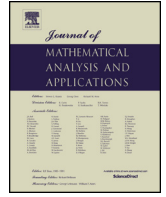




Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Journal of Mathematical Analysis and Applications

journal homepage: [www.elsevier.com/locate/jmaa](http://www.elsevier.com/locate/jmaa)



## Regular Articles

# $L^\infty$ -optimal transport for a class of strictly quasiconvex cost functions



Camilla Brizzi, Luigi De Pascale\*, Anna Kausamo

*Dipartimento di Matematica e Informatica, Università di Firenze, Viale Morgagni 67/a, 50134 Firenze, Italy*

### ARTICLE INFO

*Article history:*

Received 11 June 2021  
 Available online 20 April 2023  
 Submitted by G. Carlier

*Keywords:*

Optimal transport problem  
 $L^\infty$ -optimal transport  
 Wasserstein distances  
 Monge Kantorovich problem

### ABSTRACT

We consider the  $L^\infty$ - optimal mass transportation problem

$$\min_{\Pi(\mu, \nu)} \gamma - \text{ess sup}_{(x, y)} c(x, y),$$

for a new class of costs  $c(x, y)$  for which we introduce a tentative notion of twist condition. In particular we study the conditions under which the  $\infty$ -monotone transport plans are induced by a transportation map. We also state a uniqueness result for infinitely cyclically monotone Monge minimizers that corresponds to this class of cost functions. We compare the results to previous works.

© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

We consider two Polish spaces  $(X, d_X)$  and  $(Y, d_Y)$ . Let  $c : X \times Y \rightarrow \mathbb{R}^+$  be a cost function and  $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ ; here by  $\mathcal{P}(Z)$  we denote the space of Borel probability measures on a Polish space  $Z$ . Whereas in the classical optimal transport we minimize the integral of the cost function, in the  $L^\infty$ -optimal transport we consider the problem

$$\min_{\gamma \in \Pi(\mu, \nu)} C_\infty[\gamma] := \min_{\gamma \in \Pi(\mu, \nu)} \gamma - \text{ess sup}_{(x, y) \in X \times Y} c(x, y) \tag{P_\infty}$$

where

$$\Pi(\mu, \nu) := \{ \gamma \in \mathcal{P}(X \times Y) : \pi_\#^1 \gamma = \mu \text{ and } \pi_\#^2 \gamma = \nu \}.$$

\* Corresponding author.  
*E-mail addresses:* [camilla.brizzi@unifi.it](mailto:camilla.brizzi@unifi.it) (C. Brizzi), [luigi.depascale@unifi.it](mailto:luigi.depascale@unifi.it) (L. De Pascale), [akausamo@gmail.com](mailto:akausamo@gmail.com) (A. Kausamo).  
*URL:* <http://web.math.unifi.it/users/depascale/> (L. De Pascale).

Above  $\pi^1$  and  $\pi^2$  are the projections on the first and the second coordinate of  $X \times Y$ , respectively. Minimizers for this problem exist under mild assumptions on  $c$  such as lower-semicontinuity. In [5], Champion, De Pascale, and Juutinen carried out a comprehensive study on the  $L^\infty$ -transport in the case where the cost of transporting a point  $x$  to a point  $y$  is given by their distance:  $c(x, y) = |x - y|$  on  $\mathbb{R}^d$ . They studied the problem on compact sets of  $\mathbb{R}^d$ ,  $d \geq 1$ , and their work was generalized in 2015 by Jylhä [9] to Polish spaces with more general costs  $c$ . Champion, De Pascale, and Juutinen introduced the concept of  $\infty$ -cyclical monotonicity, which carries the well-known notion of cyclical monotonicity of optimal transportation plans to the  $L^\infty$ -case. In general, optimal  $L^\infty$ -transportation plans are not as well-behaved as the minimizers of the standard integral Monge-Kantorovich problem. For instance, their restrictions are not always optimal with respect to their marginals. To address this problem, Champion, De Pascale, and Juutinen invoked the concept of *restrictability* which is commonly used in more general  $L^\infty$ -calculus of variations problems. An optimal  $L^\infty$ -transportation plan is restrictable if, loosely speaking, its restrictions are also optimal. Moreover, they showed that restrictability and  $\infty$ -cyclical monotonicity are equivalent. This equivalence also holds for more general cost functions at least if they are continuous, as was proven by Jylhä in [9].

If  $X = Y$  and  $c = d$ , the minimal value  $C_\infty(\lambda)$  is called  $\infty$ -Wasserstein distance of the measures  $\mu$  and  $\nu$  and denoted by  $W_\infty(\mu, \nu)$ . The quantity  $W_\infty(\mu, \nu)$  is actually the  $p \rightarrow \infty$  limit of the  $p$ -Wasserstein distances  $W_p(\mu, \nu)$  of the measures  $\mu$  and  $\nu$ . Also for more general cost functions  $c$  we can consider for every  $p \geq 1$  the optimal transportation problem

$$\min_{\gamma \in \Pi(\mu, \nu)} C_p[\gamma] := \left( \int_{X \times Y} c^p(x, y) d\gamma \right)^{\frac{1}{p}}. \quad (\text{P}_p)$$

In the case  $c = d$  the  $W_\infty$ -distance provides a natural control from above for the  $W_p$ -distances. Bouchitté, Jimenez, and Rajesh [3] and, for even more general costs, Jylhä and Rajala [10] also established necessary and sufficient conditions for the existence of  $W_\infty$ -lower bounds for the integral optimal transportation costs.

The question on whether there exists a dual formulation for the  $L^\infty$ -transport, similar to the now-standard Kantorovich duality [11,12], remained open until Barron, Bocea, and Jensen stated and proved a duality theorem in 2017 [2]. The theory was further developed in the 1-dimensional case by De Pascale and Louet in [6]. Unlike in the standard integral optimal transportation, it is not immediate how to use the  $L^\infty$ -duality to prove the existence of deterministic solutions to the Monge-Kantorovich problem. By deterministic solutions we mean minimizers for the problem  $(\text{P}_\infty)$  of the form  $\gamma = (Id \times T)_\# \mu$ , where  $T$  belongs to the set

$$\mathcal{T}(\mu, \nu) := \{T : X \rightarrow Y : T \text{ is a Borel map and } \nu = T_\# \mu\}.$$

These solutions are also called Monge solutions, in honor of Gaspard Monge, who is one of the building fathers of the classical theory of optimal transportation.

In this paper we use a duality-free technique to prove that, under natural assumptions on  $\mu$ ,  $\infty$ -cyclical monotone transport plans (see Definition 2.2) – or, even less,  $\infty$ -monotone transport plans (these do not need to be optimal) – are actually induced by a transport map (see Theorem 2.19). The technique was introduced by Champion, De Pascale, and Juutinen in [5] where the authors used this property to prove on  $\mathbb{R}^d$ , in the case where the cost is given by the Euclidean distance, that there exists a deterministic solution for the problem  $(\text{P}_\infty)$  and, if the target measure  $\nu$  is atomic, that restrictable, deterministic solutions are unique. Jylhä in [9] generalized the existence result to cost functions of the form  $c(x, y) = h(y - x)$ , where  $h : \mathbb{R}^d \rightarrow \mathbb{R}^+$  is strictly quasiconvex in the sense that for all  $t \in (0, 1)$  and  $x, \bar{x} \in \mathbb{R}^d$  with  $x \neq \bar{x}$

$$h((1-t)x + t\bar{x}) < \max\{h(x), h(\bar{x})\}.$$

He also extended the uniqueness result of Champion, De Pascale and Juutinen, proving that if  $\nu$  contains an atom, say  $y_0$ , and  $T$  and  $S$  are two optimal transport maps, both corresponding to infinitely cyclically monotone transport plans, then  $\mu(T^{-1}(\{y_0\}) \setminus S^{-1}(\{y_0\})) = 0$ .

The relevant notion in the Monge-Kantorovich problem for integral costs is an invertibility condition on the partial gradient of  $c$  [4,7]. This condition is commonly used in dynamical systems, it is called the twist condition and it goes very well along with convexity properties. There is no analogue of the twist condition for  $L^\infty$ -optimal transport problem. In this study we introduce a property which could serve as twist condition in this setting (see Theorem 2.17). And, in addition to giving some examples of costs which have this property, we observe that it is satisfied by some costs studied in previous works. The property we introduce is invariant (as one could expect) by composition of  $c$  with a strictly increasing, differentiable function.

Although we do not change the essence of the proof of uniqueness, we give a new structure to it which, in our opinion, makes the arguments more transparent and may allow a better understanding of the problem.

The proof of the main results is based on measure theoretic considerations and on the construction of certain specific cones. The same happens in [9] as well as in previous works in which these techniques are applied. What is different is that the construction of [9] is based on the translation invariance of the cost while our construction relies on the notion of normal cone to the boundary of a convex set. This points to the possibility of a general construction which may be adapted to different costs.

A different technique, which applies to several costs of the form  $c(x, y) = h(x - y)$ , to prove the existence of an optimal transport map, was devised by C. Jimenez and F. Santambrogio in [8]. In that paper the authors minimize  $\int |x - y|^2 d\gamma + \chi(\gamma)$  among the  $\gamma \in \Pi(\mu, \nu)$  and

$$\chi(\gamma) = \begin{cases} 0 & \text{if } \text{supp}(\gamma) \subset \{(x, y) \mid c(x, y) \leq M\}; \\ +\infty & \text{otherwise.} \end{cases}$$

They prove that if  $\gamma$  is optimal then it is deterministic. However the optimal map of [8] is different from the one we describe here.

## 2. Basic definitions and main theorem

First we recall some basic results about  $L^\infty$ -transport, as first established in [5] and then generalized in [9]. The proofs can be found in these references.

The first important result is the existence of minimizers for the  $L^\infty$ -transportation problem. It is a direct consequence of the lower-semicontinuity of the functional  $C_\infty$  and the compactness of the set  $\Pi(\mu, \nu)$ .

**Proposition 2.1.** *Let  $c$  be lower semicontinuous. Then the problem  $(P_\infty)$  has at least one minimizer.*

In general, there can be a high level of non-uniqueness of minimizers of the functional  $C_\infty$ : we are only optimizing the “worst case” and more locally the situation can be far from optimal. It is often useful to consider a subset of better-behaving minimizers, the so-called  $\infty - c$ -cyclically monotone (ICM) transport plans. The idea is that on the support of such a plan, there is no rearrangement of destinations of a fixed finite set of points that would improve the highest cost coupling of that set. To our purpose it is also useful to have a definition for a plan that can not be improved by the interchange of any two destinations of a given pair of initial points. The formal definitions are the following:

**Definition 2.2.** A set  $\Gamma \subset X \times Y$  is  $\infty - c$ -cyclically monotone (ICM) if for every finite set of points  $\{(x_i, y_i)\}_{i=1}^k \subset \Gamma$  and for every permutation  $\sigma$  of the set  $\{1, \dots, k\}$  we have

$$\max_{1 \leq i \leq k} c(x_i, y_i) \leq \max_{1 \leq i \leq k} c(x_i, y_{\sigma(i)}).$$

We say that a transport plan  $\gamma \in \Pi(\mu, \nu)$  is ICM, if it is concentrated on an ICM set.

**Definition 2.3.** The set  $\Gamma \subset X \times Y$  is  $\infty$ -monotone (IM) if for any  $(x, y), (x', y') \in \Gamma$  we have

$$\max\{c(x, y), c(x', y')\} \leq \max\{c(x, y'), c(x', y)\}.$$

We say that a transport plan  $\gamma \in \Pi(\mu, \nu)$  is IM, if it is concentrated on an IM set.

**Remark 2.4.** Every ICM plan is IM but the reverse is not true in general. Consider, for example, the marginal measures  $\mu = \nu$  to be the uniform probability measure on the circle  $\mathbb{S}^1$ , the cost function  $c(x, y) = |x - y|$  (the Euclidean distance), and the set  $\Gamma = \text{Graph}(R_\theta)$ , where  $R_\theta$  is the rotation of a vector  $v \in \mathbb{S}^1$  by an angle  $\theta < \pi/2$ . Now  $\Gamma$  is IM, but since the transport plan supported by  $\Gamma$  is not optimal (the optimal cost being 0, given by the identity map) the set  $\Gamma$  cannot be ICM.

$\infty$ - $c$ -cyclically monotone transport plans are much better-behaved than arbitrary solutions to the problem  $(P_\infty)$ . For example, they are restrictable in the sense of the following definition.

**Definition 2.5.** A transport plan  $\gamma \in \Pi(\mu, \nu)$  is said to be a restrictable minimizer of the problem  $(P_\infty)$  if it satisfies the following condition: for any  $\gamma' \leq \gamma$ , i.e.  $\gamma'(B) \leq \gamma(B)$  for every Borel sets  $B$ , and  $\gamma'(X \times Y) > 0$  we have

$$C_\infty(\bar{\gamma}) = \min \{C_\infty(\lambda) : \lambda \in \Pi(\pi_{\#}^1 \bar{\gamma}, \pi_{\#}^2 \bar{\gamma})\},$$

$$\text{where } \bar{\gamma} = \frac{\gamma'}{\gamma'(X \times Y)}.$$

They exist under relatively mild conditions as the following lemma, taken from [9], states.

**Lemma 2.6.** *Let  $c$  be lower semicontinuous. Let  $\gamma_p$  be a solution of the problem  $(P_p)$ ,  $p \geq 1$ . Then weak\* cluster points of  $(\gamma_p)_{p \geq 1}$  exist and each of them is a solution of the problem  $(P_\infty)$ . If, in addition,  $c$  is continuous and the minimum of each problem  $(P_p)$  is finite, then the weak\* cluster points are ICM (here weak\* convergence is tested against continuous and bounded functions).*

Since in this paper we are interested in the IM plans, namely in proving that they are of the Monge-type, it is useful to state their existence:

**Corollary 2.7.** *If  $c$  is continuous and the minimum of the problem  $(P_p)$  is finite for every  $p \geq 1$ , then the problem  $(P_\infty)$  has a solution  $\gamma$  that is IM.*

In the following we state some definitions and results which are necessary for the proof of Theorem 2.17. They can also be found in [5].

**Definition 2.8.** Let  $y \in \mathbb{R}^d$ ,  $r > 0$  and let  $\gamma \in \Pi(\mu, \nu)$  be a transport plan. We define

$$\gamma^{-1}(B(y, r)) := \pi^1((\mathbb{R}^d \times B(y, r)) \cap \text{supp } \gamma)$$

In other words,  $\gamma^{-1}(B(y, r))$  is the set of points whose mass is partially or completely transported to  $B(y, r)$  by  $\gamma$ , and  $\gamma^{-1}$  corresponds to the inverse of the multimap induced by  $\text{supp } \gamma$ . Notice also that  $\gamma^{-1}(B(y, r))$  is a Borel set.

Since this notion is important in the sequel, we recall that when  $U$  is a Borel set, one has

$$\lim_{r \rightarrow 0^+} \frac{\mathcal{L}^d(U \cap B(x, r))}{\mathcal{L}^d(B(x, r))} = 1,$$

for almost every  $x$  in  $U$ : we shall call such a point  $x$  a Lebesgue point of  $U$  and we will denote by  $\text{Leb}(U)$  the Borel set of Lebesgue points.

**Definition 2.9.** We say that the couple  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$  is a  $\gamma$ -regular point if  $x \in \text{Leb}(\gamma^{-1}(B(y, r)))$  for any positive  $r$ . We denote

$$R(\gamma) := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : (x, y) \text{ is a } \gamma\text{-regular point}\}. \tag{2.1}$$

**Remark 2.10.** Notice first that by the closedness of  $\text{supp } \gamma$  we have that  $R(\gamma) \subset \text{supp } \gamma$ . Moreover it is not difficult to show that  $R(\gamma)$  is a Borel set. Indeed, if for every fixed  $n \in \mathbb{N}$  we consider a countable covering  $\mathbb{R}^d \subset \bigcup_{i \in \mathbb{N}} B(y_{i,n}, 2^{-n})$  of balls of radius  $2^{-n}$ , then

$$R(\gamma) = \bigcap_{n \geq 1} \bigcup_{i \in \mathbb{N}} (\text{Leb}(\gamma^{-1}(B(y_{i,n}, 2^{-n}))) \times B(y_{i,n}, 2^{-n})). \tag{2.2}$$

**Lemma 2.11.** Let  $\gamma \in \Pi(\mu, \nu)$  and assume that  $\mu \ll \mathcal{L}^d$ . Then  $\gamma$  vanishes outside the set  $R(\gamma)$  of  $\gamma$ -regular points.

**Proof.** Let us denote by  $S := \text{supp } \gamma \setminus R(\gamma)$  the complement of the  $\gamma$ -regular points, intersected with the support of  $\gamma$ , i.e.

$$S := \{(x, y) \in \text{supp } \gamma : x \notin \text{Leb}(\gamma^{-1}(B(y, r))) \text{ for some positive } r\}.$$

One can prove that  $S$  is, actually, equal to the set

$$\bigcup_{n \geq 1} \bigcup_{i \in \mathbb{N}} ((\gamma^{-1}(B(y_{i,n}, 2^{-n})) \setminus \text{Leb}(\gamma^{-1}(B(y_{i,n}, 2^{-n})))) \times B(y_{i,n}, 2^{-n})) \cap \text{supp } \gamma.$$

In particular we have that

$$\pi^1(S) = \bigcup_{n \geq 1} \bigcup_{i \in \mathbb{N}} \left( \gamma^{-1} \left( B(y_{i,n}, \frac{1}{2^n}) \right) \setminus \text{Leb} \left( \gamma^{-1} \left( B(y_{i,n}, \frac{1}{2^n}) \right) \right) \right),$$

and the set on the right-hand side has Lebesgue measure 0. Therefore, by absolute continuity of  $\mu$  with respect to  $\mathcal{L}^d$  we have  $\mu(\pi^1(S)) = 0$ . Finally

$$\gamma(S) \leq \gamma(\pi^1(S) \times \mathbb{R}^d) = \mu(\pi^1(S)) = 0. \quad \square$$

**Notation.** At this point, it is natural to introduce a more refined definition

$$\gamma_*^{-1}(B(y, r)) := \pi^1((\mathbb{R}^d \times B(y, r)) \cap R(\gamma)).$$

For future use we also introduce a suitable notation for a cone: let  $x_0, \xi \in \mathbb{R}^d$  such that  $|\xi| = 1$  and  $\delta \in [0, 2]$  then

$$K(x_0, \xi, \delta) := \left\{ x \in \mathbb{R}^d \setminus \{x_0\} : \frac{x - x_0}{|x - x_0|} \cdot \xi \geq 1 - \delta \right\} \cup \{x_0\},$$

and for a “truncated” cone

$$K(x_0, \xi, \delta, s) := K(x_0, \xi, \delta) \cap B(x_0, s). \quad (2.3)$$

**Proposition 2.12.** *Let  $\mu \ll \mathcal{L}^d$ , let  $(x_0, y_0) \in R(\gamma)$  and let  $r > 0$ , then for every  $\xi \in \partial B(0, 1)$ ,  $\delta \in (0, 2]$  and  $s > 0$  it holds:*

$$\mathcal{L}^d(\gamma_*^{-1}(B(y_0, r)) \cap K(x_0, \xi, \delta, s)) > 0.$$

*In general, if  $A$  is a Borel set and  $x_0 \in \text{Leb}(A)$  it holds:*

$$\mathcal{L}^d(A \cap \gamma_*^{-1}(B(y_0, r)) \cap K(x_0, \xi, \delta, s)) > 0.$$

**Proof.** By definition of  $R(\gamma)$ ,  $x_0$  is a Lebesgue point of  $\gamma^{-1}(B(y_0, r))$  which implies  $\mathcal{L}^d(\gamma^{-1}(B(y_0, r)) \cap K(x_0, \xi, \delta, s)) > 0$  for every  $\xi \in \partial B(0, 1)$ ,  $\delta \in (0, 2]$ ,  $s > 0$ . Thus, since the boundary of  $K(x_0, \xi, \delta, s)$  is a Lebesgue-null set, there exists

$$(x, y) \in (\text{int } K(x_0, \xi, \delta, s) \times B(y_0, r)) \cap \text{supp } \gamma$$

such that  $x \neq x_0$ . We take  $\rho > 0$  small enough so that  $B(x, \rho) \subset K(x_0, \xi, \delta, s)$  and  $B(y, \rho) \subset B(y_0, r)$ . We know that  $\gamma(B(x, \rho) \times B(y, \rho)) > 0$  because  $(x, y) \in \text{supp } \gamma$ , and since  $\gamma$  is concentrated in  $R(\gamma)$  we have:

$$\begin{aligned} 0 &< \gamma((B(x, \rho) \times B(y, \rho)) \cap R(\gamma)) \\ &\leq \gamma((K(x_0, \xi, \delta, s) \times B(y_0, r)) \cap R(\gamma)) \\ &\leq \gamma(K(x_0, \xi, \delta, s) \cap \gamma_*^{-1}(B(y_0, r)) \times \mathbb{R}^d) \\ &= \mu(K(x_0, \xi, \delta, s) \cap \gamma_*^{-1}(B(y_0, r))). \end{aligned}$$

Which in turn implies  $\mathcal{L}^d(K(x_0, \xi, \delta, s) \cap \gamma_*^{-1}(B(y_0, r))) > 0$  by the fact that  $\mu \ll \mathcal{L}^d$ .  $\square$

**Corollary 2.13.** *Let  $(x_0, y_0) \in R(\gamma)$  and let  $C$  be a convex set with  $\text{int } C \neq \emptyset$  such that  $x_0 \in C$ . Then for every  $r > 0$*

$$\mathcal{L}^d(\gamma_*^{-1}(B(y_0, r) \cap \text{int } C)) > 0.$$

*Moreover, if  $A$  is a Borel set and  $x_0 \in \text{Leb}(A)$*

$$\mathcal{L}^d(A \cap \gamma_*^{-1}(B(y_0, r) \cap \text{int } C)) > 0.$$

**Proof.** We observe that by convexity of  $C$  there exists a cone  $K(x_0, \xi, \delta)$ , for some  $\xi \in \partial B(0, 1)$  and  $\delta \in (0, 1)$ , such that for  $s$  sufficiently small

$$K(x_0, \xi, \delta, s) \subset \text{int } C \cup \{x_0\}.$$

By the monotonicity of the Lebesgue measure, the claim now follows from Proposition 2.12.  $\square$

Before the first main result of this paper we introduce quasiconvex and strictly quasiconvex functions. We also present a technical lemma on convex sets that will be useful in the sequel.

**Definition 2.14.** We say that a function  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  is quasiconvex if for all  $x, y \in \mathbb{R}^d$  and for all  $t \in [0, 1]$  we have

$$h((1 - t)x + ty) \leq \max\{h(x), h(y)\}.$$

We say that  $h$  is strictly quasiconvex if for all  $x \neq y$  and for all  $t \in (0, 1)$

$$h((1 - t)x + ty) < \max\{h(x), h(y)\}.$$

**Remark 2.15.** Immediate properties are the following

- (1)  $h$  is quasiconvex if and only if for all  $\lambda \in \mathbb{R}$  the sublevel set  $C_\lambda = \{x : h(x) \leq \lambda\}$  is convex (possibly empty);
- (2) if  $h$  is continuous and strictly quasiconvex, then the only sublevel sets (possibly empty) with empty interior are the sets  $C_m$  with  $m \leq \inf h$ ;
- (3) if  $h$  is continuous and strictly quasiconvex, then the level set  $\{x : h(x) = \lambda\}$  has always empty interior.

**Lemma 2.16.** Let  $B, C \subset \mathbb{R}^d$  be two closed convex sets with nonempty interiors, differentiable boundaries, and such that  $B \cap C \neq \emptyset$ . Let  $x \in \partial B \cap \partial C$  be such that  $n_B(x) \neq n_C(x)$ , where  $n_B(x)$  and  $n_C(x)$  are the unit outer normals of  $B$  and  $C$  at  $x$ , respectively. Then there exists a point  $a \in \text{int } B \cap \{w : (w - x) \cdot n_C(x) > 0\}$  and  $\delta, s > 0$  such that the intersection of the cone of direction  $a - x$  and amplitude  $\delta$  with the ball centered at  $x$  and of radius  $s$  is all contained in  $\text{int } B$ , that is

$$K\left(x, \frac{a - x}{|a - x|}, \delta, s\right) \subset \text{int } B \cup \{x\}. \tag{2.4}$$

Moreover for the reverse cone we have

$$K\left(x, -\frac{a - x}{|a - x|}, \delta, s\right) \subset \text{int } C \cup \{x\}. \tag{2.5}$$

**Proof.** By the differentiability of the boundaries of sets  $B$  and  $C$  we have that  $n_B(x)$  and  $n_C(x)$  are well-defined elements of  $\mathbb{R}^d$ . In particular, by convexity  $n_B(x)$  is the only unit vector such that

$$(b - x) \cdot n_B(x) \leq 0, \quad \text{for every } b \in B.$$

Hence there exists  $a \in \text{int } B$  such that

$$(a - x) \cdot n_C(x) > 0. \tag{2.6}$$

Choosing  $s$  and  $\delta$  close enough to 0, by the continuity of the scalar product and the convexity of  $B$  we can fix the cone  $K\left(x, \frac{a - x}{|a - x|}, \delta, s\right)$  that satisfies the first claim (2.4).

Moreover, by Condition (2.6) we have  $-\frac{a - x}{|a - x|} \cdot n_C(x) < 0$ . Thus for a suitable choice of  $\tau > 0$ , we have<sup>1</sup>  $x -$

---

<sup>1</sup> There exist  $\rho > 0$  and a concave function  $\phi : B^{d-1}(\mathbf{0}, \rho) \rightarrow \mathbb{R}$  such that  $x = (\mathbf{0}, \phi(\mathbf{0}))$  and such that  $\partial C$  is the graph of  $\phi$  in a sufficiently small neighborhood of  $x$ . In this setting  $n_C(x) = \frac{1}{\sqrt{1 + |\nabla \phi(\mathbf{0})|^2}}(-\nabla \phi(\mathbf{0}), 1)$ . We use the notation  $(\mathbf{v}, v_d) = v = -\frac{a - x}{|a - x|}$ , where  $\mathbf{v} \in \mathbb{R}^{d-1}$  and  $v_d \in \mathbb{R}$ . The condition  $v \cdot n_C(x) < 0$  becomes  $\mathbf{v} \cdot \nabla \phi(\mathbf{0}) > v_d$ . Proving that  $x + \tau v \in \text{int } C$ , for some  $\tau > 0$ , can be done by showing that  $\phi(\tau \mathbf{v}) > \phi(\mathbf{0}) + \tau v_d$ . The latter inequality follows directly by the fact that  $\lim_{t \rightarrow 0} \frac{\phi(t\mathbf{v}) - \phi(\mathbf{0})}{t} > v_d$ .

$\frac{a-x}{|a-x|} \tau \in \text{int } C$ . Arguing as before, there exists  $s'$  and  $\delta'$  close enough to 0, such that  $K\left(x, -\frac{a-x}{|a-x|}, \delta', s'\right)$  satisfies the (2.5). Therefore, possibly substituting  $s$  with  $\min\{s, s'\}$  and  $\delta$  with  $\min\{\delta, \delta'\}$ , (2.4) and (2.5) hold simultaneously.  $\square$

**Notation.** Given  $y \in \mathbb{R}^d$  and  $\lambda \in \mathbb{R}$  we denote, whenever it is possible, by

$$n_{c(\cdot, y)}(x)$$

the unit outer normal to the sublevel set  $C_\lambda = \{z : c(z, y) \leq \lambda\}$  at the point  $x \in \partial C_\lambda$ .

**Theorem 2.17.** Let  $\mu, \nu$  be two Borel probability measures on  $\mathbb{R}^d$  with compact supports and  $\mu \ll \mathcal{L}^d$ . Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$  be a continuous function satisfying the following properties:

- (i) For all  $x$  there exists a unique  $y \in \mathbb{R}^d$  such that  $c(x, y) = 0$ , and for every  $y \in \mathbb{R}^d$  we have<sup>2</sup>

$$\inf_{x \in \mathbb{R}^d} c(x, y) = 0;$$
- (ii)  $c(\cdot, y)$  is strictly quasiconvex for every  $y$ , i.e.

$$c((1-t)x + t\bar{x}, y) < \max\{c(x, y), c(\bar{x}, y)\}$$

for all  $t \in (0, 1)$  and for every  $x, \bar{x} \in \mathbb{R}^d$ ,  $x \neq \bar{x}$ ;

- (iii) for all  $\lambda > 0$  and  $y \in \mathbb{R}^d$  the convex set  $C_\lambda = \{z : c(z, y) \leq \lambda\}$  has differentiable ( $\mathcal{C}^1$ ) boundary;
- (iv)  $c$  satisfies a “twist kind” condition, that is: for every  $x, y, \tilde{y} \in \mathbb{R}^d$  and for all  $\lambda > 0$

$$\begin{cases} c(x, y) = c(x, \tilde{y}) = \lambda \\ n_{c(\cdot, y)}(x) = n_{c(\cdot, \tilde{y})}(x) \end{cases} \implies y = \tilde{y}.$$

Let  $\gamma \in \Pi(\mu, \nu)$  and  $(x, \tilde{y}) \in R(\gamma)$ . Then for every  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$  with  $y \neq \tilde{y}$ , there exists  $(x', y') \in R(\gamma)$  such that

$$\max\{c(x', y), c(x, y')\} < \max\{c(x, y), c(x', \tilde{y})\}. \quad (2.7)$$

Moreover, if  $x \in \text{Leb}(A)$ , for some Borel set  $A$ , then  $(x', y')$  above can be taken with  $x' \in A$ .

**Remark 2.18.** We call condition (iv) above “twist kind” because whenever  $c$  is differentiable with respect to  $x$ , if  $\nabla_x c(x, \tilde{y}) \neq 0$  and  $\nabla_x c(x, y) \neq 0$ , then the condition becomes

$$\begin{cases} c(x, y) = c(x, \tilde{y}) \\ \frac{\nabla_x c(x, y)}{|\nabla_x c(x, y)|} = \frac{\nabla_x c(x, \tilde{y})}{|\nabla_x c(x, \tilde{y})|} \end{cases} \implies y = \tilde{y}.$$

This looks like the twist condition in the classical integral optimal transport.

**Proof. Case 1:**  $c(x, y) > c(x, \tilde{y})$ .

We fix  $\lambda = c(x, y)$  and we consider  $C_\lambda = \{z : c(z, y) \leq \lambda\}$ . Notice that  $\text{int } C_\lambda$  is nonempty by part (2) of

<sup>2</sup> By the strict quasiconvexity of  $c(\cdot, y)$  (Assumption (ii)), if there exists a point  $x$  such that  $c(x, y) = \inf_{z \in \mathbb{R}^d} c(z, y)$ , then such a point is also unique.



Remark 2.15. Indeed  $c(x, y) = \lambda > 0 = \inf_{w \in \mathbb{R}^d} c(w, \tilde{y})$ , by Assumption (i). The continuity of the cost function allows us to fix  $r > 0$  such that

$$c(x, y') < c(x, y) \quad \text{for all } y' \in B(\tilde{y}, r). \tag{2.8}$$

Moreover, since  $(x, \tilde{y}) \in R(\gamma)$ , thanks to Corollary 2.13, we know that there exists  $(x', y') \in R(\gamma)$  with  $x' \in \text{int } C_\lambda = \{z : c(z, y) < \lambda\}$  (so that  $x \neq x'$ ) and  $y' \in B(\tilde{y}, r)$  that is

$$(x', y') \in (\text{int } C_\lambda \times B(\tilde{y}, r)) \cap R(\gamma).$$

Notice that such a  $y'$  exists by the definition of  $\gamma_*^{-1}(B(\tilde{y}, r))$ . By Condition (2.8) we have

$$c(x, y') < c(x, y).$$

We also have that

$$c(x', y) < c(x, y),$$

because  $x' \in \text{int } C_\lambda$ . These two inequalities imply that

$$\max\{c(x', y), c(x, y')\} < c(x, y) \leq \max\{c(x, y), c(x', y')\},$$

concluding the proof for the Case 1.

**Case 1 bis:** If  $x \in \text{Leb}(A)$ , for a Borel set  $A$ , then by the second part of Corollary 2.13, we can find  $(x', y') \in R(\gamma)$ , such that  $x' \in \text{int } C_\lambda \cap A$  and  $y \in B(\tilde{y}, r)$ . As before, such  $(x', y')$  satisfies the (2.7).

**Case 2:**  $c(x, y) < c(x, \tilde{y})$ .

We follow a path similar to that of [9] but with a different construction of cones.

Fix  $\lambda = c(x, \tilde{y})$ , so that  $x \in \partial \tilde{C}_\lambda$ , where  $\tilde{C}_\lambda = \{z : c(z, \tilde{y}) \leq \lambda\}$ . Again we denote by  $n_{c(\cdot, \tilde{y})}(x)$  the unit outer normal to the set  $\tilde{C}_\lambda$  at  $x$ .

Let  $s$  be a small, positive real number such that  $\tilde{a} := x - sn_{c(\cdot, \tilde{y})}(x) \in \text{int } \tilde{C}_\lambda$ , which is not empty thanks to (2) of Remark 2.15. We fix

$$\varepsilon = \frac{1}{4} \min \{c(x, \tilde{y}) - c(\tilde{a}, \tilde{y}), c(x, \tilde{y}) - c(x, y)\}.$$

Thanks to the continuity of  $c$  it is possible to find a positive radius  $r$  such that  $B(\tilde{a}, r) \subset \tilde{C}_\lambda$ , for every  $(\tilde{z}, y') \in B(\tilde{a}, r) \times B(\tilde{y}, r)$  we have

$$c(\tilde{z}, y') < c(\tilde{a}, \tilde{y}) + \varepsilon \tag{2.9}$$

and for every  $y' \in B(\tilde{y}, r)$  it holds

$$c(x, \tilde{y}) - \varepsilon < c(x, y'). \tag{2.10}$$

Consider the set<sup>3</sup>  $\tilde{K} = \{(1-t)x + tB(\tilde{a}, r) : t \in (0, 1)\}$ . Thanks to the convexity,  $\tilde{K}$  is contained in  $\tilde{C}_\lambda$ . We claim that

$$\tilde{K} \subset \{z : c(z, y') < c(x, y')\} \quad \text{for every } y' \in B(\tilde{y}, r).$$

<sup>3</sup> We use the notation  $\tilde{K}$  because also this set is a cone, even if slightly different from the cones defined by (2.3).

Let  $z \in \tilde{K}$ . Then  $z = (1-t)x + t\tilde{z}$ , for some  $\tilde{z} \in B(\tilde{a}, r)$  and some  $t \in (0, 1)$ . We observe that  $c(\tilde{z}, y') < c(x, y')$ . Indeed

$$c(\tilde{z}, y') < c(\tilde{a}, \tilde{y}) + \varepsilon \leq \frac{1}{4}c(x, \tilde{y}) + \frac{3}{4}c(\tilde{a}, \tilde{y}) < c(x, \tilde{y}) - 3\varepsilon < c(x, y'), \quad (2.11)$$

where the first inequality is due to Condition (2.9), the second and the third to the definition of  $\varepsilon$  and the last one to (2.10). We conclude by the strict quasiconvexity of  $c(\cdot, y')$ , which implies  $c(z, y') < \max\{c(x, y'), c(\tilde{z}, y')\} = c(x, y')$ .

On the other hand, if we consider the reflected cone  $K = (1-t)x + tB(\tilde{a}, r)$  with  $t \in (-1, 0)$ , again by the strict quasiconvexity of  $c(\cdot, y')$  we show that

$$K \subset \{z : c(z, y') > c(x, y')\} \quad \text{for every } y' \in B(\tilde{y}, r).$$

Indeed, let  $z \in K$ , then  $z = (1-t)x + t\tilde{z}$ , for some  $\tilde{z} \in B(\tilde{a}, r)$  and  $t \in (-1, 0)$ . So  $x = sz + (1-s)\tilde{z}$ , with  $s = \frac{1}{1-t}$ . Thus  $c(x, y') < \max\{c(z, y'), c(\tilde{z}, y')\} = c(z, y')$ , where the last equality is due to (2.11), and therefore for every  $x' \in K$  and  $y' \in B(\tilde{y}, r)$  it holds

$$c(x, y') < c(x', y').$$

Let now  $\rho > 0$  such that  $c(x', y) < c(x, y) + \varepsilon$  for every  $x' \in K \cap B(x, \rho)$ . Then for every  $x' \in K \cap B(x, \rho)$  and  $y' \in B(\tilde{y}, r)$  it also holds:

$$c(x', y) < c(x, y) + \varepsilon < \frac{1}{4}c(x, \tilde{y}) + \frac{3}{4}c(x, y) \leq c(x, \tilde{y}) - 3\varepsilon < c(x, y') < c(x', y').$$

For  $\rho$  small enough  $K \cap B(x, \rho)$  is a truncated cone of the form (2.3) and then the proof of Case 2 is concluded thanks to Proposition 2.12, for which it is possible to choose  $x'$  in the set

$$(B(x, \rho) \cap K \cap \gamma_*^{-1}(B(\tilde{y}, r)))$$

and therefore  $(x', y')$  in the set

$$((K \cap B(x, \rho)) \times B(\tilde{y}, r)) \cap R(\gamma).$$

**Case 2 bis:** If  $x \in \text{Leb}(A)$  for some Borel set  $A$ , by the last part of Proposition 2.12, we can choose  $(x', y')$  satisfying the (2.7) in the set

$$((A \cap K \cap B(x, \rho)) \times B(\tilde{y}, r)) \cap R(\gamma).$$

**Case 3:**  $c(x, y) = c(x, \tilde{y}) = \lambda > 0$  (if  $c(x, y) = c(x, \tilde{y}) = 0$  then  $y = \tilde{y}$  by Assumption (i)) and  $n_{c(\cdot, y)}(x) \neq n_{c(\cdot, \tilde{y})}(x)$ .

In Fig. 2.1 we provide a visual description of the proof of this case, where we have denoted the vector  $n_{c(\cdot, \tilde{y})}(x)$  by  $\vec{n}$ .

Since  $x \in \partial C_\lambda \cap \partial \tilde{C}_\lambda$  and since by part (2) of Remark 2.15 these sublevel sets have nonempty interiors, the assumptions of Lemma 2.16 above are satisfied. Therefore we can construct two “truncated” cones, one inside  $\text{int } C_\lambda \cup \{x\}$  and the “reverse” one contained in  $\text{int } \tilde{C}_\lambda$ . More precisely there exist  $a \in \text{int } C_\lambda$  and  $\delta, s > 0$  such that

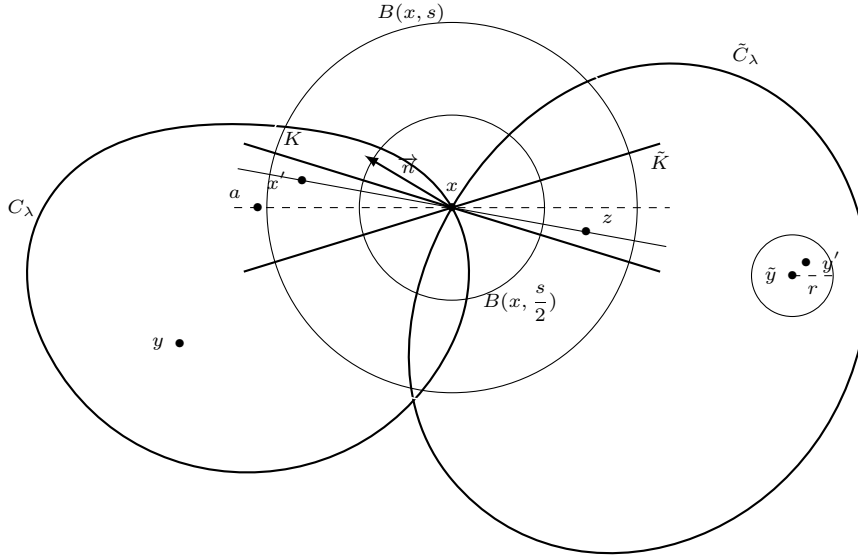


Fig. 2.1. Case 2 - Theorem 2.17.

$$K \left( x, \frac{a-x}{|a-x|}, \delta, s \right) \subset \text{int } C_\lambda \cup \{x\} \text{ and}$$

$$K \left( x, -\frac{a-x}{|a-x|}, \delta, s \right) \subset \text{int } \tilde{C}_\lambda \cup \{x\}.$$

In order to simplify the notation, let us denote by

$$K := K \left( x, \frac{a-x}{|a-x|}, \delta, s \right) \quad \text{and} \quad \tilde{K} := K \left( x, -\frac{a-x}{|a-x|}, \delta, s \right).$$

Since  $c(x, \tilde{y}) > c(z, \tilde{y})$  for every  $z \in \text{int } \tilde{C}_\lambda$ , if we take  $z$  in a set well-contained in  $\text{int } \tilde{C}_\lambda$ , for instance the portion of annulus  $\tilde{K} \setminus B(x, \frac{s}{2})$ , by the continuity of the cost there exists  $r > 0$  such that

$$c(x, y') > c(z, y'), \quad \text{for all } y' \in B(\tilde{y}, r), \text{ for all } z \in \tilde{K} \setminus B(x, \frac{s}{2}). \tag{2.12}$$

Let  $(x', y') \in (K \times B(\tilde{y}, r)) \cap R(\gamma)$  with  $x' \neq x$ , which exists by Proposition 2.12 (and definition of  $\gamma_*^{-1}(\tilde{y}, r)$ ). Consider  $z \in \tilde{K} \setminus B(x, \frac{s}{2})$  on the line passing from  $x'$  and  $x$  (the order of the points on the line being  $x', x, z$ ), by the strict quasiconvexity of  $c$  with respect to the first variable we have

$$c(x, y') < \max\{c(x', y'), c(z, y')\}.$$

By Condition (2.12), we infer that  $c(x, y') < c(x', y')$ .

Finally, since  $x' \in \text{int } C_\lambda$ , we have  $c(x', y) < c(x, y)$  (again we use that, by the continuity and the strict quasiconvexity of  $c$ ,  $\partial C_\lambda = \{z : c(z, y) = \lambda\}$ ). Hence Condition (2.7) holds.

**Case 3 bis:** Also in this case thanks to the second part of Proposition 2.12, we can choose our  $(x', y') \in ((A \cap K) \times B(\tilde{y}, r)) \cap R(\gamma)$  if  $x \in \text{Leb}(A)$ , for some Borel set  $A$ .

**Case 4:**  $c(x, y) = c(x, \tilde{y}) = \lambda > 0$  and

$$n_{c(\cdot, y)}(x) = n_{c(\cdot, \tilde{y})}(x).$$

If this case occurs, by the assumption (iv) (the “twist kind” condition) we know that  $y = \tilde{y}$  against our assumption.  $\square$

We are now ready to state and prove the main result of this section, which follows directly by Theorem 2.17.

**Theorem 2.19.** *Let  $\mu, \nu$  be two Borel probability measures on  $\mathbb{R}^d$  with compact supports and  $\mu \ll \mathcal{L}^d$  and let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$  be a continuous function satisfying the assumptions (i)-(iv) of Theorem 2.17. If  $\gamma \in \Pi(\mu, \nu)$  is  $\infty$ -monotone, then  $\gamma$  vanishes outside the graph of a Borel map  $T \in \mathcal{T}(\mu, \nu)$ .*

**Proof.** By Lemma 2.11 we know that  $\gamma$  is concentrated on the set of  $\gamma$ -regular points  $R(\gamma)$  and that the set  $R(\gamma)$  is a Borel set. Therefore, it suffices to show that the set  $R(\gamma)$  is the graph of a function  $T \in \mathcal{T}(\mu, \nu)$ . More precisely it is enough to prove that the set  $R(\gamma)$  is contained in the graph of some function  $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . Indeed, Theorem 2.3 of [1] ensures that  $T$  is a Borel map and that  $\gamma = (Id \times T)_\# \mu$ . Let us assume, by contradiction, that there exist  $(x, y), (x, \tilde{y}) \in R(\gamma)$  with  $y \neq \tilde{y}$ . Then we can apply Theorem 2.17 thanks to which there exists  $(x', y') \in R(\gamma)$  such that

$$\max\{c(x', y), c(x, y')\} < \max\{c(x, y), c(x', y')\},$$

contradicting the  $\infty$ -monotonicity of  $\gamma$ . In fact, the above inequality implies that  $R(\gamma)$  is not  $\infty$ -monotone and, since  $R(\gamma) \subset \text{supp } \gamma$ , neither is  $\text{supp } \gamma$ . By the continuity of the cost function  $c$  a transport plan  $\gamma$  is  $\infty$ - $c$ -cyclically monotone ( $\infty$ -monotone) if and only if  $\text{supp } \gamma$  is  $\infty$ - $c$ -cyclically monotone ( $\infty$ -monotone) (this is immediate but the reader may also see Lemma 2.11 in [9]).  $\square$

### 3. On the uniqueness of ICM optimal transport plans

The next lemma is slightly less general than its equivalent in [9] but we believe it makes more transparent the proof of uniqueness that will follow.

**Lemma 3.1.** *Let  $\mu$  and  $\nu$  be two probability measures on  $\mathbb{R}^d$ , with compact supports and  $\mu \ll \mathcal{L}^d$ . Assume that the function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$  satisfies the assumptions of Theorem 2.17. Let  $T, \tilde{T} \in \mathcal{T}(\mu, \nu)$  be two transport maps associated to two  $\infty$ -monotone transport plans  $\gamma$  and  $\tilde{\gamma}$ , i.e.,  $\gamma = (Id \times T)_\# \mu$  and  $\tilde{\gamma} = (Id \times \tilde{T})_\# \mu$ . Let  $A \subset \mathbb{R}^d$  be a Borel set such that for all  $x \in A$ , one has  $T(x) \neq \tilde{T}(x)$ . Define*

$$\mathcal{R}(A) = \{z \in \mathbb{R}^d : \exists x \in A \text{ s.t.}$$

$$\max\{c(x, T(z)), c(z, \tilde{T}(x))\} < \max\{c(x, \tilde{T}(x)), c(z, T(z))\}.$$

Then  $\mu(A \setminus \mathcal{R}(A)) = 0$ .

**Remark 3.2.** The definition of  $\mathcal{R}(A)$  is not symmetric with respect to the interchange of roles of  $T$  and  $\tilde{T}$ , in the sense that

$$\tilde{\mathcal{R}}(A) = \{z \in \mathbb{R}^d : \exists x \in A \text{ s.t.}$$

$$\max\{c(x, \tilde{T}(z)), c(z, T(x))\} < \max\{c(x, T(x)), c(z, \tilde{T}(z))\},$$

is not necessarily equal to  $\mathcal{R}(A)$ . However, the same result can also be proven for  $\tilde{\mathcal{R}}(A)$ .

**Proof.** First of all we observe that

$$\begin{aligned} \mu(A \setminus \mathcal{R}(A)) &= \mu(\pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})) \cap A \setminus \mathcal{R}(A)) \\ &= \mu(\pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})) \cap \text{Leb}(A) \setminus \mathcal{R}(A)), \end{aligned}$$

since  $\gamma$  and  $\tilde{\gamma}$  are respectively concentrated on  $R(\gamma)$  and  $R(\tilde{\gamma})$  and  $\mu \ll \mathcal{L}^d$ . We will prove that

$$\pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})) \cap \text{Leb}(A) \setminus \mathcal{R}(A) = \emptyset.$$

Let us assume by contradiction that there exists  $x \in \pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})) \cap \text{Leb}(A) \setminus \mathcal{R}(A)$  and consider  $(x, T(x)) \in R(\gamma)$  and  $(x, \tilde{T}(x)) \in R(\tilde{\gamma})$ . By definition of  $A$ , we have  $T(x) \neq \tilde{T}(x)$ . By the final sentence of Theorem 2.17 applied to  $R(\tilde{\gamma})$ , we can find  $(x', y') = (x', \tilde{T}(x')) \in R(\tilde{\gamma})$  such that  $x' \in A$  and

$$\max\{c(x, \tilde{T}(x')), c(x', T(x))\} < \max\{c(x, T(x)), c(x', \tilde{T}(x'))\}. \tag{3.1}$$

We conclude noticing that the inequality (3.1) implies that  $x \in \mathcal{R}(A)$ , contradicting the fact that  $x \in A \setminus \mathcal{R}(A)$ .  $\square$

**Theorem 3.3.** *Let  $\mu$  and  $\nu$  be two probability measures on  $\mathbb{R}^d$ , with compact support and  $\mu \ll \mathcal{L}^d$ . Assume that the function  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$  satisfies the assumptions (i)-(iv) of Theorem 2.17 and that  $\nu(\{y_0\}) > 0$  for some  $y_0 \in \mathbb{R}^d$ . Let  $T$  and  $\tilde{T}$  be optimal transport maps corresponding respectively to  $\infty - c$ -cyclically monotone transport plans  $\gamma$  and  $\tilde{\gamma}$ . Then*

$$\mu(T^{-1}(\{y_0\}) \setminus \tilde{T}^{-1}(\{y_0\})) = 0.$$

**Proof.** Assume, by contradiction, that

$$\mu(T^{-1}(\{y_0\}) \setminus \tilde{T}^{-1}(\{y_0\})) > 0.$$

We first restrict to the set of the full  $\mu$ -measure where both the graphs of  $T$  and  $\tilde{T}$  are  $\infty - c$ -cyclically monotone (by Theorem 2.19 we know that we may consider the set  $\pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma}))$ ). We may also assume, restricting our attention to the set of interest, that

$$T(x) \neq \tilde{T}(x) \quad \text{for all } x \in \pi^1(R(\gamma)) \cap \pi^1(R(\tilde{\gamma})). \tag{3.2}$$

In this proof we will apply Lemma 3.1 repeatedly to different subsets  $A$  of  $\text{supp } \mu$ , this is always possible by Assumption (3.2).

We define  $B_0 = \{y_0\}$ ,  $A_1 = T^{-1}(B_0)$ ,  $A'_1 = \mathcal{R}(A_1)$ , and  $B_1 = \tilde{T}(A'_1)$ . We continue recursively: assuming that for all  $j \in \{1, \dots, k-1\}$  the sets  $A_j, A'_j$ , and  $B_j$  have already been defined we set

$$\begin{aligned} A_k &= T^{-1}(B_{k-1}), \quad A'_k = \mathcal{R}(A_k), \quad \text{and} \\ B_k &= \tilde{T}(A'_k). \end{aligned}$$

We observe that by construction  $\mu(A_j) > 0$  and, by Lemma 3.1, also  $\mu(A'_j) > 0$  for every  $j$ . We continue by defining

$$P_1 = \bigcup_{k \geq 1} A_k \quad \text{and} \quad P_2 = \bigcup_{k \geq 1} B_k.$$

We prove the following intermediate claim:

**Claim**  $y_0 \in P_2$ .

*Proof.* Let us first show that

$$\mu(P_1) = \nu(P_2). \quad (3.3)$$

We prove that  $\nu(P_2) \leq \mu(P_1)$ . Right from the definitions of the sets  $P_1$ ,  $P_2$ ,  $A_k$ , and  $B_k$  we see that

$$T^{-1}(B_0 \cup P_2) = P_1.$$

Since  $\nu = T_{\#}\mu$ , we get

$$\nu(P_2) \leq \nu(B_0 \cup P_2) = \mu(T^{-1}(B_0 \cup P_2)) = \mu(P_1). \quad (3.4)$$

In order to prove the opposite inequality we first observe that, by using Lemma 3.1 with  $A = A_k$ , we have

$$\mu(A_k \setminus \mathcal{R}(A_k)) = 0 \quad \text{for all } k \geq 1; \quad (3.5)$$

and we recall that in this construction  $A'_k = \mathcal{R}(A_k)$ . Now

$$\begin{aligned} \mu(P_1) &= \mu\left(\bigcup_{k \geq 1} A_k\right) \stackrel{a)}{\leq} \mu\left(\bigcup_{k \geq 1} A'_k\right) \\ &\leq \mu\left(\bigcup_{k \geq 1} \tilde{T}^{-1}(\tilde{T}(A'_k))\right) = \mu\left(\tilde{T}^{-1}\left(\bigcup_{k \geq 1} \tilde{T}(A'_k)\right)\right) \\ &= \nu\left(\bigcup_{k \geq 1} \tilde{T}(A'_k)\right) = \nu\left(\bigcup_{k \geq 1} B_k\right) = \nu(P_2) \end{aligned}$$

where inequality a) is due to Condition (3.5). So we have  $\mu(P_1) \leq \nu(P_2)$ , which completes the proof of Condition (3.3).

If by contradiction  $y_0 \notin P_2$ , since  $\nu(\{y_0\}) > 0$  (remember that in our construction  $\{y_0\} = B_0$ ), then

$$\nu(P_2) < \nu(B_0 \cup P_2) = \mu(P_1), \quad (3.6)$$

where in the last equality we have used Condition (3.4). Inequality (3.6) contradicts (3.3) and the **Claim** is proven.

By the inclusion  $y_0 \in P_2$  we now know that  $y_0$  belongs to  $B_k$  for some  $k \geq 1$ . Therefore there exist  $x'_k \in A'_k$  (since  $B_k = \tilde{T}(A'_k)$ ) such that  $y_0 = \tilde{T}(x'_k)$ . Next we choose  $x_k \in A_k$  such that

$$\max\{c(x_k, \tilde{T}(x'_k)), c(x'_k, T(x_k))\} < \max\{c(x_k, T(x_k)), c(x'_k, \tilde{T}(x'_k))\}.$$

This is possible since  $A'_k = \mathcal{R}(A_k)$ .

Since by construction  $A_k = T^{-1}(B_{k-1})$  and  $x_k \in A_k$ , we have that  $T(x_k) \in B_{k-1} = \tilde{T}(A'_{k-1})$ , so there exists  $x'_{k-1} \in A'_{k-1}$  such that  $\tilde{T}(x'_{k-1}) \in B_{k-1}$  and  $T(x_k) = \tilde{T}(x'_{k-1})$ . And so we continue, decreasing with the indices down until having defined  $x'_1$  and  $x_1$ . Since  $x_1 \in A_1 = T^{-1}(B_0)$  we have that  $T(x_1) = y_0 = \tilde{T}(x'_1)$ . Thus, we have constructed two  $k$ -uples of points  $(x_j)_{j=1}^k$  and  $(x'_j)_{j=1}^k$  such that for all  $j$ ,  $x_j \in A_j$  and  $x'_j \in A'_j$ .

$$\begin{cases} T(x_j) = \tilde{T}(x'_{j-1}) & \text{for all } 2 \leq j \leq k; \\ T(x_1) = \tilde{T}(x'_k) = y_0, \end{cases}$$

and that for all  $1 \leq j \leq k$

$$\max\{c(x'_j, T(x_j)), c(x_j, \tilde{T}(x'_j))\} < \max\{c(x'_j, \tilde{T}(x'_j)), c(x_j, T(x_j))\}. \tag{3.7}$$

We observe that by Condition (3.7),  $x_j \neq x'_j$ . We now apply to the set  $\{x_j\}_{j=1}^k$  the cyclical permutation

$$\sigma(j) = \begin{cases} j + 1 & \text{if } j \in \{1, \dots, k - 1\} \\ 1 & \text{if } j = k \end{cases}.$$

We have

$$\begin{aligned} \max_{1 \leq j \leq k} c(x_j, T(x_j)) &\stackrel{a)}{\leq} \max_{1 \leq j \leq k} c(x_j, T(x_{\sigma(j)})) \stackrel{b)}{=} \max_{1 \leq j \leq k} c(x_j, \tilde{T}(x'_j)) \\ &\stackrel{c)}{<} \max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\} \end{aligned} \tag{3.8}$$

where inequality a) follows from the ICM-condition assumed on  $T$ . Equality b) follows from the fact that by construction  $T(x_{\sigma(j)}) = \tilde{T}(x'_j)$  for all  $j \in \{1, \dots, k\}$ . Estimate c) follows from Condition (3.7). Concerning the last term in (3.8), we have two possibilities: either

$$\max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\} = c(x_m, T(x_m)) \quad \text{for some } m$$

or

$$\max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\} = \max_{1 \leq j \leq k} c(x'_j, \tilde{T}(x'_j)).$$

The first case leads immediately to a contradiction. So we are left with the latter case. We apply to the set  $\{x'_j\}_{j=1}^k$  the cyclical permutation

$$\tau(j) = \begin{cases} j - 1 & \text{if } j \in \{2, \dots, k\} \\ k & \text{if } j = 1 \end{cases}.$$

We get

$$\begin{aligned} \max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\} &= \max_{1 \leq j \leq k} c(x'_j, \tilde{T}(x'_j)) \stackrel{a)}{\leq} \max_{1 \leq j \leq k} c(x'_j, \tilde{T}(x'_{\tau(j)})) \\ &\stackrel{b)}{=} \max_{1 \leq j \leq k} c(x'_j, T(x_j)) \stackrel{c)}{<} \max_{1 \leq j \leq k} \max\{c(x_j, T(x_j)), c(x'_j, \tilde{T}(x'_j))\}. \end{aligned} \tag{3.9}$$

Above, in a) we have used the ICM property of  $\tilde{T}$ , in b) the fact that, by construction,  $T(x_j) = \tilde{T}(x'_{\tau(j)})$  for all  $j \in \{1, \dots, k\}$ , and in c) again Condition (3.7). Estimate (3.9) is impossible, so we get a contradiction that concludes the proof.  $\square$

#### 4. Some examples

In this section we give some examples of cost functions that satisfy the assumptions of Theorem 2.17. We also present a counterexample which shows that the assumption of the strict quasiconvexity of  $c(\cdot, y)$  is necessary.

**Example 4.1.** Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$  be a cost function of the type introduced in [9], defined by

$$c(x, y) := h(y - x),$$

where  $h : \mathbb{R}^d \rightarrow \mathbb{R}^+$  is a continuous and strictly quasiconvex function. Here, we also assume that there exists a unique  $p \in \mathbb{R}^d$  such that  $h(p) = 0$ , that  $h$  is differentiable, and that  $\nabla h(v) \neq 0$  for every  $v \neq p$ . This cost function satisfies the assumptions of Theorem 2.17 from which we infer that the IM transportation plans corresponding to  $c$  are of the Monge type. The fact that  $c$  fulfills the condition (iv) can be seen as a special case of the proof of the example below.

**Example 4.2.** Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$  be defined by

$$c(x, y) := h(G(y) - F(x)),$$

where  $h : \mathbb{R}^d \rightarrow \mathbb{R}_+$  is defined as in the Example 4.1,  $F, G : \mathbb{R}^d \rightarrow \mathbb{R}^d$  affines transformations of the form

$$F(x) = Ax + b \quad \text{and} \quad G(y) = A'y + b',$$

where  $A, A' \in M^{d \times d}$  are invertible matrices and  $b, b' \in \mathbb{R}^d$ .

We observe that the cost function  $c$  satisfies the assumptions of Theorem 2.19. By the assumptions on  $h$  and the invertibility of  $F$  and  $G$ , assumption (i) is satisfied. The strict quasiconvexity of  $h$  and the linearity of  $F$  and  $G$  ensure the strict quasiconvexity of  $c(\cdot, y)$  for every  $y$ , and, for the same reason,  $c(x, \cdot)$  is strictly quasiconvex for every  $x$ .

Let us prove that the “twist kind” condition holds. Let  $y, \tilde{y} \in \mathbb{R}^d$  and  $\lambda > 0$  be such that  $c(x, y) = c(x, \tilde{y}) = \lambda$  and  $n_{c(\cdot, y)}(x) = n_{c(\cdot, \tilde{y})}(x)$ . We first observe that, by the fact that

$$\nabla_x c(x, y) = -\nabla F(x)^T \cdot \nabla h(G(y) - F(x)) = -A^T \cdot \nabla h(G(y) - F(x)),$$

$\nabla_x c(x, y) \neq 0$  for every  $x, y$  such that  $G(y) - F(x) \neq p$  (that is, for every  $x, y$  such that  $c(x, y) \neq 0$ ). Thus, the equality between the two unit normals becomes

$$\frac{\nabla_x c(x, y)}{|\nabla_x c(x, y)|} = \frac{\nabla_x c(x, \tilde{y})}{|\nabla_x c(x, \tilde{y})|}.$$

Since

$$\nabla_y c(x, y) = A'^T \cdot \nabla h(G(y) - F(x)),$$

thanks to the invertibility of  $A'$  we have that

$$\nabla_x c(x, y) = -A^T (A'^T)^{-1} \nabla_y c(x, y).$$

The computation above, the invertibility assumptions on  $A$  and  $A'$  and the condition on the normalized gradients imply



$$\frac{\nabla_y c(x, y)}{|\nabla_y c(x, y)|} = \frac{\nabla_y c(x, \tilde{y})}{|\nabla_y c(x, \tilde{y})|}.$$

We observe that both  $y$  and  $\tilde{y}$  belong to the boundary of the sublevel set with respect to  $x$ ,  $C_\lambda^x := \{z \in \mathbb{R}^d : c(x, z) \leq \lambda\}$ . Then  $y$  and  $\tilde{y}$  must coincide, since they are two points on the smooth boundary of a strictly convex set  $C_\lambda^x$  with the same normal.

The following is an example of an ICM plan that is not given by a map for a cost function that doesn't satisfy the assumptions of Theorem 2.17.

**Example 4.3.** We consider the space  $X = \mathbb{R}^2$  with the cost function given by the  $\infty$ -distance

$$d((x_1, y_1), (x_2, y_2)) = \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$

We denote by  $Q := [0, 1] \times [0, 1]$ , the unit square centered at  $(\frac{1}{2}, \frac{1}{2})$ . We take  $\mu = \mathcal{L}^2|_Q$ , the 2-dimensional Lebesgue measure restricted to  $Q$ , and  $\nu = \mathcal{L}^2|_{Q+(10,0)}$ , the 2-dimensional Lebesgue measure restricted to the translation of  $Q$  by the vector  $(10, 0)$ . Let  $\gamma \in \Pi(\mu, \nu)$  be defined by

$$\gamma = \mu \otimes \mathcal{H}^1|_{\{x_1+10\} \times [0,1]},$$

that is, the transport plan which distributes evenly every point  $(x_1, x_2) \in \text{supp } \mu$  to the vertical line segment  $\{(x_1 + 10, y_2) \mid y_2 \in \mathbb{R}\} \cap (Q + (10, 0))$ . This plan is clearly not given by a map, but it is ICM.

To prove that  $\gamma$  is ICM we fix points  $\{(x_1^i, x_2^i), (x_1^i + 10, y_2^i)\}_{i=1}^N$  in the support of  $\gamma$  and a permutation  $\sigma$  of the set  $\{1, \dots, N\}$ . Now

$$\max_{1 \leq i \leq N} \{d((x_1^i, x_2^i), ((x_1^i + 10, y_2^i)))\} = \max_{1 \leq i \leq N} \{\max\{10, |x_2^i - y_2^i|\}\} = 10$$

where in the last equality we have used the fact that

$$|x_2^i - y_2^i| \leq 1 < 10.$$

Let us prove that

$$\max_{1 \leq i \leq N} \{d((x_1^i, x_2^i), (x_1^{\sigma(i)} + 10, y_2^{\sigma(i)}))\} \geq 10.$$

It is enough to show that there exists an index  $k \in \{1, \dots, N\}$  such that

$$|x_1^k - (x_1^{\sigma(k)} + 10)| \geq 10.$$

We may assume that  $\sigma$  is not the identity: for the identity permutation the ICM condition holds as an equality. Let us denote  $I = \{1, \dots, N\}$  and let  $k_0$  be such that

$$x_1^{k_0} = \min\{x_1^k \mid k \in I\}.$$

The point  $x_1^{k_0}$  satisfies  $x_1^{k_0} \leq x_1^{\sigma(k_0)}$ . Therefore,

$$|x_1^{k_0} - (x_1^{\sigma(k_0)} + 10)| = 10 + (x_1^{\sigma(k_0)} - x_1^{k_0}) \geq 10$$

and we are done.

## Acknowledgments

The research of the first two authors is partially financed by the “*Fondi di ricerca di ateneo, ex 60%*” of the University of Firenze and is part of the project “*Alcuni problemi di trasporto ottimo ed applicazioni*” of the GNAMPA-INDAM.

The research of the third author is part of the project *Contemporary topics on multi-marginal optimal mass transportation*, funded by the Finnish Postdoctoral Pool (Suomen Kulttuurisäätiö), and of the project *Giovani ricercatori protagonisti 2020*, funded by Fondazione Cassa di Risparmio di Firenze.

The authors wish to thank the anonymous referee for the help in finding early flaws and weakness in the paper and allowing for corrections.

## References

- [1] L. Ambrosio, Lecture notes on optimal transportation, in: *Mathematical Aspects of Evolving Interfaces*, Funchal, 2000, in: *Lecture Notes in Mathematics*, Springer, New York, 2003.
- [2] E.N. Barron, M. Bocea, R.R. Jensen, Duality for the  $L^\infty$  optimal transport problem, *Trans. Am. Math. Soc.* 369 (2017) 3289–3323.
- [3] G. Bouchitté, C. Jimenez, M. Rajesh, A new  $L^\infty$  estimate in optimal mass transport, *Proc. Am. Math. Soc.* 135 (2007) 3525–3535.
- [4] T. Champion, L. De Pascale, On the twist condition and  $c$ -monotone transport plans, *Discrete Contin. Dyn. Syst.* 34 (4) (2014) 1339–1353.
- [5] T. Champion, L. De Pascale, P. Juutinen, The  $\infty$ -Wasserstein distance: local solutions and existence of optimal transport maps, *SIAM J. Math. Anal.* 40 (1) (2008) 1–20.
- [6] L. De Pascale, J. Louet, A study of the dual problem of the one-dimensional  $L^\infty$  optimal transport problem with applications, preprint, 2017.
- [7] A. Fathi, A. Figalli, Optimal transportation on non-compact manifolds, *Isr. J. Math.* 175 (2010) 1–59.
- [8] C. Jimenez, F. Santambrogio, Optimal transportation for a quadratic cost with convex constraints and applications, *J. Math. Pures Appl.* 98 (2012) 103–113.
- [9] H. Jylhä, The  $L^\infty$  optimal transport: infinite cyclical monotonicity and the existence of optimal transport maps, *Calc. Var. Partial Differ. Equ.* 52 (2015) 303–326.
- [10] H. Jylhä, T. Rajala,  $L^\infty$  estimates in optimal mass transportation, *J. Funct. Anal.* 270 (2016) 4297–4321.
- [11] H. Kellerer, Duality theorems for marginal problems, *Z. Wahrscheinlichkeitstheor. Verw. Geb.* 67 (1984) 399–432.
- [12] C. Villani, *Optimal Transport. Old and New*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, Berlin, 2009.