




On the Structure of Optimal Transportation Plans between Discrete Measures

Gennaro Auricchio¹ · Marco Veneroni¹ 

Accepted: 1 February 2022
© The Author(s) 2022

Abstract

It is well known that the optimal transportation plan between two probability measures μ and ν is induced by a transportation map whenever μ is an absolutely continuous measure supported over a compact set in the Euclidean space and the cost function is a strictly convex function of the Euclidean distance. However, when μ and ν are both discrete, this result is generally false. In this paper, we prove that, given any pair of discrete probability measures and a cost function, there exists an optimal transportation plan that can be expressed as the sum of two deterministic plans, i.e., plans induced by transportation maps. As an application, we estimate the infinity-Wasserstein distance between two discrete probability measures μ and ν with the p -Wasserstein distance, times a constant depending on μ , on ν , and on the fixed cost function.

Keywords Wasserstein distance · Discrete optimal transport · Uniform estimates · Structure of solutions · Monge–Kantorovich problem

Mathematics Subject Classification 49Q22 · 05C70 · 39B62

1 Introduction

The Optimal Transport (OT) problem is a classical minimization problem dating back to the work of Monge [24] and Kantorovich [20,21]. In this problem, we are given two probability measures, namely μ and ν , and we search for the cheapest way to reshape μ into ν . The effort needed in order to perform this transformation depends on a cost function, which describes the underlying geometry of the product space of

✉ Marco Veneroni
marco.veneroni@unipv.it

Gennaro Auricchio
gennaro.auricchio@unipv.it

¹ Department of Mathematics F. Casorati, University of Pavia, Via Ferrata 5, 27100 Pavia, Italy

the support of the two measures. In the right setting, this effort induces a distance between probability measures.

During the last century, the OT problem has been fruitfully used in many applied fields such as the study of systems of particles by Dobrushin [13], the Boltzmann equation by Tanaka in [17–19], and the field of fluidodynamics by Yann Brenier [9]. All these results pointed out that, by a qualitative description of optimal transport, it was possible to gain insightful information on many open problems. For this reason, the Optimal Transport problem has become a topic of major interest for analysts, probabilists and statisticians [3,29,31]. In particular, a plethora of results concerning the uniqueness [10,14,16], the structure [1,2,28], and the regularity [8,23] of the optimal transportation plan in the continuous framework has been proved.

In recent years, it has also become a crucial sub-problem in several applications in Computer Vision [7,25–27], Computational Statistics [22], Probability [5,6], and Machine Learning [4,11,15,30]. However, in these fields, the measures μ and ν are discrete, and therefore the optimal transportation plans lack most of the good properties their continuous counterparts enjoy.

In this paper, we study the structure of optimal transportation plans between discrete probability measures. After introducing the notion of *trim plan* between the measures μ and ν , we prove that such plans are the sum of two deterministic plans, i.e., plans that are induced by the action of two suitable push-forward maps. The first map acts on a portion $\mu^{(d)}$ of μ , while the other one acts on a portion $\nu^{(d)}$ of ν (Theorem 3). Thanks to this formula, we recover an extension of the estimate given in [8]. Namely, we estimate the *infinity*-Wasserstein distance between a pair of discrete measures (μ, ν) (see Definition 4 below) by the *c*-Wasserstein distance between μ and ν , times a quantity that only depends on μ and ν (Theorem 7).

2 Basic Notions on Optimal Transport

In this section, following [31], we recall the main definitions regarding optimal transportation and we examine the continuous counterpart [8] to our W^∞ estimate.

Given a polish space (X, d) , we denote with $\mathcal{B}(X)$ the set of Borel sets over X , while with $\mathcal{P}(X)$ we denote the set of Borel measures over X . Given a Borel measurable function $T : X \rightarrow Y$, we denote with $T_\# : \mathcal{P}(X) \rightarrow \mathcal{P}(Y)$ the push-forward operator induced by T , defined by: $(T_\#\mu)[A] = \mu[T^{-1}(A)]$. The projection maps are $\mathfrak{p}_X : X \times Y \rightarrow X$, $\mathfrak{p}_X(x, y) = x$ and $\mathfrak{p}_Y : X \times Y \rightarrow Y$, $\mathfrak{p}_Y(x, y) = y$.

Definition 1 Let μ and ν be two measures over two polish spaces X and Y . The 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 .

Given $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(Y)$, the quantity $\pi(A \times B)$ is the amount of mass that travels from the set A to the set B . By assigning a cost function c on $X \times Y$ we specify a way to measure the cost of every transportation plan.

Definition 2 Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and let $c : X \times Y \rightarrow [0, +\infty)$ be a lower semicontinuous (l.s.c.) symmetric cost function. The transportation functional $\mathbb{T}_c : \Pi(\mu, \nu) \rightarrow [0, +\infty)$ is defined as

$$\mathbb{T}_c(\pi) := \int_{X \times Y} c \, d\pi. \tag{1}$$

Given two measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and a cost function c , the optimal transportation problem consists in finding the infimum of \mathbb{T}_c over $\Pi(\mu, \nu)$, i.e.

$$\inf_{\pi \in \Pi(\mu, \nu)} \mathbb{T}_c(\pi). \tag{2}$$

By making further assumptions on c , it is possible to prove that the infimum in (2) is actually a minimum. In particular, when the cost function is nonnegative, the solution exists. For a complete discussion on the existence of the solution, we refer to [31, Chapter 4].

We can use the optimal transportation problem to define a distance over the space $\mathcal{P}(X)$. In particular, since X is a polish space, we can lift the distance d from X to $\mathcal{P}(X)$, by choosing d as a cost function in (1).

Definition 3 Let (X, d) be a polish space and $p \in [1, \infty)$. The Wasserstein distance of order p between the probability measures μ and ν on X is defined as

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \mathbb{T}_{d^p}(\pi) \right)^{\frac{1}{p}} = \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} d^p(x, y) d\pi(x, y) \right)^{\frac{1}{p}}. \tag{3}$$

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

When the cost function is not the space distance d , we denote the infimum in (2) with $W_c(\mu, \nu)$.

Remark 1 The infimum in (3) could actually be $+\infty$, it is thus customary to restrict W_p to the space of probability measures with finite p -moments.

Definition 4 Given a cost function c , the $W_c^{(\infty)}$ distance between two measures μ and ν is defined as

$$W_c^{(\infty)}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \|c\|_{L_\pi^\infty}$$

where $\|\cdot\|_{L_\pi^\infty}$ is the L^∞ norm with respect to the measure π . When c is the Euclidean distance, we use the notation: $W^{(\infty)}$.

Let μ and ν be two probability measures on a Lipschitz regular and bounded subset $\Omega \subset \mathbb{R}^n$. We define the cost function

$$c_p(\mathbf{x}, \mathbf{y}) := \left(\sqrt[n]{\sum_{i=1}^n |x_i - y_i|^2} \right)^p, \quad p > 1.$$

When μ is absolutely continuous with respect to the Lebesgue measure, it is well known (Theorem 6.3 and Theorem 6.4, [16]) that the optimal transportation plan π between μ and ν is unique and it is induced by a transportation map T_p , i.e.

$$\pi = (Id, T_p)_\# \mu.$$

In [8], Bouchitté et al. established an L^∞_μ -bound on the displacement map $Id - T_p$, which only depends on the shape of Ω , on p , and on the density of μ . This estimate allowed the authors to give the following upper bound on the $W^{(\infty)}$ distance between μ and ν .

Theorem 1 (Theorem 1.2, [8]) *Let Ω be a bounded connected open subset of \mathbb{R}^n with Lipschitz boundary and denote by $\mathcal{P}(\overline{\Omega})$ (resp. $\mathcal{P}_{ac}(\Omega)$) the set of Borel (resp. absolutely continuous) probability measures on $\overline{\Omega}$. Then, for every $p > 1$ and every pair $(\mu, \nu) \in \mathcal{P}_{ac}(\Omega) \times \mathcal{P}(\overline{\Omega})$ there holds*

$$(W^{(\infty)}(\mu, \nu))^{p+n} \leq C_{p,n}(\Omega) \|f^{-1}\|_{L^\infty(\Omega)} W_p^p(\mu, \nu), \tag{4}$$

where f is the density of μ with respect to the Lebesgue measure and $C_{p,n}(\Omega)$ is a positive constant depending only on p, n , and Ω .

The proof of this result heavily relies on the regularity of μ , hence, when μ and ν are both discrete, this result does not apply. In particular, we are no longer able to find a constant depending only on μ and the geometry of the support of μ , as the following example shows.

Example 1 Let $\mu, \nu_\epsilon \in \mathcal{P}(\mathbb{R})$ be defined as

$$\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \quad \nu_\epsilon = \frac{1-\epsilon}{2}\delta_0 + \frac{1+\epsilon}{2}\delta_1,$$

for $\epsilon \in (0, 1)$, and let $c_2(x, y) = |x - y|^2$. By a simple computation we have that

$$W_2^{(\infty)}(\mu, \nu_\epsilon) = 1, \quad W_2^2(\mu, \nu_\epsilon) = \frac{\epsilon}{2}.$$

Hence, estimate (4) does not hold true, as for every constant $C(p, n, \Omega, \mu) > 0$ (possibly depending on p, n, Ω, μ), there exists $\epsilon > 0$ such that

$$(W^{(\infty)}(\mu, \nu_\epsilon))^{2+1} = 1 > C(p, n, \Omega, \mu) W_2^2(\mu, \nu) = \epsilon C(p, n, \Omega, \mu).$$

3 Structure of Discrete Optimal Transportation Plans

In what follows, we prove the existence of an optimal transportation plan between two discrete measures that is induced by the action of two push-forward functions, one going from X to Y and one going from Y to X . This allows us to establish a bound on $W^{(\infty)}(\mu, \nu)$, similar to the one proved in [8]. We always assume $\#X = \#Y = n \in \mathbb{N}$. In

this case, we can identify the sets X and Y with $\{1, \dots, n\}$. Without loss of generality, we therefore assume $X = Y$. In this setting, a measure $\mu \in \mathcal{P}(X)$ has the form $\sum_{x \in X} \mu_x \delta_x$, we thus use the notation μ_x to denote the coefficient of μ in x and, likewise, $c_{x,y}$ (resp. $\pi_{x,y}$) stands for the value of $c : X \times Y \rightarrow \mathbb{R}$ (resp. the coefficient of $\pi \in \mathcal{P}(X \times Y)$) in the point $(x, y) \in X \times Y$.

Definition 5 Let $\mu, \nu \in \mathcal{P}(X)$ be two measures on a discrete set X and let $c : X \times X \rightarrow \mathbb{R}$ be a cost function. A minimal solution π^* of the transportation problem is said to be *trim* if

$$\#spt(\pi^*) \leq \#spt(\pi)$$

for each optimal solution π .

Lemma 2 Let $\pi \in \Pi(\mu, \nu)$ be a trim solution. Then each restriction of π is a trim solution for its marginals. In particular, if $\pi^{(1)}$ and $\pi^{(2)}$ are such that

$$\pi = \pi^{(1)} + \pi^{(2)}$$

and $spt(\pi^{(1)}) \cap spt(\pi^{(2)}) = \emptyset$, then $\pi^{(1)}$ and $\pi^{(2)}$ are trim solutions for their marginals.

Proof Let π^* be a restriction of π . By Theorem 4.6 (Chapter 4, [31]), we know that π^* is optimal between its marginals, hence we only need to prove that its support has minimal cardinality.

Arguing by contradiction, let us assume that π^* is not trim, hence there exists another optimal plan η between the marginals of π^* such that

$$\#spt(\eta) < \#spt(\pi^*).$$

We can define the measure $\hat{\pi}$ as

$$\hat{\pi} = \pi - \pi^* + \eta,$$

since $\pi \geq \pi^*$ and $\eta \geq 0$, we have $\hat{\pi} \geq 0$. Moreover, since π^* and η have the same marginals, $\hat{\pi}$ has the same marginals of π , therefore $\hat{\pi} \in \Pi(\mu, \nu)$. Moreover, since π^* and η are optimal between their marginals, we have

$$\sum_{(x,y) \in X \times X} c_{x,y} \pi_{x,y}^* = \sum_{(x,y) \in X \times X} c_{x,y} \eta_{x,y},$$

thus

$$\begin{aligned} \sum_{(x,y) \in X \times X} c_{x,y} \hat{\pi}_{x,y} &= \sum_{(x,y) \in X \times X} c_{x,y} \pi_{x,y} - \sum_{(x,y) \in X \times X} c_{x,y} \pi_{x,y}^* \\ &\quad + \sum_{(x,y) \in X \times X} c_{x,y} \eta_{x,y} \\ &= \sum_{(x,y) \in X \times X} c_{x,y} \pi_{x,y}. \end{aligned}$$

In particular, π and $\hat{\pi}$ have the same cost, therefore $\hat{\pi}$ is an optimal transportation plan between μ and ν .

To conclude, we notice that, since π^* is a restriction of π , we have

$$\#spt(\pi) = \#spt(\pi - \pi^*) + \#spt(\pi^*) > \#spt(\pi - \pi^*) + \#spt(\eta) \geq \#spt(\hat{\pi}),$$

which concludes the contradiction, since π is trim by hypothesis. □

Theorem 6.3 in [16] states that, whenever μ is an absolutely continuous measure supported over a compact set $\Omega \subset \mathbb{R}^n$ and the cost function c is a strictly convex function of the Euclidean distance, the optimal transportation plan is induced by a transportation map, regardless of the regularity of ν . When μ and ν are both discrete, this result is generally false. However, in the next Theorem 3, we show that there exists at least one optimal transportation plan between two measures that can be recreated as the action of two functions, one acting from a subset $\tilde{X} \subset spt(\mu)$ to $spt(\nu)$ and one acting from a subset $\tilde{Y} \subset spt(\nu)$ to $spt(\mu)$.

Theorem 3 *Let X be a discrete polish space and let μ and ν be two positive measures over the set X such that*

$$\begin{aligned} \mu_a > 0 & \quad \forall a \in X, \\ \nu_b > 0 & \quad \forall b \in X, \end{aligned}$$

and

$$\sum_{a \in X} \mu_a = \sum_{b \in X} \nu_b.$$

Given a cost function $c : X \times X \rightarrow \mathbb{R}$, let π be a trim solution of the transportation problem. We can then find two couples of measures $(\mu^{(d)}, \mu^{(c)})$ and $(\nu^{(d)}, \nu^{(c)})$ and a couple of functions $h^{(1)}$ and $h^{(2)}$ such that

$$\mu = \mu^{(d)} + \mu^{(c)} \quad \text{and} \quad \nu = \nu^{(d)} + \nu^{(c)}, \tag{5}$$

$$\pi = (Id, h^{(1)})_{\#} \mu^{(d)} + (h^{(2)}, Id)_{\#} \nu^{(d)}. \tag{6}$$

We say that the decomposition ensured by Theorem 3 is a *diffusive model* associated with the given (trim) solution π . We call $\mu^{(d)}$ and $\nu^{(d)}$ the *diffusive part* of μ and ν , respectively. Similarly, we denote with $\mu^{(c)}$ and $\nu^{(c)}$ the *concentrating part* of μ and ν , respectively. Finally, we call $h^{(1)}$ the *diffusive scheme* of μ and $h^{(2)}$ the *diffusive scheme* of ν .

Proof We proceed by induction on the cardinality of X . If $\#X = 1$, the thesis follows trivially.

Let us now assume that the statement holds for each couple of measures whose support has cardinality $(n - 1)$ and let μ and ν be two measures supported on a set

with cardinality n , namely X_n . Given a trim solution π , it is well known (Chapter 7, [12]) that

$$\#\text{spt}(\pi) \leq 2n - 1.$$

Since μ and ν have n points in their support, we can find $\bar{a} \in X$ such that there exists a unique $\bar{b} \in \text{spt}(\nu)$ for which

$$\pi_{\bar{a},\bar{b}} > 0,$$

hence $\mu_{\bar{a}} = \pi_{\bar{a},\bar{b}} \leq \nu_{\bar{b}}$. Similarly, we can find $\underline{b} \in X$ such that there exists a unique $\underline{a} \in \text{spt}(\mu)$ for which

$$\pi_{\underline{a},\underline{b}} > 0,$$

so that $\nu_{\underline{b}} = \pi_{\underline{a},\underline{b}} \leq \mu_{\underline{a}}$.

If $\mu_{\bar{a}} = \pi_{\bar{a},\bar{b}} = \nu_{\bar{b}}$, we can restrict the plan π to the set $\text{spt}(\pi) \setminus \{(\bar{a}, \bar{b})\}$. We denote this restriction with π_* . By definition, the marginals of π_* are

$$\mu_* = \mu - \mu_{\bar{a}}\delta_{\bar{a}}$$

and

$$\nu_* = \nu - \nu_{\bar{b}}\delta_{\bar{b}}.$$

In particular, the supports of μ_* and ν_* contain $(n - 1)$ points each. By induction we can find $(\mu_*^{(d)}, \mu_*^{(c)})$, $(\nu_*^{(d)}, \nu_*^{(c)})$, and $(h_*^{(1)}, h_*^{(2)})$ such that

$$\begin{aligned} \mu_* &= \mu_*^{(d)} + \mu_*^{(c)}, \\ \nu_* &= \nu_*^{(d)} + \nu_*^{(c)}, \end{aligned}$$

and

$$\pi_* = (Id, h_*^{(1)})\# \mu_*^{(d)} + (h_*^{(2)}, Id)\# \nu_*^{(d)}.$$

We can then define

$$\begin{aligned} \mu^{(d)} &= \mu_*^{(d)} + \mu_{\bar{a}}\delta_{\bar{a}}, & \mu^{(c)} &= \mu_*^{(c)}, \\ \nu^{(d)} &= \nu_*^{(d)}, & \nu^{(c)} &= \nu_*^{(c)} + \nu_{\bar{b}}\delta_{\bar{b}}, \end{aligned}$$

and

$$h^{(1)}(a) = \begin{cases} h_*^{(1)}(a) & \text{if } a \neq \bar{a}, \\ \bar{b} & \text{otherwise,} \end{cases}, \quad h^{(2)}(b) = h_*^{(2)}(b).$$

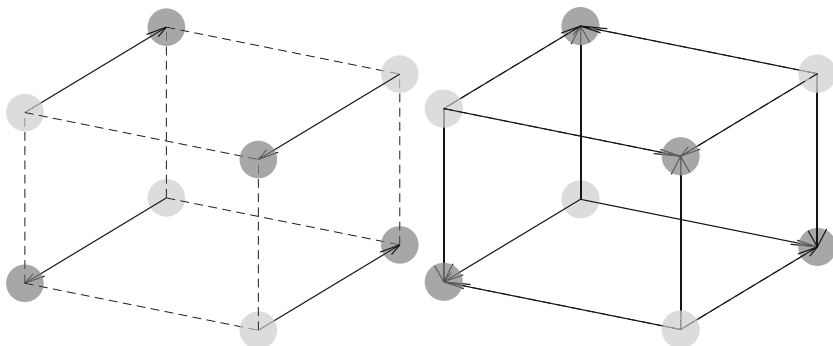


Fig. 1 (Example 2) Visual comparison between the optimal plan π , which is not trim (right) and a trim plan (left). The support of μ is indicated by light gray dots, the support of ν by dark gray dots, points i and j are connected by a solid edge if (i, j) belongs to the support of the plan

It easy to see that

$$\mu = \mu^{(d)} + \mu^{(c)}, \quad \nu = \nu^{(d)} + \nu^{(c)}$$

and, since $h_{\#}^{(1)} \delta_{\bar{a}} = \delta_{\bar{b}}$, we have

$$\pi = (Id, h^{(1)})_{\#} \mu^{(d)} + (h^{(2)}, Id)_{\#} \nu^{(d)}, \tag{7}$$

which concludes the proof in the case $\mu_{\bar{a}} = \pi_{\bar{a}, \bar{b}} = \nu_{\bar{b}}$. We proceed similarly if $\nu_{\underline{b}} = \pi_{\underline{a}, \underline{b}} = \mu_{\underline{a}}$. See Fig. 1 for a visual representation of this process in the case $n = 3$.

To conclude, consider the case in which $\mu_{\bar{a}} = \pi_{\bar{a}, \bar{b}} < \nu_{\bar{b}}$ and $\nu_{\underline{b}} = \pi_{\underline{a}, \underline{b}} < \mu_{\underline{a}}$. In this case, we restrict π to the set $\text{spt}(\pi) \setminus \{(\bar{a}, \bar{b}), (\underline{a}, \underline{b})\}$. Let us denote again with π_* the restriction and with μ_* and ν_* its marginals. Since both μ_* and ν_* have $(n - 1)$ points in their supports, we can again decompose them as

$$\mu_* = \mu_*^{(d)} + \mu_*^{(c)}, \quad \nu_* = \nu_*^{(d)} + \nu_*^{(c)}$$

and find a couple of functions $h_*^{(1)}, h_*^{(2)}$ for which

$$\pi_* = (Id, h_*^{(1)})_{\#} \mu_*^{(d)} + (h_*^{(2)}, Id)_{\#} \nu_*^{(d)}.$$

We can then define

$$\begin{aligned} \mu^{(d)} &= \mu_*^{(d)} + \mu_{\bar{a}} \delta_{\bar{a}}, & \mu^{(c)} &= \mu_*^{(c)} + \mu_{\underline{a}} \delta_{\underline{a}}, \\ \nu^{(d)} &= \nu_*^{(d)} + \nu_{\bar{b}} \delta_{\bar{b}}, & \nu^{(c)} &= \nu_*^{(c)} + \nu_{\underline{b}} \delta_{\underline{b}}, \end{aligned}$$

and

$$h^{(1)}(a) = \begin{cases} h_*^{(1)}(a) & \text{if } a \neq \bar{a}, \\ \bar{b} & \text{otherwise.} \end{cases} \quad h^{(2)}(b) = \begin{cases} h_*^{(2)}(b) & \text{if } b \neq \underline{b}, \\ \underline{a} & \text{otherwise,} \end{cases}$$

which concludes the thesis. □

Remark 2 Given two measures as in the hypothesis of Theorem 3, let $\mu^{(d)}$ and $\nu^{(d)}$ be their diffusive parts. Since $\text{spt}(\mu^{(d)}) \subset \text{spt}(\mu)$ and $\text{spt}(\nu^{(d)}) \subset \text{spt}(\nu)$, the support of the transportation plan defined by formula (7) has, at most, $2n$ points. Thus the trim condition on the optimal transportation plan is necessary, as we are going to show in the next example.

Example 2 Let us take

$$\mu = \frac{1}{4} \left(\delta_{(0,0,0)} + \delta_{(1,1,0)} + \delta_{(1,0,1)} + \delta_{(0,1,1)} \right)$$

and

$$\nu = \frac{1}{4} \left(\delta_{(1,1,1)} + \delta_{(0,0,1)} + \delta_{(0,1,0)} + \delta_{(1,0,0)} \right),$$

(see Fig. 2) and, as a cost function, we choose the Euclidean distance in \mathbb{R}^3 , i.e.

$$|\mathbf{x} - \mathbf{y}| := \sqrt{\sum_{i=1}^3 (x_i - y_i)^2}.$$

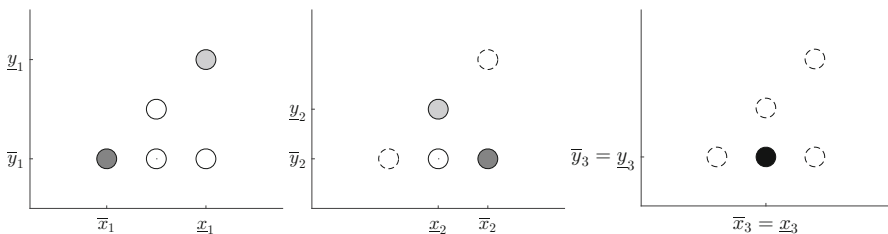


Fig. 2 Visual description of the decomposition process used in the proof of Theorem 3. Here, the measures μ and ν are one-dimensional and have 3 points each in their supports

It is easy to see that the plan

$$\begin{aligned} \pi := & \frac{1}{12} \delta_{(0,0,0)} \otimes \left(\delta_{(1,0,0)} + \delta_{(0,1,0)} + \delta_{(0,0,1)} \right) \\ & + \frac{1}{12} \delta_{(1,1,0)} \otimes \left(\delta_{(0,1,0)} + \delta_{(1,0,0)} + \delta_{(1,1,1)} \right) \\ & + \frac{1}{12} \delta_{(1,0,1)} \otimes \left(\delta_{(1,0,0)} + \delta_{(0,0,1)} + \delta_{(1,1,1)} \right) \\ & + \frac{1}{12} \delta_{(0,1,1)} \otimes \left(\delta_{(0,1,0)} + \delta_{(0,0,1)} + \delta_{(1,1,1)} \right) \end{aligned}$$

is optimal. However, according to Remark 2, it cannot be decomposed as in formula (7), since

$$\#spt(\pi) = 12 > 2\#spt(\mu) = 8.$$

Remark 3 Given a trim solution, there might be more than one diffusive model associated with it. For example, let

$$\mu = \frac{1}{2} \delta_{(0,0)} + \frac{1}{2} \delta_{(1,1)} \quad \text{and} \quad \nu = \frac{1}{4} \delta_{(-1,1)} + \frac{3}{4} \delta_{(1,0)}$$

be two discrete measures over \mathbb{R}^2 . As a cost function, we choose the Euclidean distance

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

Then, the probability measure

$$\pi = \frac{1}{4} \delta_{(0,0)} \otimes \delta_{(-1,1)} + \frac{1}{4} \delta_{(0,0)} \otimes \delta_{(1,0)} + \frac{1}{2} \delta_{(1,1)} \otimes \delta_{(1,0)}$$

is a trim plan between μ and ν . It is easy to check that

$$\begin{aligned} \mu^{(d)} &= \frac{1}{4} \delta_{(0,0)} + \frac{1}{2} \delta_{(1,1)}, & \mu^{(c)} &= \frac{1}{4} \delta_{(0,0)}, \\ \nu^{(c)} &= \frac{1}{4} \delta_{(-1,1)} + \frac{1}{2} \delta_{(1,0)}, & \nu^{(d)} &= \frac{1}{4} \delta_{(1,0)}, \end{aligned}$$

and

$$h^{(1)} := \begin{cases} (-1, 1) & \text{if } x = (0, 0), \\ (+1, 0) & \text{if } x = (1, 1), \\ (0, 0) & \text{otherwise,} \end{cases} \quad h^{(2)}(y) = (0, 0) \quad \forall y \in \mathbb{R}^2,$$

is a decomposition of the trim plan. However, we can also decompose ν as

$$\tilde{\nu}^{(d)} = \frac{1}{4}\delta_{(-1,1)}, \quad \tilde{\nu}^{(c)} = \frac{3}{4}\delta_{(1,0)},$$

define the functions as

$$h^{(1)}(\mathbf{x}) = (1, 0) \quad \forall \mathbf{x} \in \mathbb{R}^2, \quad h^{(2)}(\mathbf{y}) = (0, 0) \quad \forall \mathbf{y} \in \mathbb{R}^2,$$

and still obtain an admissible decomposition of π .

4 An Upper Bound for the Infinity Wasserstein Distance in the Discrete Setting

As an immediate consequence of the diffusive model decomposition (5)–(6) given in Theorem 3, we can decompose the Wasserstein distance associated to a cost function c and use it to estimate the infinity-Wasserstein distance.

Corollary 4 *Let $\mu, \nu \in \mathcal{P}(X)$ be two discrete measures, $c : X \times X \rightarrow \mathbb{R}$ be a cost function, and π be a trim solution of the transportation problem. Given a diffusive model for π , we have*

$$W_c(\mu, \nu) = \sum_{x \in X} c(x, h^{(1)}(x))\mu_x^{(d)} + \sum_{y \in X} c(h^{(2)}(y), y)\nu_y^{(d)}$$

and

$$T_c^{(\infty)}(\pi) = \max \left\{ \|c(x, h^{(1)}(x))\|_{L_{\mu^{(d)}}^\infty}, \|c(h^{(2)}(y), y)\|_{L_{\nu^{(d)}}^\infty} \right\}.$$

In particular, we have

$$W_c(\mu, \nu) \geq \alpha W_c^{(\infty)}(\mu, \nu), \tag{8}$$

where

$$\alpha = \min_{a \in \text{spt}(\mu^{(d)}), b \in \text{spt}(\nu^{(d)})} \{v_b^{(d)}, \mu_a^{(d)}\}. \tag{9}$$

The value α defined in (9) depends on the particular diffusive model we choose. However, since $W_c(\mu, \nu)$ and $W_c^{(\infty)}$ do not depend on the choice of the diffusive model, if we can give a lower bound on α for a particular diffusive model, we can generalize the estimate (8).

Corollary 5 Let $\mu, \nu \in \mathcal{P}(X)$ be two discrete measures and $c : X \times X \rightarrow \mathbb{R}_+$ be a cost function. For any trim plan π , there exists a diffusive model for which

$$\alpha \geq \min_{(A,B) \in K(\mu,\nu)} \left\{ \left| \sum_{x \in A} \mu_x - \sum_{y \in B} \nu_y \right| \right\}, \tag{10}$$

where α is defined in relation (9) and

$$K(\mu, \nu) := \left\{ (A, B) \subset X \times X \text{ s.t. } \left| \sum_{x \in A} \mu_x - \sum_{y \in B} \nu_y \right| > 0 \right\}.$$

Proof Let n be the cardinality of X . Since π is trim between μ and ν , we have $\#\text{spt}(\pi) \leq 2n - 1$, hence we can find \bar{x}_1 such that

$$\exists! \bar{y}_1 \text{ s.t. } \pi_{\bar{x}_1, \bar{y}_1} \neq 0$$

and \underline{y}_1 such that

$$\exists! \underline{x}_1 \text{ s.t. } \pi_{\underline{x}_1, \underline{y}_1} \neq 0.$$

If $\underline{x}_1 = \bar{x}_1$ (and hence $\underline{y}_1 = \bar{y}_1$), we have $\mu_{\bar{x}_1} = \nu_{\bar{y}_1}$ and we define

$$\mu_{\bar{x}_1}^{(d)} = \mu_{\bar{x}_1}, \quad \nu_{\bar{y}_1}^{(c)} = \mu_{\bar{x}_1},$$

and

$$\mu^{(1)} := \mu - \mu_{\bar{x}_1} \delta_{\bar{x}_1}, \quad \nu^{(1)} := \nu - \nu_{\bar{y}_1} \delta_{\bar{y}_1}, \quad \pi^{(1)} = \pi - \pi_{\bar{x}_1, \bar{y}_1} \delta_{\bar{x}_1, \bar{y}_1}.$$

Otherwise, if $\underline{x}_1 \neq \bar{x}_1$ (and hence $\underline{y}_1 \neq \bar{y}_1$), we set

$$\begin{aligned} \mu_{\bar{x}_1}^{(d)} &= \mu_{\bar{x}_1}, & \mu_{\underline{x}_1}^{(c)} &= \nu_{\underline{y}_1}, \\ \nu_{\underline{y}_1}^{(d)} &= \nu_{\underline{y}_1}, & \nu_{\bar{y}_1}^{(c)} &= \mu_{\bar{x}_1}, \end{aligned}$$

and

$$\begin{aligned} \mu^{(1)} &= \mu - \mu_{\bar{x}_1} \delta_{\bar{x}_1} - \nu_{\underline{y}_1} \delta_{\underline{x}_1}, \\ \nu^{(1)} &= \nu - \nu_{\underline{y}_1} \delta_{\underline{y}_1} - \mu_{\bar{x}_1} \delta_{\bar{y}_1} \\ \pi^{(1)} &= \pi - \pi_{\bar{x}_1, \bar{y}_1} \delta_{\bar{x}_1, \bar{y}_1} - \pi_{\underline{x}_1, \underline{y}_1} \delta_{\underline{x}_1, \underline{y}_1}. \end{aligned}$$

In both cases we find two measures, $\mu^{(1)}$ and $\nu^{(1)}$, whose support has at most $n - 1$ points. Since $\pi^{(1)}$ is a restriction of a trim plan, by Lemma 2, also $\pi^{(1)}$ is trim between

its marginals $\mu^{(1)}$ and $\nu^{(1)}$. Therefore, we can repeat the process, finding two points \bar{x}_2 and \underline{y}_2 for which

$$\exists! \bar{y}_2 \text{ s.t. } \pi_{\bar{x}_2, \bar{y}_2} \neq 0$$

and

$$\exists! \underline{x}_2 \text{ s.t. } \pi_{\underline{x}_2, \underline{y}_2} \neq 0.$$

We can then extend the definition of the measures $\mu^{(d)}$, $\mu^{(c)}$, $\nu^{(d)}$, and $\nu^{(c)}$, define the measures $\mu^{(2)}$, $\nu^{(2)}$, and $\pi^{(2)}$ and start all over again.

At each step, we define two measures $\mu^{(i)}$ and $\nu^{(i)}$ and increase the cardinality of the supports of $\mu^{(d)}$, $\mu^{(c)}$, $\nu^{(d)}$, and $\nu^{(c)}$. Given any $x \in \text{spt}(\mu^{(d)})$, we can then find $i \in \{0, 1, \dots, n - 1\}$ such that

$$\mu_x^{(d)} = \mu_x^{(i)}, \tag{11}$$

and, similarly, for any $y \in \text{spt}(\nu^{(d)})$, we can find a $j \in \{0, 1, \dots, n - 1\}$ such that

$$\nu_y^{(d)} = \nu_y^{(j)},$$

with the convention $\mu^{(0)} = \mu$ and $\nu^{(0)} = \nu$. The relation between $\mu^{(i)}$ and $\mu^{(i+1)}$ is either

$$\mu^{(i+1)} = \mu^{(i)} - \mu_{\bar{x}_{i+1}}^{(i)} \delta_{\bar{x}_{i+1}}$$

or

$$\mu^{(i+1)} = \mu^{(i)} - \mu_{\bar{x}_{i+1}}^{(i)} \delta_{\bar{x}_{i+1}} - \nu_{\underline{y}_{i+1}}^{(i)} \delta_{\underline{y}_{i+1}}.$$

Similarly, we have

$$\nu^{(i+1)} = \nu^{(i)} - \nu_{\underline{y}_{i+1}}^{(i)} \delta_{\underline{y}_{i+1}}$$

or

$$\nu^{(i+1)} = \nu^{(i)} - \nu_{\underline{y}_{i+1}}^{(i)} \delta_{\underline{y}_{i+1}} - \mu_{\bar{x}_{i+1}}^{(i)} \delta_{\bar{x}_{i+1}}.$$

Similarly, we can write $\mu^{(i)}$ and $\nu^{(i)}$ as a function of $\mu^{(i-1)}$ and $\nu^{(i-1)}$, and then express $\mu^{(i+1)}$ through $\mu^{(i-1)}$ and $\nu^{(i-1)}$ as

$$\mu_x^{(i+1)} = \sum_{a \in \bar{A}_2} \mu_a^{(i-1)} - \sum_{b \in \bar{B}_2} \nu_b^{(i-1)}, \tag{12}$$

where \tilde{A}_2 and \tilde{B}_2 are two subsets of X whose cardinality is at most two. By iterating this process, we are able to find

$$\mu_x^{(i+1)} = \sum_{a \in \tilde{A}_{n-(i+1)}} \mu_a - \sum_{b \in \tilde{B}_{n-(i+1)}} \nu_b, \tag{13}$$

where $\tilde{A}_{n-(i+1)}$ and $\tilde{B}_{n-(i+1)}$ are subsets of X , whose cardinality is $n - (i + 1)$. Since the left side of (12) is positive, we can rewrite (13) as

$$\mu_x^{(i+1)} = \left| \sum_{a \in \tilde{A}_2} \mu_a^{(i-1)} - \sum_{b \in \tilde{B}_2} \nu_b^{(i-1)} \right|. \tag{14}$$

By taking the minimum over $K(\mu, \nu)$ of the right side in (14), we find

$$\mu_x^{(i)} \geq \min_{(A, B) \in K(\mu, \nu)} \left\{ \left| \sum_{x \in A} \mu_x - \sum_{y \in B} \nu_y \right| \right\},$$

for any $i \in \{0, 1, \dots, n - 1\}$ and each $x \in \text{spt}(\mu^{(i)})$, therefore, from relation (11), we get

$$\mu^{(d)} \geq \min_{(A, B) \in K(\mu, \nu)} \left\{ \left| \sum_{x \in A} \mu_x - \sum_{y \in B} \nu_y \right| \right\}.$$

Similarly, one can prove

$$\nu_y^{(d)} \geq \min_{(A, B) \in K(\mu, \nu)} \left\{ \left| \sum_{x \in A} \mu_x - \sum_{y \in B} \nu_y \right| \right\},$$

for each $y \in \text{spt}(\nu^{(d)})$, hence relation (10) is proven. □

In Corollary 4, we bound $W_c^{(\infty)}$ from above with W_c . However, due to the properties of $W_c^{(\infty)}$, it is possible to relate this distance to the Wasserstein cost induced by any p -power of the same cost function.

Lemma 6 *Let $\mu, \nu \in \mathcal{P}(X)$ and let $c : X \times X \rightarrow \mathbb{R}_+$ be a cost function. Given any $p > 0$, it holds true*

$$W_{c^p}^{(\infty)}(\mu, \nu) = (W_c^{(\infty)}(\mu, \nu))^p.$$

Proof Let $\pi \in \Pi(\mu, \nu)$ be a plan such that

$$T_c(\pi) = W_c^{(\infty)}(\mu, \nu),$$

then

$$W_{c^p}^{(\infty)}(\mu, \nu) \leq T_{c^p}(\pi) = T_c(\pi)^p = (W_c^{(\infty)}(\mu, \nu))^p.$$

Similarly, one can prove $(W_c^{(\infty)}(\mu, \nu))^p \leq W_{c^p}^{(\infty)}(\mu, \nu)$ and conclude the thesis. \square

Thanks to Lemma 6, we are able to prove the following result.

Theorem 7 *Given a cost function $c : X \times X \rightarrow [0, \infty)$, let $\mu, \nu \in \mathcal{P}(X)$ be two discrete measures. For any $p \geq 1$,*

$$W_c^{(\infty)}(\mu, \nu) \leq \frac{W_{c_p}(\mu, \nu)}{(\alpha_p)^{\frac{1}{p}}}, \tag{15}$$

where α_p is the constant defined in (9).

Proof Given a $p \geq 1$, let us denote with $\pi^{(p)}$ the trim optimal transportation plan between μ and ν according to the cost function c_p . Given a diffusive model for $\pi^{(p)}$, we denote with α_p the constant defined in (9). From Lemma 6 we have

$$W_{c_p}^{(\infty)}(\mu, \nu) = (W_c^{(\infty)}(\mu, \nu))^p,$$

hence, for any p , we have

$$(W_c^{(\infty)}(\mu, \nu))^p = W_{c_p}^{(\infty)}(\mu, \nu) \leq \frac{W_{c_p}^p(\mu, \nu)}{\alpha_p},$$

i.e.,

$$W_c^{(\infty)}(\mu, \nu) \leq \frac{W_{c_p}(\mu, \nu)}{(\alpha_p)^{\frac{1}{p}}}.$$

\square

In particular, since the constant α from Corollary 5 bounds from below every α_p and does not depend on the cost function but only on the starting measures μ and ν , we have

$$W_c^{(\infty)}(\mu, \nu) \leq \frac{W_{c_p}(\mu, \nu)}{(\alpha)^{\frac{1}{p}}}$$

for any $p \geq 1$. In particular, if we take

$$c(\mathbf{x}, \mathbf{y}) := \sqrt{\sum_{i=1}^n |x_i - y_i|^2},$$

we recover the bound proposed in Theorem 1 for discrete measures.

Remark 4 The estimate in (15) is sharp. To prove it, let us take

$$\mu = \delta_a \quad \text{and} \quad \nu = \delta_b$$

where $a, b \in \mathbb{R}^n$. By definition (9), we have $\alpha = 1$. Moreover, it is easy to see that

$$W^{(\infty)}(\mu, \nu) = |a - b| \quad \text{and} \quad W_p(\mu, \nu) = |a - b|,$$

which proves the sharpness of inequality (8).

Acknowledgements We are deeply indebted to Filippo Santambrogio for introducing us to the work of Bouchitté, Jimenez, and Mahadevan and for several stimulating discussions and valuable suggestions. We thank Stefano Gualandi for his feedback and Gabriele Loli for enhancing the images of this paper.

Funding Open access funding provided by Università degli Studi di Pavia within the CRUI-CARE Agreement. The authors have not disclosed any funding.

Declarations

Competing interest The authors have not disclosed any competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abdellaoui, T., Heinich, H.: Caractérisation d'une solution optimale au problème de Monge-Kantorovitch. *Bull. Soc. Math. France* **127**(3), 429–443 (1999)
2. Albertos, J. C., Matrán, C., Tuero-Díaz, A.: On the monotonicity of optimal transportation plans. *J Math Anal Appl.* **215**(1), 86–94. ISSN 0022-247X (1997)
3. Ambrosio, L., Gigli, N., Savaré, G.: *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Birkhäuser, Basel (2008)
4. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. *Proc. Mach. Learn. Res.* **70**(214–223), 06–11 (2017)
5. Bassetti, F., Regazzini, E.: Asymptotic properties and robustness of minimum dissimilarity estimators of location-scale parameters. *Soc. Ind. Appl. Math.* **50**, 312–330, 01 (2005) 10.4213/tvp109
6. Bassetti, F., Bodini, A., Regazzini, E.: On minimum Kantorovich distance estimators. *Stat. Probab. Lett.* **76**(12), 1298–1302 (2006). (<https://EconPapers.repec.org/RePEc:eee:stapro:v:76:y:2006:i:12:p:1298-1302>)
7. Bassetti, F., Gualandi, S., Veneroni, M.: On the computation of Kantorovich-Wasserstein distances between 2D-histograms by uncapacitated minimum cost flows. *SIAM J. Optim.* **30**(3), 2441–2469 (2020)
8. Bouchitté, G., Jimenez, C., Mahadevan, R.: A new L^∞ estimate in optimal mass transport. *Proc Am Math Soc*, 135:3525–3535, 11 (2007) <https://doi.org/10.1090/S0002-9939-07-08877-6>
9. Brenier, Y.: On the translocation of masses. *Commun. Pure Appl. Math.* **44**(4), 375–417 (1991)

10. Caffarelli, L.A., Feldman, M., McCann, R.J.: Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Am. Math. Soc.* **15**(1), 1–26 (2002). ISSN 08940347, 10886834. <http://www.jstor.org/stable/827090>
11. Cuturi, M., Doucet, A.: Fast computation of Wasserstein barycenters. *Proc. Mach. Learn. Res.* **32**(2), 685–693, 22–24 (2014)
12. Dantzig, G.B., Thapa, M.N.: *Linear Programming 1: Introduction*. Springer, Berlin (1997) 0387948333
13. Dobrushin, R.: Vlasov equations. *Funct. Anal. Appl.* **13**(2), 115–123 (1979)
14. Figalli, A.: Existence, uniqueness, and regularity of optimal transport maps. *SIAM J. Math. Anal.* **39**(1), 126–137 (2007)
15. Frogner, C., Zhang, C., Mobahi, H., Araya, M., Poggio, T. A.: Learning with a wasserstein loss. In: *Advances in Neural Information Processing Systems*, pp. 2053–2061 (2015)
16. Gangbo, W., McCann, R.J.: The geometry of optimal transportation. *Acta Math.* **177**(177), 113–161 (1996)
17. Hiroshi, M., Hiroshi, T.: An inequality for certain functional of multidimensional probability distributions. *Hiroshima Math.* **4**(1), 75–81 (1974)
18. Hiroshi, T.: An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas. *Z. Wahrscheinlichkeitstheorie Verwandte Gebiete* **27**, 47–52 (1973)
19. Hiroshi, T.: Probabilistic treatment of the Boltzmann equation of Maxwellian molecules. *Z. Wahrscheinlichkeitstheorie Verwandte Gebiete* **46**, 67–105 (1978)
20. Kantorovich, L.V.: Mathematical methods of organizing and planning production. *Manag. Sci.* **6**(4), 366–422 (1960)
21. Kantorovich, L.V.: On the translocation of masses. *J. Math. Sci.* **133**(4), 1381–1382 (2006)
22. Levina, E., Bickel, P.: The Earth Mover's Distance is the Mallows distance: Some insights from statistics. In: *Proceedings of the IEEE International Conference on Computer Vision*, Vol. 2, pp. 251–256, 02 (2001). <https://doi.org/10.1109/ICCV.2001.937632>
23. Loeper, G., et al.: On the regularity of solutions of optimal transportation problems. *Acta Math.* **202**(2), 241–283 (2009)
24. Monge, G.: Mémoire sur la théorie des déblais et des remblais. In: *Histoire de l'Académie Royale des Sciences de Paris* (1781)
25. Pele, O., Werman, M.: Fast and robust earth mover's distances. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 460–467 IEEE (2009)
26. Rubner, Y., Tomasi, C., Guibas, L.: Metric for distributions with applications to image databases. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 59–66, 02 (1998). <https://doi.org/10.1109/ICCV.1998.710701>
27. Rubner, Y., Tomasi, C., Guibas, L.J.: The Earth Mover's Distance as a metric for image retrieval. *Int. J. Comput. Vis.* **40**(2), 99–121 (2000)
28. Rüschendorf, L., Rachev, S.T.: A characterization of random variables with minimum L2-distance. *J. Multivar. Anal.* **32**(1), 48–54 (1990)
29. Santambrogio, F.: *Optimal Transport for Applied Mathematicians*. Springer, New York (2015)
30. Solomon, J., Rustamov, R., Guibas, L., Butscher, A.: Wasserstein propagation for semi-supervised learning. *Proc. Mach. Learn. Res.* **32**(1), 306–314, 22–24 (2014)
31. Villani, C.: *Optimal Transport: Old and New*, vol. 338. Springer, Berlin (2009)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.