

# OBSTRUCTIONS TO EXTENSION OF WASSERSTEIN DISTANCES FOR VARIABLE MASSES

LUCA LOMBARDINI AND FRANCESCO ROSSI

**ABSTRACT.** We study the possibility of defining a distance on the whole space of measures, with the property that the distance between two measures having the same mass is the Wasserstein distance, up to a scaling factor. We prove that, under very weak and natural conditions, if the base space is unbounded, then the scaling factor must be constant, independently of the mass. Moreover, no such distance can exist, if we include the zero measure. Instead, we provide examples with non-constant scaling factors for the case of bounded base spaces.

## 1. INTRODUCTION AND MAIN RESULTS

Wasserstein distances are crucial modern tools in mathematics, raising an enormous interest from the community of mathematical analysis (see e.g. [1, 3, 17]). Their applications are also extremely varied, ranging from crowd dynamics [6, 13], to economics [9, 12], to computer science [14].

Yet, Wasserstein distances have several limitations, the most apparent being that they are defined between measures with the same mass only. This issue led to the definition of different possible generalizations of the Wasserstein distance between measures with different masses. The first attempt in this direction, described in [8], was related to the heat equation on a domain  $\Omega$  with Dirichlet boundary condition, that clearly does not preserve mass. A second attempt, see [7], was given by the so-called optimal partial transport problem: given two measures  $\mu, \nu$  with different masses, one fixes a smaller mass and computes the Wasserstein distance between optimal submeasures.

A third group of contributions, known under the name of *generalized Wasserstein distance* [15, 16], was based on the optimization of a mixed Wasserstein- $L^1$  cost, in which a part of the measure is transported (with Wasserstein cost) and the remaining part is removed (with a  $L^1$  cost). This contribution was then followed by several “generalizations of the generalization”, see [5, 10, 11].

One of the main drawbacks of the distances given above is that they do not coincide, in general, with the standard Wasserstein distance when the two measures have the same mass. The only exception is clearly the partial optimal transportation problem when one chooses to transport the whole mass. This observation raises a natural question: *is it possible to define a distance between measures of different masses, that coincides with the Wasserstein distance in the case measures have the same mass?*

Our main result provides a negative answer, under quite general hypotheses. We show that it is not possible to extend the  $p$ -Wasserstein distance  $W_p$  to a distance function  $d$  defined on the whole space of measures in  $\mathbb{R}^n$  and satisfying the property

$$(1.1) \quad d(\mu_1, \mu_2) = f(m)W_p\left(\frac{\mu_1}{m}, \frac{\mu_2}{m}\right),$$

whenever the mass  $|\mu| := \int_{\mathbb{R}^n} d\mu$  satisfies  $|\mu_1| = |\mu_2| = m > 0$ .

More precisely, let  $\mathcal{M}_c(\mathbb{R}^n)$  be the space of non-negative measures in  $\mathbb{R}^n$  having finite mass and compact support and  $\mathcal{M}_c^*(\mathbb{R}^n)$  the space of measures with positive finite mass and compact support. Then,  $\mathcal{M}_c(\mathbb{R}^n) = \{0\} \cup \mathcal{M}_c^*(\mathbb{R}^n)$ , where 0 denotes the zero measure. Let also  $\mathcal{P}_c(\mathbb{R}^n)$  be the space of probability measures on  $\mathbb{R}^n$ , i.e. of mass 1, having compact support. We use the notation  $\mathbb{R}_{>0}$  for the space of positive real numbers  $(0, +\infty)$ , endowed with the Euclidean distance. We stress that, as sets, we can identify

$$(1.2) \quad \mathcal{M}_c^*(\mathbb{R}^n) \simeq \mathbb{R}_{>0} \times \mathcal{P}_c(\mathbb{R}^n),$$

---

2020 *Mathematics Subject Classification.* 28A33, 49Q22.

*Key words and phrases.* Metrization of weak convergence, convergence of measures, unbalanced optimal transportation.

via the bijection  $\mu \mapsto (|\mu|, \mu/|\mu|)$ .

We investigate the possibility of defining a distance function  $d$  on  $\mathcal{M}_c(\mathbb{R}^n)$  satisfying (1.1), i.e. with the property that on each fiber  $\{m\} \times \mathcal{P}_c(\mathbb{R}^n)$  the distance  $d$  coincides with the  $p$ -Wasserstein distance, up to a multiplicative scaling factor depending on  $m > 0$  only.

Our main result is a strong obstruction to a general definition of such a distance, as it provides a very weak condition forcing the function  $f$  in (1.1) to be constant. Indeed, we have the following first result.

**Theorem 1.1** (Unbounded space). *Let  $p \in [1, +\infty)$  and let  $d_{\mathcal{M}} : \mathcal{M}_c^*(\mathbb{R}^n) \times \mathcal{M}_c^*(\mathbb{R}^n) \rightarrow [0, +\infty)$  be a distance function satisfying the following two properties:*

(i) *there exists a function  $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$  such that*

$$d_{\mathcal{M}}(m\mu_1, m\mu_2) = f(m)W_p(\mu_1, \mu_2),$$

*for every  $m \in \mathbb{R}_{>0}$  and every  $\mu_1, \mu_2 \in \mathcal{P}_c(\mathbb{R}^n)$ .*

(ii) *There exists an unbounded subset  $\Sigma \subseteq \mathbb{R}^n$  such that for every  $m_0 \in \mathbb{R}_{>0}$  there exist a radius  $r > 0$  and a positive constant  $C > 0$ , both depending only on  $m_0$ , with*

$$(1.3) \quad \sup_{x \in \Sigma} d_{\mathcal{M}}(m_0\delta_x, m\delta_x) \leq C$$

*for every  $m \in (m_0 - r, m_0 + r)$ , where  $\delta_x$  is the Dirac delta centered at  $x$ .*

*Then,  $f \equiv \lambda$ , for some positive constant  $\lambda > 0$ .*

As a consequence, it is not possible to extend such a distance to the whole space  $\mathcal{M}_c(\mathbb{R}^n)$ , i.e. when the zero measure is added.

**Corollary 1.2** (Unbounded space with zero). *Let  $d_{\mathcal{M}}$  be a distance function on  $\mathcal{M}_c^*(\mathbb{R}^n)$  as in Theorem 1.1. Then, there exists no distance function  $d_{\mathcal{M}_0}$  defined on  $\mathcal{M}_c(\mathbb{R}^n)$ , which agrees with  $d_{\mathcal{M}}$  on  $\mathcal{M}_c^*(\mathbb{R}^n)$ , and such that either*

$$(1.4) \quad \sup_{x \in \Sigma} d_{\mathcal{M}_0}(0, m_0\delta_x) < +\infty,$$

*for some  $m_0 \in \mathbb{R}_{>0}$ , or*

$$(1.5) \quad \lim_{m \searrow 0} d_{\mathcal{M}_0}(0, m\mu_i) = 0,$$

*at least for two different  $\mu_1 \neq \mu_2 \in \mathcal{P}_c(\mathbb{R}^n)$ .*

A few observations are in order. First of all, simple examples of distance functions on  $\mathcal{M}_c^*(\mathbb{R}^n)$  satisfying (1.1), with  $f$  constant, are given by the product metrics on the decomposed space (1.2), see Example 3.3 below. Roughly speaking, these distances measure separately the costs of changing mass and of transporting two equal masses, then sum up the two costs.

We also stress that condition (1.3) is very weak. It is an assumption regarding only the point masses located within  $\Sigma$ , and it represents the fact that the cost of “destroying” or “creating” mass, going from  $m\delta_x$  to  $m_0\delta_x$ , is uniformly bounded with respect to the position  $x \in \Sigma$ , locally around  $m_0$ . Furthermore, we point out that (1.3) is implied by other very natural conditions, like the invariance of  $d_{\mathcal{M}}$  with respect to isometries together with the compatibility of  $d_{\mathcal{M}}$  with weak convergence when varying the mass, see Proposition 3.2 for the precise statement.

In Theorem 1.1 and Corollary 1.2, the specific properties of the  $p$ -Wasserstein distance actually play no role, except for the fact that  $W_p$  is unbounded on  $\mathcal{M}_c^*(\mathbb{R}^n)$ —and in particular on the subset consisting of Dirac measures with base points belonging to  $\Sigma$ , an unbounded subset of  $\mathbb{R}^n$ . As a consequence, Theorem 1.1 and Corollary 1.2 can be extended to much more general settings: in place of  $\mathbb{R}^n$  we can consider a Polish space  $(X, d_X)$  and in place of the Wasserstein distance  $W_p$  we can consider any distance function  $d_{\mathcal{P}}$  defined on the space of compactly supported probability measures  $\mathcal{P}_c(X)$ .

If  $d_{\mathcal{P}}$  is unbounded, we can then translate Theorem 1.1 to this setting in a straightforward way: in point (ii) simply consider a subset  $S \subseteq \mathcal{P}_c(X)$  which is unbounded with respect to  $d_{\mathcal{P}}$ , in place of the set  $\{\delta_x : x \in \Sigma\}$ .

On the contrary, if  $d_{\mathcal{P}}$  is bounded—e.g., if  $d_{\mathcal{P}}$  is the bounded Lipschitz distance in  $\mathcal{P}_c(\mathbb{R}^n)$  (see, e.g., [16,18]) or if  $d_{\mathcal{P}}$  is the  $p$ -Wasserstein distance in  $\mathcal{P}_c(X)$ , when  $(X, d_X)$  is bounded—the picture is completely different.

**Proposition 1.3** (Bounded space). *Let  $(X, d_X)$  be a Polish space and let  $d_{\mathcal{P}} : \mathcal{P}_c(X) \times \mathcal{P}_c(X) \rightarrow [0, +\infty)$  be a bounded distance function. Let  $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$  be a Lipschitz, increasing, function such that*

$$\sup_{\substack{m_1, m_2 \in \mathbb{R}_{>0} \\ m_1 \neq m_2}} \frac{|f(m_1) - f(m_2)|}{|m_1 - m_2|} \leq \frac{1}{\text{diam } \mathcal{P}_c(X)} \quad \text{and} \quad \lim_{m \searrow 0} f(m) = 0.$$

*Then, the function  $d_{\mathcal{M}_0} : \mathcal{M}_c(X) \times \mathcal{M}_c(X) \rightarrow [0, +\infty)$  defined by setting*

$$d_{\mathcal{M}_0}(m_1\mu_1, m_2\mu_2) := |m_1 - m_2| + \min\{f(m_1), f(m_2)\}d_{\mathcal{P}}(\mu_1, \mu_2),$$

*for every  $m_1, m_2 \in \mathbb{R}_{>0}$  and  $\mu_1, \mu_2 \in \mathcal{P}_c(X)$ , and*

$$d_{\mathcal{M}_0}(0, m\mu) = d_{\mathcal{M}_0}(m\mu, 0) := m, \quad d_{\mathcal{M}_0}(0, 0) := 0,$$

*is a distance function on  $\mathcal{M}_c(X)$ .*

Clearly, the distance function  $d_{\mathcal{M}_0}$  defined in Proposition 1.3 is such that

$$d_{\mathcal{M}_0}(m\mu_1, m\mu_2) = f(m)d_{\mathcal{P}}(\mu_1, \mu_2),$$

for every  $m \in \mathbb{R}_{>0}$  and every  $\mu_1, \mu_2 \in \mathcal{P}_c(X)$ . Moreover,

$$\sup_{\mu \in \mathcal{P}_c(X)} d_{\mathcal{M}_0}(m_1\mu, m_2\mu) = |m_1 - m_2|,$$

for every  $m_1, m_2 \in [0, +\infty)$ , which is a stronger property than assumption (ii) of Theorem 1.1, and also implies (1.5) for every  $\mu \in \mathcal{P}_c(\mathbb{R}^n)$ .

The rest of the paper is organized as follows. In Section 2, we first prove some general results about distance functions defined on the direct product of two metric spaces. Then, in Section 2.1, we exploit these results to prove our main theorems, stated here above in Section 1. In Section 3 we prove some further results, which are natural consequences of Theorem 1.1, and we provide some examples of meaningful distance functions defined on  $\mathcal{M}_c^*(\mathbb{R}^n)$ .

## 2. METRICS ON PRODUCT SPACES

We begin with the following result, which is a generalization of Theorem 1.1 to product spaces  $M = X \times Y$ .

**Theorem 2.1.** *Let  $(X, d_X)$  and  $(Y, d_Y)$  be two metric spaces, and let  $M := X \times Y$  be equipped with a distance function  $d_M : M \times M \rightarrow [0, +\infty)$  satisfying the following two properties:*

(i) *there exists a function  $f : X \rightarrow (0, +\infty)$  such that*

$$(2.1) \quad d_M((x, y_1), (x, y_2)) = f(x)d_Y(y_1, y_2),$$

*for every  $x \in X$  and every  $y_1, y_2 \in Y$ .*

(ii) *For every  $x_0 \in X$  there exist a radius  $r > 0$  and a positive constant  $C > 0$ , both depending only on  $x_0$ , such that*

$$(2.2) \quad \sup_{y \in Y} d_M((x_0, y), (x, y)) \leq C$$

*for every  $x \in B_r^X(x_0)$ .*

*If  $X$  is connected and  $\text{diam } Y = +\infty$ , then  $f \equiv \lambda$ , for some positive constant  $\lambda > 0$ .*

*Proof.* Fix  $x_0 \in X$  and let  $r$  and  $C$  be as in point (ii). Notice that by (2.2) and the triangle inequality of  $d_M$  we have

$$\sup_{y \in Y} d_M((x_1, y), (x_2, y)) \leq 2C,$$

for every  $x_1, x_2 \in B_r^X(x_0)$ . Thus, exploiting again the triangle inequality of  $d_M$  and also (2.1), we obtain

$$\begin{aligned} f(x_1)d_Y(y_1, y_2) &= d_M((x_1, y_1), (x_1, y_2)) \\ &\leq d_M((x_1, y_1), (x_2, y_1)) + d_M((x_2, y_1), (x_2, y_2)) + d_M((x_2, y_2), (x_1, y_2)) \\ &\leq 4C + f(x_2)d_Y(y_1, y_2), \end{aligned}$$

for every  $x_1, x_2 \in B_r^X(x_0)$  and every  $y_1, y_2 \in Y$ . This implies that

$$|f(x_1) - f(x_2)|d_Y(y_1, y_2) \leq 4C,$$

and hence

$$\sup_{x_1, x_2 \in B_r^X(x_0)} |f(x_1) - f(x_2)| \leq \frac{4C}{\text{diam } Y}.$$

In particular, if  $\text{diam } Y = +\infty$  then  $f$  must be constant in  $B_r^X(x_0)$ , hence it is locally constant in  $X$ . Therefore, if we also assume  $X$  to be connected we conclude that  $f$  is globally constant, as claimed.  $\square$

An example of such a distance function is given by

$$d_M((x_1, y_1), (x_2, y_2)) := d_X(x_1, x_2) + \lambda d_Y(y_1, y_2),$$

for any fixed  $\lambda > 0$ .

*Remark 2.2.* On the other hand, if  $\text{diam } Y < +\infty$ , then we can easily find distances on  $M$  satisfying points (i) and (ii) of Theorem 2.1, with  $f$  not constant. For example, let  $f : X \rightarrow \mathbb{R}_{>0}$  be such that

$$M := \sup_{x \in X} f(x) < +\infty.$$

Then, the function

$$d_M((x_1, y_1), (x_2, y_2)) := \begin{cases} f(x_1) d_Y(y_1, y_2) & \text{if } x_1 = x_2, \\ M \text{ diam } Y & \text{if } x_1 \neq x_2, \end{cases}$$

is such a distance function.

We now restrict our attention to the case in which  $(X, d_X) = \mathbb{R}_{>0}$ .

**Proposition 2.3.** *Let  $(Y, d_Y)$  be a metric space with  $\text{diam } Y < +\infty$  and let  $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$  be a Lipschitz, increasing, function such that*

$$(2.3) \quad \sup_{\substack{x_1, x_2 \in \mathbb{R}_{>0} \\ x_1 \neq x_2}} \frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq \frac{1}{\text{diam } Y}.$$

Then, the function

$$d_M((x_1, y_1), (x_2, y_2)) := |x_1 - x_2| + \min\{f(x_1), f(x_2)\} d_Y(y_1, y_2),$$

is a distance function on  $M := \mathbb{R}_{>0} \times Y$  satisfying point (i) of Theorem 2.1 and

(ii') for every  $x_0 \in \mathbb{R}_{>0}$  there exists a radius  $r > 0$  and a positive constant  $C > 0$ , both depending only on  $x_0$ , such that

$$\sup_{y \in Y} d_M((x_1, y), (x_2, y)) \leq C|x_1 - x_2|,$$

for every  $x_1, x_2 \in (x_0 - r, x_0 + r)$ —which is more restrictive than point (ii) of Theorem 2.1.

*Proof.* We only need to verify that  $d_M$  satisfies the triangle inequality. By symmetry we can assume without loss of generality that  $x_1 \leq x_2$ , so that

$$d_M((x_1, y_1), (x_2, y_2)) = x_2 - x_1 + f(x_1) d_Y(y_1, y_2).$$

Let us now consider a third point  $(x_3, y_3) \in M$ . If  $x_3 \geq x_1$ , then

$$f(x_1) \leq \min\{f(x_1), f(x_3)\} \quad \text{and} \quad f(x_1) \leq \min\{f(x_2), f(x_3)\},$$

hence, exploiting also the triangle inequality of  $|\cdot|$  and  $d_Y$ , we have

$$\begin{aligned} d_M((x_1, y_1), (x_2, y_2)) &= x_2 - x_1 + f(x_1) d_Y(y_1, y_2) \\ &\leq x_3 - x_1 + |x_2 - x_3| + f(x_1) d_Y(y_1, y_3) + f(x_1) d_Y(y_2, y_3) \\ &\leq x_3 - x_1 + \min\{f(x_1), f(x_3)\} d_Y(y_1, y_3) + |x_2 - x_3| + \min\{f(x_2), f(x_3)\} d_Y(y_2, y_3) \\ &= d_M((x_1, y_1), (x_3, y_3)) + d_M((x_2, y_2), (x_3, y_3)). \end{aligned}$$

Suppose now that  $x_3 < x_1$ , so that

$$d_M((x_1, y_1), (x_3, y_3)) = x_1 - x_3 + f(x_3) d_Y(y_1, y_3),$$

and

$$d_M((x_2, y_2), (x_3, y_3)) = x_2 - x_3 + f(x_3) d_Y(y_2, y_3).$$

Therefore the triangle inequality is equivalent to

$$(2.4) \quad f(x_1)d_Y(y_1, y_2) - f(x_3)(d_Y(y_1, y_3) + d_Y(y_2, y_3)) \leq 2(x_1 - x_3).$$

Notice that by the triangle inequality of  $d_Y$  we have

$$-f(x_3)(d_Y(y_1, y_3) + d_Y(y_2, y_3)) \leq -f(x_3)d_Y(y_1, y_2).$$

Thus, by exploiting also assumption (2.3), we obtain

$$\begin{aligned} f(x_1)d_Y(y_1, y_2) - f(x_3)(d_Y(y_1, y_3) + d_Y(y_2, y_3)) &\leq (f(x_1) - f(x_3))d_Y(y_1, y_2) \\ &\leq \frac{x_1 - x_3}{\text{diam } Y} \text{diam } Y, \end{aligned}$$

establishing (2.4) and concluding the proof of the Proposition.  $\square$

**Corollary 2.4.** *Let  $(Y, d_Y)$  be a metric space,  $M := \mathbb{R}_{>0} \times Y$ , and let  $M_0 := \{0\} \cup M$  be equipped with a distance function  $d_{M_0} : M_0 \times M_0 \rightarrow [0, +\infty)$  such that the restriction of  $d_{M_0}$  to  $M$  satisfies points (i) and (ii) of Theorem 2.1. If there exist at least two distinct points  $p_1 \neq p_2 \in Y$  such that*

$$(2.5) \quad \lim_{x \rightarrow 0} d_{M_0}(0, (x, p_i)) = 0 \quad \text{for } i = 1, 2,$$

then  $\text{diam } Y < +\infty$ .

*Proof.* We argue by contradiction and we assume that  $\text{diam } Y = +\infty$ . Then, by Theorem 2.1 we know that

$$d_{M_0}((x, y_1), (x, y_2)) = \lambda d_Y(y_1, y_2),$$

for every  $x \in \mathbb{R}_{>0}$  and every  $y_1, y_2 \in Y$ , for some fixed constant  $\lambda > 0$ . Let  $x_k \searrow 0$ . Then, by (2.5) we have

$$\begin{aligned} 0 < d_Y(p_1, p_2) &= \frac{1}{\lambda} d_{M_0}((x_k, p_1), (x_k, p_2)) \\ &\leq \frac{1}{\lambda} (d_{M_0}((x_k, p_1), 0) + d_{M_0}(0, (x_k, p_2))) \xrightarrow{k \rightarrow \infty} 0, \end{aligned}$$

giving a contradiction.  $\square$

**Lemma 2.5.** *Let  $(Y, d_Y)$  be a metric space,  $M := \mathbb{R}_{>0} \times Y$ , and let  $M_0 := \{0\} \cup M$  be equipped with a distance function  $d_{M_0} : M_0 \times M_0 \rightarrow [0, +\infty)$ . Suppose that there exist  $x_0 \in \mathbb{R}_{>0}$  and  $\lambda_0 > 0$  such that*

$$d_{M_0}((x_0, y_1), (x_0, y_2)) = \lambda_0 d_Y(y_1, y_2),$$

for every  $y_1, y_2 \in Y$ , and

$$\Lambda := \sup_{y \in Y} d_{M_0}((x_0, y), 0) < +\infty.$$

Then  $\text{diam } Y < +\infty$ .

*Proof.* It is enough to notice that by the triangle inequality we have

$$d_Y(y_1, y_2) = \frac{1}{\lambda_0} d_{M_0}((x_0, y_1), (x_0, y_2)) \leq \frac{1}{\lambda_0} (d_{M_0}((x_0, y_1), 0) + d_{M_0}((x_0, y_2), 0)),$$

for every  $y_1, y_2 \in Y$ . This implies that

$$\text{diam } Y \leq \frac{2\Lambda}{\lambda_0},$$

concluding the proof.  $\square$

*Remark 2.6.* If  $\text{diam } Y < +\infty$ , then we can easily find many such distances  $d_{M_0}$  on  $M_0 = \{0\} \cup \mathbb{R}_{>0} \times Y$ . Indeed, let  $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$  be a Lipschitz, increasing, function such that

$$\sup_{\substack{x_1, x_2 \in \mathbb{R}_{>0} \\ x_1 \neq x_2}} \frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq \frac{1}{\text{diam } Y} \quad \text{and} \quad \lim_{x \searrow 0} f(x) = 0.$$

Define the function  $d_{M_0} : M_0 \times M_0 \rightarrow [0, +\infty)$  by setting

$$d_{M_0}((x_1, y_1), (x_2, y_2)) := |x_1 - x_2| + \min\{f(x_1), f(x_2)\}d_Y(y_1, y_2),$$

for every  $(x_1, y_1), (x_2, y_2) \in \mathbb{R}_{>0} \times Y$  and

$$d_{M_0}(0, (x, y)) = d_{M_0}((x, y), 0) := x, \quad d_{M_0}(0, 0) := 0.$$

Then, by arguing as in Proposition 2.3, we can easily verify that  $d_{M_0}$  is a distance function on  $M_0$ .

**2.1. Proofs of the main results.** We apply the results that we obtained in the previous Section to prove Theorem 1.1, Corollary 1.2 and Proposition 1.3.

*Proof of Theorem 1.1.* We consider the metric space  $(Y, d_Y) := (S, W_p)$ , where  $S \subseteq \mathcal{P}_c(\mathbb{R}^n)$  is the subspace given by

$$S := \{\delta_x : x \in \Sigma\},$$

and the subspace

$$M := \{m\delta_x : x \in \Sigma, m > 0\} \simeq \mathbb{R}_{>0} \times S \subseteq \mathcal{M}_c^*(\mathbb{R}^n).$$

Then,  $(M, d_M)$  satisfies the hypothesis of Theorem 2.1. Since  $W_p(\delta_x, \delta_y) = |x - y|$  and  $\Sigma$  is unbounded, also the space  $(Y, d_Y)$  is not bounded. Thus, by Theorem 2.1 we conclude that  $f \equiv \lambda$ , as claimed.  $\square$

We observe that Corollary 1.2, with hypothesis (1.4), means that we cannot extend a distance function  $d_M$  as in Theorem 1.1 to a distance on the whole space  $\mathcal{M}_c(\mathbb{R}^n)$ , while preserving the boundedness assumption of point (ii) also around the zero measure. Similarly, such an extension is not possible under the very weak continuity hypothesis (1.5)—which, roughly speaking, translates the fact that when we fix any probability measure  $\mu$  and we decrease its mass, considering  $m\mu$ , with  $m \searrow 0$ , we expect to be “approaching” the zero measure. This can be seen, e.g., as a compatibility of  $d_{M_0}$  with the weak convergence  $m\mu \rightarrow 0$  as  $m \searrow 0$ . A related result is given in Proposition 3.1, assuming invariance of the distance with respect to isometries.

*Proof of Corollary 1.2.* The claims of the Corollary follow by Lemma 2.5 and by arguing as in the proof of Corollary 2.4.  $\square$

*Proof of Proposition 1.3.* The proof follows from Proposition 2.3 and Remark 2.6.  $\square$

### 3. FURTHER RESULTS AND EXAMPLES

In this Section, we discuss other obstructions to the extension of the Wasserstein distances to general measures, based on invariance with respect to isometries.

We prove that there exists no distance function on  $\mathcal{M}_c(\mathbb{R}^n)$  which is invariant with respect to isometries and which satisfies (1.1) even on a single fiber  $\{m_0\} \times \mathcal{P}_c(\mathbb{R}^n)$ . More precisely:

**Proposition 3.1** (Unbounded space with zero, isometries invariance). *Let  $p \in [1, +\infty)$ . There exists no distance function  $d$  on  $\mathcal{M}_c(\mathbb{R}^n)$  such that*

$$d(\mu_1, \mu_2) = \lambda_0 W_p\left(\frac{\mu_1}{m_0}, \frac{\mu_2}{m_0}\right),$$

whenever  $|\mu_1| = |\mu_2| = m_0$ , for some  $m_0 \in \mathbb{R}_{>0}$  and  $\lambda_0 > 0$ , and which is invariant with respect to the isometries of  $\mathbb{R}^n$ , i.e.

$$(3.1) \quad d(T\#\mu_1, T\#\mu_2) = d(\mu_1, \mu_2)$$

for every isometry  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and for every  $\mu_1, \mu_2 \in \mathcal{M}_c(\mathbb{R}^n)$ .

*Proof.* Notice that for every  $x \in \mathbb{R}^n$  the function  $\tau_x : \mathbb{R}^n \rightarrow \mathbb{R}^n$  defined by  $\tau_x(y) := y + x$  is an isometry, and  $\delta_x = \tau_x\#\delta_0$ . By (3.1) we thus obtain

$$d(m_0\delta_x, 0) = d(m_0\tau_x\#\delta_0, \tau_x\#0) = d(\tau_x\#(m_0\delta_0), \tau_x\#0) = d(m_0\delta_0, 0),$$

for every  $x \in \mathbb{R}^n$ . The conclusion then follows from Lemma 2.5, by considering  $(Y, d_Y) = (S, W_p)$ , with  $S := \{\delta_x : x \in \mathbb{R}^n\}$ , which is not bounded, and  $d_{M_0} = d$  on  $M_0 := \{0\} \cup \mathbb{R}_{>0} \times S \subseteq \mathcal{M}_c(\mathbb{R}^n)$ .  $\square$

Once again, in Proposition 3.1 the obstruction to the existence of such a distance function  $d$  is given by the combination of the presence of the zero measure with the unboundedness of  $W_p$ . On a related note, we also observe that hypothesis (ii) in Theorem 1.1 is weaker than the combination of other very natural assumptions, like the invariance with respect to isometries and the compatibility with weak convergence when varying the mass. Thus, we obtain the following:

**Proposition 3.2** (Unbounded space, isometries invariance). *Let  $p \in [1, +\infty)$  and let  $d_{\mathcal{M}} : \mathcal{M}_c^*(\mathbb{R}^n) \times \mathcal{M}_c^*(\mathbb{R}^n) \rightarrow [0, +\infty)$  be a distance function such that*

$$d_{\mathcal{M}}(m\mu_1, m\mu_2) = f(m)W_p(\mu_1, \mu_2),$$

*for every  $m \in \mathbb{R}_{>0}$  and every  $\mu_1, \mu_2 \in \mathcal{P}_c(\mathbb{R}^n)$ , for some function  $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ . Suppose that  $d_{\mathcal{M}}$  is invariant with respect to the isometries of  $\mathbb{R}^n$ , i.e.*

$$(3.2) \quad d_{\mathcal{M}}(T\#\mu_1, T\#\mu_2) = d_{\mathcal{M}}(\mu_1, \mu_2)$$

*for every isometry  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and for every  $\mu_1, \mu_2 \in \mathcal{M}_c^*(\mathbb{R}^n)$ . Assume moreover that*

$$(3.3) \quad \lim_{|m-m_0| \rightarrow 0} d_{\mathcal{M}}(m\mu, m_0\mu) = 0,$$

*for every  $m_0 \in \mathbb{R}_{>0}$  and every  $\mu \in \mathcal{P}_c(\mathbb{R}^n)$ . Then,  $f \equiv \lambda$ , for some  $\lambda > 0$ .*

*Proof.* Fix any  $m_0 \in \mathbb{R}_{>0}$ . By the continuity hypothesis in (3.3), we know in particular that there exists  $r > 0$ , depending only on  $m_0$ , such that

$$d_{\mathcal{M}}(m\delta_0, m_0\delta_0) \leq 1,$$

for every  $m \in (m_0 - r, m_0 + r)$ . Moreover, exploiting the isometric invariance (3.3) and arguing as in the proof of Proposition 3.1, we have

$$\begin{aligned} d_{\mathcal{M}}(m\delta_x, m_0\delta_x) &= d_{\mathcal{M}}(m\tau_x\#\delta_0, m_0\tau_x\#\delta_0) = d_{\mathcal{M}}(\tau_x\#(m\delta_0), \tau_x\#(m_0\delta_0)) \\ &= d_{\mathcal{M}}(m\delta_0, m_0\delta_0), \end{aligned}$$

for every  $x \in \mathbb{R}^n$ . Therefore,  $d_{\mathcal{M}}$  satisfies hypothesis (ii) in Theorem 1.1, with  $\Sigma = \mathbb{R}^n$ , and the conclusion follows.  $\square$

**3.1. Examples.** The simplest examples of distance functions defined on  $\mathcal{M}_c^*(\mathbb{R}^n)$  which satisfy (1.1) (with  $f$  constant) are given by product metrics. Indeed, if  $d$  is any distance function on  $(0, +\infty)$ , then

$$d_{\mathcal{M}}(\mu_1, \mu_2) := \sqrt{d^2(|\mu_1|, |\mu_2|) + \lambda^2 W_p^2\left(\frac{\mu_1}{|\mu_1|}, \frac{\mu_2}{|\mu_2|}\right)},$$

is such a distance function. Of particular interest are the cases given in Example 3.3 and Example 3.4 here below, which also satisfy the boundedness condition (1.3).

**Example 3.3.** Let  $q \in [1, +\infty)$  and define

$$d_{\mathcal{M},q}(\mu_1, \mu_2) := \left( \left| |\mu_1| - |\mu_2| \right|^q + \lambda^q W_p^q\left(\frac{\mu_1}{|\mu_1|}, \frac{\mu_2}{|\mu_2|}\right) \right)^{\frac{1}{q}},$$

and also

$$d_{\mathcal{M},\infty}(\mu_1, \mu_2) := \max \left\{ \left| |\mu_1| - |\mu_2| \right|, \lambda W_p\left(\frac{\mu_1}{|\mu_1|}, \frac{\mu_2}{|\mu_2|}\right) \right\},$$

for every  $\mu_1, \mu_2 \in \mathcal{M}_c^*(\mathbb{R}^n)$ . Then,  $d_{\mathcal{M},q}$  is a distance function on  $\mathcal{M}_c^*(\mathbb{R}^n)$  satisfying points (i) and (ii) of Theorem 1.1, for every  $q \in [1, +\infty]$ . Actually, each  $d_{\mathcal{M},q}$  is such that

$$\sup_{\mu \in \mathcal{P}_c(\mathbb{R}^n)} d_{\mathcal{M},q}(m_0\mu, m\mu) = |m_0 - m|,$$

for every  $m, m_0 \in \mathbb{R}_{>0}$ , which is more restrictive than point (ii).

**Example 3.4.** Let  $d$  be a bounded distance function on  $(0, +\infty)$ , with  $d(m_1, m_2) \leq C_d$  for every  $m_1, m_2 \in (0, +\infty)$ . Then,

$$d_{\mathcal{M}}(\mu_1, \mu_2) := d(|\mu_1|, |\mu_2|) + \lambda W_p\left(\frac{\mu_1}{|\mu_1|}, \frac{\mu_2}{|\mu_2|}\right),$$

is a distance function on  $\mathcal{M}_c^*(\mathbb{R}^n)$ , satisfying point (i) of Theorem 1.1 and the global boundedness property

$$\sup_{\mu \in \mathcal{P}_c(\mathbb{R}^n)} d_{\mathcal{M}}(m_0\mu, m\mu) \leq C_d,$$

for every  $m_0, m \in (0, +\infty)$ , which is clearly stronger than point (ii). In particular, by considering as  $d$  the discrete metric on  $(0, +\infty)$ , we obtain an example of distance function  $d_{\mathcal{M}}$  for which we do not have continuity in the mass, in the sense that

$$\lim_{|m-m_0| \rightarrow 0} d_{\mathcal{M}}(m_0\mu, m\mu) = 0,$$

does not hold true for any  $\mu \in \mathcal{P}_c(\mathbb{R}^n)$ .

Other less trivial examples can be obtained as the so-called warped product metrics, by exploiting the fact that  $(\mathcal{P}_c(\mathbb{R}^n), W_p)$  is a geodesic space, see [17].

**Example 3.5** (Warped products). Let  $d_{\mathcal{M},2}$  be the product metric on  $\mathcal{M}_c^*(\mathbb{R}^n)$  defined in Example 3.3. Given a Lipschitz continuous curve  $\gamma : I \rightarrow (\mathcal{M}_c^*(\mathbb{R}^n), d_{\mathcal{M},2})$ , where  $I := [0, 1] \subseteq \mathbb{R}$ , we denote by  $m_\gamma : I \rightarrow \mathbb{R}_{>0}$  and  $\rho_\gamma : I \rightarrow (\mathcal{P}_c(\mathbb{R}^n), W_p)$  its projections, defined respectively as

$$m_\gamma(t) := |\gamma(t)| \quad \text{and} \quad \rho_\gamma(t) := \frac{\gamma(t)}{|\gamma(t)|},$$

which are Lipschitz continuous curves. Let  $\Gamma$  denote the set of all such curves. Given a continuous function  $g : (\mathcal{P}_c(\mathbb{R}^n), W_p) \rightarrow \mathbb{R}_{>0}$ , we define the warped product metric, with warping function  $g$ , as

$$d_{\mathcal{M},g}(\mu_1, \mu_2) := \inf_{\substack{\gamma \in \Gamma \\ \gamma(0)=\mu_1, \gamma(1)=\mu_2}} \int_I \sqrt{g^2(\rho_\gamma(t)) |m'_\gamma(t)|^2 + |\rho'_\gamma(t)|^2} dt.$$

Here above we have denoted, with a slight abuse of notation,

$$|\rho'_\gamma(t)| := \lim_{s \rightarrow 0} \frac{W_p(\rho_\gamma(t), \rho_\gamma(t+s))}{|s|},$$

which exists for almost every  $t \in I$ , since  $\rho_\gamma$  is Lipschitz, see, e.g., [2, Theorem 2.7.6]. This definition can be found, e.g., in [2, Section 3.6.4], and it is easy to check that it coincides with the one given in [4, Section 3.1]. By [4, Proposition 3.1],  $d_{\mathcal{M},g}$  is a distance function on  $\mathcal{M}_c^*(\mathbb{R}^n)$ , and, by [4, Lemma 3.2], it satisfies (1.1), as indeed

$$(3.4) \quad d_{\mathcal{M},g}(\mu_1, \mu_2) = W_p\left(\frac{\mu_1}{|\mu_1|}, \frac{\mu_2}{|\mu_2|}\right),$$

for every  $\mu_1, \mu_2 \in \mathcal{M}_c^*(\mathbb{R}^n)$  such that  $|\mu_1| = |\mu_2|$ . It is easy to find examples of warping functions  $g$  for which the corresponding distance  $d_{\mathcal{M},g}$  does not satisfy point (ii) of Theorem 1.1, such as  $g(\mu) := 1 + W_p(\mu, \delta_0)$ . Similarly, if we consider the warping function

$$g(\mu) := 1 + \inf_{x \in \mathbb{R}^n} W_p(\mu, \delta_x),$$

then,  $d_{\mathcal{M},g}$  satisfies point (ii) of Theorem 1.1, with  $\Sigma = \mathbb{R}^n$ . However, since we can find a sequence  $\{\mu_k\}_{k \in \mathbb{N}} \subseteq \mathcal{P}_c(\mathbb{R}^n)$  such that  $g(\mu_k) \geq k$  for every  $k$ , it is easy to verify that (1.3) does not hold on the whole of  $\mathcal{P}_c(\mathbb{R}^n)$ . That is, there exist no  $m_0, r, C > 0$  for which

$$\sup_{\mu \in \mathcal{P}_c(\mathbb{R}^n)} d_{\mathcal{M},g}(m_0\mu, m\mu) \leq C,$$

for every  $m \in (m_0 - r, m_0 + r)$ .

Furthermore, we observe that for every choice of warping function  $g$ , the distance  $d_{\mathcal{M},g}$  is compatible with weak convergence when varying the mass, as indeed

$$d_{\mathcal{M},g}(m_0\mu, m\mu) \leq g(\mu)|m - m_0|,$$

for every  $\mu \in \mathcal{P}_c(\mathbb{R}^n)$ . Nevertheless, since (3.4) holds true, by arguing as in the proof of Corollary 2.4 we see that no such distance can be extended to a distance  $d$  defined on the whole of  $\mathcal{M}_c(\mathbb{R}^n)$  and with the property that

$$\lim_{m \searrow 0} d(0, m\mu_i) = 0,$$

for at least two different  $\mu_1 \neq \mu_2 \in \mathcal{P}_c(\mathbb{R}^n)$ .



## REFERENCES

1. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR 2401600
2. Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001. MR 1835418
3. Luis A. Caffarelli, *Some regularity properties of solutions of Monge Ampère equation*, Comm. Pure Appl. Math. **44** (1991), no. 8-9, 965–969. MR 1127042
4. Chien-Hsiung Chen, *Warped products of metric spaces of curvature bounded from above*, Trans. Amer. Math. Soc. **351** (1999), no. 12, 4727–4740. MR 1466944
5. Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard, *An interpolating distance between optimal transport and Fisher-Rao metrics*, Found. Comput. Math. **18** (2018), no. 1, 1–44. MR 3749413
6. Emiliano Cristiani, Benedetto Piccoli, and Andrea Tosin, *Multiscale modeling of granular flows with application to crowd dynamics*, Multiscale Model. Simul. **9** (2011), no. 1, 155–182. MR 2769993
7. Alessio Figalli, *The optimal partial transport problem*, Arch. Ration. Mech. Anal. **195** (2010), no. 2, 533–560. MR 2592287
8. Alessio Figalli and Nicola Gigli, *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions*, J. Math. Pures Appl. (9) **94** (2010), no. 2, 107–130. MR 2665414
9. Alfred Galichon, *Optimal transport methods in economics*, Princeton University Press, Princeton, NJ, 2016. MR 3586373
10. Stanislav Kondratyev, Léonard Monsaingeon, and Dmitry Vorotnikov, *A new optimal transport distance on the space of finite Radon measures*, Adv. Differential Equations **21** (2016), no. 11-12, 1117–1164. MR 3556762
11. Matthias Liero, Alexander Mielke, and Giuseppe Savaré, *Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves*, SIAM J. Math. Anal. **48** (2016), no. 4, 2869–2911. MR 3542003
12. Daniel Matthes and Giuseppe Toscani, *Analysis of a model for wealth redistribution*, Kinet. Relat. Models **1** (2008), no. 1, 1–27. MR 2383713
13. Bertrand Maury, Aude Roudneff-Chupin, and Filippo Santambrogio, *A macroscopic crowd motion model of gradient flow type*, Math. Models Methods Appl. Sci. **20** (2010), no. 10, 1787–1821. MR 2735914
14. Gabriel Peyré and Marco Cuturi, *Computational optimal transport: With applications to data science*, Foundations and Trends in Machine Learning **11** (2019), no. 5-6, 355–607.
15. Benedetto Piccoli and Francesco Rossi, *Generalized Wasserstein distance and its application to transport equations with source*, Arch. Ration. Mech. Anal. **211** (2014), no. 1, 335–358. MR 3182483
16. ———, *On properties of the generalized Wasserstein distance*, Arch. Ration. Mech. Anal. **222** (2016), no. 3, 1339–1365. MR 3544329
17. Cédric Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR 1964483
18. ———, *Optimal transport*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR 2459454

UNIVERSITÀ DEGLI STUDI DI PADOVA, DIPARTIMENTO DI MATEMATICA “TULLIO LEVI-CIVITA”, VIA TRIESTE 63, 35121 PADOVA, ITALY

UNIVERSITÀ DEGLI STUDI DI PADOVA, DIPARTIMENTO DI MATEMATICA “TULLIO LEVI-CIVITA”, VIA TRIESTE 63, 35121 PADOVA, ITALY