Measure dynamics with Probability Vector Fields and sources

September 9, 2018

Benedetto Piccoli

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

Francesco Rossi

Dipartimento di Matematica "Tullio Levi–Civita"
Università degli Studi di Padova
Padova, Italy
francesco.rossi@math.unipd.it

Abstract

We introduce a new formulation for differential equation describing dynamics of measures on an Euclidean space, that we call Measure Differential Equations with sources. They mix two different phenomena: on one side, a transport-type term, in which a vector field is replaced by a Probability Vector Field, that is a probability distribution on the tangent bundle; on the other side, a source term. Such new formulation allows to write in a unified way both classical transport and diffusion with finite speed, together with creation of mass.

The main result of this article shows that, by introducing a suitable Wasserstein-like functional, one can ensure existence of solutions to Measure Differential Equations with sources under Lipschitz conditions. We also prove a uniqueness result under the following additional hypothesis: the measure dynamics needs to be compatible with dynamics of measures that are sums of Dirac masses.

Keywords: Measure dynamics, Probability Vector Fields, Wasserstein distance, generalized Wasserstein distance

MSC2010: 35S99, 35F20, 35F25

1 Introduction

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. [3]). For a complete introduction to Wasserstein distances, see [12, 13].

This approach has at least two main limits. The first is that the use of transport equation, together with their counterpart in terms of ordinary differential equations [1, 2], does not allow to model neither mass diffusion nor concentration phenomena. The second one is that the Wasserstein distance $W_p(\mu, \nu)$ is defined only if the two measures μ, ν have the same mass, then PDEs with sources cannot be studied with such tools.

Both limits were recently overcome by a variety of contributions. The first was addressed in [9], in which a generalization of the concept of vector fields was introduced. Such tool, called Probability Vector Field (PVF in the following), allows to model concentration and diffusion phenomena in the formalism of the transport equation, then being able to translate several useful techniques from dynamical system.

The second limit was addressed by a series of papers introducing generalizations of the Wasserstein distance to measures with different masses. In [10] we defined a generalized Wasserstein distance $W^g(\mu,\nu)$, combining the standard Wasserstein and L^1 distances. In rough words, for $W^g(\mu,\nu)$ 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 $bW_p(\delta\mu,\delta\nu)$. This distance is a generalization of the so-called *flat distance*. Other generalizations of the Wasserstein distance, with the same spirit of allowing sources of mass, are studied in [5, 6, 8, 7]. As a consequence, sources of mass can be introduced in the transport equation, even when they depend on the measure itself, see [11].

The goal of this article is to define a new class of equations, which are able to describe complex dynamics in the space of measures, including mass diffusion, concentration and sources. The idea is to merge two different dynamics, already individually described in [9, 10], and couple them.
The first contribution is given by dynamics induced by Probability Vector Fields (PVF in the following), recently introduced in [9]. There, the equation

$$\dot{\mu} = V[\mu] \tag{1}$$

is considered, where $V : \mathcal{P}(\mathbb{R}^n) \to \mathcal{P}(T\mathbb{R}^n)$ is a function from the space $\mathcal{P}(\mathbb{R}^n)$ of probability measures to the space $\mathcal{P}(T\mathbb{R}^n)$ of probability measures of the tangent space $T\mathbb{R}^n$. The idea of such function is to describe the infinitesimal spreading of the mass $\mu(x)$ in a point x along the velocities described by the measure $V[\mu](x,\cdot)$ on the fiber $T_x\mathbb{R}^n$. Given the projection $\pi : T\mathbb{R}^n \to \mathbb{R}^n$ defined by $\pi(x,v) = x$, we also require $\pi\#V[\mu] = \mu$, i.e. that the projection of $V[\mu]$ from $\mathcal{P}(T\mathbb{R}^n)$ to $\mathcal{P}(\mathbb{R}^n)$ coincides with μ. This is the measure counterpart of the fact that a vector field is a section of the tangent bundle. The main contribution of [9] is to introduce conditions ensuring existence and/or uniqueness of the solution of the Cauchy problem with dynamics (1). In particular, two key tools are defined: the first is a new non-negative operator $\mathcal{W}$, based on the Wasserstein distance and enjoying some of its properties, on the space $\mathcal{P}(T\mathbb{R}^n)$. The idea is that $\mathcal{W}$ measures the cost of the minimizing transference plan on fibers, among plans whose projections are optimal on the base space. The formal definition is given in Definition 17. If one assumes that V from $\mathcal{P}(\mathbb{R}^n)$ endowed with the Wasserstein distance to $\mathcal{P}(T\mathbb{R}^n)$ endowed with $\mathcal{W}$ is Lipschitz, then there exists at least one solution to (1). The second tool is the definition of Dirac germs, that are specific choices of solutions to (1) for measures composed of Dirac deltas only. Fixed a Dirac germ for (1), then for each initial measure there exists at most one solution to (1) that is compatible with such chosen germ. In some specific but relevant cases, the coupling of Lipschitz continuity of V with the choice of a compatible Dirac germ ensures both existence and uniqueness of a solution to (1).
The second contribution is given by sources and sinks. In this case, the dynamics reads as

$$\dot{\mu} = s, \tag{2}$$

where s is a measure on the space $\mathbb{R}^n$, representing a source/sink of mass. The description of such Partial Differential Equation with a fixed source s is very classical, since the solution is clearly $\mu_t = \mu_0 + ts$. Instead, we introduced in [10] new conditions to ensure that the dynamics (2) is well posed even when the source $s[\mu]$ depends on the whole measure μ itself. The key tool is the introduction of a new distance on the space of measures with finite mass, called the generalized Wasserstein distance W^g. If s is Lipschitz with respect to this distance, then one has existence and uniqueness of the solution to the Cauchy problem with dynamics (2).

For simplicity, from now on we restrict ourselves to the space $\mathcal{M}(\mathbb{R}^d)$ of Borel measures with bounded support and finite mass. In this space, the generalized Wasserstein distance $W^g(\mu,\nu)$ is always finite, while the standard Wasserstein distance $W(\mu,\nu)$ is defined only if the masses of the two measures coincide, i.e. $\mu(\mathbb{R}^n) = \nu(\mathbb{R}^n)$. We endow the space $\mathcal{M}(\mathbb{R}^n)$ with the topology of weak convergence; this coincides with the

topology induced by the generalized Wasserstein distance, see Proposition 11 below.
We are now ready to define Measure Differential Equations with Source:

$$\dot{\mu} = V[\mu] \oplus s[\mu], \tag{3}$$

where $V[\mu]$ is a PVF $V : \mathcal{M}(\mathbb{R}^n) \to \mathcal{M}(T\mathbb{R}^n)$ and $s[\mu]$ is a source $s : \mathcal{M}(\mathbb{R}^n) \to \mathcal{M}(\mathbb{R}^n)$. The goal is to prove existence and/or uniqueness of a solution to the associated Cauchy problem, under the joint hypotheses ensuring existence and/or uniqueness for each of the dynamics (1) and (2). More precisely, we first give the definition of a solution to (3):

Definition 1 (Solution to (3)) *A solution to (3) is a continuous curve $\mu : [0,T] \to \mathcal{M}(\mathbb{R}^n)$ satisfying the following condition: for each $f \in C_c^\infty(\mathbb{R}^n)$*

- *the integral $\int_{T\mathbb{R}^n} (\nabla f(x) \cdot v)\, dV[\mu(\tau)](x,v)$ is defined for almost every $\tau \in [0,T]$;*
- *the map $\tau \to \int_{\mathbb{R}^n} f(x) ds[\mu(\tau)](x)$ belongs to $L^1([0,T])$;*
- *the map $t \to \int f\, d\mu(t)$ is absolutely continuous, and it satisfies*

$$\frac{d}{dt} \int_{\mathbb{R}^n} f\, d\mu(t) = \int_{T\mathbb{R}^n} (\nabla f(x) \cdot v)\, dV[\mu(t)](x,v) + \int_{\mathbb{R}^n} f(x) ds[\mu(t)](x) \tag{4}$$

  for almost every $t \in [0,T]$.

Such definition is pretty weak and can not allow uniqueness results, thus we are also interested in stronger properties for solutions to (3). In particular, we focus on existence of semigroups of solutions, whose definition in this setting is given below.

Definition 2 *A Lipschitz semigroup S_t of solutions to (3) is a map $S : [0,T] \times \mathcal{M}(\mathbb{R}^n) \to \mathcal{M}(\mathbb{R}^n)$ satisfying:*

1. *$S_0\mu = \mu$ and $S_{t+s}\mu = S_t S_s \mu$;*
2. *the map $t \to S_t\mu$ is a solution to (3);*
3. *for every $R, M > 0$ there exists $C = C(R,M) > 0$ such that if $\mathrm{supp}(\mu) \cup \mathrm{supp}(\nu) \subset B(0,R)$ and $\mu(\mathbb{R}^n) + \nu(\mathbb{R}^n) \leq M$, it then holds*
   - *(a) $\mathrm{supp}(S_t\mu) \subset B(0, e^{Ct}(R+M+1))$;*
   - *(b) $W^g(S_t\mu, S_t\nu) \leq e^{Ct} W^g(\mu,\nu)$;*
   - *(c) $W^g(S_t\mu, S_s\mu) \leq C|t-s|$.*

We also need to define a natural tool, merging properties of the operator $\mathcal{W}$ on $\mathcal{P}(T\mathbb{R}^n)$ with the setting of the generalized Wasserstein distance W^g on $\mathcal{M}(\mathbb{R}^n)$. Such non-negative operator, that we denote by $\mathcal{W}^g$, measures the minimal standard Wasserstein distance on the fiber between transference plans whose projections give a minimizing decomposition for the generalized Wasserstein distance on the base space. The operator is precisely defined in Section 2.5.

We are now ready to state the two main results of this article. The first deals with existence of a solution to (3), while the second focuses on uniqueness.

Theorem 3 *Consider the Measure Differential Equation with Source (3) with the following two sets of hypotheses:*

(V) *The Probability Vector Field $V : \mathcal{M}(\mathbb{R}^n) \to \mathcal{M}(T\mathbb{R}^n)$ satisfies:*

- **(V1)** *support sublinearity: there exists $C > 0$ such that for all $\mu \in \mathcal{M}(\mathbb{R}^n)$ it holds*

$$\sup_{(x,v) \in \mathrm{supp}(V[\mu])} |v| \leq C(1 + \sup_{x \in \mathrm{supp}(\mu)} |x|);$$

(V2) *Lipschitz continuity: for each $R > 0$ there exists $K = K(R) > 0$ such that* $\mathrm{supp}(\mu) \cup \mathrm{supp}(\nu) \subset B(0, R)$ *implies*

$$\mathcal{W}^g(V[\mu], V[\nu]) \leq K W^g(\mu, \nu); \tag{5}$$

(s) *The source $s : \mathcal{M}(\mathbb{R}^n) \to \mathcal{M}(\mathbb{R}^n)$ satisfies:*

(s1) *Lipschitz continuity: there exists L such that for all $\mu, \nu \in \mathcal{M}(\mathbb{R}^n)$ it holds*

$$W^g(s[\mu], s[\nu]) \leq L W^g(\mu, \nu); \tag{6}$$

(s2) *uniform boundedness of the support: there exists R such that for all $\mu \in \mathcal{M}(\mathbb{R}^n)$ it holds* $\mathrm{supp}(s[\mu]) \subset B_R(0)$.

Then, there exists a Lipschitz semigroup of solutions to (3) in the sense of Definition 2.

Theorem 4 *Consider the Measure Differential Equation with Source (3) satisfying Hypotheses* **(V1), (s)** *recalled in Theorem 3. Choose a Dirac germ γ, as in Definition 22 below. Then, there exists at most one Lipschitz semigroup compatible with γ, in the sense of Definition 29 below.*

Several corollaries about existence and/or uniqueness of the solutions to (3) can be directly derived from corresponding results about PVFs from [9]. In particular, one can observe that the uniqueness property depends on the PVF V only, and not on the source s. We then have the two following remarkable cases:

- Let $V[\mu] = \mu \otimes \delta_{v(x)}$ with v locally Lipschitz vector field with sub-linear growth. Then, (3) admits a unique Lipschitz semigroup, obtained as the limit of the discretization described in Section 3.1.
- Fix $\phi : [0, +\infty) \to \mathbb{R}$ an increasing function. In the space $\mathbb{R}$, define $V_\phi[\mu] = \mu \otimes J_\phi(x)$, where

$$J_\phi(x) = \begin{cases} \delta_{\phi(F_\mu(x))} & \text{if } F_\mu(x^-) = F_\mu(x), \\ \frac{\phi\#\left(\chi_{[F_\mu(x^-), F_\mu(x)]}\lambda\right)}{F_\mu(x) - F_\mu(x^-)} & \text{otherwise,} \end{cases}$$

  $F_\mu(x) = \mu((-\infty, x])$ is the cumulative distribution of μ, and λ is the Lebesgue measure. This choice of the PVF allows to have solutions that diffuse with finite velocities, see [9, Section 7.1] for more details. In this case, for any choice of the source s satisfying **(s)**, one has existence of a solution to (3). Even though this solution is not unique, in general, there exists a unique semigroup obtained by the limit of the discretization algorithm described in Section 3.1.

Remark 5 Observe that hypotheses in Theorem 3 are not sharp, in general. For example, in **(V2)**, the Lipschitz constant K in (5) can depend on $|\mu|$, with the only requirement of having $\sup_{m \in [0,M]} K(m) < +\infty$ for all finite M.

Similarly, condition **(s2)** can be replaced by any condition ensuring uniform boundedness of the supports, such as the existence of a radius R such that $\mathrm{supp}(\mu) \subseteq B(0, R')$ with $R' > R$ implies $\mathrm{supp}(s[\mu]) \subseteq B(0, R')$.

The structure of the article is the following. In Section 2 we fix the notation and recall main properties of the tools used later: the Wasserstein distance, the generalized Wasserstein distance and Measure Differential Equations with Probability Vector Fields. In the main Section 3, we prove the results of this paper. In Section 3.1, we prove Theorem 3 about existence of a solution to (3), while in Section 3.2, we prove Theorem 4 about uniqueness.

2 Dynamics in generalized Wasserstein Spaces

In this section, we fix the notation and define the main tools used in the rest of the article: the Wasserstein distance, the generalized Wasserstein distance and Measure Differential Equations with Probability Vector Fields.

2.1 The Wasserstein distance

We use $\mathcal{M}(\mathbb{R}^d)$ to denote the space of positive Borel regular measures with bounded support and finite mass on $\mathbb{R}^d$. Given μ, μ_1 Radon measures (i.e. positive Borel measures with locally finite mass), we write $\mu_1 \ll \mu$ if μ_1 is absolutely continuous with respect to μ, while we write $\mu_1 \leq \mu$ if $\mu_1(A) \leq \mu(A)$ for every Borel set A. We denote by $|\mu| := \mu(\mathbb{R}^d)$ the norm of μ (also called its mass). More generally, if $\mu = \mu^+ - \mu^-$ is a signed Borel measure, we define $|\mu| := |\mu^+| + |\mu^-|$.

Given a Borel map $\gamma : \mathbb{R}^d \to \mathbb{R}^d$, the push forward of a measure $\mu \in \mathcal{M}(\mathbb{R}^d)$ is defined by:

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

Note that the mass of μ is identical to the mass of $\gamma\#\mu$. Therefore, given two measures μ, ν with the same mass, one may look for γ such that $\nu = \gamma\#\mu$ and it minimizes the cost

$$I[\gamma] := |\mu|^{-1} \int |x - \gamma(x)|^p \, d\mu(x).$$

This means that each infinitesimal mass $\delta\mu$ is sent to $\delta\nu$ and that its infinitesimal cost is the p-th power of the distance between them. Such minimization problem is known as the Monge problem. A generalization of the Monge problem is defined as follows. Given a probability measure π on $\mathbb{R}^d \times \mathbb{R}^d$, one can interpret π 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:

$$|\mu| \int_{\mathbb{R}^d} d\pi(x, \cdot) = d\mu(x), \qquad |\nu| \int_{\mathbb{R}^d} d\pi(\cdot, y) = d\nu(y). \tag{7}$$

Such π is called a transference plan from μ to ν and we denote the set of transference plans from μ to ν by $P(\mu, \nu)$. A condition equivalent to (7) is that, for all $f, g \in C_c^\infty(\mathbb{R}^d)$ it holds $|\mu| \int_{\mathbb{R}^d \times \mathbb{R}^d} (f(x) + g(y)) \, d\pi(x, y) = \int_{\mathbb{R}^d} f(x) \, d\mu(x) + \int_{\mathbb{R}^d} g(y) \, d\nu(y)$.

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 $P(\mu, \nu)$. Such problem is called the Monge-Kantorovich problem. It is important to observe that such problem is a generalization of the Monge problem. The main advantage of this approach is that a minimizer of J in $P(\mu, \nu)$ always exists. We then denote by $P^{opt}(\mu, \nu)$ the set of transference plans that are minimizers of J, that is always non-empty.

One can thus define on $\mathcal{M}_1(\mathbb{R}^d)$ the following operator between measures of the same mass, called the **Wasserstein distance**:

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

It is indeed a distance on the subspace of measures in $\mathcal{M}(\mathbb{R}^d)$ with a given mass, see [12]. It is easy to prove that $W_p(k\mu, k\nu) = W_p(\mu, \nu)$ for $k \geq 0$, by observing that $P(k\mu, k\nu) = P(\mu, \nu)$ and that $J[\pi]$ does not depend on the mass.

From now on, we only consider the Wasserstein distance with parameter $p = 1$, that will then be denoted by $W(\mu, \nu)$. It satisfies the following fundamental dual property.

Proposition 6 *[Kantorovich-Rubinstein duality] Let $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^d)$. It then holds*

$$W(\mu, \nu) = \sup \left\{ \int f d(\mu - \nu) \quad s.t. \quad Lip(f) \leq 1 \right\} \tag{8}$$

Such property plays a crucial role in the theory of PVF, see [9]. It is then unclear if a corresponding theory can be generalized to any $p > 1$.

2.2 The generalized Wasserstein distance

In this section, we provide a definition of the generalized Wasserstein distance, introduced in [10, 11], together with some useful properties. We consider here the generalized Wasserstein distance with parameters $a = b = 1$, to simplify the notation, and $p = 1$.

Definition 7 *Let $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$ be two measures. We define the functional*

$$W^g(\mu, \nu) := \inf_{\tilde{\mu}, \tilde{\nu} \in \mathcal{M}(\mathbb{R}^d),\, |\tilde{\mu}| = |\tilde{\nu}|} |\mu - \tilde{\mu}| + |\nu - \tilde{\nu}| + W(\tilde{\mu}, \tilde{\nu}). \tag{9}$$

We now provide some properties of W^g. Proofs can be adapted from those given in [10].

Proposition 8 *The following properties hold:*
1. The infimum in (9) coincides with

$$\inf_{\tilde{\mu} \leq \mu, \tilde{\nu} \leq \nu,\, |\tilde{\mu}| = |\tilde{\nu}|} |\mu - \tilde{\mu}| + |\nu - \tilde{\nu}| + W(\tilde{\mu}, \tilde{\nu}),$$

where we have added the constraint $\tilde{\mu} \leq \mu, \tilde{\nu} \leq \nu$.
2. The infimum in (9) is attained by some $\tilde{\mu}, \tilde{\nu}$.
3. The functional W^g is a distance on $\mathcal{M}(\mathbb{R}^d)$.
4. It holds $W^g(\mu, 0) \leq |\mu|$.
5. It holds

$$||\mu| - |\nu|| \leq W^g(\mu, \nu). \tag{10}$$

6. If $|\mu| = |\nu|$, it holds

$$W^g(\mu, \nu) \leq W(\mu, \nu) \tag{11}$$

We recall now some useful topological results related to the metric space $\mathcal{M}(\mathbb{R}^d)$ when endowed with the generalized Wasserstein distance. We first recall the definition of tightness in this context.

Definition 9 *A set of measures M is tight if for each $\varepsilon > 0$ there exists a compact K_ε such that $\mu(\mathbb{R}^d \backslash K_\varepsilon) < \varepsilon$ for all $\mu \in M$.*

We now recall the definition of weak convergence of measures, as well as an important result about convergence with respect to the generalized Wasserstein distance, see [10, Theorem 13].

Definition 10 *Let $\{\mu_n\}$ be a sequence of measures in $\mathbb{R}^d$, and μ a measure. We say that μ^n converges to μ with respect to the weak topology, and we write $\mu_n \rightharpoonup \mu$, if for all functions $f \in C_c^\infty$ it holds*

$$\lim_{n \to \infty} \int f \, d\mu_n = \int f \, d\mu.$$

Proposition 11 *Let $\{\mu_n\}$ be a sequence of measures in $\mathbb{R}^d$, and $\mu \in \mathcal{M}(\mathbb{R}^d)$. Then*

$$W^g(\mu_n, \mu) \to 0 \qquad \textit{is equivalent to} \qquad \mu_n \rightharpoonup \mu \;\; \textit{and} \;\; \{\mu_n\} \textit{ is tight.}$$

We also recall the result of completeness, see [10, Proposition 15].

Proposition 12 *The space $\mathcal{M}(\mathbb{R}^d)$ endowed with the distance W^g is a complete metric space.*

The generalized Wasserstein distance also satisfies a useful dual formula, showing that it coincides with the so-called flat distance. See [11].

Proposition 13 *Let $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$. It then holds*

$$W^g(\mu,\nu) = \sup\left\{\int f d(\mu-\nu) \quad s.t. \quad \|f\|_C^0 \le 1,\, Lip(f) \le 1\right\} \tag{12}$$

We recall that the L^1 distance satisfies a dual formula too, that is

$$|\mu-\nu| = \sup\left\{\int f d(\mu-\nu) \quad \text{s.t.} \quad \|f\|_C^0 \le 1\right\} \tag{13}$$

We also have this useful estimate to bound integrals. See [10].

Lemma 14 *Let $f \in C^0(\mathbb{R}^d) \cap \mathrm{Lip}(\mathbb{R}^d)$. It then holds*

$$\int f d(\mu-\nu) \le \max\left\{\|f\|_{C^0}, \mathrm{Lip}(f)\right\} W^g(\mu,\nu). \tag{14}$$

We end this section by giving useful estimates both for the standard and generalized Wasserstein distances W_p and W^g under flow actions. Proofs are given in [10, 11].

Proposition 15 *Let v_t, w_t be two time-varying vector fields, uniformly Lipschitz with respect to the space variable, and ϕ^t, ψ^t the flow generated by v, w respectively. Let L be the Lipschitz constant of v and w, i.e. $|v_t(x) - v_t(y)| \le L|x-y|$ for all t, and similarly for w. Let $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$. We have the following estimates for the standard Wasserstein distance*

- $W_p\left(\phi^t\#\mu, \phi^t\#\nu\right) \le e^{Lt} W_p\left(\mu,\nu\right)$,
- $W_p\left(\mu, \phi^t\#\mu\right) \le t\|v\|_{C^0}|\mu|$,
- $W_p\left(\phi^t\#\mu, \psi^t\#\nu\right) \le e^{Lt} W_p\left(\mu,\nu\right) + \frac{e^{Lt}-1}{L}|\mu| \sup_{\tau\in[0,t]} \|v_t - w_t\|_{C^0}$.

We have the following estimates for the generalized Wasserstein distance

- $W^g(\phi^t\#\mu, \phi^t\#\nu) \le e^{Lt} W^g(\mu,\nu)$,
- $W^g(\mu, \phi^t\#\mu) \le t\|v\|_{C^0}|\mu|$,
- $W^g(\phi^t\#\mu, \psi^t\#\nu) \le e^{Lt} W^g(\mu,\nu) + \frac{e^{Lt}-1}{L}|\mu| \sup_{\tau\in[0,t]} \|v_t - w_t\|_{C^0}$.

2.3 Measure Differential Equations with Probability Vector Fields

In this section, we summarize the main results and tools about PVFs, introduced in [9]. We slightly enlarge the setting of [9], since we consider general measures with finite mass and not only probability measures.

We first recall the definition of a solution to the Cauchy problem

$$\dot{\mu} = V[\mu], \qquad \mu(0) = \mu_0. \tag{15}$$

Definition 16 *Fix a final time $T > 0$. A solution to (15) is a map $\mu : [0,T] \to \mathcal{M}(\mathbb{R}^n)$ such that $\mu(0) = \mu_0$ and the following holds: for each $f \in C_c^\infty(\mathbb{R}^n)$*

- *the integral $\int_{T\mathbb{R}^n} (\nabla f(x)\cdot v)\, dV[\mu(s)](x,v)$ is defined for almost every $s \in [0,T]$;*
- *the map $s \to \int_{T\mathbb{R}^n} (\nabla f(x)\cdot v)\, dV[\mu(s)](x,v)$ belongs to $L^1([0,T])$;*
- *the map $t \to \int f\, d\mu(t)$ is absolutely continuous, and it satisfies*

$$\frac{d}{dt}\int_{\mathbb{R}^n} f\, d\mu(t) = \int_{T\mathbb{R}^n} (\nabla f(x)\cdot v)\, dV[\mu(t)](x,v) \tag{16}$$

for almost every $t \in [0,T]$.

We now recall the definition of the pseudo-distance $\mathcal{W}$, that will be useful in the following.

Definition 17 *Let $V_1, V_2 \in \mathcal{M}(T\mathbb{R}^n)$ with $|V_1| = |V_2|$. Denote by $\mu_1 = \pi_1 \# V_1$ and $\mu_2 = \pi_1 \# V_2$ the projection of the PVF on the base space. Define*

$$\mathcal{W}(V_1, V_2) := \inf \left\{ \int_{T\mathbb{R}^n \times T\mathbb{R}^n} |v - w| \, dp(x, v, y, w) \ \text{ such that } p \in P(V_1, V_2) \text{ and } \pi_{13} \# p \in P^{opt}(\mu_1, \mu_2) \right\}.$$

Clearly, such functional is not a distance, see examples in [9]. Nevertheless, we will see in the following that the local Lipschitz condition **(V2)** will ensure existence of solutions to (15). Observe that it also holds

$$W(V_1, V_2) \leq \mathcal{W}(V_1, V_2) + W(\pi_1 \# V_1, \pi_1 \# V_2). \tag{17}$$

See [9] for more details.

We now address the problem of existence of solutions to (15). The idea developed in [9] is to define a semigroup of solutions as the limit of approximated ones. We first describe precisely the discretization method, that will be also useful in the following.

Definition 18 *Fix $N \in \mathbb{N}$ and define the time step size $\Delta_N = \frac{1}{N}$, the velocity step size $\Delta_N^v = \frac{1}{N}$ and the space step size $\Delta_N^x = \Delta_N^v \Delta_N = \frac{1}{N^2}$. Define x_i the $(2N^3 + 1)^n$ equispaced discretization points of $(\mathbb{Z}^n/(N^2)) \cap [-N, N]^n$, and v_j the $(2N^2 + 1)^n$ equispaced discretization points of $(\mathbb{Z}^n/N) \cap [-N, N]^n$.*

Define $\mathcal{M}_N^x \subset \mathcal{M}(\mathbb{R}^n)$ the space of measures of $\mathbb{R}^n$ with support on the set of points x_i, and $\mathcal{M}_N^v \subset \mathcal{M}(\mathbb{R}^{2n})$ the space of measures of $\mathbb{R}^{2n}$ with support on the set of points (x_i, v_j),

Define the discretization operator in the space variable $A_N^x : \mathcal{M}(\mathbb{R}^n) \to \mathcal{M}_N^x$ as follows

$$\mathcal{A}_N^x(\mu) := \sum_i m_i^x(\mu) \delta_{x_i},$$

where $m_i^x(\mu) := \mu(x_i + Q)$ with $Q = \left[0, \frac{1}{N^2}\right)^n$. Define the discretization operator in the velocity variable $A_N^v : \mathcal{M}(\mathbb{R}^{2n}) \to \mathcal{M}_N^v$ as follows

$$\mathcal{A}_N^v(V) := \sum_{i,j} m_{ij}^v(V) \delta_{(x_i, v_j)}, \tag{18}$$

where $m_{ij}^v(V) := V((x_i + Q) \times (v_j + Q'))$ with $Q' = \left[0, \frac{1}{N}\right)^n$.

The first property of such discretization is that it introduces an arbitrarily small error in the Wasserstein distance.

Proposition 19 *Given $\mu \in \mathcal{M}_c(\mathbb{R}^n)$ and $V[\mu] \in \mathcal{M}_c(T\mathbb{R}^n)$, for a sufficiently large N it holds*

$$W(\mu, \mathcal{A}_N^x(\mu)) \leq |\mu| \Delta_N^x, \qquad W(V[\mu], \mathcal{A}_N^v(V[\mu]) \leq |\mu| \Delta_N^v.$$

Proof. The proof with μ and $V[\mu]$ being probability measures is given in [9]. The generalization to measures with finite mass is straightforward. □

One can then define an approximated solution (called the Lattice Approximate Solution) to (15) via an explicit Euler scheme.

Definition 20 *Given the Cauchy problem (15), we define the following Lattice Approximate Solution μ^N: we set $\mu^N(0) := \mathcal{A}_N^x(\mu_0)$, then recursively*

$$\mu^N((k+1)\Delta_N) := \sum_{i,j} m_{ij}^v(V[\mu^N(k\Delta_N)]) \delta_{x_i + \Delta_N v_j},$$

and for intermediate times $t \in [0, \Delta_N)$ we define

$$\mu^N(k\Delta_N + t) := \sum_{i,j} m_{ij}^v(V[\mu^N(k\Delta_N)]) \delta_{x_i + t v_j}.$$

We are now ready to state the existence of a solution to (15) as a limit of the Lattice Approximate Solutions introduced above.

Theorem 21 *Let a PVF $V : \mathcal{M}(\mathbb{R}^n) \to \mathcal{M}(T\mathbb{R}^n)$ be given, satisfying* **(V)** *where $\mathcal{W}^g$ is replaced by $\mathcal{W}$. Then, there exists a Lipschitz semigroup of solutions to (15), obtained as uniform-in-time limit of Lattice Approximate Solutions for the Wasserstein Metric.*

Proof. The first key observation is that both A_N^x and A_N^v are operators preserving the mass for N sufficiently large, i.e. $\mathcal{A}_N^x(\mu)(\mathbb{R}^n) = \mu(\mathbb{R}^n)$ and similarly for the PVF. As a consequence, the mass of $\mu^N(t)$ coincides with $\mu^N(0)$, that in turn coincides with μ_0 for N sufficiently large.

If $\mu_0(\mathbb{R}^n) = 1$, then the whole sequence $\mu^N(t)$ is in $\mathcal{P}_c(\mathbb{R}^n)$, and one can apply the proof of [9, Theorem 4.1]. Otherwise, rescale the mass by defining $\nu^N(t) = \frac{1}{\mu_0(\mathbb{R}^N)}\mu^N(t)$, apply the previous case to define $\nu(t)$ and prove that $\mu(t) = \mu_0(\mathbb{R}^N)\nu(t)$ is a solution to (15). □

We now recall the definition of Dirac germs, that permits to address the problem of uniqueness of the solution to (15). We also give the definition of semigroup compatible with the germ.

Definition 22 *Fix a PVF V. Define $\mathcal{M}^D := \{\mu \in \mathcal{M}(\mathbb{R}^n)$ such that $\mu = \sum_{l=1}^m m_l \delta_{x_l}\}$ the space of measures composed of Dirac deltas. A Dirac germ γ compatible with V is a map assigning to each $\mu \in \mathcal{M}^D$ a Lipschitz curve $\gamma_\mu : [0, \varepsilon(\mu)] \to \mathcal{M}(\mathbb{R}^n)$, with the following conditions:*

- *$\varepsilon(\mu) > 0$ is uniformly positive for measures with uniformly bounded support;*
- *γ_μ is a solution to (1).*

Definition 23 *Fix a PVF V satisfying* **(V1)**, *a final time $T > 0$ and a Dirac germ γ. A semigroup for (1) is said to be compatible with γ if one has the following property: for each $R, M > 0$ there exists $C(R, M)$ such that the space $\mathcal{M}^D_{R,M} := \left\{\mu \in \mathcal{M}^D \quad s.t. \ \operatorname{supp}(\mu) \in B(0, R), |\mu| \le M\right\}$ satisfies*

$$\text{for all } t \in [0, \inf_{\mu \in \mathcal{M}^D_{R,M}} \varepsilon(\mu)] \text{ one has } \sup_{\mu \in \mathcal{M}^D_R} W(S_t\mu, \gamma_\mu(t)) \le C(R, M)t^2. \tag{19}$$

We are now ready to prove the main result about uniqueness of solutions to (15).

Theorem 24 *Consider a PVF satisfying (V1) and fix a Dirac germ γ. There exists at most one Lipschitz semigroup S_t of solutions to (15) compatible with γ.*

Proof. First observe that uniform boundedness of the support and the weak formulation (16) when choosing $f = 1$ on $\cup_{t\in[0,T]}\operatorname{supp}(\mu(t))$ imply that the mass $\mu(t)(\mathbb{R}^n)$ is constant along trajectories of (1). Thus, the Dirac germ satisfies conservation of mass too.

Apply now the proof of Theorem 5.1 in [9] for an initial data being a probability measure, with the Dirac germ restricted to probability measures. For initial data with general finite mass, apply the rescaling trick described in the proof of Theorem 21 both to the initial data and the Dirac germ. □

2.4 Measure Equations with sources

In this section, we briefly study the measure equation with source

$$\dot{\mu} = s[\mu], \qquad \mu(0) = \mu_0. \tag{20}$$

The goal is to prove that condition **(s)** in Theorem 3 ensures existence and uniqueness of a solution to (20). This is indeed a particular case of a more general result, stated in [10], in which a transport term is added too. For our future use, we prove the statement with the same discretization method of Lattice Approximate Solution introduced in Definition 18.

Proposition 25 *Fix $T > 0$. Let the source $s : \mathcal{M}(\mathbb{R}^n) \to \mathcal{M}(\mathbb{R}^n)$ satisfy Hypotheses* **(s)** *in Theorem 3. Then, there exists a unique solution to (20).*

Moreover, such solution is the uniform-in-time Wasserstein limit for $N \to \infty$ of Lattice Approximate Solutions $\mu^N : [0,T] \to \mathcal{M}(\mathbb{R}^n)$ defined as follows: Define $\mu^N(0) := \mathcal{A}^x_N(\mu_0) = \sum_i m^x_i(\mu_0)\delta_{x_i}$, then recursively

$$\mu^N((k+1)\Delta_N) = \mu^N(k\Delta_N) + \Delta_N \mathcal{A}^x_N(s[\mu^N(k\Delta_N)]). \tag{21}$$

We also define the time-interpolated solution for $t \in [0, \Delta_N]$ as follows: $\mu^N(k\Delta_N + t) = \mu^N(k\Delta_N) + t\mathcal{A}^x_N(s[\mu(k\Delta_N)])$.

Proof. We first prove existence of a solution, based on the Lattice Approximate Solution. We prove that μ^N is a sequence of equi-Lipschitz and equi-bounded curves in $C^0([0,T], \mathcal{M}(X))$, where X is a compact subset of $\mathbb{R}^n$ and the space $\mathcal{M}(X)$ is endowed with the generalized Wasserstein distance W^g. For $\tau, \sigma \in [0, \Delta_N]$ it holds

$$\begin{aligned} &W^g(\mu^N(k\Delta_N + \tau), \mu^N(k\Delta_N + \sigma)) = W^g(\tau A^x_N(s[\mu^N(k\Delta_N)]), \sigma A^x_N(s[\mu^N(k\Delta_N)])) \leq \\ &|\tau - \sigma| \left|A^x_N(s[\mu^N(k\Delta_N)])\right| \leq |\tau - \sigma| \left|s[\mu^N(k\Delta_N)]\right|. \end{aligned} \tag{22}$$

We are then left to prove that $\left|s[\mu^N(k\Delta_N)]\right|$ is uniformly bounded for $k\Delta_N \in [0,T]$. It is sufficient to observe that (10) and hypothesis **(s1)**, together with (22), imply

$$\left|s[\mu^N((k+1)\Delta_N)]\right| \leq \left|s[\mu^N(k\Delta_N)]\right| + W^g(s[\mu^N(k\Delta_N + \tau)], s[\mu^N(k\Delta_N + \sigma)]), \leq (1 + L\Delta_N)\left|s[\mu^N(k\Delta_N)]\right|.$$

hence recursively $\left|s[\mu^N((k+1)\Delta_N)]\right| \leq e^{LT}\left|s[\mu^N(0)]\right| \leq e^{LT}|s[\mu_0]|$.

We now prove that there exists R' such that $\mathrm{supp}(\mu^N(t)) \subset B(0, R')$ for all N and $t \in [0,T]$. Eventually enlarging the radius R given in hypothesis **(s2)**, one can assume that $\mathrm{supp}(\mu_0) \subset B(0,R)$. Thus, the approximation operator $\mathcal{A}^x_N$ satisfies $\mathrm{supp}(A^x_N(\mu_0)) \subset B(0, R+1)$, as well as $\mathrm{supp}(A^x_N(s[\mu])) \subset B(0, R+1)$ for any $\mu \in \mathcal{M}(\mathbb{R}^n)$. Since sum of measures with the same support gives a measure with the same support, one can easily prove by induction that measures $\mu^N(k\Delta_N + \tau)$ defined by the scheme (21) all have support contained in $B(0, R')$ with $R' = R + 1$.

Choose now $X = \overline{B(0, R')}$, that is a compact space. Then, $\mathcal{M}(X)$ is complete when endowed with the generalized Wasserstein distance W^g, see Proposition 12. The sequence μ^N is equi-Lipschitz in $\mathcal{M}(X)$, due to (22), and equi-bounded, since masses are equi-bounded. Then, there exists a converging subsequence, converging to some $\mu^* \in C^0([0,T], \mathcal{M}(X))$. Recall that such convergence with respect to W^g coincides with weak convergence of measures.

We now prove that μ^* is a solution to (20). We first observe that $W^g(\mu^N(0), \mu_0) \leq |\mu_0|\Delta^x_N$ for N sufficiently large implies $\mu^*(0) = \mu_0$. We now prove that for all $f \in C^\infty_c(\mathbb{R}^n)$ and $\tau, \sigma \in [0,T]$ with $\sigma > \tau$, it holds

$$\int f(x)\, d(\mu^*(\sigma) - \mu^*(\tau)) - \int_\tau^\sigma dt \int f(x)\, ds[\mu^*(t)] = 0 \tag{23}$$

The definition (21) implies that, for $\sigma, \tau \in [0, \Delta_N]$, it holds

$$\int f(x)\, d(\mu^N(k\Delta_N + \sigma) - \mu^*(k\Delta_N + \tau)) = (\sigma - \tau) \int f(x)\, dA^x_N(s[\mu^N(k\Delta_N)]). \tag{24}$$

We then have

$$\begin{aligned} &\left|\int f(x)\, d(\mu^*(\sigma) - \mu^*(\tau)) - \int_\tau^\sigma dt \int f(x)\, ds[\mu^*(t)]\right| \leq \left|\int f(x)\, d(\mu^*(\sigma) - \mu^N(\sigma))\right| + \\ &\left|\int f(x)\, d(\mu^*(\tau) - \mu^N(\tau))\right| + \int_\tau^\sigma dt \left|\int f(x)\, d(s[\mu^N(k^t_N\Delta_N)] - s[\mu^*(t)])\right| + \\ &\left|\int f(x)\, d(\mu^N(\sigma) - \mu^N(\tau)) - \int_\tau^\sigma dt \int f(x)\, ds[\mu^N(k^t_N\Delta_N)]\right|, \end{aligned} \tag{25}$$

where k_N^t is the largest integer such that $t \geq k\Delta_N$, i.e. $k_N^t = \lfloor \frac{t}{\Delta_N} \rfloor$. The first two terms converge to zero since $\mu^N \rightharpoonup \mu^*$, while the last term is identically zero due to (24). For the third term, observe that it holds

$$\left| \int f(x)\, d(s[\mu^N(k_N^t \Delta_N)] - s[\mu^*(t)]) \right| \leq \|f\|_{C^1} L \left(W^g(\mu^N(k_N^t \Delta_N), \mu^N(t)) + W^g(\mu^N(t), \mu^*(t)) \right), \tag{26}$$

where we used condition **(s1)** about the Lipschitz continuity of s, as well as the dual formulation (14) for W^g. The proof now follows from observing that both the terms in the right hand side of (26) converge to zero: the first satisfies

$$W^g(\mu^N(k_N^t \Delta_N), \mu^N(t)) \leq K' |t - k_N^t \Delta_N| \leq K' \Delta_N \to 0$$

for the constant $K' = e^{LT} |s[\mu_0]|$ given by (22). The second converges to zero since W^g metrizes weak convergence.

We now prove uniqueness of the solution, by proving continuous dependence on the initial data for (20). Consider two solutions $\mu(t), \nu(t)$ to (20) with initial data μ_0, ν_0 respectively. Using the the weak formulation of (20), it holds

$$\int f(x)\, d(\mu(t) - \nu(t)) = \int f(x)\, d(\mu_0 - \nu_0) + \int_0^t d\tau f(x)\, d(s[\mu(\tau)] - s[\nu(\tau)]). \tag{27}$$

Choose now a sequence f_n with $\|f_n\|_C^0 \leq 1$, $Lip(f_n) \leq 1$ realizing $W^g(\mu(t), \nu(t))$ in its dual formulation (12). Then, equation (27) reads as $W^g(\mu(t), \nu(t)) \leq W^g(\mu_0, \nu_0) + \int_0^t L W^g(\mu(\tau), \nu(\tau))$, where we used Lipschitz continuity of s. Since both $\mu(t), \nu(t)$ are Lipschitz with respect to time, a direct application of the Gronwall lemma implies $W^g(\mu(t), \nu(t)) \leq e^{Lt} W^g(\mu_0, \nu_0)$, that in turn implies uniqueness of the solution to (20). □

2.5 The operator $\mathcal{W}^g$

In this section, we define the operator $\mathcal{W}^g$. The Lipschitz condition **(V2)** with respect to such operator will be crucial to ensure existence of a solution to (3). Then we can define:

Definition 26 *Let $V_1, V_2 \in \mathcal{M}(T\mathbb{R}^n)$. Denote by $\mu_1 = \pi_1 \# V_1$ and $\mu_2 = \pi_1 \# V_2$ the projection of the PVF on the base space. For each pair $(\tilde{V}_1, \tilde{V}_2)$ satisfying $\tilde{V}_1 \leq V_1$ and $\tilde{V}_2 \leq V_2$, denote by $\tilde{\mu}_i = \pi_1 \# \tilde{V}_i$. Define*

$$\begin{aligned} \mathcal{W}^g(V_1, V_2) \quad := \quad & \inf \Big\{ \int_{T\mathbb{R}^n \times T\mathbb{R}^n} |v - w|\, dp(x, v, y, w) \ \text{ such that } \tilde{V}_1 \leq V_1,\ \tilde{V}_2 \leq V_2, p \in P(\tilde{V}_1, \tilde{V}_2), \\ & W^g(\mu_1, \mu_2) = |\mu_1 - \tilde{\mu}_1| + W(\tilde{\mu}_1, \tilde{\mu}_2) + |\mu_2 - \tilde{\mu}_2| \ \text{ and } \pi_{13} \# p \in P^{opt}(\tilde{\mu}_1, \tilde{\mu}_2) \Big\}. \end{aligned} \tag{28}$$

In the definition above, one can observe that the condition $W^g(\mu_1, \mu_2) = |\mu_1 - \tilde{\mu}_1| + W(\tilde{\mu}_1, \tilde{\mu}_2) + |\mu_2 - \tilde{\mu}_2|$ is equivalent to state that $\tilde{\mu}_1, \tilde{\mu}_2$ is a minimizer in Definition 7.

Remark 27 One might require the minimization of the functional $|V_1 - \tilde{V}_1| + \int_{T\mathbb{R}^n \times T\mathbb{R}^n} |v - w|\, dp(x, v, y, w) + |V_2 - \tilde{V}_2|$ in (28), that seems more close to the definition of W^g. Nevertheless, recall that that $|V_i - \tilde{V}_i| = |\mu_i - \tilde{\mu}_i|$. As a consequence, when the choice of the minimizer for $W^g(\mu_1, \mu_2)$ is unique, there is no difference for the minimization of the two functionals. When minimizers for $W^g(\mu_1, \mu_2)$ are not unique, this would introduce two additional terms $|V_i - \tilde{V}_i|$ in the right hand side of (29), thus providing a less restricitve inequality.

Moreover, the chosen definition of $\mathcal{W}^g$ is correct to prove existence of a solution to (3), that is the main goal of this paper. This definition would indeed be natural in the estimate (51) below.

When two measures μ_1, μ_2 have the same mass $|\mu_1| = |\mu_2|$ and have sufficiently near supports, one can choose $\tilde{\mu}_i = \mu_i$ among the minimizers of $W^g(\mu_1, \mu_2)$. Moreover, if $\mu_1 \perp \mu_2$, i.e. the two measures have no shared mass, such choice is the unique minimizer. In this case, the choice $\tilde{V}_i = V_i$ is unique too, and the operator $\mathcal{W}^g$ coincides with $\mathcal{W}$.

Similarly to $\mathcal{W}$, the operator $\mathcal{W}^g$ is then not a distance: the same counterexamples with sufficiently near supports and no shared mass can be found. For example, choose $\mu_1 = \delta_0, \mu_2 = \delta_\epsilon, V_1 = \delta_0 \otimes \delta_0, V_2 = \delta_\epsilon \otimes \delta_0$, with $\epsilon > 0$ sufficiently small. The unique minimizer in (28) is given by $\tilde{\mu}_i = \mu_i$ and $\tilde{V}_i = V_i$, that in turn gives $\mathcal{W}^g(V_1, V_2) = 0$, even though $V_1 \neq V_2$.

Again, similarly to the estimate (17) between the standard generalized Wasserstein and the operator $\mathcal{W}$, one can easily prove the following proposition.

Proposition 28 *Given $V_1, V_2 \in \mathcal{M}(T\mathbb{R}^n)$ two PVFs, it holds*

$$W^g(V_1, V_2) \leq \mathcal{W}^g(V_1, V_2) + W^g(\pi_1 \# V_1, \pi_1 \# V_2). \tag{29}$$

Proof. First estimate $W^g(V_1, V_2)$ from above by choosing $\tilde{V}_1, \tilde{V}_2$ minimizers of $\mathcal{W}^g(V_1, V_2)$ in the right-hand side of (9). Similarly, estimate the Wasserstein distance $W(\tilde{V}_1, \tilde{V}_2)$ from above by using the transference plan p realizing $\mathcal{W}^g(V_1, V_2)$. It then holds

$$\begin{aligned} W^g(V_1, V_2) &\leq |V_1 - \tilde{V}_1| + W(\tilde{V}_1, \tilde{V}_2) + |V_2 - \tilde{V}_2| \leq |\mu_1 - \tilde{\mu}_1| + \mathcal{W}(\tilde{V}_1, \tilde{V}_2) + W(\tilde{\mu}_1, \tilde{\mu}_2) + |\mu_2 - \tilde{\mu}_2| = \\ &= W^g(\mu_1, \mu_2) + \mathcal{W}(\tilde{V}_1, \tilde{V}_2). \end{aligned}$$

We used here that $|V_i - \tilde{V}_i| = |\mu_i - \tilde{\mu}_i|$ since $\tilde{V}_i \leq V_i$ and $\tilde{\mu}_i \leq \mu_i$. We also used (17) and the fact that $\tilde{\mu}_1, \tilde{\mu}_2$ is a minimizer for $W^g(\mu_1, \mu_2)$.

We are now left to prove that $\mathcal{W}(\tilde{V}_1, \tilde{V}_2) = \mathcal{W}^g(\tilde{V}_1, \tilde{V}_2)$. It clearly holds $\mathcal{W}(\tilde{V}_1, \tilde{V}_2) \geq \mathcal{W}^g(\tilde{V}_1, \tilde{V}_2)$, since the minimization on the right hand side takes place in a larger space. By contradiction, if a strict inequality holds true, there exists a decomposition $(\hat{V}_1, \hat{V}_2)$ satisfying $\hat{V}_i < \tilde{V}_i$ and minimizing $\mathcal{W}^g(\tilde{V}_1, \tilde{V}_2)$. This means that there exists $q \in P(\hat{V}_1, \hat{V}_2)$ such that

$$\int_{T\mathbb{R}^n \times T\mathbb{R}^n} |v - w| \, dq(x, v, y, w) < \mathcal{W}(\tilde{V}_1, \tilde{V}_2)$$

and, by defining $\hat{\mu}_i = \pi_1 \# \hat{V}_i$, it holds

$$W^g(\tilde{\mu}_1, \tilde{\mu}_2) = |\tilde{\mu}_1 - \hat{\mu}_1| + W(\hat{\mu}_1, \hat{\mu}_2) + |\tilde{\mu}_2 - \hat{\mu}_2|,$$

with $\pi_{13} \# q \in P^{opt}(\hat{\mu}_1, \hat{\mu}_2)$. Also recall that $W^g(\tilde{\mu}_1, \tilde{\mu}_2) \leq W(\tilde{\mu}_1, \tilde{\mu}_2)$, by (11).

Observe now that $\hat{V}_1, \hat{V}_2$ is a possible decomposition to estimate $\mathcal{W}^g(V_1, V_2)$ in (28), with transference plan q. Indeed, it first holds $\hat{V}_i \leq \tilde{V}_i \leq V_i$, thus $|\mu_i - \hat{\mu}_i| = |\mu_i - \tilde{\mu}_i| + |\tilde{\mu}_i - \hat{\mu}_i|$. This implies that

$$\begin{aligned} &|\mu_1 - \hat{\mu}_1| + W(\hat{\mu}_1, \hat{\mu}_2) + |\mu_2 - \hat{\mu}_2| = \\ &|\mu_1 - \tilde{\mu}_1| + |\tilde{\mu}_1 - \hat{\mu}_1| + W(\hat{\mu}_1, \hat{\mu}_2) + |\tilde{\mu}_2 - \hat{\mu}_2| + |\mu_2 - \tilde{\mu}_2| = \\ &|\mu_1 - \tilde{\mu}_1| + W^g(\tilde{\mu}_1, \tilde{\mu}_2) + |\mu_2 - \tilde{\mu}_2| \leq |\mu_1 - \tilde{\mu}_1| + W(\tilde{\mu}_1, \tilde{\mu}_2) + |\mu_2 - \tilde{\mu}_2| = W^g(\mu_1, \mu_2), \end{aligned}$$

i.e. the decomposition $\hat{\mu}_1, \hat{\mu}_2$ realizes the minimizer of $W^g(\mu_1, \mu_2)$. Since $\pi_{13} \# q \in P^{opt}(\hat{\mu}_1, \hat{\mu}_2)$, one can write by the contradiction hypothesis that it holds

$$\int_{T\mathbb{R}^n \times T\mathbb{R}^n} |v - w| \, dq(x, v, y, w) = \mathcal{W}^g(\tilde{V}_1, \tilde{V}_2) < \mathcal{W}(\tilde{V}_1, \tilde{V}_2) = \mathcal{W}^g(V_1, V_2).$$

This contradicts the definition of $\mathcal{W}^g$ as the infimum of the functional $\int_{T\mathbb{R}^n \times T\mathbb{R}^n} |v - w| \, dp(x, v, y, w)$. □

3 Proof of the main theorems

In this section, we prove the main results of this article, that are Theorem 3 about existence of solutions to (3) and Theorem 4 about uniqueness.

3.1 Existence - Proof of Theorem 3

In this section, we prove Theorem 3 about existence of solutions to (3). The idea is to define Lattice Approximate Solutions described in Definition 20, then pass to the limit. This procedure proved useful for each term in (3) separately, namely the PVF studied in (1) and the source in (2).

Proof of Theorem 3. We first fix an initial data μ_0 and prove the existence of a solution to (3) with initial data μ_0, that has bounded support and is Lipschitz with respect to time. This corresponds to prove Properties 1-2-3a-3c in the Definition 2 of semigroups for (3). We will then prove Property 3b.

Fix an initial data μ_0. For each N, define the following approximated solution μ^N, based on the discretization in Definition 18: set $\mu^N(0) := \mathcal{A}^x_N(\mu_0)$, then recursively

$$\mu^N((k+1)\Delta_N) := \sum_{i,j} m^v_{ij}(V[\mu^N(k\Delta_N)])\delta_{x_i+\Delta_N v_j} + \Delta_N \mathcal{A}^x_N(s[\mu^N(k\Delta_N)]). \tag{30}$$

Also define the interpolated measure for $\tau \in [0, \Delta_N]$ as follows:

$$\mu^N(k\Delta_N + \tau) := \sum_{i,j} m^v_{ij}(V[\mu^N(k\Delta_N)])\delta_{x_i+\tau v_j} + \tau \mathcal{A}^x_N(s[\mu^N(k\Delta_N)]). \tag{31}$$

Clearly, the first term on the right hand side corresponds to the transport by the PVF V, while the second term corresponds to the source term s. We now prove that the sequence $\mu^N(t)$ is equi-bounded and equi-integrable on a complete space, to apply the Ascoli-Arzelà theorem.

We first prove that, for a fixed $T > 0$, the measures $\mu^N(t)$ are all supported in a compact set. Choose the radius R in hypothesis **(s2)** giving the maximal support of $s[\mu]$. Then, eventually enlarging R, one can assume that $\text{supp}(\mu(0)) \subset B(0, R)$. This implies $\text{supp}(\mu^N(0)) \subset B(0, R + \sqrt{\frac{n}{2}\Delta^x_N}) \subset B(0, R+1)$, where n is the dimension of the space $\mathbb{R}^n$ and N is chosen sufficiently large. Observe the following simple estimate: if $\text{supp}(\mu^N(k\Delta_N)) \subset B(0, r)$ with $r > R+1$, then it holds $\text{supp}(\mu^N(k\Delta_N + \tau)) \subset B(0, r + \Delta_N C(1+r))$ for $\tau \in (0, \Delta_N)$. Indeed:

- for each term i, j in the first term it holds that $(x_i, v_j) \in \text{supp}(V[\mu^N(k\Delta_N)])$ implies $|v_j| \le C(1+r)$ by Hypothesis **(V1)**, hence $\text{supp}(\delta_{x_i+\tau v_j}) \subset B(0, r + \Delta_N C(1+r))$;
- for the second term it holds $\text{supp}(s[\mu^N(k\Delta_N)]) \subset B(0, R)$, hence $\Delta_N \mathcal{A}^x_N(s[\mu^N(k\Delta_N)]) \subset B(0, R+1) \subset B(0, r)$.

Since summing measures with the same support is a closed operation, it holds $\text{supp}(\mu^N(k\Delta_N + \tau)) \subset B(0, r + \Delta_N C(1+r))$. Eventually replacing r by $\max\{1, r\}$, it holds by induction $\text{supp}(\mu^N(k\Delta_N + \tau)) \subset B(0, r(1+2C\Delta_N)^{k+1})$. Since $k \le \frac{T}{\Delta_N} + 1$, this implies $\text{supp}(\mu^N(t)) \subset B(0, r(1+2C)^2 e^{2CT})$ for all $t \in [0, T]$. Observe that the space $X = \overline{B(0, r(1+2C)^2 e^{2CT})}$ is compact. Then, the space $M(X)$ of measures with finite mass endowed with the generalized Wasserstein distance W^g is complete, see Proposition 12. Moreover, if we prove that the limit of a subsequence of μ^N exists, then it satisfies **Property 3a** in Definition 2.

We now prove that the sequences $\mu^N(t)$ are equi-Lipschitz in time with respect to the distance W^g. We also prove that the masses $|\mu^N(t)|$ are uniformly bounded. First observe that the operator $\mu \to \mathcal{A}^x_N(\mu)$ does not increase the mass of μ. The same property holds for the operator $\mu \to \sum_{i,j} m^v_{ij}(V[\mu])\delta_{x_i+\tau v_j}$. Then, by the explicit expressions (30)-(31) for $\mu^N(t)$, it holds

$$W^g(\mu^N(k\Delta_N + \tau), \mu^N(k\Delta_N + \sigma)) \le \sum_{i,j} m^v_{ij}(V[\mu^N(k\Delta_N)])|\tau - \sigma||v_j| + |\tau - \sigma||s[\mu^N(k\Delta_N)]| \le$$
$$\le |\tau - \sigma|\,|\mu^N(k\Delta_N)| \sup_j |v_j| + |\tau - \sigma|(W^g(s[\mu^N(k\Delta_N)], s[\mu_0]) + |s[\mu_0]|),$$

for $\tau, \sigma \in [0, \Delta_N]$. Here, we estimated the generalized Wasserstein distance by decomposing it in the Wasserstein distance for the transport term given by the PVF V, and the L^1 distance for the source term given by s. Use now (10) to estimate $|\mu^N(k\Delta_N)| \le |\mu_0| + W^g(\mu^N(k\Delta_N), \mu_0)$. Use uniform boundedness of

the supports for $\mu^N(k\Delta_N)$ and Hypothesis **(V1)** to estimate $\sup_j |v_j| \le C_1 := C(1+\mathrm{diam}(X))$. Also use Hypothesis **(s1)** to estimate $W^g(s[\mu^N(k\Delta_N)], s[\mu_0]) \le LW^g(\mu^N(k\Delta_N), \mu_0)$. This gives

$$\begin{aligned} &W^g(\mu^N(k\Delta_N+\tau), \mu^N(k\Delta_N+\sigma)) \le |\tau-\sigma| C_1(|\mu_0| + W^g(\mu^N(k\Delta_N), \mu_0)) + \\ &|\tau-\sigma|(|s[\mu_0]| + LW^g(\mu^N(k\Delta_N), \mu_0)) = |\tau-\sigma|(C_2 W^g(\mu^N(k\Delta_N), \mu_0) + C_2), \end{aligned} \tag{32}$$

where we choose C_2 a constant depending on $|\mu_0|$ and X only, i.e. independent on N.

We now prove that $W^g(\mu^N(k\Delta_N), \mu_0)$ is bounded, uniformly in N, k. Observe that the sequence $\mu^N(0)$ for N sufficiently large satisfies $W^g(\mu^N(0), \mu_0) \le W(\mu^N(0), \mu_0) \le |\mu_0|\Delta_N$, as a consequence of (11) and Proposition 19. Thus, there exists a constant C_3 such that $W^g(\mu^N(0), \mu_0) \le C_3$ for all N. We now prove the estimate

$$W^g(\mu^N(k\Delta_N), \mu_0) \le (1+C_2\Delta_N)^k C_3 + ((1+C_2\Delta_N)^k - 1), \tag{33}$$

by induction on k. The case $k=0$ is already proved. We now prove that, if the estimate holds for k, then it holds for $k+1$ too. Use (32) with $\tau = 0, \sigma = \Delta_N$, that gives

$$\begin{aligned} &W^g(\mu^N((k+1)\Delta_N), \mu_0) \le W^g(\mu^N((k+1)\Delta_N), \mu^N(k\Delta_N)) + W^g(\mu^N(k\Delta_N), \mu_0) \le \\ &\le (1+C_2\Delta_N) W^g(\mu^N(k\Delta_N), \mu_0) + C_2\Delta_N \le (1+C_2\Delta_N)^{k+1} C_3 + (1+C_2\Delta_N)((1+C_2\Delta_N)^k - 1) + \\ &C_2\Delta_N = (1+C_2\Delta_N)^{k+1} C_3 + ((1+C_2\Delta_N)^{k+1} - 1). \end{aligned}$$

The estimate (33) is now proved. Since $k \le T/\Delta_N$, it also holds $W^g(\mu^N(k\Delta_N), \mu_0) \le C_4 := e^{C_2 T} C_3 + (e^{C_2 T} - 1)$. Then, again by (32) and triangular inequalities, it holds

$$W^g(\mu^N(t), \mu^N(s)) \le |t-s| C_5, \tag{34}$$

with $C_5 := C_2 C_4 + C_2$, i.e. uniform Lipschitz continuity with respect to t of the sequence μ^N. Moreover, (10) also implies $|\mu^N(t)| \le C_6 := |\mu_0| + TC_5$, hence uniform boundedness of the mass. As a consequence, Ascoli-Arzelà theorem implies existence of a converging subsequence μ^N, that we do not relabel. Such limit $\mu^*(t)$ clearly satisfies **Property 1** and **Property 3c** in Definition 2.

We now prove **Property 2**, i.e. the fact that the limit is a solution to (3). Since the limit is uniformly Lipschitz and with bounded mass, it is easy to prove that the two first properties in Definition 1 are satisfied, as well as the fact that the function $t \to \int_{\mathbb{R}^n} f d\mu(t)$ is absolutely continuous (and even Lipschitz) for each $f \in C_c^\infty(\mathbb{R}^n)$. We are left to prove that the limit μ^* satisfies (4) for each $f \in C_c^\infty(\mathbb{R}^n)$ and almost each $t \in [0,T]$. For a fixed $f \in C_c^\infty(\mathbb{R}^n)$, define the operator F^N for $\tau, \sigma \in [0,T]$ as follows:

$$F^N(\tau,\sigma) := \int f(x)\, d(\mu^N(\sigma) - \mu^N(\tau)) - \int_\tau^\sigma dt \left(\int \nabla f(x) \cdot v \, dV[\mu^N(t)]) + \int f(x)\, ds[\mu^N(t)] \right). \tag{35}$$

For each $\tau, \sigma \in [k\Delta_N, (k+1)\Delta_N]$, the first term in the right hand side of (35) coincides with

$$\int f(x) \left(\sum_{i,j} m_{ij}^v (V[\mu^N(k\Delta_N)]) d(\delta_{x_i + (\sigma - k\Delta_N)v_j} - \delta_{x_i + (\tau - k\Delta_N)v_j}) + (\sigma-\tau)\, d\mathcal{A}_N^x(s[\mu^N(k\Delta_N)]) \right) \tag{36}$$

Define

$$g_{ij}(\alpha) := f(x_i + \alpha(\sigma - k\Delta_N)v_j) - f(x_i + \alpha(\tau - k\Delta_N)v_j), \tag{37}$$

that satisfies $g_{ij}(1) = \int f(x) d(\delta_{x_i + (\sigma - k\Delta_N)v_j} - \delta_{x_i + (\tau - k\Delta_N)v_j})$. By applying Taylor's theorem with Lagrange remainder to each g_{ij}, there exist $\alpha_{ij} \in (0,1)$ such that (37) coincides with

$$g_{ij}(1) \quad = \quad (\sigma-\tau)\nabla f(x_i) \cdot v_j + v_j \cdot ((\sigma - k\Delta_N)^2 Hf(x_i + \alpha_{ij}\sigma v_j) - (\tau - k\Delta_N)^2 Hf(x_i + \alpha_{ij}\tau v_j)) \cdot v_j,$$

where Hf is the Hessian of f.

We now estimate $F^N(\tau,\sigma)$ with $\tau,\sigma\in[k\Delta_N,(k+1)\Delta_N]$. It holds

$$\begin{aligned}|F^N(\tau,\sigma)| \quad &\leq \quad \left|\sum_{i,j} g_{ij}(1)m^v_{ij}(V[\mu^N(k\Delta_N)]) - \int_\tau^\sigma dt\int \nabla f(x)\cdot v\, dV[\mu^N(t)])\right| + \\ &\qquad \left|(\sigma-\tau)\int f(x)\, d\mathcal{A}^x_N(s[\mu^N(k\Delta_N)]) - \int_\tau^\sigma dt\int f(x)\, ds[\mu^N(t)]\right|. \end{aligned} \tag{38}$$

The first term on the right hand side of (38) is bounded from above by $\int_\tau^\sigma I_1(t)\,dt + I_2(\tau,\sigma)$, where

$$\begin{aligned} I_1(t) \quad &:= \quad \left|\sum_{i,j}\nabla f(x_i)\cdot v_j\, m^v_{ij}(V[\mu^N(k\Delta_N)]) - \int \nabla f(x)\cdot v\, dV[\mu^N(t)])\right| = \\ &= \quad \left|\int \nabla f(x)\cdot v\, d\mathcal{A}^v_N(V[\mu^N(k\Delta_N)]) - \int \nabla f(x)\cdot v\, dV[\mu^N(t)])\right|, \\ I_2(\tau,\sigma) \quad &:= \quad \left|\sum_{i,j} v_j\cdot((\sigma-k\Delta_N)^2 Hf(x_i+\alpha_{ij}\sigma v_j) - (\tau-k\Delta_N)^2 Hf(x_i+\alpha_{ij}\tau v_j))\cdot v_j\, m^v_{ij}(V[\mu^N(k\Delta_N)])\right| \end{aligned}$$

The second term on the right hand side of (38) is bounded from above by $\int_\tau^\sigma I_3(t)\,dt$, where

$$I_3(t) := \left|\int f(x)\, d\mathcal{A}^x_N(s[\mu^N(k\Delta_N)]) - \int f(x)\, ds[\mu^N(t)]\right|.$$

By the duality formulas for the generalized Wasserstein and the L^1 distance (14)-(13), the Lipschitz condition (34) and the estimate $|t-k\Delta_N|\leq\Delta_N$, for a sufficiently large N it holds

$$\begin{aligned} &I_1(t) \leq \|\nabla f\cdot v\|_{C^1} W^g(\mathcal{A}^v_N(V[\mu^N(k\Delta_N)]), V[\mu^N(t)]) \leq \\ &(\|f\|_{C^2}\|v\|_{C^1}(W^g(\mathcal{A}^v_N(V[\mu^N(k\Delta_N)]), V[\mu^N(k\Delta_N)]) + W^g(V[\mu^N(k\Delta_N)], V[\mu^N(t)])) \leq \\ &\|f\|_{C^2}(C_1+1)(\Delta_N|V[\mu^N(k\Delta_N)]| + |t-k\Delta_N|C_5) \leq \Delta_N\|f\|_{C^2}(C_1+1)(C_6+C_5). \end{aligned} \tag{39}$$

We recall that the support of the velocities is bounded ($\sup_j|v_j|\leq C_1$), and that the projection condition $\pi\#V[\mu]=\mu$ implies $|V[\mu]|=|\mu|$. Similarly, for $I_2(\tau,\sigma)$, by removing and adding the term $(\tau-k\Delta_N)^2 Hf(x_i+\alpha_{ij}\sigma v_j))$, it holds

$$\begin{aligned} I_2(\tau,\sigma) \quad &\leq \quad \sum_{i,j}\|f\|_{C^2}(nC_1)^2((\sigma-k\Delta_N)^2-(\tau-k\Delta_N)^2)m^v_{ij}(V[\mu^N(k\Delta_N)]) + \\ &\qquad \sum_{i,j}(nC_1)^2(\tau-k\Delta_N)^2\|Hf(x_i+\alpha_{ij}\sigma v_j) - Hf(x_i+\alpha_{ij}\tau v_j)\|_{\mathbb{R}^n,\mathbb{R}^n} m^v_{ij}(V[\mu^N(k\Delta_N)]), \end{aligned}$$

where $\|\cdot\|_{\mathbb{R}^n,\mathbb{R}^n}$ is the operator norm. Apply the mean-value theorem to Hf, to have

$$\|Hf(x_i+\alpha_{ij}\sigma v_j) - Hf(x_i+\alpha_{ij}\tau v_j)\|_{\mathbb{R}^n,\mathbb{R}^n} \leq \alpha_{ij}(\sigma-\tau)v_j\|f\|_{C^3}.$$

Recall that $\sigma-k\Delta_N,\tau-k\Delta_N\leq\Delta_N$ as well as $\alpha_{ij}\in(0,1)$. It then holds

$$\begin{aligned} I_2(\tau,\sigma) \quad &\leq \quad \|f\|_{C^2}(nC_1)^2(\sigma-\tau)2\Delta_N|V[\mu^N(k\Delta_N)]| + (nC_1)^3\Delta_N^2(\sigma-\tau)\|f\|_{C^3}|V[\mu^N(k\Delta_N)]| \\ &\leq \quad (\sigma-\tau)\Delta_N(nC_1)^3C_6\|f\|_{C^3}(2+\Delta_N). \end{aligned} \tag{40}$$

Finally, for $I_3(t)$ it holds

$$\begin{aligned} I_3(t) \quad &\leq \quad \|f\|_{C^1}W^g(\mathcal{A}^x_N(s[\mu^N(k\Delta_N)]), s[\mu^N(t)]) \leq \\ &\qquad \|f\|_{C^1}(W(\mathcal{A}^x_N(s[\mu^N(k\Delta_N)]), s[\mu^N(k\Delta_N)]) + W^g(s[\mu^N(k\Delta_N)], s[\mu^N(t)])) \leq \\ &\qquad \|f\|_{C^1}(\Delta_N|s[\mu^N(k\Delta_N)]| + LC_5\Delta_N). \end{aligned} \tag{41}$$

Observe that (10) implies

$$\begin{aligned}\left|s[\mu^N(k\Delta_N)]\right| \quad &\leq \quad |s[\mu_0]| + W^g(s[\mu^N(k\Delta_N)]|, s[\mu_0]) \leq |s[\mu_0]| + \\ &L(W^g(\mu^N(k\Delta_N), \mu^N(0)) + W^g(\mu^N(0), \mu_0)) \leq |s[\mu_0]| + LC_5T + L|\mu_0|\Delta_N. \qquad (42)\end{aligned}$$

Merging (39)-(40)-(41)-(42), it then holds

$$|F^N(\tau, \sigma)| \leq |\tau - \sigma|\Delta_N \|f\|_{C^3} C_7$$

for a suitable constant C_7. Take now a general pair $\tau, \sigma \in [0, T]$. For simplicity, assume that $\tau < \sigma$ and that N is sufficiently large to have

$$(k_1 - 1)\Delta_N \leq \tau < k_1\Delta_N \leq k_2\Delta_N < \sigma \leq (k_2 + 1)\Delta_N$$

for some $k_1, k_2 \in \mathbb{N} \setminus \{0\}$. Our aim is to prove that it holds $\lim_{\sigma\to\tau} \mathcal{L} = 0$, with

$$\mathcal{L} := \frac{1}{\sigma - \tau}\left(\int f(x)\, d(\mu^*(\sigma) - \mu^*(\tau)) - \int_\tau^\sigma dt \int \nabla f(x) \cdot v\, dV[\mu^*(t)] - \int_\tau^\sigma dt \int f(x)\, ds[\mu^*(t)]\right).$$

Observe that, restricting ourselves to the converging subsequence $\mu^N \rightharpoonup \mu^*$, it holds

$$\mathcal{L} = \lim_{N\to\infty} \frac{1}{\sigma - \tau}\left(\int f(x)\, d(\mu^N(\sigma) - \mu^N(\tau)) - \int_\tau^\sigma dt \int \nabla f(x) \cdot v\, dV[\mu^N(t)] - \int_\tau^\sigma dt \int f(x)\, ds[\mu^N(t)]\right).$$

We used here continuity of both V and s with respect to weak convergence of measures, as a consequence of (29) and **(V2)** for V, and **(s1)** for s. Observe that $\mathcal{L}$ coincides with

$$\lim_{N\to\infty} \frac{1}{\sigma - \tau}\left(F^N(\tau, k_1\Delta_N) + \sum_{k=k_1}^{k_2-1} F^N(k\Delta_N, (k+1)\Delta_N) + F^N(k_2\Delta_N, \sigma)\right),$$

thus

$$\lim_{\sigma\to\tau} |\mathcal{L}| \leq \lim_{\sigma\to\tau} \frac{1}{|\sigma - \tau|} \lim_{N\to\infty} (|\tau - k_1\Delta_N| + \ldots + |k_2\Delta_N - \sigma|)\Delta_N \|f\|_{C^3} C_7 = 0.$$

We finally prove **Property 3b** in Definition 2. Take two different data μ_0, ν_0 and build the Lattice Approximate Solutions μ^N, ν^N according with scheme (30)-(31). Assume to have N sufficiently large so that it holds

$$\mathcal{A}_N^x[\mu^N(k\Delta_N)] = \mu^N(k\Delta_N) \qquad (43)$$

for all $k \in \mathbb{N}$ with $k \leq T/\Delta_N$, and similarly for ν^N. Such N exists, for two reasons: first, μ^N, ν^N have uniformly bounded support (Property 3a proved above), thus for N sufficiently large $\mathcal{A}_N^x$ conserves the mass. Second, observe that $\mu^N(0)$ satisfies (43) by construction, and that if $\mu^N(k\Delta_N)$ satisfies it, then $\mu^N((k+1)\Delta_N)$ satisfies it too. Then, by induction, this holds for all k.

Similarly, we assume to have N sufficiently large to have

$$\pi_1\#\mathcal{A}_N^v(V[\mu^N(k\Delta_N)]) = \pi_1\#V[\mu^N(k\Delta_N)], \qquad (44)$$

and similarly for ν^N. This is first based on the fact that uniform bounded supports of the measure μ_t^N (Property 3a proved above), together with support sublinearity **(V1)** of V implies uniform boundedness of the supports of the PVF $V[\mu_t^N]$, thus $\mathcal{A}_N^v$ conserves the mass for N sufficiently large. As soon as $|\mathcal{A}_N^v(V[\mu^N(k\Delta_N)])| = |\pi_1\#V[\mu^N(k\Delta_N)]|$, one has that the support of $\pi_1\#\mathcal{A}_N^v(V[\mu^N(k\Delta_N)])$ coincides with $\mathcal{A}_N^x(\mu^N(k\Delta_N))$, since $\pi_1\#V[\mu] = \mu$ and the discretization (18) has the same effect of $\mathcal{A}_N^x$ on the base space. Then, (43) implies (44).

Moreover, since (44) holds, one has that any transference plan $p \in P(\mathcal{A}^v_N(V[\mu^N(k\Delta_N)]), V[\mu^N(k\Delta_N)])$ that satisfies $\pi_{13}\#p \in P^{opt}(\mu^N(k\Delta_N), \mu^N(k\Delta_N))$ is indeed a.e.-$\mu^N(k\Delta_N)$ concentrated on the diagonal $\{x = y\}$ of $\mathbb{R}^n \times \mathbb{R}^n$. As a consequence, one can estimate $\mathcal{W}^g(\mathcal{A}^v_N(V[\mu^N(k\Delta_N)]), V[\mu^N(k\Delta_N)])$ by choosing in the right hand side of (28) the decomposition $\tilde{V}_1 = \mathcal{A}^v_N(V[\mu^N(k\Delta_N)])$, $\tilde{V}_2 = V[\mu^N(k\Delta_N)]$ and the transference plan p realizing $W(V[\mu^N(k\Delta_N)], \mathcal{A}^v_N(V[\mu^N(k\Delta_N)]))$. Such choice implies

$$\mathcal{W}^g(\mathcal{A}^v_N(V[\mu^N(k\Delta_N)]), V[\mu^N(k\Delta_N)]) \leq W(\mathcal{A}^v_N(V[\mu^N(k\Delta_N)]), V[\mu^N(k\Delta_N)]) \leq |\mu^N(k\Delta_N)|\Delta^v_N. \tag{45}$$

We now estimate recursively $W^g(\mu^N((k+1)\Delta_N), \nu^N((k+1)\Delta_N))$ starting from $W^g(\mu^N(k\Delta_N), \nu^N(k\Delta_N))$. Consider the PVFs

$$V_1 := \mathcal{A}^v_N(V[\mu^N(k\Delta_N)]), \qquad V_2 = \mathcal{A}^v_N(V[\nu^N(k\Delta_N)]),$$

and the operator $\mathcal{W}^g(V_1, V_2)$. By definition of $\mathcal{W}^g$, there exist a choice $\tilde{V}_1 \leq V_1, \tilde{V}_2 \leq V_2$, and a transference plan $p \in P(\tilde{V}_1, \tilde{V}_2)$ with the two following properties: one one side, it holds $\mathcal{W}^g(V_1, V_2) = \int_{T\mathbb{R}^n \times T\mathbb{R}^n} |v - w|\, dp(x, v, y, w)$; on the other side, denoting with $\tilde{\mu}_i = \pi_1\#\tilde{V}_i$, it holds $W^g(\mu^N(k\Delta_N), \nu^N(k\Delta_N)) = |\mu^N(k\Delta_N) - \tilde{\mu}_1| + W(\tilde{\mu}_1, \tilde{\mu}_2) + |\nu^N(k\Delta_N) - \tilde{\mu}_2|$, and $W(\tilde{\mu}_1, \tilde{\mu}_2)$ is realized by the transference plan $\pi_{13}\#p \in P(\tilde{\mu}_1, \tilde{\mu}_2)$.

Consider now the following corresponding decomposition: write $\mu^N(k\Delta_N) = \sum_i m_i\delta_{x_i} + \sum_j n_j\delta_{y_j}$ and $\nu^N(k\Delta_N) = \sum_l p_l\delta_{z_i} + \sum_j n_j\delta_{t_j}$, where $\tilde{\mu}_1 = \sum_j n_j\delta_{y_j}$, $\tilde{\mu}_2 = \sum_j n_j\delta_{t_j}$, and the optimal transference plan $\pi_{13}\#p$ sends each $n_j\delta_{y_j}$ to each $n_j\delta_{t_j}$. Decompose accordingly the PFV as follows:

$$V_1 = \sum_{i,k} m_{ik}\delta_{(x_i,v_k)} + \tilde{V}^1 \quad \text{with } \tilde{V}^1 := \sum_{j,k} n_{jk}\delta_{(y_j,v_k)},$$

and similarly

$$V_2 = \sum_{l,k} p_{lk}\delta_{(z_i,v'_k)} + \tilde{V}^2 \quad \text{with } \tilde{V}^2 := \sum_{j,k} n_{jk}\delta_{(t_j,v'_k)},$$

with the additional requirement that the optimal transference plan $p \in P(\tilde{V}^1, \tilde{V}^2)$ sends $n_{jk}\delta_{(y_j,v_k)}$ to $n_{jk}\delta_{(t_j,v'_k)}$.

By definition of μ^N, it then holds $\mu^N((k+1)\Delta_N) = \sum_{ik} m_{ik}\delta_{x_i+\Delta_N v_k} + \sum_{jk} n_{jk}\delta_{y_j+\Delta_N v_k}$, and similarly for $\nu^N((k+1)\Delta_N)$. Estimate the distance $W^g(\mu^N((k+1)\Delta_N), \nu^N((k+1)\Delta_N))$ by choosing the first component for mass removal and the second one for transport. It then holds

$$\begin{aligned} W^g(\mu^N((k+1)\Delta_N), \nu^N((k+1)\Delta_N)) &\leq \sum_{ik} |m_{ik}| + \sum_{jk} n_{jk}|y_j + \Delta_N v_k - t_j - \Delta_N v'_k| + \sum_{lk} |p_{lk}| \leq \\ &\leq \sum_i m_i + \sum_j n_j|y_j - t_j| + \Delta_N \sum_{jk} n_{jk}|v_k - v'_k| + \sum_l p_l = \\ &= W^g(\mu^N(k\Delta_N), \nu^N(k\Delta_N)) + \Delta_N \mathcal{W}^g(V_1, V_2). \end{aligned} \tag{46}$$

We now need to compare $\mathcal{W}^g(V_1, V_2)$ with $\mathcal{W}^g(V[\mu^N(k\Delta_N)], V[\nu^N(k\Delta_N)])$. Denote with $\tilde{V}_1, \tilde{V}_2$ the decomposition and $p \in P(\tilde{V}_1, \tilde{V}_2)$ the transference plan in (28) realizing $\mathcal{W}^g(V[\mu^N(k\Delta_N)], V[\nu^N(k\Delta_N)])$. Observe that $\pi_1\#\tilde{V}_1 \leq \pi_1\#(V[\mu^N(k\Delta_N)]) = \mu^N(k\Delta_N)$ that is a finite sum of Dirac deltas, and the same holds for $\pi_1\#\tilde{V}_2$. Thus, p can be decomposed as follows

$$p := \sum_{i,k} p_{ik}(v, w)\delta_{x_i,y_k}, \tag{47}$$

where each p_{ik} is a transference plan on $T_{x_i}\mathbb{R}^n \times T_{y_k}\mathbb{R}^n$.

We are now ready to define a decomposition $\bar{V}_1 \leq V_1, \bar{V}_2 \leq V_2$ and a transference plan $q \in P(\bar{V}_1, \bar{V}_2)$ to estimate $\mathcal{W}^g(V_1, V_2)$ from above. For each transference plan p_{ik}, define

$$q_{ik} := \sum_{jl} p_{ik}((v_j + Q') \times (w_l + Q'))\delta_{v_j,w_l},$$

where the v_j, w_l are the equispaced discretized points on $T_x\mathbb{R}^n, T_y\mathbb{R}^n$, respectively, and Q' is defined in Definition 18. Then define

$$q := \sum_{i,k} q_{ik}\delta_{x_i,y_k}. \tag{48}$$

Define now $\bar{V}_1 := \pi_{12}\#q$. By the definition of $\mathcal{A}_N^v$ in (18) and of q in (48), it is easy to prove that it holds

$$\bar{V}_1 = \pi_{12}\#q = \mathcal{A}_N^v(\pi_{12}\#p) = \mathcal{A}_N^v(\tilde{V}_1) \leq A_N^v(V[\mu^N(k\Delta_N)]) = V_1.$$

One can equivalently prove that it holds $\bar{V}_2 := \pi_{34}\#q \leq V_2$. Moreover, it holds $\pi_{13}\#q = \pi_{13}\#p \in P^{opt}(\mu^N(k\Delta_N)], \nu^N(k\Delta_N))$, where we used that p is a minimizer of $\mathcal{W}^g(V[\mu^N(k\Delta_N)], V[\nu^N(k\Delta_N)])$. Then, the decomposition $\bar{V}_1, \bar{V}_2$ with the transference plan q is admissible in the right hand side of (28). It then holds

$$\mathcal{W}^g(V_1, V_2) \leq \int_{T\mathbb{R}^n\times T\mathbb{R}^n} |v-w|\, dq(x,v,y,w) = \sum_{ij}\int_{T_{x_i}\mathbb{R}^n\times T_{y_k}\mathbb{R}^n} |v-w|\, dq_{ik}(v,w). \tag{49}$$

We estimate each term by following the definition of q_{ik}, as follows

$$\begin{aligned}\int_{T_{x_i}\mathbb{R}^n\times T_{y_k}\mathbb{R}^n} |v-w|\, dq_{ik}(v,w) &= \sum_{jl} |v_j - w_l|\, dq_{ik}(v,w) = \sum_{jl} |v_j - w_l| p_{ik}((v_j+Q')\times(w_l+Q')) \leq \\ &\sum_{jl}(|v-w| + 2\mathrm{diam}(Q'))(p_{ik})_{|((v_j+Q')\times(w_l+Q'))}.\end{aligned}$$

Here, we used that $|v_j - w_k| \leq |v_j - v| + |v-w| + |v-w_k|$ for any $v \in v_j + Q'$ and $w \in w_l + Q'$. Observe now that, by decomposition, it holds

$$\sum_{jl}(|v-w| + 2\mathrm{diam}(Q'))(p_{ik})_{|((v_j+Q')\times(w_l+Q'))} = \int |v-w|\, dp_{ik} + 2\mathrm{diam}(Q')|p_{ik}|.$$

Summing over ik in (49) and recalling the definition of p in (47), we have

$$\mathcal{W}^g(V_1,V_2) \leq \int |v-w|\, dp + 2\mathrm{diam}(Q')|p| \leq \mathcal{W}^g(V[\mu^N(k\Delta_N)], V[\nu^N(k\Delta_N)]) + 2\sqrt{n}\Delta_N|\mu^N(k\Delta_N)|. \tag{50}$$

We used here that $\mathrm{diam}(Q') \leq \sqrt{n}\Delta_N$ as a consequence of Definition 18, as well as the fact that it holds

$$|p| = |\pi_{12}\#p| \leq |V[\mu^N(k\Delta_N)]| = |\mu^N(k\Delta_N)| \leq C_6.$$

Going back to (46), and using the Lipschitz continuity hypothesis **(V2)** in (50), it holds

$$W^g(\mu^N((k+1)\Delta_N), \nu^N((k+1)\Delta_N)) \leq (1+K)W^g(\mu^N(k\Delta_N), \nu^N(k\Delta_N)) + 2\sqrt{n}(\Delta_N)^2 C_6 \tag{51}$$

Clearly, this estimate implies

$$W^g(\mu^N(t), \nu^N(t)) \leq e^{Kt}W^g(\mu^N(0), \nu^N(0)) + 2\sqrt{n}(\Delta_N)^2 C_6 \frac{e^{Kt}-1}{K},$$

as long as $t = k\Delta_N$. By uniform Lipschitz continuity given by (34), and Proposition 19, this implies

$$W^g(\mu^N(t), \nu^N(t)) \leq e^{Kt}W^g(\mu_0, \nu_0) + 2C_6\Delta_N + 2\sqrt{n}(\Delta_N)^2 C_6 \frac{e^{Kt}-1}{K}. \tag{52}$$

We now use a density argument to pass to the limit. Consider the countable set

$$\mathcal{D} = \left\{\mu_0 \in \mathcal{M} \ \text{ s.t. } \ \mu_0 = \sum_i m_i\delta_{x_i}, 0 < m_i \in \mathbb{Q}, x_i \in \mathbb{Q}^n\right\}.$$

Choose a $\mu_0 \in \mathcal{D}$, define the corresponding sequence μ_0^N and choose a subsequence N_k such that $\mu_0^{N_k}$ uniformly converges to a solution $\mu(t)$ to (3). Then choose $\mu_1 \in \mathcal{D}$, consider the subsequence $\mu_1^{N_k}$ and choose a converging subsequence $\mu_1^{N_{k_l}}$. Repeat this diagonal argument for the countable set of initial data in $\mathcal{D}$, and observe that, passing to the limit in (52) for $N \to \infty$, it holds

$$W^g(\mu(t), \nu(t)) \leq e^{Kt} W^g(\mu_0, \nu_0)$$

for all $\mu_0, \nu_0 \in \mathcal{D}$. Observe now that $\mathcal{D}$ is dense in $\mathcal{M}$, thus the continuous semigroup $\mu_0 \to S_t\mu_0 = \mu(t)$ can be uniquely extended from $\mathcal{D}$ to $\mathcal{M}$. □

3.2 Uniqueness

We now prove Theorem 4, i.e. uniqueness of a solution to (3) when a Dirac germ γ is fixed. We first need to define compatibility of a semigroup for the dynamics (3), that is the following.

Definition 29 *Fix a PVF V satisfying* **(V1)**, *a final time $T > 0$ and a Dirac germ γ as in Definition 22. A semigroup for (3) is said to be compatible with γ if one has the following property: for each $R, M > 0$ there exists $C(R, M)$ such that the space $\mathcal{M}^D_{R,M} := \{\mu \in \mathcal{M}^D \ \ s.t. \ \ \mathrm{supp}(\mu) \in B(0, R), |\mu| \leq M\}$ satisfies*

$$\text{for all } t \in [0, \inf_{\mu \in \mathcal{M}^D_{R,M}} \varepsilon(\mu)] \text{ one has } \sup_{\mu \in \mathcal{M}^D_R} W^g(S_t\mu, \gamma_\mu(t)) \leq C(R, M)t^2. \tag{53}$$

Observe that this definition coincides with Definition 23, where the dynamics (1) is replaced by (3) and the metric W is replaced by W^g.

The proof of Theorem 4 is then based on the following Lemma.

Lemma 30 *Let S be a Lipschitz semigroup and $\mu : [0, T] \to \mathcal{M}$ a Lipschitz continuous curve. It then holds*

$$W^g(S_t\mu(0), \mu(t)) \leq e^{Ct} \int_0^t \liminf_{h \to 0^+} W^g(S_h\mu(s), \mu(s + h))\, ds,$$

where C is the Lipschitz constant in Property 3b of Definition 2.

Proof. The original proof in Banach spaces can be found in [4, Thm 2.9]. Its adaptation to metric spaces can be found in [9, Appendix A]. □

We are now ready to prove Theorem 4.

Proof of Theorem 4. Assume to have two semigroups S^1_t, S^2_t for solutions to (3), both compatible with a given germ γ. Fix an initial data $\mu_0 \in \mathcal{M}$ and $T \geq 0$. By Property 3a in the Definition 2, there exists $R > 0$ such that $\mathrm{supp}(S^1_t\mu_0) \cup \mathrm{supp}(S^2_t\mu_0) \subset B(0, R)$ for all $t \in [0, T]$. By applying Lemma 30 with $\mu(t) = S^2_t$, it holds

$$W^g(S^1_t\mu_0, S^2_t\mu_0) \leq e^{Ct} \int_0^t \liminf_{h \to 0^+} W^g(S^1_h S^2_s\mu_0, S^2_{s+h}\mu_0)\, ds, \tag{54}$$

Fix $s > 0$ and define $\nu = S^2_s\mu_0$. By density of $\mathcal{M}^D$ in $\mathcal{M}$ with respect to the topology induced by W^g, for each $\varepsilon > 0$ there exists $\bar{\nu} \in \mathcal{M}^D$ such that $W^g(\nu, \bar{\nu}) < \varepsilon$. Property (53) applied to both S^1_h and S^2_h implies the existence of a constant $C_1(R, M)$ such that

$$W^g(S^1_h\bar{\nu}, \gamma_{\bar{\nu}}(h)) \leq C_1(R, M)h^2, \qquad W^g(S^2_h\bar{\nu}, \gamma_{\bar{\nu}}(h)) \leq C_1(R, M)h^2.$$

It then holds

$$W^g(S^1_h S^2_s\mu_0, S^2_{s+h}\mu_0) \leq W^g(S^1_h\nu, S^1_h\bar{\nu}) + W^g(S^1_h\bar{\nu}, \gamma_{\bar{\nu}}(h)) + W^g(\gamma_{\bar{\nu}}(h), S^2_h\bar{\nu}) + W^g(S^2_h\bar{\nu}, S^2_h\nu) \leq$$
$$2(e^{Ch}\varepsilon + C_1(R, M)h^2).$$

Since s has been chosen arbitrarily and ε is arbitrarily small for the density property, it holds

$$\liminf_{h \to 0^+} \frac{1}{h} W^g(S^1_h S^2_s\mu_0, S^2_{s+h}\mu_0) = 0$$

for each $s > 0$. This estimate applied to (54) implies $S^1_t\mu_0 = S^2_t\mu_0$ for any $t \geq 0$ and any initial data μ_0, i.e. $S^1 = S^2$. □

References

[1] Luigi Ambrosio. Transport equation and Cauchy problem for BV vector fields. *Invent. Math.*, 158(2):227–260, 2004.

[2] Luigi Ambrosio, Maria Colombo, and Alessio Figalli. Existence and uniqueness of maximal regular flows for non-smooth vector fields. *Archive for Rational Mechanics and Analysis*, 218(2):1043–1081, Nov 2015.

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[4] Alberto Bressan. *Hyperbolic systems of conservation laws*, volume 20. Oxford University Press, Oxford, 2000.

[5] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. An interpolating distance between optimal transport and Fisher–Rao metrics. *Foundations of Computational Mathematics*, pages 1–44, 2016.

[6] Stanislav Kondratyev, Léonard Monsaingeon, and Dmitry Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11-12):1117–1164, 2016.

[7] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal transport in competition with reaction: The hellinger–kantorovich distance and geodesic curves. *SIAM Journal on Mathematical Analysis*, 48(4):2869–2911, 2016.

[8] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, Mar 2018.

[9] Benedetto Piccoli. Measure differential equations. *arXiv preprint arXiv:1708.09738*, 2017.

[10] Benedetto Piccoli and Francesco Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.*, 211(1):335–358, 2014.

[11] Benedetto Piccoli and Francesco Rossi. On properties of the generalized wasserstein distance. *Archive for Rational Mechanics and Analysis*, 222(3):1339–1365, 2016.

[12] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[13] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.