# SCALAR CONSERVATION LAWS SEEN AS GRADIENT FLOWS: KNOWN RESULTS AND NEW PERSPECTIVES.

MARCO DI FRANCESCO

ABSTRACT. We review some results in the literature which attempted (only partly successfully) at linking the theory of scalar conservation laws with the Wasserstein gradient flow theory. In particular, we consider the problem of writing a scalar conservation law within the Wasserstein gradient flow theory. As a related problem, we also review results on contraction properties of scalar conservation laws in the $p$-Wasserstein distances. Moreover, we provide a particle-based approach to view a scalar conservation law as a gradient flow in a nonlinear-mobility sense. Finally, we propose a semi-implicit particle method based on the standard 2-Wasserstein distance.

## 1. INTRODUCTION

In 2000, a preprint by Felix Otto (published in 2001 in [52]) laid the basis for a theory which was going to catch the interest of a wide community of fine applied mathematicians in the later 15 years. By improving and extending an approach previously developed in [45], and possibly inspired by old results by De Giorgi [34], Otto proposed a new way to look at linear and nonlinear diffusion equations as gradient flows. The novelty of Otto's approach was mostly in the use of the Wasserstein space of probability measures as the chosen metric framework in which the new gradient flow approach was going to be set. Such approach has become a well established and structured theory in the later years mostly thanks to the monograph [2]. In short, the theory works as follows: let $\mathcal{F}$ be a 'good-mannered' functional defined on the space $\mathcal{P}_2(\mathbb{R}^d)$ of probability measures with finite second moment. Consider its 'sub-differential' $\partial\mathcal{F}[\rho]$, which is interpreted as an element in the tangent space of $\mathcal{P}_2(\mathbb{R}^d)$ around $\rho$. Then, the continuity equation

$$\frac{\partial \rho}{\partial t} - \operatorname{div}\left(\rho \partial\mathcal{F}[\rho]\right) = 0 \tag{1}$$

can be interpreted as the gradient flow of $\mathcal{F}$ with respect to the $\left(P_2(\mathbb{R}^d), d_2\right)$ metric structure, where $d_2$ is the 2-Wasserstein metric, see [56]. With the results in [2], which were partly anticipated in [26], the following class of functionals turned out to be feasible for this theory, namely

$$\mathcal{F}[\rho] = \int \phi(\rho)dx + \int \rho V(x)dx + \frac{1}{2}\iint W(x-y)\rho(x)\rho(y)dxdy, \tag{2}$$

with

- $\phi$ an increasing function with certain additional properties, having in mind $\phi(\rho) = \frac{1}{m-1}\rho^m$ for $m > 1 - 1/d$ and $\phi(\rho) = \rho\log\rho$ as prototype examples,
- $V$ and $W$ two continuous external potentials which are convex up to a quadratic perturbation.

With the choice (2) for the functional $\mathcal{F}$, the equation (1) becomes

$$\rho_t = \operatorname{div}\left(\rho\nabla\left(\phi'(\rho) + W*\rho + V\right)\right). \tag{3}$$

We remark the important role played by [49] to establish a suitable notion of convexity for this theory. The gradient flow approach of [45] was also used in chemotaxis modelling, see [6], and in nonlocal interaction equations with non smooth kernels, see [20, 8].

Already in their review paper [28], Carrillo and Toscani started investigating the role of the Lagrangian formulation for a continuity equation of the form (1), with the aim of proving simple 'quasi-contraction inequalities' of the form

$$d_2(\rho_1(t), \rho_2(t)) \leq d_2(\rho_1(0), \rho_2(0))e^{Ct}, \qquad C \in \mathbb{R}.$$

Their approach is restricted to the one dimensional case, in which the 2-Wasserstein space is isomorphic to the space $L^2([0,1])$ via the bijection

$$\mathcal{P}_2(\mathbb{R}) \ni \mu \mapsto X_\mu(z) = \inf\left\{x \in \mathbb{R} \ : \ \mu((-\infty, x]) > z\right\} \in L^2([0,1]).$$

If one assumes that $\rho(\cdot, t)$ is a solution to the gradient flow equation (1), then the *pseudo-inverse* variable $X = X_\rho$ satisfies another partial differential equation which, despite being often seemingly hard to handle, turns out to

provide such contraction estimates almost for free. The simplest example considered in [28] concerns with the Porous Medium equation (PME) $\rho_t = (\rho^m)_{xx}$, with $m > 1$.

Inspired by the results of [4] in probability theory, the gradient flow structure of (3) led to several results on the large time asymptotic behaviour of solutions. A first result by Toscani [55] on the heat equation was the basis of a systematic theory developed in [3] for linear Fokker-Planck type equations, in which exponential convergence to equilibrium was proven. A few months later, Carrillo and Toscani [27], Otto [52], and Del Pino and Dolbeault [35] (independently) found the optimal rate of convergence to Barenblatt solutions of the Porous Medium equation for solutions with finite second moment. The techniques in [27] and [52] were both (more or less directly) exploiting the Wasserstein gradient flow structure of the PME. Connections with functional inequalities arising in probability theory, such as the Log-Sobolev inequality, the Talagrad inequality, and the Csiszar-Kullback inequality, were established in [3, 52, 27, 53, 23] to mention a few. The two papers [25, 26] extended the theory to the full class of equations (3) including also nonlocal potentials $W$. In the meantime, the book [56] by Cedric Villani collected the most relevant facts on optimal transportation theory related with PDEs, and became a fundamental reference to this subject. In 2005, Ambrosio, Gigli, and Savaré [2] provided an extensive theory of gradient flows on metric spaces and in the Wasserstein space, which contained the previous results as special cases, and made the Riemannian structure of the Wasserstein space developed by Otto in [52] rigorous.

Already at this stage, some researchers started wondering on whether or not a nonlinear local *convection term* could be inserted in (3) in a way to make the whole theory still work, or even more ambitiously, whether a *nonlinear conservation law*

$$\rho_t + f(\rho)_x = 0 \tag{4}$$

could be seen as a Wasserstein gradient flow. This was mainly due to three reasons:

- Scalar conservation laws have basically the structure of a continuity equation with a nonlinear flux. Very naively, if one chooses

$$W(x) = H(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \,, \end{cases}$$

  then $H' * \rho = \rho$ (formally), and one may be tempted to view the Burgers equation

$$\rho_t + (\rho^2)_x = 0$$

  as a gradient flow of

$$\rho \mapsto \iint \rho(x)\rho(y)H(x-y)dydx.$$

  Unfortunately (apart from $H$ being too singular for the theory to work), only the symmetric part of the interaction potential $W$ survives in the functional $\mathcal{F}$ in (2), and therefore this functional would reduce to a constant. So this interpretation is clearly *false*.
- Scalar conservation laws with $f(\rho) = \rho^q$, $q > 1$, exhibit a very similar large time behaviour to that of the Porous Medium Equation, in that they decay to zero for large times with a polynomial rate, and most importantly they are well approximated by a compactly supported *self-similar* profile $t^{-\alpha}U(xt^{-\alpha})$ as $t \to +\infty$. Such profiles are called $N$-waves. Their occurrence as intermediate states for scalar conservation laws was well known since the pioneering works by Lax [47] and Di Perna [40] for strictly convex fluxes. These results were extended to more general power law cases in [48].
- Of course, the heat equation can be interpreted as $L^2$ gradient flow of $\rho \mapsto \int |\nabla\rho|^2 dx$, as well as many other PDEs which can be written as gradient flows in the classical theory on Hilbert spaces started by Brezis [12]. Gradient flows on a Hilbert spaces are a special case of the theory of maximal monotone operators, a branch of semigroup theory which (among other achievements) unified the mathematical theories of nonlinear diffusion equations and nonlinear scalar conservation laws, see e. g. [31]. As a byproduct, the semigroup generated by the transport operator of a scalar conservation law (in the entropy solution sense of [46]) is non-expansive in $L^1$, as it happens for nonlinear diffusion equations.

The first results in the direction of linking conservation laws with Wasserstein gradient flows go back to the early 2000 years. Since then, some interesting achievements have been carried out, but the theory is still incomplete.

The goal of this note is twofold. We shall first review the existing literature on this subject, trying to identify an overall advancement toward the current state of the art. As a second goal we shall propose some open problems and new possible directions (see in particular sections 7 and 8), mostly based on a microscopic (deterministic) particle formulation for a scalar conservation law.

Due to the fact the structure of a scalar conservation law changes drastically when passing from one space dimension to many space variables, all the results (as well as the coverage of this note) are restricted to the one dimensional case.[1]

As for the deterministic particle formulation mentioned above, we stress that several results have been achieved in the case of stochastic particles, see e. g. [54, 30]. We will not deal with such approach here, and send the reader to the proposed references.

This note is structured as follows. We will follow a partly chronological order, and collect the results under review according to the following scheme:

- We shall first introduce some basic facts about scalar conservation laws in section 2. We shall then recall the isomorphism of the 2-Wasserstein space with $L^2([0, 1])$, and introduce (in short) the technique of [28] on the study of the pseudo-inverse equation. In particular, we establish a technical result which allows to formulate a scalar conservation laws via a pseudo-inverse equation, see Theorem 3.1. We shall also introduce another possible pseudo-inverse-based approach which deals with *increasing* solutions.
- Although less natural and less intuitive, the afore-mentioned approach for increasing solutions was the first one to see some developments in this theory. Therefore, we shall explain the main ideas and the main results in sections 4 (based on the result of [7]) and 5. The latter shows a link between the Burgers equation and a class of nonlocal interaction equations. It was developed in [8].
- The problem of establishing a contraction in the Wasserstein space for a scalar conservation laws, this time for nonnegative integrable solutions, will be the main goal of the remaining part of the note. In section 6 we shall review an encouraging result proven in [17], namely an $\infty$-Wasserstein contraction result. As a minor original contribution of this note, we prove with a simple example that such result *cannot* be extended to $p$-distances with finite $p$. This fact suggests that the classical Wasserstein distance approach is not the best one to define a notion of gradient flow for conservation laws.
- To bypass such a problem, in section 7 we introduce a new approach of a gradient flow in the generalised Wasserstein space with a nonlinear mobility, see [42]. We show a rigorous gradient flow structure at a microscopic Lagrangian particle level, and we conjecture a variational structure based on results in [39]. The content of this section is completely new, though mostly formal.
- In section 8 we propose an alternative semi-implicit approach which is rigorous at a microscopic level, but totally open at the macroscopic one.
- Finally, in section 9 we quickly review some results in which nonlinear convection has been coupled with linear and nonlinear diffusion terms in a Wasserstein gradient flow approach.

## 2. Preliminaries on scalar conservation laws on the real line

We shall assume that the reader is familiar with the basic theory of scalar conservation laws in one space dimension. We refer to the book [11] and to the theory in Chapter 6 of [33] for further details and references. Here we only recall the most important concepts.

It is well known that the equation (4) produces discontinuous solutions even for some smooth initial data. A notion of weak solution is not enough to provide a well posed semigroup. The seminal work by Oleinik [50] and Kruzkov [46] established the correct (and physically relevant) notion of solution for (4) called *entropy solution*.

**Definition 2.1.** Let $\rho_0 \in L^\infty(\mathbb{R})$. A function $\rho \in L^\infty([0, +\infty) \times \mathbb{R})$ is called an *entropy solution* to (4) with initial condition $\rho_0$ if it satisfies the entropy inequality

$$\int_0^{+\infty} \int_{\mathbb{R}} [|\rho(t, x) - k|\varphi_t(t, x) + \text{sgn}(\rho(t, x) - k)[f(\rho(t, x)) - f(k)]\varphi_x(t, x)] \, dt \, dx$$
$$+ \int_{\mathbb{R}} |\rho_0(x) - k|\varphi(0, x) \, dx \geq 0$$

for all $\varphi \in C_c^\infty([0, +\infty) \times \mathbb{R})$, $\varphi \geq 0$, and for all constants $k \in \mathbb{R}$.

The existence and uniqueness of entropy solutions in the sense of Definition 2.1 for a locally Lipschitz flux $f$ and for a fixed initial condition $\rho_0 \in L^\infty(\mathbb{R})$ is well known since [46]. A later result [29] states that uniqueness can be obtained also by assuming a weaker condition at the initial time provided that $f$ is not affine on sets of positive measure: more precisely, under that assumption on $f$, $\varphi$ in definition 2.1 can be taken in $C_c^\infty((0, +\infty) \times \mathbb{R})$ and $\rho(t, \cdot)$ can be supposed to converge to $\rho_0$ *narrowly* as $t \searrow 0$, in order to still get a unique solution.

For the purpose of this note, and in all the results presented here, a key feature that is assumed is that solutions to (4) are *nonnegative*. The role of this extra assumption is well explained in the paper [48]. Here we want to

---

[1]A major exception to this rule is a paper by Yann Brenier [10], with results valid on a multi-dimensional torus.

stress the following aspect: if one wants to replace the $L^\infty$ assumption for the initial condition by $\rho_0 \in L^1_+(\mathbb{R})$, the uniqueness result for entropy solutions is still true for fluxes of the form

$$f(\rho) = \rho^q, \qquad q > 1,$$

provided we require additionally

- $\rho \geq 0$,
- $\rho \in L^\infty([0, +\infty); L^1(\mathbb{R})) \cap L^\infty([\tau, +\infty) \times \mathbb{R})$ for all $\tau > 0$.

We refer to [48, Theorem 1.1] for the details. The importance of attaining an $L^1$ initial condition in a narrow sense is that we are allowed to consider *source-type* solutions, i. e. solutions with a single Dirac mass as initial condition. Such solutions are *self-similar* in the case $f(u) = u^q$ with $q > 1$, and they are called generalised $N$-waves, see [48].

## 3. PRELIMINARIES ON ONE DIMENSIONAL WASSERSTEIN DISTANCES

Throughout the whole note, $\mathcal{P}(\mathbb{R})$ denotes the space of probability measures on $\mathbb{R}$. Let $p \in [1, +\infty)$. We set

$$\mathcal{P}_p(\mathbb{R}) = \left\{ \mu \in \mathcal{P}(\mathbb{R}) : \int |x|^p d\mu(x) < +\infty \right\}.$$

We also set

$$\mathcal{P}_\infty(\mathbb{R}) = \{ \mu \in \mathcal{P}(\mathbb{R}) : \mu \text{ has compact support} \}.$$

For $p \in [1, +\infty]$, the $p$-Wasserstein distance between $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$ is usually defined via an optimal transportation problem, see e. g. [56, 2]. However, the definition is much simpler in the one dimensional case. Let

$$\mathcal{K} = \left\{ X \in L^2([0, 1]) : X \text{ is non decreasing} \right\}.$$

We define the map $\mathcal{P}(\mathbb{R}) \ni \mu \mapsto X_\mu \in \mathcal{K}$ via

$$X_\mu(z) = \inf \left\{ x \in \mathbb{R} : \mu((-\infty, x]) \geq z \right\}, \tag{5}$$

i. e. $X_\mu$ is the *inverse cumulative distribution function* corresponding to the probability measure $\mu$. Then (see [56]), let $p \in [1, +\infty]$ and let $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$. The $p$-Wasserstein distance between $\mu$ and $\nu$ is recovered as

$$d_p(\mu, \nu) = \|X_\mu - X_\nu\|_{L^p([0,1])}. \tag{6}$$

Let us note in particular that

- When $p = 1$, this distance coincides with the $L^1(\mathbb{R})$ difference

$$d_1(\mu, \nu) = \int_{\mathbb{R}} |F_\mu(x) - F_\nu(x)| \, dx, \qquad F_\mu(x) = \mu((-\infty, x]), \quad F_\nu(x) = \nu((-\infty, x]).$$

- When $p = +\infty$, this distance controls the set distance between the two supports of $\mu$ and $\nu$, namely

$$\text{dist}(\text{supp}(\mu), \text{supp}(\nu)) \leq d_\infty(\mu, \nu).$$

We recall that an atomic measure $\mu = \delta_x$, $x \in \mathbb{R}$ yields

$$X_\mu(z) \equiv x, \quad \text{for all } z \in [0, 1].$$

On the other hand, if $\mu \in \mathcal{P}(\mathbb{R})$ has no atoms, this implies that $X_\mu$ is strictly increasing on $[0, 1]$.

### 3.1. **Contraction properties in Wasserstein spaces.** The main advantage of working in one space dimension and using the identification (6) is explained as follows. Consider an evolutionary partial differential equation of the form

$$\rho_t + (\rho v(\rho, \rho_x))_x = 0 \tag{7}$$

with initial datum in $\mathcal{P}_\infty(\mathbb{R})$. Assume for simplicity that the solution $\rho$ to (7) is smooth and bounded for all times in such a way that the pseudo-inverse $X_\rho$ (in which, by abuse of notation, we identify a measure with its absolutely continuous part) is the actual inverse of $F(t, x) = \int_{-\infty}^x \rho(t, y) dy$. We then have

$$X(t, F(t, x)) = x. \tag{8}$$

Due to the regularity we have assumed, we can differentiate (8) once with respect to $x$,

$$X_z(t, F(t, x))\rho(t, x) = 1,$$

and obtain

$$\rho(t, x) = \frac{1}{X_z(t, F(t, x))}. \tag{9}$$

Note that $\rho$ being bounded is equivalent to have $X_z$ uniformly bounded from below by a positive constant. By differentiating (8) twice with respect to $x$ we obtain

$$X_{zz}(t, F(t,x))\rho(t,x)^2 + X_z(t, F(t,x))\rho_x(t,x) = 0,$$

which gives

$$\rho_x(t,x) = -\frac{X_{zz}(t, F(t,x))}{X_z^3(t, F(t,x))} = \left(\frac{1}{2X_z^2}\right)_z (t, F(t,x)). \tag{10}$$

Now, by integrating (7) on $(-\infty, x]$, we formally get the equation for $F(t,x) = \int_{-\infty}^x \rho(t,y)dy$

$$F_t + \rho v(\rho, \rho_x) = 0, \tag{11}$$

and by differentiating (8) with respect to $t$ we get

$$X_z(t, F(t,x))F_t(t,x) + X_t(t, F(t,x)) = 0,$$

which gives, after changing variable $X(t,z) = x$,

$$X_t(t,z) = -\frac{F_t(t, X(t,z))}{\rho(t, X(t,z))},$$

and (11), (9), and (10) imply the following PDE for $X = X(t,z)$:

$$X_t = v\left(X_z^{-1}, \left(\frac{1}{2}X_z^{-2}\right)_z\right). \tag{12}$$

The form (12) for the original PDE (7) is often called *pseudo-inverse equation*. Roughly speaking, estimating the $p$-Wasserstein distance between two solutions $\rho_1$ and $\rho_2$ to (7) is equivalent to computing the $L^p$ distance between the two corresponding pseudo-inverse variables $X_{\rho_1}$ and $X_{\rho_2}$ upon a direct use of the PDE (12).

The most striking example of application of the pseudo-inverse form for an evolutionary PDE is provided in [28, 22] for the Porous Medium Equation (PME) (see also [44])

$$\rho_t = (\rho^m)_{xx}.$$

In the $X = X_\rho$ variable, the PME can be formally rewritten as

$$\partial_t X(t,z) = -\partial_z \left((\partial_z X(t,z))^{-m}\right).$$

Let $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R})$ be two (bounded) solutions to the PME. A direct application of the above strategy in [28, 22] leads to the contraction property

$$d_2(\rho_1(t), \rho_2(t)) \le d_2(\rho_1(0), \rho_2(0)).$$

The examples considered in [28, 22] showed that the pseudo-inverse form is a very powerful tool to investigate contraction properties of a semigroup in one-dimensional Wasserstein spaces. In the special case of a gradient flow (1) on $\mathcal{P}_2$, in which $v = \partial \mathcal{F}[\rho]$, one can formally see that the pseudo-inverse form of (1) becomes

$$X_t = -\partial \tilde{\mathcal{F}}[X], \tag{13}$$

where

$$\tilde{\mathcal{F}}[X] = \mathcal{F}[X_z^{-1}(F(\cdot))], \quad F(x) = \int_{-\infty}^x \rho(y)dy.$$

While the gradient of the functional $\mathcal{F}$ in (1) has to interpreted as a sub-differential on the $\mathcal{P}_2$ Riemannian structure, the expression $\partial \tilde{\mathcal{F}}[X]$ in (13) denotes the *classical $L^2$ sub-differential*, see [12]. The identification between these two structures is not always immediate, and we shall return to it later on in this note, see section 5.

We now prove that the pseudo inverse equation (12) is well justified in the case of a scalar conservation law with a reasonable flux. In order to point out this property, we rely on some existing results on the SBV regularity of scalar conservation laws. For simplicity we shall work under the assumption of compactly supported initial data.

**Theorem 3.1.** *Let $f : [0, +\infty) \to [0, +\infty)$ be such that $f \in C^2$ with $f'' > 0$ on $[0, +\infty)$. Let $\overline{\rho} \in \mathcal{P}(\mathbb{R}) \cap L^\infty(\mathbb{R})$ be compactly supported, and let $\rho$ be the corresponding unique entropy solution to (4). Then, the pseudo-inverse variable $X = X_\rho$ satisfies*

$$X_t(t,z) = v(X_z(t,z)^{-1}), \qquad \text{almost everywhere on } (t,z) \in [0, +\infty) \times [0,1], \tag{14}$$

*with $v(\rho) = f(\rho)/\rho$.*

*Proof.* The proof is provided in the Appendix A. $\square$

3.2. **Contraction properties of space-derivatives.** An advantage of the approach in subsection 3.1 is that it describes a contraction property of the semigroup in its *Lagrangian* formulation, i. e. the pseudo-inverse equation (12) can be seen as the equation for the Lagrangian position of particles (e. g. in a fluid) computed as a function of the *mass* variable and the time variable. On the other hand, this approach provides a contraction estimate in a Wasserstein distance, which induces a topology only slightly stronger than the weak convergence in the sense of measures.

There is an alternative way to still use the Wasserstein distance but at the same time describe the semigroup in a stronger topology. Unfortunately, it only applies to non decreasing solutions on the real line (hence, solutions with infinite mass). Let $F_0 \in L^\infty(\mathbb{R})$, $F_0(x) \in [0,1]$, $F_0$ monotone non-decreasing with

$$\lim_{x \to -\infty} F_0(x) = 0, \qquad \lim_{x \to +\infty} F_0(x) = 1. \tag{15}$$

Consider an evolutionary PDE of the form

$$F_t + f(F, F_x)_x = 0, \tag{16}$$

and assume that (16) preserves the monotonicity property, i. e. $F(t, \cdot)$ is a non-decreasing function, the property $F(x,t) \in [0,1]$, and the limiting properties (15) for all $t \geq 0$. Hence, the distributional space derivative $\mu(t) = F_x(t, \cdot)$ of $F$ is a probability measure, and it makes sense to compute the evolution of the quantity $d_p(F_x(t, \cdot), G_x(t, \cdot))$ for two solutions $F$ and $G$ to (16) with the above properties. Here $p \in [1, +\infty]$.

With the choice $p = 1$, we remark that

$$d_1(F_x(t, \cdot), G_x(t, \cdot)) = \|F(t, \cdot) - G(t, \cdot)\|_{L^1(\mathbb{R})},$$

which suggests that this approach provides a description of the semigroup in stronger topologies.

## 4. INCREASING SOLUTIONS OF SCALAR CONSERVATION LAWS

A first class of interesting results trying to link scalar conservation laws and Wasserstein gradient flows concerns with *increasing* solutions.

Consider the one dimensional scalar conservation law

$$\frac{\partial F}{\partial t} + \frac{\partial}{\partial x} f(F) = 0, \qquad f \text{ locally Lipschitz}, \tag{17}$$

with initial datum $F(0, x) = F_0(x)$, $F_0 \in \mathcal{U}$ with

$$\mathcal{U} := \{F : \mathbb{R} \to \mathbb{R} : F \text{ non decreasing, right-continuous, with limits 0 at } -\infty \text{ and 1 at } +\infty\}.$$

The first result in this direction goes back to 2005 in [7], where the authors consider Lipschitz fluxes $f$ and entropy solutions to (17). By a simple maximum principle, the unique entropy solution $F$ to (17) satisfies $F(t, \cdot) \in \mathcal{U}$ for all $t \geq 0$. So we are in the context of subsection 3.2. The idea of [7] is to consider the formal PDE for the pseudo inverse of $F$

$$X(t, w) = \inf\{x \in \mathbb{R} : F(x) > w\}.$$

Assuming for simplicity that $F$ is *strictly* increasing and smooth enough on a given nontrivial interval $x \in [a, b]$, similar to the computation in subsection 3.1 one can use (17) directly to get

$$X_t(t, w) = -\frac{F_t(t, X(t, w))}{F_x(t, X(t, w))} = f'(F(t, X(t, w))) = f'(w). \tag{18}$$

The equation in (18) is just an ordinary differential equation with a parameter $w \in [0, 1]$. Its solution is easily computed as

$$X(t, w) = X(0, w) + t f'(w), \tag{19}$$

which is very reminiscent of the expression for characteristic lines of (17). Now, with the notation of subsection 3.1,

$$X(t, w) = X_{\mu(t)}(w) \qquad \mu(t) = \partial_x F(t, \cdot), \qquad F \text{ solution to (17).}$$

Therefore, given two non decreasing initial conditions $F_{0,1}, F_{0,2} \in \mathcal{U}$, given $F_1$ and $F_2$ the two corresponding entropy solutions with initial data $F_{0,1}$ and $F_{0,2}$ respectively, we clearly have that the two distributional derivatives $(F_1)_x$ and $(F_2)_x$ belong to $\mathcal{P}(\mathbb{R})$. Therefore, we can use (19) and compute for $p \in [1, +\infty]$,

$$\begin{aligned} d_p((F_1)_x(t, \cdot), (F_2)_x(t \cdot)) &= \|X_{(F_1)_x}(t, \cdot) - X_{(F_2)_x}(t, \cdot)\|_{L^p([0,1])} \\ &= \|X_{(F_1)_x}(0, \cdot) - X_{(F_2)_x}(0, \cdot) + t f'(\cdot) - t f'(\cdot)\|_{L^p([0,1])} \\ &= \|X_{(F_{1,0})_x} - X_{(F_{2,0})_x}\|_{L^p([0,1])} = d_p((F_{1,0})_x, (F_{2,0})_x). \end{aligned} \tag{20}$$

The above computation is only formal, in that it only holds if $F$ is strictly increasing and smooth. Due to possible shocks and rarefaction waves, $F$ may exhibit angular points and intervals in which it is constant in $x$. It turns out that the equality (20) becomes an inequality in the general case.

**Theorem 4.1** ([7]). *Given a locally Lipschitz $f : \mathbb{R} \to \mathbb{R}$ and two initial data $F_{1,0}, F_{2,0} \in \mathcal{U}$, let $F_1$ and $F_2$ be the two corresponding entropy solutions to* (17). *Then, for all $t \geq 0$ and $p \geq 1$, we have (with possibly infinite values)*

$$d_p((F_1)_x(t, \cdot), (F_2)_x(t\cdot)) \leq d_p((F_{1,0})_x, (F_{2,0})_x). \tag{21}$$

The results in [7] provide a contraction property in the $d_p$ distances for scalar conservation laws. Although comparing *derivatives* of solutions (and not the solutions themselves), this result has the merit of being the first one to link the theory of scalar conservation laws with Wasserstein spaces. In particular, (21) in the case $p = 1$ is the classical $L^1$ contraction property for scalar conservation laws, see [33]. However, no relation with gradient flows came out until Yann Brenier beautifully extended the approach of [7] to the multidimensional case in 2009 in [10]. In his paper, the author expresses the solution to the conservation laws by a combination of level-set, kinetic, and transport-collapse approximation. The connection with maximal monotone operator theory appears here for the first time, though only indirectly.

A clear connection of a scalar conservation law with Wasserstein gradient flows in one space dimension was established years later in [8]. We shall briefly review this result in the next section.

## 5. THE GRADIENT FLOW OF THE ONE-DIMENSIONAL NEWTON ENERGY

Consider the nonlocal interaction equation

$$\partial_t \mu = \partial_x (\mu \, \partial_x W * \mu), \qquad x \in \mathbb{R}, \quad t > 0, \tag{22}$$

where $W$ is either the repulsive or the attractive Newton potential in one space dimension

$$W(x) = -|x| \qquad \text{or} \qquad W(x) = |x|. \tag{23}$$

Then (22) can be interpreted as the 2-Wasserstein gradient flow of the functional

$$\mathcal{W}[\mu] = \frac{1}{2} \int_{\mathbb{R} \times \mathbb{R}} W(x - y) d\mu(x) d\mu(y),$$

defined on the space $\mathcal{P}_2(\mathbb{R})$.

*Remark* 5.1 (Particle solutions). The equation (22) is the PDE counterpart of the system of interacting particles

$$\dot{x}_i(t) = -\frac{1}{N} \sum_{k \neq i} W'(x_i(t) - x_k(t)), \qquad i = 1, \ldots, N. \tag{24}$$

In the attractive case $W(x) = |x|$, the results in [20] guarantee that a *moving particle solution*

$$\mu(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i(t)}, \qquad x_1(t), \ldots, x_N(t) \text{ solutions to } (24),$$

is a rigorous solution to (22) in the 2-Wasserstein gradient flow sense. We omit the details for simplicity. We only remark that the main reason behind this fact is that the functional $\mathcal{W}$ is displacement convex in the Wasserstein sense, and this allows uniqueness of gradient flow solutions. At the microscopic level (i. e. at the particle ODE level (24)), the functional $\mathcal{W}$ is replaced by

$$\hat{\mathcal{W}}(X) = \frac{1}{2N^2} \sum_i \sum_j W(x_i - x_j), \qquad X = (x_1, \ldots, x_N) \in \mathbb{R}^N,$$

which is convex on $\mathbb{R}^N$, and therefore the system (24) can be rephrased as

$$\dot{X}(t) \in -\nabla \hat{\mathcal{W}}(X(t)),$$

where the gradient above should be interpreted as a classical sub-differential on the finite dimensional Hilbert space $(\mathbb{R}^N, \| \cdot \|_N)$, where $\|X\|_N^2 = \frac{1}{N} \sum_{i=1}^{N} x_i^2$. Once again, convexity implies uniqueness of particle solutions.

Now, for a given solution $\mu(t)$ of (22), set

$$F(t, x) = \mu(t)((-\infty, x]).$$

By integrating (22) on $(-\infty, x]$ we obtain

$$\partial_t F(t,x) = F_x(t,x) \left(\partial_x W * \mu\right)(t,x) = \pm F_x(t,x) \left(\text{sgn} * \mu\right)(t,x)$$

$$= \pm F_x(t,x) \int_{-\infty}^{+\infty} \text{sgn}(x-y) d\mu(y) = \pm F_x(t,x) \left[\int_{-\infty}^{x} d\mu(y) - \int_{x}^{+\infty} d\mu(y)\right]$$

$$= \pm F_x(t,x) \left[2\mu(t)((-\infty,x]) - 1\right] = \pm F_x(t,x)(2F(t,x)-1) = \pm \partial_x(F^2 - F)(t,x).$$

Therefore, $F$ is formally a solution to the Burgers' type equation

$$F_t \pm (F - F^2)_x = 0. \tag{25}$$

We remark that $F(t, \cdot)$ belongs to the set $\mathcal{U}$ considered in section 4.

The natural question arises on whether $F$ is an entropy solution for (25) provided $\mu$ is a Wasserstein gradient flow for (22) and viceversa. This task has been addressed in [8].

### 5.1. **Identification with $L^2$ gradient flows.**
Let us consider the pseudo-inverse equation for (22). Let $\mu$ be a gradient flow solution to (22) and let $X : [0, +\infty] \times [0,1] \to \mathbb{R}$ be the pseudo inverse variable $X = X_{\mu(t)}$. Formally, $X$ satisfies

$$X_t(t,z) = -\int_0^1 W'(X(t,z) - X(t,\zeta)) d\zeta = \mp \int_0^1 \text{sign}(X(t,z) - X(t,\zeta)) d\zeta. \tag{26}$$

We observe that the above right hand side is the $L^2$ gradient of the functional $\tilde{\mathcal{W}}[X] : L^2([0,1]) \to \mathbb{R}$

$$\tilde{\mathcal{W}}[X] = \frac{1}{2} \int_0^1 \int_0^1 W(X(z) - Z(\zeta)) \, dz \, d\zeta.$$

In the attractive case $W(x) = |x|$, we compute for $X \in \mathcal{K}$

$$\tilde{\mathcal{W}}[X] = \frac{1}{2} \int_0^1 \int_0^1 |X(z) - X(\zeta)| \, dz \, d\zeta$$

$$= \frac{1}{2} \iint_{X(z) \geq X(\zeta)} (X(z) - X(\zeta)) \, dz \, d\zeta - \frac{1}{2} \iint_{X(z) \leq X(\zeta)} (X(z) - X(\zeta)) \, dz \, d\zeta$$

$$= \iint_{X(z) \geq X(\zeta)} (X(z) - X(\zeta)) \, dz \, d\zeta.$$

Since $X$ is non decreasing, we clearly have

$$\tilde{\mathcal{W}}[X] = \iint_{z \geq \zeta} (X(z) - X(\zeta)) \, dz \, d\zeta = \int_0^1 \int_0^z X(z) d\zeta \, dz - \int_0^1 \int_\zeta^1 X(\zeta) \, dz \, d\zeta$$

$$= \int_0^1 z X(z) \, dz - \int_0^1 (1 - \zeta) X(\zeta) \, d\zeta = \int_0^1 (2z - 1) X(z) \, dz,$$

and since the above is the expression for a *linear* functional, the $L^2$ gradient is easily computed as

$$\partial \tilde{\mathcal{W}}[X](z) = 2z - 1$$

and the equation (26) becomes in this case

$$X_t(t,z) = 1 - 2z.$$

Therefore, it is immediately seen that a bounded strictly increasing initial datum $X(0,z)$ will evolve to a constant state at a finite time $t = t^*$. At $t = t^*$ the dynamical system (26) reaches the boundary of the convex cone $\mathcal{K} \subset L^2([0,1])$, and it could possibly exit $\mathcal{K}$ after $t^*$. To bypass this problem, we consider the new functional $\mathcal{W} : L^2([0,1]) \to \mathbb{R}$

$$\overline{\mathcal{W}}[X] = \tilde{\mathcal{W}}[X] + \mathcal{I}_{\mathcal{K}}[X],$$

where $\mathcal{I}_{\mathcal{K}}$ is the indicator function of the convex cone $\mathcal{K}$,

$$\mathcal{I}_{\mathcal{K}} = \begin{cases} 1 & \text{if } X \in \mathcal{K} \\ 0 & \text{otherwise}. \end{cases}$$

Clearly, $\overline{\mathcal{W}}$ is convex on $L^2([0,1])$, and therefore it generates a unique solution for the gradient flow inclusion

$$X_t \in -\partial \overline{\mathcal{W}}[X]$$

in which $\partial$ denotes the $L^2$ subdifferential, see [12] for more details. For a given initial condition $\mu_0 \in \mathcal{P}_2(\mathbb{R})$, one has $X_0 = X_{\mu_0} \in \mathcal{K}$. The presence of the indicator function $\mathcal{I}_{\mathcal{K}}$ in the functional $\mathcal{W}$ avoids the solution $X$ with initial datum $X_0$ to escape from the convex cone $\mathcal{K}$ when two (or more) particles concentrate to one point. In particular,

it is easily seen that all particles collapse to a unique Dirac's delta at a finite time. We also notice that particles cannot split once they have collided.

A different situation arises in the repulsive case. There the driving functional has the expression

$$\tilde{\mathcal{W}}[X] = \int_0^1 (1 - 2z)X(z)\,\mathrm{d}z$$

and (26) is written as

$$X_t(t, z) = (2z - 1)\,,\tag{27}$$

for non-decreasing solutions. It is immediately seen that (27) preserves the strict monotonicity for all times. Roughly speaking,

$$X_z(t, z) = X_{0,z}(z) + 2t \geq X_{0,z}(z)$$

which implies indeed that any non increasing initial datum $X_0 \in \mathcal{K}$ turns into a strictly increasing one for positive times. Therefore, no indicator function needs to be considered here to keep the $L^2$ dynamical system (26) within the convex cone $\mathcal{K}$. We observe that in this case the driving functional of the $L^2$ gradient flow

$$X_t \in -\partial\tilde{\mathcal{W}}[X]$$

is linear, and therefore convex.

*Remark* 5.2 (Piecewise constant pseudo-inverses in the attractive case). Consider a piecewise constant initial condition $X_0 \in \mathcal{K}$

$$X_0(z) = \sum_{i=1}^N x_{0,i}\mathbf{1}_{[i/N, (i+1)/N)}\,, \qquad x_{0,1} < \ldots < x_{0,N}.$$

Then, it is immediately seen that the solution to the *attractive* gradient flow

$$X_t \in -\partial\overline{\mathcal{W}}[X]$$

is given by

$$X(t, z) = \sum_{i=1}^N x_i(t)\mathbf{1}_{[i/N, (i+1)/N)}\,,$$

where $x_1(t), \ldots, x_N(t)$ is the solution to the particle system

$$\dot{x}_i(t) = -\frac{1}{N}\sum_{x_k(t) \neq x_i(t)} \mathrm{sign}(x_i(t) - x_k(t))\,, \qquad i = 1, \ldots, N\,.\tag{28}$$

In both the attractive and the repulsive case, the identification between the $L^2$ gradient flow and the Wasserstein gradient flow are rigorously recovered in [8].

5.2. **Shocks vs. moving particles, rarefaction waves vs. diffused particles.** Consider the initial condition

$$F_0(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

for the Burgers' type equation (25). In the attractive case $W(x) = |x|$, (25) reads

$$F_t + f(F)_x = 0\,, \qquad f(F) = F - F^2\,,$$

a scalar conservation law with uniformly concave flux. Hence, the stationary shock wave

$$F(t, x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

is the unique entropy solution to (25). Due to remark 5.1, the distributional spatial derivative of $F$, i. e. the stationary Dirac's delta

$$\mu(t) = \partial_x F(t, x) = \delta_0\,,$$

is a gradient flow solution to (22).

Consider now the repulsive case $W(x) = -|x|$. Here, something slightly different happens compared to the attractive case, in terms of what we explained in remark 5.1. In particular, the situation described in remark 5.2 is no longer valid in this case. Due to the non zero repulsive force acting on particles on the same position, the *self-interaction* force cannot be neglected here when considering the particle version of (22). This is also easily seen in the $L^2$ equation (27), in which all piecewise constant initial data becomes strictly increasing at positive

times. This simply means that *particles do not remain particles* in the repulsive case. This fact reflects also in the behaviour of the solution to the Burgers equation (25), which reads in this case

$$F_t + f(F)_x = 0, \qquad f(F) = -F + F^2,$$

i. e. the flux $f$ is convex, and initial discontinuities are solved through a *rarefaction wave*. Indeed, with the initial condition $F_0$ we can provide the explicit solution

$$F(t, x) = \begin{cases} 0 & \text{if } x < -t \\ \frac{1}{2} + \frac{x}{2t} & \text{if } -t \le x \le t \\ 1 & \text{if } x > t. \end{cases}$$

The space derivative of $F$ can be computed as

$$\mu(t) = \partial_X F(t, \cdot) = \frac{1}{2t} \mathbf{1}_{[-t, t]}$$

Physically speaking, in terms of the nonlocal equation (22), the initial particle set at $x = 0$ *diffuses* to a continuum $L^1$ state for all positive times. Such state decays to zero as $t \to +\infty$.

5.3. **A particle method for the repulsive case.** The example considered in subsection 5.2 suggests a deeper link between the repulsive PDE

$$\mu_t + (\mu(|\cdot| * \mu)_x)_x = 0 \tag{29}$$

and the scalar conservation law

$$F_t + (F^2 - F)_x = 0. \tag{30}$$

It is well known that the solution semigroup for the equation (30) in the entropy sense can be constructed via the so-called *wave front tracking* (WFT) algorithm, see [32, 11]. In the present context, this method works as follows. Consider an initial condition $F_0 \in \mathcal{U}$. Suppose for simplicity that

$$\inf\{x \in \mathbb{R} : F(x) > 0\} = \underline{x} > -\infty, \qquad \inf\{x \in \mathbb{R} : F(x) = 1\} = \overline{x} < +\infty.$$

For $N \in \mathbb{N}$ large, let us set for $i = 0, \ldots, N - 1$,

$$x_{0,i} = \inf\{x \in \mathbb{R} : F_0(x) > i/N\}.$$

Note that $x_{0,0} = \underline{x}$. We then approximate $F_0$ by the piecewise constant function

$$F_0^N = \sum_{i=0}^{N-1} \frac{i}{N} \mathbf{1}_{[x_{0,i}, x_{0,i+1})}.$$

The goal of the WFT algorithm is that of approximating the solution to a conservation law by the superposition of solutions of Riemann problems at the discontinuity points of $F_0$. The idea is that of having the approximated solution be piecewise constant too for all times. Clearly, this is not possible in our case if we let the discontinuous initial condition $F_0^N$ evolve via (30), because a convex flux smooths out all increasing discontinuities to rarefaction waves. To bypass this problem, the WFT algorithm replaces the uniformly convex flux $f(F) = F^2 - F$ in (30) by the piecewise linear one

$$f^N(F) = \sum_{j=1}^{N} N \left[ f\left(\frac{j}{N}\right) - f\left(\frac{j-1}{N}\right) \right] \left( F - \frac{j-1}{N} \right) \mathbf{1}_{[\frac{j}{N}, \frac{j-1}{N})}(F).$$

Then, we set $F^N$ as the unique entropy solution to the Cauchy problem

$$F_t + f^N(F)_x = 0, \qquad F(0, x) = F_0^N(x). \tag{31}$$

The flux $f^N$ is designed on purpose in a way that all discontinuities of $F_0^N$ are solved by a contact discontinuity travelling with constant velocity

$$\lambda_j = \frac{f\left(\frac{j}{N}\right) - f\left(\frac{j-1}{N}\right)}{\frac{1}{N}} = \frac{j-1}{N} - \frac{N-j}{N}, \quad j = 1, \ldots, N, \tag{32}$$

obtained via the Rankine-Hugoniot condition (see [33]). This is due to the fact that the flux $f^N$ is linearly degenerate on each interval $[(j-1)/N, j/N]$ on which the Riemann problems are solved. We can therefore write an explicit formula for the discontinuity points

$$x_i(t) = x_{0,i} + t\lambda_j,$$

with $\lambda_j$ as in (32), and for the solution $F^N$ to (31)

$$F^N(t,x) = \sum_{j=0}^{N-1} \frac{j}{N} \mathbf{1}_{[x_j(t), x_{j+1}(t))} + \mathbf{1}_{[x_N(t), +\infty)}.$$

We note that the jumps of $F^N$ are all of size $1/N$, which is reasonable since in the $N \gg 1$ limit they should be able to approximate rarefaction waves. In view of well known results on the convergence of the WFT algorithm [32, 11] (which actually sound a bit redundant in this context, since the convergence of the scheme is quite straightforward in this case), we have (up to possibly extracting a subsequence),

$$F^N \to F, \qquad \text{strongly in } L^1_{loc}([0, +\infty) \times \mathbb{R}), \tag{33}$$

with $F$ being the unique entropy solution to (30) with initial datum $F_0$.

Let us now interpret this result from the point of view of the nonlocal equation (29). The initial condition for (29) is $\mu_0 = \partial_x F_0$, an arbitrary probability measure with compact support. The piecewise constant approximation $F_0^N$ for $F_0$ implies an *atomization* for the initial condition $\mu_0$, i. e.

$$\mu_0^N = \frac{1}{N} \sum_{i=0}^{N-1} \delta_{x_{0,i}},$$

with $\mu_0^N \to \mu_0$ narrowly as $N \to +\infty$. As for the convergence result (33), this means that the curve of atomic measures

$$\mu^N(t) = \partial_x F^N(t, \cdot) = \frac{1}{N} \sum_{i=0}^{N-1} \delta_{x_i(t)}$$

converges in the 1-Wasserstein distance towards the unique gradient flow solution $\mu(t)$ of (29). In short, $\mu^N$ provides an atomic particle approximation for the gradient flow solution of (29). The formula (32) shows that the correct discrete counterpart of the repulsive equation (29) is a particle system in which *self-interaction is neglected*.

Although quite simple to prove, and related to known results in the theory of conservation laws, the above results are relevant in that

- it allows to atomize a continuum solution (we recall that no particle solutions exist for (29)),
- at least in the one dimensional case, it allows to stretch the particle approximation theory for nonlocal interaction equations developed in [16] up to the Newtonian potential.

We refer to [8] for further details.

## 6. Contraction in Wasserstein spaces

The results reviewed in sections 4 and 5 reveal an intriguing connection between scalar conservation laws and gradient flows, which however turns out to be only partly satisfactory in that it only applies to *increasing* solutions. The original question, i. e. whether or not it is possible to write a scalar conservation law

$$\rho_t + f(\rho)_x = 0, \tag{34}$$

as a Wasserstein gradient flow in the space of probability measures, still remains without an answer at this stage. In 2006, the paper [17] provided a promising result on a *contraction property* for scalar conservation laws with convex fluxes $f$ such that $f(0) = f'(0) = 0$, in the $\infty$-Wasserstein distance. We briefly review this result.

Assume the flux function $f : [0, +\infty) \to \mathbb{R}$ is uniformly convex and increasing, and let $\rho_0 \in L^\infty(\mathbb{R}) \cap \mathcal{P}(\mathbb{R})$, $\rho_0$ compactly supported for simplicity. It is well known that the primitive variable

$$F(t,x) = \int_{-\infty}^x \rho(t,y) \, \mathrm{d}y$$

satisfies the Cauchy problem for the Hamilton-Jacobi equation

$$\begin{cases} F_t + f(F_x) = 0 \\ F(0,x) = \int_{-\infty}^x \rho_0(y) \, \mathrm{d}y. \end{cases} \tag{35}$$

Due to a simple minimum principle for the scalar conservation law (34), $F(t, \cdot)$ is an $L^\infty(\mathbb{R})$ non decreasing function with

$$\lim_{x \to -\infty} F(t,x) = 0, \qquad \lim_{x \to +\infty} F(t,x) = 1.$$

The unique semi-concave solution (see e. g. [14]) to (35) is provided via the famous Hopf-Lax formula

$$F(t,x) = \min_{y \in \mathbb{R}} \left\{ tf^* \left( \frac{x-y}{t} \right) + F_0(y) \right\} , \tag{36}$$

where $f^*$ stands for the Legendre transform of $f$. Then, well known results on the relation between the Hamilton-Jacobi equation and scalar conservation laws imply that distributional derivative $\rho(t,x) = \partial_x F(t,x)$ is the unique entropy solution to (34) with initial condition $\rho_0$.

Following the ideas described in subsection 3.1, the information (36) on the primitive of the solution $\rho$ to (34) may turn out to be essential if one wants to investigate contraction properties of the conservation law semigroup in Wasserstein spaces. In order to achieve this task, we have to use (36) to recover a similar formula for the pseudo-inverse of $F$, namely for the function $X : [0, +\infty) \times [0,1] \to \mathbb{R}$ defined by

$$X(t,z) = \inf\{x \in \mathbb{R} : F(t,x) > z\}.$$

We assume as an extra requirement that

$$\mathrm{supp}(\rho_0) \text{ is a connected interval,}$$

which implies that $F_0$ is strictly increasing on some interval $(a,b)$, $F_0(x) = 0$ for $x \leq a$, and $F_0(x) = 1$ for $x \geq b$. Under this set of assumptions, the following formula holds:

$$X(t,z) = \max_{0 \leq w \leq z} \left\{ tg \left( \frac{z-w}{t} \right) + X_0(w) \right\} , \tag{37}$$

where $X_0$ is the pseudo-inverse of $F_0$ and $g : [0, +\infty) \to \mathbb{R}$ is the inverse of the positive branch of $f^*$, which is well defined due to the uniform convexity of $f$. The proof of (37) is performed in [17, Lemma 2.3]. Here we only sketch the main idea behind it in the appendix B.

Now, consider two compactly supported initial conditions $\overline{\rho}_1, \overline{\rho}_2 \in \mathcal{P}(\mathbb{R})$. Assume further that the supports of $\overline{\rho}_1$ and $\overline{\rho}_2$ are connected intervals. Consider the two corresponding primitive variables $\overline{F}_1$ and $\overline{F}_2$, and their pseudo inverses $\overline{X}_1, \overline{X}_2 \in L^\infty([0,1])$. From (37) we have

$$X_i(t,z) = \max_{0 \leq w \leq z} \left\{ tg \left( \frac{z-w}{t} \right) + \overline{X}_i(w) \right\} , \qquad i = 1, 2 .$$

Now, for fixed $(t,z) \in [0, +\infty) \times [0,1]$, let $\bar{w}_1 = \bar{w}_1(t,z)$ be the point in which the maximum above is attained for $X_1$. We clearly have

$$X_1(t,z) - X_2(t,z) \leq tg \left( \frac{z - \overline{w}_1}{t} \right) + \overline{X}_1(\overline{w}_1) - tg \left( \frac{z - \overline{w}_1}{t} \right) - \overline{X}_2(\overline{w}_1)$$

$$= \overline{X}_1(\overline{w}_1) - \overline{X}_2(\overline{w}_1) \leq \|\overline{X}_1 - \overline{X}_2\|_{L^\infty([0,1])}$$

By interchanging the roles of $\overline{X}_1$ and $\overline{X}_2$ and taking the supremum over $z \in [0,1]$, we get

$$\|X_1(t\cdot) - X_2(t, \cdot)\|_{L^\infty([0,1])} \leq \|\overline{X}_1 - \overline{X}_2\|_{L^\infty([0,1])}.$$

An approximation procedure (see the appendix of [17]) allows to extend the above formula to all initial data in $BV(\mathbb{R})$. Therefore we have the following contraction property.

**Theorem 6.1** ([17]). *Let $\overline{\rho}_1, \overline{\rho}_2 \in \mathcal{P}(\mathbb{R})$ be compactly supported and with bounded variation. Let $\rho_1$ and $\rho_2$ be the two corresponding entropy solutions to (34), with $f$ a uniformly convex flux with $f(0) = f'(0) = 0$. Then,*

$$d_\infty(\rho_1(t, \cdot), \rho_2(t, \cdot)) \leq d_\infty(\overline{\rho}_1, \overline{\rho}_2)$$

*for all $t \geq 0$.*

The result in Theorem 6.1 is used in [17] to detect the $t \to +\infty$ $L^1$ intermediate asymptotic profile in the $d_\infty$ distance under the extra assumption $\rho \mapsto f(\rho)^{1-\alpha}$ is convex for some $\alpha \in (0,1)$. Such intermediate profiles coincide with the classical $N$-waves considered in [48] recalled in section 2. We omit the details and send the reader to [17, Section 3]. Moreover, similarly to what proven in [22] for the porous medium equation, the above result provides a control for the speed of propagation of the support of any compactly solution to the scalar conservation law (34) in terms of the support of the $N$-waves.

Already at the time of [17], the authors wondered about the possibility that a scalar conservation law defined a contraction semigroup in the $d_p$ distances for $p < +\infty$. Actually, the answer to this question is negative in general. We shall prove that in the next (new) result.

**Theorem 6.2.** *Let $f : [0, +\infty) \to [0, +\infty)$ be a $C^2$ function, with $f'' > 0$ and such that $v(\rho) = f(\rho)/\rho$ is strictly decreasing. Let $p \in [1, +\infty)$. Then, there exist $\overline{\rho}_1, \overline{\rho}_2 \in \mathcal{P}(\mathbb{R}) \cap L^\infty(\mathbb{R})$ such that the two corresponding entropy solutions $\rho_1$ and $\rho_2$ to (34) satisfy*

$$\frac{d}{dt} d_p(\rho_1(t, \cdot), \rho_2(t, \cdot))\big|_{t=0} > 0.$$

*Proof.* From the pseudo-inverse equation (14), we have that the pseudo inverse variables $X_1 = X_{\rho_1}$ and $X_2 = X_{\rho_2}$ satisfy

$$\partial_t X_i(t, z) = v((\partial_z X_i(t, z))^{-1}), \tag{38}$$

with $v(\rho) = f(\rho)/\rho$. Now, consider

$$\overline{\rho}_1(x) = \mathbf{1}_{[0,1]}(x), \qquad \overline{\rho}_2(x) = \frac{1}{2}\mathbf{1}_{[0,2]}.$$

The corresponding pseudo-inverses $\overline{X}_i = X_{\overline{\rho}_i}$ are given by

$$\overline{X}_1(z) = z, \qquad \overline{X}_2(z) = 2z.$$

Therefore, by using (38) we have

$$\frac{d}{dt} d_p(\rho_1(t, \cdot), \rho_2(t, \cdot))^p \big|_{t=0} = \frac{d}{dt} \int_0^1 |X_1(t, z) - X_2(t, z)|^p dz \big|_{t=0}$$

$$= \int_0^1 \operatorname{sgn}(\overline{X}_1(z) - \overline{X}_2(z)) |\overline{X}_1(z) - \overline{X}_2(t, z)|^{p-1} \left( v(\partial_z \overline{X}_1(z)^{-1}) - v(\partial_z \overline{X}_2(z)^{-1}) \right) dz$$

$$= -\int_0^1 z^{p-1} \left( v(1) - v\left(\frac{1}{2}\right) \right) dz > 0,$$

in view of $v$ being strictly decreasing. $\qquad\square$

We remark that the use of Theorem 3.1 is not necessary here, since the two solutions $\rho_1$ and $\rho_2$ in the above proof are explicitly computable.

The result in Theorem 6.2 reveals that the Wasserstein topology is not the best one to deal with conservation laws (at least if one wants to prove a contraction property). The result in Theorem 6.1 says that a contraction property holds in a transport distance which only accounts for *global* properties of solutions, since $d_\infty(\rho_1, \rho_2)$ measures the largest path needed to transport mass from $\rho_1$ to $\rho_2$. When the cost in the transport distance is local as in the case of the $d_p$'s with $p < +\infty$, then no contraction property holds.

## 7. The nonlinear mobility approach: a microscopic viewpoint

The correct way to look at a scalar conservation law as a gradient flow is replacing the classical 2-Wasserstein distance with an alternative transport distance with a *nonlinear mobility*, see [42, 24, 15]. We shall describe such approach here. Already in [43] (see also [51]) the authors realized that this is the correct approach to a scalar conservation law. However, similarly to what happens in section 4, their approach does not apply to probability measures, but rather to a special class of solutions with limit zero at $-\infty$ and limit 1 at $+\infty$. We shall propose here a strategy based on the Lagrangian particle approximation developed in [39], which somehow follows some ideas proposed in an enlightening note written by Y. Brenier in 2000, see [9]. We shall start with a toy model description, which is presented in a more general framework in [57], see also clear connections with typical problems of calculus of variations arising in [14].

### 7.1. A finite dimensional 'extended' gradient flow.

Let $\ell > 0$ be fixed, and let $n \in \mathbb{N}$. Let $v : [0, +\infty) \to \mathbb{R}$ be a $C^1$ strictly decreasing function. Consider the convex cone

$$\mathcal{K}_{n+1} = \{X = (X_0, \ldots, X_n) \in \mathbb{R}^{n+1}, \ X_0 < \cdots < X_n\}, .$$

With the notation $P = (P_0, \ldots, P_n) \in \mathbb{R}^{n+1}$, consider the *Lagrangian* function $L : \mathcal{K}_{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}$ defined by

$$L(X, P) = \frac{\ell}{2} \sum_{i=0}^{n-1} \frac{1}{v(R_i)} P_i^2 + \frac{\ell}{2v(0)} P_n^2, \qquad R_i = \frac{\ell}{X_{i+1} - X_i}. \tag{39}$$

For $X, Y \in \mathcal{K}_{n+1}$, we consider the *action* functional

$$\mathcal{A} : C^1([0, 1], \mathbb{R}^{n+1}) \to \mathbb{R}, , \qquad \mathcal{A}[X] = \int_0^1 L(X(t), \dot{X}(t)) \, dt,$$

and the *cost* functional

$$\mathcal{C}(X, Y) = \inf \left\{ \mathcal{A}[X], \ X \in C^1\left([0, 1], \mathcal{K}_{n+1}\right), \ X(0) = X, \ X(1) = Y \right\}.$$

As last ingredient, let us consider the *linear drift* functional $F : \mathcal{K}_{n+1} \to \mathbb{R}$ defined by

$$F(X) = -\sum_{i=0}^{n} \ell X_i.$$

In the simple case $v(R) \equiv v(0)$ for all $R \geq 0$, the cost functional $\mathcal{C}$ coincides with the weighted squared Euclidean distance

$$\mathcal{C}(X, Y) = \frac{\ell}{v(0)} |X - Y|^2.$$

Hence, for a given time step $\tau > 0$, the optimality condition for the minimising scheme

$$\mathcal{K}_{n+1} \ni X_{k-1} \mapsto X_k = \operatorname{argmin} \left\{ \frac{1}{2\tau} \mathcal{C}(X_{k-1}, X) + F(X) \right\}$$

produces the implicit Euler approximation

$$\frac{X_k - X_{k-1}}{\tau} = -\frac{v(0)}{\ell} \nabla F(X_k) = -v(0)$$

of the gradient flow equation

$$\dot{X}(t) = -v(0). \tag{40}$$

When $v$ is non constant, the solution to the geodesic equation for $\mathcal{A}$ is more involved, and the above time-step approximation is less obvious to obtain. As noticed in [9], a better way to obtain a generalized notion of gradient flow in a distance induced by the action $\mathcal{A}$ is the following.

For a given $T > 0$, consider the *penalised action*

$$\mathcal{B} : C^1\left([0, T], \mathcal{K}_{n+1}\right) \times \mathcal{K}_{n+1} \to \mathbb{R}, \qquad \mathcal{B}[X(\cdot), Y] = \int_0^T L(X(t), \dot{X}(t)) \, \mathrm{d}t + F[Y],$$

and the value function

$$\mathcal{V}_T(X_0) = \inf \left\{ \mathcal{B}[X(\cdot), Y], \ (X(\cdot), Y) \in C^1\left([0, T], \mathcal{K}_{n+1}\right) \times \mathcal{K}_{n+1}, \ X(0) = X_0, \ X(T) = Y \right\}. \tag{41}$$

The optimality conditions for the minimisation problem (41) are

$$\frac{d}{dt} D_p L(X, \dot{X}) = D_X L(\dot{X}, X), \qquad \text{almost everyhere on } [0, T], \tag{42}$$

$$D_p L(Y, \dot{Y}) = -D_Y F[Y] \tag{43}$$

The set of conditions (42)-(43) and the condition $X(0) = X_0$ can be interpreted as a second order differential equation for $X(t)$ on $[0, T]$ with two boundary conditions, i. e.

$$\begin{cases} \frac{d}{dt} D_p L(X(t), \dot{X}(t)) = D_X L(\dot{X}(t), X(t)) & t \in (0, T) \\ X(0) = X_0 \\ D_p(X(T), \dot{X}(T)) = -DF(X(T)) \end{cases} \tag{44}$$

In the spirit of the JKO scheme [45], the final time $T$ in the Cauchy problem (44) should be interpreted as a small time step. Hence, one needs to solve (44) with initial condition $X(0) = X_0$, then repeat the procedure on $t \in [T, 2T]$ with $X(T)$ as initial condition and the same final condition as in (44) at time $2T$, and so on recursively on time intervals $t \in [kT, (k+1)T]$. This will define a sequence $X_0, X(T), X(2T), \dots, X(kT)$ which is expected to converge (upon a suitable interpolation) to a curve in $\mathbb{R}^n$ as $T \searrow 0$. Formally, the limit curve should satisfy the *generalised gradient flow equation*

$$D_p(X(t), \dot{X}(t)) = -DF(X(t)). \tag{45}$$

Before writing (45) explicitly, we point out that there is an alternative way to formally derive such an equation by interpreting the Lagrangian function $L$ as a Riemannian metric tensor, in the spirit of [52]. Let us write

$$L(X, P) = g_X(P, P), \tag{46}$$

where, for a given $X \in \mathcal{K}_{n+1}$ we have $g_X : \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}$ is the *inner tensor product*

$$g_X(U, V) = \frac{1}{2} \sum_{i=0}^{n} c_i(X) U_i V_i, \qquad c_i(X) = \begin{cases} \frac{\ell}{v(R_i)} & i = 0, \dots, n-1 \\ \frac{\ell}{v(0)} & i = N. \end{cases}$$

We then consider the Riemannian manifold $M$ having as tangent space on the point $X \in \mathcal{K}_{n+1}$ the linear space $T_X M$ given by $\mathbb{R}^{n+1}$ equipped with the tensor product $g$ above. Hence, the manifold $M$ is naturally equipped with the metric

$$d(X, Y) = \inf \left\{ \int_0^1 g_{X(t)}(\dot{X}(t), \dot{X}(t))^{1/2} dt, \ X(\cdot) \in C^1([0,1], \mathcal{K}_{n+1}), \ X(0) = X, \ X(1) = Y \right\}.$$

Now, for a given smooth functional $F$ on $M$, the notion of gradient of $F$ is uniquely defined as the vector field $\nabla F$ on $M$ with the property that $\nabla F(X) \in T_X M$ for all $X \in \mathcal{K}_{n+1}$ and such that

$$g_X(\nabla F(X), U) = \mathrm{diff}\, F_{X(t)}(U),$$

where $\mathrm{diff}_X F : T_X M \to \mathbb{R}$ is the differential of $F$,

$$\mathrm{diff}_X F(U) = \sum_{i=0}^n U_i \cdot \frac{\partial F}{\partial x_i}, \qquad \text{where } U = \sum_{i=0}^n U_i \frac{\partial}{\partial x_i}.$$

Then, the gradient flow $t \mapsto X(t)$ of the functional $F$ on $M$ is defined by

$$g_{X(t)}(\dot{X}, U) = g_{X(t)}(-\nabla F(X(t)), U), \qquad \text{for all } U \in C^1([0, +\infty); TM) \text{ such that } U(t) \in T_{X(t)} M, \tag{47}$$

which in view of the above definitions can be rewritten as

$$\sum_{i=0}^n c_i(X) \dot{X}_i(t) U_i = -\sum_{i=0}^n U_i \frac{\partial F}{\partial x_i}(X(t)), \qquad \text{for arbitrary } U \in C^1([0, +\infty); \mathbb{R}^{n+1}),$$

and this yields

$$c_i(X(t)) \dot{X}_i(t) = -\frac{\partial F}{\partial x_i}(X(t)). \tag{48}$$

We note that

$$D_P L(X(t), \dot{X}(t)) = (c_i(X) \dot{X}_i(t))_{i=1}^n,$$

so that (48) is in fact the same as the generalised gradient flow equation (45).

We now compute explicitly

$$D_{P_i} L(X, P) = \ell \left( \frac{1}{v(R_1)} P_1, \dots, \frac{1}{v(R_{n-1})} P_{n-1}, P_n \right).$$

Therefore, (45) can be written as the unique solution to the system of ODEs

$$\begin{cases} \dot{X}_i(t) = v \left( \frac{\ell}{X_{i+1}(t) - X_i(t)} \right) & i = 0, \dots, n-1 \\ \dot{X}_n(t) = v(0), \end{cases} \tag{49}$$

which is known in the literature as the Follow-the-leader system, see [39]. Such a system can be interpreted as a gradient flow on a *non-flat* metric. The agents tend to minimise $F$, i. e. they tend to maximise their barycenter in the positive direction. On a 'flat' metric structure, they would simply move with a constant positive speed, see (40). On the other hand, the presence of an agent $X_{i+1}$ before $X_i$ changes the metric perception for the agent $X_i$, and the latter is induced to move slower in a way which depends on the distance to $X_{i+1}$.

A major result in [39] is that the many particle limit of the system (49) with mass $\ell = \ell_n = 1/n$ is given by the unique entropy solution to the scalar conservation law (4) with $f(\rho) = \rho v(\rho)$ under the additional condition that $v$ is strictly increasing. This fact clearly suggests a gradient flow structure for a scalar conservation law, which should be inherited as a many particle limit from the structure we described above in this subsection. We explain this issue in the next subsection.

7.2. **A generalised gradient flow structure for scalar conservation laws.** The gradient flow structure found in subsection 7.1 is inherited by its continuum $n \to +\infty$ limit given by the scalar conservation law

$$\rho_t + (\rho v(\rho))_x = 0$$

in its Lagrangian formulation provided by Theorem 3.1

$$X_t = v \left( \frac{1}{X_z} \right).$$

To see this (at least formally), we introduce the metric tensor

$$g_X(U, V) = \int_0^1 \frac{1}{v(X_z^{-1}(z))} U(z) V(z) dz,$$

defined on the differential manifold

$$M = \left\{ X \in L^2([0,1]) \ : \ X \text{ is increasing} \right\},$$

i. e. $M$ is the interior of the convex cone $\mathcal{K}$ defined in section 3, with tangent space $T_X \mathcal{K} = L^2([0,1])$ for every $X \in M$.

We can therefore define the corresponding induced distance on $M$, i. e. for $X_0, X_1 \in M$ we have

$$d_v(X_0, X_1) = \inf \left\{ \int_0^1 \int_0^1 \frac{1}{v(X_z^{-1}(s,z))} U(s,z)^2 \, \mathrm{d}z \, \mathrm{d}s \, , \ X \in C^1([0,1]\,;M)\, , \ X(0) = X_0\, , \ X(1) = X_1 \right\},$$

which coincides, in its Eulerian formulation, to the generalised *Wasserstein distance with nonlinear mobility*

$$W(\rho_0, \rho_1) = \int \left\{ \int_0^1 \int_{\mathbb{R}} m(\rho(s,x)) w^2(s,x) \, \mathrm{d}x \, \mathrm{d}s \, , \ \rho_s + (m(\rho)w)_x = 0\, , \ \rho(0,x) = \rho_0(x)\, , \ \rho(1,x) = \rho_1(x) \right\}, \quad (50)$$

with the nonlinear function

$$[0, +\infty) \ni \rho \mapsto m(\rho) = \rho v(\rho)$$

called *mobility function*. Here $X = X_\rho$ according to the notation established in section 3. The class of distances (50) has been studied in [42, 24, 15], together with their relation with certain applied PDEs in which a nonlinear mobility effect is relevant in the dynamics, see e. g. [13] for chemotaxis movements.

Let us now consider the functional $\mathcal{F} : M \to \mathbb{R}$

$$\mathcal{F}[X] = - \int X(z) dz\,.$$

In order to define the generalised gradient flow of $\mathcal{F}$ with respect to the distance induced on $M$ by the metric tensor $g_X(U,V)$ defined above, we use an argument similar to the one in subsection 7.1, i. e. the gradient flow $t \mapsto X(t)$ of the functional $\mathcal{F}$ is defined by the 'weak' formulation

$$g_{X(t)}(X_t, U) = \int_0^1 \frac{1}{v\left(X_z(t,z)^{-1}\right)} X_t(t,z) U(t,z) \, \mathrm{d}z = \int_0^1 (\partial \mathcal{F})[X(t)](z) U(t,z) \, \mathrm{d}z\,, \quad (51)$$

for all

$$U \in \left\{ C^1([0,1], L^2([0,1]))\, , \ g_{X(t)}(X_t, U) < +\infty \right\}.$$

the symbol $\partial \mathcal{F} \in L^2([0,1])$ in (51) denotes the Frechét subdifferential of $\mathcal{F}$, easily computed as

$$(\partial \mathcal{F})[X](z) = 1\,, \qquad \text{for all } z \in [0,1]\,.$$

Due to the arbitrariness of $U$ in (51), we have that the generalised gradient flow $X(t,z)$ satisfies

$$X_t = v\left(X_z(t,z)^{-1}\right),$$

which is the Lagrangian formulation of the scalar conservation law $\rho_t + (\rho v(\rho))_x = 0$ according to Theorem 3.1.

*Remark* 7.1. The result in this section is, to a certain extent, formal. More precisely, one would like to construct the solution to a scalar conservation law in a variational way in the spirit of [45]. This task could be performed by taking the variational formulation in the discrete follow-the-leader case described in subsection 7.1, let the number of particles $n$ tend to infinity, use somehow the results in [39] and at the same time prove that the discrete metric structure constructed in subsection 7.1 converges to the continuum one proposed in this subsection. This task will be tackled in the future.

## 8. A SEMI-IMPLICIT APPROACH

The approach proposed in section 7 shows that, in the language of gradient flows, the nonlinearity effects in a scalar conservation laws are *due to the metric structure*, not to the driving functional. On the other hand, we show here that the follow-the-leader system can also be expressed via a semi-implicit time discretisation which uses the discrete version of the standard 2-Wasserstein distance.

Consider once again the $n+1$-particle follow-the-leader system with fixed mass $\ell = 1/n > 0$ of each particle, i. e.

$$\begin{cases} \dot{X}_i(t) = v\left(\dfrac{\ell}{X_{i+1}(t) - X_i(t)}\right) & i = 0, \ldots, n-1 \\ \dot{X}_n(t) = v(0)\,. \end{cases} \quad (52)$$

Let $x, y \in \mathbb{R}$. We set

$$\mathcal{V}(x;y) = \int_x^y v\left(\frac{\ell}{y-z}\right) \, \mathrm{d}z\,.$$

We define the *relative energy* of the state $X = (X_0, \ldots, X_n) \in \mathcal{K}_{n+1}$ with respect to the state $Y = (Y_0, \ldots, Y_n) \in \mathcal{K}_{n+1}$ as

$$\mathcal{R}[X;Y] = \sum_{i=0}^{n-1} \ell \mathcal{V}(X_i; Y_{i+1}) - v(0)\ell X_n \,. \tag{53}$$

In the constant velocity case $v(r) = c \in \mathbb{R}$, one simply has

$$\mathcal{R}[X;Y] = \sum_{i=0}^{n-1} \ell(Y_{i+1} - X_i) - v(0)\ell X_n \,,$$

i. e. minimising $\mathcal{R}[X;Y]$ means aiming at a shorter possible distance for a particle $X_i$, $i < n$, with respect to the $i+1$-th particle of a given reference distribution, whereas the $n$-th particle travels with speed $v(0)$. We therefore use the relative energy $\mathcal{R}$ to describe system (52) by a *variational splitting method*.

Let $\tau > 0$ be a given time step. For a given initial state $X_0 = (\bar{x}_0, \ldots, \bar{x}_n) \in \mathbb{R}^n$ and a given state $X_{k-1} = (x_{k-1,0}, \ldots, x_{k-1,n})$ at time $(k-1)\tau$, we define recursively $X_k = (x_{k,0}, \ldots, x_{k,n})$ as

$$X_k = \mathrm{argmin}\left\{ \frac{1}{2\tau} d(X, X_{k-1})^2 + \mathcal{R}[X; X_{k-1}], \ X \in \mathbb{R}^{n+1} \right\}, \tag{54}$$

where we are using the notation

$$d(X,Y)^2 = \sum_{i=0}^{n} \ell |X_i - Y_i|^2 \,.$$

The scheme (54) is defined in the spirit of a semi-implicit version of the JKO scheme which was used for the first time in [36] for a system of nonlocal interaction equations. This strategy allows to use a JKO type strategy for a system without a neat gradient flow structure, for which a non autonomous version of a gradient flow is possible for short times by 'freezing' some terms in the system. A simple computation yields the following Euler-Lagrange equations for the scheme (54)

$$\frac{X_{k,i} - X_{k-1,i}}{\tau} = v\left( \frac{\ell}{X_{k-1,i+1} - X_{k,i}} \right), \quad i = 0, \ldots, n-1 \,, \tag{55}$$

$$\frac{X_{k,n} - X_{k-1,n}}{\tau} = v(0) \tag{56}$$

for all $k \in \mathbb{N}$. It is immediately seen that (55)-(56) is a semi-implicity Euler discretisation of the system (52). By definition of minimum in (54), we get

$$\frac{1}{2\tau} d(X_k, X_{k-1})^2 \leq \mathcal{R}(X_{k-1}, X_{k-1}) - \mathcal{R}(X_k, X_{k-1})$$

$$= \ell \sum_{i=0}^{n-1} [\mathcal{V}(X_{k-1,i}; X_{k-1,i+1}) - \mathcal{V}(X_{k,i}; X_{k-1,i+1})] - \ell v(0)(X_{k-1,n} - X_{k,n})$$

$$= \ell \sum_{i=0}^{n-1} \int_{X_{k-1,i}}^{X_{k,i}} v\left( \frac{\ell}{X_{k-1,i+1} - z} \right) dz - \ell v(0)(X_{k-1,n} - X_{k,n})$$

$$\leq \frac{1}{4\tau} d(X_k, X_{k-1})^2 \, \tau \|v\|_{L^\infty} \,,$$

which yields, in the spirit of [2, Chapter 3],

$$d(X_k, X_h)^2 \leq 2(k-h)^2 \tau^2 \|v\|_{L^\infty}^2 \,,$$

and, given $T > 0$ and the piecewise interpolation

$$\overline{X}_\tau : [0, T] \mapsto L^2([0,1]), \qquad \overline{X}_\tau(t) = \sum_{i=0}^{n-1} X_{k,i} \mathbf{1}_{[i\ell, (i+1)\ell)}, \qquad \text{for } t \in [(k-1)\tau, k\tau), \quad k = 1, \ldots, [T/\tau],$$

we have

$$\|\overline{X}_\tau(t) - \overline{X}_\tau(s)\|_{L^2([0,1])} \leq \|v\|_{L^\infty}(|t-s| + o(1)), \qquad \text{as } \tau \searrow 0 \,,$$

which, by a suitable generalisation of Arzelá-Ascoli (see [2, Chapter 3]) implies that the family $\overline{X}_\tau$ has a strong limit (up to a subsequence) $\overline{X} \in C([0,T]; L^2([0,1]))$.

An important feature in the computations above is that the compactness estimates are independent of the number of particles $n$, which suggests the possibility that this result could be stretched to the $n \to +\infty$ case. However, the

$n \to +\infty$ is quite tricky here, since the formal $n \to +\infty$ behaviour of the scheme would be described by the formal equation

$$X_t = v \left( \frac{1}{X_z - \frac{\tau}{\ell} X_t} \right) ,$$

in which $\tau/\ell$ stands for the $n \to +\infty$ limit of this ratio. Hence, a tricky CFL condition may arise in the limit. We leave the identification of the $n \to +\infty$ limit of this semi-implicit scheme as an open problem for future work.

## 9. Coupling with diffusion terms and other results

As we said in the introduction, one of the first questions posed by the experts in Wasserstein gradient flows concerned with the possibility of considering a nonlinear convection term coupled with a diffusion operator in such a way to obtain a contraction property in some Wasserstein space, at least in one space dimension with the help of the pseudo-inverse approach recalled in section 3. The first result in this direction was proven in [21], and it is about an equation of the form

$$\rho_t = (\rho^m)_{xx} + (\rho^q)_x$$

under the assumption $q > m + 1$, which implies that the diffusion term is dominant in the large time asymptotics, and therefore the convection term only plays a role of a small perturbation (for large times) of the diffusion operator. The first Lipschitz continuity property in a Wasserstein space of the semigroup generated by a convection-diffusion operator in a regime in which both effects are relevant in the time-asymptotics is that of [37] on the viscous Burgers' equation

$$\rho_t = \rho_{xx} + \rho \rho_x ,$$

in which a quasi contraction estimate is obtained by means of the Hopf-Cole transformation. Hinted by the results in [17], in [18] the authors provide a contraction result in the $\infty$-Wasserstein distance for a large class of convection-diffusion equations using a splitting method. We also mention the result in [41], which found an algebraic rate of convergence toward $N$-waves by means of the same time dependent scaling strategy considered in [27].

This subject was not considered for some years, until it was partly revived in [38]. In that paper the authors consider a special class of convection diffusion equation of the form

$$\rho_t = (\rho(a(x)\rho^{m-1})_x)_x, \tag{57}$$

which is the formal gradient flow of the functional

$$\mathbb{F}[\rho] = \frac{1}{m} \int a(x)\rho^m \, dx .$$

The main result of [38] proves that entropy solutions to (57) can be obtained as solutions to the JKO scheme of the functional $\mathbb{F}$. This results links for the first time the notion of entropy solution in the sense of Kruzkov with the notion of gradient flow in the Wasserstein sense.

We finally mention that the use of the pseudo-inverse equation to study the well-posedness (in a measure sense) of a heterogeneous conservation law arising in the context of Bose-Einstein condensates was performed recently in [19].

## Acknowledgments

## Appendix A. Proof of Theorem 3.1

It is well known (cf. e.g. [33, Chapter 11]) that the primitive variable

$$F(t,x) = \int_{-\infty}^{x} \rho(t,y) dy$$

of $\rho$ entropy solution to (4) is a viscosity solution to the Hamilton-Jacobi equation

$$F_t + f(F_x) = 0. \tag{58}$$

Now, standard results on the SBV regularity of scalar conservation laws with genuinely nonlinear flux, cf. [1, 5], imply that $\rho$ has jumps on countably many Lipschitz curves of the form $\gamma_i : [0, +\infty) \to \mathbb{R}$, $i \in \mathbb{N}$, and features SBV regularity elsewhere, i.e. $\rho_x$ has no Cantor part between two consecutive curves $\gamma_i$ and $\gamma_{i+1}$. On each region
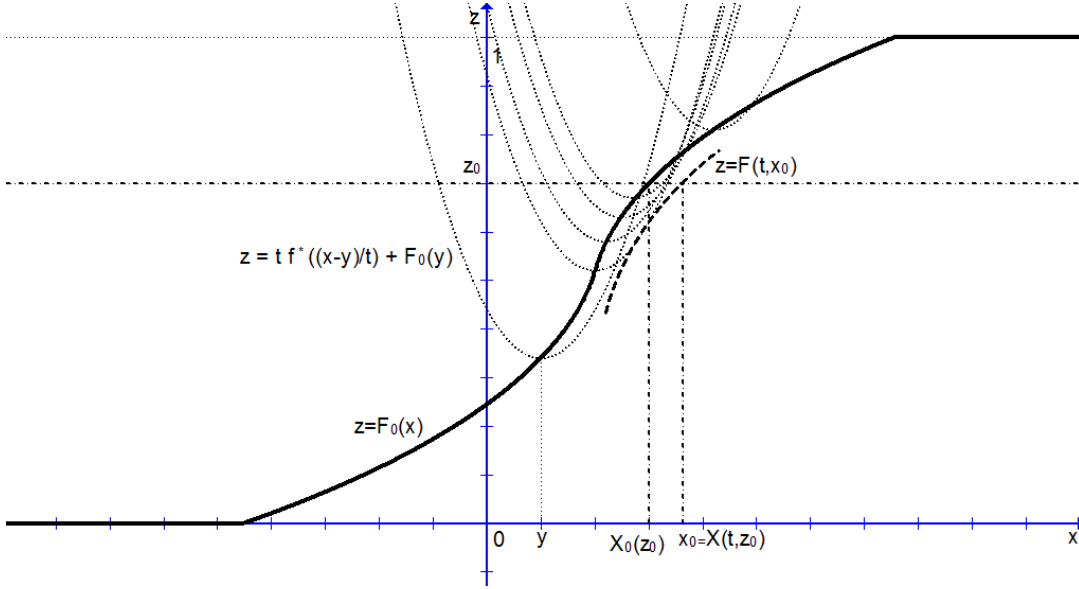
FIGURE 1. The idea behind the proof of (59). The bold curve is the initial condition $F_0$ on the $(x, z)$ plane. The solution $F = F(t, x)$ to (35), computed via (36), is the convex envelop of a set of 'parabolas'. It is plotted 'locally' near a point $x_0 \in \mathbb{R}$. For a given $z_0 \in [0, 1]$, the value of $X(t, z_0)$ is clearly seen as the *supremum* over a set of 'inverse parabolas' centered at $(y, F_0(y))$, with $y \in \mathbb{R}$. Only the parabolas with $F_0(y) \leq z$ are affected.

$I_i$ between $\gamma_i$ and $\gamma_{i+1}$, the primitive variable is $W^{1,\infty}$ with respect to $x$, the Hamilton-Jacobi equation (58) is satisfied almost everywhere, and $F_t$ is well defined almost everywhere. This implies that the pseudo-inverse

$$X(t, z) = \inf\{x \in \mathbb{R} \ : \ F(t, x) > z\}$$

is piecewise smooth with respect to $z$ for a fixed $t$ on each interval $z \in J_i(t) = (F(t, \gamma_i(t)), F(t, \gamma_{i+1}(t)))$, with (at most) countably many increasing jumps on $J_i$ due to the possible vacuum regions of $\rho(t, \cdot)$ in $I_i$. Summing up, the set $[0, +\infty) \times [0, L]$ (with $L$ being the total mass of $\rho$) can be split in a sequence $\{K_j\}_j$ of pair-wise disjoint sets of the form $K_j = \bigcup_{t \geq 0} H_j(t)$, with $H_j(t)$ being a subinterval of $[0, L]$ moving continuously in time. On the interior of each set $K_j$, $\rho$ is strictly positive, which implies $F$ is strictly increasing with respect to $x$, and therefore the pseudo-inverse $X(t, z)$ satisfies

$$X_t = v(X_z^{-1})$$

as explained in subsection 3.1. The proof is complete, since the set on which the above equation is possibly not satisfied has zero Lebesgue measure.

## APPENDIX B. A DUAL EXPRESSION OF LAX-HOPF FORMULA

We provide here a sketch of the proof of the formula (37), and refer to [17, Lemma 2.3] for the details.

The first step in the proof is

$$X(t, z) = \sup_{\{y \in \mathbb{R} \,:\, F_0(y) \leq z\}} \left\{ x \ : \ tf^* \left( \frac{x - y}{t} \right) + F_0(y) = z \right\}. \tag{59}$$

We outline the main ideas behind (59) in figure 1. The graph of $z = F(t, x)$ in the $(x, z)$ plane is obtained as an infimum of a set of parabolas centered at $(y, F_0(y))$ with $y$ ranging on the real line. The width of the parabolas gets larger with time. Since $F_0$ is increasing, only the positive branches of the parabolas contribute to the value of $F(t, x)$. Now, if we fix a $z_0 \in [0, 1]$, the value of $X(t, z_0)$ is obtained as the *supremum* as in the formula (59).

Once (59) is proven, we can invert the formula $tf^* \left( \frac{x-y}{t} \right) + F_0(y) = z$ with respect to $x$ and get

$$x = y + tg \left( \frac{z - F_0(y)}{t} \right), \tag{60}$$

due to the fact that $f^*$ is invertible on $[0, +\infty)$. Hence, for a given $z \in [0, 1]$, the value $X(t, z)$ is determined by taking the supremum of all $x$ in (60), where only the $y$ such that $F_0(y) \leq z$ are involved. This is still easily seen from figure 1.

## References

[1] L. Ambrosio and C. De Lellis. A note on admissible solutions of 1d scalar conservation laws and 2d hamilton-jacobi equations. *J. Hyperbolic Diff. Equ.*, 1(4):813–826, 2004.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.

[3] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26(1-2):43–100, 2001.

[4] D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[5] S. Bianchini. SBV regularity of genuinely nonlinear hyperbolic systems of conservation laws in one space dimension. *Acta Mathematica Scientia*, 32:380–388, 2012.

[6] A. Blanchet, V. Calvez, and J. A. Carrillo. Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model. *SIAM J. Numer. Anal.*, 46(2):691–721, 2008.

[7] F. Bolley, Y. Brenier, and G. Loeper. Contractive metrics for scalar conservation laws. *J. Hyperbolic Differ. Eq.*, 2:91–107, 2005.

[8] G. A. Bonaschi, J. A. Carrillo, M. Di Francesco, and M. A. Peletier. Equivalence of gradient flows and entropy solutions for singular nonlocal interaction equations in 1D. *ESAIM Control Optim. Calc. Var.*, 21(2):414–441, 2015.

[9] Y. Brenier. Extended Monge-Kantorovich theory. In *Optimal transportation and applications (Martina Franca, 2001)*, volume 1813 of *Lecture Notes in Math.*, pages 91–121. Springer, Berlin, 2003.

[10] Y. Brenier. $L^2$ formulation of multidimensional scalar conservation laws. *Arch. Ration. Mech. Anal.*, 193(1):1–19, 2009.

[11] Alberto Bressan. *Hyperbolic systems of conservation laws*, volume 20 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2000. The one-dimensional Cauchy problem.

[12] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[13] M. Burger, M. Di Francesco, and Y. Dolak-Struss. The Keller-Segel model for chemotaxis with prevention of overcrowding: linear vs. nonlinear diffusion. *SIAM J. Math. Anal.*, 38(4):1288–1315 (electronic), 2006.

[14] P. Cannarsa and C. Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*. Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston, Inc., Boston, MA, 2004.

[15] P. Cardaliaguet, G. Carlier, and B. Nazaret. Geodesics for a class of distances in the space of probability measures. *Calc. Var. Partial Differential Equations*, 48(3-4):395–420, 2013.

[16] J. A. Carrillo, Y.-P. Choi, and M. Hauray. The derivation of swarming models: mean-field limit and Wasserstein distances. In *Collective dynamics from bacteria to crowds*, volume 553 of *CISM Courses and Lectures*, pages 1–46. Springer, Vienna, 2014.

[17] J. A. Carrillo, M. Di Francesco, and C. Lattanzio. Contractivity of Wasserstein metrics and asymptotic profiles for scalar conservation laws. *J. Differential Equations*, 231(2):425–458, 2006.

[18] J. A. Carrillo, M. Di Francesco, and C. Lattanzio. Contractivity and asymptotics in Wasserstein metrics for viscous nonlinear scalar conservation laws. *Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat. (8)*, 10(2):277–292, 2007.

[19] J. A. Carrillo, M. Di Francesco, and G. Toscani. Condensation phenomena in nonlinear drift equations. *Ann. Sc. Normale Super. Pisa Cl. Sci.* To appear.

[20] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev. Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.*, 156(2):229–271, 2011.

[21] J. A. Carrillo and K. Fellner. Long-time asymptotics via entropy methods for diffusion dominated equations. *Asymptot. Anal.*, 42(1-2):29–54, 2005.

[22] J. A. Carrillo, M. P. Gualdani, and G. Toscani. Finite speed of propagation in porous media by mass transportation methods. *Comptes Rendus Mathematique*, 338:815–818, 2004.

[23] J. A. Carrillo, A. Jüngel, P. A. Markowich, G. Toscani, and A. Unterreiter. Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. *Monatsh. Math.*, 133(1):1–82, 2001.

[24] J. A. Carrillo, S. Lisini, G. Savaré, and D. Slepčev. Nonlinear mobility continuity equations and generalized displacement convexity. *J. Funct. Anal.*, 258(4):1273–1309, 2010.

[25] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.

[26] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[27] J. A. Carrillo and G. Toscani. Asymptotic $L^1$-decay of solutions of the porous medium equation to self-similarity. *Indiana Univ. Math. J.*, 49(1):113–142, 2000.

[28] J. A. Carrillo and G. Toscani. Wasserstein metric and large-time asymptotics of nonlinear diffusion equations. In *New trends in mathematical physics*, pages 234–244. World Sci. Publ., Hackensack, NJ, 2004.

[29] G.-Q. Chen and M. Rascle. Initial layers and uniqueness of weak entropy solutions to hyperbolic conservation laws. *Arch. Ration. Mech. Anal.*, 153(3):205–220, 2000.

[30] I. Corwin, P. L. Ferrari, and S. Péché. Limit processes for TASEP with shocks and rarefaction fans. *J. Stat. Phys.*, 140(2):232–267, 2010.

[31] M. G. Crandall and T. M. Liggett. Generation of semi-groups of nonlinear transformations on general Banach spaces. *Amer. J. Math.*, 93:265–298, 1971.

[32] C. M. Dafermos. Polygonal approximations of solutions of the initial value problem for a conservation law. *J. Math. Anal. Appl.*, 38:33–41, 1972.

[33] C. M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2000.

[34] E. De Giorgi. New problems on minimizing movements. In *Boundary value problems for partial differential equations and applications*, volume 29 of *RMA Res. Notes Appl. Math.*, pages 81–98. Masson, Paris, 1993.

[35] M. Del Pino and J. Dolbeault. Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions. *J. Math. Pures Appl. (9)*, 81(9):847–875, 2002.

[36] M. Di Francesco and S. Fagioli. Measure solutions for non-local interaction PDEs with two species. *Nonlinearity*, 26(10):2777–2808, 2013.

[37] M. Di Francesco and P. A. Markowich. Entropy dissipation and Wasserstein metric methods for the viscous Burgers' equation: convergence to diffusive waves. In *Partial differential equations and inverse problems*, volume 362 of *Contemp. Math.*, pages 145–165. Amer. Math. Soc., Providence, RI, 2004.

[38] M. Di Francesco and D. Matthes. Curves of steepest descent are entropy solutions for a class of degenerate convection-diffusion equations. *Calc. Var. Partial Differential Equations*, 50(1-2):199–230, 2014.

[39] M. Di Francesco and M.D. Rosini. Rigorous derivation of nonlinear scalar conservation laws from follow-the-leader type models via many particle limit. *Archive for Rational Mechanics and Analysis*, 217(3):831–871, 2015.

[40] R. J. DiPerna. Decay and asymptotic behavior of solutions to nonlinear hyperbolic systems of conservation laws. *Indiana Univ. Math. J.*, 24(11):1047–1071, 1974/75.

[41] J. Dolbeault and M. Escobedo. $L^1$ and $L^\infty$ intermediate asymptotics for scalar conservation laws. *Asymptot. Anal.*, 41(3-4):189–213, 2005.

[42] J. Dolbeault, B. Nazaret, and G. Savaré. A new class of transport distances between measures. *Calc. Var. Partial Differential Equations*, 34(2):193–231, 2009.

[43] N. Gigli and F. Otto. Algebraic contraction rate for distance between entropy solutions of scalar conservation laws. journal of mathematical analysis and applications. Submitted.

[44] L. Gosse and G. Toscani. Identification of asymptotic decay to self-similarity for one-dimensional filtration equations. *SIAM J. Numer. Anal.*, 43(6):2590–2606 (electronic), 2006.

[45] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[46] S. N. Kružkov. First order quasilinear equations with several independent variables. *Mat. Sb. (N.S.)*, 81 (123):228–255, 1970.

[47] P. D. Lax. Hyperbolic systems of conservation laws. II. *Comm. Pure Appl. Math.*, 10:537–566, 1957.

[48] T.-P. Liu and M. Pierre. Source-solutions and asymptotic behavior in conservation laws. *J. Differential Equations*, 51(3):419–441, 1984.

[49] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[50] Olga Arsenievna Oleinik. Discontinuous solutions of non-linear differential equations. *Uspekhi Matematicheskikh Nauk*, 12(3):3–73, 1957.

[51] F. Otto. Evolution of microstructure in unstable porous media flow: a relaxational approach. *Comm. Pure Appl. Math.*, 52(7):873–915, 1999.

[52] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[53] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[54] M. Prähofer and H. Spohn. Current fluctuations for the totally asymmetric simple exclusion process. In *In and out of equilibrium (Mambucaba, 2000)*, volume 51 of *Progr. Probab.*, pages 185–204. Birkhäuser Boston, Boston, MA, 2002.

[55] G. Toscani. Kinetic approach to the asymptotic behaviour of the solution to diffusion equations. *Rend. Mat. Appl. (7)*, 16(2):329–346, 1996.

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

[57] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.