Rend. Lincei Mat. Appl. 34 (2023), 745–771
DOI 10.4171/RLM/1026

Calculus of Variations. – *On the pythagorean structure of the optimal transport for separable cost functions*, by Gennaro Auricchio, communicated on 10 November 2023.

Abstract. – In this paper, we study the optimal transport problem induced by two measures supported over two polish spaces, namely, X and Y, which are the product of n smaller polish spaces, that is, $X = \times_{j=1}^{n} X_j$ and $Y = \times_{j=1}^{n} Y_j$. In particular, we focus on problems induced by a cost function $c : X \times Y \to [0, +\infty)$ that is separable; i.e., c is such that $c = c_1 + \cdots + c_n$, where each c_j depends only on the couple (x_j, y_j), and thus $c_j : X_j \times Y_j \to [0, +\infty)$. Noticeably, if $X = Y = \mathbb{R}^n$, this class of cost functions includes all the l_p^p costs. Our main result proves that the optimal transportation plan with respect to a separable cost function between two given measures can be expressed as the composition of n different lower-dimensional transports, one for each pair of coordinates (x_i, y_i) in $X \times Y$. This allows us to decompose the entire Wasserstein cost as the sum of n lower-dimensional Wasserstein costs and to prove that there always exists an optimal transportation plan whose random variable enjoys a conditional independence property with respect to its marginals. We then show that our formalism allows us to explicitly compute the optimal transportation plan between two probability measures when each measure has independent marginals. Finally, we focus on two specific frameworks. In the first one, the cost function is a separable distance, i.e., $d = d_1 + d_2$, where both d_1 and d_2 are distances themselves. In the second one, both measures are supported over $\mathbb{R}^n$ and the cost function is of the form $c(x, y) = h(|x_1 - y_1|) + h(|x_2 - y_2|)$, where h is a convex function such that $h(0) = 0$.

Keywords. – Wasserstein distance, optimal transport, structure of the optimal plan.

Mathematics Subject Classification 2020. – 49Q22 (primary); 49Q20 (secondary).

1. Introduction

The optimal transport problem is a classical minimization problem that dates back to the pioneering works of Monge [23] and Kantorovich [19, 20]. Given two probability measures, namely, μ and ν, the objective of the problem is to identify the cheapest way to reshape μ into ν. The expenditure of the transformation depends on a cost function, which encapsulates the geometric characteristics of the underlying space. Within an appropriate framework, this minimization problem induces a metric over the space of probability measures.

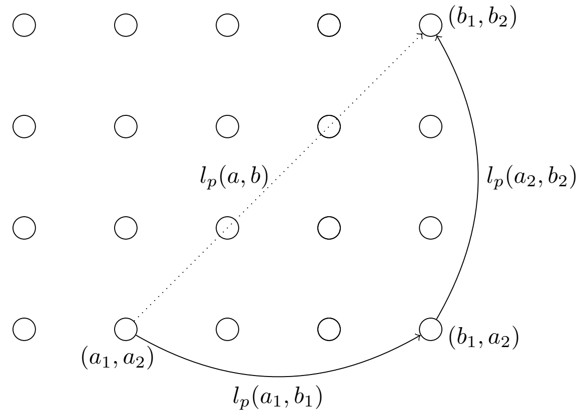

FIGURE 1. An exemplification of the $(d+1)$-partite approach presented in [5] when $d = 2$. Instead of moving a unit of mass from node $a = (a_1, a_2)$ to node $b = (b_1, b_2)$, we first move the unit from a to (b_1, a_2) and then from (b_1, a_2) to b. The cost of performing this two-step movement is the same as moving the mass from a to b since $l_p^p(a, b) = l_p^p(a_1, b_1) + l_p^p(a_2, b_2)$.

Throughout the last century, the optimal transport (OT) problem has emerged as a highly valuable tool across several applied domains, such as the study of systems of particles by Dobrushin [13], the Boltzmann equation by Tanaka [24, 33, 34], and the field of fluidodynamics by Yann Brenier [9]. These contributions highlighted the power of qualitative descriptions of optimal transport, providing insights into a multitude of longstanding problems. Consequently, the optimal transport problem has emerged as a topic of interest for analysts, probabilists, and statisticians [2, 31, 35]. Notably, a plethora of results pertaining to the uniqueness [10, 14, 16], structure [1, 6, 11, 30], and regularity [8, 22] of the optimal transportation plan within the continuous framework have been established.

More recently, it has been observed that harnessing the properties of the transportation plan is advantageous also from a computational standpoint. This revelation aligns with the growing prominence of the optimal transport problem as a crucial subproblem across various domains, including computer vision [4, 26, 28, 29] and machine learning [3, 12, 15, 32]. Consequently, there has been an effort within the community to develop efficient methods for solving the minimization problem associated with the optimal transport problem. For instance, as shown in [5], if the two probability measures are supported over two regular grids, the classic uncapacitated minimum cost flow problem [17, 25] can be reformulated as a $(d+1)$-partite graph, a structure amenable to efficient handling. The core concept behind the $(d+1)$-partite graph formulation lies in the fact that the l_p^p cost between two nodes on a bidimensional grid, that is, $l_p^p(x, y) = |x_1 - y_1|^p + |x_2 - y_2|^p$, can be decomposed as two l_p^p costs along the two cardinal directions, i.e., $l_p^p(x, y) = l_p^p(x_1, y_1) + l_p^p(x_2, y_2)$, as shown in Figure 1.

In this paper, we expand upon the framework introduced in [5], originally designed for measures supported over regular grids. Our extension pertains to a more general scenario where the measures are supported on two Polish spaces, denoted as X and Y, both of which are products of n smaller Polish spaces. Specifically, X is the Cartesian product of n Polish spaces, defined as $X = \times_{j=1}^{n} X_j$. Likewise, Y is defined as $Y = \times_{j=1}^{n} Y_j$. Instead of considering only the l_p^p costs, we generalize the discussion to the wider class of separable cost functions. A cost function c is separable if c can be decomposed into the sum of n functions, i.e., $c = c_1 + \cdots + c_n$, with each c_j depending only on the pair (x_j, y_j); thus, $c_j : X_j \times Y_j \to [0, +\infty)$.

We then introduce the notion of cardinal flows, formulate the related minimization problem, and show that it is equivalent to the classic optimal transport problem. We demonstrate that the separability of the cost function enables the separation of the total transportation cost. In particular, given two probability measures μ and ν supported over $X = X_1 \times X_2$ and $Y = Y_1 \times Y_2$, respectively, it holds that

$$\mathscr{C}_c(\mu, \nu) = \int_{X_2} \mathscr{C}_{c_1}(\mu_{|x_2}, \zeta_{|x_2})\, d\mu_2 + \int_{Y_1} \mathscr{C}_{c_2}(\zeta_{|y_1}, \nu_{|y_1})\, d\nu_1,$$

where ζ is a feasible probability measure, which we name *pivot measure*, and $\mathscr{C}_c(\mu, \nu)$ is the optimal transportation cost between μ and ν associated with c. We then focus on the relations between the pivot measures and the optimal transportation plan between two measures. We show how to build an optimal transportation plan given a pivot measure and prove that there always exists an optimal transportation plan whose associated random variable enjoys a conditional independence property with respect to its marginals. We conclude our study by considering two specific scenarios. In the first scenario, the cost function is a separable distance $d : X \times Y \to [0, +\infty)$, specifically $d = d_1 + d_2$, where both $d_1 : X_1 \times Y_1 \to [0, +\infty)$ and $d_2 : X_2 \times Y_2 \to [0, +\infty)$ are distances. In the second scenario, both measures are defined over the Euclidean space $\mathbb{R}^n$, and the cost function takes the form $c(x, y) = h(|x_1 - y_1|) + h(|x_2 - y_2|)$, where h represents a convex function such that $h(0) = 0$.

1.1. Preliminaries and notations

We now fix our notation and recall the optimal transport problem. To keep the discussion as general as possible, we only require X and Y to be Polish spaces. For a complete introduction to the theory of optimal transportation, we refer to [2, 7, 35].

Given a Polish space (X, d), we denote with $\mathscr{P}(X)$ the set of all the Borel probability measures over X and with $\mathscr{P}_p(X)$ the subset of probability measures that have finite p-moment. We denote with $\mathrm{spt}(\mu)$ the support of the measure μ. For any given measurable function $T : X \to Y$, we denote with $T_{\#}\mu \in \mathscr{P}(Y)$ the push-forward of μ through T, defined as $T_{\#}\mu(A) = \mu(T^{-1}(A))$ for every $A \subset Y$. We recall that if $T : X \to Y$ and $S : Y \to Z$ are both measurable functions, then, for any given $\mu \in \mathcal{P}(X)$, the following chain rule $(S \circ T)_{\#}\mu = S_{\#}(T_{\#}\mu)$ holds.

Throughout the paper, we assume that every Polish space X is the direct product of at least two Polish spaces, namely, X_1 and X_2, so that $X = X_1 \times X_2$. In this framework, the projections over X_1 and X_2 are then

$$(\mathfrak{p}_{X_1})(\boldsymbol{x}) := x_1 \quad \text{and} \quad (\mathfrak{p}_{X_2})(\boldsymbol{x}) := x_2,$$

respectively, for $\boldsymbol{x} = (x_1, x_2) \in X = X_1 \times X_2$.

The i-th marginal of $\mu \in \mathcal{P}(X)$ is the probability measure $\mu_i \in \mathcal{P}(X_i)$ defined as $\mu_i := (\mathfrak{p}_{X_i})_{\#}\mu$. We say that μ is an independent measure if its marginals, that is, $\mu_1 \in \mathcal{P}(X_1)$ and $\mu_2 \in \mathcal{P}(X_2)$, are independents. In this case, we write $\mu = \mu_1 \otimes \mu_2$, where $\otimes$ denotes the product between measures, so that $\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for every $A_i \subset X_i$. Moreover, we denote with $\{\mu_{|x_i}\}_{x_i \in X_i}$ the disintegration of μ with respect to the function $(\mathfrak{p}_{X_i})$. The measure $\mu_{|x_i}$ is called the conditional law of μ given x_i and, with a slight abuse of notation, we write $\mu = \mu_{|x_i} \otimes \mu_i$. We will also call $\mu_{|x_i}$ the conditional law of μ with respect to its i-th marginal. For a complete discussion on the existence and uniqueness of the conditional laws, we refer to [7, Chapter 10].

The first formulation of the transportation problem is due to Monge and, in modern language, to Kantorovich. In [18], the author modelized the transhipment of mass through a probability measure over the product space $X \times Y$. He called these measures transportation plans.

DEFINITION 1 (Transportation plan). Let μ and ν be two measures over two Polish spaces X and Y. A probability measure $\pi \in \mathcal{P}(X \times Y)$ is a transportation plan between μ and ν if

$$(\mathfrak{p}_X)_{\#}\pi = \mu \quad \text{and} \quad (\mathfrak{p}_Y)_{\#}\pi = \nu.$$

We denote with $\Pi(\mu, \nu)$ the set of all the transportation plans between μ and ν.

DEFINITION 2 (Transportation functional). Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and let $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous function such that there exist two upper semi-continuous functions $a \in L^1_\mu$ and $b \in L^1_\nu$ for which it holds true that $c(x, y) \geq a(x) + b(x)$ for each $(x, y) \in X \times Y$. In this framework, the transportation functional $\mathbb{T}_c : \Pi(\mu, \nu) \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\mathbb{T}_c(\pi) := \int_{X \times Y} c \, d\pi. \tag{1.1}$$

The conditions imposed on the cost function in Definition 2 are the minimal ones for which the integral in (1.1) is well defined. Under these assumptions, we define the following minimum problem.

DEFINITION 3 (Minimal transportation cost). Let us take a cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ as in Definition 2. The minimal transportation cost functional $\mathscr{C}_c : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$(\mu, \nu) \to \mathscr{C}_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \mathbb{T}_c(\pi). \tag{1.2}$$

The value $\mathscr{C}_c(\mu, \nu)$ is also called the Wasserstein cost between μ and ν, with respect to the cost function c.

By making further assumptions on c, it is possible to prove that the infimum in (1.2) is a minimum. In particular, when the cost function is non-negative, a minimizing solution exists. We denote with $\Gamma_o(\mu, \nu)$ the set of minimizers.

2. THE CARDINAL FLOW AND THE PIVOT MEASURE FORMULATION

From now on, we assume X and Y to be the product of two Polish spaces, i.e., $X = X_1 \times X_2$ and $Y = Y_1 \times Y_2$. In this framework, we can introduce the separable cost function and reformulate the optimal transport problem as an optimal cardinal flow problem.

DEFINITION 4 (Separable cost function). Let $X = X_1 \times X_2$ and $Y = Y_1 \times Y_2$ be two Polish spaces. We say that $c : X \times Y \to \mathbb{R}$ is *separable* if there exists a pair of functions $c_1 : X_1 \times Y_1 \to \mathbb{R}$ and $c_2 : X_2 \times Y_2 \to \mathbb{R}$ such that

$$c(\boldsymbol{x}, \boldsymbol{y}) := c_1(x_1, y_1) + c_2(x_2, y_2)$$

for each $\boldsymbol{x} = (x_1, x_2) \in X$ and for each $\boldsymbol{y} = (y_1, y_2) \in Y$.

DEFINITION 5 (Cardinal flow). Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We say that the couple of measures $(f^{(1)}, f^{(2)}) \in \mathcal{P}(X \times Y_1) \times \mathcal{P}(X_2 \times Y)$ is a *cardinal flow* between μ and ν if it satisfies the following conditions.

- The marginal on X of $f^{(1)}$ is equal to μ, i.e.,

$$\mu = (\mathfrak{p}_X)_\# f^{(1)}.$$

- The marginal on Y of $f^{(2)}$ is equal to ν, i.e.,

$$\nu = (\mathfrak{p}_Y)_\# f^{(2)}.$$

- The flows $f^{(1)}$ and $f^{(2)}$ have the same marginal on $Y_1 \times X_2$, i.e.,

$$(\mathfrak{p}_{Y_1 \times X_2})_\# f^{(1)} = (\mathfrak{p}_{Y_1 \times X_2})_\# f^{(2)}.$$

We call the measures $f^{(1)}$ and $f^{(2)}$ first and second cardinal flow, respectively. Moreover, we denote with $\mathcal{F}(\mu,\nu)$ the set of all cardinal flows between μ and ν.

REMARK 1. For any couple of probability measures μ and ν, the set $\mathcal{F}(\mu,\nu)$ is nonempty. In fact, the couple $(f^{(1)}, f^{(2)})$, defined as

$$f^{(1)} = \mu \otimes \nu_1 \quad \text{and} \quad f^{(2)} = \mu_2 \otimes \nu,$$

belongs to $\mathcal{F}(\mu,\nu)$. Moreover, the sets $\mathcal{F}(\mu,\nu)$ and $\mathcal{F}(\nu,\mu)$ are, in general, different. For instance, let $X = Y = \mathbb{R}^2$, $\mu = \delta_{(0,0)}$, and $\nu = \delta_{(1,1)}$. In this case, $\mathcal{F}(\mu,\nu) = \{\delta_{((0,0);1)}\}$, while $\mathcal{F}(\nu,\mu) = \{\delta_{((1,1);0)}\}$.

DEFINITION 6 (Cardinal flow functional). Given two probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and a separable cost function $c = c_1 + c_2$ over $X \times Y$, we define the first and second cardinal transportation functionals as

$$\mathbb{CT}_c^{(1)}(f^{(1)}) = \int_{X\times Y_1} c_1\, df^{(1)} \quad \text{and} \quad \mathbb{CT}_c^{(2)}(f^{(2)}) = \int_{X_2\times Y} c_2\, df^{(2)}$$

where $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$. The total cardinal flow functional is then defined as

$$\mathbb{CT}_c(f^{(1)}, f^{(2)}) = \mathbb{CT}_c^{(1)}(f^{(1)}) + \mathbb{CT}_c^{(2)}(f^{(2)}).$$

THEOREM 1. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and let $c : X \times Y \to [0,+\infty)$ be a separable cost function. Then,*

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) = \min_{(f^{(1)},f^{(2)})\in\mathcal{F}(\mu,\nu)} \mathbb{CT}_c(f^{(1)}, f^{(2)}).$$

PROOF. Let us consider two measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Given $\pi \in \Pi(\mu,\nu)$, let $f^{(1)}$ and $f^{(2)}$ be the marginals of π over $X \times Y_1$ and $X_2 \times Y$, respectively. Then, the following identity holds:

$$\int_{X\times Y} c\, d\pi = \int_{X\times Y_1} c_1\, df^{(1)} + \int_{X_2\times Y} c_2\, df^{(2)};$$

thus, $\mathbb{T}_c(\pi) = \mathbb{CT}_c(f^{(1)}, f^{(2)})$. In particular, we infer

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) \geq \min_{(f^{(1)},f^{(2)})\in\mathcal{F}(\mu,\nu)} \mathbb{CT}_c(f^{(1)}, f^{(2)}). \tag{2.1}$$

To conclude, we show the inverse inequality. Let us now consider a cardinal flow $(f^{(1)}, f^{(2)})$. By definition of cardinal flow, it holds that

$$(\mathfrak{p}_{Y_1\times X_2})_\# f^{(1)} = (\mathfrak{p}_{Y_1\times X_2})_\# f^{(2)}.$$

Thus, owing to the gluing lemma (see [35, Section 1]), there exists a probability $\pi \in \mathcal{P}(X \times Y)$ whose marginal on $X \times Y_1$ is $f^{(1)}$ and whose marginal on $X_2 \times Y$ is $f^{(2)}$.

Moreover, we have that $\pi \in \Pi(\mu,\nu)$. Indeed, we have that $(\mathfrak{p}_X)_\# \pi = (\mathfrak{p}_X)_\#((\mathfrak{p}_{X\times Y_1})_\# \pi) = (\mathfrak{p}_X)_\# f^{(1)} = \mu$. Similarly, it holds that $(\mathfrak{p}_Y)_\# \pi = \nu$. We have then shown that, for any given cardinal flow $(f^{(1)}, f^{(2)})$, there exists a transportation plan π for which it holds that $\mathbb{T}_c(\pi) = \mathbb{C}\mathbb{T}_c(f^{(1)}, f^{(2)})$; thus,

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{T}_c(\pi) \le \min_{(f^{(1)},f^{(2)})\in\mathcal{F}(\mu,\nu)} \mathbb{C}\mathbb{T}_c(f^{(1)}, f^{(2)}),$$

which, combined with (2.1), concludes the proof. ■

Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, we define the function $L : \Pi(\mu,\nu) \to \mathcal{F}(\mu,\nu)$ as

$$L(\pi) = \big((\mathfrak{p}_{X\times Y_1})_\#(\pi), (\mathfrak{p}_{X_2\times Y})_\#(\pi)\big).$$

By the chain rule for push-forwards, we infer that $L(\pi) \in \mathcal{F}(\mu,\nu)$ for each $\pi \in \Pi(\mu,\nu)$. Vice-versa, let us take $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$. Owing to the gluing lemma [35, Chapter 1], we find $\pi \in \Pi(\mu,\nu)$ such that $L(\pi) = (f^{(1)}, f^{(2)})$. We therefore conclude that $\mathcal{F}(\mu,\nu) = L(\Pi(\mu,\nu))$. We notice that the functionals $\mathbb{T}_c$ and $\mathbb{C}\mathbb{T}_c$ are related through the function L as follows:

$$\mathbb{T}_c(\pi) = \mathbb{C}\mathbb{T}_c\big(L(\pi)\big), \quad \forall \pi \in \Pi(\mu,\nu).$$

This relation, together with the identity $L(\Pi(\mu,\nu)) = \mathcal{F}(\mu,\nu)$, allows us to conclude that, in the framework described in Definition 2, the infimum of $\mathbb{C}\mathbb{T}_c$ is a minimum and that the set of minimizers of $\mathbb{C}\mathbb{T}_c$ coincides with the image of $\Gamma_o(\mu,\nu)$ through L. As a straightforward consequence, we infer that any set of conditions that ensures the uniqueness of the optimal transportation plan ensures also the uniqueness of the optimal cardinal flow.

COROLLARY 1. *Whenever the optimal transportation plan is unique, so is the optimal cardinal flow.*

REMARK 2. Since the operator L is only surjective and not injective, the reverse implication is not true; i.e., given an optimal cardinal flow $(f^{(1)}, f^{(2)})$, there might exist several optimal transportation plans π such that $L(\pi) = (f^{(1)}, f^{(2)})$.

DEFINITION 7. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We define the set of intermedium measures between μ and ν as

$$\mathcal{I}(\mu,\nu) := \big\{\lambda \in \mathcal{P}(Y_1 \times X_2) \text{ s.t. } (\mathfrak{p}_{X_2})_\#(\lambda) = \mu_2 \text{ and } (\mathfrak{p}_{Y_2})_\#(\lambda) = \nu_1\big\}.$$

Given $\lambda \in \mathcal{I}(\mu,\nu)$, we say that the cardinal flow $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)$ glues on λ if

$$(\mathfrak{p}_{Y_1\times X_2})_\# f^{(1)} = (\mathfrak{p}_{Y_1\times X_2})_\# f^{(2)} = \lambda.$$

As from a cardinal flow we are always able to retrieve a transportation plan, from any intermediate measure we are able to retrieve at least a cardinal flow that glues on it.

LEMMA 1. *Let* $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, *and* $\lambda \in \mathcal{I}(\mu, \nu)$. *Then, there exists* $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$ *such that*

$$(\mathfrak{p}_{X_2 \times Y_1})_\# f^{(1)} = \lambda = (\mathfrak{p}_{X_2 \times Y_1})_\# f^{(2)}.$$

PROOF. Let $\lambda \in \mathcal{I}(\mu, \nu)$. By disintegrating λ, we get

$$\lambda = \lambda_{|x_2} \otimes \lambda_2 = \lambda_{|x_2} \otimes \mu_2 \quad \text{and} \quad \mu = \mu_{|x_2} \otimes \mu_2.$$

We define $f^{(1)} \in \mathcal{P}(X \times Y_1)$ as

$$f^{(1)} = (\lambda_{|x_2} \otimes \mu_{|x_2}) \otimes \mu_2.$$

It is easy to see that $(\mathfrak{p}_X)_\# f^{(1)} = \mu$ and $(\mathfrak{p}_{X \times Y_1})_\# f^{(1)} = \lambda$. Similarly, we define $f^{(2)}$ as

$$f^{(2)} = (\lambda_{|y_1} \otimes \nu_{|y_1}) \otimes \nu_1,$$

so that

$$\nu = (\mathfrak{p}_Y)_\# f^{(2)} \quad \text{and} \quad \lambda = (\mathfrak{p}_{Y_1 \times X_2})_\# f^{(2)};$$

hence, $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$. ■

DEFINITION 8 (Pivot measure). Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and let c be a separable cost function. We say that $\zeta \in \mathcal{P}(Y_1 \times X_2)$ is a *pivot measure* between μ and ν if there exists at least an optimal cardinal flow $(f^{(1)}, f^{(2)})$ that glues on it.

REMARK 3. Using the chain rule for push-forwards, we have that all the pivot measures are also intermediate measures.

We are now ready to state our main result, which allows us to decompose the Wasserstein cost along the coordinates of the space on which the measures μ and ν are supported.

THEOREM 2. *Let* $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, *and let* $c = c_1 + c_2$ *be a separable cost function. For any pivot measure* ζ, *it holds that*

$$\mathscr{C}_c(\mu, \nu) = \int_{X_2} \mathscr{C}_{c_1}(\mu_{|x_2}, \zeta_{|x_2})\, d\mu_2 + \int_{Y_1} \mathscr{C}_{c_2}(\zeta_{|y_1}, \nu_{|y_1})\, d\nu_1. \tag{2.2}$$

PROOF. Let ζ be a pivot measure between μ and ν. From Remark 3, we know that $\zeta \in \mathcal{I}(\mu, \nu)$. The disintegration theorem (see [7, Chapter 10]) allows us to write

$$\zeta = \sigma^{(1)}_{|x_2} \otimes \mu_2 \quad \text{and} \quad \zeta = \sigma^{(2)}_{|y_1} \otimes \nu_1. \tag{2.3}$$

Similarly, we decompose μ and ν as

$$\mu = \mu_{|x_2} \otimes \mu_2 \quad \text{and} \quad \nu = \nu_{|y_1} \otimes \nu_1, \tag{2.4}$$

respectively. Notice that, for μ_2-almost every $x_2 \in X_2$, the following quantity is well defined

$$\mathscr{C}_{c_1}(\mu_{|x_2}, \sigma^{(1)}_{|x_2}) = \inf_{\pi^{(1)}_{|x_2} \in \Pi(\mu_{|x_2}, \sigma^{(1)}_{|x_2})} \int_{X_1 \times Y_1} c_1 \, d\pi^{(1)}_{|x_2}.$$

Moreover, since the maps $x_2 \to \mu_{|x_2}$ and $x_2 \to \sigma^{(1)}_{|x_2}$ are both measurable, there exists a measurable selection of optimal plans $\pi^{(1)}_{|x_2}$ for which it holds true that

$$\mathscr{C}_{c_1}(\mu_{|x_2}, \sigma^{(1)}_{|x_2}) = \int_{X_1 \times Y_1} c_1 \, d\pi^{(1)}_{|x_2} \tag{2.5}$$

for μ_2-almost every $x_2 \in X_2$ [35, Corollary 5.21]. Similarly, there exists a measurable selection $\pi^{(2)}_{|y_1}$ for which, for ν_1-almost every $y_1 \in Y_1$,

$$\mathscr{C}_{c_2}(\sigma^{(2)}_{|y_1}, \nu_{|y_1}) = \int_{X_2 \times Y_2} c_2 \, d\pi^{(2)}_{|y_1}. \tag{2.6}$$

Let us now consider the measures $f^{(1)} \in \mathscr{P}(X \times Y_1)$ and $f^{(2)} \in \mathscr{P}(X_2 \times Y)$, defined as

$$f^{(1)} = \pi^{(1)}_{|x_2} \otimes \mu_2 \quad \text{and} \quad f^{(2)} = \pi^{(2)}_{|y_1} \otimes \nu_1. \tag{2.7}$$

The couple $(f^{(1)}, f^{(2)})$ is a cardinal flow between μ and ν: in fact, given $\phi \in L^1_\mu$, we have

$$\begin{aligned}
\int_X \phi \, d\mu &= \int_{X_2} \left(\int_{X_1} \phi \, d\mu_{|x_2} \right) d\mu_2 \\
&= \int_{X_2} \left(\int_{X_1} \phi \, d\left((\mathfrak{p}_{X_1})_\# \pi^{(1)}_{|x_2} \right) \right) d\mu_2 \\
&= \int_{X_2} \left(\int_{X_1 \times Y_1} \phi \circ \left((\mathfrak{p}_{X_1}), \mathrm{Id}_{X_2} \right) d\pi^{(1)}_{|x_2} \right) d\mu_2 \\
&= \int_{X \times Y_1} \phi \circ (\mathfrak{p}_X) \, df^{(1)} \\
&= \int_X \phi \, d\left((\mathfrak{p}_X)_\# f^{(1)} \right);
\end{aligned}$$

hence, $(\mathfrak{p}_X)_\# f^{(1)} = \mu$. Similarly, we get

$$(\mathfrak{p}_{Y_1 \times X_2})_\# f^{(1)} = \zeta, \quad (\mathfrak{p}_{Y_1 \times X_2})_\# f^{(2)} = \zeta, \quad (\mathfrak{p}_Y)_\# f^{(2)} = \nu;$$

hence, $(f^{(1)}, f^{(2)}) \in \mathscr{F}(\mu,\nu)$. From the identities (2.5) and (2.6), we have

$$\begin{aligned}\int_{X_2} \mathscr{C}_{c_1}(\mu_{|x_2}, \zeta_{|x_2})\, d\mu_2 + \int_{Y_1} \mathscr{C}_{c_2}(\zeta_{|y_1}, \nu_{|y_1})\, d\nu_1 \\ = \int_{X_2}\int_{X_1\times Y_1} c_1\, d\pi^{(1)}_{|x_2} d\mu_2 + \int_{Y_1}\int_{X_2\times Y_2} c_2\, d\pi^{(2)}_{|y_1} d\nu_1 \\ = \int_{X\times Y_1} c_1\, df^{(1)} + \int_{X_2\times Y} c_2\, df^{(2)},\end{aligned}$$

so that

$$\begin{aligned}\int_{X_2} \mathscr{C}_{c_1}(\mu_{|x_2}, \zeta_{|x_2})\, d\mu_2 + \int_{Y_1} \mathscr{C}_{c_2}(\zeta_{|y_1}, \nu_{|y_1})\, d\nu_1 \\ \geq \min_{(f^{(1)}, f^{(2)})} \mathbb{CT}_c(f^{(1)}, f^{(2)}).\end{aligned} \tag{2.8}$$

To prove the other inequality, let us take $(f^{(1)}, f^{(2)}) \in \mathscr{F}(\mu,\nu)$. By definition, we have

$$(\mathfrak{p}_{X_2})_\# f^{(1)} = (\mathfrak{p}_{X_2})_\# \mu = \mu_2 \quad \text{and} \quad (\mathfrak{p}_{Y_1})_\# f^{(2)} = (\mathfrak{p}_{Y_1})_\# \nu = \nu_1.$$

By disintegrating $f^{(1)}$ and $f^{(2)}$ with respect to x_2 and y_1, respectively, we find

$$f^{(1)} = \psi_{|x_2} \otimes \mu_2 \quad \text{and} \quad f^{(2)} = \phi_{|y_1} \otimes \nu_1.$$

Let ζ be the measure on which $f^{(1)}$ and $f^{(2)}$ glue, and we have $\psi_{|x_2} \in \Pi(\mu_{|x_2}, \zeta_{|x_2})$ μ_2-a.e. and $\phi_{|y_1} \in \Pi(\zeta_{|y_1}, \nu_{|y_1})$ ν_1-a.e. Indeed,

$$\mu = (\mathfrak{p}_X)_\# f^{(1)} = (\mathfrak{p}_X)_\#(\psi_{|x_2} \otimes \mu_2) = \big((\mathfrak{p}_X)_\#\psi_{|x_2}\big) \otimes \mu_2$$

so that, by the uniqueness of the conditional law, we have

$$(\mathfrak{p}_X)_\#\psi_{|x_2} = \mu_{|x_2} \quad \text{and} \quad (\mathfrak{p}_{Y_1\times X_2})_\#\psi_{|x_2} = \zeta_{|x_2};$$

therefore, $\psi_{|x_2} \in \Pi(\mu_{|x_2}, \zeta_{|x_2})$, for μ_2-a.e. $x_2 \in X_2$. Similarly, we can find a family of measures such that $\phi_{|y_1} \in \Pi(\zeta_{|y_1}, \nu_{|y_1})$, for ν_1-a.e. $y_1 \in Y_1$. Finally, we observe that

$$\int_{X\times Y_1} c_1\, df^{(1)} = \int_{X_2}\int_{X_1\times Y_1} c_1\, d\psi_{|x_2} d\mu_2 \geq \int_{X_2} \mathscr{C}_{c_1}(\mu_{|x_2}, \zeta_{|x_2})\, d\mu_2 \tag{2.9}$$

and

$$\int_{X_2\times Y} c_2\, df^{(2)} = \int_{Y_1}\int_{X_2\times Y_2} c_2\, d\phi_{|y_1} d\nu_1 \geq \int_{Y_1} \mathscr{C}_{c_2}(\zeta_{|y_1}, \nu_{|y_1})\, d\nu_1. \tag{2.10}$$

By summing up the relations (2.9) and (2.10), we conclude that

$$\mathscr{C}_c(\mu,\nu) \geq \int_{X_2} \mathscr{C}_{c_1}(\mu_{|x_2}, \zeta_{|x_2})\, d\mu_2 + \int_{Y_1} \mathscr{C}_{c_2}(\zeta_{|y_1}, \nu_{|y_1})\, d\nu_1,$$

which, along with (2.8), concludes the proof. ■

REMARK 4 (The Monge minimizers). A Monge minimizer is defined as an optimal transportation plan of the form $\pi_T = (\mathrm{Id}_X, T)_\# \mu$, where $\mathrm{Id}_X : X \to X$ is the identity function over the set X, and $T : X \to Y$ is a measurable function satisfying the condition $T_\# \mu = \nu$. The function T represents the optimal way to transfer the mass from X to Y; that is, the mass that μ assigns to $x \in X$ is relocated to $T(x) \in Y$. We will also say that π_T is induced by T. If the optimal transportation plan is a Monge minimizer, the optimal cardinal flows and the pivot measure associated with it can be described through the function $T = (T_1, T_2)$ as well. Indeed, owing to the chain rule, we have that $f^{(1)} = (\mathfrak{p}_{X \times Y_1})_\#((\mathrm{Id}_X, T)_\# \mu) = (\mathrm{Id}_X, T_1)_\# \mu$. Similarly, it holds that $f^{(2)} = (\mathrm{Id}_{X_2}, T)_\# \mu$ and $\zeta = (T_1, \mathrm{Id}_{X_2})_\# \mu$. Furthermore, according to Theorem 2, there exist two functions, denoted as $H : X \times Y_1 \to Y_1 \times X_2$ and $V : X_2 \times Y_1 \to Y$, such that

(i) $x_2 \to H_2(x_1, x_2) = x_2$ for μ_2-a.e. $x_2 \in X_2$ and $y_1 \to V_1(y_1, x_2) = y_1$ for ν_1-a.e. $y_1 \in Y_1$;

(ii) for μ_2-a.e. $x_2 \in X_2$, the function

$$H_{x_2} : x_1 \to H_1(x_1, x_2)$$

is such that $(H_{x_2}, \mathrm{Id}_{X_2})_\# \mu_{|x_2}$ is an optimal transportation plan between $\mu_{|x_2}$ and $\zeta_{|x_2}$, with respect to c_1. Similarly, for ν_1-a.e. $y_1 \in Y_1$, the function

$$V_{y_1} : x_2 \to V_2(y_1, x_2)$$

is such that $(\mathrm{Id}_{Y_1}, V_{y_1})_\# \zeta_{|y_1}$ is an optimal transportation plan between $\zeta_{|y_1}$ and $\nu_{|y_1}$, with respect to c_2.

Using again the chain rule, it is easy to see that $H(x_1, x_2) = (T_1(x_1, x_2), x_2)$; however, V has no explicit form in terms of T, but it is characterized by the identity $V \circ H = T$. In particular, it holds that $V_1(y_1, x_2) = y_1$ and $V_2(T_1(x_1, x_2), x_2) = T_2(x_1, x_2)$.

When both X and Y are the product of more than two Polish spaces, we can iteratively use Theorem 2 and retrieve the following more general result.

THEOREM 3. *Let* $X = \bigtimes_{i=1}^n X_i$ *and* $Y = \bigtimes_{i=1}^n Y_i$ *be Polish spaces such that* X_i *and* Y_i *are Polish spaces for every* $i = 1, \dots, n$. *Moreover, let* $c : X \times Y \to \mathbb{R}$ *be a separable cost function, that is,* $c(x, y) = \sum_{i=1}^n c_i(x_i, y_i)$. *Then, given* $\mu \in \mathcal{P}(X)$ *and* $\nu \in \mathcal{P}(Y)$, *there exists a family of* $(n+1)$ *probability measures, namely,* $\{\zeta^{(i)}\}_{i=0,1,\dots,n}$, *such that*

- $\zeta^{(0)} = \mu$ *and* $\zeta^{(n)} = \nu$,
- $\zeta^{(i)} \in \mathcal{P}((\bigtimes_{j=1}^{i} Y_j) \times (\bigtimes_{j=i+1}^{n} X_j))$, *for every* $i = 1, 2, \dots, n-1$,
- $\mathcal{C}_c(\mu, \nu) = \sum_{i=1}^{n} \int_{(\bigtimes_{j=1}^{i-1} Y_j)\times(\bigtimes_{j=i+1}^{n} X_j)} \mathcal{C}_{c_i}(\zeta_{|-i}^{(i-1)}, \zeta_{|-i}^{(i)})\, d\zeta_{-i}^{(i-1)}$,

where $\zeta_{-i}^{(i-1)}$ *is the marginal of* $\zeta^{(i-1)}$ *over* $(\bigtimes_{j=1}^{i-1} Y_j) \times (\bigtimes_{j=i+1}^{n} X_j)$ *and* $\zeta_{|-i}^{(i-1)}$ *is the conditional law of* $\zeta^{(i-1)}$ *given*

$$z_{-i} = (y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n) \in \left(\bigtimes_{j=1}^{i-1} Y_j\right) \times \left(\bigtimes_{j=i+1}^{n} X_j\right).$$

Similarly, $\zeta_{|-i}^{(i)}$ *is the conditional law of the marginal* $\zeta^{(i)}$ *given*

$$z_{-i} \in \left(\bigtimes_{j=1}^{i-1} Y_j\right) \times \left(\bigtimes_{j=i+1}^{n} X_j\right).$$

PROOF. For the sake of simplicity, we prove the statement for $n = 3$. Let us set $X_1' = X_1$, $X_2' = X_2 \times X_3$, $Y_1' = Y_1$, and $Y_2' = Y_2 \times Y_3$. Moreover, let us consider the functions $c_1' = c_1$ and $c_2' = c_2 + c_3$, so that $c = c_1' + c_2'$. From Theorem 2, we find a measure $\zeta \in \mathcal{P}(Y_1' \times X_2')$ such that

$$\mathcal{C}_c(\mu, \nu) = \int_{X_2'} \mathcal{C}_{c_1'}(\mu_{|x_2'}, \zeta_{|x_2'})\, d\mu_2 + \int_{Y_1'} \mathcal{C}_{c_2'}(\zeta_{|y_1'}, \nu_{|y_1'})\, d\nu_1, \tag{2.11}$$

where μ_2 is the marginal of μ over $X_2' = X_2 \times X_3$ and ν_1 is the marginal of ν over $Y_1' = Y_1$. For ν_1-almost every y_1', we have $\zeta_{|y_1'} \in \mathcal{P}(X_2')$ and $\nu_{|y_1'} \in \mathcal{P}(Y_2')$. Then, since $X_2' = X_2 \times X_3$, $Y_2' = Y_2 \times Y_3$, and $c_2' = c_2 + c_3$, we can apply Theorem 2 again and find a family of measures depending on y_1', namely, $\rho^{(y_1')}$, such that $\rho^{(y_1')} \in \mathcal{P}(Y_2 \times X_3)$ and

$$\begin{aligned} \mathcal{C}_{c_2'}(\zeta_{|y_1'}, \nu_{|y_1'}) = &\int_{X_3} \mathcal{C}_{c_2}\big((\zeta_{|y_1'})_{|x_3}, (\rho^{(y_1')})_{|x_3}\big)\, d(\zeta_{|y_1'})_3 \\ &+ \int_{Y_2} \mathcal{C}_{c_3}\big((\rho^{(y_1')})_{|x_2}, (\nu_{|y_1'})_{|y_2}\big)\, d(\nu_{|y_1'})_2, \end{aligned} \tag{2.12}$$

for ν_1-almost everywhere. We then define

$$\rho = \rho^{(y_1')} \otimes \nu_1 \in \mathcal{P}(Y_1 \times Y_2 \times X_3).$$

Notice that ρ is well defined since the family of measures $\{\rho^{(y_1')}\}_{y_1' \in Y_1'}$ is measurable (it follows from [35, Corollary 5.21] and from the fact that being measurable is preserved by the pushforward operation). Then, by combining (2.12) with (2.11), we find

$$\begin{aligned}\mathscr{C}_c(\mu,\nu) = &\int_{X_2\times X_3} \mathscr{C}_{c_1}(\mu_{|(x_2,x_3)}, \zeta_{|(x_2,x_3)})\, d\mu_{-1} \\ &+ \int_{Y_1\times X_3} \mathscr{C}_{c_2}(\zeta_{|(y_1,x_3)}, \rho_{|(y_1,x_3)})\, d\zeta_{-2} \\ &+ \int_{Y_1\times Y_2} \mathscr{C}_{c_3}(\rho_{|(y_1,y_2)}, \nu_{|(y_1,y_2)})\, d\rho_{-3}.\end{aligned}$$

Moreover, by construction, $\zeta \in \mathcal{P}(Y_1' \times X_2') = \mathcal{P}(Y_1 \times X_2 \times X_3)$. We therefore conclude the thesis by setting $\zeta^{(0)} = \mu$, $\zeta^{(1)} = \zeta$, $\zeta^{(2)} = \rho$, $\zeta^{(3)} = \nu$, and by recalling that $x_1' = x_1$, $y_1' = y_1$, and $c_1' = c_1$. The result for $n \geq 3$ is recovered by applying the previous argument $n-1$ times. ■

From formula (2.3), we deduce that solving the optimal transport problem between two generic measures can be achieved in two steps: detecting the pivotal measure and solving a family of lower dimensional problems. In particular, if we are able to solve the lower dimensional transportation problems, the only unknown left to determine is the pivot measure.

DEFINITION 9 (Pivoting functional). Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and a separable cost function $c = c_1 + c_2$, we define the pivoting functional $\mathbb{Z}_c : \mathcal{I}(\mu,\nu) \to \mathbb{R}$ as

$$\mathbb{Z}_c : \zeta \to \int_{X_2} \mathscr{C}_{c_1}(\mu_{|x_2}, \zeta_{|x_2})\, d\mu_2 + \int_{Y_1} \mathscr{C}_{c_2}(\zeta_{|y_1}, \nu_{|y_1})\, d\nu_1.$$

Due to the convexity and the lower semi-continuity of the Wasserstein costs (see [35, Chapter 4]), we infer the following lemma.

LEMMA 2. *For every $\mu \in \mathcal{P}(X)$ and every $\nu \in \mathcal{P}(Y)$, the functional $\mathbb{Z}_c$ is convex and lower semi-continuous over $\mathcal{I}(\mu,\nu)$. Therefore, $\mathbb{Z}_c$ admits a minimizer over $\mathcal{I}(\mu,\nu)$.*

To conclude, we prove that the minimum of $\mathbb{Z}_c$ over $\mathcal{I}(\mu,\nu)$ is the Wasserstein cost between μ and ν.

THEOREM 4. *Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$, and a separable cost function $c = c_1 + c_2$, it holds that*

$$\mathscr{C}_c(\mu,\nu) = \min_{\zeta\in\mathcal{I}(\mu,\nu)} \mathbb{Z}_c(\zeta).$$

PROOF. Since any pivot measure ζ is an element of $\mathcal{I}(\mu,\nu)$, Theorem 2 ensures that

$$\mathscr{C}_c(\mu,\nu) \geq \min_{\zeta\in\mathcal{I}(\mu,\nu)} \mathbb{Z}_c(\zeta).$$

To conclude the proof, we prove the other inequality.

Let us fix $\zeta \in \mathcal{I}(\mu, \nu)$. Following the steps of the proof of Theorem 2, we disintegrate μ, ζ, and ν (see (2.3)–(2.4)), find the optimal transportation plans between the conditional measures, and define the cardinal flow as done in (2.7). Finally, since the couple $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$, we have

$$\mathbb{Z}_c(\zeta) = \mathbb{C}\mathbb{T}_c(f^{(1)}, f^{(2)}) \geq \min_{(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)} \mathbb{C}\mathbb{T}_c(f^{(1)}, f^{(2)});$$

hence,

$$\min_{\zeta \in \mathcal{I}(\mu,\nu)} \mathbb{Z}_c(\zeta) \geq \min_{(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu,\nu)} \mathbb{C}\mathbb{T}_c(f^{(1)}, f^{(2)}) = \mathscr{C}_c(\mu, \nu),$$

which concludes the proof. ∎

2.1. *Independence of the optimal coupling*

In Remark 2, we noticed that any optimal transportation plan determines one and only one optimal cardinal flow, while the opposite is not true. The following example showcases how, even if we have a unique pivot measure and a unique optimal cardinal flow, we might retrieve an infinite number of optimal transportation plans.

EXAMPLE 1. Let us take two probability measures on $\mathbb{R}^2$, μ and ν, defined as

$$\mu := \frac{1}{2}(\delta_{(1,0)} + \delta_{(2,0)}) \quad \text{and} \quad \nu := \frac{1}{2}(\delta_{(0,1)} + \delta_{(0,2)}),$$

where $\delta_{(x_1,x_1)}$ is the Dirac delta centered in $(x_1, x_2) \in \mathbb{R}^2$. Since $\mathcal{I}(\mu, \nu) = \{\delta_{(0,0)}\}$, the only possible pivot measure is $\zeta = \delta_{(0,0)}$; hence, the only (and therefore optimal) cardinal flow is

$$f^{(1)} := \frac{1}{2}(\delta_{((1,0);0)} + \delta_{((2,0);0)}) \quad \text{and} \quad f^{(2)} := \frac{1}{2}(\delta_{((0,0);1)} + \delta_{((0,0);2)}).$$

However, the transportation plans π, π_1, and π_2, defined as

$$\pi := \frac{1}{4}(\delta_{((1,0);(0,1))} + \delta_{((1,0);(0,2))} + \delta_{((2,0);(0,1))} + \delta_{((2,0);(0,2))}),$$
$$\pi_1 := \frac{1}{2}(\delta_{((1,0);(0,1))} + \delta_{((2,0);(0,2))}),$$
$$\pi_2 := \frac{1}{2}(\delta_{((1,0);(0,2))} + \delta_{((2,0);(0,1))}),$$

are all optimal transportation plans between μ and ν since they all induce the same cardinal flow $(f^{(1)}, f^{(2)})$. Moreover, it is worthy of notice that since $\mathbb{T}_c$ is convex, any convex combination of π_1 and π_2 is an optimal transportation plan between μ and ν.

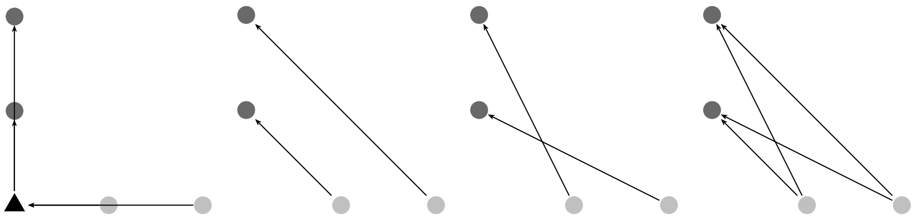

FIGURE 2. The lack of uniqueness we showcased in Example 1. The support of μ is indicated by light grey circles, the support of ν by dark grey circles. In Figure (A), we showcase the optimal cardinal flow. The support of the pivot measure ζ is indicated by the black triangle. Figures (B), (C), and (D) showcase the transportation plans π_1, π_2, and π, respectively. Each of these transportation plans induces the cardinal flow described in (A).

The lack of uniqueness is due to a natural lack of memory. Roughly speaking, once the first cardinal flow $f^{(1)}$ allocates the mass into $(0,0)$, the mass coming from $(1,0)$ and $(2,0)$ merges into one point and loses its identity. Therefore, when the second cardinal flow $f^{(2)}$ reallocates the mass in $(0,0)$ and moves it vertically to complete the transportation, we are unable to tell how much of the mass that ended in $(0,1)$ came from the point $(1,0)$ or $(2,0)$. The plans π, π_1, and π_2 are different for this reason: for π, just half of the mass in $(1,0)$ goes to $(0,1)$, for π_1, all the mass in $(1,0)$ goes to $(0,1)$, and, for π_2, none of the mass in $(1,0)$ goes to $(0,1)$ (see Figure 2).

As Example 1 shows, for any pivot measure ζ, there might exist more than one optimal transportation plan whose marginal on $Y_1 \times X_2$ is equal to ζ. In the following, we characterize this lack of uniqueness and prove that, given a pivot measure ζ, there always exists a unique optimal transportation plan π that induces ζ and enjoys a conditional independence property with respect to its marginals.

LEMMA 3. *Let us take $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to [0, +\infty)$ a separable cost function. Moreover, let $(f^{(1)}, f^{(2)})$ be an optimal cardinal flow, let ζ be the pivotal measure related to $(f^{(1)}, f^{(2)})$, and let $f^{(1)}_{|(y_1,x_2)}$ and $f^{(2)}_{|(y_1,x_2)}$ be the conditional laws of $f^{(1)}$ and $f^{(2)}$ given (y_1, x_2). Then, any measurable family of probability measures $\gamma_{(y_1,x_2)}$ satisfying*

$$\gamma_{(y_1,x_2)} \in \Pi(f^{(1)}_{|(y_1,x_2)}, f^{(2)}_{|(y_1,x_2)}), \quad \zeta\text{-a.e. on } Y_1 \times X_2, \tag{2.13}$$

is such that

$$\gamma_{(y_1,x_2)} \otimes \zeta \in \Gamma_o(\mu, \nu).$$

PROOF. Let $\pi \in \Pi(\mu, \nu)$ be defined as in (2.13). Since it holds that $L(\pi) = (f^{(1)}, f^{(2)})$ and $(f^{(1)}, f^{(2)})$ is optimal, we conclude that $\pi \in \Gamma_o(\mu, \nu)$. ∎

THEOREM 5. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and let $c = c_1 + c_2$ be a separable cost function. Then, for any given pivot measure ζ, there exists a unique optimal transportation plan π such that $(\mathfrak{p}_{Y_1 \times X_2})_{\#}\pi = \zeta$ and such that the conditional law of π given (y_1, x_2) is an independent measure for ζ-a.e. on $Y_1 \times X_2$. Moreover, the plan π is given by*

$$\pi := \big(f^{(1)}_{|(y_1,x_2)} \otimes f^{(2)}_{|(y_1,x_2)}\big) \otimes \zeta \tag{2.14}$$

where $(f^{(1)}, f^{(2)}) = L(\pi)$.

PROOF. It follows from the fact that $\{f^{(1)}_{|(y_1,x_2)} \otimes f^{(2)}_{|(y_1,x_2)}\}_{(y_1,x_2)\in Y_1\times X_2}$ is the family of measures that satisfies the requirements of Lemma 5 and it is independent ζ-a.e. over $Y_1 \times X_2$. ■

REMARK 5. Going back to Example 1, the optimal transportation plan described in Theorem 5 is

$$\pi := \frac{1}{4}\big(\delta_{((1,0);(0,1))} + \delta_{((1,0);(0,2))} + \delta_{((2,0);(0,1))} + \delta_{((2,0);(0,2))}\big).$$

When the optimal transportation plan is unique, we can enhance the previous result as follows.

COROLLARY 2. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and let $c = c_1 + c_2$ be a separable cost function. Let us assume that the transportation problem between μ and ν has a unique solution π. Then, if (X_1, X_2, Y_1, Y_2) is the optimal coupling inducing the law π, we have the following:*

(1) *X_1 and Y_2 are conditionally independent given X_2 and Y_1,*

(2) *X_2 and Y_1 are conditionally independent given X_1 and Y_2.*

PROOF. Let π be the optimal transportation plan, $(f^{(1)}, f^{(2)}) = L(\pi)$, and let ζ be the pivot measure. Since the plan (2.14) is optimal, we have

$$\pi = \big(f^{(1)}_{|(y_1,x_2)} \otimes f^{(2)}_{|(y_1,x_2)}\big) \otimes \zeta.$$

By the uniqueness of the disintegration, we find $\pi_{|(y_1,x_2)} = f^{(1)}_{|(y_1,x_2)} \otimes f^{(2)}_{|(y_1,x_2)}$, which proves (1). To prove (2), it suffice to swap the roles of μ and ν. ■

COROLLARY 3. *Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. If there exists $\bar{x}_2 \in X_2$ for which*

$$\mu_2\big(\{\bar{x}_2\}\big) = \mu\big(X_1 \times \{\bar{x}_2\}\big) = 1, \tag{2.15}$$

then $\zeta = \nu_1 \otimes \delta_{\bar{x}_2}$ is the unique pivot measure between μ and ν. Similarly, if there exists $\bar{y}_1$ such that

$$\nu_1\big(\{\bar{y}_1\}\big) = \nu\big(\{\bar{y}_1\} \times Y_2\big) = 1, \tag{2.16}$$

the unique pivot measure is given by $\zeta = \delta_{\bar{y}_1} \otimes \mu_2$. *Moreover, if there exists a couple* $(\bar{y}_1, \bar{x}_2)$ *satisfying both* (2.15) *and* (2.16), *then* $\zeta = \delta_{(\bar{y}_1, \bar{x}_2)}$ *is the pivot measure and every* $\pi \in \Pi(\mu, \nu)$ *is optimal.*

PROOF. The first two statements follow from the fact that $\mathcal{I}(\mu, \nu)$ contains only one measure, which is $\nu_1 \otimes \delta_{\bar{x}_2}$ in the first case and $\delta_{\bar{y}_1} \otimes \mu_2$ in the second one. If both (2.15) and (2.16) hold, also $\mathcal{F}(\mu, \nu)$ contains only one element,

$$(f^{(1)}, f^{(2)}) = (\mu \otimes \delta_{\bar{y}_1}, \delta_{\bar{x}_2} \otimes \nu).$$

In particular, $L(\pi) = (f^{(1)}, f^{(2)})$ for each $\pi \in \Pi(\mu, \nu)$, and therefore each $\pi \in \Pi(\mu, \nu)$ is optimal. ∎

We conclude the section by showing that if both μ and ν are independent measures, that is, $\mu = \mu_1 \otimes \mu_2$ and $\nu = \nu_1 \otimes \nu_2$, Corollary 2 can be enhanced further as it follows.

COROLLARY 4. *In the framework of Corollary* 2, *assume that both* μ *and* ν *are independent, that is,* $\mu = \mu_1 \otimes \mu_2$ *and* $\nu = \nu_1 \otimes \nu_2$. *Then, there exists an optimal transportation plan between* μ *and* ν, *namely,* π, *such that if* (X_1, X_2, Y_1, Y_2) *is the coupling inducing the law* π, *the couple* (X_1, Y_1) *is independent from* (X_2, Y_2). *Moreover, we have*

$$\mathscr{C}_c(\mu, \nu) = \mathscr{C}_{c_1}(\mu_1, \nu_1) + \mathscr{C}_{c_2}(\mu_2, \nu_2).$$

PROOF. Let us define the intermedium measure $\zeta := \nu_1 \otimes \mu_2$. By definition, ζ is independent and therefore we have that $\mu_{|x_2} = \mu_1$ and $\zeta_{|x_2} = \nu_1$ for any $x_2 \in X_2$. Let $\pi^{(1)}$ be the optimal transportation plan between μ_1 and ν_1 with respect to c_1, and let us define $f^{(1)} := \pi^{(1)} \otimes \mu_2$. Similarly, we define $f^{(2)} := \pi^{(2)} \otimes \nu_1$, where $\pi^{(2)}$ is the optimal transportation plan between μ_2 and ν_2 with respect to c_2.

It is easy to see that $(f^{(1)}, f^{(2)})$ is a cardinal flow between μ and ν that glues on ζ. Following formula (2.14), the measure defined as

$$\pi = \left(f^{(1)}_{|(y_1, x_2)} \otimes f^{(2)}_{|(y_1, x_2)}\right) \otimes \zeta \tag{2.17}$$

is a transportation plan between μ and ν. Moreover, according to the definitions of ζ, $f^{(1)}$, and $f^{(2)}$, equation (2.17) boils down to

$$\pi = \left(f^{(1)}_{|y_1} \otimes f^{(2)}_{|x_2}\right) \otimes (\nu_1 \otimes \mu_2) = \pi^{(1)} \otimes \pi^{(2)}.$$

We now show that π is c-cyclically monotone, which implies that π is an optimal transportation plan (see [31, Theorem 1.49]). By contradiction, let us assume that there exists $N \in \mathbb{N}$ and a set of points

$$\left\{(x^{(i)}, y^{(i)})\right\}_{i=1,2,\ldots,N} = \left\{(x_1^{(i)}, x_2^{(i)}, y_1^{(i)}, y_2^{(i)})\right\}_{i=1,2,\ldots,N} \subset \operatorname{spt}(\pi)$$

such that

$$\sum_{i=1}^{N} c(x^{(i)}, y^{(i+1)}) < \sum_{i=1}^{N} c(x^{(i)}, y^{(i)}), \tag{2.18}$$

where $x^{(N+1)} = x^{(1)}$ and $y^{(N+1)} = y^{(1)}$. Since $c = c_1 + c_2$ is separable, from (2.18), we infer

$$\sum_{i=1}^{N} c(x^{(i)}, y^{(i+1)}) = \sum_{i=1}^{N} \left(c_1(x_1^{(i)}, y_1^{(i+1)}) + c_2(x_2^{(i)}, y_2^{(i+1)})\right)$$

and, similarly,

$$\sum_{i=1}^{N} c(x^{(i)}, y^{(i)}) = \sum_{i=1}^{N} \left(c_1(x_1^{(i)}, y_1^{(i)}) + c_2(x_2^{(i)}, y_2^{(i)})\right).$$

Therefore, we rewrite (2.18) as

$$\sum_{i=1}^{N} c_1(x_1^{(i)}, y_1^{(i+1)}) + \sum_{i=1}^{N} c_2(x_2^{(i)}, y_2^{(i+1)}) < \sum_{i=1}^{N} c_1(x_1^{(i)}, y_1^{(i)}) + \sum_{i=1}^{N} c_2(x_2^{(i)}, y_2^{(i)}),$$

which allows us to conclude that either one of the two inequalities must hold:

$$\begin{aligned} &\sum_{i=1}^{N} c_1(x_1^{(i)}, y_1^{(i+1)}) < \sum_{i=1}^{N} c_1(x_1^{(i)}, y_1^{(i)}) \quad \text{or} \\ &\sum_{i=1}^{N} c_2(x_2^{(i)}, y_2^{(i+1)}) < \sum_{i=1}^{N} c_2(x_2^{(i)}, y_2^{(i)}). \end{aligned} \tag{2.19}$$

By definition of π, we have $\{(x_1^{(i)}, x_2^{(i)}, y_1^{(i)}, y_2^{(i)})\}_{i=1,2,\dots,N} \subset \mathrm{spt}(\pi)$ if and only if

$$\{(x_1^{(i)}, y_1^{(i)})\}_{i=1,2,\dots,N} \subset \mathrm{spt}(\pi^{(1)}) \quad \text{and} \quad \{(x_2^{(i)}, y_2^{(i)})\}_{i=1,2,\dots,N} \subset \mathrm{spt}(\pi^{(2)}).$$

We therefore conclude that both inequalities in (2.19) can not hold, as the support of $\pi^{(1)}$ is c_1-cyclically monotone and the support of $\pi^{(2)}$ is c_2-cyclically monotone.

Finally, we notice that, in this framework, Formula (2.2) simplifies to

$$\begin{aligned} \mathscr{C}_c(\mu, \nu) &= \int_{X_2} \mathscr{C}_{c_1}(\mu_{|x_2}, \zeta_{|x_2})\, d\mu_2 + \int_{Y_1} \mathscr{C}_{c_2}(\zeta_{|y_1}, \nu_{|y_1})\, d\nu_1 \\ &= \int_{X_2} \mathscr{C}_{c_1}(\mu_1, \nu_1)\, d\mu_2 + \int_{Y_1} \mathscr{C}_{c_2}(\mu_2, \nu_2)\, d\nu_1 \\ &= \mathscr{C}_{c_1}(\mu_2, \nu_2) + \mathscr{C}_{c_2}(\mu_2, \nu_2), \end{aligned}$$

which concludes the proof. ■

3. Two examples

We conclude our paper with two examples. In the first one, the cost function is a separable distance over a polish space $X = X_1 \times X_2$, i.e., $c := d_1 + d_2$, where $d_1 : X_1 \times X_1 \to [0, +\infty)$ is a distance over X_1 and $d_2 : X_2 \times X_2 \to [0, +\infty)$ is a distance over X_2. In the second one, the measures are supported over $\mathbb{R}^2$ and the cost function c has the form $c(x, y) = h(|x_1 - y_1|) + h(|x_2 - y_2|)$, where $h : \mathbb{R} \to \mathbb{R}$ is a convex function such that $h(0) = 0$. Albeit the second example is presented in $\mathbb{R}^2$, Theorem 3 allows us to extend all the results to the case $\mathbb{R}^n$ with $n \geq 2$.

3.1. Wasserstein distance

If we take $X = Y$ and choose the distance d as a cost function, the optimal transport problem lifts the distance d over the space $\mathcal{P}_p(X)$. The resulting distance is called the Wasserstein distance.

Definition 10 (Wasserstein distance). Let (X, d) be a Polish space and $p \in [1, +\infty)$. The p-order Wasserstein distance between the probability measures μ and ν on X is defined as

$$W^p_{d^p}(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \mathbb{T}_{d^p}(\pi). \tag{3.1}$$

When $p = 1$, the 1-Wasserstein distance is also known as the Kantorovich–Rubistein distance.

Remark 6. The infimum in (3.1) could actually be $+\infty$, to avoid that it suffices to restrict W_p to $\mathcal{P}_p(X)$.

Theorem 6 ([35, Chapter 6]). *The W_{d^p} distance is a finite distance over $\mathcal{P}_p(X)$. When the set X is bounded, the W_{d^p} distance induces the weak topology on the space $\mathcal{P}_p(X)$. Moreover, $(\mathcal{P}_p(X), W_{d^p})$ is a Polish space.*

We now prove that when d is separable, the induced distance W_d inherits a weaker version of the separability, that is,

$$W_d(\mu, \nu) = W_d(\mu, \zeta) + W_d(\zeta, \nu), \tag{3.2}$$

for any pivot measure ζ.

Since $X = Y$, we need to slightly change the notation in order to avoid confusion. We denote the generic point $(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) \in X \times X$ with

$$(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = \big((x_1^{(1)}, x_2^{(1)}), (x_1^{(2)}, x_2^{(2)})\big);$$

hence, we denote with $\boldsymbol{x}^{(i)}$ the i-th component in the space $X \times X$ and we denote with $x_j^{(i)}$ the j-th component of $\boldsymbol{x}^{(i)}$. The projections $(\mathfrak{p})^{(i)} : X \times X \to X$ and $p_j^{(i)} : X \times X \to X_j$ are defined as

$$(\mathfrak{p})^{(i)}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) := \boldsymbol{x}^{(i)} \quad \text{and} \quad p_j^{(i)}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) := x_j^{(i)}$$

for $i = 1, 2$ and $j = 1, 2$. In particular, we have

$$(\mathfrak{p})^{(i)} = (p_1^{(i)}, p_2^{(i)}).$$

Theorem 7. *Let us take $\mu, \nu \in \mathcal{P}(X)$, where $X = X_1 \times X_2$, and let d be a separable distance over X, i.e.,*

$$d(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}) = d_1(x_1^{(1)}, x_1^{(2)}) + d_2(x_2^{(1)}, x_2^{(2)}), \quad \forall \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in X,$$

where $d_1 : X_1 \times X_1 \to \mathbb{R}$ and $d_2 : X_2 \times X_2 \to \mathbb{R}$ are two distances. Then, for any pivot measure $\zeta \in \mathcal{I}(\mu, \nu)$, we have

$$W_d(\mu, \nu) = W_d(\mu, \zeta) + W_d(\zeta, \nu).$$

Proof. Since W_d is a distance over $\mathcal{P}(X)$, by the triangular inequality, we have

$$W_d(\mu, \nu) \leq W_d(\mu, \zeta) + W_d(\zeta, \nu), \tag{3.3}$$

for any $\zeta \in \mathcal{P}(X)$ and, in particular, for any pivot measure ζ.

To prove the other inequality, let $(f^{(1)}, f^{(2)}) \in \mathcal{F}(\mu, \nu)$ be an optimal cardinal flow. We define

$$\pi^{(1)} = \big((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)})\big)_{\#} f^{(1)} \quad \text{and} \quad \pi^{(2)} = \big((p_1^{(2)}, p_2^{(1)}), (\mathfrak{p})^{(2)}\big)_{\#} f^{(2)}.$$

By definition, we have

$$(\mathfrak{p})^{(1)} \circ \big((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)})\big) = (p_1^{(1)}, p_2^{(1)}) \tag{3.4}$$

and

$$(\mathfrak{p})^{(2)} \circ \big((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)})\big) = (p_1^{(2)}, p_2^{(1)}). \tag{3.5}$$

Through the relations (3.4), (3.5), and the chain rule for the push-forwards, we get

$$(\mathfrak{p})_{\#}^{(1)} \pi^{(1)} = \quad (p_1^{(1)}, p_2^{(1)})_{\#}\Big(\big((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)})\big)_{\#} f^{(1)}\Big) = (p_1^{(1)}, p_2^{(1)})_{\#} f^{(1)} = \mu$$

and

$$(\mathfrak{p})_{\#}^{(2)} \pi^{(1)} = \zeta;$$

hence, $\pi^{(1)} \in \Pi(\mu, \zeta)$. Similarly, we get $\pi^{(2)} \in \Pi(\zeta, \nu)$. Finally, by Theorem 1, we have that

$$\begin{aligned} W_d(\mu,\nu) &= \int_{X\times X_1} d_1\, df^{(1)} + \int_{X_2\times X} d_2\, df^{(2)} \\ &= \int_{X\times X_1} d\circ\big((\mathfrak{p})^{(1)}, (p_1^{(2)}, p_2^{(1)})\big) df^{(1)} \\ &\quad + \int_{X_2\times X} d\circ\big((p_1^{(2)}, p_2^{(1)}), (\mathfrak{p})^{(2)}\big) df^{(2)} \\ &= \int_{X\times X} d\, d\pi^{(1)} + \int_{X\times X} d\, d\pi^{(2)} \\ &\geq W_d(\mu,\zeta) + W_d(\zeta,\nu). \end{aligned}$$

The latter inequality, in conjunction with (3.3), allows us to conclude that

$$W_d(\mu,\nu) = W_d(\mu,\zeta) + W_d(\zeta,\nu)$$

for any pivot measure ζ. ■

It is easy to see that when both μ and ν are supported over $X = \bigtimes_{i=1}^n X_i$ and the distance d is separable, that is,

$$d(x,y) = \sum_{i=1}^n d_i(x_i, y_i),$$

we can combine Theorems 3 and 7 to generalize equation (3.2) to

$$W_d(\mu,\nu) = \sum_{i=1}^n W_{d_i}(\zeta^{(i-1)}, \zeta^{(i)}),$$

where $\{\zeta^{(i)}\}_{i=0,1,\dots,n}$ is a suitable family of measures.

Remark 7. As a straightforward consequence of Theorem 7, we infer that any pivot measure is a minimizer of the functional

$$\Theta : \lambda \to W_d(\mu,\lambda) + W_d(\lambda,\nu),$$

over the space $\mathcal{P}(X)$. However, the reverse implication is not true. Let us consider, for instance,

$$X = \mathbb{R}^2, \quad \mu = \frac{1}{2}[\delta_{(0,0)} + \delta_{(7,1)}], \quad \nu = \frac{1}{2}[\delta_{(1,1)} + \delta_{(8,0)}].$$

It is easy to see that the only pivot measure is

$$\zeta = \frac{1}{2}[\delta_{(1,0)} + \delta_{(8,1)}].$$

Moreover, since $\mu_2 = \nu_2$, we have $\nu \in \mathcal{I}(\mu, \nu)$ and, therefore,

$$W_d(\mu, \nu) = \inf_{\lambda \in \mathcal{I}(\mu,\nu)} W_d(\mu, \lambda) + W_d(\lambda, \nu) \leq W_d(\mu, \nu) + W_d(\nu, \nu) = W_d(\mu, \nu);$$

hence, ν minimizes Θ although it is not the pivot measure.

3.2. Cardinal flow in the Euclidean space

When both the measures μ and ν are supported on $\mathbb{R}$ and the cost function c is convex, the solution is unique and is characterized by the pseudo-inverse function of μ and ν.

DEFINITION 11. Given $\mu, \nu \in \mathcal{P}(\mathbb{R})$, the co-monotone transportation plan γ_{mon} between μ and ν is defined as

$$\gamma_{\text{mon}} := (F_\mu^{[-1]}, F_\nu^{[-1]})_\# \mathcal{L}_{|[0,1]},$$

where $F_\mu^{[-1]}$ and $F_\nu^{[-1]}$ are the pseudo-inverse of the cumulative functions of μ and ν, respectively, and $\mathcal{L}_{|[0,1]}$ is the Lebesgue measure restricted on $[0, 1]$.

THEOREM 8 ([31, Chapter 2, Theorem 2.9]). *Let $h : \mathbb{R} \to \mathbb{R}_+$ be a strictly convex function such that $h(0) = 0$ and $\mu, \nu \in \mathcal{P}(\mathbb{R})$. Consider the cost*

$$c(x, y) = h\big(|x - y|\big)$$

and suppose that this cost is feasible for the transportation problem. Then, the optimal transport problem has a unique solution which is γ_{mon}.

REMARK 8. We recall that if h is convex, but not strictly convex, the monotone transportation plan is still a solution of the optimal transport problem; however, it might not be the only solution.

A straightforward consequence of Theorem 8 allows us to compute the Wasserstein cost through the pseudo-inverse of the cumulative functions of μ and ν.

COROLLARY 5 ([31, Chapter 2, Proposition 2.17]). *Let us take $\mu, \nu \in \mathcal{P}(\mathbb{R})$. If $c(x, y) = |x - y|^p$, with $p \geq 1$, then*

$$W_p^p(\mu, \nu) = \int_{[0,1]} \big|F_\mu^{[-1]} - F_\nu^{[-1]}\big|^p d\mathcal{L},$$

where $\mathcal{L}$ is the Lebesgue measure over $[0, 1]$. Moreover, for $p = 1$, we have

$$W_1(\mu, \nu) = \int_{\mathbb{R}} \big|F_\mu(t) - F_\nu(t)\big| dt.$$

Corollary 5 allows us to rewrite (2.2) in terms of pseudo-inverse function if $c = c_1 + c_2$, with

$$c_i(x_i, y_i) = h\big(|x_i - y_i|\big),$$

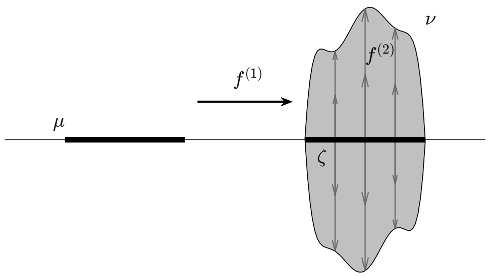

FIGURE 3. The cardinal flow found in Theorem 9.

where $h : \mathbb{R} \to [0, +\infty)$ is a convex function such that $h(0) = 0$.[1] In particular, it holds that

$$W_c(\mu, \nu) = \int_{\mathbb{R}} \int_{[0,1]} h\Big(\big|F^{[-1]}_{\mu_{|x_2}}(s) - F^{[-1]}_{\zeta_{|x_2}}(s)\big|\Big) ds\, d\mu_2 + \int_{\mathbb{R}} \int_{[0,1]} h\Big(\big|F^{[-1]}_{\zeta_{|y_1}}(t) - F^{[-1]}_{\nu_{|y_1}}(t)\big|\Big) dt\, d\nu_1,$$

where $F^{[-1]}_{\mu_{|x_2}}, F^{[-1]}_{\zeta_{|x_2}}, F^{[-1]}_{\zeta_{|y_1}}$, and $F^{[-1]}_{\nu_{|y_1}}$ are the pseudo-inverse functions of the cumulative function of $\mu_{|x_2}, \zeta_{|x_2}, \zeta_{|y_1}$, and $\nu_{|y_1}$, respectively.

THEOREM 9. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$ and assume that μ satisfies*

$$\mu\big(\{x_2 = 0\}\big) = 1.$$

Then, the optimal cardinal flow $(f^{(1)}, f^{(2)})$ between μ and ν is defined as

$$f^{(1)} := (F^{[-1]}_{\mu_1}, F^{[-1]}_{\nu_1})_{\#}\mathcal{L}_{|[0,1]} \otimes \delta_0 \tag{3.6}$$

and

$$f^{(2)} := \big(\delta_0 \otimes (F^{[-1]}_{\nu_{|y_1}})_{\#}\mathcal{L}_{|[0,1]}\big) \otimes \nu_1. \tag{3.7}$$

In particular, we have

$$\begin{aligned} \mathcal{C}_c(\mu, \nu) &= \mathcal{C}_{c_1}(\mu_1, \nu_1) + \int_{\mathbb{R}} \mathcal{C}_{c_2}(\delta_0, \nu_{|y_1})\, d\nu_1 \\ &= \mathcal{C}_{c_1}(\mu_1, \nu_1) + \int_{\mathbb{R}} \int_{\mathbb{R}} c_2(0, y_2)\, d\nu_{|y_1} d\nu_1 \\ &= \mathcal{C}_{c_1}(\mu_1, \nu_1) + \int_{\mathbb{R}^2} c_2(0, y_2)\, d\nu. \end{aligned} \tag{3.8}$$

(1) Notice that we do not need c_1 and c_2 to be induced by the same function h. Hence, all the results are valid whenever $c(x, y) = h_1(|x_1 - y_1|) + h_2(|x_2 - y_2|)$ where $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ are convex and such that $h_1(0) = h_2(0) = 0$.

PROOF. From Corollary 3, we have that the pivot measure between μ and ν is

$$\zeta = \nu_1 \otimes \delta_0.$$

Then, we have

$$f^{(1)} = \gamma^{(1)}_{|x_2} \otimes \delta_0 \quad \text{and} \quad f^{(2)} = \gamma^{(2)}_{|y_1} \otimes \nu_1,$$

where $\gamma^{(1)}_{|x_2}$ is the measurable family of optimal transportation plans between $\mu_{|x_2} = \mu_1$ and $\zeta_{|x_2} = \nu_1$. Similarly, $\gamma^{(2)}_{|y_1}$ is the measurable family of optimal transportation plans between $\zeta_{|y_1} = \delta_0$ and $\nu_{|y_1}$. Finally, using Theorem 8, we retrieve the identities in (3.6) and (3.7). Identity (3.8) follows by evaluating the functional $\mathbb{CT}_c$ in $(f^{(1)}, f^{(2)})$. ■

REMARK 9. The cardinal flow defined in (3.6) and (3.7) is the cardinal flow induced by the Knothe–Rosenblatt rearrangement that moves μ into ν [21, 27]. In particular, from Theorem 9, we infer that the Knothe–Rosenblatt rearrangement is an optimal transportation plan if the cost function is separable and the starting measure μ is such that $\mu(\{x_1 = \bar{x}\}) = 1$, for a given $\bar{x} \in \mathbb{R}$.

Theorem 9 becomes particularly useful when we take the squared Euclidean distance as cost function, i.e.,

$$c(\boldsymbol{x}, \boldsymbol{y}) := |x_1 - y_1|^2 + |x_2 - y_2|^2.$$

Indeed, since c is invariant under isometries, Theorem 9 generalizes to any μ supported on a straight line.

COROLLARY 6. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^2)$, and $c(\boldsymbol{x}, \boldsymbol{y}) := |x_1 - y_1|^2 + |x_2 - y_2|^2$. If there exists a triple $a, b, q \in \mathbb{R}$ such that*

$$\mu\big(\{ax_1 + bx_2 = q\}\big) = 1,$$

then the optimal transportation plan between μ and ν is $\pi_O := (O, O)_\# \pi$, where $O : \mathbb{R}^2 \to \mathbb{R}^2$ is a rotation that sends the set $\{x_2 = 0\}$ in $\{ax_1 + bx_2 = q\}$ and π is the optimal transportation plan between $(O^{(-1)})_\# \mu$ and $(O^{(-1)})_\# \nu$.

PROOF. Since the Euclidean distance is invariant under isometries, we have

$$\mathbb{T}_c(\pi) = \mathbb{T}_c\big((O^{(-1)}, O^{(-1)})_\# \pi\big)$$

for any rotation O and, therefore,

$$W_2^2\big((O^{(-1)})_\# \mu, (O^{(-1)})_\# \nu\big) = W_2^2(\mu, \nu).$$

Since $(O^{(-1)})_\# \mu$ satisfies the hypothesis of Theorem 9, we conclude the thesis. ■

ACKNOWLEDGEMENTS. – We thank Stefano Gualandi and Marco Veneroni for their valuable feedback. We also thank Gabriele Loli for enhancing the images of this paper.

FUNDING. – This project is partially supported by a Leverhulme Trust Research Project Grant (2021–2024).

REFERENCES

[1] T. ABDELLAOUI – H. HEINICH, Caractérisation d'une solution optimale au problème de Monge-Kantorovitch. *Bull. Soc. Math. France* **127** (1999), no. 3, 429–443. Zbl 0940.60013 MR 1724403

[2] L. AMBROSIO – N. GIGLI – G. SAVARÉ, *Gradient flows: In metric spaces and in the space of probability measures*. 2nd edn., Lectures in Math. ETH Zürich, Birkhäuser, Basel, 2008. Zbl 1145.35001 MR 2401600

[3] M. ARJOVSKY – S. CHINTALA – L. BOTTOU, Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 214–223, Proc. Mach. Learn. Res. (PMLR) 70, 2017.

[4] G. AURICCHIO – F. BASSETTI – S. GUALANDI – M. VENERONI, Computing Wasserstein barycenters via linear programming. In *Integration of constraint programming, artificial intelligence, and operations research*, pp. 355–363, Lecture Notes in Comput. Sci. 11494, Springer, Cham, 2019. Zbl 07116704

[5] G. AURICCHIO – S. GUALANDI – M. VENERONI – F. BASSETTI, Computing Kantorovich-Wasserstein distances on d-dimensional histograms using $(d+1)$-partite graphs. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 5798–5808, Adv. Neural Inf. Process. Syst. 31, 2018.

[6] G. AURICCHIO – M. VENERONI, On the structure of optimal transportation plans between discrete measures. *Appl. Math. Optim.* **85** (2022), no. 3, article no. 29. Zbl 1497.49055 MR 4429318

[7] V. I. BOGACHEV, *Measure theory. Vol. I*. Springer, Berlin, 2007. Zbl 1120.28001 MR 2267655

[8] G. BOUCHITTÉ – C. JIMENEZ – M. RAJESH, A new L^∞ estimate in optimal mass transport. *Proc. Amer. Math. Soc.* **135** (2007), no. 11, 3525–3535. Zbl 1120.49040 MR 2336567

[9] Y. BRENIER, Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** (1991), no. 4, 375–417. Zbl 0738.46011 MR 1100809

[10] L. A. CAFFARELLI – M. FELDMAN – R. J. MCCANN, Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.* **15** (2002), no. 1, 1–26. Zbl 1053.49032 MR 1862796

[11] J. A. CUESTA-ALBERTOS – C. MATRÁN – A. TUERO-DÍAZ, On the monotonicity of optimal transportation plans. *J. Math. Anal. Appl.* **215** (1997), no. 1, 86–94. Zbl 0892.60020 MR 1478852

[12] M. Cuturi – A. Doucet, Fast computation of Wasserstein barycenters. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 685–693, Proc. Mach. Learn. Res. (PMLR) 32, 2014.

[13] R. L. Dobrushin, Vlasov equations. *Funct. Anal. Appl.* **13** (1979), 115–123. Zbl 0422.35068 MR 541637

[14] A. Figalli, Existence, uniqueness, and regularity of optimal transport maps. *SIAM J. Math. Anal.* **39** (2007), no. 1, 126–137. Zbl 1132.28322 MR 2318378

[15] C. Frogner – C. Zhang – H. Mobahi – M. Araya-Polo – T. Poggio, Learning with a Wasserstein loss. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, pp. 2053–2061, Adv. Neural Inf. Process. Syst. 28, 2015.

[16] W. Gangbo – R. J. McCann, The geometry of optimal transportation. *Acta Math.* **177** (1996), no. 2, 113–161. Zbl 0887.49017 MR 1440931

[17] A. V. Goldberg – E. Tardos – R. E. Tarjan, Network flow algorithms. In *Paths, flows, and VLSI-layout (Bonn, 1988)*, pp. 101–164, Algorithms Combin. 9, Springer, Berlin, 1990. Zbl 0728.90035 MR 1083378

[18] L. V. Kantorovich, On a problem of Monge (in Russian). *Uspekhi Mat. Nauk* **3** (1948), 225–226.

[19] L. V. Kantorovich, Mathematical methods of organizing and planning production. *Management Sci.* **6** (1959/60), 366–422. Zbl 0995.90532 MR 129016

[20] L. V. Kantorovich, On the translocation of masses. *J. Math. Sci. (N.Y.)* **133** (2006), 1381–1382. Zbl 1080.49507

[21] H. Knothe, Contributions to the theory of convex bodies. *Michigan Math. J.* **4** (1957), 39–52. Zbl 0077.35803 MR 83759

[22] G. Loeper, On the regularity of solutions of optimal transportation problems. *Acta Math.* **202** (2009), no. 2, 241–283. Zbl 1219.49038 MR 2506751

[23] G. Monge, *Mémoire sur la théorie des déblais et des remblais*. Imprimerie Royale, Paris, 1781.

[24] H. Murata – H. Tanaka, An inequality for certain functional of multidimensional probability distributions. *Hiroshima Math. J.* **4** (1974), 75–81. Zbl 0287.60021 MR 365656

[25] J. B. Orlin, A faster strongly polynomial minimum cost flow algorithm. *Oper. Res.* **41** (1993), no. 2, 338–350. Zbl 0781.90036 MR 1214540

[26] O. Pele – M. Werman, Fast and robust Earth Mover's Distances. In *2009 IEEE 12th International Conference on Computer Vision*, pp. 460–467, 2009.

[27] M. Rosenblatt, Remarks on a multivariate transformation. *Ann. Math. Statistics* **23** (1952), 470–472. Zbl 0047.13104 MR 49525

[28] Y. Rubner – C. Tomasi – L. J. Guibas, A metric for distributions with applications to image databases. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 59–66, 1998.

[29] Y. Rubner – C. Tomasi – L. J. Guibas, The Earth Mover's Distance as a metric for image retrieval. *Int. J. Comput. Vis.* **40** (2000), no. 2, 99–121. Zbl 1012.68705

[30] L. Rüschendorf – S. T. Rachev, A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.* **32** (1990), no. 1, 48–54. Zbl 0688.62034 MR 1035606

[31] F. Santambrogio, *Optimal transport for applied mathematicians: Calculus of variations, PDEs, and modeling*. Progr. Nonlinear Differential Equations Appl. 87, Birkhäuser, Cham, 2015. Zbl 1401.49002 MR 3409718

[32] J. Solomon – R. Rustamov – L. Guibas – A. Butscher, Wasserstein propagation for semi-supervised learning. In *Proceedings of the 31st International Conference on Machine Learning*, pp. 306–314, Proc. Mach. Learn. Res. (PMLR) 32, 2014.

[33] H. Tanaka, An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas. *Z. Wahrsch. Verw. Gebiete* **27** (1973), 47–52. Zbl 0302.60005 MR 362442

[34] H. Tanaka, Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Gebiete* **46** (1978/79), no. 1, 67–105. Zbl 0389.60079 MR 512334

[35] C. Villani, *Optimal transport: Old and new*. Grundlehren Math. Wiss. 338, Springer, Berlin, 2009. Zbl 1156.53003 MR 2459454

Received 7 August 2023,
and in revised form 15 September 2023

Gennaro Auricchio
Department of Computer Science, University of Bath
Claverton Down, Bath BA2 7AY, UK
ga647@bath.ac.uk