OPTIMAL TRANSPORTATION OF VECTOR-VALUED MEASURES

Xavier Bacon

OPTIMAL TRANSPORTATION OF VECTOR-VALUED MEASURES

Xavier Bacon *†

Abstract

Given two n-dimensional measures $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ on Polish spaces, we propose an optimal transportation's formulation, inspired by classical Kantorovitch's formulation in the scalar case. In particular, we established a strong duality result and as a consequence, optimality conditions are investigated. Wasserstein's metrics induced by our formulation are also investigated.

Key Words: Optimal Transport, Calculus of variations, Wasserstein distance.

1 Introduction and notations

1.1 Introduction

Starting from the article of G. Monge [1], many mathematical formulations of optimal transportation have been offered ([2], [3] and [4]). In Monge's formulation, given two Polish spaces X and Y, if μ (resp. ν) is a Borelian probability on X (resp. Y) and if $c : X \times Y \to \mathbb{R}$, then the Monge's formulation consists on a minimization of the total cost among all Borelian maps which push forward μ to ν, more precisely for T a Borelian function between X and Y and m a positive measure on X, $T\#\mu$ stands for the push forward measure which is the measure on Y defined for all measurable set B by $T\#\mu(B) := \mu\left[T^{-1}(B)\right]$. Let $M(\mu,\nu)$ be the set of such maps, Monge transportation problem is then

$$\mathcal{M}(\mu,\nu) := \inf\left\{\int_X c\left[x,T(x)\right]\,\mathrm{d}\mu(x) : T \in M(\mu,\nu)\right\}. \tag{1}$$

In the middle of the 20th century, L. Kantorovitch proposed a relaxation of (1) in [2] by allowing mass splitting. Thinking of μ and ν as piles of sands, grains located at x can be sent at different places at the same time. Formally,

*Statistique, Analyse et Modélisation Multidisciplinaire (SAMM), Centre Pierre Mendès France 90, rue de Tolbiac 75634 Paris cedex 13, France.

†E-mail adress: xavier.bacon@etu.univ-paris1.fr

Kantorovitch's problem consists on minimizing a new total cost among all transference plans $\gamma \in \Pi(\mu,\nu)$, where $\Pi(\mu,\nu)$ is the set of couplings between μ and ν *i.e.* $\gamma \in \Pi(\mu,\nu)$ if for all A Borelian subset of X, $\gamma(A\times Y)=\mu(A)$ and for all B Borelian subset of Y, $\gamma(X\times B)=\nu(B)$. Kantorovitch's transportation problem is then

$$\mathcal{K}(\mu,\nu) := \inf\left\{ \iint_{X\times X} c(x,y)\,\mathrm{d}\gamma(x,y) : \gamma\in\Pi(\mu,\nu)\right\} \tag{2}$$

and for reasons that are discussed below, (2) is more accurate to extend the classical theory to vector-valued measures.

When c is the power of a distance, these two problems induce a metric on the set of probabilities, called here Wasserstein metric (see [5], [6], [7] or [8]). In the recent years, extensions of optimal transportation to more general objects have been proposed, such as multimarginal transportation ([9],[10]) or density functional theory ([11]). Notice also that optimal transportation of matricial and tensorial measures (see [12], [13]) or vector-valued densities in [14] have already been investigated.

In the present paper, we propose an extension to vector-valued measures. This one is deeply based on Kantorovich's formulation of *scalar* optimal transportation (section 2). Given two probabilities μ and ν and two decompositions of them (say) $\mu=\mu_1+\cdots+\mu_n$ and $\nu=\nu_1+\cdots+\nu_n$, more than a transportation between μ and ν, we are interested in a description of a transportation between these two decompositions. A naive strategy would be to study the n subproblems of classical optimal transportation between μ_i and ν_i for $i\in[\![1,n]\!]$, assuming that for all i, μ_i and ν_i share the same mass. If this new transportation problem leads to a metric, then the toplogy induced is the product one, due to the independance of each *phasis*. This problem has been explored in [15],[16] and more recently in [17]. To remove the independance of each phasis, we allow transformation similarly as explored in [18]. Introducing n^2 particular costs c_{ij} and n^2 particular transference plans γ_{ij} which describe the transport of a piece of μ_i into a piece of ν_j, we consider that the cost transportation to move $\mathrm{d}\mu_i(x)$ to $\mathrm{d}\nu_j(y)$ is $c_{ij}(x,y)\mathrm{d}\gamma_{ij}(x,y)$. Compatibility constraints are described by the set $\boldsymbol{\Pi}(\boldsymbol{\mu},\boldsymbol{\nu})$ where we ask that the n^2 transport plans γ_{ij} clear each μ_i and fill each ν_j. The new minimization problem is given by

$$\inf\left\{ \sum_{(i,j)\in[\![1,n]\!]^2} \iint_{X\times Y} c_{ij}(x,y)\,\mathrm{d}\gamma_{ij}(x,y), \boldsymbol{\gamma}\in\boldsymbol{\Pi}\left(\begin{bmatrix}\mu_1\\ \vdots\\ \mu_n\end{bmatrix},\begin{bmatrix}\nu_1\\ \vdots\\ \nu_n\end{bmatrix}\right)\right\}.$$

In section 2, we give an existence result for this problem as well as various examples. Then, following the shipper's problem interpretation of optimal transportation from L. Caffarelli (presented in [6]), we introduce a dual formulation in section 3 and prove strong duality theorem. As a consequence of the duality, optimality conditions for primal-dual optimizers are derived. Finally,

assuming that costs (c_{ij}) are all the same power of different distances, a metric on vector-valued measures is presented in section 5.

1.2 Notations

In this article, we differenciate vectorial objects from scalar ones by using bold type character like $\mathbf{\Pi}$ for the first one and non-bold type character like Π for the latter one.

- Given X a measurable space, $\mathcal{P}(X)$ stands for the set of probability measures on X and for all $n \in \mathbb{N}^*$, $\mathcal{M}^n(X)$ (resp. $\mathcal{M}^{n\times n}(X)$) refers to the set of vectorial measure on $(X, \mathcal{X})$ valued in $\mathbb{R}^n$ (resp. in $\mathbb{R}^{n\times n}$) meaning that each coordinate is a signed measure. $\mathcal{M}_+^n(X)$ (resp. $\mathcal{M}_+^{n\times n}(X)$) stands for the subset where each coordinate is positive measure. Recall that for T a measurable function between X and Y and m a positive measure on X, $T\#m$ stands for the push forward measure which is the measure on Y defined for all measurable set B by $T\#m(B) = m\left[T^{-1}(B)\right]$.

- Given $X_1 \times \cdots \times X_n$ a product space and $k \in [\![1, n]\!]$, π_k denotes the canonical projection on X_k *i.e*,

$$\begin{array}{rccc} \pi_k & : & X_1 \times \cdots \times X_n & \longrightarrow & X_k \\ & & (x_1, \cdots, x_n) & \mapsto & x_k \end{array}$$

  and for $l \in [\![1, n]\!]$ and $l > k$, $\pi_{k,l}$ denotes the canonical projection on $X_k \times X_l$ *i.e*,

$$\begin{array}{rccc} \pi_{k,l} & : & X_1 \times \cdots \times X_n & \longrightarrow & X_k \times X_l \\ & & (x_1, \cdots, x_n) & \mapsto & (x_k, x_l) \end{array}$$

- For A a borelian subset of $\mathbb{R}$, $\mathcal{L}_A$ stands for the Lebesgue measure on A. If $m, M \in \mathcal{M}_+(X)$ satisfy for all $A \in \mathcal{X}, m(A) \leqslant M(A)$, m is called a submeasure of M and this property will be written $m \leqslant M$. Note that being a submeasure of M implies the absolute continuity w.r.t. M.

- Given $(X, \mathcal{T})$ a topological space and (Y, d) a metric space, $C_b(X, Y)$ refers to the set of bounded continuous functions between $(X, \mathcal{T})$ and (Y, d).

- Given $(i, j) \in [\![1, n]\!]^2$, $\boldsymbol{E_{ij}}$ refers to the matrix $n \times n$ whose coordinates are all equal to 0 except (i, j) which is equal to 1.

- Given a set X and S a subset of X, ι_S denotes for the function equals to 0 on S and $+\infty$ on its complementary.

- The notation $\wedge$ will be used to denote the minimum of two reals, and $\vee$ for the maximum.

2 Kantorovitch's problem

2.1 Presentation

In the remainder of the paper, n will denote an element of $\mathbb{N}^*$.

Definition 2.1. Given $(X, \mathcal{X})$ a mesurable space, $\mathcal{P}^n(X)$ denotes the set of admissible distributions of n species defined by

$$\mathcal{P}^n(X) = \left\{ \mathbf{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} \in \mathcal{M}_+^n(X) : \sum_{i=1}^n m_i \in \mathcal{P}(X) \right\}.$$

It is straightforward that $\mathcal{P}^n(X)$ is a non-empty convex subset of $\mathcal{M}^n(X)$.

Inspired by Kantorovitch's formulation of optimal transportation, an extension of the notion of transference plan between two scalar measures is now proposed. For a well understanding of the next definition, let us make a short digression and present our model. Given $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ two measurable spaces and $\boldsymbol{\mu} \in \mathcal{P}^n(X), \boldsymbol{\nu} \in \mathcal{P}^n(Y)$ two distributions of n species, since the total amount of each specy is not equal transformations between species are allowed. Given $(i, j) \in [\![1, n]\!]^2$, the "transportation" (with "transformation" if $i \neq j$) of a piece of μ_i into a piece of ν_j is described by a transference plan $\gamma_{ij} \in \mathcal{M}_+(X \times Y)$. Constraints on $\boldsymbol{\gamma} = (\gamma_{ij})_{1 \leqslant i,j \leqslant n}$ are given by

$$\text{(Clear } \boldsymbol{\mu}) \ \forall i \in [\![1, n]\!], \forall A \in \mathcal{X}, \mu_i(A) = \sum_{k=1}^n \gamma_{ik}(A \times Y)$$

$$\text{(Fill } \boldsymbol{\nu}) \ \forall j \in [\![1, n]\!], \forall B \in \mathcal{Y}, \nu_j(B) = \sum_{k=1}^n \gamma_{kj}(X \times B),$$

or in other words, for all $(i, j) \in [\![1, n]\!]^2, \sum_{k=1}^n \gamma_{ik}$ has μ_i as first marginal and $\sum_{k=1}^n \gamma_{kj}$ has ν_j as second marginal. This naturally leads to the following definition.

Definition 2.2. Given $\boldsymbol{\mu} \in \mathcal{P}^n(X)$ and $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$, $\boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ denotes the set of transference plans between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ defined by

$$\boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \left\{ \begin{array}{l} \\ \boldsymbol{\gamma} \in \mathcal{M}_+^{n \times n}(X \times Y) : \forall (i, j) \in [\![1, n]\!]^2, \\ \\ \end{array} \begin{array}{r} \mu_i = \pi_1 \# \left(\sum_{k=1}^n \gamma_{ik} \right) \\ \\ \nu_j = \pi_2 \# \left(\sum_{k=1}^n \gamma_{kj} \right) \end{array} \right\}.$$

Remark 2.1. According to the Definition 2.2, every $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ induces a canonical transference plan (for $n = 1$, the two definitions of transference plan are the same) between $\sum_{i=1}^n \mu_i$ and $\sum_{j=1}^n \nu_j$ given by $\sum_{i,j=1}^n \gamma_{ij}$. However the converse is not true since given $\gamma \in \Pi\left(\sum_{i=1}^n \mu_i, \sum_{j=1}^n \nu_j\right)$ and $(x, y) \in X \times Y$,

there is still a choice to make: is the first specy sent into the first or the second one or both? And in what proportions? Let us give a short example to clarify this remark. Taking

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathcal{L}_{[-1,0]} \\ \mathcal{L}_{[-1,0]} \end{bmatrix}, \boldsymbol{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathcal{L}_{[0,1]} \\ \mathcal{L}_{[0,1]} \end{bmatrix}$$

and writting $\tau_1 : x \to x + 1$, it is known that $\gamma = (I, \tau_1) \# (\mu_1 + \mu_2)$ is a transference plan between $\mathcal{L}_{[-1,0]} = \mu_1 + \mu_2$ and $\mathcal{L}_{[0,1]} = \nu_1 + \nu_2$. Given a such γ, μ_1 can be sent towards ν_1, or towards ν_2. A mix is even possible and μ_1 can be sent towards $\mathbb{1}_{[0,\frac{1}{2}]} \, \mathrm{d}\nu_1 + \mathbb{1}_{[\frac{1}{2},1]} \, \mathrm{d}\nu_2$. In other words, the following matrix measures are transference plans,

$$\begin{bmatrix} (I, \tau_1)\#\mu_1 & 0 \\ 0 & (I, \tau_1)\#\mu_2 \end{bmatrix},$$
$$\begin{bmatrix} 0 & (I, \tau_1)\#\mu_1 \\ (I, \tau_1)\#\mu_2 & 0 \end{bmatrix},$$
$$\frac{1}{2} \begin{bmatrix} (I, \tau_1)\#\mu_1 & (I, \tau_1)\#\mu_1 \\ (I, \tau_1)\#\mu_2 & (I, \tau_1)\#\mu_2 \end{bmatrix}.$$

We also introduce matrix-valued cost $\mathbf{c}$ as a function from $X \times Y \to \mathcal{M}_n(\mathbb{R})$, integrable w.r.t. $\boldsymbol{\gamma}$ or positive measurable. The associated total cost is given by the following definition.

Definition 2.3. Given $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ and a cost matrix $\mathbf{c}$, $K(\boldsymbol{\gamma})$ denotes the total transportation cost according to $\boldsymbol{\gamma}$ defined by

$$K(\boldsymbol{\gamma}) = \sum_{(i,j) \in [\![1,n]\!]^2} \iint_{X \times Y} c_{ij}(x, y) \, \mathrm{d}\gamma_{ij}(x, y).$$

The Kantorovich's transportation problem between two distributions of n species $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ for $\mathbf{c}$ is given by

$$\inf \{K(\boldsymbol{\gamma}) : \boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})\} =: \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu}) \in [-\infty, +\infty] \tag{KP}$$

Example 2.1. Note that if $c_{ij} = c$ for all $(i, j) \in [\![1, n]\!]^2$ then (KP) shares the same value as the *scalar* optimal transportation between $\sum_{i=1}^n \mu_i$ and $\sum_{j=1}^n \nu_j$ for the cost c. The most simple example of non trivial matrix cost is given by the following one: let c be a *scalar* cost and κ be a real and define the following matrix cost:

$$\forall (x, y) \in X \times Y, \mathbf{c}(x, y) = \begin{bmatrix} c(x, y) & c(x, y) + \kappa \\ c(x, y) + \kappa & c(x, y) \end{bmatrix}.$$

In other words, a constant cost is requiered for any transformation. See the example 3.1 below for a study of this special cost.

Example 2.2. Let $\boldsymbol{p}, \boldsymbol{q} \in \Delta_n := \left\{\boldsymbol{x} \in \mathbb{R}_+^n, \sum_{i=1}^n x_i = 1\right\}$ and $\boldsymbol{x}, \boldsymbol{y} \in X^n$. Define

$$\boldsymbol{\mu} = \begin{bmatrix} p_1\delta_{x_1} \\ \vdots \\ p_n\delta_{x_n} \end{bmatrix}, \boldsymbol{\nu} = \begin{bmatrix} q_1\delta_{y_1} \\ \vdots \\ q_n\delta_{y_n} \end{bmatrix}.$$

Let us first notice that for all (i,j), since $\operatorname{supp}(\sum_{l=1}^n \gamma_{il}) \subseteq \operatorname{supp}(\mu_i) \times Y$ and $\operatorname{supp}(\sum_{l=1}^n \gamma_{lj}) \subseteq X \times \operatorname{supp}(\nu_j)$ then $\operatorname{supp}(\gamma_{ij}) \subseteq (x_i, y_j)$ and hence $\gamma_{ij} = t_{ij}\delta_{(x_i,y_j)}$ for some $t_{ij} \in [0,1]$. Constraints on $\boldsymbol{\gamma}$ give us that for all $(i,j) \in [\![1,n]\!]^2$, $\sum_{l=1}^n t_{il} = p_i$ and $\sum_{l=1}^n t_{lj} = q_j$ and finally

$$\boldsymbol{K}(\boldsymbol{\gamma}) = \sum_{(i,j)\in[\![1,n]\!]^2} t_{ij}c_{ij}(x_i, y_j),$$

(KP) becomes in that case

$$\inf\left\{\sum_{(i,j)\in[\![1,n]\!]^2} t_{ij}c_{ij}(x_i,y_j), \boldsymbol{t} \in \mathcal{M}_n(\mathbb{R}) : \sum_{l=1}^n t_{il} = p_i, \sum_{l=1}^n t_{lj} = q_j\right\}$$

which reduces to the discret optimal transportation.

2.2 Existence of a minimizer

Let X and Y be two Polish spaces. In this subsection, we prove an existence result for the problem (KP). Arguments used to establish it are the same as in scalar case (see [6] or [7] for instance). Let us first gather the main structural properties of problem (KP).

Lemma 2.1. *Given $\boldsymbol{\mu} \in \mathcal{P}^n(X)$, $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$ and $\boldsymbol{c}$ a cost matrix, following assertions are satisfied:*

[1] *$\boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is a non-empty convex subset of $\mathcal{M}^{n\times n}(X \times Y)$.*

[2] *$\boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is a weakly sequentially compact [1] subset of $\mathcal{M}^{n\times n}(X \times Y)$.*

[3] *If for all $(i,j) \in [\![1,n]\!]^2$, c_{ij} is bounded from below, then $K : \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu}) \to \mathbb{R} \cup \{+\infty\}$ is bounded from below.*

[4] *If for all $(i,j) \in [\![1,n]\!]^2$, c_{ij} is l.s.c. and bounded from below then $K : \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu}) \to \mathbb{R} \cup \{+\infty\}$ is weakly l.s.c. with respect to the tight convergence.*

Proof. [1] Convexity is clear and it is easy to check that $(\mu_i \otimes \nu_j)_{(i,j)\in[\![1,n]\!]^2} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$.
[2] Let $(\boldsymbol{\gamma^k})_{k\in\mathbb{N}} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})^{\mathbb{N}}$ and $(i,j) \in [\![1,n]\!]^2$. We claim that $(\gamma_{ij}^k)_{k\in\mathbb{N}}$ is tight. Indeed, let $\varepsilon \in \mathbb{R}_+^*$ and K_X (resp. K_Y) a compact of X (resp. Y) such that[2]

$$\forall i \in [\![1,n]\!], \mu_i(X\backslash K_X) \leqslant \varepsilon \text{ (resp. } \forall j \in [\![1,n]\!], \nu_j(Y\backslash K_Y) \leqslant \varepsilon). \tag{3}$$

[1] w.r.t. the test function space $C_b(X \times Y, \mathbb{R}^{n\times n})$.

[2] These two compacts exist: all μ_i are finite measure on Polish spaces then it exists K_X^i verifying these inequalities and then we just have to take union of them.

Let $k \in \mathbb{N}$, following inequalities are satisfied,

$$\begin{aligned}\gamma_{ij}^k\left[(X\times Y)\backslash(K_X\times K_Y)\right] &\leqslant \gamma_{ij}^k[(X\backslash K_X)\times Y)] + \gamma_{ij}^k[X\times(Y\backslash K_Y)]\\ &\leqslant \sum_{l=1}^n \gamma_{il}^k[(X\backslash K_X)\times Y)] + \sum_{\tilde{l}=1}^n \gamma_{\tilde{l}j}^k[X\times(Y\backslash K_Y)]\\ &= \mu_i(X\backslash K_X) + \nu_j(Y\backslash K_Y) \text{ since } \boldsymbol{\gamma}\in\boldsymbol{\Pi}(\boldsymbol{\mu},\boldsymbol{\nu})\\ &\leqslant 2\varepsilon.\end{aligned}$$

This proves the claim and thanks to Prokhorov theorem, there exists a non-negative finite measure on $X\times Y, \gamma_{ij}^\infty$ and a subsequence of $(\gamma_{ij}^k)_{k\in\mathbb{N}}$ (still written $(\gamma_{ij}^k)_{k\in\mathbb{N}}$) such as $(\gamma_{ij}^k)_{k\in\mathbb{N}}$ tightly converges towards γ_{ij}^∞. In order to conclude, we only have to check that $\boldsymbol{\gamma}^\infty \in \Gamma(\boldsymbol{\mu},\boldsymbol{\nu})$ Let $\phi \in C_b(X\times Y,\mathbb{R})$ and notice that for all $i\in[\![1,n]\!]$ and $k\in\mathbb{N}$,

$$\int_X \phi(x)\,\mathrm{d}\mu_i(x) = \sum_{l=1}^n \iint_{X\times Y} \phi(x)\,\mathrm{d}\gamma_{il}^k(x,y) \to \sum_{l=1}^n \iint_{X\times Y} \phi(x)\,\mathrm{d}\gamma_{il}^\infty(x,y)$$

.

[3] Straightforward.

[4] Let $(\boldsymbol{\gamma}^k)_{k\in\mathbb{N}} \in \Pi(\boldsymbol{\mu},\boldsymbol{\nu})^{\mathbb{N}}$ and $\boldsymbol{\gamma}^\infty \in \Pi(\boldsymbol{\mu},\boldsymbol{\nu})$ such that $(\boldsymbol{\gamma^k})_{k\in\mathbb{N}}$ tightly converges towards $\boldsymbol{\gamma}^\infty$ in that for all $(i,j)\in[\![1,n]\!]^2, (\gamma_{ij}^k)_{k\in\mathbb{N}}$ weakly converges in duality with C_b towards γ_{ij}^∞. Then, by lower semi-continuity of $\gamma_{ij} \mapsto < \gamma_{ij}, c_{ij} >$ (see [7], Lemma 1.6), for all $(i,j)\in[\![1,\cdots,n]\!]^2$,

$$\iint_{X\times Y} c_{ij}(x,y)\,\mathrm{d}\gamma_{ij}^\infty(x,y) \leqslant \liminf_{k\to\infty} \iint_{X\times Y} c_{ij}(x,y)\,\mathrm{d}\gamma_{ij}^k(x,y)$$

and since sum of lim inf is less or equal to lim inf of sum, it ends the proof. □

With these facts in hand, our main result easily follows.

Theorem 2.1. *Given* $\mathbf{c}$ *a cost matrix such as for all* $(i,j)\in[\![1,n]\!]^2, c_{ij}$ *is bounded from below and l.s.c., it exists* $\boldsymbol{\gamma}\in\boldsymbol{\Pi}(\boldsymbol{\mu},\boldsymbol{\nu})$ *such as* $K(\boldsymbol{\gamma}) = \mathcal{K}(\boldsymbol{\mu},\boldsymbol{\nu})$.

Proof. This proof follows the classical direct method of calculus of variations. Let $(\boldsymbol{\gamma}^k)_{k\in\mathbb{N}}$ be a minimizing sequence for the problem (KP) *i.e*

$$\forall k\in\mathbb{N}, K(\boldsymbol{\gamma}^k) \leqslant \mathcal{K}(\boldsymbol{\mu},\boldsymbol{\nu}) + \frac{1}{k}.$$

Compactness of $\boldsymbol{\Pi}(\boldsymbol{\mu},\boldsymbol{\nu})$ according to Lemma 2.1 implies that $(\boldsymbol{\gamma}^k)_{k\in\mathbb{N}}$ can be assumed to converge towards (say) $\boldsymbol{\gamma}^\infty$. Lower semi-continuity implies that

$$K(\boldsymbol{\gamma}^\infty) \leqslant \liminf_{k\mapsto\infty} K(\boldsymbol{\gamma}^k) \leqslant \mathcal{K}(\boldsymbol{\mu},\boldsymbol{\nu}),$$

and then $\boldsymbol{\gamma}^\infty$ is a minimum. □

3 Duality

3.1 Presentation

In this section, we look for a dual formulation of (KP). In order to find it, consider the following situation[3]: mines full of different metals (n kinds) and refineries (n kinds) are distributed in space. For each kind of metal corresponds a kind of refinery, for instance a kind refinery for iron, a kind of refinery for gold etc. On the one hand we want to minimize the travel cost *i.e.* minimize the associated Kantorovich's problem, on the other hand a character suggests to supervise the travelling operation for us and propose that contract: for each ton of metal i located in x, its price will be $\varphi_i(x)$ to extract it and for each ton of metal j located in y its price will be $\psi_j(y)$ to drop it off. To guarantee our interrest, its contraints will be that for all (i,j) and (x,y), $\varphi_i(x)+\psi_j(y) \leqslant c_{ij}(x,y)$. All these considerations suggest to give following definitions.

Definition 3.1. Given $\mathbf{c}$ a cost matrix, $\boldsymbol{\Delta}(\mathbf{c})$ denotes the set of potential couples for cost $\boldsymbol{c}$ defined by

$$\boldsymbol{\Delta}(\mathbf{c}) = \left\{ \begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} = \begin{bmatrix} \varphi_1 & \cdots & \varphi_n \\ \psi_1 & \cdots & \psi_n \end{bmatrix}, \forall (i,j) \in [\![1,n]\!]^2, \quad \begin{array}{c} \varphi_i \in C_b(X) \\ \psi_j \in C_b(Y) \\ \varphi_i \oplus \psi_j \leqslant c_{ij} \end{array} \right\}$$

and if there is no ambiguity on $\mathbf{c}$, we will write $\boldsymbol{\Delta}$ instead of $\boldsymbol{\Delta}(\mathbf{c})$.

Definition 3.2. Given $\boldsymbol{\mu} \in \mathcal{P}^n(X), \boldsymbol{\nu} \in \mathcal{P}^n(Y)$, $\mathbf{c}$ a cost matrix and $\begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} = \begin{bmatrix} \varphi_1 & \cdots & \varphi_n \\ \psi_1 & \cdots & \psi_n \end{bmatrix} \in \Delta(\mathbf{c})$, $D(\boldsymbol{\varphi}, \boldsymbol{\psi})$ denotes the dual cost of $\begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix}$ defined by

$$D(\boldsymbol{\varphi}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \int_X \varphi_i(x) \, \mathrm{d}\mu_i(x) + \sum_{j=1}^{n} \int_Y \psi_j(y) \, \mathrm{d}\nu_j(y). \tag{4}$$

Finally, the dual transportation problem is given $\boldsymbol{\mu} \in \mathcal{P}^n(X)$, $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$ and a cost matrix $\mathbf{c}$,

$$\sup \left\{ D(\boldsymbol{\varphi}, \boldsymbol{\psi}) : \begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \Delta(\boldsymbol{c}) \right\} =: \mathcal{D}(\mu, \nu) \in [-\infty, +\infty] \tag{DP}$$

We establish now a weak duality result.

Proposition 3.1. *Given $\boldsymbol{\mu} \in \mathcal{P}^n(X)$, $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$, a cost matrix $\mathbf{c}, \boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $\begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \boldsymbol{\Delta}(\boldsymbol{c})$, the following inequality is satisfied,*

$$\boldsymbol{D}(\boldsymbol{\varphi}, \boldsymbol{\psi}) \leqslant \boldsymbol{K}(\boldsymbol{\gamma}).$$

[3] This interpretation is due to L. Caffareli in scalar case, according to [6].

Proof. Let $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $\begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \boldsymbol{\Delta}(\boldsymbol{c})$. Compute:

$$\begin{aligned} \boldsymbol{D}(\boldsymbol{\varphi}, \boldsymbol{\psi}) &= \sum_{i=1}^{n} \int_X \varphi_i \, \mathrm{d}\mu_i + \sum_{j=1}^{n} \int_Y \psi_j \, \mathrm{d}\nu_j \\ &= \sum_{i=1}^{n} \iint_{X \times Y} \varphi_i \, \mathrm{d} \left(\sum_{j=1}^{n} \gamma_{ij} \right) + \sum_{j=1}^{n} \iint_{X \times Y} \psi_j \, \mathrm{d} \left(\sum_{i=1}^{n} \gamma_{ij} \right). \end{aligned}$$

The last equality coming from the fact that $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$. And then,

$$\begin{aligned} \boldsymbol{D}(\boldsymbol{\varphi}, \boldsymbol{\psi}) &\leqslant \sum_{(i,j) \in [\![1,n]\!]^2} \iint_{X \times Y} c_{ij} \, \mathrm{d}\gamma_{ij} \text{ since } \begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \Delta(\boldsymbol{c}) \\ &= \boldsymbol{K}(\boldsymbol{\gamma}). \end{aligned}$$

That concludes the proof. □

3.2 An extension of c-transformation

In order to prove that (DP) is attained, at least in compact case, we propose an extension of the classical c-transform (see the recall below). First, we make a short digression about modulus of continuity.

Definition 3.3. Given (X, d) a metric space and $f : X \to \mathbb{R}$, a uniform modulus of continuity for f according to d is a function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ such that the following conditions are satisfied:

[1] $\lim_{t \to 0^+} \omega(t) = 0$

[2] $\forall (x, x') \in X^2 : |f(x) - f(x')| \leqslant \omega \left[d(x, x') \right]$.

Lemma 3.1. *If f admits a uniform modulus of continuity ω_f and g admits a uniform modulus of continuity ω_g then $\omega_f + \omega_g$ is a uniform modulus of continuity for* $\min(f, g)$.

Proof. Let $(x, x') \in X^2$, we have

$$\begin{aligned} &|\min(f, g)(x) - \min(f, g)(x')| \\ &\quad \leqslant \frac{|f(x) - f(x')| + |g(x) - g(x')|}{2} + \frac{||f(x') - g(x')| - |f(x) - g(x)||}{2} \\ &\quad \leqslant \frac{\omega_f[d(x, x')] + \omega_g[d(x, x')]}{2} + \frac{|f(x') - f(x) + g(x) - g(x')|}{2} \\ &\quad \leqslant \omega_f[d(x, x')] + \omega_g[d(x, x')]. \end{aligned}$$

This proves the lemma. □

Recall that when f is a function between X (resp. Y) and $\mathbb{R} \cup \{-\infty\}$ and c a cost function, we can define its c-transform f^c (resp. $\overline{c}$-transform) by:

$$\begin{array}{rccl} f^c & : & Y & \to & \mathbb{R} \cup \{-\infty, +\infty\} \\ & & y & \mapsto & \inf\{c(x,y) - f(x) : x \in X\} \end{array}$$

$$\left(\text{resp. } \begin{array}{rccl} f^c & : & X & \to & \mathbb{R} \cup \{-\infty, +\infty\} \\ & & x & \mapsto & \inf\{c(x,y) - f(y) : y \in Y\} \end{array} \right)$$

We introduce a new transformation and to motivate it just remark than in our case, we have $2n$ potentials and n^2 inequalities in the dual formulation. A naive idea would be to first subsitute φ_1 by $\psi_1^{c_{11}}$ but there is no guarantee that our new couple of potentials $\begin{bmatrix} \psi_1^{c_{11}} & \varphi_2 & \cdots & \varphi_n \\ \psi_1 & \cdots & \cdots & \psi_n \end{bmatrix}$ will still be in $\boldsymbol{\Delta}(\boldsymbol{c})$. The following definition answers this problem.

Definition 3.4. Given $\boldsymbol{f} = (f_1, \cdots, f_n) : X \to (\mathbb{R} \cup \{-\infty\})^n$ and $\boldsymbol{c} = (c_1, \cdots, c_n) : X \times Y \to (\mathbb{R} \cup \{+\infty\})^n$, $\boldsymbol{f^c}$ (resp. $\boldsymbol{f^{\bar{c}}}$) denotes the $\boldsymbol{c}$-transform of f (resp. $\bar{\boldsymbol{c}}$-transform of f) defined by

$$\forall y \in Y : \boldsymbol{f^c}(y) = \min\left(f_1^{c_1}(y), \cdots, f_n^{c_n}(y)\right)$$
$$\left(\text{resp. } \forall x \in X : \boldsymbol{f^{\bar{c}}}(x) = \min\left(f_1^{\bar{c_1}}(x), \cdots, f_n^{\bar{c_n}}(x)\right)\right)$$

All benefits of this transformation is contained in the next proposition.

Proposition 3.2. *Let $\boldsymbol{f} = (f_1, \cdots, f_n) : X \to (\mathbb{R} \cup \{-\infty\})^n$ and $\boldsymbol{c} = (c_1, \cdots, c_n) : X \times Y \to (\mathbb{R} \cup \{+\infty\})^n$, then*

***[1]** Following inequalities are satisfied,*

$$\forall i \in [\![1, n]\!], f_i \oplus \boldsymbol{f^c} \leqslant c_i \tag{5}$$
$$\forall j \in [\![1, n]\!], \boldsymbol{f^{\bar{c}}} \oplus f_j \leqslant c_j \tag{6}$$

***[2]** If $h : Y \to \mathbb{R} \cup \{-\infty\}$ is such that for all $i \in [\![1, n]\!], f_i \oplus h \leqslant c_i$ then $h \leqslant \boldsymbol{f^c}$. If $h : X \to \mathbb{R} \cup \{-\infty\}$ is such that for all $j \in [\![1, n]\!], h \oplus g_j \leqslant c_j$ then $h \leqslant \boldsymbol{f^{\bar{c}}}$.*

Proof. [1] Let $(i, j) \in [\![1, n]\!]^2$ and $(x, y) \in X \times Y$. Since $f_i(x) + f_j^{c_j}(y) \leqslant c_i(x, y)$ and $\boldsymbol{f^c} \leqslant f_i^{c_i}$ the first inequality is deduced and note that the second inequality can be proved following the same way.
[2] If such a function exists, we deduce from $f_i \oplus h \leqslant c_i$ that for all $(x, y) \in X \times Y, h(y) \leqslant c_i(x, y) - f_i(x)$, then take infimum with respect to x and arbitrary on i concludes for the first inequality. The same proof also works for the second inequality. □

We will show next that this process is a natural way to improve the dual cost while staying in the constraint $\Delta(\boldsymbol{c})$, at least in compact case and continuous costs. Moreover, it provides a common uniform modulus of continuity for all the potentials.

Lemma 3.2. *Let X, Y two compact metric spaces, $\boldsymbol{c}$ a continuous cost matrix and $(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Delta}(\boldsymbol{c})$. It exists $(\boldsymbol{\varphi}, \underline{\boldsymbol{\psi}}) \in \boldsymbol{\Delta}(\boldsymbol{c})$ such that*

[1] *$D(\boldsymbol{\varphi}, \boldsymbol{\psi}) \leqslant D(\boldsymbol{\varphi}, \underline{\boldsymbol{\psi}})$.*

[2] *$\underline{\varphi_1}, \cdots, \underline{\varphi_n}, \underline{\psi_1}, \cdots, \underline{\psi_{n-1}}$ and $\underline{\psi_n}$ admit a common uniform modulus of continuity which depends only on $\boldsymbol{c}$.*

Proof. First, make the following substitutions:

$$\forall j \in [\![1, n]\!] : \psi_j \leftarrow \varphi^{(c_{1j}, \cdots, c_{nj})} := \underline{\psi_j},$$

then, thanks to Proposition 3.2, $(\boldsymbol{\varphi}, \underline{\boldsymbol{\psi}}) \in \boldsymbol{\Delta}(\boldsymbol{c})$ and $D(\boldsymbol{\varphi}, \boldsymbol{\psi}) \leqslant D(\boldsymbol{\varphi}, \underline{\boldsymbol{\psi}})$. Denoting $\omega_{c_{ij}}$ a uniform modulus of continuity of c_{ij} for $(i, j) \in [\![1, n]\!]^2$, $\omega_{c_{ij}}$ is also a uniform modulus of continuity of $\varphi_i^{c_{ij}}$ according to [7] (Box. 1.8). Thanks to Lemma 3.1, we conclude that $\omega_{\underline{\psi_j}} = \omega_{c_{1j}} + \cdots + \omega_{c_{nj}}$ is a uniform modulus of continuity of $\underline{\psi_j}$. Then, make the following substitutions:

$$\forall i \in [\![1, n]\!] : \varphi_i \leftarrow \underline{\psi}^{\overline{(c_{i1}, \cdots, c_{in})}} := \underline{\varphi_i}$$

and of course the new couple of potentials is still in $\boldsymbol{\Delta}(c)$ and the dual cost is increased. To conclude, we just have to check that $\sum_{1 \leqslant i,j \leqslant n} \omega_{c_{ij}}$ is a common uniform modulus of continuity for $(\underline{\boldsymbol{\varphi}}, \underline{\boldsymbol{\psi}})$, which is clear. □

Example 3.1. Coming back to the example 2.1, let us compute this new $\boldsymbol{c}$-transform to reduce the problem. Fix κ to be strictly non-negative and assume that $X = Y$ and $\boldsymbol{c}$ is symetric (then, $\boldsymbol{c}$-transform is equivalent to $\overline{\boldsymbol{c}}$-transform). Constraints of (DP) are given by the following system:

$$\begin{cases} \varphi_1(x) + \psi_1(y) \leqslant c(x, y) \\ \varphi_1(x) + \psi_2(y) \leqslant c(x, y) + \kappa \\ \varphi_2(x) + \psi_1(y) \leqslant c(x, y) + \kappa \\ \varphi_2(x) + \psi_2(y) \leqslant c(x, y) \end{cases}$$

First step: it is easy to check that:

$$\begin{aligned} (f_1, f_2)^{c, c+\kappa} &= [f_1 \wedge (f_2 - \kappa)]^c \\ (f_1, f_2)^{c+\kappa, c} &= [(f_1 - \kappa) \wedge f_2]^c, \end{aligned}$$

then make the following substitutions:

$$\begin{aligned} \psi_1 \leftarrow (\varphi_1, \varphi_2)^{c, c+\kappa} &= [\varphi_1 \wedge (\varphi_2 - \kappa)]^c =: \tilde{\psi}_1 \\ \psi_2 \leftarrow (\varphi_1, \varphi_2)^{c+\kappa, c} &= [(\varphi_1 - \kappa) \wedge \varphi_2]^c =: \tilde{\psi}_2 \end{aligned}$$

Second step: following the proof below, we make the following substitutions:

$$\begin{aligned}
\varphi_1 \leftarrow (\tilde{\psi}_1, \tilde{\psi}_2)^{c, c+\kappa} &= \left[\tilde{\psi}_1 \wedge (\tilde{\psi}_2 - \kappa)\right]^c \\
&= \tilde{\psi}_1^{\,c} \vee (\tilde{\psi}_2 - \kappa)^c \text{ since } (\sup_\alpha f_\alpha)^c = \inf_\alpha f_\alpha^c \\
&= \left[\varphi_1 \wedge (\varphi_2 - \kappa)\right]^{cc} \vee \left([(\varphi_1 - \kappa) \wedge \varphi_2] - \kappa\right)^{cc} \\
&= \left[\varphi_1 \wedge (\varphi_2 - \kappa)\right]^{cc} \vee \left[(\varphi_1 - 2\kappa) \wedge (\varphi_2 - \kappa)\right]^{cc} \\
&= \left[\varphi_1 \wedge (\varphi_2 - \kappa)\right]^{cc} \text{ since if } f \leqslant g \text{ then } g^c \leqslant f^c \\
&= \tilde{\psi}_1^{\,c} \\
\varphi_2 \leftarrow (\tilde{\psi}_1, \tilde{\psi}_2)^{c+\kappa, c} &= \tilde{\psi}_2^{\,c} \text{ for the same reasons.}
\end{aligned}$$

When $c = d$ is a distance, according to [7] (Proposition 3.1):

$$\begin{cases}
\tilde{\psi}_1(y) - \tilde{\psi}_1(x) \leqslant d(x, y) \\
\tilde{\psi}_2(y) - \tilde{\psi}_1(x) \leqslant d(x, y) + \kappa \\
\tilde{\psi}_1(y) - \tilde{\psi}_2(x) \leqslant d(x, y) + \kappa \\
\tilde{\psi}_2(y) - \tilde{\psi}_2(x) \leqslant d(x, y),
\end{cases}$$

which is equivalent to the following system, thanks to the symmetry of d:

$$\begin{cases}
|\tilde{\psi}_1(x) - \tilde{\psi}_1(y)| \leqslant d(x, y) \\
|\tilde{\psi}_1(x) - \tilde{\psi}_2(y)| \leqslant d(x, y) + \kappa \\
|\tilde{\psi}_2(x) - \tilde{\psi}_2(y)| \leqslant d(x, y)
\end{cases}$$

i.e. $(\tilde{\psi}_1, \tilde{\psi}_2)$ are solution to the system below if and only if they are 1-Lipschitz w.r.t. to d and satisfy $\|\tilde{\psi}_1 - \tilde{\psi}_2\|_\infty \leqslant \kappa$.

3.3 Existence of a maximizer

Theorem 3.1. *Given X et Y two compact metric spaces, $\boldsymbol{\mu} \in \mathcal{P}^n(X)$, $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$ and $\boldsymbol{c}$ a continuous cost matrix, there exists $(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Delta}(\boldsymbol{c})$ such as $\mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\nu}) = D(\boldsymbol{\varphi}, \boldsymbol{\psi})$.*

Proof. The constraint set is non-empty since $\boldsymbol{c}$ is bounded by below (continuous on compact). Let:

$$\begin{bmatrix} \boldsymbol{\varphi^k} \\ \boldsymbol{\psi^k} \end{bmatrix}_{k \in \mathbb{N}} = \begin{bmatrix} \varphi_1^k & \cdots & \varphi_n^k \\ \psi_1^k & \cdots & \psi_n^k \end{bmatrix}_{k \in \mathbb{N}}$$

be a maximizing sequence for (DP). According to Lemma 3.2, we may assume that our $2n$ sequences share a common uniform modulus of continuity. We now prove that the sequence is uniformly bounded with respect to n. Indeed, setting for all $k \in \mathbb{N}$:

$$m_k := \min \left[\inf_{x \in X} \varphi_1^k(x), \cdots, \inf_{x \in X} \varphi_n^k(x) \right],$$

and since m_k is finite, we can substitute:

$$\forall i \in [\![1, \cdots, n]\!] : \varphi_i^k \leftarrow \varphi_i^k - m_k \text{ still written } \varphi_i^k$$
$$\forall j \in [\![1, \cdots, n]\!] : \psi_j^k \leftarrow \psi_j^k + m_k \text{ still written } \psi_j^k,$$

and these new potentials are still admissible, have the same dual cost and for all $i \in [\![1, n]\!], k \in \mathbb{N}, \varphi_i^k \geqslant 0$. Therefore we have:

$$\forall i \in [\![1, n]\!], k \in \mathbb{N} : \varphi_i^k \leqslant \omega\left[diam(X)\right],$$

which concludes the case of $\boldsymbol{\varphi}$. Next, let us make new following substitutions:

$$\forall j \in [\![1, n]\!] : \psi_j \leftarrow \varphi^{\overline{(c_{1j}, \cdots, c_{nj})}} \text{ still written } \psi_j.$$

We have for all $y \in Y, j \in [\![1, n]\!]$ and $k \in \mathbb{N}$,

$$\begin{aligned}
&\min\left(c_{1j}, \cdots c_{nj}\right) - \omega\left[diam(X)\right] \\
&\qquad \leqslant \min\left[\inf_{x\in X} c_{1j}(x, y) - \varphi_1^k(x), \cdots, \inf_{x\in X} c_{nj}(x, y) - \varphi_n^k(x)\right] := \psi_j^k(y) \\
&\text{and } \psi_j^k(y) := \min\left[\inf_{x\in X} c_{1j}(x, y) - \varphi_1^k(x), \cdots, \inf_{x\in X} c_{nj}(x, y) - \varphi_n^k(x)\right] \\
&\qquad \leqslant \max\left(c_{1j}, \cdots, c_{nj}\right),
\end{aligned}$$

which leads to the conclusion on $\boldsymbol{\psi}$. Finally, the Ascoli-Arzelà theorem applied to each sequence provides the existence of a continuous couple

$$\begin{bmatrix} \boldsymbol{\varphi}^\infty \\ \boldsymbol{\psi}^\infty \end{bmatrix} = \begin{bmatrix} \varphi_1^\infty & \cdots & \varphi_n^\infty \\ \psi_1^\infty & \cdots & \psi_n^\infty \end{bmatrix}$$

which belong to $\Delta(\boldsymbol{c})$ thanks to pointwise convergence and $D(\boldsymbol{\varphi}^\infty, \boldsymbol{\psi}^\infty) = \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\nu})$ thanks to uniform convergence on finite measure sets. □

3.4 Strong duality

We establish a strong duality result. The proof follows the one of strong duality theorem for scalar optimal transportation proposed by C. Jimenez (see [7]).

Definition 3.5. Given $\boldsymbol{\mu} \in \mathcal{P}^n(X)$, $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$ and $\boldsymbol{c}$ a cost matrix, we denote by H the value function of the perturbated dual problem, *i.e.*

$$\forall \boldsymbol{\varepsilon} \in C(X \times Y, \mathbb{R}^{n\times n}), H(\boldsymbol{\varepsilon}) = \sup\left\{D(\boldsymbol{\varphi}, \boldsymbol{\psi}) : \begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \Delta(\boldsymbol{c} - \boldsymbol{\varepsilon})\right\}$$

Lemma 3.3. *Let X and Y two metric compact spaces. H satisfy the following properties:*

[1] *H is concave.*

[2] *Suppose that $\boldsymbol{c}$ is continuous, then H is u.s.c. with respect to the uniform norm.*

Proof. [1] Let $t \in [0,1]$, $\boldsymbol{\varepsilon^0} \in C(X \times Y, \mathbb{R}^{n\times n})$ (resp. $\boldsymbol{\varepsilon^1} \in C(X \times Y, \mathbb{R}^{n\times n})$) and let $(\boldsymbol{\varphi^0}, \boldsymbol{\psi^0})$ (resp. $(\boldsymbol{\varphi^1}, \boldsymbol{\psi^1})$) be optimal in (DP) associated to $\boldsymbol{c} - \boldsymbol{\varepsilon^0}$ (resp. to $\boldsymbol{c} - \boldsymbol{\varepsilon^1}$). Note that they exist thanks to the existence result below. Define $\boldsymbol{\varepsilon_t} = (1-t)\boldsymbol{\varepsilon^0} + t\boldsymbol{\varepsilon^1}, \boldsymbol{\varphi_t} = (1-t)\boldsymbol{\varphi^0} + t\boldsymbol{\varphi^1}, \boldsymbol{\psi_t} = (1-t)\boldsymbol{\psi^0} + t\boldsymbol{\psi^1}$. Therefore $(\boldsymbol{\varphi_t}, \boldsymbol{\psi_t})$ is admissible for the dual problem associated to $\boldsymbol{c} - \boldsymbol{\varepsilon_t}$ and then by definition of H we have

$$H(\boldsymbol{\varepsilon_t}) \geqslant D(\boldsymbol{\varphi_t}, \boldsymbol{\psi_t}) = (1-t)H(\boldsymbol{\varepsilon^0}) + tH(\boldsymbol{\varepsilon^1})$$

And the conclusion follows.
[2] Let $(\boldsymbol{\varepsilon^k})_{k\in\mathbb{N}} \in C(X \times Y, \mathbb{R}^{n\times n})^{\mathbb{N}}$ and $\boldsymbol{\varepsilon^\infty} \in C(X \times Y, \mathbb{R}^{n\times n})$ such that for all $(i,j) \in [\![1,n]\!]^2, \varepsilon_{ij}^k \xrightarrow{\|\cdot\|_\infty} \varepsilon_{ij}^\infty$. Let $(\boldsymbol{\varepsilon^k})_{k\in\mathbb{N}}$ a subsequence $(\boldsymbol{\varepsilon^{k_l}})_{l\in\mathbb{N}}$ satisfying for all $(i,j) \in [\![1,n]\!]^2, \limsup_k H(\varepsilon_{ij}^k) = \lim_l H(\varepsilon_{ij}^{k_l})$. Ascoli-Arzelà theorem ensures that for all $(i,j) \in [\![1,n]\!]^2, (\varepsilon_{ij}^{k_l})_{l\in\mathbb{N}}$ are equicontinuous and uniformly bounded with respect to l therefore chose the corresponding optimal potentials $(\boldsymbol{\varphi^{k_l}}, \boldsymbol{\psi^{k_l}})$ equicontinuous and uniformly bounded in l, thanks to the Ascoli-Arzelà theorem again, there is a uniform convergence towards (say) $(\boldsymbol{\varphi^\infty}, \boldsymbol{\psi^\infty})$ up to an extraction, therefore thanks to pointwise convergence there is for all $(i,j) \in [\![1,n]\!]^2, \varphi_i^\infty \oplus \psi_j^\infty \leqslant c_{ij} - \varepsilon_{ij}$ and so:

$$H(\boldsymbol{\varepsilon}) \geqslant D(\boldsymbol{\varphi^\infty}, \boldsymbol{\psi^\infty}) = \lim_l H(\boldsymbol{\varepsilon^{k_l}}) = \limsup_k H(\boldsymbol{\varepsilon^k}).$$

This concludes the proof. □

Finally, the strong duality theorem follows.

Theorem 3.2. *Suppose that X and Y are both metric compact spaces and that $\boldsymbol{c}$ is continuous, then for all $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{P}^n(X) \times \mathcal{P}^n(Y), \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\nu})$.*

Proof. Let $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{P}^n(X) \times \mathcal{P}^n(Y)$, since $(-H)$ is convex and l.s.c. and according to the Fenchel-Moreau theorem, we have:

$$\begin{aligned}
\mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\nu}) &= H(\mathbf{0}) \\
&= -[-H(\mathbf{0})] \\
&= -[-H]^{**}(\mathbf{0}) \\
&= -\sup_{\boldsymbol{\gamma} \in \mathcal{M}^{n\times n}(X\times Y)} < \mathbf{0}, \boldsymbol{\gamma} > -[-H]^*(\boldsymbol{\gamma}) \\
&= \inf_{\boldsymbol{\gamma} \in \mathcal{M}^{n\times n}(X\times Y)} [-H]^*(\boldsymbol{\gamma}).
\end{aligned}$$

Next, we compute for all $\boldsymbol{\gamma} \in \mathcal{M}^{n\times n}(X \times Y)$,

$$\begin{aligned}
[-H]^*(\boldsymbol{\gamma}) &= \sup_{\boldsymbol{\varepsilon}} \left(\sum_{1\leqslant i,j\leqslant n} \iint_{X\times Y} \varepsilon_{ij} \, \mathrm{d}\gamma_{ij} + \sup_{\boldsymbol{\varphi}\oplus\boldsymbol{\psi} \leqslant \boldsymbol{c}-\boldsymbol{\varepsilon}} \sum_{i=1}^n \int_X \varphi_i \, \mathrm{d}\mu_i + \sum_{j=1}^n \int_Y \psi_j \, \mathrm{d}\nu_j \right) \\
&= \sup_{\boldsymbol{\varepsilon}} \left(\sup_{\boldsymbol{\varphi}\oplus\boldsymbol{\psi} \leqslant \boldsymbol{c}-\boldsymbol{\varepsilon}} \left[\sum_{1\leqslant i,j\leqslant n} \iint_{X\times Y} \varepsilon_{ij} \, \mathrm{d}\gamma_{ij} + \sum_{i=1}^n \int_X \varphi_i \, \mathrm{d}\mu_i + \sum_{j=1}^n \int_Y \psi_j \, \mathrm{d}\nu_j \right] \right).
\end{aligned}$$

If it exists (i_0, j_0) such as $\gamma_{i_0 j_0} \notin \mathcal{M}_+(X \times Y)$ and $\varepsilon^0_{i_0 j_0}$ such as $\iint_{X\times Y} \varepsilon^0_{i_0 j_0} \mathrm{d}\gamma_{i_0 j_0} > 0$, then take $(\varepsilon^k_{i_0 j_0})_{k\in\mathbb{N}*}$ such as $\varepsilon^k_{i_0 j_0} = c_{i_0 j_0} + k\varepsilon^0_{i_0 j_0}$ for $k \in \mathbb{N}^*$ and take $\varphi_{i_0} = 0$ and $\psi_{j_0} = 0$. Then putting all the other potentials equals at the value 0 and find (ε_{ij}) such that all the contraints are still satisfied ($\mathbf{c}$ is bounded), we get $[-H]^*(\boldsymbol{\gamma}) = +\infty$ if $\boldsymbol{\gamma} \notin \mathcal{M}^{n\times n}_+(X \times Y)$. Now, suppose that $\boldsymbol{\gamma} \in \mathcal{M}^{n\times n}_+(X \times Y)$, when $(\boldsymbol{\varphi}, \boldsymbol{\psi})$ are fixed, we are interested in taking the largest ε_{ij} possible for every $(i, j) \in [\![1, n]\!]^2$, that is $\varepsilon_{ij} = c_{ij} - \varphi_i - \psi_j$ and we get

$$
\begin{aligned}
[-H]^*(\boldsymbol{\gamma}) = \sup_{\boldsymbol{\varphi}\oplus\boldsymbol{\psi}\leqslant \boldsymbol{C}-\boldsymbol{\varepsilon}} & \sum_{1\leqslant i,j\leqslant n} \iint_{X\times Y} c_{ij} - \varphi_i - \psi_j \,\mathrm{d}\gamma_{ij} \\
& + \sum_{i=1}^n \int_X \varphi_i \,\mathrm{d}\mu_i + \sum_{j=1}^n \int_Y \psi_j \,\mathrm{d}\nu_j \\
= \sup_{(\boldsymbol{\varphi},\boldsymbol{\psi})} K(\boldsymbol{\gamma}) & + \left(\sum_{i=1}^n \int_X \varphi_i \,\mathrm{d}\mu_i - \sum_{1\leqslant i,j\leqslant n} \iint_{X\times Y} \varphi_i \,\mathrm{d}\gamma_{ij} \right) \\
& + \left(\sum_{j=1}^n \int_Y \psi_j \,\mathrm{d}\nu_j - \sum_{1\leqslant i,j\leqslant n} \iint_{X\times Y} \psi_j \,\mathrm{d}\gamma_{ij} \right) \\
= \sup_{(\boldsymbol{\varphi},\boldsymbol{\psi})} K(\boldsymbol{\gamma}) & + \sum_{i=1}^n \left(\int_X \varphi_i \,\mathrm{d}\mu_i - \iint_{X\times Y} \varphi_i \,\mathrm{d} \sum_{j=1}^n \gamma_{ij} \right) \\
& + \sum_{j=1}^n \left(\int_Y \psi_j \,\mathrm{d}\nu_j - \iint_{X\times Y} \psi_j \,\mathrm{d} \sum_{i=1}^n \gamma_{ij} \right) \\
& = \iota_{\Pi(\boldsymbol{\mu},\boldsymbol{\nu})}(\boldsymbol{\gamma}) \text{ according to [7], Lemma 1.45.}
\end{aligned}
$$

This ends the proof. □

4 Optimality conditions

In this subsection, X and Y are two metric compact spaces. As a direct consequence of Theorem 3.2, we deduce optimality contraints linking (KP) and (DP).

Proposition 4.1. *Given $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Delta}(\boldsymbol{c})$, the following assertions are equivalent:*

[1] *$\boldsymbol{\gamma}$ is optimal in (KP) and $(\boldsymbol{\varphi}, \boldsymbol{\psi})$ is optimal in (DP).*

[2] *$\forall (i, j), \varphi_i \oplus \psi_j = c_{ij}$ γ_{ij}-a.e.*

Proof. If **[1]** is satisfied, according to Theorem 3.2, $K(\boldsymbol{\gamma}) = D(\boldsymbol{\varphi}, \boldsymbol{\psi})$. We then

compute $D(\boldsymbol{\varphi}, \boldsymbol{\psi})$ as a function of $\boldsymbol{\gamma}$.

$$\begin{aligned} D(\boldsymbol{\varphi}, \boldsymbol{\psi}) : &= \sum_{i=1}^{n} \int_X \varphi_i(x)\, \mathrm{d}\mu_i(x) + \sum_{j=1}^{n} \int_Y \psi_j(x)\, \mathrm{d}\nu_j(x) \\ &= \sum_{1 \leqslant i,j \leqslant n} \iint_{X \times Y} \varphi_i(x)\, \mathrm{d}\gamma_{ij}(x,y) + \sum_{1 \leqslant i,j \leqslant n} \iint_{X \times Y} \psi_j(x)\, \mathrm{d}\gamma_{ij}(x,y) \\ &= \sum_{1 \leqslant i,j \leqslant n} \iint_{X \times Y} \left[\varphi_i(x) + \psi_j(y)\right]\, \mathrm{d}\gamma_{ij}(x,y). \end{aligned}$$

Comparing the latter expression with $K(\boldsymbol{\gamma})$ gives

$$\begin{aligned} 0 &= K(\boldsymbol{\gamma}) - D(\boldsymbol{\varphi}, \boldsymbol{\psi}) \\ &= \sum_{ij} \iint_{X \times Y} \left(c_{ij}(x,y) - \left[\varphi_i(x) + \psi_j(y)\right]\right)\, \mathrm{d}\gamma_{ij}(x,y). \end{aligned}$$

The conclusion follows from the fact that $(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \boldsymbol{\Delta}(\boldsymbol{c})$.

Conversely, if **[2]** is satisfied, it is clear that $K(\boldsymbol{\gamma}) = D(\boldsymbol{\varphi}, \boldsymbol{\psi})$ which implies that both $\boldsymbol{\gamma}$ and $(\boldsymbol{\varphi}, \boldsymbol{\psi})$ are optimal according to Proposition 3.1. □

The result above is not surprising since any given $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ induces n^2 *scalar* optimal transportation problems between each marginals (say) $\pi_1 \# \gamma_{ij} := f_{ij} \mathrm{d}\mu_i$ and $\pi_2 \# \gamma_{ij} := g_{ij} \mathrm{d}\nu_j$

$$\inf \left\{ \iint_{X \times Y} c_{ij}(x,y)\, \mathrm{d}\gamma_{ij}(x,y) : \gamma_{ij} \in \Pi(f_{ij} \mathrm{d}\mu_i, g_{ij} \mathrm{d}\nu_j) \right\}, \qquad (KP_{ij})$$

and looking at contraints in vectorial Kantorovitch's problem, it is easy to see that $\boldsymbol{\gamma}$ has to be optimal in every subproblems (KP_{ij}) to be optimal between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ (if not, take a better one and compare the total cost, which is nothing less than another proof of the result above).

5 Induced metrics

In this section, we take $X = Y$ a Polish space. We investigate how to extend the well-known Wasserstein distance and answer the question "does the problem (KP) define a distance on the space $\mathcal{P}^n(X)$?".

Let $(d_{ij})_{(i,j) \in [\![1,n]\!]^2}$ be n^2 finite, symmetric and non negative functions on $X \times X$ satisfying the triangle inequality (we do not assume that they are distances). Then let $p \in [1, \infty)$, $x_0 \in X$ and define

$$\mathcal{P}_p^n(X) = \left\{ \boldsymbol{m} \in \mathcal{P}^n(X), \forall (i,j) \in [\![1,n]\!]^2, \int_X d_{ij}^p(x_0,x) + d_{ji}^p(x_0,x)\, \mathrm{d}m_i(x) < \infty \right\}$$

and notice that as in scalar case, this set does not depend on x_0.

Definition 5.1. Given $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{P}_p^n(X)$, the p-transportation distance between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ is defined by

$$W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \left(\inf \left\{ \sum_{ij} \iint d_{ij}(x,y)^p \, \mathrm{d}\gamma_{ij}(x,y), \boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu}) \right\} \right)^{\frac{1}{p}}$$
$$:= (\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu}))^{\frac{1}{p}} .$$

The symmetry of W_p is clear provided the costs are symmetric themselves. However, the fact that $W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ implies that $\boldsymbol{\mu} = \boldsymbol{\nu}$ is never satisfied if all costs are (power of) distances. In place of it, if $W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ then $\sum_{i=1}^n \mu_i = \sum_{j=1}^n \nu_j$. In other words, $\boldsymbol{W_p}$ is pseudodistance in that case. To prevent that, we add new hypothesis on (d_{ij}) described in the next proposition.

Proposition 5.1. *Let $(d_{ij})_{(i,j)\in[\![1,n]\!]^2}$ be n^2 symetric finite non negative functions on $X \times X$ satisfyong the triangle inequality. Assume moreover that for all $(i,j) \in [\![1,n]\!]^2, i \neq j, d_{ij}$ is strictly non negative and d_{ii} is a distance. Then for all $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{P}_p^n(X)$, if $W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ then $\boldsymbol{\mu} = \boldsymbol{\nu}$.*

Proof. Let $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{P}_p^n(X)$ be such as $W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ and let $\boldsymbol{\gamma}^*$ be optimal in (KP), then

$$0 = \sum_{k=1}^n \iint_{X^2} d_{kk}(x,y)^p \, \mathrm{d}\gamma_{kk}^*(x,y) + \sum_{i \neq j} \iint_{X^2} d_{ij}(x,y)^p \, \mathrm{d}\gamma_{ij}^*(x,y)$$

According to the strict positivity of non diagonal distances, for all $i \neq j$, $\gamma_{ij}^* = 0$ and then for all $k \in [\![1,n]\!]$, γ_{kk}^* is a transport plan between μ_k and ν_k. The proof of Theorem 7.3. in [6] concludes. □

However, without any other constraints on (d_{ij}), the following example shows that the triangle inequality fails.

Example 5.1. Let $X = \mathbb{R}, n = 2$ and set:

$$\mu = \begin{bmatrix} \delta_0 \\ 0 \end{bmatrix}, \boldsymbol{\nu} = \begin{bmatrix} 0 \\ \delta_1 \end{bmatrix}, \boldsymbol{\lambda} = \begin{bmatrix} \delta_2 \\ 0 \end{bmatrix}$$

Then set $p = 1, d_{11} = |\cdot|$ and $d_{12} = d_{21} := d_\varepsilon$ the ε-discrete distance on $\mathbb{R}$ (with $\varepsilon \in \mathbb{R}_+^*$) defined by $d_\varepsilon(x,y) = \varepsilon$ if $x = y$ and 0 otherwise and an arbitrary distance for d_{22}. Clearly,

$$W_1(\boldsymbol{\mu}, \boldsymbol{\lambda}) = 2, W_1(\boldsymbol{\mu}, \boldsymbol{\nu}) = W_1(\boldsymbol{\nu}, \boldsymbol{\lambda}) = \varepsilon \tag{7}$$

And these three numbers do not satisfy to triangle inequality as soon as ε is smaller enough.

The main problem in the example above is the lack of comparison between all (d_{ij}). To give a everyday-life example, it could be more expansive to travel

between Paris and Berlin using plane than to first travel between Paris and Amsterdam using car and then going to Berlin from Amsterdam using train. To avoid this phenomenon above, we add new constraints on (d_{ij}):

$$\forall (i,j,k) \in [\![1,n]\!]^3, \forall (x,y,z) \in X^3, d_{ik}(x,z) \leqslant d_{ij}(x,y) + d_{jk}(y,z) \qquad \text{(MTI)}$$

and from now on we assume that these contraints are satisfied.

Remark 5.1. Note that (MTI) (for Mixed Triangle Inequalities) contain the fact that all costs satisfy triangle inequality (take $i = j = k$) and if one of theses inequalities is false for some (x_0, y_0, z_0) then one can exhibit a counterexample to fail the triangle inequality on W_p similar to the (counter)Example 5.1 above.

Example 5.2. An easy way to construct objects that satisfy (MTI) is (and then, we do not work on empty set) given a distance d on X and a non negative scalar t (for transformation), $d_{ii} = d$ for all i and $d_{ij} = d + t$ for all (i,j) with $i \neq j$.

Proposition 5.2. *Let $p \in [1,\infty)$ and $(d_{ij})_{(i,j)\in[\![1,n]\!]^2}$ be such that (MTI) are satisfied. Then W_p satisfies the triangle inequality.*

Proof. Let $\boldsymbol{\gamma}^* = (\gamma^*_{ij})$ (resp. $\tilde{\boldsymbol{\gamma}}^* = (\tilde{\gamma}^*_{jk})$) be optimal[4] between $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ (resp. $\boldsymbol{\nu}$ and $\boldsymbol{\lambda}$). Let $j \in [\![1,n]\!]$ and define for all $i,k \in [\![1,n]\!]$ the marginals $\nu_j^{i,\leftarrow} := \pi_2 \# \gamma^*_{ij}$ and $\nu_j^{k,\rightarrow} := \pi_1 \# \tilde{\gamma}^*_{jk}$. These marginals are all submeasures of ν_j and then, according to Radon-Nikodym theorem, we denote by $f_j^{i,\leftarrow}$ (resp. $f_j^{k,\rightarrow}$) the density of $\nu_j^{i,\leftarrow}$ (resp. $\nu_j^{k,\rightarrow}$) w.r.t. ν_j. Finally, define for each $i,j,k \in [\![1,n]\!]$ the following transference plans

$$\gamma^*_{ijk} \text{ is defined as the measure with density } (x,y) \to f_j^{k,\rightarrow}(y) \text{ w.r.t. } \gamma^*_{ij}, \quad (8)$$

$$\tilde{\gamma}^*_{ijk} \text{ is defined as the measure with density } (y,z) \to f_j^{i,\leftarrow}(y) \text{ w.r.t. } \tilde{\gamma}^*_{jk}, \quad (9)$$

these definitions imply that

$$\forall (i,j) \in [\![1,n]\!]^2, \gamma^*_{ij} = \sum_{k=1}^{n} \gamma^*_{ijk}, \quad (10)$$

$$\forall (j,k) \in [\![1,n]\!]^2, \tilde{\gamma}^*_{jk} = \sum_{i=1}^{n} \tilde{\gamma}^*_{ijk}, \quad (11)$$

$$\forall (i,j,k) \in [\![1,n]\!]^3, \pi_2 \# \gamma^*_{ijk} = \pi_1 \# \tilde{\gamma}^*_{ijk}. \quad (12)$$

[4] they exist according to 2.1, even if it is not necessary here: passing to supremum bound *aposteriori* otherwise.

To obtain the last equality, fix B a measurable subset of Y and compute

$$\begin{aligned}
\pi_2\#\gamma^*_{ijk}(B) &= \gamma^*_{ijk}(X\times B)\\
&= \iint_{X\times B} f_j^{k,\rightarrow}(y)\,\mathrm{d}\gamma^*_{ij}(x,y) \text{ by (8)}\\
&= \int_B f_j^{k,\rightarrow}(y)\,\mathrm{d}\nu_j^{i,\leftarrow}(y) \text{ by definition of } \nu_j^{i,\leftarrow}\\
&= \int_B f_j^{k,\rightarrow}(y) f_j^{i,\leftarrow}(y)\,\mathrm{d}\nu_j(y) \text{ by definition of } f_j^{i,\leftarrow}\\
&= \int_B f_j^{i,\leftarrow}(y)\,\mathrm{d}\nu_j^{k,\rightarrow}(y) \text{ by definition of } f_j^{k,\rightarrow}\\
&= \iint_{B\times Z} f_j^{i,\leftarrow}(y)\,\mathrm{d}\tilde{\gamma}^*_{ij}(y,z) \text{ by definition of } \nu_j^{k,\rightarrow}\\
&= \pi_1\#\tilde{\gamma}^*_{ijk}(B) \text{ by (9)}.
\end{aligned}$$

Then, equalities (12) allow us to apply the Gluing Lemma (see [6], Lemma 7.6) and guarantee the existence of Π_{ijk} a measure on $X\times Y\times Z$ such that

$$\pi_{1,2}\#\Pi_{ijk} = \gamma^*_{ijk} \text{ and } \pi_{2,3}\#\Pi_{ijk} = \tilde{\gamma}_{ijk}.$$

We next define for all $(i,k)\in[\![1,n]\!]^2, \Pi_{ik} = \sum_{j=1}^n \Pi_{ijk}$ and compute

$$\begin{aligned}
\pi_1\#\sum_{k=1}^n \Pi_{ik} &= \sum_{k=1}^n\sum_{j=1}^n \pi_1\#\Pi_{ijk} \text{ by definition}\\
&= \sum_{k=1}^n\sum_{j=1}^n \pi_1\#\gamma^*_{ijk} \text{ by (13)}\\
&= \sum_{j=1}^n \pi_1\#\left(\sum_{k=1}^n \gamma^*_{ijk}\right)\\
&= \sum_{j=1}^n \pi_1\#\gamma^*_{ij} \text{ by (10)}\\
&= \mu_i.
\end{aligned}$$

For identical reasons $\pi_3\#\sum_{i=1}^n \Pi_{ik} = \lambda_k$ and as a consequence:

$$\boldsymbol{\gamma} := (\pi_{1,3}\#\Pi_{ik})_{1\leqslant i,k\leqslant n} \in \boldsymbol{\Pi}(\boldsymbol{\mu},\boldsymbol{\lambda}).$$

Finally, we have:

$$
\begin{aligned}
W_p(\boldsymbol{\mu},\boldsymbol{\lambda}) &\leqslant \left(\sum_{ik}\iint d_{ik}(x,z)^p \,\mathrm{d}\gamma_{ik}(x,y,z)\right)^{\frac{1}{p}} \\
&= \left(\sum_{ijk}\iiint d_{ik}(x,z)^p \,\mathrm{d}\Pi_{ijk}(x,y,z)\right)^{\frac{1}{p}} \text{ by definition of } \boldsymbol{\gamma} \text{ and } \boldsymbol{\Pi} \\
&\leqslant \left(\sum_{ijk}\iiint (d_{ij}(x,y)+d_{jk}(y,z))^p \,\mathrm{d}\Pi_{ijk}(x,y,z)\right)^{\frac{1}{p}} \text{ by (MTI)} \\
&\leqslant \left(\sum_{ijk}\iiint d_{ij}(x,y)^p \,\mathrm{d}\Pi_{ijk}(x,y,z)\right)^{\frac{1}{p}} + \left(\sum_{ijk}\iiint d_{jk}(y,z)^p \,\mathrm{d}\Pi_{ijk}(x,y,z)\right)^{\frac{1}{p}} \\
&= \left(\sum_{ijk}\iint d_{ij}(x,y)^p \,\mathrm{d}\gamma^*_{ijk}(x,z)\right)^{\frac{1}{p}} + \left(\sum_{ijk}\iint d_{jk}(y,z)^p \,\mathrm{d}\tilde{\gamma}^*_{ijk}(x,y,z)\right)^{\frac{1}{p}} \text{ by (13)} \\
&= \left(\sum_{ij}\iint d_{ij}(x,y)^p \,\mathrm{d}\gamma^*_{ij}(x,y)\right)^{\frac{1}{p}} + \left(\sum_{jk}\iint d_{jk}(y,z)^p \,\mathrm{d}\tilde{\gamma}^*_{jk}(y,z)\right)^{\frac{1}{p}} \\
&= W_p(\boldsymbol{\mu},\boldsymbol{\nu}) + W_p(\boldsymbol{\nu},\boldsymbol{\lambda}).
\end{aligned}
$$

□

Theorem 5.1. *Let X be a Polish space. Let $p \in [1,\infty)$. Let (d_{ij}) be n^2 functions on $X \times X$ valued in $\mathbb{R}_+$ such that:*
[1] *$\forall (i,j) \in [\![1,n]\!]^2, d_{ij}$ is symmetric,*
[2] *(MTI) is satisfied,*
[3] *$\forall i \in [\![1,n]\!], \forall x \in X, d_{ii}(x,x) = 0$.*
[4] *$\forall (i,j) \in [\![1,n]\!]^2, i \neq j, \forall (x,y) \in X \times Y, d_{ij}(x,y) \neq 0$.*
W_p is a distance on $\mathcal{P}^n_p(X)$.

Example 5.3. Coming back to the Example 2.2, let us fix all weights

$$\boldsymbol{p} = \boldsymbol{q} = \frac{1}{n}\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix} \tag{13}$$

and given $x, y \in X^n$, we define the distance $w_p(\boldsymbol{x},\boldsymbol{y})$ between them by

$$w_p(\boldsymbol{x},\boldsymbol{y}) = W_p\left(\frac{1}{n}\begin{bmatrix}\delta_{x_1}\\ \vdots \\ \delta_{x_n}\end{bmatrix}, \frac{1}{n}\begin{bmatrix}\delta_{y_1}\\ \vdots \\ \delta_{y_n}\end{bmatrix}\right).$$

According to 2.2, w_p is given by

$$w_p(\boldsymbol{x},\boldsymbol{y})^p = \inf\left\{\sum_{(i,j)\in[\![1,n]\!]^2} t_{ij}d_{ij}(x_i,y_j)^p, \boldsymbol{t}\in\mathcal{M}_n(\mathbb{R}), \sum_{l=1}^n t_{il}=\frac{1}{n}, \sum_{l=1}^n t_{lj}=\frac{1}{n}\right\}$$
$$= \frac{1}{n}\min\left\{\sum_{(i,j)\in[\![1,n]\!]^2} t_{ij}d_{ij}(x_i,y_j)^p, \boldsymbol{t} \text{ bistochastic matrix n } \times \text{ n}\right\},$$

the last equality providing from classical arguments of linear programing. This example show a way to define new distances on a finite product of spaces using n^2 distances.

6 Conclusion

The aim of this paper was to present a new point of view in vector-valued optimal transportation. Writing this paper, we discover that in [18] that these authors suggest to use the same idea to treat this problem and allowed mixing of species. Their point of view follows a dynamical formulation of optimal transportation (presented in [4]) while in our paper, Kantorovitch's point of view of optimal transportation was our approach angle.

Concerning this approach angle, let us make another small digression about Monge's optimal transportation's problem and present it. Given two probabilities μ and ν we are interrested in knowing if the optimal tranference plan between μ and ν split mass *i.e* if (formaly) the support of γ^* is included in a function's graph, say T^*. A natural question here is if there exists a similar problem associated to (KP)? An possible answer is given $\boldsymbol{\gamma}\in\Pi(\boldsymbol{\mu},\boldsymbol{\nu})$ we declare that $\boldsymbol{\gamma}$ has a $Monge's\ form$ if for all $(i,j),\gamma_{ij}$ is included in a function's graph say T_{ij}. The main problem here is given (T_{ij}), it is not possible to build the associated γ_{ij}. Indeed, the knowledge of (T_{ij}) does not include which parts of μ_i is transported into ν_j or in other words we have still to fix (f_{ij}) and (g_{ij}). This remark makes hard to use only entropic relaxations (see [19]) to solve our problem since the main data to find optimal transference plan is to find these (f_{ij}) and (g_{ij}).

Acknowledgements

The author would like to thank B. Nazaret for his advices and his re-reading of the paper.

References

[1] G Monge. Mémoire sur la théorie des déblais et des remblais: Histoire de 1'academieroyale des sciences. 1781.

[2] Leonid Kantorovitch. On the translocation of masses. *Management Science*, 5(1):1–4, 1958.

[3] Martin Beckmann. A continuous model of transportation. *Econometrica: Journal of the Econometric Society*, pages 643–660, 1952.

[4] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.

[6] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

[7] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 2015.

[8] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[9] Guillaume Carlier and Bruno Nazaret. Optimal transportation for the determinant. *ESAIM: Control, Optimisation and Calculus of Variations*, 14(4):678–698, 2008.

[10] Jun Kitagawa and Brendan Pass. The multi-marginal optimal partial transport problem. In *Forum of Mathematics, Sigma*, volume 3. Cambridge University Press, 2015.

[11] Codina Cotar, Gero Friesecke, and Brendan Pass. Infinite-body optimal transport with coulomb cost. *Calculus of Variations and Partial Differential Equations*, 54(1):717–742, 2015.

[12] Eric A Carlen and Jan Maas. An analog of the 2-wasserstein metric in non-ommutative probability under which the fermionic fokker–planck equation is gradient flow for the entropy. *Communications in mathematical physics*, 331(3):887–926, 2014.

[13] Yongxin Chen, Tryphon T Georgiou, Lipeng Ning, and Allen Tannenbaum. Matricial wasserstein-1 distance. *IEEE control systems letters*, 1(1):14–19, 2017.

[14] Jonathan Zinsl and Daniel Matthes. Transport distances and geodesic convexity for systems of degenerate diffusion equations. *Calculus of Variations and Partial Differential Equations*, 54(4):3397–3438, 2015.

[15] Jean-David Benamou, Yann Brenier, and Kévin Guittet. *Numerical resolution of a multiphasic optimal mass transport problem*. PhD thesis, INRIA, 2000.

[16] Jean-David Benamou, Yann Brenier, and Kevin Guittet. Numerical analysis of a multi-phasic mass transport problem. *Contemporary Mathematics*, 353:1–18, 2004.

[17] Hugo Lavenant. Time-convexity of the entropy in the multiphasic formulation of the incompressible euler equation. *Calculus of Variations and Partial Differential Equations*, 56(6):170, 2017.

[18] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Vector-valued optimal mass transport. *SIAM Journal on Applied Mathematics*, 78(3):1682–1696, 2018.

[19] Luca Nenna. *Numerical methods for multi-marginal optimal transportation.* PhD thesis, PSL Research University, 2016.