. The (negative) entropy of a probability measure is given by

$$H(\mu) = \begin{cases} \int_{\mathbb{R}^d} \Big(\log(\mu(x)) - 1 \Big) \mu(x) dx \ if \ \mu \ll \mathcal{L}. \\ +\infty \ if \ not. \end{cases}$$

Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ absolutely continuous, the regularized Monge-Kantorovich problem is given by

$$(MK_\varepsilon) : \inf_{\gamma \in \Gamma(\mu,\nu)} \int c(x,y) d\gamma(x,y) + \varepsilon H(\gamma) \tag{5.1}$$

It turns out that this problem is linked to the Schrödinger's Brige Problem about the flow density of particles between two points. We refer the interested reader to [19, 20] and the references therein.

5.2 Around Γ-convergence

Given a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, we define its entropy $H_1(\mu)$ as above and it's second moment by $M_2(\mu) = \int_{\mathbb{R}^d} |x|^2 dx$. We denote by $\mathcal{P}_2(\mathbb{R}^d)$ the set of all probability measures on $\mathbb{R}^d$ with finite second order moment. Given $\mu, \nu \in \mathcal{P}_2^{ac}(\mathbb{R}^d)$, such that $H_1(\mu), H_1(\nu) < \infty$, consider $\gamma \in \Gamma(\mu, \nu)$. We similarly define it's entropy by

$$H_2(\gamma(x,y)) = \begin{cases} \int_{\mathbb{R}^d \times \mathbb{R}^d} \gamma(x,y) \log(\gamma(x,y)) dxdy, \ if \ \gamma \in \mathcal{P}^{ac}(\mathbb{R}^d \times \mathbb{R}^d) \\ +\infty \ otherwise. \end{cases}$$

Given $\varepsilon \geq 0$ we define the functional

$$K_\varepsilon(\gamma) = \begin{cases} \int c d\gamma + \varepsilon H_2(\gamma) \ if \ \gamma \in \Gamma(\mu, \nu) \\ +\infty \ otherwise \end{cases}$$

and $K := K_0$. A natural question is whether K_ε converges in some sense to K as $\varepsilon \to 0$. This kind of questions can be naturally adressed in the framework of Γ-convergence.

Definition 5.1 *We say that K_k Γ-converges to K if for every $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, one has:*

- $\Gamma - \liminf$ *condition:*

  For and any sequence $\gamma_k \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\gamma_k \rightharpoonup \gamma$

$$K(\gamma) \leq \liminf_{k \to \infty} K_k(\gamma_k).$$

- $\Gamma - \limsup$ *condition:*

  There exists a sequence $(\gamma_k)_{k\in\mathbb{N}}$ *such that* $\gamma_k \rightharpoonup \gamma$ *and*

$$K(\gamma) \geq \limsup_{k\to\infty} K_k(\gamma_k).$$

We note that if $(K_k)_{k\in\mathbb{N}}$ *is equi-coercive and* $\Gamma-$*converges to* K, *then* $\lim_{k\to\infty} \inf K_k = \inf K$. *Moreover if the minimizer* γ *of* K *is unique, then the sequence* $(\gamma_k)_{k\in\mathbb{N}}$ *of minimizers of* K_k *converges to* γ. *More details about* $\Gamma-$*convergence can be found in [15].*

It has been proven in [10] that the regularized functional K_ε Γ-converges to K as $\varepsilon \to 0$.

Theorem 5.2 ([10]) *Given two probability measures on* $\mathbb{R}^d$ μ *and* ν *with finite entropy and finite moment of order* $p > 1$. *Assume that the cost function is given by* $c(x,y) = h(|x-y|)$ *with* $|x|^p \leq h(x) \leq |x|^p + 1$. *Then* K_ε Γ*-convergences towards* K *as* $\varepsilon \to 0$ *with respect to narrow topology.*

Here, we consider a variant of this problem where $\varepsilon > 0$ is fixed but where the margins μ, ν are approximated by a family μ_η, ν_η $(\eta > 0)$. We then introduce

$$K_\varepsilon^\eta(\gamma) = \begin{cases} \int c d\gamma + \varepsilon H_2(\gamma) \ if \ \gamma \in \Gamma(\mu_\eta, \nu_\eta) \\ +\infty \ otherwise \end{cases}$$

and the associated minimization problems

$$\inf_{\gamma\in\Gamma(\mu,\nu)} K_\varepsilon(\gamma) \tag{5.2}$$

$$\inf_{\gamma\in\Gamma(\mu_\eta,\nu_\eta)} K_\varepsilon^\eta(\gamma) \tag{5.3}$$

A natural question one may ask is : Does the minima (and the minimizers) of K_ε^η converge to the ones of K_ε as η goes to zero ? A first remark is that this result is false for arbitrary approximations $(\mu_\eta)_{\eta>0}, (\nu_\eta)_{\eta>0}$ of μ, ν. Indeed, if μ_η and ν_η are finitely supported, then K_ε^η is constant equal to $+\infty$! This motivates the introduction of the so called block-approximation at scale η. From now on, we will work on $[0,1]^d$ instead of $\mathbb{R}^d$.

Definition 5.3 (Block approximation) *Given* $k \in \mathbb{Z}^d$, $\eta > 0$, *define* $C_k = \prod_{i=1}^d [k_i, k_i + 1)$ *and* $C_k^\eta = \eta C_k$. *Given a probability measure* μ *on* $\mathbb{R}^d$ *with finite second moment, we consider the following block approximations of* μ *at scale* η*:*

$$\mu_\eta = \sum_{j\in\mathbb{Z}^d} \frac{\mu(C_j^\eta)}{\lambda(C_j^\eta)} \mathbb{1}_{C_j^\eta}$$

We can easily check that $\mu_\eta \in \mathcal{P}_2^{ac}(\mathbb{R}^d)$ *with finite entropy.*

Theorem 5.4 *Fix a regularization parameter* $\varepsilon > 0$, *and let* $\eta > 0$ *be the parameter of discretization and let* μ_η *and* ν_η *be block approximations of* μ, ν. *Then* K_ε^η $\Gamma-$*converges towards* K_ε *as* $\eta \to 0$ *with respect to narrow topology. Moreover, if* γ *is the unique minimizer of* K_ε *and* γ_η *is the unique minimizer of* K_ε^η. *Then* $\gamma_\eta \rightharpoonup \gamma$ *as* η *goes to zero.*

The proof is divided in two steps. We start by establishing the $\Gamma - \liminf$ condition, which is somehow the easy part, and it follows mostly form the lower semicontinuity of the entropy. To establish the $\Gamma - \limsup$ condition, we consider a solution γ of (5.2) and we construct its approximation γ_η at scale η in an appropriate way so that $K_\varepsilon(\gamma_\eta) \leq K_\varepsilon(\gamma) + f(\eta)$ for some function $f \underset{\eta\to 0}{\to} 0$.

Take a sequence $(\eta_k)_k$ tending to 0 as k goes to infinity. We shall prove that $\lim_{k\to\infty} \inf_{\gamma_\eta} K_\varepsilon^{\eta_k}(\gamma) = \inf_{\gamma\in\Gamma(\mu,\nu)} K_\varepsilon(\gamma)$. First, we start by showing the existence of minimizers of (5.2).

Lemma 5.5 *Suppose that $\gamma \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ with bounded support. Then $\gamma \mapsto H_2(\gamma)$ is lower semicontinuous.*

Proof. See [22, Proposition 7.7]. □

This gives the following:

Proposition 5.6 ($\Gamma - \liminf$ condition) *Let $\gamma_k \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ converging narrowly to a certain $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Then*

$$K_\varepsilon(\gamma) \leq \liminf_{k\to\infty} K_\varepsilon(\gamma_k)$$

where $\eta_k \to 0$ as $k \to \infty$

Proof. It's a direct consequence of the previous results. Particularly, lsc of $\gamma \mapsto \int c d\gamma$ and $\gamma \mapsto H_2(\gamma)$, and narrow compactness of $\Gamma(\mu,\nu)$. □

Proposition 5.7 ([10]) *Let $\mu, \nu \in \mathcal{P}_2^{ac}(\mathbb{R}^d)$ with finite entropy. Then, there exists $\gamma \in \Gamma^{ac}(\mu,\nu)$ with finite entropy, minimizing* (5.2).

Proof. This a direct consequence of narrow compactness of $\Gamma(\mu,\nu)$, narrow lower semi-continuity of $\gamma \to \int c d\gamma$ and lsc of the entropy H_2. □

The proof of the Γ-limsup part relies mainly on the two following lemmas.

Lemma 5.8 *If $\gamma \in \Gamma(\mu,\nu)$ and γ_η is its block approximation,*

$$\gamma_\eta = \sum_{j,k\in\mathbb{Z}^d} \gamma(C_j^\eta \times C_k^\eta) \frac{\mathbb{1}_{C_j^\eta}}{\lambda(C_j^\eta)} \otimes \frac{\mathbb{1}_{C_k^\eta}}{\lambda(C_k^\eta)},$$

then $\gamma_\eta \in \Gamma(\mu_\eta, \nu_\eta)$.

Proof. Take a Borelian $A \subset \mathbb{R}^d$, we have

$$\begin{aligned}\gamma_\eta(A \times \mathbb{R}^d) &= \sum_{j,k\in\mathbb{Z}^d} \gamma(C_j^\eta \times C_k^\eta) \left(\frac{\mathbb{1}_{C_j^\eta}}{\lambda(C_j^\eta)} \otimes \frac{\mathbb{1}_{C_k^\eta}}{\lambda(C_k^\eta)} \right) \Big(A \times \mathbb{R}^d \Big) \\ &= \sum_{j\in\mathbb{Z}^d} \frac{\mathbb{1}_{C_j^\eta \cap A}}{\lambda(C_j^\eta)} \sum_{k\in\mathbb{Z}^d} \gamma(C_j^\eta \times C_k^\eta) = \sum_{j\in\mathbb{Z}^d} \frac{\mathbb{1}_{C_j^\eta\cap A}}{\lambda(C_j^\eta)} \mu(C_j^\eta) \\ &= \mu_\eta(A).\end{aligned} \tag{5.4}$$

We verify similarly that $\gamma_\eta(\mathbb{R}^d \times A) = \nu_\eta(A)$. □

Lemma 5.9 *Let $\rho \in \mathcal{P}^{ac}([0,1]^d)$ and ρ_η it's approximation at scale η. Then*

$$\int_{[0,1]^d} \rho(x)\log(\rho(x))dx \geq \int_{[0,1]^d} \rho_\eta(x)\log(\rho_\eta(x))dx.$$

Proof. It's a direct consequence of Jensen inequality. Indeed, since $r \mapsto r\log r$ is convex, we have

$$\begin{aligned}\int_{[0,1]^d} \rho\log\rho dx &= \sum_j \int_{C_j^\eta} \rho\log\rho dx \geq \sum_j \int_{C_j^\eta} \rho(x)dx \log\left(\int_{C_j^\eta}\rho(x)dx\right)\\ &= \sum_j \rho(C_j^\eta)\frac{\mathbb{1}_{C_j^\eta}}{\lambda(C_j^\eta)}\log\left(\rho(C_j^\eta)\frac{\mathbb{1}_{C_j^\eta}}{\lambda(C_j^\eta)}\right) = \int_{[0,1]^d}\rho_\eta(x)\log(\rho_\eta(x))dx \quad (5.5)\end{aligned}$$

□

Corollary 5.10 *We have*

$$H_2(\gamma) \geq H_2(\gamma_\eta)$$

Proposition 5.11 *We have $\gamma_\eta \rightharpoonup \gamma$ as $\eta \to 0$.*

Proof. To do so, we only need to prove that $W_2(\gamma_\eta, \gamma) \to 0$ as η goes to zero, where W_2 is the 2-Wasserstein distance between γ_η and γ. As constructed before, $(C_i \times C_j)_{(i,j)\in I\times J}$ is a countable partition of $\mathbb{R}^d \times \mathbb{R}^d$, where I, J are countable sets of indices. It is clear that $\sup_{i,j} diam(C_i \times C_j) \leq C$ with C a positive constant. Moreover, since $\gamma \ll \mathcal{L}^{d\times d}$, with $\mathcal{L}^{d\times d}$ is the Lebesgue measure on $\mathbb{R}^d \times \mathbb{R}^d$, we have $\gamma(C_i \times C_j) \leq \tilde{C}\eta^2$, with $\tilde{C}$ is a constant depending only on the density of γ. Consider $\tilde{I} \subset I$, and $\tilde{J} \subset J$, such that $\gamma_\eta(C_i \times C_j) = \gamma(C_i \times C_j)$ on $\tilde{I} \times \tilde{J}$. Define

$$\gamma_\eta^{ij}(A) = \frac{\gamma_\eta(A \cap (C_i \times C_j))}{\gamma_\eta(C_i \times C_j)} \quad and \quad \gamma^{ij}(A) = \frac{\gamma(A \cap (C_i \times C_j))}{\gamma(C_i \times C_j)}.$$

for every Borelian $A \subset \mathbb{R}^d \times \mathbb{R}^d$. Clearly, $\gamma_\eta^{ij}, \gamma^{ij} \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$, and are supported in $\overline{C_i \times C_j}$. Take $\Lambda \in \Gamma(\gamma_\eta^{ij}, \gamma^{ij})$ such that $supp(\Lambda) \subset (\overline{C_i \times C_j})^2$. We have

$$\int_{(\mathbb{R}^d\times\mathbb{R}^d)^2} |x-y|^2 d\Lambda = \int_{(C_i\times C_j)^2} |x-y|^2 d\Lambda \leq C^2.$$

Define $\tilde{\Lambda} = \sum_{(i,j)\in(\tilde{I}\times\tilde{J})} \gamma(C_i \times C_j)\Lambda$. We can verify easily that $\tilde{\Lambda}$ is a transport plan between γ_η and γ. Indeed, for every Borelian $A \subset \mathbb{R}^d \times \mathbb{R}^d$ we have

$$\begin{aligned}\tilde{\Lambda}((\mathbb{R}^d\times\mathbb{R}^d)\times A) = \sum_{i,j}\gamma(C_i\times C_j)\Lambda((\mathbb{R}^d\times\mathbb{R}^d)\times A) &= \sum_{ij}\gamma(C_i\times C_j)\gamma_\eta^{ij}(A)\\ &= \sum_{ij}\gamma_\eta(A\cap(C_i\times C_j)) = \gamma_\eta(A). \quad (5.6)\end{aligned}$$

Similarly, we show that the other marginal is γ. Hence

$$W_2^2(\gamma_\eta,\gamma) \leq \int |x-y|^2 d\tilde{\Lambda} = \sum_{ij}\gamma(C_i\times C_j)\int|x-y|^2 d\Lambda \lesssim \eta^2$$

So as $\eta \to 0$, $W_2^2(\gamma_\eta, \gamma) \to 0$, and narrow convergence follow. □

Lemma 5.12 *Let γ, γ_η be respectively the optimal plans for* (5.2) *and* (5.3). *Then* $\int cd(\gamma - \gamma_\eta) \to 0$ *as* $\eta \to 0$.

Proof. Since c is assumed to be 1-Lipschitz we have $\int cd(\gamma - \gamma_\eta) \leq W_1(\gamma, \gamma_\eta)$, where W_1 is the 1-Wasserstein distance defined by

$$W_1(\gamma, \gamma_\eta) = \sup_{\phi \text{ } 1-Lipschitz} \int \phi d(\gamma - \gamma_\eta)$$

We know that $W_1(\gamma_\eta, \gamma) \leq W_2(\gamma_\eta, \gamma)$, hence by Proposition 1, $W_1(\gamma_\eta, \gamma) \to 0$ as η goes to zero which implies that $\int cd\gamma_\eta \to \int cd\gamma$ as $\eta \to 0$. □

Proposition 5.13 ($\Gamma - \limsup$ condition) *There exists $\eta_k \to 0$ as $k \to \infty$ such that*

$$K_\varepsilon(\gamma) \geq \limsup K_\varepsilon^{\eta_k}(\gamma_{\eta_k})$$

Proof. For every no-negative sequence $(\eta_k)_k$ converging to zero, $\gamma_{\eta_k} \rightharpoonup \gamma$. Using the previous results we have $H_2(\gamma_{\eta_k}) \leq H_2(\gamma)$. By the lemma 4.3.6 $\int cd\gamma_{\eta_k} \to \int cd\gamma$, we deduce that

$$\limsup_{k \to \infty} \int cd\gamma_{\eta_k} + H_2(\gamma_{\eta_k}) \leq \int cd\gamma + H_2(\gamma).$$

ad desired. □

Remark 5.14 The problem (5.3) is not fully discrete since we are minimizing on $\gamma \in \Gamma(\mu_\eta, \nu_\eta)$. To make it a fully discrete problem, one has to add the following functional

$$F^\eta(\gamma) = \begin{cases} 0 & if \ \gamma = \sum_{ij} \gamma_{ij} 1_{C_i^\eta} \otimes 1_{C_j^\eta} \\ +\infty & if \ not \end{cases}$$

The problem becomes an optimization problem in finite dimension with unknown γ_{ij}. Using the same arguments as before, we can show that the minimizers of $K_\varepsilon^\eta + F^\eta$ converge to the minimizers of K_ε as $\eta \to 0$.

5.3 Fully discrete setting

We go back to the "standard" Kantorovich problem, *i.e*, the two marginal problem with $\mu \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$, $\gamma \in \mathbb{R}^{n \times n}$

$$(\mathcal{KP}_2) : \inf_{\gamma \in \Gamma(\mu,\nu)} \left\{ K(\gamma) = \int_{X \times Y} cd\gamma \right\} \tag{5.7}$$

which is a linear optimization problem. So it is important to look at it discretization. We replace μ and ν by a finite combination of Dirac masses:

$$\mu = \sum_{i=1}^n \alpha_i \delta_{x_i}, \quad \nu = \sum_{i=1}^n \beta_i \delta_{y_i}.$$

where α_i, β_i are non-negative numbers and for simplicity we have assumed that $X = \{x_1, \cdots, x_n\}$ and $Y = \{y_1, \cdots, y_n\}$ have the same cardinality n. Suppose that we are transferring mass $a_{ij} \geq 0$ form x_i to y_j. Then transport plans are of the form $\gamma = \sum_{i,j=1}^n a_{ij} \delta_{x_i} \otimes \delta_{y_j}$. Since we send all the mass at x_i, we must have $\sum_{i=1} a_{ij} = \alpha_i$, and since the mass transported to y_j must be equal to the original mass at y_j, we have $\sum_{j=1} a_{ij} = \beta_j$. Every transport plan costs $\sum_{i,j=1}^{n} a_{ij}c_{ij}$ where $c_{ij} = c(x_i, y_j)$. The discrete Kantorovich problem writes

$$(Dis\mathcal{KP}) : \inf_{\gamma\in\Gamma_n(\mu,\nu)} \left\{ \sum_{i,j}^{n} c_{ij}\gamma_{ij} \right\} \tag{5.8}$$

where $\Gamma(\mu,\nu) = \{\gamma \in \mathbb{R}^n \times \mathbb{R}^n,\ \gamma_{ij} \geq 0, \sum_{j=1}\gamma_{ij} = \alpha_i,\ \sum_{i=1}\gamma_{ij} = \beta_j\}$

Remark 5.15 The dual problem of 5.8 which can be obtained similarly to the continuous setting. We have

$$\begin{aligned}
\min_{\gamma\in\Gamma(\mu,\nu)} \sum_{ij} \gamma_{ij}c_{ij} &= \min_{\gamma\in\Gamma(\mu,\nu)} \max_{\phi,\psi\in\mathbb{R}^n} \sum_{ij} \gamma_{ij}c_{ij} + \sum_i \Big(\alpha_i - \sum_j \gamma_{ij}\Big)\phi_i + \sum_j \Big(\beta_j - \sum_i \gamma_{ij}\Big)\psi_j \\
&= \min_{\gamma\in\Gamma(\mu,\nu)} \max_{\phi,\psi\in\mathbb{R}^n} \sum_{ij} \gamma_{ij}(c_{ij} - \phi_i - \psi_j) + \sum_i \alpha_i\phi_i + \sum_j \beta_j\psi_j \\
&= \max_{\phi,\psi,\ \phi\oplus\psi_j\leq c} \sum_i \alpha_i\phi_i + \sum_j \beta_j\psi_j
\end{aligned}$$

where $\phi \oplus \psi \leq c$ means that $\phi_i + \psi_j \leq c_{ij}$ for all $i, j = 1, \cdots, n$.

To define the entropic regularization of (5.8), we define the discrete entropy by

$$E(\gamma) = \begin{cases} \gamma(\log(\gamma) - 1),\ if\ \gamma \geq 0 \\ 0 \quad if\ \gamma = 0 \\ +\infty \quad if\ not \end{cases} \tag{5.9}$$

and we consider, for fixed ε, the following penalized problem

$$(\mathcal{KP}_\varepsilon) : \min_{\gamma\in\Gamma(\mu,\nu)} \left\{ \sum_{ij} c_{ij}\gamma_{ij} + \varepsilon E(\gamma_{\mathrm{ij}}) \right\} \tag{5.10}$$

The entropic regularization allows to get ride of the positivity constraint on γ. Moreover, this problem becomes a problem of projection the constraint subspaces $\{\sum_{i=1}\gamma_{ij} = \mu_i\}$ and $\{\sum_{j=1}\gamma_{ij} = \nu_j\}$, and as we will see in the following sections, we can compute explicitly the projections.

Remark 5.16

- The regularized problem $(\mathcal{KP}_\varepsilon)$ has a unique solution γ_ε. Moreover, if the solution γ of the non-regularised problem is unique then we have $\gamma_\varepsilon \to \gamma$ as ε goes to zero. Moreover, the convergence is exponential:

$$\|\gamma_\varepsilon - \gamma\|_{\mathbb{R}^d\times\mathbb{R}^d} \leq ke^{-\lambda/\varepsilon}$$

  where k and λ depend on the cost function, the marginals mu and ν, and the discretization number n. More details can be found in [12].

- If the non-regularized problem has more than one solution, then γ_ε converges to the one with minimal entropy as ε goes to zero [14, Proposition 2.1].

- When $\varepsilon \to \infty$, we find that $\gamma_\varepsilon = \mu \otimes \nu$. Indeed, if$\gamma_\varepsilon$ is the minimizer of the regularized problem then

$$\gamma_\varepsilon = argmin_{\gamma\in\Gamma(\mu,\nu)} KL(\gamma \mid \bar{\gamma})$$

  with $\bar{\gamma} = e^{-c/\varepsilon}\mu\otimes\nu$, and hence $\frac{\partial KL(\gamma_\varepsilon|\bar{\gamma})}{\partial_{\gamma_\varepsilon}} = 0$. This gives $\gamma_\varepsilon = e^{-c/\varepsilon}\mu\otimes\nu \to \mu\otimes\nu$ as ε goes to ∞.

5.4 Algorithms and Numerical Results

In this section we present some algorithms used to solve regularized OT problems. We distinguish two points of view: geometric projection vs dual optimization point of view. In the first category we handle transport plans, *i.e*, we are dealing with the primal transport problem. Whereas in the second category we deal with the dual problem, and so we handle Kantorovich potentials. We present this in details and we will show the advantages and the drawbacks of each method.

5.4.1 Geometric projection point of view

Given $\gamma \in \mathbb{R}_{+}^{d\times d}$ and $\xi \in \mathbb{R}_{+*}^{d\times d}$, $i.e$, $\xi_{ij} > 0$ for all i, j, we define the Kullback-Leibler divergence between γ and ξ by

$$KL(\gamma \mid \xi) = \sum_{i,j=1}^{d} \gamma_{ij}\Big(\log(\frac{\gamma_{ij}}{\xi_{ij}}) - 1\Big).$$

If $\mathcal{C}$ is a convex subset of $\mathbb{R}^{d\times d}$, the projection according to the KL divergence is defined as

$$\mathcal{P}_{\mathcal{C}}^{KL} = argmin_{\gamma\in\mathcal{C}} KL(\gamma \mid \xi)$$

Optimal transport The next lemma shows that the regularized optimal transport problem is equivalent to the computation of the projection of some density (Gibbs kernel) on the intersection of some convex sets with respect to the $KL-$divergence.

Lemma 5.17 *[3]*

- *The regularized optimal transport problem* (5.10) *can be rewritten in the form*

$$\min_{\gamma\in\mathcal{C}} KL(\gamma \mid \xi), \quad \mathcal{C} = \cap_{i=1}^{2}\mathcal{C}_i. \tag{5.11}$$

where $\xi_{ij} = e^{-c_{ij}/\varepsilon}$ *and where the* $\mathcal{C}_i$ *are the affine subspaces of* $\mathbb{R}^{d\times d}$ *defined by*

$$\mathcal{C}_1 = \{\gamma \in \mathbb{R}_{+}^{d\times d},\ \gamma\mathbb{1} = \mu\}, \quad \mathcal{C}_2 = \{\gamma \in \mathbb{R}_{+}^{d\times d},\ \gamma^{\dagger}\mathbb{1} = \nu\}.$$

with $\mathbb{1} = (1, \cdots, 1)^{\dagger} \in \mathbb{R}^n$ *and* $\gamma\mathbb{1}$ *is a matrix-vector product.*

$$\mathcal{P}_{\mathcal{C}_1}^{KL}(\gamma) = diag\Big(\frac{\mu}{\gamma\mathbb{1}}\Big)\gamma, \ and\ \mathcal{P}_{\mathcal{C}_2}^{KL}(\gamma) = \gamma diag\Big(\frac{\nu}{\gamma^{\dagger}\mathbb{1}}\Big).$$

While there are explicit formulas for the projection on $\mathcal{C}_1$ and $\mathcal{C}_1$, it is not as straightforward to project ξ on $\mathcal{C}_1 \cap \mathcal{C}_2$. The KL-projection on the intersection of two affine subspaces can however be computed using the *alternating projection algorithm*:

$$\begin{cases} \gamma_0 := \xi \\ \gamma_{(n)} = \mathcal{P}_{\mathcal{C}_1}^{KL}(\gamma_{(n-1)}) & \text{for } n \text{ odd },\\ \gamma_{(n)} = \mathcal{P}_{\mathcal{C}_2}^{KL}(\gamma_{(n-1)}) & \text{for } n \text{ even }, \end{cases}$$

This procedure is illustrated by Figure 4, and it is possible to prove [5] that as $n \to \infty$, $\gamma_{(n)}$ converges to $\mathcal{P}^{KL}_{\mathcal{C}}(\xi)$, i.e. the unique solution of (5.11). Figure 2 shows the transport plans computed by this algorithm for various values of ε.

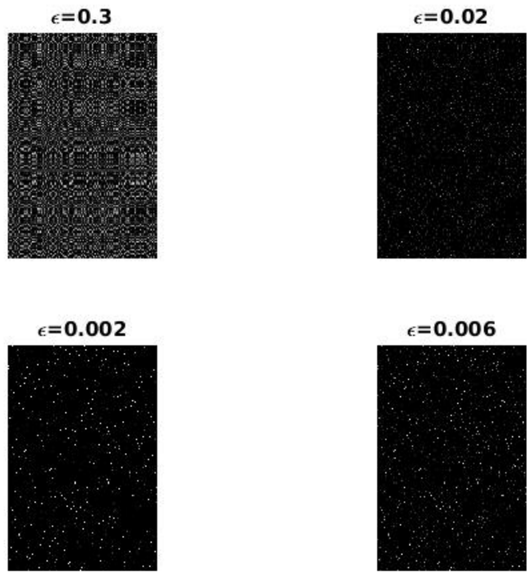

Figure 2: Regularized Transport for different values of the regularization parameters

We verify that the regularized is sufficiently close to the initial marginals.

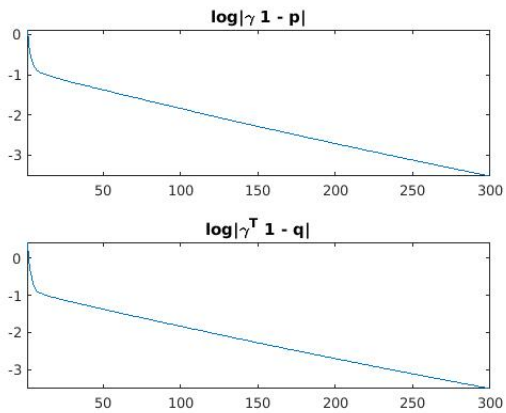

Figure 3: The error $\|\mathbb{1}\gamma - \mu\|$ and $\|\gamma^{\dagger}\mathbb{1} - \nu\|$ at $\log_{(10)}$ scale

Remark 5.18 Combining this formulas with the Iterative Bregman Projection (affine case), we find that the iterates satisfy

$$\gamma_{(n)} = diag(a_{(n)})\xi diag(b_{(n)}).$$

with $(a_{(n)}, b_{(n)}) \in \mathbb{R}^d \times \mathbb{R}^d$ satisfies $b_{(0)} = \mathbb{1}$ and

$$a_{(n)} = \frac{\mu}{\xi b_{(n)}}, \; and \; b_{(n+1)} = \frac{\nu}{\xi^{\dagger} a_{(n)}}.$$

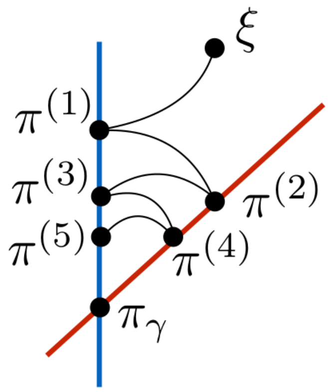

Figure 4: The projection of Gibbs kernel ξ on the intersection of two convex sets (in red and blue).

and this is a "transport plan version" of Sinkhorn-Knopp's algorithm which will be introduced in the next section.

Remark 5.19 The KL-divergence is a particular case of the notion of a divergence induced by an entropy functional wich is a convex lower semi-continuous function ϕ such that $dom\phi \subset [0, \infty[$ and $dom\phi \cap [0, \infty[\neq \emptyset$. The speed of growth of ϕ at ∞ is defined by

$$\phi'_\infty = \lim_{x \to +\infty} \frac{\phi(x)}{x} \in \mathbb{R} \cup \{\infty\}.$$

Given an entropy ϕ and two measures μ and ν on $\mathbb{R}^d$, we suppose that the Lebesgue decomposition of μ with respect to ν is $\frac{d\mu}{d\nu}\nu + \mu^\perp$. The divergence induced by ϕ is defined by

$$\mathcal{D}_\phi(\mu \mid \nu) = \begin{cases} \int_{\mathbb{R}^d} \phi\left(\frac{d\mu}{d\nu}\right) d\nu + \phi'_\infty \mu^\perp(\mathbb{R}^d) \ if\ \mu, \nu \ are\ non-negative. \\ \infty\ otherwise. \end{cases}$$

We can easily see that the KL-divergence is associated to the entropy function

$$\phi_{KL}(s) = \begin{cases} s\log(s) - s + 1,\ if\ s > 0. \\ 1\ if\ s = 0. \\ \infty\ otherwise. \end{cases}$$

Partial optimal transport In the case of partial transport we start with two marginals $\mu, \nu \in \mathbb{R}^d_+$ and we wish to transport a mass portion $0 \leq m \leq \min(\mu^\dagger \mathbb{1}, \nu^\dagger \mathbb{1})$. The corresponding regularized problem reads

$$\min_{\gamma \in \mathbb{R}^d_+} < c \mid \gamma > + \varepsilon E(\gamma),\ with\ \gamma \mathbb{1} \leq \mu, \gamma^\dagger \mathbb{1} \leq \nu, \mathbb{1}^\dagger \gamma \mathbb{1} = m. \tag{5.12}$$

We follow the same strategy as before, rewriting this problem as the KL-projection on the intersection of a family of convex subsets of $\mathbb{R}^{d\times d}$.

Lemma 5.20 *[3] The regularized optimal partial transport problem* (5.12) *can be rewritten in the form*

$$\min_{\gamma \in \mathcal{C}} KL(\gamma \mid \xi), \ \mathcal{C} = \cap_{i=1}^{3} \mathcal{C}_i.$$

with $\xi = e^{-c/\varepsilon}$ *and*

$$\mathcal{C}_1 = \{\gamma \in \mathbb{R}_+^d, \gamma\mathbb{1} \leq p\}, \mathcal{C}_2 = \{\gamma \in \mathbb{R}_+^d, \gamma^\dagger\mathbb{1} \leq q\}, \mathcal{C}_3 = \{\gamma \in \mathbb{R}_+^d, \mathbb{1}^\dagger\gamma\mathbb{1} = m\}.$$

Moreover we have explicit formulas for the projection $\mathcal{P}_{\mathcal{C}_i}^{KL}$*:*

$$\mathcal{P}_{\mathcal{C}_1}^{KL}(\gamma) = diag\Big(\min(\frac{\mu}{\gamma\mathbb{1}}, \mathbb{1})\Big)\gamma, \ \mathcal{P}_{\mathcal{C}_2}^{KL}(\gamma) = \gamma diag\Big(\min(\frac{\nu}{\gamma^\dagger\mathbb{1}}, \mathbb{1})\Big), \ \mathcal{P}_{\mathcal{C}_3}^{KL}(\gamma) = \gamma\frac{m}{\mathbb{1}^\dagger\gamma\mathbb{1}}.$$

with $\min(x, y) \triangleq (\min(x_i, y_i))_i$.

Since the $\mathcal{C}_i$ are not all affine subspaces, the alternating projection algorithm doesn't converge in general to $\mathcal{P}_{\mathcal{C}}^{KL}(\xi)$. Fortunately, an extension of Dykstra's algorithm [16] to the KL setting convergences to the projection. This can be done as follows:

$$\begin{cases} \gamma_0 := \xi, \ q_{(0)} = q_{(-1)} = \cdots = q_{(-k+1)} = \mathbb{1}. \\ \gamma_{(n)} = \mathcal{P}_{\mathcal{C}_n}^{KL}(\gamma_{n-1} \odot q_{n-k}). \\ q_{(n)} = q_{(n-k)} \odot \frac{\gamma_{(n-1)}}{\gamma_{(n)}}. \\ \mathcal{C}_{i+3k} = \mathcal{C}_i \ for \ i = 1, 2, 3 \ and \ k \in \mathbb{N}. \end{cases}$$

With $x \odot y = (x_i y_i)_i \in \mathbb{R}^d$ and $\frac{x}{y} = \left(\frac{x_i}{y_i}\right)_i \in \mathbb{R}^d$, for $x, y \in \mathbb{R}^d$. We note that Dykstra's algorithm is used in the two cases: $m = 1$ and $m < 1$. Moreover, we can prove [2] that $\gamma_{(n)} \to \mathcal{P}_{\mathcal{C}}^{KL}$ as $n \to \infty$. Figure 5 shows the solution computed via this algorithm.

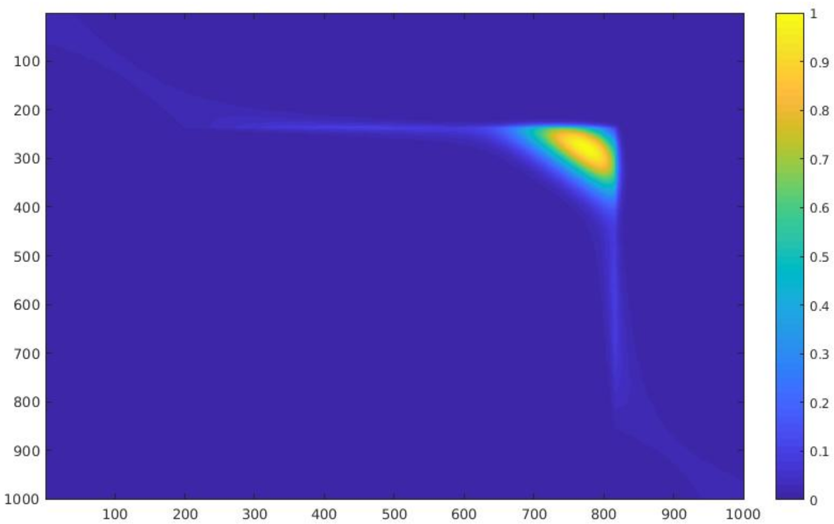

Figure 5: The solution of the partial transport between two Gaussians

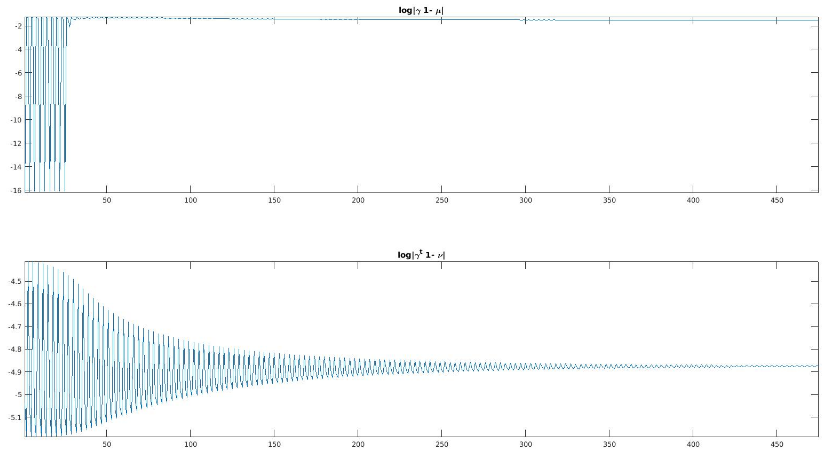

Figure 6: The error $\|\mathbb{1}\gamma - \mu\|$ and $\|\gamma^{\dagger}\mathbb{1} - \nu\|$ at $\log_{(10)}$ scale

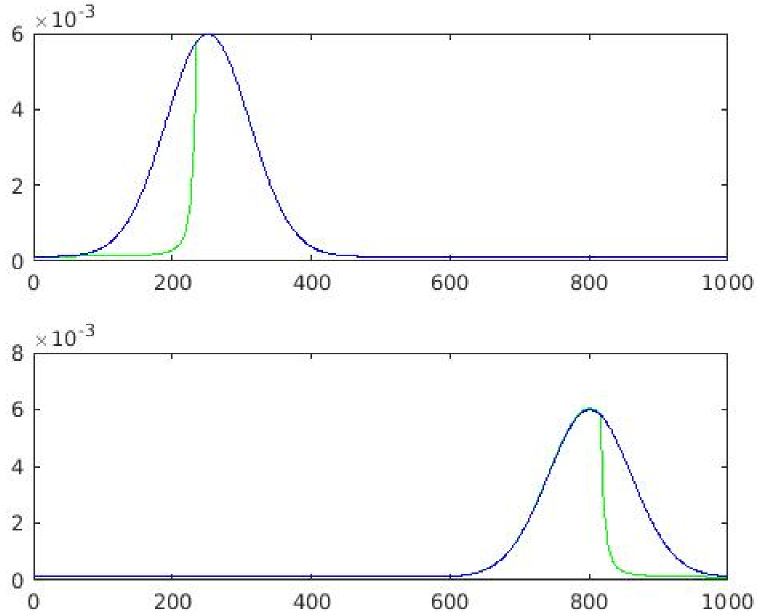

Figure 7: The marginals of the computed solution to the partial problem (In green) plotted alongside with μ and ν.

5.4.2 Dual Optimization point of view

In this section we start from a regularized optimal transport problem and we derive its dual formulation. This makes Kantorovich potentials appear, and thanks to the algebraic properties of the entropy, the dual problem is solved by maximizing alternatively in the potentials. This approach is privileged since handling potentials is numerically easier that using transport plans which appear need more memory, and the algebraic operations are more complicated (in the case of $N = 3$ marginals we deal with a 3−tensor). Let $\gamma \in \mathbb{R}^{n\times n}$, we define the discrete entropy by

$$E(\gamma) = \begin{cases} \gamma(\log(\gamma) - 1), \ if \ \gamma \geq 0 \\ 0 \quad if \ \gamma = 0 \\ +\infty \quad if \ not \end{cases} \tag{5.13}$$

For $\varepsilon > 0$ we consider the following penalized problem

$$(\mathcal{KP}_\varepsilon) : \min_{\gamma \in \Gamma(\mu,\nu)} \Big\{ \sum_{ij} c_{ij}\gamma_{ij} + \varepsilon E(\gamma_{\text{ij}}) \Big\} \tag{5.14}$$

To obtain the dual problem, we take $\phi, \psi \in \mathbb{R}^n$ to express the two marginal constraints:

$$\begin{aligned} \min_{\gamma \in \Gamma(\mu,\nu)} \Big\{ \sum_{ij} c_{ij}\gamma_{ij} + \varepsilon E(\gamma) \Big\} &= \min_{\gamma_{ij}} \sup_{\phi_i,\psi_j} \sum_{ij} \gamma_{ij} c_{ij} + \varepsilon E(\gamma_{ij}) \\ &+ \sum_i (\sum_j \mu_i - \gamma_{ij})\phi_i + \sum_j (\sum_i \nu_j - \gamma_{ij})\psi_j \end{aligned} \tag{5.15}$$

by interchanging the inf-sup we obtain

$$\begin{aligned} (\mathcal{KP}_\varepsilon) = \sup_{\phi_i,\psi_j} \min_{\gamma_{ij}} \sum_{ij} \gamma_{ij} c_{ij} + \varepsilon E(\gamma_{ij}) \\ + \sum_i (\sum_j \mu_i - \gamma_{ij})\phi_i + \sum_j (\sum_i \nu_j - \gamma_{ij})\psi_j \end{aligned} \tag{5.16}$$

We can obtain an explicit formula of the optimal plan γ in terms of c, ϕ, ψ, by taking the derivative with respect to γ_{ij}:

$$c_{ij} + \varepsilon \partial_{\gamma_{ij}} E(\gamma_{ij}) - \phi_i + \psi_j = 0$$

$i.e$, $\gamma_{ij} = \exp(\frac{\phi_i + \psi_j - c_{ij}}{\varepsilon})$. We plug theses values in $(\mathcal{KP}_\uparrow)$ to get the dual problem:

$$(\mathcal{DK}_\varepsilon) \sup_{\phi,\psi \in \mathbb{R}^n} \Psi_\varepsilon(\phi,\psi) \tag{5.17}$$

wherer $\Psi_\varepsilon(\phi,\psi) = \Big\{ \sum_{i=1}^n \phi_i\mu_i + \sum_{j=1}^n \psi_j\mu_j - \varepsilon \sum_{i,j=1}^n \exp(\frac{\phi_i+\psi_j-c_{ij}}{\varepsilon}) \Big\}$. For $\psi \in \mathbb{R}^n$, the maximizer of $\Psi_\varepsilon(.,\psi)$ can be obtained by taking the derivative of Ψ_ε with respect to ϕ:

$$\partial_{\phi_i}\Psi_\varepsilon(\phi,\psi) = \mu_i - \sum_j^n \exp(\frac{\phi_i + \psi_j - c_{ij}}{\varepsilon}) = 0, \quad \forall i = 1, \cdots, n$$

So we obtain that:

$$\frac{\phi_i}{\varepsilon} = \log(\mu_i) - \log\Big(\sum_j^n \exp(\frac{\psi_j - c_{ij}}{\varepsilon})\Big), \ \forall i = 1, \cdots, n.$$

and similarly by fixing $\phi \in \mathbb{R}^n$ and taking the derivative with of Ψ_ε with respect to ψ we obtain:

$$\frac{\psi_j}{\varepsilon} = \log(\nu_j) - \log\Big(\sum_i^n \exp(\frac{\phi_i - c_{ij}}{\varepsilon})\Big), \ \forall j = 1, \cdots, n.$$

Before introducing the Sinkhorn-Knopp algorithm which allows us to solve the dual problem $(\mathcal{KP}_\varepsilon)$ my maximizing alternatively Ψ_ε in ϕ and ψ. We recall that duo to strict convexity of $\sum_{ij} c_{ij}\gamma_{ij} + \varepsilon E(\gamma_{ij})$ the solution γ of $(\mathcal{KP}_\varepsilon)$ is necessarily unique, and form the previous computations it's characterized by solving the equation $c_{ij} + \varepsilon \log(\gamma_{ij}) - \phi_i - \psi_j = 0$. In other words $\gamma_{ij} = \exp(\phi_i/\varepsilon)\exp(-c_{ij}/\varepsilon)\exp(\psi_j/\varepsilon)$ for $i, j = 1, \cdots, n$. Using the marginal constraints $\sum_j \gamma_{ij} = \mu_i, \sum_i \gamma_{ij} = \nu_j$, we find that the Kantorovich potentials ϕ and ψ are uniquely determined by

$$a_i \triangleq \exp(\phi_i/\varepsilon) = \mu_i/(\sum_j \exp(-c_{ij}/\varepsilon))b_j$$

$$b_j \triangleq \exp(\psi_j/\varepsilon) = \nu_j/(\sum_i \exp(-c_{ij}/\varepsilon))a_i.$$

We define as before Gibbs kernel G_ε defined by

$$\begin{array}{rccc} G_\varepsilon & : & \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ & & \xi & \longmapsto & \sum_{j=1}^n \exp(-c_{ij}/\varepsilon)\xi_j, \forall i = 1, \cdots, n \end{array}$$

and similarly it's adjoint can be defined as

$$\begin{array}{rccc} G_\varepsilon^\dagger & : & \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ & & \zeta & \longmapsto & \sum_{i=1}^n \exp(-c_{ij}/\varepsilon)\zeta_j, \forall j = 1, \cdots, n \end{array}$$

Finally, the Sinkhorn-Knopp algorithm is given by:

Algorithm 1 Sinkhorn-Knopp algorithm

function S-K$(\mu, \nu, G\varepsilon)$
  $a_0 \leftarrow 1_{\{1,\cdots,n\}},\ b_0 \leftarrow 1_{\{1,\cdots,n\}}$
  for $0 \leq k \leq k_{max}$ **do**
    $a_{k+1} \leftarrow \mu/G_\varepsilon b_k$
    $b_{k+1} \leftarrow \nu/G_\varepsilon^\dagger a_k$
  end for
end function

Figure 8 displays examples of transport plans γ_ε solving the regularized optimal transport problem where each measure μ, ν is the sum of two Gaussian densities with $N = 200$ and as a cost $c_{ij} = |i - j|^2$.

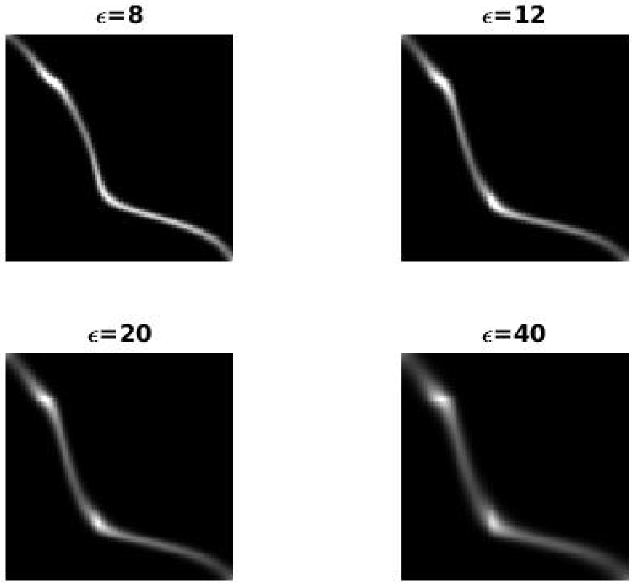

Figure 8: The solution γ_ε for several values of ε.

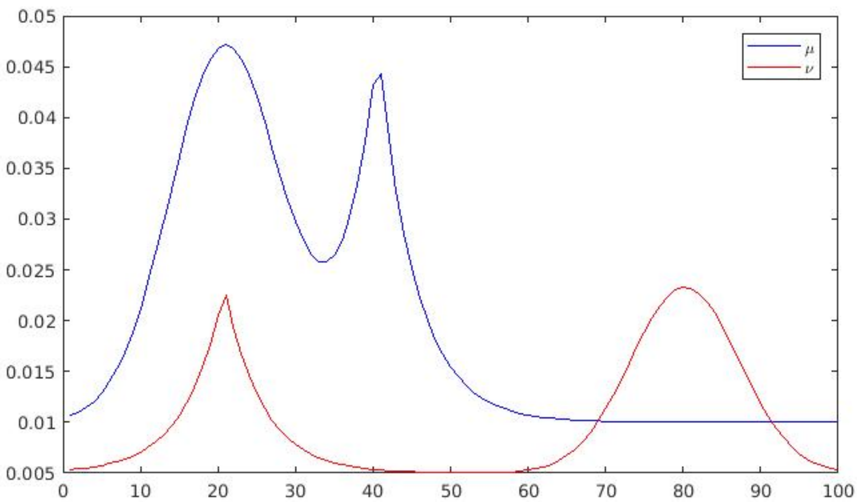

Figure 9: The two marginals μ and ν

6 Appendix

The most known version of Prokhorov's theorem makes a link between tightness of measures to compactness in the space of probability measures. We give here a more general version of this theorem which gives conditions for compactness in the weak topology $\sigma(\mathcal{M}, C_b(X)))$ of families of measures. This result can be found in [4] and we give it here for completeness and because it is not often presented in a such general way in the literature on optimal transport.

Theorem 6.1 *Les X be a compact metric space and let $\mathcal{F}$ be a family of Borel measures on X. Then the following conditions are equivalent:*

- *Every sequence $(\mu_m) \subset \mathcal{F}$ contains a weakly convergent subsequence.*
- *$\mathcal{F}$ is uniformly tight and uniformly bounded in the variation norm.*

References

[1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011.

[2] H. H. Bauschke and A. S. Lewis. Dykstra's algorithm with Bregman projections: A convergence proof. *Optimization*, 48(4):409–427, 2000.

[3] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):a1111–a1138, 2015.

[4] V. I. Bogachev. *Measure theory. Vol. I and II.* Berlin: Springer, 2007.

[5] L. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *Zh. Vychisl. Mat. Mat. Fiz.*, 7:620–631, 1967.

[6] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. (Polar decomposition and increasing rearrangement of vector fields). *C. R. Acad. Sci., Paris, Sér. I*, 305:805–808, 1987.

[7] G. Buttazzo, T. Champion, and L. De Pascale. Continuity and estimates for multimarginal optimal transportation problems with singular costs. *Applied Mathematics & Optimization*, Feb 2017.

[8] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi. Optimal-transport formulation of electronic density-functional theory. *Phys. Rev. A*, 85:062502, Jun 2012.

[9] L. A. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Ann. Math. (2)*, 171(2):673–730, 2010.

[10] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.

[11] G. Carlier and I. Ekeland. Matching for teams. *Econ. Theory*, 42(2):397–418, 2010.

[12] R. Cominetti and J. San Martín. Asymptotic analysis of the exponential penalty trajectory in linear programming. *Math. Program.*, 67(2 (A)):169–187, 1994.

[13] C. Cotar, G. Friesecke, and C. Klüppelberg. Density functional theory and optimal transportation with Coulomb cost. *Commun. Pure Appl. Math.*, 66(4):548–599, 2013.

[14] M. Cuturi and G. Peyré. A smoothed dual approach for variational Wasserstein problems. *SIAM J. Imaging Sci.*, 9(1):320–343, 2016.

[15] G. Dal Maso. *An introduction to Γ-convergence.* Basel: Birkhäuser, 1993.

[16] R. L. Dykstra. An algorithm for restricted least squares regression. *J. Am. Stat. Assoc.*, 78:837–842, 1983.

[17] A. Figalli. The optimal partial transport problem. *Arch. Ration. Mech. Anal.*, 195(2):533–560, 2010.

[18] J. Kitagawa and B. Pass. The multi-marginal optimal partial transport problem. *Forum Math. Sigma*, 3:28, 2015.

[19] C. Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(4):1879–1920, 2012.

[20] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.

[21] L. Nenna. *Numerical Methods for Multi-Marginal Optimal Transportation.* Theses, PSL Research University, Dec. 2016.

[22] F. Santambrogio. *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling.* Cham: Birkhäuser/Springer, 2015.

[23] C. Villani. *Topics in optimal transportation.* Providence, RI: American Mathematical Society (AMS), 2003.

[24] C. Villani. *Optimal transport. Old and new.* Berlin: Springer, 2009.