Università degli Studi di Padova

Padua Research Archive - Institutional Repository

Generalized Wasserstein Distance and its Application to Transport Equations with Source

(Article begins on next page)

Generalized Wasserstein distance and its application to transport equations with source

BENEDETTO PICCOLI, FRANCESCO ROSSI

Abstract

In this article, we generalize the Wasserstein distance to measures with different masses. We study the properties of such distance. In particular, we show that it metrizes weak convergence for tight sequences.

We use this generalized Wasserstein distance to study a transport equation with source, in which both the vector field and the source depend on the measure itself. We prove existence and uniqueness of the solution to the Cauchy problem when the vector field and the source are Lipschitzian with respect to the generalized Wasserstein distance.

Key words. transport equation − evolution of measures − Wasserstein distance − pedestrian modelling

MSC code: 35F25

The problem of optimal transportation, also called Monge-Kantorovich problem, has been intensively studied in the mathematical community. Related to such problem, the definition of the Wasserstein distance in the space of probability measure has revealed itself to be a powerful tool, in particular for dealing with dynamics of measures (like the transport PDE, see e.g. [2]). For a complete introduction to Wasserstein distances, see [19, 20].

The main limit of this approach, at least for its application to dynamics of measures, is that the Wasserstein distance $W_p(\mu, \nu)$ is defined only if the two measures μ, ν have the same mass. For this reason, in this article we first define a generalized Wasserstein distance $W_p^{a,b}(\mu, \nu)$, combining the standard Wasserstein and L^1 distances. In rough words, for $W_p^{a,b}(\mu, \nu)$ an infinitesimal mass $\delta\mu$ of μ (or ν) can either be removed at cost $a|\delta\mu|$, or moved from μ to ν at cost $bW_p(\delta\mu, \delta\nu)$.

After proving simple but important properties of this distance, we show that it is the natural distance to state existence and uniqueness of the solution to the following dynamics of measures:

$$\begin{cases} \partial_t \mu + \nabla \cdot (v\,[\mu]\,\mu) = h\,[\mu]\,, \\ \mu_{|t=0} = \mu_0. \end{cases} \tag{1}$$

Such equation is intensively used for modelling of crowd dynamics, where μ_t represents a pedestrian density (see e.g. [6, 8, 12–14, 16–18]). Other examples in one dimension for structured populations are given in [4, 10, 11]. Several authors have studied (1) without source terms, i.e. $h \equiv 0$, showing that it is very convenient to use the standard Wasserstein distance in this framework. In particular, in [1] the authors prove existence and uniqueness of the solution under Lipschitzianity of v with respect to μ, when the space of measures (of given fixed mass) is endowed with the Wasserstein distance. See also [5]. We also showed that these hypotheses give convergence of numerical schemes [14].

The main limit of the approach based on the standard Wasserstein distance is that it cannot encompass the case of a source h. Indeed, in this case the mass of the measure μ_t varies in time, hence

in general $W_p(\mu_t, \mu_s)$ is not defined for $t \neq s$. Sources (and sinks) are nevertheless very interesting for models of pedestrian, for instance in the case of people entering or exiting a door. It is interesting to recall that the L^1 distance (that one could try to use, since it is defined even between two measures with different masses) is not suitable in this context, since Lipschitzianity of v with respect to μ measured in L^1 does not guarantee uniqueness (see [14]).

In this article, to deal with a source in (1), we focus on the space of Borel measures with finite mass on $\mathbb{R}^d$ (denoted with $\mathcal{M}$), that we endow with the generalized Wasserstein distance $W_p^{a,b}$. We also denote with $\mathcal{M}_0^{ac}$ the subspace of $\mathcal{M}$ of measures that are absolutely continuous with respect to the Lebesgue measure and with bounded support. In this framework, we prove existence and uniqueness of the solution of (1) with $\mu_0 \in \mathcal{M}_0^{ac}$ under the following hypotheses:

(H)

The function
$$v\,[\mu] : \begin{cases} \mathcal{M} \to C^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d) \\ \mu \mapsto v\,[\mu] \end{cases}$$

satisfies

- $v\,[\mu]$ is uniformly Lipschitz and uniformly bounded, i.e. there exist L, M not depending on μ, such that for all $\mu \in \mathcal{M}, x, y \in \mathbb{R}^d$,

$$|v\,[\mu]\,(x) - v\,[\mu]\,(y)| \leq L|x - y| \qquad |v\,[\mu]\,(x)| \leq M.$$

- v is a Lipschitz function, i.e. there exists N such that

$$\|v\,[\mu] - v\,[\nu]\,\|_{\mathrm{C}^0} \leq N W_p^{a,b}\,(\mu, \nu).$$

The function
$$h\,[\mu] : \begin{cases} \mathcal{M} \to \mathcal{M}_0^{ac} \\ \mu \mapsto h\,[\mu] \end{cases}$$

satisfies

- $h\,[\mu]$ has uniformly bounded mass and support, i.e. there exist P, R such that

$$h\,[\mu]\,(\mathbb{R}^d) \leq P, \qquad \mathrm{supp}\,(h\,[\mu]) \subseteq B_R(0).$$

- h is a Lipschitz function, i.e. there exists Q such that

$$W_p^{a,b}\,(h\,[\mu], h\,[\nu]) \leq Q W_p^{a,b}\,(\mu, \nu).$$

Remark 1. The hypotheses **(H)** can be relaxed in standard ways. For example, the results still hold if we remove the uniform boundedness of v on $\mathbb{R}^d$ and ask for uniform boundedness only in a point x_0. See e.g. [3, Chap. 2].

Remark 2. The application to pedestrian dynamics also explains the choice of the basic assumptions **(H)**, namely that we deal with measures with bounded support.

The structure of the paper is the following. In Section 1, we define the generalized Wasserstein distance $W_p^{a,b}$ and we prove some important properties related to that. In particular, this distance metrizes the weak topology for tight sequences. Moreover, $\mathcal{M}$ is complete with respect to such distance. We also compare $W_p^{a,b}$ with other distances, like the Levy-Prokhorov distance. We then restrict ourselves to the study of $W_p^{a,b}$ in $\mathcal{M}_0^{ac}$ and provide Gronwall-like estimates under flow action.

In Section 2, we describe the complete picture for (1) under **(H)**. We first provide a candidate solution for (1) via a semi-discrete Lagrangian scheme and using the sample-and-hold method. We then show that it is indeed a solution, and finally that it is unique.

More results on the generalized Wasserstein distance will be proposed in the forthcoming paper [15]. There, we prove a duality formula of equality between the $W_1^{1,1}$ distance and the flat distance, provide the analog of the Benamou-Brenier formula for $W_2^{a,b}$ and generalize results of Section 2.

1. Generalized Wasserstein distance

In this section, we define the generalized Wasserstein distance $W_p^{a,b}(\mu, \nu)$ that we study in this article, and prove some useful properties. We first recall basic definitions and notations about measure theory and Wasserstein distance. For a complete introduction, see [7, 20].

1.1. Notation and standard Wasserstein distance

In this section, we fix the notation that we use throughout the paper, and recall definitions and properties related to measure theory and the Wasserstein distance, like push-forward of measures $\gamma \# \mu$ and transference plans.

Let μ be a positive Borel measure with locally finite mass. If μ_1 is absolutely continuous with respect to μ, we write $\mu_1 \ll \mu$. If $\mu_1 \ll \mu$ and $\mu_1(A) \le \mu(A)$ for all Borel sets, we write $\mu_1 \le \mu$. Given a measure with finite mass, we denote with $|\mu| := \mu(\mathbb{R}^d)$ its norm. More in general, if $\mu = \mu^+ - \mu^-$ is a signed Borel measure, we have $|\mu| := |\mu^+| + |\mu^-|$. Such norm defines a distance in $\mathcal{M}$, that is $|\mu - \nu|$. It is useful to recall that, if $\mu_1 \ll \mu$ and $d\mu_1 = f \, d\mu$ with $f \in L^1(d\mu)$, then $|\mu_1| = \int |f| \, d\mu$.

Given two measures μ, ν, one can always write in a unique way $\mu = \mu_{ac} + \mu_s$ such that $\mu_{ac} \ll \nu$ and $\mu_s \perp \nu$, i.e. there exists B such that $\mu_s(B) = 0$ and $\nu(\mathbb{R}^n \setminus B) = 0$. This is the Lebesgue's decomposition Theorem. Then, it exists a unique $f \in L^1(d\nu)$ such that $d\mu_{ac}(x) = f(x) \, d\nu(x)$. Such function is called the Radon-Nikodym derivative of μ with respect to ν. We denote it with $D_\nu \mu$. For more details, see e.g. [7].

Given a Borel map $\gamma : \mathbb{R}^d \to \mathbb{R}^d$, one can consider the following action on a measure $\mu \in \mathcal{M}$:

$$\gamma \# \mu(A) := \mu(\gamma^{-1}(A)).$$

An evident property is that the mass of μ, i.e. $\mu(\mathbb{R}^d)$ is identical to the mass of $\gamma \# \mu$.

Given two measures μ, ν with the same mass, it is thus possible to ask if there exists a γ such that $\nu = \gamma \# \mu$. We say that γ sends μ to ν. Moreover, one can add a cost to such γ, given by $I[\gamma] := |\mu|^{-1} \int_{\mathbb{R}^d} |x - \gamma(x)|^p \, d\mu(x)$. This means that each infinitesimal mass $\delta\mu$ is sent to $\delta\nu$ and that its infinitesimal cost is related to the p-th power of the distance between them. The, one can consider the map γ minimizing such cost, if it exists. This is known as the Monge problem, stated by Monge in 1791.

In general, this procedure works only for special μ, ν and p. Indeed, there exist simple examples of μ, ν for which a γ that sends μ to ν does not exist. For example $\mu = 2\delta_1$, $\nu = \delta_0 + \delta_2$ on the real line have the same mass, but there exists no γ with $\nu = \gamma \# \mu$, since γ cannot separate masses. Moreover, one can have a sequence γ_n of maps such that $I[\gamma_n]$ is a minimizing sequence, but the limit is not a map γ^*. A simple condition that ensures the existence of a minimizing γ is that μ and ν are absolutely continuous with respect to the Lebesgue measure.

For such reason, one can generalize the problem to the following setting. Given a probability measure π on $\mathbb{R}^d \times \mathbb{R}^d$, one can interpret it as a method to transfer a measure μ on $\mathbb{R}^d$ to another measure on $\mathbb{R}^d$ as follows: each infinitesimal mass on a location x is sent to a location y with a probability given by $\pi(x, y)$. Formally, μ is sent to ν if the following properties hold:

$$|\mu| \int_{\mathbb{R}^d} d\pi(x, \cdot) = d\mu(x), \qquad |\nu| \int_{\mathbb{R}^d} d\pi(\cdot, y) = d\nu(y). \qquad (2)$$

Such π is called a transference plan from μ to ν. We denote the set of such transference plans as $\Pi(\mu, \nu)$. Since one usually deals with probability measures μ, ν, the terms $|\mu|, |\nu|$ are usually neglected in the literature. A condition equivalent to (2) is that, for all $f, g \in C_c^\infty(\mathbb{R}^d)$ it holds $|\mu| \int_{\mathbb{R}^d \times \mathbb{R}^d} (f(x) + g(y)) \, d\pi(x, y) = \int_{\mathbb{R}^d} f(x) \, d\mu(x) + \int_{\mathbb{R}^d} g(y) \, d\nu(y)$. Then, one can define a cost for π as follows $J[\pi] := \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \, d\pi(x, y)$ and look for a minimizer of J in $\Pi(\mu, \nu)$. Such problem is called the Monge-Kantorovich problem.

It is important to observe that such problem is a generalization of the Monge problem. Indeed, given a γ sending μ to ν, one can define a transference plan $\pi = (\mathrm{Id} \times \gamma)\#\mu$, i.e. $d\pi(x,y) = \mu(\mathbb{R}^n)^{-1} d\mu(x)\delta_{y=\gamma(x)}$. It also holds $J[\mathrm{Id} \times \gamma] = I[\gamma]$. The main advantages of such approach are the following: first, the existence of at least one π satisfying (2) is easy to check, since one can choose $\pi(A \times B) = |\mu|^{-1} \mu(A)\nu(B)$, i.e. the mass from μ is proportionally split to ν. Moreover, a minimizer of J in $\Pi(\mu,\nu)$ always exists.

A natural space in which J is finite is the space of Borel measures with finite p-moment, that is

$$\mathcal{M}^p := \left\{ \mu \in \mathcal{M} \mid \int |x|^p \, d\mu(x) < \infty \right\}.$$

In the following, we also denote with $\mathcal{P}$ the space of probability measures, i.e. the measures in $\mathcal{M}$ with unit mass. We also deal with $\mathcal{P}^p := \mathcal{M}^p \cap \mathcal{P}$, i.e. the space of probability measures with finite p-moment. One can thus define on $\mathcal{M}^p$ the following operator between measures of the same mass, called the **Wasserstein distance**:

$$W_p(\mu,\nu) = \left(|\mu| \min_{\pi \in \Pi(\mu,\nu)} J[\pi] \right)^{1/p}.$$

It is indeed a distance on the subspace of measures in $\mathcal{M}^p$ with a given mass, see [20]. It is easy to prove that $W_p(k\mu, k\nu) = k^{1/p} W_p(\mu,\nu)$ for $k \geq 0$, by observing that $\Pi(k\mu, k\nu) = \Pi(\mu,\nu)$ and that $J[\pi]$ does not depend on the mass. We will recall some other properties all along the paper, when useful.

1.2. Definition of generalized Wasserstein distance

We are now ready to define the **generalized Wasserstein distance** $W_p^{a,b}(\mu,\nu)$. We first give a rough description of the idea. Imagine to have three different admissible actions on μ, ν: either add/remove mass to μ, or add/remove mass to ν or transport mass from μ to ν. The three techniques have their cost: add/remove mass has a unitary cost a (in both cases); transport of mass has the classic Monge-Kantorovich cost J, multiplied by a fixed constant b. The distance is the minimal cost of a mix of such techniques. We will show in the following that, depending on μ, ν, all mixes are possible: either remove all the mass of μ and ν (if they are very far), or transport the whole μ to ν (if they have the same mass and are close enough), or a mix of the two (for example when μ and ν are very close but with different masses). Instead, we will prove that add mass is never optimal.

Remark 3. The fact that the unitary cost of adding/removing mass is identical for the two terms μ, ν is to ensure symmetry of $W_p^{a,b}$.

We now formally define the generalized Wasserstein distance.

Definition 1. Let $\mathcal{M}$ be the space of Borel measures with finite mass on $\mathbb{R}^d$. Then, given $a, b \in (0, \infty)$ and $p \geq 1$, the **generalized Wasserstein distance** is

$$W_p^{a,b}(\mu,\nu) = \inf_{\substack{\tilde{\mu}, \tilde{\nu} \in \mathcal{M}^p \\ |\tilde{\mu}| = |\tilde{\nu}|}} \left(a|\mu - \tilde{\mu}| + a|\nu - \tilde{\nu}| + b W_p(\tilde{\mu}, \tilde{\nu}) \right). \tag{3}$$

Proposition 1. The operator $W_p^{a,b}$ is a distance. Moreover, one can restrict the computation in (3) to $\tilde{\mu} \leq \mu, \tilde{\nu} \leq \nu$ and the infimum is always attained.

To prove this Proposition, we need to recall some known concepts and results related to tightness and weak convergence. For more details, see [19,20].

We first recall results about tightness. We always deal with measures on $\mathbb{R}^d$.

Definition 2. A set of measures M is tight if for each $\varepsilon > 0$ there exists a compact K_ε such that $\mu(\mathbb{R}^d \setminus K_\varepsilon) < \varepsilon$ for all $\mu \in M$.

Lemma 1. The following holds:

1. A measure with finite mass is tight.

2. A set of measures M such that all measures are bounded above by a tight measure $\tilde{\mu}$ is tight.

3. The union of a finite number of tight sets is tight. In particular, a finite set of measures with finite mass is tight.

Proof. The first result is proved by taking a sequence of invading compacts $\{K^n\}$. Since $\mu(\mathbb{R}^d) = \lim_n \mu(K^n) = \sum_{n=1}^{\infty} \mu(K^n \setminus K^{n-1}) < \infty$, then $\mu(\mathbb{R}^d \setminus K^i) = \sum_{n=i+1}^{\infty} \mu(K^n \setminus K^{n-1})$, as small as needed, since it is the remainder of a converging sequence.

For the second property, for each ε take K_ε that gives tightness of $\tilde{\mu}$. Since $\mu \in M$ implies $\mu \leq \tilde{\mu}$, hence $\mu(\mathbb{R}^d \setminus K_\varepsilon) < \varepsilon$.

For the third result, take the sets $M_1, \ldots, M_n$, each of them being tight. For each $\varepsilon > 0$, there exists a corresponding compact K_ε^i such that for each $\mu \in M_i$ satisfies $\mu(\mathbb{R}^d \setminus K_\varepsilon^i)$. Then, define $K_\varepsilon = \cup_{i=1}^n K_\varepsilon^i$, that is compact because it is a finite union of compacts. Now take $\mu \in \cup_{i=1}^n M_i$ and observe that $\mu \in M_i$ for some i, thus $\mu(\mathbb{R}^d \setminus K_\varepsilon) \leq \mu(\mathbb{R}^d \setminus K_\varepsilon^i) < \varepsilon$.

We now recall results about weak convergence and give some technical lemmas we use in the following.

Theorem 1 (Prokhorov's theorem). *Let X be a Polish space. A set P in the space of probabilities $\mathcal{P}(X)$ is precompact for the weak topology if and only if it is tight.*

We recall that $\mathbb{R}^d$ is a Polish space (see a definition in [20]), thus Prokhorov's theorem can be applied in our setting.

Theorem 2 (Weak compactness [7]). *Let μ^n be a sequence of Radon measures in $\mathbb{R}^d$ with uniformly bounded mass[1], i.e. there exists M such that $|\mu^k| \leq M$ for all k. Then there exists a subsequence μ^{n_j} and a Radon measure μ^* such that $\mu^{n_j} \rightharpoonup \mu^*$.*

From now on, we denote with the asterisk $\mu^*, c^*, \ldots$ the limit (or the weak limit) of a sequence $\mu_n, c_n, \ldots$

Lemma 2. *Let μ^n, ν^n be two sequences of Borel measures such that $\mu^n(\mathbb{R}^d) = \nu^n(\mathbb{R}^d)$ for each n. For each n, let π^n be a transference plan with marginals μ^n, ν^n. If we have $\mu^n \rightharpoonup \mu^*$, $\nu^n \rightharpoonup \nu^*$, $\pi^n \rightharpoonup \pi^*$ for some μ^*, ν^*, π^*, then π^* is a transference plan with marginals μ^*, ν^*.*

Proof. The weak convergence of π^n means that, for each $f \in C_c^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ we have $\int f(x,y) \, d\pi^n(x,y) \to \int f(x,y) \, d\pi^*(x,y)$. In particular, choose $f(x,y) = g(x)$ and observe that

$$\int g(x) \, d\pi^*(x,y) \leftarrow \int g(x) \, d\pi^n(x,y) = \int g(x) \, d\mu^n(x) \to \int g(x) \, d\mu^*(x).$$

By uniqueness of the limit, we have the proof for the marginal μ^*. Using $f(x,y) = h(y)$, we have the same for ν^*.

We are now ready to prove Proposition 1.

Proof of Proposition 1. The **symmetry property** $W_p^{a,b}(\mu,\nu) = W_p^{a,b}(\nu,\mu)$ is evident.

We first prove that **we can always restrict to** $\tilde{\mu} \leq \mu$. Define

$$\mathcal{C}(\tilde{\mu}, \tilde{\nu}) := a|\mu - \tilde{\mu}| + a|\nu - \tilde{\nu}| + b W_p(\tilde{\mu}, \tilde{\nu}).$$

First assume that the infimum of $\mathcal{C}$ is attained by $\tilde{\mu} \not\leq \mu$ and a certain $\tilde{\nu}$. Let $\pi \in \Pi(\tilde{\mu}, \tilde{\nu})$ be the transference realizing $W_p(\tilde{\mu}, \tilde{\nu})$. Let d be the Radon-Nikodym derivative $f = D_\mu \tilde{\mu}$ and $\mu_\perp := \tilde{\mu} - f\mu$ the orthogonal of $\tilde{\mu}$ with respect to μ. Define $\bar{\mu} := \min\{f, 1\} \mu$ and $\bar{\nu}$ the image of $\bar{\mu}$ under π. Since $\bar{\mu} \leq \tilde{\mu}$ and $\pi \in \Pi(\tilde{\mu}, \tilde{\nu})$, then $\bar{\nu} \leq \tilde{\nu}$. Moreover, $|\tilde{\nu} - \bar{\nu}| = |\tilde{\mu} - \bar{\mu}|$, since $|\tilde{\mu}| = |\tilde{\nu}|$, $|\bar{\mu}| = |\bar{\nu}|$ by construction. Observe that

$$|\mu - \tilde{\mu}| = \int |1 - f| \, d\mu + \mu_\perp(\mathbb{R}^d) = \int_{1-f \geq 0} (1-f) \, d\mu + \int_{1-f < 0} (f-1) \, d\mu + \mu_\perp(\mathbb{R}^d) =$$

$$= \int_{f \leq 1} (d\mu - d\bar{\mu}) + \left(\int_{f > 1} (d\tilde{\mu} - d\bar{\mu}) - \mu_\perp(\{f > 1\}) \right) + \mu_\perp(\mathbb{R}^d) \geq |\mu - \bar{\mu}| + |\bar{\mu} - \tilde{\mu}|.$$

---

[1] The result in [7] is stated with a weaker condition, that is uniformly boundedness of mass on each compact.

Hence

$$|\mu - \bar{\mu}| + |\nu - \bar{\nu}| \le |\mu - \tilde{\mu}| - |\bar{\mu} - \tilde{\mu}| + |\nu - \tilde{\nu}| + |\tilde{\nu} - \bar{\nu}| = |\mu - \tilde{\mu}| + |\nu - \tilde{\nu}|. \tag{4}$$

The fact that $W_p(\bar{\mu}, \bar{\nu}) \le W_p(\tilde{\mu}, \tilde{\nu})$ is a direct consequence of the fact that $\pi \in \Pi(\tilde{\mu}, \tilde{\nu})$ can be restricted to $\pi' \in \Pi(\bar{\mu}, \bar{\nu})$, by construction of $\bar{\nu}$. Moreover, the cost of π' is smaller[2] that the cost of π. Using this inequality and (4), we have that $\mathcal{C}(\bar{\mu}, \bar{\nu}) \le \mathcal{C}(\tilde{\mu}, \tilde{\nu})$. Since $\bar{\mu} \le \mu$, then the result is proven.

If the infimum is not attained, consider a minimizing sequence $\tilde{\mu}^n$ and construct each $\bar{\mu}^n := \min\{D_\mu \tilde{\mu}^n, 1\} \mu$, that gives another minimizing sequence with $\bar{\mu}^n \le \mu$. By symmetry, the same property can be proved for the term ν.

We now prove that **the infimum in (3) is always attained**. To prove it, we restrict ourselves to $\tilde{\mu} \le \mu, \tilde{\nu} \le \nu$. Take a sequence $\tilde{\mu}^n, \tilde{\nu}^n$ such that $\mathcal{C}(\tilde{\mu}^n, \tilde{\nu}^n) \to W_p^{a,b}(\mu, \nu)$ and $\tilde{\mu}^n \le \mu, \tilde{\nu}^n \le \nu$. Since both $\mu, \nu \in \mathcal{M}$ have finite mass, then $\{\tilde{\mu}^n\}, \{\tilde{\nu}^n\}$ have both uniformly bounded masses. Thus, due to Theorem 2, passing to sub-sequences we have $\tilde{\mu}^n \rightharpoonup \mu^*, \tilde{\nu}^n \rightharpoonup \nu^*$. We now prove that $\mathcal{C}(\mu^*, \nu^*) \le \lim_n \mathcal{C}(\tilde{\mu}^n, \tilde{\nu}^n) = W_p^{a,b}(\mu, \nu)$. First recall that weak convergence gives $|\mu - \mu^*| \le \liminf_n |\mu - \tilde{\mu}^n|$ and equivalently for $|\nu - \nu^*|$. We are left to prove that

$$W_p(\mu^*, \nu^*) \le \lim_n W_p(\tilde{\mu}^n, \tilde{\nu}^n). \tag{5}$$

If $\mu^* = \nu^* = 0$, then we are done. Otherwise, the sequence $c_n := |\tilde{\mu}^n| = |\tilde{\nu}^n|$ (eventually passing to a sub-sequence) converges to $c^* > 0$. For n such that $c_n \ne 0$, define the probability measures $\bar{\mu}^n = c_n^{-1} \tilde{\mu}^n$, $\bar{\nu}^n = c_n^{-1} \tilde{\nu}^n$. It is clear that $\bar{\mu}^n \rightharpoonup \bar{\mu}^* = (c^*)^{-1} \mu^*$, and similarly for $\bar{\nu}^n$. Denote with π^n the optimal transference plan in $\Pi(\tilde{\mu}^n, \tilde{\nu}^n) = \Pi(\bar{\mu}^n, \bar{\nu}^n)$. Since $\bar{\mu}^n \le \frac{1}{\sup c_n} \mu$ and $\bar{\nu}^n \le \frac{1}{\sup c_n} \nu$, then $M := \{\bar{\mu}^n\}$, $N := \{\bar{\nu}^n\}$ are both tight, hence the set of transference plans $\Pi(M, N)$ is tight (see e.g. [19, Lemma 4.4]). Hence, due to Prokhorov's theorem, up to sub-sequences we have $\pi^n \rightharpoonup \pi^*$ for some π^*. Using Lemma 2, we have that π^* is a transference plan with marginals $\bar{\mu}^*$ and $\bar{\nu}^*$ (not necessarily optimal). Since the distance is non-negative, then the functional $J : \pi \to \int |x - y|^p \, d\pi(x, y)$ is lower semicontinuous with respect to the weak topology, see [19, Lemma 4.3], thus

$$W_p(\mu^*, \nu^*) = (c^*)^{1/p} W_p(\bar{\mu}^*, \bar{\nu}^*) \le (c^*)^{1/p} J(\pi^*)^{1/p} \le \lim_n c_n^{1/p} J(\pi^n)^{1/p} = \lim_n c_n^{1/p} W_p(\bar{\mu}^n, \bar{\nu}^n) = \lim_n W_p(\tilde{\mu}^n, \tilde{\nu}^n).$$

We now prove that $W_p^{a,b}(\mu, \nu) = 0$ **implies** $\mu = \nu$. Since the infimum is attained for some $\tilde{\mu}, \tilde{\nu}$, then $|\mu - \tilde{\mu}| = W_p(\tilde{\mu}, \tilde{\nu}) = |\nu - \tilde{\nu}| = 0$ implies $\mu = \tilde{\mu} = \tilde{\nu} = \nu$.

We now prove **triangle inequality** $W_p^{a,b}(\mu, \eta) \le W_p^{a,b}(\mu, \nu) + W_p^{a,b}(\nu, \eta)$. Denote with $\tilde{\mu}, \tilde{\nu}^1$ the minimizers in (3) for $W_p^{a,b}(\mu, \nu)$ and $\tilde{\nu}^2, \tilde{\eta}$ the minimizers for $W_p^{a,b}(\nu, \eta)$. Observe that we have $\tilde{\nu}^1 \ne \tilde{\nu}^2$ in general. Also call π^1, π^2 the transference plans realizing $W_p(\tilde{\mu}, \tilde{\nu}^1), W_p(\tilde{\nu}^2, \tilde{\eta})$, respectively. Define now $\bar{\nu} := \min\{D_{\tilde{\nu}^2} \tilde{\nu}^1, 1\} \tilde{\nu}^2$. Define $\bar{\mu}, \bar{\eta}$ the marginals of $\bar{\nu}$ with respect to π^1, π^2 respectively, i.e. $\pi^1 \in \Pi(\bar{\mu}, \bar{\nu})$ and $\pi^2 \in \Pi(\bar{\nu}, \bar{\eta})$. Thus we have $\pi^2 \circ \pi^1 \in \Pi(\bar{\mu}, \bar{\eta})$. By construction, we have $|\tilde{\mu} - \bar{\mu}| = |\tilde{\nu}^1 - \bar{\nu}|$. Moreover, $W_p(\bar{\mu}, \bar{\nu}) \le W_p(\tilde{\mu}, \tilde{\nu}^1)$, because we use the same transference plan π^1 with a smaller mass. Similar properties hold for $\tilde{\eta}, \bar{\eta}$. We thus have

$$\begin{aligned} W_p^{a,b}(\mu, \eta) &\le a|\mu - \bar{\mu}| + a|\eta - \bar{\eta}| + bW_p(\bar{\mu}, \bar{\eta}) \le \\ &\le a|\mu - \tilde{\mu}| + a|\tilde{\mu} - \bar{\mu}| + a|\eta - \tilde{\eta}| + a|\tilde{\eta} - \bar{\eta}| + bW_p(\bar{\mu}, \bar{\nu}) + bW_p(\bar{\nu}, \bar{\eta}) \le \\ &\le a|\mu - \tilde{\mu}| + a|\tilde{\nu}^1 - \bar{\nu}| + a|\eta - \tilde{\eta}| + a|\tilde{\nu}^2 - \bar{\nu}| + bW_p(\tilde{\mu}, \tilde{\nu}^1) + bW_p(\tilde{\nu}^2, \tilde{\eta}). \end{aligned} \tag{6}$$

Define $f := D_{\tilde{\nu}^2} \tilde{\nu}^1$ and $\tilde{\nu}_\perp^1 := \tilde{\nu}^1 - f \tilde{\nu}^2$ the orthogonal part of $\tilde{\nu}^1$ with respect to $\tilde{\nu}^2$. Observe that both $\tilde{\nu}^1, \tilde{\nu}^2 \le \nu$, hence $D_{\tilde{\nu}^2} \nu \ge \max\{f, 1\}$ and $\nu_\perp := \nu - (D_{\tilde{\nu}^2} \nu) \tilde{\nu}^2$ satisfies $\nu_\perp \ge \tilde{\nu}_\perp^1$. Then one has

$$|\tilde{\nu}^1 - \bar{\nu}| = \int (f - \min\{f, 1\}) \, d\tilde{\nu}^2 + \tilde{\nu}_\perp^1(\mathbb{R}^d) = \int_{f \ge 1} (f - 1) d\tilde{\nu}^2 + \int_{f < 1} 0 \, d\tilde{\nu}^2 + \tilde{\nu}_\perp^1(\mathbb{R}^d) =$$

$$= \int_{f \ge 1} (\max\{f, 1\} - 1) \, d\tilde{\nu}^2 + \tilde{\nu}_\perp^1(\mathbb{R}^d) \le \int_{f \ge 1} (d\nu - d\tilde{\nu}^2) + \nu_\perp(\mathbb{R}^d) = |\nu - \tilde{\nu}^2|,$$

and similarly $|\tilde{\nu}^2 - \bar{\nu}| \le |\nu - \tilde{\nu}^1|$. Plugging them into (6), one has the proof. $\qquad\square$

One interesting feature of this distance is that the $|\cdot|$ term and the Wasserstein term W_p have different degree of homogeneity with respect to translation in $\mathbb{R}^n$, thus the optimal strategy for $W_p^{a,b}$

---

[2] Theorem 4.6 in [19] also shows that π' is optimal, but this is not crucial here.

varies when translating one measure. For example, compute $W_p^{a,b}(\delta_0, \delta_x)$ as a function of $x \geq 0$. We have $|\delta_0 - \delta_x| = 2$ and $W_p(\delta_0, \delta_x) = x$. Hence $W_p^{a,b}(\delta_0, \delta_x) = \min\{2a, bx\}$. If $2a < bx$, i.e. measures are "far", then the optimal strategy is to delete both masses δ_0 and δ_x, otherwise it is optimal to move δ_0 to δ_x with a translation.

Another simple example permits us to show that optimal strategies can be either based on removing mass only (L^1 strategy), or on transporting mass only (W_p strategy), or by a mix of them. We give an example in Figure 1. Take the measures μ, ν on the real line with densities $d\mu = \chi_{[-1,0]}\, d\lambda$, $d\nu = \chi_{[x,1+x]}\, d\lambda$ with $x \geq 0$, where λ is the Lebesgue measure. It is clear that the optimal strategy amounts to choose $\tilde{\mu}, \tilde{\nu}$ with densities $\chi_{[-y,0]}, \chi_{[x,x+y]}$ respectively, for a certain parameter $y \in [0, 1]$. The L_1 strategy is given by choosing $y = 0$, while the W_p strategy is given by $y = 1$. We now prove that all values of y can be optimal, depending on x and the parameters a, b, p. We have $|\mu - \tilde{\mu}| = |\nu - \tilde{\nu}| = 1 - y$ and $W_p(\tilde{\mu}, \tilde{\nu}) = |y|^{1/p}(x + y)$. We choose for simplicity $a = b = p = 1$. Thus $W_p^{a,b}(\mu, \nu) = \min_{y \in [0,1]} 2 - 2y + xy + y^2$. A simple computation shows that the minimum is attained by $y = \frac{2-x}{2}$ if $x \in [0, 2]$, and $y = 0$ for $x \geq 2$. This clearly shows that if the measures are very close ($x = 0$), then the best strategy is the W_p (since $y = 1$), while for measures that are far (x big) the best strategy is the L^1. As stated above, varying $x \in (0, 2)$ one has mixed strategies.
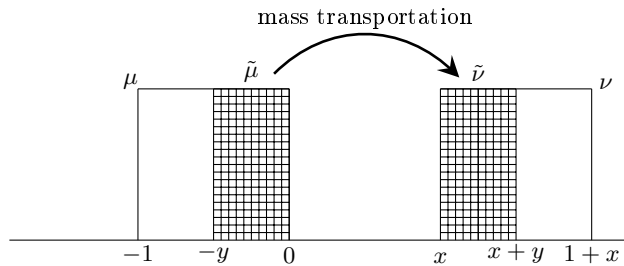

Fig. 1. Choice of $\tilde{\mu}, \tilde{\nu}$ (shaded) for the computation of $W_p^{a,b}$.

We now state some simple properties of $W_p^{a,b}$.

Proposition 2. *The following properties hold:*

- $W_p^{a,b}(k\mu, k\nu) \leq \max\{k^{1/p}, k\} W_p^{a,b}(\mu, \nu)$ *for $k \geq 0$,*
- $W_p^{a,b}(\mu_1 + \mu_2, \nu_1 + \nu_2) \leq W_p^{a,b}(\mu_1, \nu_1) + W_p^{a,b}(\mu_2, \nu_2)$.
- $a\left||\mu| - |\nu|\right| \leq W_p^{a,b}(\mu, \nu) \leq a(|\mu| + |\nu|)$

Proof. The first two properties are direct consequences of similar properties for $|\cdot|$ and W_p.

For the third, we first prove the inequality $a\left||\mu| - |\nu|\right| \leq W_p^{a,b}(\mu, \nu)$. Without loss of generality, we assume $|\mu| \geq |\nu|$. Take any $\tilde{\mu} \leq \mu$, $\tilde{\nu} \leq \nu$ and observe that $|\mu - \tilde{\mu}| = |\mu| - |\tilde{\mu}|$, and similarly for $\nu, \tilde{\nu}$. Also recall that $|\tilde{\mu}| = |\tilde{\nu}| \leq |\nu|$ by construction. Now choose $\tilde{\mu} \leq \mu, \tilde{\nu} \leq \nu$ realizing $W_p^{a,b}(\mu, \nu)$ and observe that

$$W_p^{a,b}(\mu, \nu) \geq a|\mu - \tilde{\mu}| + a|\nu - \tilde{\nu}| = a\left(|\mu| - |\tilde{\mu}| + |\nu| - |\tilde{\nu}|\right) \geq a\left(|\mu| - |\nu| + 0\right).$$

We now prove the inequality $W_p^{a,b}(\mu, \nu) \leq a(|\mu| + |\nu|)$. Choose $\tilde{\mu} = \tilde{\nu} = 0$ and observe that $\mathcal{C}(0,0) = a(|\mu| + |\nu|)$. Since $W_p^{a,b}$ is the infimum on all $\tilde{\mu}, \tilde{\nu}$, we have the inequality.

1.3. Topology of the generalized Wasserstein distance

In this section, we prove that $W_p^{a,b}$ metrizes weak convergence for tight sequences. We also prove that $\mathbb{R}^d$ is complete with respect to $W_p^{a,b}$.

We first prove a simple lemma for $W_p^{a,b}$, stating that optimal choices $\tilde{\mu}, \tilde{\nu}$ are very close to each other in $\mathbb{R}^d$. The basic idea is that, if we want to transfer far mass between $\tilde{\mu}$ and $\tilde{\nu}$, then it is cheaper to remove such masses from both measures.

Proposition 3. *Let $\mu, \nu \in \mathcal{M}$ and $\tilde{\mu}, \tilde{\nu}$ be the choices realizing $W_p^{a,b}(\mu, \nu)$. If $\tilde{\mu}$ has support contained in a compact K, then $\tilde{\nu}$ has bounded support, contained in the enlarged compact*

$$K^d := \cup_{x \in K} B(x, d) \tag{7}$$

with $d = \frac{2a}{b}$. Here we denote with $B(x, r)$ the closed ball centered in x with radius r.

Similarly, let $\tilde{\mu}, \tilde{\nu}$ be the choices realizing $W_p^{a,b}(\mu, \nu)$, and π the transference plan realizing $W_p(\tilde{\mu}, \tilde{\nu})$. Let $\mu' \leq \tilde{\mu}$ have support contained in a compact K. Then the corresponding marginal ν' with respect to π has compact support contained in K^d.

Proof. We prove the second statement, since the first can be recovered by choosing $\mu' = \tilde{\mu}$.

With the notations given in the statement, we prove that the support of ν' is contained in K^d, by contradiction. Assume that there exists $d' > d$ and a Borel set $D \subset \mathbb{R}^d \setminus K^{d'}$ such that $\nu'(D) > 0$. Thus define $\bar{\nu} := k\nu'_{|_D}$ with $k \leq 1$ chosen in such a way that $\bar{\nu}(D) < 1$. Observe that $\bar{\nu} \leq \nu' \leq \tilde{\nu}$. Define $\nu^* := \tilde{\nu} - \bar{\nu} \leq \tilde{\nu}$ and μ^* the corresponding marginal given by π. We now prove that the choice μ^*, ν^* in (3) gives a cost that is strictly less than the optimal choice $\tilde{\mu}, \tilde{\nu}$. First observe that

$$|\mu - \mu^*| + |\nu - \nu^*| = |\mu - \tilde{\mu}| + |\tilde{\mu} - \mu^*| + |\nu - \tilde{\nu}| + |\tilde{\nu} - \nu^*| =$$
$$= |\mu - \tilde{\mu}| + |\nu - \tilde{\nu}| + 2|\tilde{\nu} - \nu^*| = |\mu - \tilde{\mu}| + |\nu - \tilde{\nu}| + 2|\bar{\nu}|. \tag{8}$$

Now observe that $W_p^p(\tilde{\mu}, \tilde{\nu}) = W_p^p(\mu^*, \nu^*) + W_p^p(\tilde{\mu} - \mu^*, \tilde{\nu} - \nu^*)$, by construction of μ^*, ν^* via the optimal transference plan π. Observe that $\text{supp}(\tilde{\mu} - \mu^*) \subseteq \text{supp}(\mu') \subseteq K$ and that $\text{supp}(\tilde{\nu} - \nu^*) = \text{supp}(\bar{\nu}) \subseteq D \subseteq \mathbb{R}^d \setminus K^{d'}$. In particular, if $x \in \text{supp}(\tilde{\mu} - \mu^*), y \in \text{supp}(\tilde{\nu} - \nu^*)$, then $|x - y| \geq d'$. Thus, given the optimal transference plan $\pi' \in \Pi(\tilde{\mu} - \mu^*, \tilde{\nu} - \nu^*)$, we have

$$W_p^p(\tilde{\mu} - \mu^*, \tilde{\nu} - \nu^*) = \int |x - y|^p d\pi'(x, y) \geq \int d'^p \, d(\tilde{\nu} - \nu^*) = |\bar{\nu}| d'^p. \tag{9}$$

Putting together (8) and (9), and using $|\bar{\nu}|^{1/p} \geq |\bar{\nu}|$ since $|\bar{\nu}| \leq 1$, we have

$$\mathcal{C}(\mu^*, \nu^*) \leq a|\mu - \tilde{\mu}| + a|\nu - \tilde{\nu}| + 2a|\bar{\nu}| + bW_p(\tilde{\mu}, \tilde{\nu}) - bW_p(\tilde{\mu} - \mu^*, \tilde{\nu} - \nu^*) \leq$$
$$\leq a|\mu - \tilde{\mu}| + a|\nu - \tilde{\nu}| + bW_p(\tilde{\mu}, \tilde{\nu}) + 2a|\bar{\nu}| - bd'|\bar{\nu}|^{1/p} \leq \mathcal{C}(\tilde{\mu}, \tilde{\nu}) - b\left(d' - 2\frac{a}{b}\right)|\bar{\nu}| < \mathcal{C}(\tilde{\mu}, \tilde{\nu}).$$

Thus $\tilde{\mu}, \tilde{\nu}$ is not optimal. Contradiction. Thus $\text{supp}(\nu') \subseteq K^{d'}$ for all $d' > d$, hence $\text{supp}(\nu') \subseteq K^d$.

We now prove the following convergence theorem, stating that $W_p^{a,b}$ metrizes weak convergence for tight sequences.

Theorem 3. *Let $\{\mu_n\}$ be a sequence of measures in $\mathbb{R}^d$, and $\mu_n, \mu \in \mathcal{M}$. Then*

$$W_p^{a,b}(\mu_n, \mu) \to 0 \qquad \text{is equivalent to} \qquad \mu_n \rightharpoonup \mu \quad \text{and} \quad \{\mu_n\} \text{ is tight.}$$

Proof. Fix the following notation: for each n, let $\tilde{\mu}_n, \tilde{\nu}_n$ be optimal choices in (3) for $W_p^{a,b}(\mu_n, \mu)$ with $\tilde{\mu}_n \leq \mu_n, \tilde{\nu}_n \leq \mu$, and π_n the transference plan realizing $W_p(\tilde{\mu}_n, \tilde{\nu}_n)$.

We first prove $\Leftarrow$. Fix $\varepsilon > 0$. Let N be such that $W_p^{a,b}(\mu_n, \mu) < \varepsilon$ for all $n > N$. Since $\{\mu_n\}$ is tight, then the set $M := \{\mu_n\} \cup \{\mu\}$ is tight too. Given $\delta > 0$, consider the corresponding K_δ giving tightness of M. By definition, all $m \in M$ satisfy $m(\mathbb{R}^d \setminus K_\delta) < \delta$. Due to weak convergence of μ_n to μ, we also have $\mu_n(K_\delta) \to \mu(K_\delta)$. Choose N' such that $\left||\mu_n(K_\delta)| - |\mu(K_\delta)|\right| < \varepsilon$ for all $n > N'$. It means that there exists ν_n (positive or negative, supported in K_δ) such that $(\mu_n + \nu_n)(K_\delta) = \mu(K_\delta)$. Define $\tilde{\mu}_n := (\mu_n + \nu_n)_{|_{K_\delta}}$ and $\tilde{\mu} := \mu_{|_{K_\delta}}$. It is clear that $\nu_n \rightharpoonup 0$, hence $\tilde{\mu}_n \rightharpoonup \tilde{\mu}$. Since K_δ has bounded diameter and W_p metrizes weak convergence in bounded spaces (see [20, 7.12], recalled below in Theorem 4), then $W_p(\tilde{\mu}_n, \tilde{\mu}) \to 0$. Take $N > N'$ such that $W_p(\tilde{\mu}_n, \tilde{\mu}) < \delta$ and $|\nu_n| < \delta$ for all $n > N$. We now estimate

$$W_p^{a,b}(\mu_n, \mu) \leq \mathcal{C}(\tilde{\mu}_n, \tilde{\mu}) = a|\mu_n - \tilde{\mu}_n| + a|\mu - \tilde{\mu}| + bW_p(\tilde{\mu}_n, \tilde{\mu}) <$$
$$< a|\mu_n(\mathbb{R}^d \setminus K_\delta)| + a|\nu_n| + a|\mu(\mathbb{R}^d \setminus K_\delta)| + b\delta < (3a + b)\delta.$$

Choose $\delta = \varepsilon/(3a + b)$ and have the result.

We now prove $\Rightarrow$. As a first, main step, we prove that $W_p^{a,b}(\mu_n, \mu) \to 0$ **implies that $\{\mu_n\}$ is tight**. Fix now $\varepsilon > 0$. There exists a corresponding N such that $W_p^{a,b}(\mu_n, \mu) < \varepsilon$ for all $n > N$, and a compact K_ε such that $\mu(\mathbb{R}^d \setminus K_\varepsilon) < \varepsilon$, since μ is tight. Define $\nu_n' := \tilde{\nu}_n$ restricted to K_ε, and μ_n' the corresponding measure given by the transference plan π_n realizing $W_p(\tilde{\mu}_n, \tilde{\nu}_n)$. Using Proposition 3, we have that the support of μ_n' is contained in K_ε^d with $d = 2\frac{a}{b}$. Observe that such set does not depend on n. By construction, we have

$$|\tilde{\mu}_n - \mu_n'| = |\tilde{\mu}_n| - |\mu_n'| = |\tilde{\nu}_n| - |\nu_n'| = \tilde{\nu}_n(\mathbb{R}^d \setminus K_\varepsilon) < \varepsilon.$$

We now estimate

$$\mu_n(\mathbb{R}^d \setminus K_\varepsilon^d) \le |\mu_n - \tilde{\mu}_n| + \tilde{\mu}_n(\mathbb{R}^d) - \tilde{\mu}_n(K_\varepsilon^d) \le \frac{1}{a} W_p^{a,b}(\mu_n, \mu) + |\tilde{\mu}_n| - |\mu_n'| \le \frac{\varepsilon}{a} + \varepsilon.$$

This is the tightness of $\{\mu_n\}_{n > N}$. Observe now that the finite set of measures $\{\mu_1, \ldots, \mu_N\}$ with finite masses is tight. Since a finite union of tight sets is tight, we have that $\{\mu_n\}$ is tight.

We now observe that $|\mu_n| \le |\mu| + \frac{1}{a} W_p(\mu_n, \mu)$ by Proposition 2, that is a converging sequence of real numbers, thus μ_n have uniformly bounded mass. We now apply Theorem 2, that gives **weak convergence** of μ_n to a certain μ^*. We prove that $\mu = \mu^*$. Using the implication $\Leftarrow$, we have $W_p^{a,b}(\mu_n, \mu^*) \to 0$. Thus $W_p^{a,b}(\mu, \mu^*) \le \lim_n W_p^{a,b}(\mu, \mu_n) + W_p^{a,b}(\mu_n, \mu^*) = 0$, hence $\mu = \mu^*$.

It is interesting to compare such result with a similar result for the standard Wasserstein distance, that we recall here.

Theorem 4. [20, 7.12] *Let μ_k be a sequence of probability measures in $\mathcal{P}^p$, and $\mu \in \mathcal{P}$. Then the following statements are equivalent:*

- *$W_p(\mu_k, \mu) \to 0$;*
- *$\mu_k \rightharpoonup \mu$ and the following condition holds*

$$\lim_{R \to \infty} \limsup_k \int_{|x|>R} |x|^p \, d\mu_k(x) = 0; \tag{10}$$

- *$\mu_k \rightharpoonup \mu$ and the p-moment converges, i.e.*

$$\int |x|^p \, d\mu_k(x) \to \int |x|^p \, d\mu(x). \tag{11}$$

Condition (10) is called "tightness" condition, a notation that could create some confusion with respect to Definition 2. Anyway, we remark that condition (10) is stronger than Definition 2, thus our Theorem 3 applies on a wider class than Theorem 4. First of all, Theorem 4 applies for measures that have all the same mass, otherwise $W_p(\mu_k, \mu)$ is not defined. Moreover, even taking a sequence of probability measures μ_k, one can have convergence in $W_p^{a,b}$ and no convergence in W_p. This occurs exactly in the case in which $\{\mu_k\}$ is tight according to Definition 2 and not according to condition (10). For example, take the following sequence of probability measures: $\mu_k := (1 - k^{-p})\delta_0 + k^{-p} \delta_k$. It is clear that such sequence converges weakly to $\mu^* = \delta_0$, and that $W_p(\mu_k, \mu^*) = W_p(k^{-p}\delta_0, k^{-p}\delta_k) = k^{-1} \int k^p \delta_0 = k^{p-1} \not\to 0$. Indeed, condition (11) is not satisfied, since the p-moment of μ^* is 0, while the p-moment of μ_k is 1. Similarly, condition (10) is not satisfied, since for each R the measures μ_k with $k > R$ satisfy $\int_{|x|>R} d\mu_k = k^{-p} \int_{|x|>R} k^p \delta_k = 1 \not\to 0$. Instead, estimate $W_p^{a,b}(\mu_k, \mu^*)$ by choosing $\tilde{\mu}_k = \mu_k - k^{-p} \delta_k$, $\tilde{\mu}^* = \mu^* - k^{-p} \delta_0$, that gives $W_p^{a,b}(\mu_k, \mu^*) = a|k^{-p} \delta_k| + a|k^{-p} \delta_0| = 2k^{-p} \to 0$. Thus, $\{\mu_k\}$ is tight according to Definition 2, as it is easy to prove by observing that $\mu_k(\mathbb{R} \setminus [-n, n]) < n^{-p}$.

It is clear that, on the contrary, if $\{\mu_k\}$ satisfies (10), then it satisfies Definition 2. Indeed, let $\{\mu_k\}$ satisfy (10). For each $\varepsilon > 0$ there exist R, m such that $\int_{|x|>R} |x|^p \, d\mu_k < \varepsilon$ for all $k > m$. One can always assume $R > 1$. Thus $\mu_k(\mathbb{R}^n \setminus B(0, R)) < \varepsilon$ for all $k > m$, that ensures tightness of $\{\mu_k\}$.

We now prove completeness of $\mathcal{M}$.

Proposition 4. *$\mathcal{M}$ is complete with respect to $W_p^{a,b}$.*

Proof. Take a sequence μ_n that is Cauchy with respect to $W_p^{a,b}$. We first show that $\{\mu_n\}$ is tight. Fix $\delta > 0$ and let N be such that $W_p^{a,b}(\mu_n, \mu_{n+k}) < \delta$ for all $n \geq N, k \geq 0$. Fix the notation $\tilde{\mu}_k, \tilde{\nu}_k$ to denote the choices in (3) realizing the minimizer $W_p^{a,b}(\mu_N, \mu_{N+k})$, and π_k the transference plan realizing $W_p(\tilde{\mu}_k, \tilde{\nu}_k)$. Let K_δ be such that $\mu_N(\mathbb{R}^d \setminus K_\delta) < \delta$, that exists since μ_N has finite mass. Since $\tilde{\mu}_k \leq \mu_N$, then $\tilde{\mu}_k(\mathbb{R}^d \setminus K_\delta) < \delta$. Define now μ'_k the restriction of $\tilde{\mu}_k$ to K_δ and ν'_k the corresponding marginal with respect to π_k. Using Proposition 3, we have that $\mathrm{supp}(\nu'_k) \subseteq K_\delta^d$ with $d = \frac{2a}{b}$.

As a consequence $\tilde{\nu}_k(K_\delta^d) \geq \nu'_k(K_\delta^d) = \mu'_k(K_\delta) = \tilde{\mu}_k(K_\delta)$. Since $\tilde{\mu}_k(\mathbb{R}^d) = \tilde{\nu}_k(\mathbb{R}^d)$ by construction, we have $\tilde{\nu}_k(\mathbb{R}^d \setminus K_\delta^d) \leq \tilde{\mu}_k(\mathbb{R}^d \setminus K_\delta) < \delta$. Since $a|\mu_{N+k} - \tilde{\nu}_k| \leq W_p^{a,b}(\mu_N, \mu_{N+k}) < \delta$, we have that $\mu_{N+k}(\mathbb{R}^d \setminus K_\delta^d) \leq \tilde{\nu}_k(\mathbb{R}^d \setminus K_\delta^d) + \delta/a < \delta + \delta/a$. Choose $\delta = \frac{\varepsilon}{1+1/a}$ and $K'_\varepsilon = K_\delta^d$, then $\mu_n(\mathbb{R}^d \setminus K'_\varepsilon) < \varepsilon$ for $n \geq N$. Thus, we have tightness for $n \geq N$. Since $\{\mu_1, \ldots, \mu_N\}$ is a finite family of measures with finite mass, then it is tight, hence the whole $\{\mu_n\}$ is tight.

Observe that $\{\mu_n\}$ has also uniformly bounded mass, thus there exists a subsequence $\mu_{n_k} \rightharpoonup \mu^*$ for a certain μ^*, due to Theorem 2. Using Theorem 3, we have that $W_p^{a,b}(\mu_{n_k}, \mu^*) \to 0$, and by triangular inequality we have $W_p^{a,b}(\mu_n, \mu^*) \to 0$.

1.4. Comparison with other distances

In this section, we compare $W_p^{a,b}$ with two distances proposed in the literature, namely the Levy-Prokhorov distance (see e.g. [20]) and the distance W_{b_2} defined in [9] by A. Figalli and N. Gigli.

We first recall the definition of the Levy-Prokhorov distance d_{LP} between two probability measures μ, ν:

$$d_{LP}(\mu, \nu) := \inf \{\alpha > 0 \text{ such that for any closed } A \text{ it holds } \mu(A) \leq \nu(A^\alpha) + \alpha\},$$

where A^α is the enlarged set $A^\alpha := \cup_{x \in A} B(x, \alpha)$. It is clear that the two terms $\nu(A^\alpha)$ and α represent a W_p perturbation and a L_1 perturbation, respectively. For such reason, there are some common ideas between d_{LP} and $W_p^{a,b}$.

The main difference here is that d_{LP} was defined for probability measures only, while we deal with measures with different (finite) masses. Nevertheless, even restricting to the space of probability measures, the two distances have different values. We study a remarkable case, that is the distance between $\mu = \delta_0$ and $\nu = \frac{1}{2}\delta_{-d_1} + \frac{1}{2}\delta_{d_2}$ on the real line. We study the values of d_{LP} and $W_p^{a,b}$ as functions of d_1, d_2. For both distances, the goal is to choose the optimal coupling, on one side between $\frac{1}{2}\delta_0$ and $\frac{1}{2}\delta_{-d_1}$, and on the other side between $\frac{1}{2}\delta_0$ and $\frac{1}{2}\delta_{d_2}$.

We now compare d_{LP} with $W_p^{a,b}$ of parameters $a = \frac{1}{2}, b = 1$. Without loss of generality, we assume $d_1 \leq d_2$. The set of different cases is the following:

- The two masses $\frac{1}{2}\delta_{-d_1}, \frac{1}{2}\delta_{d_2}$ are both "far" from δ_0. The best choice for both couplings is to focus on L_1 distance. For the d_{LP}, it means that $d_1 \geq 1$. In this case $d_{LP}(\mu, \nu) = 1$. For $W_p^{a,b}$, it means that $d_1 \geq \frac{2a}{b} = 1$. In this case $W_p^{a,b}(\mu, \nu) = \mathcal{C}(0, 0) = 1$.
- One of the masses is "close" and the other is "far". This means that $d_1 \leq 1 \leq d_2$. In this case, we use the W_p distance for the coupling $\frac{1}{2}\delta_0, \frac{1}{2}\delta_{-d_1}$ and the L_1 distance for the coupling $\frac{1}{2}\delta_0, \frac{1}{2}\delta_{d_2}$. Thus $d_{LP}(\mu, \nu) = \sup\{\frac{1}{2}, d_1\}$ and $W_p^{a,b}(\mu, \nu) = \mathcal{C}\left(\frac{1}{2}\delta_0, \frac{1}{2}\delta_{-d_1}\right) = \frac{1}{2} + 2^{-1/p}d_1$.
- Both masses are "close, but one is not very close". This is true for $\frac{1}{2} \leq d_2 \leq 1$. In this case, the optimal strategies are different for the two distances:
  - for d_{LP}, we use the W_p distance for the coupling $\frac{1}{2}\delta_0, \frac{1}{2}\delta_{-d_1}$ and the L_1 distance for the coupling $\frac{1}{2}\delta_0, \frac{1}{2}\delta_{d_2}$. Thus $d_{LP}(\mu, \nu) = \sup\{\frac{1}{2}, d_1\}$.
  - for $W_p^{a,b}$ we use the W_p distance for both couplings. This gives $W_p^{a,b}(\mu, \nu) = C(\mu, \nu) = W_p(\mu, \nu) = \left(\frac{d_1^p + d_2^p}{2}\right)^{1/p}$.
- Both masses are "very close". This is the case of $d_2 \leq \frac{1}{2}$. For both distance the best choice is to use the W_p distance. Thus $d_{LP}(\mu, \nu) = d_2$ and $W_p^{a,b}(\mu, \nu) = C(\mu, \nu) = W_p(\mu, \nu) = \left(\frac{d_1^p + d_2^p}{2}\right)^{1/p}$.

It is easy to prove that the two distances $d_{LP}, W_p^{a,b}$ are equivalent as norms on the space (d_1, d_2). Moreover, they have the same level lines around $d_1 = 0, d_2 = 0$ when choosing $p = \infty$.

We now recall the distance W_{b_2} defined in [9]. Such distance is defined on a subset Ω with non-empty boundary $\partial\Omega$. The idea is that such boundary is an infinite reserve of mass, in the following sense. Given two measures μ, ν, even with different masses, one can either send a part of mass of μ to ν or to the boundary $\partial\Omega$. Similarly, the mass of ν that does not receive mass from μ goes to the boundary $\partial\Omega$. In both cases, the cost is computed via the Wasserstein distance, either from μ to ν or from μ to $\partial\Omega$.

This cost is based on an approach that is rather different than ours. First of all, it deals with a space Ω that has boundary, otherwise one could not deal with measures with different masses. Moreover, the cost of sending mass to the boundary (that is similar to delete mass in our approach) is computed as a Wasserstein distance and not with L_1, like in our case.

1.5. Estimates of generalized Wasserstein distance under flow actions

In this section, we study properties of $W_p^{a,b}$ when restricted to $\mathcal{M}_0^{ac}$. In particular, we are interested in estimates about the variation of $W_p^{a,b}(\mu, \nu)$ under action of flows on μ, ν. These Gronwall-like properties will be useful for the study of solutions of (1).

We first recall a connection between flows actions on measures and transport equation. Take a Lipschitz vector field v, that generates a flow Φ_t^v for $t \geq 0$. The flow is a diffeomorphism for the space $\mathbb{R}^d$, thus one can define $\mu_t := \Phi_t^v \# \mu_0$ for a given measure μ_0. One has the following theorem.

Theorem 5. *Take a Lipschitz vector field v, and the flow Φ_t^v it generates. Given $\mu_0 \in \mathcal{M}_0^{ac}$, and $\mu_t := \Phi_t^v \# \mu_0$, then $\mu = \mu_{[0,T]}$ is the unique solution of the linear transport equation*

$$\begin{cases} \partial_t \mu_t + \nabla \cdot (v\mu_t) = 0 \\ \mu_{|t=0} = \mu_0 \end{cases} \tag{12}$$

in $C\left([0,T], \mathcal{M}_0^{ac}\right)$, where $\mathcal{M}_0^{ac}$ is endowed with the weak topology.

Proof. The proof is a direct consequence of [20, Thm 5.34]. See details in [14]. $\quad\blacksquare$

To study (1) in the setting of the generalized Wasserstein distance, it is interesting to check the evolution of $W_p^{a,b}$ under flow action. We first recall here some properties about standard Wasserstein distance, that we proved in [14]. Here, we emphasize the fact that the Wasserstein distance is computed between two measures μ, ν with same mass, in general different than 1.

Proposition 5. *Given v, w bounded and Lipschitz vector fields of Lipschitz constant L and $\mu, \nu \in \mathcal{M}_0^{ac}$, the following holds:*

1. $W_p(\Phi_t^v \# \mu, \Phi_t^v \# \nu) \leq e^{\frac{p+1}{p}Lt} W_p(\mu, \nu)$,
2. $W_p(\mu, \Phi_t^v \# \mu) \leq t \, \|v\|_{C^0} \mu(\mathbb{R}^d)^{1/p}$,
3. $W_p(\Phi_t^v \# \mu, \Phi_t^w \# \nu) \leq e^{\frac{p+1}{p}Lt} W_p(\mu, \nu) + \mu(\mathbb{R}^d)^{1/p} \frac{e^{Lt/p}(e^{Lt}-1)}{L} \|v - w\|_{C^0}$.

We now prove similar properties for the generalized Wasserstein distance.

Proposition 6. *Given v, w bounded and Lipschitz vector fields of Lipschitz constant L, the following holds:*

1. $W_p^{a,b}(\Phi_t^v \# \mu, \Phi_t^v \# \nu) \leq e^{\frac{p+1}{p}Lt} W_p^{a,b}(\mu, \nu)$,
2. $W_p^{a,b}(\mu, \Phi_t^v \# \mu) \leq t \, \|v\|_{C^0} \mu(\mathbb{R}^d)^{1/p}$,
3. $W_p^{a,b}(\Phi_t^v \# \mu, \Phi_t^w \# \nu) \leq e^{\frac{p+1}{p}Lt} W_p^{a,b}(\mu, \nu) + \mu(\mathbb{R}^d)^{1/p} \frac{e^{Lt/p}(e^{Lt}-1)}{L} \|v - w\|_{C^0}$.

Proof. For the first property, take $\tilde{\mu}, \tilde{\nu}$ realizing (3) for $W_p^{a,b}(\mu, \nu)$ with $\tilde{\mu} \leq \mu, \tilde{\nu} \leq \nu$. Then

$$W_p^{a,b}(\Phi_t^v \# \mu, \Phi_t^v \# \nu) \leq a|\Phi_t^v \# \mu - \Phi_t^v \# \tilde{\mu}| + a|\Phi_t^v \# \nu - \Phi_t^v \# \tilde{\nu}| + b W_p(\Phi_t^v \# \tilde{\mu}, \Phi_t^v \# \tilde{\nu}) \leq$$
$$\leq a|\mu - \tilde{\mu}| + a|\nu - \tilde{\nu}| + b e^{\frac{p+1}{p}Lt} W_p(\tilde{\mu}, \tilde{\nu}) \leq e^{\frac{p+1}{p}Lt} \left(a|\mu - \tilde{\mu}| + a|\nu - \tilde{\nu}| + b W_p(\tilde{\mu}, \tilde{\nu}) \right).$$

The proofs of the second and the third properties are equivalent, based on proofs of Proposition 5 given in [14].

2. Existence and uniqueness of solutions of (1)

In this section we prove the existence and uniqueness of the solution of (1), under the hypothesis **(H)**. The key tool is the construction of a candidate solution by sample-and-hold; we then prove that it is indeed a solution, and finally prove that it is unique.

2.1. Construction of an approximated solution

In this section, we apply the sample-and-hold method to construct a sequence of functions in $C\left([0,T],\mathcal{M}_0^{ac}\right)$ such that the limit exists and it is a solution of (1). With no loss of generality, we assume that the $T = 1$.

We first define an approximated solution μ^k for each $k \in \mathbb{N}$. Given a fixed k, define $\Delta t := \frac{1}{2^k}$ and consider the decomposition of the time interval in $[0, \Delta t], [\Delta t, 2\Delta t], [2\Delta t, 3\Delta t], \dots, \left[(2^k - 1)\Delta t, 2^k \Delta t\right]$.

The idea is based on a Lagrangian scheme with the time-discretization provided above. In particular, we define a solution for each time $n\Delta t$, then consider that v and h are computed at this time and fixed on the next interval $[n\Delta t, (n + 1)\Delta t)$. More precisely, for a fixed k, we define

- $\mu_0^k := \mu_0$;
- $\mu_{(n+1)\Delta t}^k := \Phi_{\Delta t}^{v\left[\mu_{n\Delta t}^k\right]} \# \mu_{n\Delta t}^k + \Delta t\, h\left[\mu_{n\Delta t}^k\right]$;
- $\mu_t^k := \Phi_\tau^{v\left[\mu_{n\Delta t}^k\right]} \# \mu_{n\Delta t}^k + \tau\, h\left[\mu_{n\Delta t}^k\right]$ with n the maximum integer such that $t - n\Delta t \geq 0$ and $\tau := t - n\Delta t$.

We now prove that, given v, h satisfying **(H)**, the sequence μ^k is a Cauchy sequence for the space $(C([0,1], \mathcal{M}), \mathcal{D})$, where
$$\mathcal{D}(\mu, \nu) := \sup_{t \in [0,1]} W_p^{a,b}(\mu_t, \nu_t).$$

Since we have proved that $(\mathcal{M}, W_p^{a,b})$ is a complete space, then also $(C([0,1], \mathcal{M}), \mathcal{D})$ is complete. Observe that we deal with the whole $\mathcal{M}$ to have completeness, and that we will subsequently prove that the limit is indeed an element of $\mathcal{M}_0^{ac}$.

We first make three simple observation:

- At each time $t \in [0,1]$, the mass $\mu_t^k(\mathbb{R}^d)$ is bounded. Indeed, first observe that
$$\mu_{(n+1)\Delta t}^k(\mathbb{R}^d) \leq \Phi_{\Delta t}^{v\left[\mu_{n\Delta t}^k\right]} \# \mu_{n\Delta t}^k(\mathbb{R}^d) + \Delta t\, h\left[\mu_{n\Delta t}^k\right](\mathbb{R}^d) \leq \mu_{n\Delta t}^k(\mathbb{R}^d) + \Delta t P.$$

As a consequence, we have
$$\mu_t^k(\mathbb{R}^d)^{1/p} \leq m, \quad \text{with} \quad m := (\mu_0(\mathbb{R}^d) + P)^{1/p}.$$

This holds for all k, t. We use such constant in the following estimates.

- Given a fixed k, the evolution of the scheme satisfies $W_p^{a,b}\left(\mu_{(n+1)\Delta t}^k, \mu_{n\Delta t}^k\right) \leq \Delta t M m + \Delta t P$, and more in general
$$W_p^{a,b}\left(\mu_t^k, \mu_s^k\right) \leq |t - s|(Mm + P).$$

- We will use in the following the estimate
$$W_p^{a,b}\left(\mu_{(n+\frac{1}{2})\Delta t}^{k+1}, \mu_{n\Delta t}^k\right) \leq W_p^{a,b}\left(\mu_{(n+\frac{1}{2})\Delta t}^{k+1}, \mu_{n\Delta t}^{k+1}\right) + W_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right) \leq$$
$$\leq \tfrac{\Delta t}{2}(Mm + P) + W_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right), \tag{13}$$

that is a simple consequence of Proposition 6-2.

We now estimate the distance $W_p^{a,b}\left(\mu_{(n+1)\Delta t}^{k+1}, \mu_{(n+1)\Delta t}^k\right)$ with respect to $W_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right)$, i.e. the evolution of $W_p^{a,b}$ at the discretization points of the Lagrangian scheme for μ^k.

We use the following compact notations:
$$\mathcal{V}_m^j := \Phi_{\Delta t/2}^{v\left[\mu_{m\Delta t}^j\right]} \qquad \mathcal{H}_m^j := h\left[\mu_{m\Delta t}^j\right].$$

We have

$$W_p^{a,b}\left(\mu_{(n+1)\Delta t}^{k+1}, \mu_{(n+1)\Delta t}^k\right) =$$

$$= W_p^{a,b}\left(\mathcal{V}_{n+\frac{1}{2}}^{k+1} \# \left(\mathcal{V}_n^{k+1} \# \mu_{n\Delta t}^{k+1} + \tfrac{\Delta t}{2}\mathcal{H}_n^{k+1}\right) + \tfrac{\Delta t}{2}\mathcal{H}_{n+\frac{1}{2}}^{k+1}, \mathcal{V}_n^k \# \mathcal{V}_n^k \# \mu_{n\Delta t}^k + \tfrac{\Delta t}{2}\mathcal{H}_n^k + \tfrac{\Delta t}{2}\mathcal{H}_n^k\right) \leq$$

$$\leq W_p^{a,b}\left(\mathcal{V}_{n+\frac{1}{2}}^{k+1} \# \mathcal{V}_n^{k+1} \# \mu_{n\Delta t}^{k+1}, \mathcal{V}_n^k \# \mathcal{V}_n^k \# \mu_{n\Delta t}^k\right) + \tfrac{\Delta t}{2} W_p^{a,b}\left(\mathcal{V}_{n+\frac{1}{2}}^{k+1} \# \mathcal{H}_n^{k+1}, \mathcal{H}_n^k\right) + \tfrac{\Delta t}{2} W_p^{a,b}\left(\mathcal{H}_n^{k+1}, \mathcal{H}_n^k\right).$$

We estimate the three parts using Proposition 6. Estimates will be given for sufficiently large k, i.e. for sufficiently small Δt. Moreover, to simplify the notation, we write estimates that hold for all $p \geq 1$. We estimate the first term as follows:

$$W_p^{a,b}\left(\mathcal{V}_{n+\frac{1}{2}}^{k+1} \# \mathcal{V}_n^{k+1} \# \mu_{n\Delta t}^{k+1}, \mathcal{V}_n^k \# \mathcal{V}_n^k \# \mu_{n\Delta t}^k\right) \leq$$

$$\leq (1 + 2L\Delta t)W_p^{a,b}\left(\mathcal{V}_n^{k+1} \# \mu_{n\Delta t}^{k+1}, \mathcal{V}_n^k \# \mu_{n\Delta t}^k\right) + m(1 + L\Delta t)\Delta t\|v\left[\mu_{(n+\frac{1}{2})\Delta t}^{k+1}\right] - v\left[\mu_{n\Delta t}^k\right]\|_{C^0} \leq$$

$$\leq (1 + 2L\Delta t)\left((1 + 2L\Delta t)W_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right) + m\Delta t(1 + L\Delta t)\|v\left[\mu_{n\Delta t}^{k+1}\right] - v\left[\mu_{n\Delta t}^k\right]\|_{C^0}\right) +$$

$$+ m\Delta t(1 + L\Delta t)N\left(\tfrac{\Delta t}{2}(Mm + P) + W_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right)\right) \leq$$

$$\leq (1 + (5L + 4mN)\Delta t)W_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right) + mN(Mm + P)\Delta t^2.$$

The second term is estimated as follows, using (13):

$$W_p^{a,b}\left(\mathcal{V}_{n+\frac{1}{2}}^{k+1} \# \mathcal{H}_n^{k+1}, \mathcal{H}_n^k\right) \leq W_p^{a,b}\left(\mathcal{V}_{n+\frac{1}{2}}^{k+1} \# \mathcal{H}_n^{k+1}, \mathcal{H}_n^{k+1}\right) + W_p^{a,b}\left(\mathcal{H}_n^{k+1}, \mathcal{H}_n^k\right) \leq \tfrac{\Delta t}{2}MP + QW_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right)$$

The third term is simply estimated by $W_p^{a,b}\left(\mathcal{H}_n^{k+1}, \mathcal{H}_n^k\right) \leq QW_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right)$. Summing up, we have

$$W_p^{a,b}\left(\mu_{(n+1)\Delta t}^{k+1}, \mu_{(n+1)\Delta t}^k\right) \leq (1 + C_1\Delta t)W_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right) + C_2\Delta t^2$$

with $C_1 := 5L + 4mN + Q$, $C_2 := mN(Mm + P) + \frac{MP}{4}$, both independent on k, n. Applying it recursively in n and recalling that $W_p^{a,b}\left(\mu_0^{k+1}, \mu_0^k\right) = 0$, we have

$$W_p^{a,b}\left(\mu_{n\Delta t}^{k+1}, \mu_{n\Delta t}^k\right) \leq C_2\Delta t^2 \frac{(1 + C_1\Delta t)^n - 1}{1 + C_1\Delta t - 1} \leq 2nC_2\Delta t^2. \tag{14}$$

We use such estimate to prove the convergence of μ_t^k for each $t \in [0, 1]$. We study the three following cases:

$t = 1$: it corresponds to $n = \Delta t^{-1}$. Applying the triangular inequality and (14), we have

$$W_p^{a,b}\left(\mu_1^k, \mu_1^{k+l}\right) \leq 2C_2\left(\frac{1}{2^k} + \frac{1}{2^{k+1}} + \ldots + \frac{1}{2^{k+l}}\right) \leq \frac{4C_2}{2^k}.$$

Thus, μ_1^k is a Cauchy sequence in a complete space, hence it is convergent.

$t = \frac{r}{2^l}$ for some r, l integers: Consider μ_t^k starting from $k = l$ and apply estimates similar to the previous case.

any $t \in [0, 1]$: For each k, let n_k be the maximum integer such that $t - n_k\Delta t \geq 0$. We have

$$W_p^{a,b}\left(\mu_t^k, \mu_t^{k+l}\right) \leq W_p^{a,b}\left(\mu_t^k, \mu_{n_k 2^{-k}}^k\right) + W_p^{a,b}\left(\mu_{n_k 2^{-k}}^k, \mu_{n_k 2^{-k}}^{k+l}\right) + W_p^{a,b}\left(\mu_{n_{k+l} 2^{-k-l}}^{k+l}, \mu_t^{k+l}\right) \leq$$

$$\leq 2^{-k}(Mm + P) + W_p^{a,b}\left(\mu_{n_k 2^{-k}}^k, \mu_{n_k 2^{-k}}^{k+l}\right) + 2^{-k}(Mm + P).$$

The first and third term are converging with respect to k, uniformly in l. The same holds for the second term, as proved at the previous step.

2.2. Existence and uniqueness of the solution

In this section we prove that the measure $\bar{\mu}_t := \lim_k \mu_t^k$ given as a limit of the sequence of the Lagrangian scheme is indeed a weak solution of (1). We will then prove that it is unique.

First observe that $\bar{\mu}_t$ exists, since it is the limit of a Cauchy sequence in $(\mathcal{M}, W_p^{a,b})$, that is complete. It is evident that, since $\mu_0^k = \mu_0$ for all k, then $\bar{\mu}_0 = \mu_0$. We have now to prove that, for each $f \in C_c^\infty((0,1) \times \mathbb{R}^d)$, it holds

$$\int_0^1 dt \int_{\mathbb{R}^d} \left(d\bar{\mu}_t \left(\partial_t f + v\left[\bar{\mu}_t\right] \cdot \nabla f \right) + dh\left[\bar{\mu}_t\right] f \right) = 0 \tag{15}$$

The idea is to use again μ^k. First of all, we recall the solution of the continuity equation with time-independent v or h and with the other part (h or v respectively) being identically zero. For $h \equiv 0$, the solution of (1) with time-independent v is $\mu_t = \Phi_t^v \# \mu_0$. For $v \equiv 0$, the solution of (1) with time-independent h is $\mu_t = \mu_0 + t\,h$. This gives

$$\sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} d\mu_t^k \left(\partial_t f + v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f \right) + dh\left[\mu_{n\Delta t}^k\right] f =$$

$$\sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} \left(d\left(\Phi_{t-n\Delta t}^{v\left[\mu_{n\Delta t}^k\right]} \# \mu_{n\Delta t}^k + (t-n\Delta t)h\left[\mu_{n\Delta t}^k\right] \right) \left(\partial_t f + v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f \right) + dh\left[\mu_{n\Delta t}^k\right] f \right) =$$

$$= \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} d\left(\Phi_{t-n\Delta t}^{v\left[\mu_{n\Delta t}^k\right]} \# \mu_{n\Delta t}^k \right) \left(\partial_t f + v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f \right) +$$

$$+ \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \left((t-n\Delta t) \int_{\mathbb{R}^d} dh\left[\mu_{n\Delta t}^k\right] \partial_t f + \int_{\mathbb{R}^d} dh\left[\mu_{n\Delta t}^k\right] f \right) +$$

$$+ \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt(t-n\Delta t) \int_{\mathbb{R}^d} dh\left[\mu_{n\Delta t}^k\right] v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f =$$

$$= 0 + 0 + \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt(t-n\Delta t) \int_{\mathbb{R}^d} dh\left[\mu_{n\Delta t}^k\right] v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f.$$

Due to the boundedness of $h, v, \nabla f$, it exists C_3 such that $\left| \int_{\mathbb{R}^d} dh\left[\mu_{n\Delta t}^k\right] v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f(t,.) \right| \leq C_3$, that gives

$$\left| \lim_k \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt(t-n\Delta t) \int_{\mathbb{R}^d} dh\left[\mu_{n\Delta t}^k\right] v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f \right| \leq$$

$$\leq \lim_k C_3 \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt(t-n\Delta t) = C_3\, 2^k \frac{\Delta t^2}{2} = 0$$

Going back to prove (15), we prove equivalently that

$$\lim_k \left| \int_0^1 dt \int_{\mathbb{R}^d} \left(d\bar{\mu}_t \left(\partial_t f + v\left[\bar{\mu}_t\right] \cdot \nabla f \right) + dh\left[\bar{\mu}_t\right] f \right) + \right.$$

$$\left. - \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} d\mu_t^k \left(\partial_t f + v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f \right) + dh\left[\mu_{n\Delta t}^k\right] f \right| = 0.$$

Since $\lim_k W_p^{a,b}\left(\mu_t^k, \bar{\mu}_t \right) = 0$ implies $\mu_t^k \rightharpoonup \bar{\mu}_t$ (Theorem 3), then

$$\lim_k \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} d\mu_t^k \partial_t f = \int_0^1 dt \int_{\mathbb{R}^d} d\bar{\mu}_t \partial_t f.$$

The same applies to h, for which we have $\lim_k W_p^{a,b}\left(h\left[\mu_t^k\right], h\left[\bar{\mu}_t\right]\right) = 0$, thus

$$\lim_k \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} dh\left[\mu_t^k\right] f = \int_0^1 dt \int_{\mathbb{R}^d} dh\left[\bar{\mu}_t\right] f.$$

We are now left to prove

$$\lim_k \left| \int_0^1 dt \int_{\mathbb{R}^d} d\bar{\mu}_t v\left[\bar{\mu}_t\right] \cdot \nabla f - \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} d\mu_t^k v\left[\mu_{n\Delta t}^k\right] \cdot \nabla f \right| = 0.$$

Using the triangular inequality, we prove it by proving the three following:

- $\lim_k \left| \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} \left(d\bar{\mu}_t - d\mu_t^k\right) v\left[\bar{\mu}_t\right] \cdot \nabla f \right| = 0$. Since $\bar{\mu}_t$ is continuous and v is Lipschitz with respect to its argument and bounded, then $v\left[\bar{\mu}_t\right] \cdot \nabla f \in C_c^\infty((0,1) \times \mathbb{R}^d)$, thus $\mu_t^k \rightharpoonup \bar{\mu}_t$ implies the result.
- $\lim_k \left| \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} d\mu_t^k \left(v\left[\bar{\mu}_t\right] - v\left[\mu_t^k\right]\right) \cdot \nabla f \right| = 0$. Since it exists a constant C_4 such that $W_p^{a,b}\left(\mu_t^k, \bar{\mu}_t\right) \le 2^{-k}C_4$, then the previous limit can be proved by recalling that μ_t^k has finite mass smaller than m and observing that ∇f is a bounded function. Thus

$$\lim_k \left| \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} d\mu_t^k \left(v\left[\bar{\mu}_t\right] - v\left[\mu_t^k\right]\right) \cdot \nabla f \right| \le \lim_k \left| \int_0^1 dt\, m\, 2^{-k} N C_4 \max_{[0,1]\times\mathbb{R}^d} (\nabla f) \right| = 0.$$

- $\lim_k \left| \sum_{n=0}^{2^k-1} \int_{n\Delta t}^{(n+1)\Delta t} dt \int_{\mathbb{R}^d} d\mu_t^k \left(v\left[\mu_t^k\right] - v\left[\mu_{n\Delta t}^k\right]\right) \cdot \nabla f \right| = 0$. The proof is similar to the previous case, recalling that $\|v\left[\mu_t^k\right] - v\left[\mu_{n\Delta t}^k\right]\|_{C_0} \le (t - n\Delta t)(Mm + P)$.

We have thus proved the first part of the following proposition

Proposition 7. *The measure $\bar{\mu}_t := \lim_k \mu_t^k$ is a weak solution of (1). Moreover, $\bar{\mu}_t \in \mathcal{M}_0^{ac}$.*

Proof. We have to prove that $\bar{\mu}_t \in \mathcal{M}_0^{ac}$. Define the non-autonomous vector field $v_t := v\left[\bar{\mu}_t\right]$ and the time-dipendent measure $h_t := h\left[\mu_t\right]$. Recall that $\bar{\mu}_t$ is continuous with respect to time, thus v_t is a continuous vector field with respect to time, each h_t is absolutely continuous and h_t is continuous with respect to time. The corresponding unique solution of (1) is thus in $\mathcal{M}_0^{ac}$, see Theorem 5. Hence, $\bar{\mu}_t \in \mathcal{M}_0^{ac}$.

We now prove that we have continuous dependence of the solution of (1) from the initial data. The estimate also give uniqueness of the solution of (1).

Theorem 6. *Let μ_t, ν_t be two solutions of (1), with initial data $\mu_0, \nu_0 \in \mathcal{M}_0^{ac}$ respectively. Let v, h satisfy* **(H)**. *Then*

$$W_p^{a,b}(\mu_t, \nu_t) \le e^{t\left(\frac{p+1}{p}L + 2mN + Q + 1\right)} W_p^{a,b}(\mu_0, \nu_0). \tag{16}$$

In particular, under assumption **(H)**, *the solution of (1) is unique.*

Proof. The key observation is that a solution of (1) at time $t + s$ is of the form $\mu_{t+s} = \Phi_s^{v\left[\mu_t\right]} \# \mu_t + s\, h\left[\mu_t\right] + o(s)$. Also observe that the mass of each solution satisfies $\mu_t(\mathbb{R}^d) \le \mu_0(\mathbb{R}^d) + tP \le m^p$, with $m := (\mu_0(\mathbb{R}^d) + P)^{1/p}$.

Take now v, h satisfying **(H)**, and μ_t, ν_t two solutions of (1), with initial data μ_0, ν_0 respectively (eventually coinciding). Define $\varepsilon(t) := W_p^{a,b}(\mu_t, \nu_t)$ and observe that it is a Lipschitz function, since

$$e(t + s) - e(t) = W_p^{a,b}(\mu_{t+s}, \nu_{t+s}) - W_p^{a,b}(\mu_t, \nu_t) \le$$
$$\le W_p^{a,b}(\mu_{t+s}, \mu_t) + W_p^{a,b}(\mu_t, \nu_t) + W_p^{a,b}(\nu_t, \nu_{t+s}) - W_p^{a,b}(\mu_t, \nu_t) \le 2s(Mm + P)$$

Now observe the following

$$\varepsilon(t+s) \le W_p^{a,b}\left(\Phi_s^{v[\mu_t]}\#\mu_t, \Phi_s^{v[\nu_t]}\#\nu_t\right) + sQW_p^{a,b}\left(\mu_t,\nu_t\right) + o(s) \le \left(1 + \tfrac{p+1}{p}Ls + o(s)\right)W_p^{a,b}\left(\mu_t,\nu_t\right) +$$

$$+\mu_t(\mathbb{R}^d)^{1/p}(1 + Ls/p + o(s))(s + o(s))NW_p^{a,b}\left(\mu_t,\nu_t\right) + sQW_p^{a,b}\left(\mu_t,\nu_t\right) + o(s) \le$$

$$\le \varepsilon(t) + s\left(\tfrac{p+1}{p}L + 2mN + Q\right)\varepsilon(t) + o(s) \le \varepsilon(t) + s\left(\tfrac{p+1}{p}L + 2mN + Q + 1\right)\varepsilon(t).$$

The last estimate holds for sufficiently small s. Using the integral form of the Gronwall inequality, we have (16). The uniqueness of the solution of (1) is a direct consequence.

References

1. L. AMBROSIO, W. GANGBO, Hamiltonian ODEs in the Wasserstein Space of Probability Measures, *Communications on Pure and Applied Mathematics*, Volume 61, Issue 1, pp. 18–53, 2008.
2. L. AMBROSIO, N. GIGLI, G. SAVARÉ, Gradient Flows: In Metric Spaces and in the Space of Probability Measures, Lectures in Mathematics Eth Zurich, Birkhäuser Basel, 2008.
3. A. BRESSAN, B. PICCOLI, Introduction to the Mathematical Theory of Control, AIMS Ser. Appl. Math., AIMS, Springfield, MO, 2007.
4. J.A. CARRILLO, R.M. COLOMBO, P. GWIAZDA, A. ULIKOWSKA, *Structured populations, cell growth and measure valued balance laws*, J. Differential Equations 252, pp. 3245–3277, 2012.
5. G. CRIPPA, M. LECUREUX-MERCIER, Existence and uniqueness of measure solutions for a system of continuity equations with non-local flow, *Nonlinear Differential Equations and Applications NoDEA*, Volume 20, Issue 3, pp. 523–537, 2013.
6. E. CRISTIANI, B. PICCOLI, A. TOSIN, Multiscale Modeling of Granular Flows with Application to Crowd Dynamics, *Multiscale Model. Simul.* 9, pp. 155–182, 2011.
7. L. C. EVANS, R. F. GARIEPY, *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.
8. J. EVERS, A. MUNTEAN, Modeling Micro-Macro Pedestrian Counterflow in Heterogeneous Domains, *Nonlinear Phenomena in Complex Systems*, Volume 14, Number 1, pp. 27–37, 2011.
9. A. FIGALLI, N. GIGLI, A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions, *Journal de Mathématiques Pures et Appliquées*, Volume 94, Issue 2, August 2010, pp. 107–130.
10. P. GWIAZDA, T. LORENZ, A. MARCINIAK-CZOCHRA, A nonlinear structured population model: Lipschitz continuity of measure-valued solutions with respect to model ingredients, *J. Differential Equations* 248, pp. 2703–2735, 2010.
11. P. GWIAZDA, A. MARCINIAK-CZOCHRA, Structured population equations in metric spaces, *J. Hyperbolic Differential Equations*, Vol. 7, No. 4, pp. 733–773, 2010.
12. B. MAURY, A. ROUDNEFF-CHUPIN, F. SANTAMBROGIO, A macroscopic crowd motion model of gradient flow type, *Mathematical Models and Methods in Applied Sciences* 20 (10), pp. 1787–1821, 2010.
13. B. MAURY, A. ROUDNEFF-CHUPIN, F. SANTAMBROGIO, J. VENEL, Handling congestion in crowd motion modeling, *Networks and Heterogeneous Media* 6(3): pp. 485–519, 2011.
14. B. PICCOLI, F. ROSSI, Transport equation with nonlocal velocity in Wasserstein spaces: convergence of numerical schemes, *Acta Applicandae Mathematicae*, 124, pp. 73–105, 2013.
15. B. PICCOLI, F. ROSSI, arXiv:1304.7014.
16. B. PICCOLI, A. TOSIN, Time-Evolving Measures and Macroscopic Modeling of Pedestrian Flow, *Archive for Rational Mechanics and Analysis*, Volume 199, Issue 3, pp. 707–738, 2011.
17. B. PICCOLI, A. TOSIN, Pedestrian flows in bounded domains with obstacles, *Continuum Mechanics and Thermodynamics*, Volume 21, Number 2, pp. 85–107, 2009.
18. A. TOSIN, P. FRASCA, Existence and approximation of probability measure solutions to models of collective behaviors, *Networks and Heterogeneous Media*, 6 (3), pp. 561–596, 2011.
19. C. VILLANI, Optimal Transport: Old and New, Grundlehren der mathematischen Wissenschaften, 2008.
20. C. VILLANI, Topics in Optimal Transportation, Graduate Studies in Mathematics, Vol. 58, 2003.

Department of Mathematical Sciences, Rutgers University - Camden, Camden, NJ. `piccoli@camden.rutgers.edu`

and

Aix-Marseille Univ, LSIS, 13013, Marseille, France. `francesco.rossi@lsis.org`