Journal of Mathematical Sciences, Vol. 181, No. 6, March, 2012

A DESCRIPTION OF TRANSPORT COST FOR SIGNED MEASURES

E. Mainini*

UDC 517.958

In this paper, we develop the analysis started in a paper by Ambrosio, Mainini, and Serfaty about the extension of the optimal transport framework to the space of real measures. The main motivation comes from the study of nonpositive solutions to some evolution PDEs. Although a canonical optimal transport distance does not seem to be available, we may describe the cost for transporting signed measures in various ways and with interesting properties. Bibliography: 22 titles.

1. Introduction

Transportation problem. Consider the Euclidean space $\mathbb{R}^d$, and let $\mathscr{P}(\mathbb{R}^d)$ denote the space of probability measures over $\mathbb{R}^d$. Moreover, given a probability measure γ on the product space $\mathbb{R}^d \times \mathbb{R}^d$, denote by $\pi^1_\#\gamma$ its first marginal and by $\pi^2_\#\gamma$ its second marginal. Now, we are given two probabilities μ, ν with finite pth moment over $\mathbb{R}^d$, $p \geq 1$. Let $\gamma \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d)$ be a coupling of μ and ν, that is, a joint measure with these marginals. It is called a transport plan. If we suppose to transport a unit quantity of mass from $x \in \operatorname{supp}\mu$ to $y \in \operatorname{supp}\nu$ with cost $|x-y|^p$, the global cost associated to the coupling γ is

$$\int\limits_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p\, d\gamma(x,y).$$

It is then natural to consider the following linear minimization problem, called the optimal transportation problem:

$$\inf\left\{ \int\limits_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p\, d\gamma(x,y) : \gamma \in \mathscr{P}(\mathbb{R}^d\times\mathbb{R}^d),\ \pi^1_\#\gamma = \mu,\ \pi^2_\#\gamma = \nu \right\}. \tag{1.1}$$

This formulation is due to Kantorovich. The work of Kantorovich goes back to the forties [15, 16], and the problem (1.1) is itself a reformulation of the original Monge problem, presented in the eighteenth century. Nonetheless, in the recent years, since the beginning of the nineties, optimal transportation has become (and is becoming) a very topical research field. Starting from the paper by Brenier [6], the study of the regularity of solutions and their characterization has drawn the attention of many mathematicians. Moreover, optimal transportation has proved to be a useful tool for many applications in different mathematical contexts. A description of the literature would be long, only few authors are cited in the bibliography. But for a general and complete overview on the topic, we refer the reader to the books by Villani [20, 21].

In this paper, we would like to present a problem that naturally arose during the analysis in one of the fields of application. This problem could be interesting on its own, at the level of the basic formulation of the optimal transportation problem. In order to get to the point, we simply have to recall the first elementary facts about the problem (1.1). First, we have the existence of solutions, guaranteed by standard direct method techniques. The attained minimum defines the optimal transport cost between the measures μ and ν. Second, it is standard stuff to show, using the properties of transport plans, that $W_p(\mu,\nu)$, the $(1/p)$th power of this cost, is a distance on $\mathscr{P}_p(\mathbb{R}^d)$ (the space of probability measures with finite pth moment), therefore named the optimal transport distance. It is also referred to in the literature as the Wasserstein distance, sometimes the Kantorovich–Rubinstein–Wasserstein distance.

Main task. We are already able to formulate the task. The optimal transportation problem contains the definition of a distance on the space of probability measures, the Wasserstein distance. *What if we want to generalize the problem to the space of signed measures? Can we find a consistent generalization, without losing all the good properties that are needed in applications? Can we endow the space of real measures with a kind of Wasserstein distance, in a canonical way?*

*Dipartimento di Matematica "F. Casorati," Università degli Studi di Pavia, Italy, e-mail: `edoardo.mainini@unipv.it`.

Published in *Zapiski Nauchnykh Seminarov POMI*, Vol. 390, 2011, pp. 147–181. Original article submitted September 22, 2011.

The answer to these questions is not straightforward. Indeed, we will see that some extensions are available, the basic one consisting in

$$\mathbb{W}_p(\mu,\nu) := W_p(\mu^+ + \nu^-, \nu^+ + \mu^-),$$

where $+$ and $-$ denote the positive and negative parts. The price to pay is the loss of some properties: among these, the feature of the cost being really a distance (unless we consider the degenerate case $p = 1$). For the details, we refer to the subsequent sections, and here we limit ourselves to stressing one of the basic reasons for the difficulties that arise. In the standard transportation problem among positive measures, the basic constraint to be satisfied is the mass conservation

$$\int_{\mathbb{R}^d} \mu = \int_{\mathbb{R}^d} \nu.$$

Then, in general, one may reduce the problem to the case of probabilities. On the other hand, if we have signed measures μ and ν, we again have to impose the same constraint. But, certainly, this will not imply the same for $|\mu|$ and $|\nu|$. Hence, in the signed case, we have to deal with a possible variation of mass. This changes a bit the nature of the problem.

Main motivations and applications. Let us now describe in some more detail the context in which these questions arose. In the study of measure solutions to some nonlinear partial differential equations of evolution type, some techniques from optimal transportation can be used for constructing suitable approximation schemes. Such an approach was developed by Otto, for the analysis of the heat and porous media equations, and continued in various other papers (see [14, 19] and the references in [2, 20, 21]). The approach consists in viewing a conservative equation

$$\partial_t \mu + \operatorname{div}(\mathbf{v}\mu) = 0, \tag{1.2}$$

with velocity vector field $\mathbf{v}$ being a gradient (and possibly depending nonlinearly on μ), as the "gradient flow" of the corresponding energy Φ with respect to the optimal transport structure. The gradient flow of a functional $\Phi : \mathscr{P}_p(\mathbb{R}^d) \to \mathbb{R}$ should be, as expected, a descent curve along the direction of the negative gradient. Some work is needed to formalize this concept in probability spaces. Indeed, using optimal transportation, one can develop a differential calculus in $\mathscr{P}_p(\mathbb{R}^d)$. This is done in [2]. Within this framework, one associates an energy Φ to Eq. (1.2) by saying that the vector field $\mathbf{v}$ is the "Wasserstein gradient" of Φ. To sketch the theory in a simpler way, we take advantage of another point of view. We may consider the following Euler implicit approximation scheme corresponding to Eq. (1.2) in the framework of probability measure solutions: given $\mu^0 \in \mathscr{P}_p(x)$, find recursively the solution μ_τ^{k+1} of

$$\min_{\nu \in \mathscr{P}_p(X)} \Phi(\nu) + \frac{1}{p\tau^{p-1}} W_p^p(\mu_\tau^k, \nu), \tag{1.3}$$

where τ is the discretization parameter. Interpolating the discrete minimizers and passing to the limit as $\tau \to 0$, one expects to find solutions to the continuity equation. In the general metric setting, this is known as the De Giorgi (see [10]) minimizing movements scheme. Here we are working in the metric space $(\mathscr{P}_p(\mathbb{R}^d), W_p)$. The most common setting is $p = 2$. It is worth pointing out, in view of the subsequent discussion, that this scheme might work even if the term perturbing Φ is not a distance. As in the seminal paper [1], one could also use a nontriangular or nonsymmetric object.

The Wasserstein gradient flow approach is very useful for obtaining well-posedness and stability results. On the other hand, the way we presented it is quite general, and Eq. (1.2) is a model for describing a wide variety of physical evolution models. For these reasons, it seems worth trying to extend the theory in order to have the possibility of attacking new problems. One of these possible extensions is the study of nonpositive solutions. Hence the goal is to find a way to approach a problem of the form (1.2) with sign-changing solutions through a transport-like scheme.

Let us show some examples of problems where it would be natural to consider signed measure solutions, for which the desired generalization could be useful.

- The evolution problem describing the motion of a density under the effect of a continuous interaction potential $W : \mathbb{R}^d \to \mathbb{R}$:

$$\partial_t \mu + \operatorname{div}((\nabla W * \mu)\mu) = 0. \tag{1.4}$$

  This equation appears in the study of granular media and aggregation phenomena. It is also a standard model for Wasserstein gradient flows (in the case of a smooth convex potential). See, for instance, [8] and the general discussions in [20, 2]. It is also suitable for a description of the dynamics of a system of particles. It would be quite natural to consider the case in which particles possess different charges. This way, if μ is the density of particles, it will be a signed measure.

- The Chapman–Rubinstein–Schatzman–E (see [9, 11]) model for Ginzburg–Landau vortices in two dimensions:
$$\partial_t \mu + \operatorname{div}\left((\nabla(\Delta^{-1}\mu))|\mu|\right) = 0. \tag{1.5}$$
Here μ represents the vortex density, and it is suitable, from the physical point of view, to consider vortices with equal and opposite topological degrees which may cancel each other during the evolution. Again, μ is a signed measure. Note that this model can be regarded as an interaction model as well, since $\Delta^{-1}\mu$ can be written as the convolution with a suitable Green kernel. The difference lies in the fact that the velocity field is less regular, being multiplied by $\operatorname{sgn}\mu$. In $\mathscr{P}_2(\mathbb{R}^d)$ this problem was studied as a gradient flow in [4, 17].

Next we list some difficulties that one encounters when passing to signed measures.

- The initial task of this paper: there is no standard definition of optimal transport distance on the space of signed measures.
- There is no standard relation between solutions to the continuity equation and absolutely continuous curves, whereas the relation is clear in the space of probability measures (see [2]).
- It is reasonable to expect more difficulties when searching for suitable compactness estimates within approximation schemes.

All these problems arose in the paper [3], during the analysis of an evolution model for signed measures of the form (1.5).

Plan of the paper. Motivated by the above applications, and by the interest to the general optimal transportation problem, in this paper we aim to consider the basic question, that of finding a suitable transport cost among signed measures. Following the ideas of [3, Sec. 2], we will give different possible definitions. Essentially, this paper does not contain new results. Rather, we tried to give a linear and exhaustive presentation of the subject, with many examples and details on transport of measures and its mathematical description. The only exception is Sec. 2.3, where we will give a result on the behavior of the Wasserstein distance on different mass scales: this will also give some insight into the difficulties arising for signed measures.

The rest of the paper is organized in two sections. In Sec. 2, we recall the definition of the Wasserstein distance and give examples of the corresponding transport paths, with particular attention to the case of atomic measures. We add a brief discussion on the behavior of the distance when the involved masses change. In Sec. 3, we describe some possible ways to define the transport cost in the case of signed measures, again with different examples, paying attention to the relations with the standard Wasserstein distance and to various geometric and topological properties.

Acknowledgments. This paper was written for the special volume of proceedings of the conference on *Optimization and Stochastic Methods for Spatially Distributed Information*, held in St.Petersburg in May 2010. The author wishes to thank Andrei Sobolevski and Eugene Stepanov for the kind invitation. This paper contains a development and a more complete exposition of some results sketched in [3], hence the author wishes to thank Luigi Ambrosio and Sylvia Serfaty as well.

2. Transport cost: the standard definition

2.1. Basic notions. We begin with the basic definitions. Let Ω be a Banach space with norm $|\cdot|$. The theory could be developed in more general settings (for instance, Ω could be a complete separable metric space, whose distance would replace the norm $|\cdot|$), but for our discussion it is not needed to be specific at this level. Denote by $\mathscr{P}(\Omega)$ the space of probability measures over Ω. This space is naturally endowed with the narrow topology, defined by the duality with $C^0_b(\Omega)$, the space of continuous and bounded functions over Ω. That is, we say that a sequence $(\mu_n) \subset \mathscr{P}(\Omega)$ narrowly converges to $\mu \in \mathscr{P}(\Omega)$ (and write $\mu_n \rightharpoonup \mu$) if
$$\int_\Omega \varphi\, d\mu_n \to \int_\Omega \varphi\, d\mu \quad \text{for every} \quad \varphi \in C^0_b(\Omega).$$

Given a measure $\mu \in \mathscr{P}(\Omega)$ and a map $\mathbf{t}: \Omega \to \Omega$, we define the *push-forward* measure $\mathbf{t}_\#\mu$ of μ through $\mathbf{t}$ by the standard relation
$$\mathbf{t}_\#\mu(\Omega_0) := \mu(\mathbf{t}^{-1}(\Omega_0)),$$

where Ω_0 is any Borel set in Ω.

Given a measure γ in the product space $\Omega \times \Omega$ and the projection maps π^1 and π^2 to its factors, $\pi^1_{\#}\gamma$ and $\pi^2_{\#}\gamma$ will be the first and the second marginal of γ, respectively. This means that for any Borel set $\Omega_0 \subset \Omega$,

$$\pi^1_{\#}\gamma(\Omega_0) = \gamma(\Omega_0 \times \Omega) \quad \text{and} \quad \pi^2_{\#}\gamma(\Omega_0) = \gamma(\Omega \times \Omega_0).$$

Given two measures $\mu, \nu \in \mathscr{P}(\Omega)$, of course, there are many ways to couple them through a measure $\gamma \in \mathscr{P}(\Omega \times \Omega)$ such that the marginals of γ are μ and ν. We define the set of transport plans between μ and ν as

$$\Gamma(\mu, \nu) := \{\gamma \in \mathscr{P}(\Omega \times \Omega) : \pi^1_{\#}\gamma = \mu, \pi^2_{\#}\gamma = \nu\}.$$

Given a measure $\mu \in \mathscr{P}(\Omega)$, its pth moment is defined as $\int_\Omega |x|^p \, d\mu$. The set of probability measures over Ω that have finite pth moment is denoted by $\mathscr{P}_p(\Omega)$. We also say that a sequence $(\mu_n) \subset \mathscr{P}_p(\Omega)$ converges to μ in $\mathscr{P}_p(\Omega)$ if it narrowly converges to μ and the corresponding pth moments converge to the pth moment of μ. This topology will be referred to as the $\mathscr{P}_p(\Omega)$ topology. On bounded subsets of Ω, it coincides with the narrow topology.

The set $\mathscr{M}^+(\Omega)$ of positive Borel measures over Ω may also be endowed with the narrow topology. Moreover, we recall that a subset Ξ of $\mathscr{M}^+(\Omega)$ is said to be tight if for any $\varepsilon > 0$ there exists a compact subset $\Omega_0 = \Omega_0(\varepsilon)$ of Ω such that $\sup_{\mu \in \Xi} \mu(\Omega \setminus \Omega_0(\varepsilon)) \le \varepsilon$. By the classical Prokhorov theorem, a subset Ξ of $\mathscr{M}^+(\Omega)$ is narrowly compact if it is tight and $\sup_{\mu \in \Xi} \mu(\Omega) < +\infty$. We also let $\mathscr{M}^+_p(\Omega)$ denote the subset of positive measures with finite pth moment. A sufficient condition for a set $\Xi \subset \mathscr{M}^+_p(\Omega)$ to be tight is the uniform boundedness of the pth moments on Ξ. For the details on narrow topologies, we refer to measure theory textbooks, such as [5].

2.2. The Wasserstein distance. Kantorovich's formulation of the optimal transportation problem can be given according to (1.1). Note that in such a variational problem, both the functional and the constraints are linear. Moreover, the set $\Gamma(\mu, \nu)$ is always nonempty, since it contains at least the product measure $\mu \times \nu$, and it is not difficult to show that it is a tight set in the narrow topology on $\mathscr{P}(\Omega \times \Omega)$ (see, for instance, [2, Chap. 5]). The narrow lower semicontinuity of the integral functional is also a standard fact, since $| \cdot |^p$ is a continuous and nonnegative function, so that, applying the direct method of the calculus of variations, we see that the Kantorovich problem admits a solution.

The achieved infimum defines the *Wasserstein distance*. Hence, letting $p \ge 1$, the definition of the Wasserstein distance is

$$W_p(\mu, \nu) = \left(\int_{\Omega \times \Omega} |x - y|^p \, d\gamma(x, y) \right)^{1/p}, \tag{2.1}$$

for $\gamma \in \Gamma^p_o(\mu, \nu)$, which is the convex set of transport plans where the infimum of (1.1) is achieved:

$$\Gamma^p_o(\mu, \nu) := \left\{ \gamma \in \Gamma(\mu, \nu) : \int_{\Omega \times \Omega} |x - y|^p \, d\gamma(x, y) \le \int_{\Omega \times \Omega} |x - y|^p \, d\tilde{\gamma}(x, y) \quad \text{for every} \quad \tilde{\gamma} \in \Gamma(\mu, \nu) \right\}. \tag{2.2}$$

We stress that in general this set is not independent of the choice of the exponent p, even if sometimes the superscript is omitted. Moreover, a distinguished role is played by the "rotund" case $p = 2$.

The fact that the transport cost given by (2.1) indeed defines a distance is standard. For the proof of the triangle inequality, we refer, for instance, to [2, 20, 21]. Some other usual properties, for the proof of which we refer the reader to the same references, are the following.

- The distance W_p metrizes the $\mathscr{P}_p(\Omega)$ topology.
- Given $\mu \in \mathscr{P}_p(\Omega)$, the map $\nu \mapsto W(\mu, \nu)$ is lower semicontinuous with rrespect to the narrow topology and continuous with respect to the $\mathscr{P}_p(\Omega)$ topology. Moreover, if $\mu_n \to \mu$ in $\mathscr{P}_p(\Omega)$ and $\nu_n \to \nu \in \mathscr{P}_p(\Omega)$, then $W_p(\mu_n, \nu_n) \to W_p(\mu, \nu)$.

The next step consists in defining the distance between positive measures over Ω with the same mass α, possibly different from 1. Let $\mathscr{M}^\alpha(\Omega) \subset \mathscr{M}^+(\Omega)$, $\alpha > 0$, denote the set of such measures. As usual, $\mathscr{M}^\alpha_p(\Omega)$ is the corresponding subset of measures with bounded pth moment. Given μ and ν in $\mathscr{M}^\alpha(\Omega)$, we again have

$$\Gamma(\mu, \nu) := \{\gamma \in \mathscr{M}^\alpha(\Omega \times \Omega) : \pi^1_{\#}\gamma = \mu, \ \pi^2_{\#}\gamma = \nu\}.$$

This is a nonempty set, because the measure $\frac{1}{\alpha}(\mu \times \nu)$ belongs to it (while the product $\mu \times \nu$ does not). Then $\Gamma^p_o(\cdot,\cdot)$ and $W_p(\cdot,\cdot)$ are defined as in (2.2) and (2.1). All the properties holding for probability measures trivially extend to this case.

It is easy to see that if either μ or ν is concentrated in a single point, then $\Gamma(\mu,\nu)$ contains the unique element $\frac{1}{\alpha}(\mu \times \nu)$, as in the first example of Fig. 1. We also recall that in the case of Dirac masses, the transportation can be described in a simple way, which is the following. Let $M, N \in \mathbb{N}$, and let $\mu, \nu \in \mathscr{M}^\alpha(\Omega)$ be of the form

$$\mu = \sum_{i=1}^{N} u_i \delta_{x_i}, \quad \nu = \sum_{j=1}^{M} v_j \delta_{y_j}, \quad \text{with} \quad \sum_{i=1}^{N} u_i = \sum_{j=1}^{M} v_j = \alpha,\ u_i > 0,\ v_j > 0. \tag{2.3}$$

Here the x_i's are N distinct points in Ω and the y_j's are M distinct points in Ω. Then any element γ of $\Gamma(\mu,\nu)$ can be written as

$$\gamma = \sum_{i=1}^{N}\sum_{j=1}^{M} w_{i,j}(\delta_{x_i} \times \delta_{y_j}). \tag{2.4}$$

Here $w_{i,j} \geq 0$ is a suitable weight indicating the (possibly null) quantity of mass that is transported from x_i to y_j. We have to add the constraints

$$\sum_{j=1}^{M} w_{i,j} = u_i, \quad \sum_{i=1}^{N} w_{i,j} = v_j,$$

which say that the total mass leaving x_i is equal to u_i and the total mass arriving to y_j is equal to v_j. The optimal transport plans are then given by suitable choices of the weights. If $w_{i,j}$ are optimal weights, then the Wasserstein distance is given by

$$W_p(\mu,\nu) = \left(\sum_{i=1}^{N}\sum_{j=1}^{M} w_{i,j}\, |x_i - y_j|^p \right)^{1/p}.$$

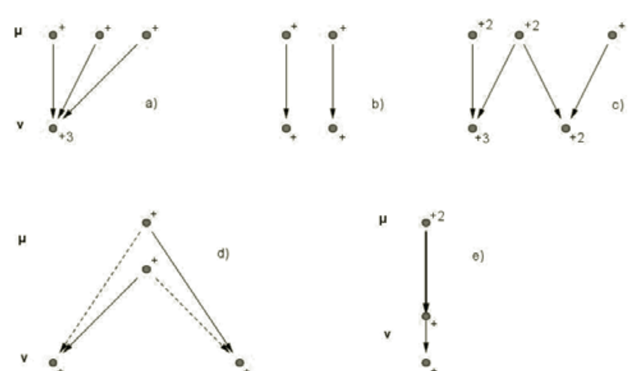

Fig. 1. Examples of optimal mass transportation among positive measures.
a) $\mu = \delta_{(0,2)} + \delta_{(1,2)} + \delta_{(2,2)}$, $\nu = 3\delta_{(0,0)}$. The unique optimal transport plan is $\delta_{(0,2)} \times \delta_{(0,0)} + \delta_{(1,2)} \times \delta_{(0,0)} + \delta_{(2,2)} \times \delta_{(0,0)}$, hence $W_p(\mu,\nu) = (2^p + \sqrt{5}^p + 2^p\sqrt{2}^p)^{1/p}$.
b) $\mu = \delta_{(0,2)} + \delta_{(1,2)}$, $\nu = \delta_{(0,0)} + \delta_{(1,0)}$. The unique optimal plan: $\delta_{(0,2)} \times \delta_{(0,0)} + \delta_{(1,2)} \times \delta_{(1,0)}$. The associated cost: $W_p(\mu,\nu) = 2^{(p+1)/p}$.
c) $\mu = 2\delta_{(0,2)} + 2\delta_{(1,2)} + \delta_{(3,2)}$, $\nu = 3\delta_{(0,0)} + 2\delta_{(2,0)}$. The unique optimal plan: $2(\delta_{(0,2)} \times \delta_{(0,0)}) + \delta_{(1,2)} \times \delta_{(0,0)} + \delta_{(1,2)} \times \delta_{(2,0)} + \delta_{(3,2)} \times \delta_{(2,0)}$. This is an example where the mass splits.
d) This is an example where an optimal transport plan is not unique: $\mu = \delta_{(0,2)} + \delta_{(0,3)}$, $\nu = \delta_{(-2,0)} + \delta_{(2,0)}$. The solid line and the dashed one are two optimal transport plans, both corresponding to the cost $W_p(\mu,\nu) = (\sqrt{8}^p + \sqrt{13}^p)^{1/p}$, and any convex combination of them is another optimal transport plan.
e) $\mu = 2\delta_{(0,3)}$, $\nu = \delta_{(0,1)} + \delta_{(0,0)}$. The optimal plan: $\delta_{(0,3)} \times \delta_{(0,1)} + \delta_{(0,3)} \times \delta_{(0,0)}$. The cost: $(3^p + 2^p)^{1/p}$. Note that the plan cannot be written as $\delta_{(0,3)} \times (\delta_{(0,1)} + \delta_{(0,0)})$.

In Fig. 1, various examples of transportation according to this framework (in $\mathbb{R}^2$) are shown. In the examples of Fig. 1, the sets $\Gamma_o(\mu,\nu)$ do not depend on the exponent p. An important example where things are different is the following one: suppose that we work on the real line, and let

$$\mu = \delta_0 + \delta_1, \quad \nu = \delta_1 + \delta_2.$$

Two transport plans in $\Gamma(\mu,\nu)$ are

$$\gamma_1 = \delta_0 \times \delta_1 + \delta_1 \times \delta_2 \quad \text{and} \quad \gamma_2 = \delta_0 \times \delta_2 + \delta_1 \times \delta_1.$$

Let us evaluate the corresponding costs. We find

$$\left(\int\limits_{\mathbb{R}\times\mathbb{R}} |x-y|^p \, d\gamma_1 \right)^{1/p} = 2^{1/p},$$

so that the cost of γ_1 does depend on p; but this is always an optimal plan, hence $\gamma_1 \in \Gamma_o^p(\mu,\nu)$ for any $p \geq 1$. On the other hand,

$$\left(\int\limits_{\mathbb{R}\times\mathbb{R}} |x-y|^p \, d\gamma_2 \right)^{1/p} = 2,$$

so that the cost does not depend on p. Comparing the costs, we see that γ_2 is not optimal if $p > 1$. It is optimal in the unique case $p = 1$.

2.3. Scaling properties. The latter example illustrates a particular feature of the 1-distance: it does not increase if we add the same measure to the source and to the target. We stress that this is an important property, and it will play a role when dealing with signed measures. It can be seen very easily in the general framework, invoking the Kantorovich dual formulation of the optimal transportation problem. Indeed, since the Kantorovich problem is linear, we can define a dual problem, which is

$$\sup \Big\{ \int\limits_\Omega \phi \, d\mu + \int\limits_\Omega \psi \, d\nu : \phi(x) + \psi(y) \leq |x-y|^p, \phi \in L^1(X,\mu), \psi \in L^1(Y,\nu) \Big\}, \tag{2.5}$$

and the supremum equals the infimum in the original problem. A significant particular instance of the Kantorovich duality is obtained for $p = 1$, namely,

$$W_1(\mu,\nu) = \sup_{\varphi \in \mathrm{Lip}(\Omega), \|\varphi\|_{\mathrm{Lip}} \leq 1} \int\limits_\Omega \varphi \, d(\mu - \nu). \tag{2.6}$$

Looking at (2.6), the claimed property is immediate. The common mass of μ and ν may stay in place, in the solution of the optimal transportation problem. This is also known as the *book-shifting* example: we are given n books strung together, and we shift the whole line by a given distance d. We also obtain the same target configuration (if the order is not to be preserved) by simply moving the first book to the top of the queue. Hence we do not move the $n-1$ books in the positions common for the starting and final configurations. For the W_1 cost, this is not more expensive. See [13, Proposition 2.9] for the properties ensuring that the common mass does not move (and for a general discussion on duality, we refer again to [20, 21]). Things are different for the case $p > 1$, as clarified by the next theorem.

Proposition 2.1. *Let $\alpha, \beta \geq 0$, let $\mu, \nu \in \mathscr{M}^\alpha(\Omega)$, and let $\sigma \in \mathscr{M}^\beta(\Omega)$. Then for $p \geq 1$,*

$$W_p(\mu,\nu) \geq W_p(\mu+\sigma, \nu+\sigma),$$

and the equality holds for any σ if $p = 1$.

Proof. Let $\gamma_1 \in \Gamma_o^p(\mu,\nu)$. Let $\mathbf{1}$ be the identity map on Ω and $(\mathbf{1},\mathbf{1})$ be the corresponding vector map with values in $\Omega \times \Omega$. It is clear that $\gamma_1 + (\mathbf{1},\mathbf{1})_\#\sigma \in \Gamma(\mu+\sigma, \mu+\sigma)$ is a plan with the same p-cost, so that

$$W_p^p(\mu+\sigma,\nu+\sigma) \leq \int\limits_{\Omega\times\Omega} |x-y|^p \, d(\gamma_1 + (\mathbf{1},\mathbf{1})_\#\sigma) = \int\limits_{\Omega\times\Omega} |x-y|^p \, d\gamma_1 = W_p^p(\mu,\nu).$$

The equality for $p = 1$ follows by (2.6). □

On the other hand, we have the following result.

Theorem 2.2. *Let $p \geq 1$, $\alpha, \beta \geq 0$, and let $\mu, \nu \in \mathscr{M}_p^\alpha(\Omega)$ For any $p \geq 1$,*

$$\sup_{\sigma \in \mathscr{M}_p^\beta(\Omega)} W_p(\mu+\sigma, \nu+\sigma) = W_p(\mu, \nu)$$

and there exists $\sigma \in \mathscr{M}_p^{n\alpha}(\Omega)$ such that

$$W_p^p(\mu+\sigma, \nu+\sigma) \leq \frac{1}{(1+n)^{p-1}} W_p^p(\mu, \nu). \tag{2.7}$$

Proof. The first equality is trivial. Indeed, if μ, ν have compact support, then it is enough to choose σ with support on a set sufficiently far from them so as to make it inconvenient to move it. This way, if $\gamma_\mu^\nu \in \Gamma_o(\mu, \nu)$, then $\gamma_\mu^\nu + (\mathbf{1}, \mathbf{1})_\# \sigma \in \Gamma_o(\mu+\sigma, \nu+\sigma)$, and the cost of the diagonal term $(\mathbf{1}, \mathbf{1})_\# \sigma$ is zero. The general case is obtained by a simple approximation argument.

Let us prove (2.7). First assume that μ and ν are atomic and with finite supports, that is, they have the form (2.3). Let $w_{i,j} \leq \min\{u_i, v_j\}$, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, M\}$, be optimal weights, so that the plan (2.4) belongs to $\Gamma_o(\mu, \nu)$, as discussed above. In particular, let $\mathcal{A} = \{(i,j) \in \{1, \ldots, N\} \times \{1, \ldots, M\}\}$ be the set of the associated indices. If $(i,j) \in \mathcal{A}$, the p-Wasserstein distance between the measures $w_{i,j}\delta_{x_i}$ and $w_{i,j}\delta_{y_j}$ is, to the power p, $w_{i,j}|x_i - y_j|^p$. Then

$$W_p^p(\mu, \nu) = \sum_{(i,j) \in \mathcal{A}} w_{i,j}\, |x_i - y_j|^p.$$

Next, define an element $\overline{\sigma} \in \mathscr{M}^{n\alpha}(\Omega)$ as follows:

$$\overline{\sigma} := \sum_{(i,j) \in \mathcal{A}} \sum_{k=1}^{n} w_{i,j} \delta_{z_{i,j}^k}, \tag{2.8}$$

where $z_{i,j}^k$ are the points given by

$$z_{i,j}^k = \frac{(1+n-k)x_i + k y_j}{1+n}.$$

That is, we are uniformly partitioning each transport segment $[x_i, y_j]$ into $1+n$ parts, according to the available mass. The measure $\overline{\sigma}$ has a finite p-moment. Indeed,

$$\begin{aligned} \int_\Omega |x|^p \, d\overline{\sigma}(x) &= \sum_{(i,j) \in \mathcal{A}} \sum_{k=1}^{n} w_{i,j}\, |z_{i,j}^k|^p = \sum_{(i,j) \in \mathcal{A}} \sum_{k=1}^{n} w_{i,j} \left| \frac{(1+n-k)x_i - k y_j}{1+n} \right|^p \\ &\leq pn \sum_{(i,j) \in \mathcal{A}} w_{i,j}\, (|x_i|^p + |y_j|^p) = pn \sum_{i=1}^{n} u_i |x_i|^p + pn \sum_{j=1}^{M} v_j |y_j|^p \\ &= pn \int_\Omega |x|^p \, d(\mu+\nu)(x) < +\infty, \end{aligned} \tag{2.9}$$

since μ and ν have finite p-moments. Besides, it is clear that for any $(i,j) \in \mathcal{A}$,

$$\sum_{k=1}^{1+n} w_{i,j}\, (\delta_{z_{i,j}^{k-1}} \times \delta_{z_{i,j}^k}) \in \Gamma_o \left(w_{i,j}\delta_{x_i} + \sum_{k=1}^{n} w_{i,j}\delta_{z_{i,j}^k},\; w_{i,j}\delta_{y_j} + \sum_{k=1}^{n} w_{i,j}\delta_{z_{i,j}^k} \right).$$

Since computing the marginal is a linear operation, we deduce

$$\sum_{(i,j) \in \mathcal{A}} \sum_{k=1}^{1+n} w_{i,j}\, (\delta_{z_{i,j}^{k-1}} \times \delta_{z_{i,j}^k}) \in \Gamma(\mu+\overline{\sigma}, \nu+\overline{\sigma}), \tag{2.10}$$

and the (pth power of the) p-cost associated to this plan is

$$\int_{\Omega\times\Omega} |x-y|^p \, d\left(\sum_{(i,j)\in\mathcal{A}} \sum_{k=1}^{1+n} w_{i,j}\left(\delta_{z_{k,i,j}} \times \delta_{z_{k+1,i,j}}\right)\right)(x,y)$$
$$= \sum_{(i,j)\in\mathcal{A}} w_{i,j} \sum_{k=1}^{1+n} |z_{k,i,j} - z_{k+1,i,j}|^p = \sum_{(i,j)\in\mathcal{A}} w_{i,j} \sum_{k=1}^{1+n} \frac{|x_i - y_j|^p}{(1+n)^p}$$
$$= \frac{1}{(1+n)^{p-1}} \sum_{(i,j)\in\mathcal{A}} w_{i,j}|x_i - x_j|^p = \frac{1}{(1+n)^{p-1}} W_p^p(\mu,\nu).$$

We infer that

$$W_p^p(\mu+\overline{\sigma}, \nu+\overline{\sigma}) \le \frac{1}{(1+n)^{p-1}} W_p^p(\mu,\nu). \tag{2.11}$$

Let us pass to the general case. If μ,ν are two generic measures in $\mathscr{M}_p^\alpha(\Omega)$, let $(\mu_l) \subset \mathscr{M}_p^\alpha(\Omega)$ and $(\nu_l) \subset \mathscr{M}_p^\alpha(\Omega)$ be two sequences of atomic measures with finite supports converging, respectively, to μ and ν in $\mathscr{M}_p^\alpha(\Omega)$. Starting from μ_l and ν_l in place of μ and ν, we may define, for any $l \in \mathbb{N}$, a measure $\overline{\sigma}_l$ exactly as in (2.8) and a plan $\gamma_l \in \Gamma(\mu_l + \overline{\sigma}_l, \nu_l + \overline{\sigma}_l)$ exactly as in (2.10). It is immediate to verify that $(\overline{\sigma}_l)$ is a tight sequence, hence narrowly converging (up to a subsequence, which we do not relabel) to some $\overline{\sigma} \in \mathscr{M}^{n\alpha}(\Omega)$. Indeed, it is enough to repeat the computation (2.9) for the measure $\overline{\sigma}_l$, obtaining

$$\int_\Omega |x|^p \, d\overline{\sigma}_l(x) \le pn \int_\Omega |x|^p \, d(\mu_l + \nu_l)(x),$$

and the quantity in the right-hand side is uniformly bounded with respect to l, since μ_l and ν_l converge in $\mathscr{M}_p^\alpha(\Omega)$. This yields the tightness. Also, the computation for the plan can be repeated, which yields

$$\int_{\Omega\times\Omega} |x-y|^p \, d\gamma_l(x,y) = \frac{1}{(1+n)^{p-1}} W_p^p(\mu_l,\nu_l),$$

and the right-hand side is again bounded, since the Wasserstein distance is continuous with respect to the convergence in $\mathscr{M}_p^\alpha(\Omega)$. The same is then true for any sequence $(\widetilde{\gamma}_l)$ of optimal plans between the same marginals. Since $\mu_l + \overline{\sigma}_l$ and $\nu_l + \overline{\sigma}_l$ are narrowly converging, and since the optimal plans $\widetilde{\gamma}_l$ have uniformly bounded p-cost, by the standard lower semicontinuity results on the Wasserstein distance (see, for instance, [2, Proposition 7.1.3]), we find

$$W_p^p(\mu+\overline{\sigma}, \nu+\overline{\sigma}) \le \liminf_{l\to\infty} W_p^p(\mu_l + \overline{\sigma}_l, \nu_l + \overline{\sigma}_l).$$

But for any fixed l inequality (2.11) holds, and we conclude that

$$W_p^p(\mu+\overline{\sigma}, \nu+\overline{\sigma}) \le \liminf_{l\to\infty} \frac{1}{(1+n)^p} W_p^p(\mu_l,\nu_l) = \frac{1}{(1+n)^p} W_p^p(\mu,\nu),$$

where we have once again made use of the continuity of W_p. □

The above result shows that there is a "wrong scaling" in the p-distance if $p > 1$: by adding more and more mass to the source and to the target, the distance can be made arbitrarily small, as stated in the following straightforward corollary.

Corollary 2.3. *Let $p > 1$. Let $\mu,\nu \in \mathscr{M}_p^\alpha(\Omega)$. Then*

$$\inf\left\{W_p(\mu+\sigma,\nu+\sigma) : \sigma \in \mathscr{M}_p^+(\Omega)\right\} = 0.$$

Proof. Simply let $\sigma_n \in \mathscr{M}_p^{n\alpha}(\Omega)$ for any $n \in \mathbb{N}$. By Theorem 2.2, each σ_n can be chosen so that (2.7) holds, and then $\lim_{n\to\infty} W_p(\mu+\sigma_n, \nu+\sigma_n) = 0$. □

Remark 2.4. The scaling behavior of the optimal transport distance is an interesting question in itself. Some finer results on the asymptotics of W_p (say, when the mass of σ increases) have been recently obtained by G. Wolanski (see [22]). On the other hand, it would be interesting to give a sharp characterization of solutions to the minimization problem

$$\inf_{\sigma\in\mathscr{M}^\alpha(\Omega)} \frac{W_p^p(\mu+\sigma,\nu+\sigma)}{W_p^p(\mu,\nu)},$$

for two given probabilities μ, ν. About this problem we infer that, if μ and ν are two Dirac masses, at least if α is an integer, the minimal value is given by

$$\frac{1}{(1+\alpha)^{p-1}},$$

which is the value coming from the Hölder inequality.

3. The case of signed measures

3.1. The setting. Let $\mathscr{M}(\Omega)$ denote the set of bounded Radon measures over Ω. We also endow $\mathscr{M}(\Omega)$ with the standard narrow convergence, given by the duality with continuous and bounded functions. We recall the Jordan–Hahn decomposition for a real measure μ: μ^+ and μ^- denote the positive and negative part of μ, respectively, so that $\mu = \mu^+ - \mu^-$ (μ^+ and μ^- are two positive, orthogonal measures). Of course, there are many pairs of positive measures whose difference is μ. This decomposition is the minimal one: for any other pair of positive measures σ^1, σ^2 such that $\sigma^1 - \sigma^2 = \mu$, we have $\mu^+ \leq \sigma^1$ and $\mu^- \leq \sigma^2$. Here the notation $\mu \leq \nu$ means that $\mu(A) \leq \nu(A)$ for any Borel set $A \in \Omega$ (μ is a submeasure). Given $\mu \in \mathscr{M}(\Omega)$, the total variation measure is standardly defined as

$$|\mu|(B) := \sup\left\{\sum_{i=1}^{N}|\mu(B_i)|,\ B_i \text{ are pairwise disjoint, } \bigcup_{i=1}^{N} B_i = B,\ N \in \mathbb{N}\right\};$$

$|\mu|$ is a positive measure, equal to $\mu^+ + \mu^-$. The quantity $|\mu|(\Omega)$ will be referred to as the total mass, whereas $\mu(\Omega)$ will be called the total integral.

If we are given two measures $\mu, \nu \in \mathscr{M}(\Omega)$ with the same total mass and the same total integral, the issue of defining a p-Wasserstein distance is trivial. Indeed, in this case we have $\mu^+(\Omega) = \nu^+(\Omega)$ and $\mu^-(\Omega) = \mu^-(\Omega)$, so that we can simply compare the positive parts and the negative parts separately. Hence we are left with the Wasserstein distance in the product space, that is, we can define the Wasserstein distance between μ and ν as

$$\left(W_p^p(\mu^+,\nu^+) + W_p^p(\mu^-,\nu^-)\right)^{1/p}. \tag{3.1}$$

This is indeed the definition used for the minimizing movements scheme in [18]. Alternatively, one could consider $W_p(|\mu|,|\nu|)$.

On the other hand, by analogy with the standard theory of transport, one should define the distance for measures with the same total integral (this accounts for mass conservation in the transportation), but possibly with different total masses. Let us define the following subset of $\mathscr{M}(\Omega)$:

$$\mathscr{M}^{\alpha,\, M}(\Omega) := \{\mu \in \mathscr{M}(\Omega) : \mu(\Omega) = \alpha, |\mu|(\Omega) \leq M\}, \tag{3.2}$$

where $\alpha \in \mathbb{R}$, $M \geq |\alpha|$. In the positive case, the bound on the total mass is implicit in the fixed value of the total integral. Later, we will see that it is fundamental to impose a bound on the total mass. We also define the corresponding space of measures with bounded p-moments, $p \geq 1$:

$$\mathscr{M}_p^{\alpha,\, M}(\Omega) := \left\{\mu \in \mathscr{M}^{\alpha,\, M}(\Omega) : \int_\Omega |x|^p\, d|\mu| < +\infty\right\}.$$

Moreover, we say that a sequence $(\mu_n) \subset \mathscr{M}_p^{\alpha,\, M}(\Omega)$ converges to μ in $\mathscr{M}_p^{\alpha,\, M}(\Omega)$ if μ_n converges narrowly to μ and

$$\int_\Omega |x|^p\, d\mu_n(x) \to \int_\Omega |x|^p\, d\mu(x). \tag{3.3}$$

Note that we do not require the much stronger condition

$$\int_\Omega |x|^p\, d|\mu_n|(x) \to \int_\Omega |x|^p\, d|\mu|(x).$$

Here we come to the real issue: how to define a cost of transportation in $\mathscr{M}_p^{\alpha,\, M}(\Omega)$? Before going into details, we remark that in principle there could be more ways to treat transport strategies in the context of real measures. And different strategies could be suitable for different applications. For instance, one could proceed with one of the following points of view.

- If the total masses are equal, then one can make use of the "product distance" defined by (3.1).

- One could be interested in transporting as much mass as possible, independently of the sign. In this case, one should perform a "partial transport," that is, considering $|\mu|$ and $|\nu|$, one should solve the problem

$$\min\left\{\int\limits_{\Omega\times\Omega}|x-y|^p\,d\gamma : \pi^1_{\#}\gamma\le|\mu|,\ \pi^2_{\#}\gamma\le|\nu|,\ \gamma(\Omega\times\Omega)=\mathfrak{M}\right\}.$$

Here the fixed value of the mass carried by a transport plan γ corresponds to the maximum mass that can be transferred, which is, of course, $\mathfrak{M}=\min\{|\mu|(\Omega),|\nu|(\Omega)\}$. Regarding the optimal partial transportation problem, we refer to the seminal papers [7, 12].
- For dealing with all the given mass, one should allow for cancellation between positive and negative masses.

Since we are interested in a global transportation problem, we will deal with the last instance of the three, following the discussion in [3, Sec. 2]. Next we list the definitions we will present (all consistent with the standard Wasserstein distance when computed on positive measures), paying attention to their main features. In the following, let μ and ν be real measures in $\mathscr{M}^{\alpha,M}_p(\Omega)$.

♢ **The "global" cost**

$$W_p(\mu^+ + \nu^-, \nu^+ + \mu^-).$$

It is symmetric, not narrowly lower semicontinuous, and it does not satisfy the triangle inequality.

♢ **The "relaxed" cost**

$$\begin{aligned}\inf\{W_p(\sigma^1+\theta^2,\theta^1+\sigma^2) : \sigma^1(\Omega)\le M^+_{\mu,\nu},\ \sigma^2(\Omega)\le M^-_{\mu,\nu},\ \sigma^1-\sigma^2=\nu,\\ \theta^1(\Omega)\le M^+_{\mu,\nu},\ \theta^2(\Omega)\le M^-_{\mu,\nu},\ \theta^1-\theta^2=\mu\},\end{aligned}$$

where $M^{\pm}_{\mu,\nu} := \max\{\mu^{\pm}(\Omega),\nu^{\pm}(\Omega)\}$. Compared to the previous one, this cost only gains the narrow lower semicontinuity.

♢ **The "unilateral" cost**

$$\inf\left\{\left(W_p^p(\sigma^1,\mu^+)+W_p^p(\sigma^2,\mu^-)\right)^{1/p} : \sigma^1-\sigma^2=\nu,\ \sigma^1(\Omega)=\mu^+(\Omega),\ \sigma^2(\Omega)=\mu^-(\Omega)\right\}.$$

Still not triangular, still narrowly lower semicontinuous (but only with respect to the argument ν, with μ fixed). It is also nonsymmetric, and suitable for describing targets with less mass than the source: $|\mu|(\Omega)\ge|\nu|(\Omega)$.

3.2. The global cost. Let $\mu,\nu\in\mathscr{M}^{\alpha,M}(\Omega)$. In order to take into account the possible positive/negative interaction, a definition that seems natural is

$$\mathbb{W}_p(\mu,\nu) := W_p(\mu^+ + \nu^-, \nu^+ + \mu^-). \tag{3.4}$$

Note that this is a good definition, since the constraint $\mu(\Omega)=\nu(\Omega)$ gives $\mu^+(\Omega)-\mu^-(\Omega)=\nu^+(\Omega)-\nu^-(\Omega)$, hence $\mu^+(\Omega)+\nu^-(\Omega)=\nu^+(\Omega)+\mu^-(\Omega)$. It is also immediate to check that if μ and ν are nonnegative, then $\mathbb{W}_p$ reduces to the Wasserstein distance between positive measures of a given mass α on Ω. By definition, the value of $\mathbb{W}_p(\mu,\nu)$ corresponds to an optimal transport plan γ in the set

$$\Gamma^p_o(\mu^+ + \nu^-, \nu^+ + \mu^-).$$

A transport plan in $\Gamma(\mu^+ + \nu^-, \nu^+ + \mu^-)$ may be seen as accounting for four transports: 1) a part of μ^+ that goes to ν^+; 2) the other part of μ^+, which goes to μ^-; 3) a part of ν^- that is transported to μ^-; 4) the remaining part of ν^- going to the remaining part of ν^+. Here a part is, of course, a submeasure. We have some remarks about two of these points.

- In order to connect μ to ν, it can be convenient to transport some part of μ^+ onto μ^-, this correspond to auto-annihilation of mass. See Fig. 2 and Fig. 3 a), d).
- On the other hand, if the total mass of ν is larger than that of μ, one expects that, in the transport given by $\mathbb{W}_2$, a nonzero part will come from moving some part of ν^- to ν^+. From the dynamic point of view, this corresponds to some fake zero charge mass which is created and separated into positive and negative mass and then transported at a certain cost. See Fig. 3 b) and d).

The next lemma makes the above discussion rigorous. In order to define plan splitting and subplans, we introduce the following notation.

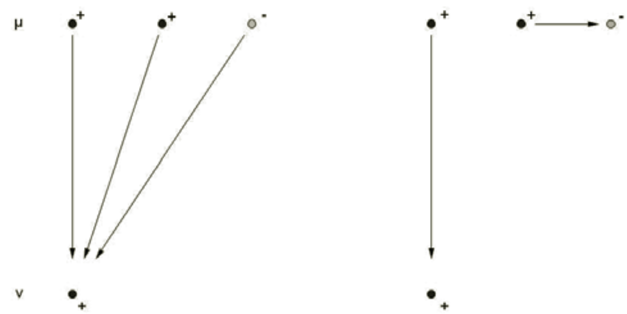

Fig. 2. In the first transport path, all the measure μ is transported to ν, in the same way one would transport three positive deltas to a single point. It is clear that, taking into account the charges of the particles, the second path is more convenient, and corresponds to an optimal plan in $\Gamma_o(\mu^+ + \nu^-, \nu^+ + \mu^-)$.

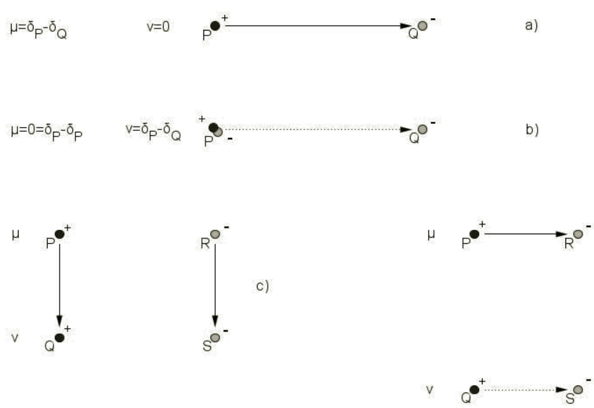

Fig. 3. a) How to transport a positive and a negative delta to the null measure? Simply transport the positive delta to the negative one, so that $\mathbb{W}_p(\delta_P - \delta_Q, 0) = |P - Q|$.
b) This time we are to transport the null mass to $\delta_P - \delta_Q$! Then we need a fake mass. Think of $0 = \delta_P - \delta_P$, leave δ_P there, and transport $-\delta_P$ to $-\delta_Q$. Again $\mathbb{W}_p(\delta_P - \delta_Q, 0) = |P - Q|$.
c) Standard mass transport: neither annihilation nor creation of mass.
d) Annihilation and creation, this is a)+b); $\mathbb{W}_p^p(\delta_P - \delta_R, \delta_Q - \delta_S) = |P - R|^p + |Q - S|^p$.

Definition 3.1 (Transport partition). *Consider partitions of the positive and negative parts of ν and μ of the form*

$$\begin{aligned} \nu_0^+ + \nu_1^+ &= \nu^+, \quad \nu_0^- + \nu_1^- = \nu^-, \\ \mu_0^+ + \mu_1^+ &= \mu^+, \quad \mu_0^- + \mu_1^- = \mu^-, \end{aligned} \tag{3.5}$$

where all the terms are positive measures. We say that a partition of this form is admissible if the following compatibility conditions hold:

$$\nu_0^+(\Omega) = \mu_0^+(\Omega), \quad \mu_0^-(\Omega) = \nu_0^-(\Omega), \quad \mu_1^-(\Omega) = \mu_1^+(\Omega), \quad \nu_1^+(\Omega) = \nu_1^-(\Omega). \tag{3.6}$$

In the above definition, μ_0^+, μ_0^- correspond to the parts that will move to ν_0^+, ν_0^-, respectively, and μ_1^+, μ_1^- (respectively, ν_1^+, ν_1^-), to the self-cancelling parts. Of course, there are many partitions of this kind. By the way, a splitting can be chosen so as to preserve optimality, as shown in the next lemma.

Lemma 3.2 (Plan splitting). *Let $\gamma \in \Gamma(\nu^+ + \mu^-, \mu^+ + \nu^-)$. Then there exists an admissible partition of the form* (3.5) *such that γ can be written as the sum of four plans γ_+^+, γ_-^-, γ_-^+, γ_+^- satisfying*

$$\begin{aligned} \gamma_+^+ &\in \Gamma(\nu_0^+, \mu_0^+), \quad \gamma_-^- \in \Gamma(\mu_0^-, \nu_0^-), \\ \gamma_-^+ &\in \Gamma(\mu_1^-, \mu_1^+), \quad \gamma_+^- \in \Gamma(\nu_1^+, \nu_1^-). \end{aligned} \tag{3.7}$$

Moreover, if γ is optimal, then the four plans (3.7) *can be chosen to be optimal as well.*

Proof. Let $\vartheta_1 = \nu^+ + \mu^-$ and $\vartheta_2 = \mu^+ + \nu^-$. It is clear that ν^+ and μ^- are both absolutely continuous with respect to ϑ_1. Let $f_1, g_1 \in L^1(\Omega, \vartheta_1)$ denote the respective densities. Similarly, let f_2, g_2 be the densities of ν^- and μ^+ with respect to ϑ_2, so that

$$\nu^+ = f_1\vartheta_1, \quad \mu^- = g_1\vartheta_1, \quad \mu^+ = g_2\vartheta_2, \quad \nu^- = f_2\vartheta_2.$$

Clearly, $f_1 + g_1 = f_2 + g_2 = 1$, so that we can write

$$\gamma = (f_1 \circ \pi^1)(g_2 \circ \pi^2)\gamma + (f_1 \circ \pi^1)(f_2 \circ \pi^2)\gamma + (g_1 \circ \pi^1)(g_2 \circ \pi^2)\gamma + (g_1 \circ \pi^1)(f_2 \circ \pi^2)\gamma. \tag{3.8}$$

Then we define the four desired plans as

$$\begin{aligned} \gamma_+^+ &:= (f_1 \circ \pi^1)(g_2 \circ \pi^2)\gamma, \quad \gamma_+^- := (f_1 \circ \pi^1)(f_2 \circ \pi^2)\gamma, \\ \gamma_-^+ &:= (g_1 \circ \pi^1)(g_2 \circ \pi^2)\gamma, \quad \gamma_-^- := (g_1 \circ \pi^1)(f_2 \circ \pi^2)\gamma, \end{aligned} \tag{3.9}$$

and we claim that this is a consistent definition. To prove the claim, let us analyze the marginals of these four plans, recalling the elementary equality $\pi^i_\#((\varphi \circ \pi^i)\gamma) = \varphi\pi^i_\#\gamma$ holding for a density $\varphi : \Omega \to \mathbb{R}$. For the first one, we have

$$\begin{aligned} \pi^1_\# \big((f_1 \circ \pi^1)(g_2 \circ \pi^2)\gamma\big) &= f_1\, \pi^1_\# \big((g_2 \circ \pi^2)\gamma\big) \le f_1\, \pi^1_\#\gamma = f_1\vartheta_1 = \nu^+, \\ \pi^2_\# \big((f_1 \circ \pi^1)(g_2 \circ \pi^2)\gamma\big) &= g_2\, \pi^2_\# \big((f_1 \circ \pi^1)\gamma\big) \le g_2\, \pi^2_\#\gamma = g_2\vartheta_2 = \mu^+, \end{aligned}$$

where we use the fact that the densities f_1, f_2, g_1, g_2 are less than or equal to 1. This shows that the first and the second marginals of γ_+^+ are nonnegative submeasures of ν^+ and μ^+, respectively. Analogously,

$$\begin{aligned} \pi^1_\#\gamma_+^- &= f_1\, \pi^1_\# \big((f_2 \circ \pi^2)\gamma\big) \le f_1\, \pi^1_\#\gamma = f_1\vartheta_1 = \nu^+, \\ \pi^2_\#\gamma_+^- &= f_2\, \pi^2_\# \big((f_1 \circ \pi^1)\gamma\big) \le f_2\, \pi^2_\#\gamma = f_2\vartheta_2 = \nu^-, \\ \pi^1_\#\gamma_-^+ &= g_1\, \pi^1_\# \big((g_2 \circ \pi^2)\gamma\big) \le g_1\, \pi^1_\#\gamma = g_1\vartheta_1 = \mu^-, \\ \pi^2_\#\gamma_-^+ &= g_2\, \pi^2_\# \big((g_1 \circ \pi^1)\gamma\big) \le g_2\, \pi^2_\#\gamma = g_2\vartheta_2 = \mu^+, \\ \pi^1_\#\gamma_-^- &= g_1\, \pi^1_\# \big((f_2 \circ \pi^2)\gamma\big) \le g_1\, \pi^1_\#\gamma = g_1\vartheta_1 = \mu^-, \\ \pi^2_\#\gamma_-^- &= f_2\, \pi^2_\# \big((g_1 \circ \pi^1)\gamma\big) \le f_2\, \pi^2_\#\gamma = f_2\vartheta_2 = \nu^-. \end{aligned}$$

From these relations we see that the marginals of the other three plans are also submeasures of the positive and negative parts of ν and μ, as required. The only thing left is to check that these marginals form an admissible partition of ν and μ as in (3.5), (3.6). But notice that for $\gamma_+^+ + \gamma_+^-$ we have

$$\begin{aligned} \pi^1_\#(\gamma_+^+ + \gamma_+^-) &= \pi^1_\#((f_1 \circ \pi^1)(g_2 \circ \pi^2)\gamma) + \pi^1_\#((f_1 \circ \pi^1)(f_2 \circ \pi^2)\gamma) \\ &= f_1\, \pi^1_\#((g_2 \circ \pi^2 + f_2 \circ \pi^2)\gamma) = f_1\, \pi^1_\#\gamma = f_1\vartheta_1 = \nu^+. \end{aligned}$$

In the same way,

$$\begin{aligned} \pi^1_\#(\gamma_-^+ + \gamma_-^-) &= g_1\pi^1_\#((g_2 \circ \pi^2 + f_2 \circ \pi^2)\gamma) = g_1\pi^1_\#\gamma = \mu^-, \\ \pi^2_\#(\gamma_+^- + \gamma_-^-) &= f_2\pi^2_\#((f_1 \circ \pi^1 + g_1 \circ \pi^1)\gamma) = f_2\pi^2\gamma = \nu^-, \\ \pi^2_\#(\gamma_-^+ + \gamma_+^+) &= g_2\pi^2_\#((f_1 \circ \pi^1 + g_1 \circ \pi^1)\gamma) = g_2\pi^2_\#\gamma = \mu^+. \end{aligned}$$

We see that the marginals of the four plans do satisfy (3.5), whereas relations (3.6) trivially hold true. Hence the claim follows: by (3.9) we have indeed defined a splitting of the desired form. Finally, if γ is optimal, each of these plans is optimal as well, since their sum is. ☐

Remark 3.3. By the previous result, the cost $\mathbb{W}_p(\mu, \nu)$ can also be written as

$$\mathbb{W}_p(\nu, \mu) = \inf \left(W_p^p(\nu_0^+, \mu_0^+) + W_p^p(\nu_1^+, \nu_1^-) + W_p^p(\mu_0^-, \nu_0^-) + W_p^p(\mu_1^-, \mu_1^+) \right)^{1/p},$$

where the infimum is taken among all admissible partitions of the form (3.5).

Plan splitting according to the cost $\mathbb{W}_p$ and the above notation is sketched in Fig. 4.

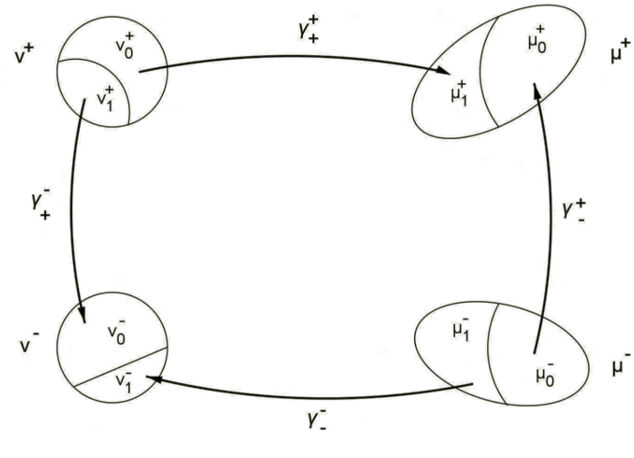

Fig. 4. Plan splitting according to Lemma 3.2.

The next step is to analyze the topological properties of the cost $\mathbb{W}_p$. We will see that many properties of the original Wasserstein distance are lost.

Proposition 3.4. *The cost $\mathbb{W}_p$ is symmetric and vanishes if and only if $\mu = \nu$. However, $\mathbb{W}_p$ is not a distance on $\mathscr{M}_p^{\alpha, M}(\Omega)$, unless $p = 1$. Besides,*

$$\mathbb{W}_p(\mu,\nu) \geq \left(\frac{1}{2M}\right)^{(p-1)/p} \mathbb{W}_1(\mu,\nu). \tag{3.10}$$

Proof. Since W_p is a distance, $W_p(\mu^+ + \nu^-, \nu^+ + \mu^-)$ vanishes if and only if $\mu^+ + \nu^- = \nu^+ + \mu^-$, which is equivalent to $\mu = \nu$. The symmetry is obvious. The following example shows that the triangle inequality fails for $p > 1$. It is enough to work on the real line: let $\mu = \delta_0$, $\nu = \delta_4$, and $\eta = \delta_1 - \delta_2 + \delta_3$. Clearly, $\mathbb{W}_2(\mu,\nu) = W_2(\mu,\nu) = 4$. But the optimal transport plan between $\mu^+ + \eta^-$ and $\eta^+ + \mu^-$ is $\delta_0 \times \delta_1 + \delta_2 \times \delta_3$, so that

$$\mathbb{W}_p^p(\mu,\eta) = \int_{\mathbb{R}} |x-y|^p \, d(\delta_0 \times \delta_1) + \int_{\mathbb{R}} |x-y|^p \, d(\delta_2 \times \delta_3) = 2.$$

Symmetrically, $\mathbb{W}_p(\nu,\eta) = \sqrt[p]{2}$, so that

$$\mathbb{W}_p(\mu,\nu) > \mathbb{W}_p(\mu,\eta) + \mathbb{W}_p(\nu,\eta).$$

On the other hand, we notice that if $\gamma \in \Gamma_o^p(\mu^+ + \nu^-, \nu^+ + \mu^-)$, then, by the Hölder inequality, we have

$$\mathbb{W}_1(\mu,\nu) \leq \int_{\Omega\times\Omega} |x-y| \, d\gamma \leq \gamma(\Omega\times\Omega)^{(p-1)/p} \left(\int_{\Omega\times\Omega} |x-y|^p \, d\gamma \right)^{1/p} \leq (2M)^{(p-1)/p} \, \mathbb{W}_p(\mu,\nu),$$

which is (3.10).

Finally, we show that

$$\mathbb{W}_1(\mu,\nu) := W_1(\mu^+ + \nu^-, \nu^+ + \mu^-) = \inf_{\gamma\in\Gamma(\mu^+ + \nu^-, \nu^+ + \mu^-)} \int_{\Omega\times\Omega} |x-y| \, d\gamma \tag{3.11}$$

is indeed a distance between signed measures. This can be seen by the formula (2.6), which gives

$$\begin{aligned} W_1(\mu^+ + \nu^-, \nu^+ + \mu^-) &= \sup_{\varphi \in \mathrm{Lip}(\Omega), \|\varphi\|_{\mathrm{Lip}} \le 1} \int_\Omega \varphi \, d((\mu^+ + \nu^-) - (\nu^+ + \mu^-)) \\ &= \sup_{\varphi \in \mathrm{Lip}(\Omega), \|\varphi\|_{\mathrm{Lip}} \le 1} \int_\Omega \varphi \, d(\mu - \nu). \end{aligned} \tag{3.12}$$

Let $\mu, \nu, \eta \in \mathscr{M}_1^{\alpha, M}(\Omega)$. We have

$$\begin{aligned} \mathbb{W}_1(\mu, \eta) + \mathbb{W}_1(\eta, \nu) &= \sup_{\varphi \in \mathrm{Lip}(\Omega), \|\varphi\|_{\mathrm{Lip}} \le 1} \int_\Omega \varphi \, d(\mu - \eta) + \sup_{\varphi \in \mathrm{Lip}(\Omega), \|\varphi\|_{\mathrm{Lip}} \le 1} \int_\Omega \varphi \, d(\eta - \nu) \\ &\ge \sup_{\varphi \in \mathrm{Lip}(\Omega), \|\varphi\|_{\mathrm{Lip}} \le 1} \left(\int_\Omega \varphi \, d(\mu - \eta) + \int_\Omega \varphi \, d(\eta - \nu) \right) \\ &= \sup_{\varphi \in \mathrm{Lip}(\Omega), \|\varphi\|_{\mathrm{Lip}} \le 1} \int_\Omega \varphi \, d(\mu - \nu) = \mathbb{W}_1(\mu, \nu), \end{aligned}$$

so that the triangle inequality holds. □

Remark 3.5. We stress again that $\mathbb{W}_1$ is not sensitive to the addition of equal masses to the source and to the target, and this is the key fact for showing that $\mathbb{W}_1$ is a distance. This fails for $p > 1$: indeed, since for the triangle inequality in $\mathscr{M}^{\alpha, M}(\Omega)$ we need to compare measures with possibly different masses, the bad scaling behavior for a strictly convex cost ($p > 1$) discussed in Sec. 2.3 causes the functional $\mathbb{W}_p$ to violate the triangle inequality. However, the bound on the total mass allows one to establish (3.10), that is, $\mathbb{W}_p$ is bounded from below by a nontrivial distance. This is an important estimate. For instance, it is a key ingredient for the convergence of the approximation scheme (1.3). See [3].

Remark 3.6. Without a bound on the total mass, in the case $p > 1$ the $\mathbb{W}_p$ cost can be made arbitrarily small, in the same spirit as in Corollary 2.3. Indeed, let, for simplicity, $\mu = \delta_0 - \delta_1$. Let $n \in \mathbb{N}$ be odd, and define a measure $\nu_n \in \mathscr{M}_p^{0, n-1}(\mathbb{R})$ by

$$\nu_n = \sum_{j=1}^{n-1} (-1)^{j+1} \delta_{j/n}.$$

Then

$$\sum_{j=0}^{(n-1)/2} \delta_{2j/n} \times \delta_{(2j+1)/n} \in \Gamma_o(\mu^+ + \nu_n^-, \nu_n^+ + \mu^-)$$

and

$$\mathbb{W}_p(\mu, \nu_n) = \left(\sum_{j=0}^{(n-1)/2} \frac{1}{n^p} \right)^{1/p} = \left(\frac{n+1}{2n^p} \right)^{1/p}.$$

If $p > 1$, it is clear that letting $n \to \infty$ yields $|\nu_n|(\mathbb{R}) \to \infty$ and $\mathbb{W}_p(\mu, \nu_n) \to 0$.

Though not a distance, as we will see in the next two propositions, $\mathbb{W}_p$ has some "metrizability" properties for the $\mathscr{M}_p^{\alpha, M}(\Omega)$ topology. First of all, we should underline that given a sequence $(\mu_n) \subset \mathscr{M}_p^{\alpha, M}(\Omega)$ and a measure $\mu \in \mathscr{M}_p^{\alpha, M}(\Omega)$, the uniform bound $\sup_n \mathbb{W}_p(\mu_n, \mu) < +\infty$ does not imply the uniform boundedness for the pth moments of (μ_n). That is, the $\mathbb{W}_p$-boundedness of a set is not equivalent to the uniform boundedness of its pth moments, in clear contrast with the case of the standard Wasserstein distance. For instance, consider the sequence $\mu_n := \delta_n - \delta_{n+\frac{1}{n}}$, which is even not tight but $\mathbb{W}_p(\mu_n, 0)$ converges to 0. For this reason, in the next proposition we have to explicitly restrict ourselves to sets of measures with uniformly bounded pth moments. Another possibility is to consider the case of a compact metric space Ω, yielding automatically the bound on moments.

Proposition 3.7. *Let μ_n, μ belong to $\mathscr{M}_p^{\alpha, M}(\Omega)$. Let $\sup_n \int_\Omega |x|^p \, d|\mu_n| < +\infty$. Then μ_n converge to μ in $\mathscr{M}_p^{\alpha, M}(\Omega)$ if $\mathbb{W}_p(\mu_n, \mu) \to 0$.*

Proof. Assume that $\mathbb{W}_p(\mu_n,\mu)\to 0$, that is, $W_p(\mu_n^+ + \mu^-, \mu^+ + \mu_n^-)\to 0$. Note that (μ_n^+) and (μ_n^-) are tight sequences, thanks to the bound on the pth moments. Then let σ_1, σ_2 be the narrow limits along subsequences $(\mu_{n_k}^+)$, $(\mu_{n_k}^-)$, respectively, and let $\sigma := \sigma_1 - \sigma_2$. By the semicontinuity of the standard Wasserstein distance with respect to the narrow topology, and thanks to Proposition 2.1 and (3.10), we have

$$\mathbb{W}_1(\sigma,\mu) = W_1(\sigma_1+\mu^-, \mu^+ + \sigma_2) \le \liminf_k W_1(\mu_{n_k}^+ + \mu^-, \mu^+ + \mu_{n_k}^-)$$
$$\le \left(\frac{1}{2M}\right)^{(p-1)/p} \liminf_k W_p(\mu_{n_k}^+ + \mu^-, \mu^+ + \mu_{n_k}^-) = 0,$$

whence $\sigma=\mu$. Since the selected subsequence was arbitrary, we get the narrow convergence to μ for the whole sequence (μ_n). In order to get the convergence of the pth moments, we let $\gamma_n \in \Gamma_o(\mu_n^+ + \mu^-, \mu^+ + \mu_n^-)$, and, by the Young and triangle inequality, we deduce

$$\int_{\Omega\times\Omega} |x|^p\, d\gamma_n \le (1+\varepsilon)\int_{\Omega\times\Omega} |y|^p\, d\gamma_n + \left(1+\frac{1}{\varepsilon}\right)\int_{\Omega\times\Omega} |x-y|^p\, d\gamma_n,$$

that is to say,

$$\int_{\Omega} |x|^p\, d(\mu_n - \mu) \le \varepsilon \int_{\Omega} |x|^p\, d(\mu_n^- + \mu^+) + \left(1+\frac{1}{\varepsilon}\right)\mathbb{W}_p(\mu_n,\mu).$$

Taking the limit as $n\to\infty$, using the uniform bound on the pth moments, and then the arbitrariness of ε, we conclude that the left-hand side goes to zero. □

Proposition 3.8. *Let μ_n, μ belong to $\mathscr{M}_p^{\alpha,M}(\Omega)$, and let $\mu_n\to\mu$ in $\mathscr{M}_p^{\alpha,M}(\Omega)$. Assume that any subsequence of (μ_n^+) and (μ_n^-) has narrow limit points with also the convergence of the corresponding pth moments (this is the case, for instance, if Ω is a compact metric space). Then $\mathbb{W}_p(\mu_n,\mu)\to 0$.*

Proof. Let $(\mu_{n_k}^+)$ be a suitable subsequence of (μ_n^+) such that $\mu_{n_k}^+\to\sigma_1$ narrowly and the corresponding pth moments converge. Therefore $\mu_{n_k}^-\to\sigma_2$ narrowly, and the pth moments also converge, where $\sigma_2=\sigma_1-\mu$. Let also $\widetilde{\sigma} := \sigma_1 - \mu^+$. By the continuity of the standard Wasserstein distance, we have

$$W_p(\mu_{n_k}^+ + \mu^-, \mu^+ + \mu_{n_k}^-) \to W_p(\sigma_1 + \mu^-, \mu^+ + \sigma_2) = W_p(\mu^+ + \mu^- + \widetilde{\sigma}, \mu^+ + \mu^- + \widetilde{\sigma}) = 0,$$

which gives the desired result. □

On the other hand, a property failing for $\mathbb{W}_p$ is the semicontinuity in the narrow topology, as a further consequence of the bad scaling.

Proposition 3.9. *Let $p>1$ and $\mu\in\mathscr{M}_p^{\alpha,M}(\Omega)$. The map $\nu\mapsto\mathbb{W}_p(\nu,\mu)$ is not narrowly lower semicontinuous.*

Proof. A counterexample on the real line is again sufficient. Let $\mu=\delta_{-1}-\delta_1$ and $\nu_n=\delta_{-1/n}-\delta_{1/n}$, so that (ν_n) narrowly converges to $\nu=0$. Clearly, $\mathbb{W}_p(\nu^+ + \mu^-, \mu^+ + \nu^-) = W_p(\mu^-,\mu^+) = 2$. But

$$\liminf_{n\to\infty} \mathbb{W}_p(\nu_n^+ + \mu^-, \mu^+ + \nu_n^-) = \liminf_{n\to\infty} \sqrt[p]{2}\,\frac{n-1}{n} = \sqrt[p]{2}.$$

The point is that (ν_n) narrowly converges to ν, (ν_n^+) and (ν_n^-) are tight, but their limits do not in general coincide with ν^+ and ν^- (in this example, they are not zero). □

3.3. The "relaxed" cost. Let us consider the first variant of $\mathbb{W}_p$. As usual, μ,ν are two measures in $\mathscr{M}_p^{\alpha,M}(\Omega)$. In order to overcome the lack of semicontinuity of the map $\nu\mapsto\mathbb{W}_p(\nu,\mu)$, we may define a relaxation, that is,

$$\widetilde{\mathbb{W}_p}^-(\nu,\mu) := \inf_{\substack{\nu_n^+(\Omega)\le\max\{\mu^+(\Omega),\nu^+(\Omega)\}\\ \nu_n^-(\Omega)\le\max\{\mu^-(\Omega),\nu^-(\Omega)\}}} \left\{ \liminf_{n\to\infty}\mathbb{W}_p(\nu_n,\mu) : \nu_n \rightharpoonup \nu,\ \sup_{n\in\mathbb{N}}\int_\Omega |x|^p\, d|\nu_n| < +\infty \right\}. \tag{3.13}$$

By the tightness (ensured by the bounds on the pth moments), again we have subsequences such that $\nu_{n_k}^+\rightharpoonup\sigma^1$ and $\nu_{n_k}^-\rightharpoonup\sigma^2$, with $\sigma^1(\Omega)\le M$, $\sigma^2(\Omega)\le M$, and $\sigma^1-\sigma^2=\nu$. Here σ^1 and σ^2 are not the positive and negative parts of ν, but simply two measures such that $\sigma^1-\sigma^2=\nu$ (a nonminimal decomposition). Hence we can write this kind of lower semicontinuous envelope as

$$\widetilde{\mathbb{W}_p}^-(\nu,\mu) = \inf_{\substack{\sigma^1(\Omega)\le M_{\mu,\nu}^+\\ \sigma^2(\Omega)\le M_{\mu,\nu}^-}} \left\{ W_p(\sigma^1+\mu^-, \mu^+ + \sigma^2) : \sigma^1,\sigma^2\in\mathscr{M}_p^+(\Omega),\ \sigma^1-\sigma^2=\nu \right\}, \tag{3.14}$$

where $M^+_{\mu,\nu} = \max\{\mu^+(\Omega), \nu^+(\Omega)\}$ and $M^-_{\mu,\nu} = \max\{\mu^-(\Omega), \nu^-(\Omega)\}$. Note that the bounds on $\sigma^1(\Omega)$ and $\sigma^2(\Omega)$ prevent the envelope from being identically zero. These bounds cannot be chosen to be simply M, since we would define a cost depending on M itself. By the continuity properties of W_p, the infimum above is attained: the functional does not have narrowly compact sublevels, because of Proposition 2.1, but one can show that minimizing sequences do have uniformly bounded pth moments. The bound on moments can be omitted in the definition. Finally, by construction, the map $\nu \mapsto \widetilde{\mathbb{W}_p}^-(\nu,\mu)$ is narrowly lower semicontinuous, and, of course, $\widetilde{\mathbb{W}_p}^- \leq \mathbb{W}_p$. For instance, we can compute $\widetilde{\mathbb{W}_p}^-$ for the case of Proposition 3.9. We have

$$\widetilde{\mathbb{W}_p}^-(\delta_{-1}-\delta_1,0) = \inf\{W_p(\delta_{-1}+\sigma^1,\delta_1+\sigma^2) : \sigma^1(\Omega) \leq 1, \sigma^2=\sigma^1\} = \inf_{\sigma(\Omega)\leq 1} W_p(\delta_{-1}+\sigma,\delta_1+\sigma).$$

Here the infimum has to be computed over positive measures with mass less than or equal to 1. Hence, to compute $\mathbb{W}_p$, we have to solve a mass scaling problem like the one introduced in Remark 2.4. It is trivial to show that the infimum can be equivalently taken over measures with mass equal to 1 and that a solution for this particular case is $\sigma = \delta_0$. Correspondingly, we have $\mathbb{W}^-_p(\delta_{-1}-\delta_1,0) = \sqrt[p]{2}$, as expected.

We gave a definition like (3.14), because we were concerned with the map $\nu \mapsto \mathbb{W}_p(\nu,\mu)$, hence we only cared about semicontinuity with respect to one of the arguments. Therefore, we may define a more appropriate, symmetric object as follows:

$$\mathbb{W}^-_p(\mu,\nu) := \inf\Big\{W_p(\sigma^1+\theta^2,\theta^1+\sigma^2) : \sigma^1(\Omega) \leq M^+_{\mu,\nu},\ \sigma^2(\Omega) \leq M^-_{\mu,\nu},\ \sigma^1-\sigma^2=\nu,$$
$$\theta^1(\Omega) \leq M^+_{\mu,\nu},\ \theta^2(\Omega) \leq M^-_{\mu,\nu},\ \theta^1-\theta^2=\mu\Big\}.$$

This is the actual form of the relaxed cost. However, we point out that, even after relaxing, the cost is still not triangular. The same counterexample as in Proposition 3.4 works. Indeed, let $\mu,\nu,\eta \in \mathscr{M}^{1,3}_p(\Omega)$ be as in that example. To compute $\mathbb{W}^-_p(\mu,\nu)$, observe that the bounds $\theta^1(\Omega) \leq M^+_{\mu,\nu}$ and $\sigma^1(\Omega) \leq M^-_{\mu,\nu}$ imply $\theta^1 = \mu^+$, $\sigma^1 = \nu^+$, $\theta^2 = \mu^-$, $\sigma^2 = \nu^-$. Then we have $\mathbb{W}^-_p(\delta_0,\delta_4) = W_p(\delta_0,\delta_4) = 4$. On the other hand, the obvious inequality $\mathbb{W}^-_p \leq \mathbb{W}_p$ entails $\mathbb{W}^-_p(\mu,\eta) \leq 2^{1/p}$ and $\mathbb{W}_p(\eta,\nu) \leq 2^{1/p}$.

3.4. The "unilateral" cost. We have seen that the $\mathbb{W}_p$ cost accounts for cancellation of mass in the source and cancellation/creation of mass in the target. Now assume that we want to describe a phenomenon in which only one of the two processes occurs. That is, we want to allow, for instance, only cancellations in the source. We are going to see how to construct a suitable cost, also preserving the narrow semicontinuity property of $\mathbb{W}^-_p$. Having cancellations only within the source, we expect to lose also the symmetry of the cost, and it is reasonable to assume that the target always has less mass.

Let $\mu,\nu \in \mathscr{M}^{\alpha,M}_p(\Omega)$, with $|\nu|(\Omega) \leq |\mu|(\Omega)$. Define

$$\mathcal{W}^p_p(\nu,\mu) := \inf\Big\{W^p_p(\sigma^1,\mu^+) + W^p_p(\sigma^2,\mu^-) \,:\, \sigma^1-\sigma^2=\nu, \sigma^1(\Omega)=\mu^+(\Omega),\ \sigma^2(\Omega)=\mu^-(\Omega)\Big\}.$$

Since any weak limit point of ν^+_n, ν^-_n is a pair of positive measures σ^1,σ^2 satisfying $\sigma^1-\sigma^2=\nu$, $\mathcal{W}^p_p(\nu,\mu)$ can also be written as

$$\inf\Big\{\liminf_{n\to\infty}\left(W^p_p(\nu^+_n,\mu^+) + W^p_p(\nu^-_n,\mu^-)\right) \,:\, \nu_n \rightharpoonup \nu,\ \nu^+_n(\Omega)=\mu^+(\Omega),\ \nu^-_n(\Omega)=\mu^-(\Omega)\Big\}.$$

This way, it is clear that the map $\nu \mapsto \mathcal{W}_p(\nu,\cdot)$ is narrowly lower semicontinuous. The tightness of minimizing sequences and the semicontinuity of the standard Wasserstein distance also show that there exists an optimal pair ϑ^+,ϑ^- such that

$$\mathcal{W}^p_p(\nu,\mu) = W^p_p(\vartheta^+,\mu^+) + W^p_p(\vartheta^-,\mu^-), \tag{3.15}$$

where $\vartheta^+ - \vartheta^- = \nu$. We let $\widetilde{\vartheta}$ denote the common part of ϑ^+ and ϑ^-, so that $\vartheta^+ = \nu^+ + \widetilde{\vartheta}$ and $\vartheta^- = \nu^- + \widetilde{\vartheta}$.

Remark 3.10. In contrast to the case of $\mathbb{W}_p$, which we have already discussed, we stress that a uniform bound on $\mathcal{W}_p(\nu_n,\mu)$ does imply the uniform boundedness of the pth moments for the sequence (ν_n), as can be seen, for instance, from (3.15).

Let us discuss the other properties. First of all, $\mathcal{W}_p$ is not a distance. Indeed, it is not symmetric. Moreover, one can show that it does not satisfy the triangle inequality, a counterexample can be easily constructed as for the case of $\mathbb{W}_p$. Let us analyze plan splitting (see also Fig. 5).

Proposition 3.11. *Let $p \geq 1$ and $\mu, \nu \in \mathscr{M}_p^{\alpha, M}(\Omega)$. Let $\gamma^+ \in \Gamma_0(\vartheta^+, \mu^+)$ and $\gamma^- \in \Gamma_0(\vartheta^-, \mu^-)$ be two optimal transport plans corresponding to the Wasserstein distances in the right-hand side of* (3.15)*. Then we can write these plans as*

$$\gamma^+ = \gamma_0^+ + \gamma_1^+ \quad \textit{and} \quad \gamma^- = \gamma_0^- + \gamma_1^-,$$

where

$$\gamma_0^+ \in \Gamma_0(\widetilde{\vartheta}, \mu_0^+), \quad \gamma_1^+ \in \Gamma_0(\nu^+, \mu_1^+), \quad \gamma_0^- \in \Gamma_0(\widetilde{\vartheta}, \mu_0^-), \quad \gamma_1^- \in \Gamma_0(\nu^-, \mu_1^-), \tag{3.16}$$

and $\mu_0^+ + \mu_1^+ = \mu^+$ and $\mu_0^- + \mu_1^- = \mu^-$.

Proof. We have two plans to split. Let us consider γ^+. In the same spirit as in Lemma 3.2, let f_1 be the density of ν^+ with respect to $\widetilde{\vartheta} + \nu^+$ and f_0 be the density of $\widetilde{\vartheta}$ with respect to $\widetilde{\vartheta} + \nu^+$, so that $f_1 \leq 1$, $f_0 \leq 1$, and $f_1 + f_0 = 1$. We may define

$$\gamma_0^+ := (f_0 \circ \pi^1)\gamma^+, \quad \gamma_1^+ := (f_1 \circ \pi^1)\gamma^+.$$

Indeed, the sum of these two plans is γ, and we have

$$\pi^1_\# \gamma_0^+ = f_0 \pi^1_\# \gamma^+ = \widetilde{\vartheta}, \quad \pi^1_\# \gamma_1^+ = f_1 \pi^1_\# \gamma^+ = \nu^+,$$

and

$$\pi^2_\# \gamma_0^+ + \pi^2_\# \gamma_1^+ = \pi^2_\# ((f_0 \circ \pi^1)\gamma^+ + (f_1 \circ \pi^1)\gamma^+) = \pi^2_\# \gamma^+ = \mu^+,$$

so that the second marginals of γ_0^+ and γ_1^+ are indeed two positive submeasures μ_0^+ and μ_1^+ of μ^+ whose sum is μ^+ itself. The optimality of γ_0^+ and γ_1^+ follows from the optimality of their sum. One proceeds in the same way for splitting γ^-. □

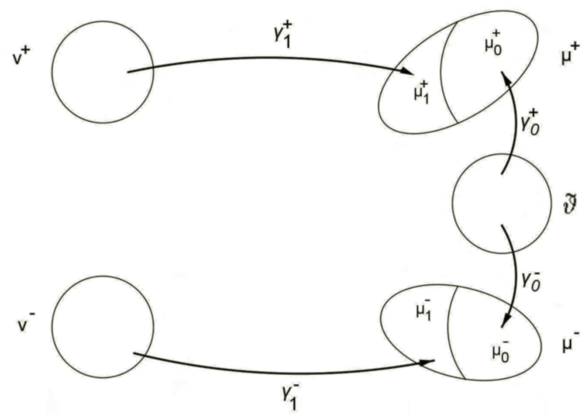

Fig. 5. Plan splitting according to Proposition 3.11.

In this case, we want to give some more information on the optimal splitting. Hence we perform the following first variation argument.

Proposition 3.12. *Let $\mu, \nu \in \mathscr{M}_p^{\alpha, M}(\Omega)$. Let ϑ^+, ϑ^- be a solution of the minimization problem defining $\mathcal{W}_p(\nu, \mu)$, and let $\widetilde{\vartheta}$ be the common part of these measures: $\vartheta^+ = \widetilde{\vartheta} + \nu^+$, $\vartheta^- = \widetilde{\vartheta} + \nu^-$. With the same splitting notation as in the previous lemma (see* Fig. 5*),*

$$\pi^1_\# ((x - y)\gamma_0^+) + \pi^1_\# ((x - y)\gamma_0^-) = 0.$$

Proof. Let us define, for $\varepsilon > 0$, the competitor

$$\vartheta_\varepsilon^+ := \nu^+ + (\mathbf{1} + \varepsilon \boldsymbol{\xi})_\# \widetilde{\vartheta}, \quad \vartheta_\varepsilon^- := \nu^- + (\mathbf{1} + \varepsilon \boldsymbol{\xi})_\# \widetilde{\vartheta},$$

where $\boldsymbol{\xi} : \Omega \to \Omega$ is a bounded vector field with bounded support. It is immediate to verify, computing the marginals, that

$$\gamma_1^+ + (\mathbf{1} + \varepsilon \boldsymbol{\xi}, \mathbf{1})_\# \gamma_0^+ \in \Gamma(\vartheta_\varepsilon^+, \mu^+) \quad \text{and} \quad \gamma_1^- + (\mathbf{1} + \varepsilon \boldsymbol{\xi}, \mathbf{1})_\# \gamma_0^- \in \Gamma(\vartheta_\varepsilon^-, \mu^-).$$

Therefore,

$$
\begin{aligned}
W_p^p(\vartheta_\varepsilon^+,\mu^+)+W_p^p(\vartheta_\varepsilon^-+\mu^-) &\le \int\limits_{\Omega\times\Omega} |x-y|^p\, d\left(\gamma_1^+ + (\mathbf{1}+\varepsilon\boldsymbol{\xi},\mathbf{1})_\#\gamma_0^+ + \gamma_1^- + (\mathbf{1}+\varepsilon\boldsymbol{\xi},\mathbf{1})_\#\gamma_0^-\right)\\
&\le \int\limits_{\Omega\times\Omega} |x-y|^p\, d(\gamma_1^+ + \gamma_1^-) + \int\limits_{\Omega\times\Omega} |x-y+\varepsilon\boldsymbol{\xi}(x)|^p\, d(\gamma_0^+ + \gamma_0^-)\\
&= \mathcal{W}_p^p(\nu,\mu) + 2\varepsilon \int\limits_{\Omega\times\Omega} \langle x-y,\boldsymbol{\xi}(x)\rangle\, d(\gamma_0^+ + \gamma_0^-) + o(\varepsilon).
\end{aligned}
$$

Since $W_p^p(\vartheta_\varepsilon^+,\mu^+)+W_p^p(\vartheta_\varepsilon^-+\mu^-) \ge \mathcal{W}_p^p(\nu,\mu)$, we obtain

$$
2\varepsilon \int\limits_{\Omega\times\Omega} \langle x-y,\boldsymbol{\xi}(x)\rangle\, d(\gamma_0^+ + \gamma_0^-) + o(\varepsilon) \ge 0.
$$

But $\boldsymbol{\xi}$ is arbitrary, so that in fact we have an equality, after dividing by ε and letting ε go to 0. We obtain

$$
\int\limits_{\Omega\times\Omega} \langle x-y,\boldsymbol{\xi}(x)\rangle\, d(\gamma_0^+ + \gamma_0^-) = 0
$$

for any $\boldsymbol{\xi}$. Again the arbitrariness of $\boldsymbol{\xi}$ gives

$$
\pi^1_\#((x-y)(\gamma_0^+ + \gamma_0^-)) = 0,
$$

where $(x-y)(\gamma_0^+ + \gamma_0^-)$ is a vector measure (with values in Ω). □

Remark 3.13. The latter proposition tells us that the optimal auxiliary measure $\widetilde{\vartheta}$ (the common part of ϑ^+ and ϑ^-) is placed somehow in the middle of μ^+ and μ^-. For instance, if $A,B\in\Omega$, $\mu^+=\delta_A$, and $\mu^-=\delta_B$, we have $\gamma_0^+ = \widetilde{\vartheta}\times\delta_A$ and $\gamma_0^- = \widetilde{\vartheta}\times\delta_B$. This way, the condition is

$$
0 = \pi^1_\#((x-y)(\widetilde{\vartheta}\times\delta_A + \widetilde{\vartheta}\times\delta_B)) = \pi^1_\#((x-A)\widetilde{\vartheta}\times\delta_A) + \pi^1_\#((x-B)\widetilde{\vartheta}\times\delta_B) = (x-A)\widetilde{\vartheta} + (x-B)\widetilde{\vartheta},
$$

therefore $x=\frac{A+B}{2}$ in the support of $\widetilde{\vartheta}$. That is, $\widetilde{\vartheta}$ is the Dirac mass at the middle point between A and B. Its weight is the excess of mass of μ with respect to ν.

We conclude with a simple result relating the relaxed cost $\mathcal{W}_p^-$ to the global cost $\mathbb{W}_p$, and showing that they are also bounded from below by a distance.

Proposition 3.14. *Let $p\ge 1$, $\mu,\nu\in\mathscr{M}_p^{\alpha,M}(\Omega)$, and $|\nu|(\Omega)\le|\mu|(\Omega)$. Then*

$$
\mathcal{W}_p(\nu,\mu) \ge \widetilde{\mathbb{W}_p}^-(\nu,\mu) \ge \mathbb{W}_p^-(\nu,\mu) \ge \left(\frac{1}{2M}\right)^{p/(p-1)} \mathbb{W}_1(\mu,\nu).
$$

Proof. Let ϑ^+, ϑ^- be, as usual, a pair realizing the infimum in the definition of $\mathcal{W}_p(\nu,\mu)$. Let $\gamma^+\in\Gamma_o^p(\vartheta^+,\mu^+)$, $\gamma^-\in\Gamma_o^p(\vartheta^-,\mu^-)$. Then

$$
(\gamma^2)^{-1}\in\Gamma_o^p(\mu^-,\vartheta^-) \quad\text{and}\quad \gamma^+ + (\gamma^-)^{-1}\in\Gamma(\mu^-+\vartheta^+,\vartheta^-+\mu^+).
$$

Hence

$$
\begin{aligned}
\mathcal{W}_p(\mu,\nu) &= \left(W_p^p(\mu^+,\vartheta^+) + W_p^p(\mu^-,\vartheta^-)\right)^{1/p} = \left(\int\limits_{\Omega\times\Omega} |x-y|^p\, d(\gamma^+ + \gamma^-)\right)^{1/p}\\
&= \left(\int\limits_{\Omega\times\Omega} |x-y|^p\, d(\gamma^+ + (\gamma^-)^{-1})\right)^{1/p} \ge W_p(\mu^-+\vartheta^+,\vartheta^-+\mu^+) \ge \widetilde{\mathbb{W}_p}^-(\nu,\mu).
\end{aligned}
$$

The inequality $\widetilde{\mathbb{W}_p^-} \geq \mathbb{W}_p$ is obvious, since there are more degrees of freedom in the minimization problem defining $\mathbb{W}_p$. On the other hand, if ς^1, ς^2 and ϑ^1, ϑ^2 solve this problem, and $\gamma \in \Gamma_o^p(\varsigma^1 + \vartheta^2, \vartheta^1 + \varsigma^2)$, we have

$$\mathbb{W}_p^-(\nu, \mu) = \left(\int\limits_{\Omega\times\Omega} |x-y|^p \, d\gamma \right)^{1/p} \geq \left(\frac{1}{2M}\right)^{p/(p-1)} \int\limits_{\Omega\times\Omega} |x-y| \, d\gamma$$
$$\geq \left(\frac{1}{2M}\right)^{p/(p-1)} W_1(\varsigma^1 + \vartheta^2, \vartheta^1 + \varsigma^2) = \left(\frac{1}{2M}\right)^{p/(p-1)} \mathbb{W}_1(\nu, \mu),$$

since $\varsigma^1 - \varsigma^2 = \nu$ and $\vartheta^1 - \vartheta^2 = \mu$. □

REFERENCES

1. F. J. Almgren, J. Taylor, and L. Wang, "Curvature-driven flows: a variational approach," *SIAM J. Control Optim.*, **31**, No. 2, 387–438 (1993).
2. L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Spaces of Probability Measures*, Birkhäuser Verlag, Basel (2005).
3. L. Ambrosio, E. Mainini, and S. Serfaty, "Gradient flow of the Chapman–Rubinstein–Schatzman model for signed vortices," *Ann. Inst. H. Poincaré Anal. Non Linéaire*, **28**, No. 2, 217–246 (2011).
4. L. Ambrosio and S. Serfaty, "A gradient flow approach to an evolution problem arising in superconductivity," *Comm. Pure Appl. Math.*, **61**, No. 11, 1495–1539 (2008).
5. V. I. Bogachev, *Measure Theory*, 2 volumes, Springer-Verlag, Berlin (2007).
6. Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Comm. Pure Appl. Math.*, **44**, 375–417 (1991).
7. L. A. Caffarelli and R. J. McCann, "Free boundaries in optimal transport and Monge–Ampère obstacle problems," *Ann. Math.* (2), **171**, No. 2, 673–730 (2010).
8. J. A. Carrillo, R. J. McCann, and C. Villani, "Contractions in the 2-Wasserstein length space and thermalization of granular media," *Arch. Ration. Mech. Anal.*, **179**, No. 2, 217–263 (2006).
9. J. S. Chapman, J. Rubinstein, and M. Schatzman, "A mean-field model for superconducting vortices," *European J. Appl. Math.*, **7**, No. 2, 97–111 (1996).
10. E. De Giorgi, "New problems on minimizing movements," in: C. Baiocchi and J. L. Lions (eds.), *Boundary Value Problems for PDE and Applications*, Masson, Paris (1993), pp. 81–98.
11. W. E, "Dynamics of vortex-liquids in Ginzburg–Landau theories with applications to superconductivity," *Phys. Rev. B*, **50**, No. 3, 1126–1135 (1994).
12. A. Figalli, "The optimal partial transport problem," *Arch. Ration. Mech. Anal.*, **195**, No. 2, 533–560 (2010).
13. W. Gangbo and R. McCann, "The geometry of optimal transport," *Acta Math.*, **177**, 113–161 (1996).
14. R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation," *SIAM J. Math. Anal.*, **29**, 1–17 (1998).
15. L. V. Kantorovich, "On the transfer of masses," *Dokl. Akad. Nauk SSSR*, **37**, 227–229 (1942).
16. L. V. Kantorovich, "On a problem of Monge," *Uspekhi Mat. Nauk*, **3**, 225–226 (1948).
17. E. Mainini, "A global uniqueness result for an evolution problem arising in superconductivity," *Boll. Unione Mat. Ital.* (9), **2**, No. 2, 509–528 (2009).
18. E. Mainini, "Well-posedness for a mean field model of Ginzburg–Landau vortices with opposite degrees," *NoDEA Nonlinear Differential Equations Appl.*, in press.
19. F. Otto, "The geometry of dissipative evolution equations: the porous-medium equation," *Comm. Partial Differential Equations*, **26**, 101–174 (2001).
20. C. Villani, *Topics in Optimal Transportation*, Amer. Math. Soc., Providence, Rhode Island (2003).
21. C. Villani, *Optimal Transport, Old and New*, Springer-Verlag, Berlin (2008).
22. G. Wolanski, "Limit theorems for optimal mass transportation," *Calc. Var. Partial Differential Equations*, in press.