Università degli Studi di Pavia
Dipartimento di Matematica "Felice Casorati"
Dottorato in Matematica e Statistica

Applications of Optimal Transport to Evolution Problems: Sticky Particles System and Fokker Planck Equations

Luca Natile

Advisor
Prof. Giuseppe Savaré

Contents

Chapter 1

Introduction

In this PhD Thesis we deal with some aspects of the applications of the theory of optimal mass transportation (and related Wasserstein distance) to evolution partial differential equations. Essentially, the Thesis is made up of two parts.

In the first part (Chapter 3), we present a simple approach to study the one–dimensional pressureless Euler system via adhesion dynamics in the Wasserstein space $\mathcal{P}_2(\mathbb{R})$. Starting from a discrete system of a finite number of "sticky" particles, we obtain new explicit estimates of the solution in terms of the initial mass and momentum and we are able to construct an evolution semigroup in a measure-theoretic phase space, allowing mass distributions in $\mathcal{P}_2(\mathbb{R})$ and corresponding L^2-velocity fields. We investigate various interesting properties of this semigroup, in particular its link with the gradient flow of the (opposite) squared Wasserstein distance.

In the second part (Chapter 4 and Chapter 5), we focus our attention on the Fokker-Planck equation in $\mathbb{R}^d$, when the drift is a monotone (or λ-monotone) operator. In Chapter 4, we prove new contraction properties of general transportation costs along nonnegative measure-valued solutions of this equation. Chapter 5 is devoted to find such solutions among the absolutely continuous curves in $\mathcal{P}_2(\mathbb{R}^d)$: our main result ensures the existence of these solutions and provides a metric characterization.

Let us now explain in greater detail the main results, referring to the next sections for more precise definitions, statements and proofs.

Sticky particles model in the one-dimensional case.

In recent years considerable attention has been devoted to the unidimensional pressureless Euler system

$$\begin{cases} \partial_t \rho + \partial_x(\rho\, v) = 0 \\ \partial_t(\rho\, v) + \partial_x(\rho\, v^2) = 0 \end{cases} \quad \text{in } \mathbb{R}\times(0,+\infty), \quad \rho|_{t=0} = \rho_0, \quad v|_{t=0} = v_0 \quad (1.1)$$

in connection with the Zel′dovich model [54] for the evolution of a sticky particle system (SPS) via adhesion dynamics. This model describes the behavior of a finite collection of particles, freely moving in the absence of forces and sticking under collision. They can be mathematically represented by a time-dependent discrete measure $\rho_t^N := \sum_{i=1}^n m_i \delta_{x_i(t)}$ concentrated in a finite set of N particles

$P_i(t) := (m_i, x_i(t), v_i(t))$, $i = 1, \dots, N$ with positive mass m_i, ordered positions $x_1(t) \leq x_2(t) \leq \dots \leq x_{N-1}(t) \leq x_N(t)$, and velocities $v_i(t)$.

Denoting by $J_i(t) := \{j : x_j(t) = x_i(t)\}$ the collection of (the indices of) the particles $P_j(t)$ coinciding with $P_i(t)$ at time t, the adhesion dynamic imposes that the sets $J_i(t)$ are nondecreasing in time, so that $v_j(t+) = v_i(t+)$ for every $j \in J_i(t)$. We can thus order in a finite and monotone sequence $0 < t_1 < t_2 < \cdots$ the collection of times when the cardinality of some $J_i(t)$ has a discontinuity (corresponding to some collision). In each open interval $[t_k, t_{k+1})$ the (right-continuous) velocities $v_i(t) = \dot{x}_i(t)$ are thus supposed to be constant and, at each collision time t_k, the conservation of mass and momentum yields the updated equation for the velocities

$$v_i(t_k+) = \frac{\sum_{j \in J_i(t_k)} m_j v_j(t_k-)}{\sum_{j \in J_i(t_k)} m_j}, \qquad i = 1, \dots N. \tag{1.2}$$

It is not difficult to check that the measures ρ^N and $(\rho v)^N_t := \sum_{i=1}^N m_i v_i(t)\, \delta_{x_i(t)}$ solve (1.1). Starting from the discrete SPS, the existence of measure-valued solutions to (1.1) with general initial data and satisfying suitable entropy conditions [12] has been proved by Grenier [29] and E, Rykov, and Sinai [26] (see also the contribution of Martin and Piasecki [32]) as limits (in the sense of weak convergence of measures) of the discrete particle evolutions ρ^N_t as $N \uparrow +\infty$. Here we also cite the different approaches of Bouchut and James [13], Poupaud and Rascle [41], and Sever [47] in the multidimensional case; viscous regularizations of (1.1) have been studied by Sobolevskiĭ [49] and Boudin [14], and a different model, starting from particles of finite size, has been considered by Wolansky [53].

The convergence result has been extended further and refined by Brenier and Grenier [18], Huang and Wang [30], and Nguyen and Tudorascu [37]. (By a different probabilistic approach, Moutsinga [34] has recently been able to consider initial velocities with nonpositive jumps at each point of the support of ρ_0.) The basic assumption is that the discrete initial velocity v_i is the value in x_i of a given continuous function v with at most quadratic growth and (the total mass being normalized to 1) the sequence ρ^N_0 converges to ρ_0 with respect to the L^2-Wasserstein distance in the space $\mathcal{P}_2(\mathbb{R})$ of probability measures with finite quadratic moment. This includes the case (considered in [18]) of a sequence ρ^N_0 with uniformly bounded support and weakly converging to ρ_0 in the duality with continuous real functions.

All these results depend on a remarkable characterization of the solution ρ found by Brenier and Grenier [18]: by introducing the cumulative distribution function M_ρ associated to a probability measure $\rho \in \mathcal{P}(\mathbb{R})$

$$M_\rho(x) := \rho((-\infty, x]) \quad \forall x \in \mathbb{R}, \quad \text{so that } \rho = \partial_x M_\rho \quad \text{in } \mathscr{D}'(\mathbb{R}), \tag{1.3}$$

they prove that the function $M(t, \cdot) := M_{\rho_t}(\cdot)$ is the unique entropy solution of the scalar conservation law

$$\partial_t M + \partial_x A(M) = 0 \quad \text{in } \mathbb{R} \times (0, +\infty), \tag{1.4}$$

where $A : [0, 1] \to \mathbb{R}$ is a continuous flux function depending only on ρ_0 and v_0 (see Theorem 3.28 for a precise statement).

It can also be shown [37] that this solution satisfies the Oleinik entropy condition

$$v_t(x_2) - v_t(x_1) \le \frac{1}{t}(x_2 - x_1) \quad \text{for } \rho_t\text{-a.e. } x_1, x_2 \in \mathbb{R},\ x_1 \le x_2.$$

In **Chapter 3** we discuss various refinements of the Brenier–Grenier result by a different approach. Our starting point (**Theorem 3.2**) is an explicit Lipschitz estimate of the dependence of ρ_t w.r.t. the initial data $\rho_0, (\rho v)_0$ in the so called L^p-Wasserstein distance W_p for every $p \ge 1$. This distance is defined by

$$W_p^p(\rho^1, \rho^2) := \min\Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \big|x_1 - x_2\big|^p \,\mathrm{d}\boldsymbol{\rho}(x_1, x_2) : \boldsymbol{\rho} \in \mathcal{P}(\mathbb{R} \times \mathbb{R}) \\ \boldsymbol{\rho} \text{ is a coupling between } \rho^1 \text{ and } \rho^2 \Big\}$$

in terms of couplings, i.e. measures $\boldsymbol{\rho}$ on the product space $\mathbb{R}\times\mathbb{R}$ whose marginals are ρ^1 and ρ^2 respectively, so that $\boldsymbol{\rho}(E \times \mathbb{R}) = \rho^1(E)$ and $\boldsymbol{\rho}(\mathbb{R} \times E) = \rho^2(E)$ for every Borel subset $E \subset \mathbb{R}$. For $p = 2$ it shows that $(\rho_t^N)_{N\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{P}_2(\mathbb{R})$ and, in particular, yields the convergence results of [26, 18, 37], allowing general initial measures in $\mathcal{P}_2(\mathbb{R})$ and (possibly discontinuous) velocity fields $v_0 \in L^2(\rho_0)$. We also show that a suitable L^2-like integral distance between the momentum ρv of two solutions can be controlled in terms of the initial data, and we prove further precise representation properties of the solution and its velocity field (**Theorem 3.3**).

This leads to the construction of a semigroup $\mathscr{S}_t$ associated to the evolution of the SPS, which exhibits interesting links with another semigroup (recently studied by Ambrosio, Gigli, and Savaré [2]), obtained as the gradient flow in $\mathcal{P}_2(\mathbb{R})$ of the (opposite) squared Wasserstein distance from a fixed reference measure.

This link (which at first glance may be unexpected) can be better understood in the simpler case when the initial velocity field v satisfies a one-sided monotonicity condition (see section 5.4.2 of Villani's book [51] for more details). Still, considering the simpler discrete case, if

$$-\delta^{-1} := \min_{x_i \neq x_j} \frac{v(x_i) - v(x_j)}{x_i - x_j} < 0, \quad v(x_i) := v_i \tag{1.5}$$

for $t \in [0, \delta)$, then the map $\mathsf{x}_0^t(x) := x + tv(x)$ is nondecreasing on the support of ρ_0 (the finite set $\{x_i : i = 1, \dots N\}$), so that the first collision occurs at $t := \delta$ and, in the interval $[0, \delta)$, one has the freely moving measures

$$\rho_t := (\mathsf{x}_0^t)_\# \rho_0 = \sum_{i=1}^N m_i \delta_{x_i + tv_i}, \quad (\rho v)_t = \sum_{i=1}^N m_i v_i \delta_{x+tv_i},\ t \in [0, \delta), \tag{1.6}$$

solving the pressureless Euler system (1.1). On the other hand, the curve $t \mapsto \rho_t$, $t \in [0, \delta]$ is a constant speed minimal geodesic in $\mathcal{P}_2(\mathbb{R})$ connecting ρ_0 with $\eta := \rho_\delta$; as in any Riemannian manifold, it coincides (up to a suitable rescaling; see [2, Theorem 11.2.10]) with the gradient flow in $\mathcal{P}_2(\mathbb{R})$ of the functional $\phi^{\rho_0}(\rho) := -\frac{1}{2} W_2^2(\rho, \rho^0)$. After the collision at time $t = \delta$ the trajectory of the gradient flow no longer coincides with the free motion (1.6), since its velocity has a jump which can be described exactly by (1.2) [2, Theorem 10.4.12]. At a later time, the velocity field induced by the (rescaled) Wasserstein gradient flow can be characterized by the formula

$$v_i(t+) = t^{-1}\left(x_i(t) - \frac{\sum_{j\in J_i(t)} m_j x_j(0)}{\sum_{j\in J_i(t)} m_j}\right), \qquad i = 1,\dots N. \tag{1.7}$$

There is an interesting property (stated in **Theorems 3.4** and **3.5**) that the two different laws (1.2) and (1.7) give rise to the same evolution, even for arbitrary initial data.

In order to obtain these results, we adopt the viewpoint of one-dimensional optimal transportation and represent each probability measure $\rho \in \mathcal{P}_2(\mathbb{R})$ by its monotone rearrangement X_ρ, which is the pseudoinverse of the distribution function M_ρ of (1.3):

$$X_\rho(w) := \inf\{x : M_\rho(x) > w\} = \inf\{x : \rho((-\infty, x]) > w\}, \quad w \in (0,1).$$

(A similar approach, in a probabilistic framework, was used in [34]; see also [28] for other applications.) The map $\rho \mapsto X_\rho$ is an isometry between $\mathcal{P}_2(\mathbb{R})$ (endowed with the L^2-Wasserstein distance) and the convex cone $\mathcal{K}$ of nondecreasing functions in the Hilbert space $L^2(0,1)$ (see §2.3.3). Through this isometry, any gradient flow with respect to W_2 in $\mathcal{P}_2(\mathbb{R})$ can be rephrased as a gradient flow in $\mathcal{K}$ with respect to the $L^2(0,1)$-distance and one can use the powerful tools of the classical theory of variational evolution inequalities in Hilbert spaces (we refer the reader to §2.1). It turns out (see **Theorem 3.6**) that in this Lagrangian formulation the solution X_{ρ_t} admits three simple characterizations—in terms of the $L^2(0,1)$-projection $\mathsf{P}_\mathcal{K}$ onto $\mathcal{K}$:

$$X_{\rho_t} = \mathsf{P}_\mathcal{K}(X_{\rho_0} + tV_0), \quad V_0 = v_0 \circ X_{\rho_0}, \tag{1.8}$$

and in terms of the differential inclusions:

$$\frac{\mathrm{d}}{\mathrm{d}t}X_{\rho_t} + \partial I_\mathcal{K}(X_{\rho_t}) \ni V_0, \qquad t\frac{\mathrm{d}}{\mathrm{d}t}X_{\rho_t} + \partial I_\mathcal{K}(X_{\rho_t}) \ni X_{\rho_t} - X_{\rho_0}, \tag{1.9}$$

where $I_\mathcal{K}$ is the indicator function associated to $\mathcal{K}$

$$I_\mathcal{K}(X) = \begin{cases} 0 & \text{if } X \in \mathcal{K}, \\ +\infty & \text{otherwise.} \end{cases}$$

Equations (1.8) and (1.9) encode all the qualitative information on the measure-valued solution ρ_t, and their proof in the case of the discrete SPS constitutes the core of our argument. The proof relies on an elementary but careful description of the $L^2(0,1)$-projection operator $\mathsf{P}_\mathcal{K}$ and on the subdifferential of $I_\mathcal{K}$, which we discuss in section 3.2. Once ρ_t has been determined, its velocity $v_t \in L^2_{\rho_t}(\mathbb{R})$ can be recovered from the right derivative $V(t) := \frac{\mathrm{d}^+}{\mathrm{d}t}X_{\rho_t} \in L^2(0,1)$. In fact, as a byproduct of the second differential inclusion of (1.9), $V(t)$ is a function of $X(t)$ and, therefore, one obtains

$$V(t) = v_t \circ X_{\rho_t}. \tag{1.10}$$

The projection formula (1.8) (which was introduced by Shnirel′man [48] (see also [5]) in a slightly different form; see Remark 3.9) lies more or less explicitly at the core of the formulations by [26] and [18]. As was nicely explained by Andrievsky, Gurbatov, and Sobolevsky [5], elaborating on the contribution of [48], (1.8) is equivalent to the *generalized variational principle* of [26], which can be expressed through the convex envelope of the primitive function of the map $X_{\rho_0} + tV_0$. As stated in full generality by **Theorem 3.10**, this convexification characterizes the L^2-projection on $\mathcal{K}$. On the other hand, a convexification is also involved in the second Hopf formula for the solutions of the Hamilton–Jacobi equation associated to (1.4), as has already been observed in [18, sect. 4]. We will detail this point in **Theorem 3.28**.

The link between the formulation based on the scalar conservation law (1.4) and the Hilbertian theory of gradient flows such as (1.9) is not at all surprising after the illuminating paper by Brenier [17] (whose ideas, in particular concerning the SPS, could be traced back to [15, 16]). Wasserstein contraction properties of solutions of one-dimensional scalar conservation laws have also been recently obtained by Bolley, Brenier, and Loeper [8] (see also the further contribution by Carrillo, Di Francesco, and Lattanzio [20]). So it would be possible, in principle, to approach the SPS starting from (1.4) and try to apply the techniques developed there. Note, however, that two solutions originating from different initial distributions of position and velocity give rise to two scalar conservation laws differing not only by the initial data but also by the flux functions, so that their comparison does not look immediate. Moreover, the present self-contained approach is very simple, since it relies on elementary tools of convex analysis and direct computations on the discrete case; the simultaneous characterization of the evolution by (1.8) and (1.9) provides a more refined description of the solution and, as a byproduct, a new direct proof of the Brenier–Grenier theorem.

Fokker-Planck equation with a monotone drift term.

The aim of this second part is to obtain new uniqueness, contractivity and existence results for nonnegative measure-valued solutions to the Fokker-Planck equation

$$\partial_t \rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho_{|t=0} = \rho_0, \tag{1.11}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel λ-monotone operator, $\lambda \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y\rangle \geq \lambda \left|x - y\right|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

Here we consider a weakly continuous family of probability measures $(\rho_t)_{t\geq 0} \subset \mathcal{P}(\mathbb{R}^d)$ satisfying the equation (1.11) in the sense of distributions

$$\int_0^{+\infty}\int_{\mathbb{R}^d} \Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\,\mathrm{d}\rho_t\mathrm{d}t = 0 \quad \forall\,\zeta \in C^\infty_{\mathrm{c}}(\mathbb{R}^d \times (0,+\infty)), \tag{1.12}$$

with the initial datum ρ_0.

Equations of this type are the subject of several papers by Bogachev, Da Prato, Krylov, Röckner, and Stannat, who consider a very general situation where the Laplacian is replaced by a second order elliptic operator with variable coefficients and B is locally bounded. Existence of solutions has been proved by [10, Cor. 3.3], uniqueness has been considered in [9] under general growth-coercivity conditions on B, and regularity has been investigated by [11]: in particular, it has been shown that ρ_t is absolutely continuous with respect to the Lebesgue measure for $\mathscr{L}^1$-a.e. t. When B is Lipschitz continuous, uniqueness can be obtained by standard duality arguments, see e.g. [3, Sec. 3].

The Wasserstein approach in the gradient case. When B is the gradient of a λ-convex function $V : \mathbb{R}^d \to \mathbb{R}$ then (1.11) can be considered as the *gradient flow* of the perturbed entropy functional

$$\mathscr{H}_V(\rho) := \int_{\mathbb{R}^d} u(x) \log u(x) \,\mathrm{d}x + \int_{\mathbb{R}^d} V(x) \,\mathrm{d}\rho(x) \quad \rho = u\mathscr{L}^d \tag{1.13}$$

in the space $\mathcal{P}_2(\mathbb{R}^d)$ of probability measures with finite quadratic moments endowed with the L^2-Kantorovich-Rubinstein-Wasserstein distance $W_2(\cdot,\cdot)$. This distance can be defined by

$$W_2^2(\rho^1,\rho^2) := \min\Big\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} \big|x_1 - x_2\big|^2 \,\mathrm{d}\boldsymbol{\rho}(x_1,x_2) : \boldsymbol{\rho} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \\ \boldsymbol{\rho} \text{ is a coupling between } \rho^1 \text{ and } \rho^2\Big\}$$

in terms of couplings, i.e. measures $\boldsymbol{\rho}$ on the product space $\mathbb{R}^d \times \mathbb{R}^d$ whose marginals are ρ^1 and ρ^2 respectively, so that $\boldsymbol{\rho}(E\times\mathbb{R}^d) = \rho^1(E)$ and $\boldsymbol{\rho}(\mathbb{R}^d\times E) = \rho^2(E)$ for every Borel subset $E \subset \mathbb{R}^d$.

This remarkable interpretation found in [31] gave rise to a series of studies on the relationships between certain classes of diffusion equations and distances between probability measures induced by optimal transport problems (see e.g. the general overviews of [51, 2, 52]). One of the strengths of this approach is a new geometric insight (developed in [38]) in the evolution process: in the case of (1.11) the λ-convexity of the potential V reflects a λ-convexity property (also called *displacement convexity*) of the functional $\mathscr{H}_V$ along the geodesics of $\mathcal{P}_2(\mathbb{R}^d)$.

Contraction estimate in the gradient case. This gradient flow interpretation has been carried out (see e.g. [2]) and, among the most interesting estimates, it provides the λ-contraction property

$$W_2(\rho_t^1,\rho_t^2) \le \mathrm{e}^{-\lambda t}\, W_2(\rho_0^1,\rho_0^2) \qquad \text{for every } t \ge 0, \tag{1.14}$$

where ρ_t^i, $i = 1, 2$, are the solutions to (1.11) starting from the initial data $\rho_0^i \in \mathcal{P}_2(\mathbb{R}^d)$. In order to prove (1.14) when $B = \nabla V$, essentially two basic strategies have been proposed:

1. A first approach, developed by [22] for smooth evolutions and by [2] in a measure-theoretic setting, starts from equation (1.11) written in the form

$$\partial_t\rho + \nabla\cdot(\rho\mathbf{v}) = 0, \quad \mathbf{v} = -\Big(\frac{\nabla u}{u} + \nabla V\Big), \quad \rho = u\,\mathscr{L}^d, \tag{1.15}$$

   and it is based on two ingredients: the first one is the formula which evaluates the derivative of the squared Wasserstein distance from a fixed measure σ along the (absolutely continuous) curve ρ in $\mathcal{P}_2(\mathbb{R}^d)$

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t,\sigma) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \mathbf{v}_t(x), y-x\rangle \,\mathrm{d}\boldsymbol{\rho}_t(x,y) \quad \text{for } \mathscr{L}^1\text{-a.e. } t > 0 \tag{1.16}$$

   where $\boldsymbol{\rho}_t$ is an optimal coupling between ρ_t and σ.

The second ingredient is the "subgradient" property of the vector field $\mathbf{v}_t$ given by (1.15), related to the displacement convexity of $\mathscr{H}_V$: in the case $\lambda = 0$ it reads as

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \mathbf{v}_t(x), y-x\rangle \,\mathrm{d}\boldsymbol{\rho}_t(x,y) \le \mathscr{H}_V(\sigma)-\mathscr{H}_V(\rho_t) \quad \text{if } \mathbf{v}_t = -\Big(\frac{\nabla u_t}{u_t}+\nabla V\Big). \tag{1.17}$$

Combination of (1.16) and (1.17) yields the so called Evolution Variational Inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t,\sigma) \le \mathscr{H}_V(\sigma) - \mathscr{H}_V(\rho_t) \quad \text{for every } \sigma \in \mathcal{P}_2(\mathbb{R}^d) \tag{1.18}$$

which easily yields (1.14) for $\lambda = 0$ by a variable-doubling argument (see [2, Theorem 11.1.4]).

The main technical point here is that (1.17) requires $\mathbf{v}_t \in L^2(\rho_t)$ and (1.16) holds if for every $0 < t_0 < t_1 < +\infty$

$$\int_{t_0}^{t_1}\int_{\mathbb{R}^d} |\mathbf{v}_t|^2 \,\mathrm{d}\rho_t\,\mathrm{d}t = \int_{t_0}^{t_1}\int_{\mathbb{R}^d} \Big|\frac{\nabla u_t}{u_t} + \nabla V\Big|^2 \,\mathrm{d}\rho_t\,\mathrm{d}t < +\infty,$$

which should be imposed (in a suitable distributional sense) as an *a priori* regularity assumption on the solution of (1.11). We do not know if solutions to (1.12) exhibit a similar regularization effect. A second, even more difficult point prevents a simple extension of (1.18) to the general non-gradient case: it is the lack of a potential V and therefore of an entropy-like functional $\mathscr{H}_V$ satisfying an inequality similar to (1.17).

2. A second approach has been proposed by [39] and further developed in [25, 21]: it is based on the Benamou-Brenier [7] representation formula for the Wasserstein distance

$$W_2^2(\rho_0,\rho_1) = \inf\Big\{ \int_0^1\int_{\mathbb{R}^d} |\mathbf{v}_t|^2\,\mathrm{d}\rho_t\,\mathrm{d}t \ : \\ \partial_t\rho_t + \nabla\cdot(\rho_t\mathbf{v}_t) = 0 \text{ in } \mathbb{R}^d\times(0,1), \quad \rho_0 = \rho_{|t=0}, \quad \rho_1 = \rho_{|t=1}\Big\}$$

and on a careful analysis of the effect of the evolution semigroup generated by the equation on curves in $\mathcal{P}_2(\mathbb{R}^d)$ and its Riemannian tensor $\int_{\mathbb{R}^d} |\mathbf{v}|^2\,\mathrm{d}\rho$. This technique involves various repeated differentiations and works quite well if a nice semigroup preserving smoothness and strict positivity of the densities has already been defined. Once contraction has been proved on smooth initial data, the evolution can be extended to more general ones but it seems hard to extend the uniqueness result to cover a general distributional solution to the equation.

Chapter 4: contraction property of the Fokker-Planck equation. Our purpose is to show (1.14) working directly on measure-valued solutions solutions to (1.11) just satisfying the usual distributional formulation (1.12).
We note that in general (1.11) does not exhibit the same regularization effect of the heat equation. Even in the gradient case $B = \nabla V$, there exist solutions ρ_t to (1.12) which are not of class $C^1(\mathbb{R}^d)$ for every $t \ge 0$: take, e.g., the

invariant measure $\rho_t \equiv Z^{-1}\mathrm{e}^{-V}$ for a suitable convex function $V \notin C^1(\mathbb{R}^d)$ with $\mathrm{e}^{-V} \in L^1(\mathbb{R}^d)$. Moreover, contraction property for such a weak class of solutions provides a stability result useful to obtain distributional solutions via approximation arguments, as regularization or splitting methods: we have an example in chapter 5.

Our approach, developed from [40], relies on the well-known dual Kantorovich formulation [51] and a comparison result for the backward Kolmogorov equation. We detail our strategy in a simple case: let us assume that B is monotone ($\lambda = 0$), bounded and smooth; estimate (1.14) reads for a certain time T as

$$W_2^2(\rho_T^1, \rho_T^2) \le W_2^2(\rho_0^1, \rho_0^2). \tag{1.19}$$

The dual Kantorovich formulation (see §2.3.1) of the L^2-Wasserstein distance

$$W_2^2(\rho^1, \rho^2) = \sup\Big\{ \int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\rho^1 + \int_{\mathbb{R}^d} \phi^2 \,\mathrm{d}\rho^2 : \\ \phi^1, \phi^2 \in C_{\mathrm{b}}(\mathbb{R}^d),\ \phi^1(x_1) + \phi^2(x_2) \le |x_1 - x_2|^2 \Big\}$$

reduces the estimate of the cost $W_2^2(\rho_T^1, \rho_T^2)$ to the estimate of

$$\Sigma(\phi^1, \phi^2; T) := \int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\rho_T^1 + \int_{\mathbb{R}^d} \phi^2 \,\mathrm{d}\rho_T^2$$

for an arbitrary pair of functions ϕ^1, ϕ^2 admissible, i.e. satisfying the constraint

$$\phi^1(x_1) + \phi^2(x_2) \le |x_1 - x_2|^2 \quad \text{for every } x_1, x_2 \in \mathbb{R}^d. \tag{1.20}$$

Estimate (1.19) is proved if for any admissible pair ϕ_T^1, ϕ_T^2, there exists an admissible pair ϕ_0^1, ϕ_0^2 such that $\Sigma(\phi_T^1, \phi_T^2; T) \le \Sigma(\phi_0^1, \phi_0^2; 0)$. We can obtain this pair by solving the final-value problem for the adjoint equation

$$\partial_t \phi^i + \Delta\phi^i - B \cdot \nabla\phi^i = 0 \quad \text{in } \mathbb{R}^d \times (0, T), \quad \phi^i(\cdot, T) := \phi^i :$$

the distributional formulation (1.12) yields

$$\Sigma(\phi_T^1, \phi_T^2; T) = \Sigma(\phi^1(\cdot, 0), \phi^2(\cdot, 0); 0).$$

It remains to show that the pair $\phi^1(\cdot, 0), \phi^2(\cdot, 0)$ still satisfies the constraint (1.20): it follows from a comparison result for the backward Kolmogorov equation (**Theorem 4.5**) based on a "variable-doubling technique".

Through this approach it is also possible to cover the case of an arbitrary monotone field B, without any growth restriction, and to extend contraction estimates to more general transportation costs. The main result of the chapter, **Theorem 4.1** presents the contraction property for transportation costs as

$$\mathcal{C}_h(\rho^1, \rho^2) := \inf\Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} h(|x_1 - x_2|) \,\mathrm{d}\boldsymbol{\rho}(x_1, x_2) : \boldsymbol{\rho} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \\ \boldsymbol{\rho} \text{ is a coupling between } \rho^1 \text{ and } \rho^2 \Big\},$$

where the cost function $h : [0, +\infty) \to [0, +\infty)$ is a continuous and **non-decreasing** function with $h(0) = 0$. It provides the general contraction estimate

$$\mathcal{C}_{h_{\lambda t}}(\rho_t^1, \rho_t^2) \le \mathcal{C}_h(\rho_0^1, \rho_0^2) \quad \text{for every } t \ge 0, \tag{1.21}$$

formulated in terms of rescaled cost $h_{\lambda t}(r) := h(r\,\mathrm{e}^{\lambda t})$, $r \ge 0$: (1.14) is directly implied by (1.21) (see **Corollary 4.2**).

Chapter 5: Wasserstein solutions to the Fokker-Planck equation. Since B is not in general a gradient, we loose the gradient flow interpretation: the vector

$$-\Big(\frac{\nabla u}{u} + B\Big)$$

could not be interpreted as subgradient of a λ-convex functional. However, we still can study problem (1.11) in the Optimal transport framework.

Inspired by (1.15), we provide a new definition for Wasserstein solution of (1.11) (see **Definition 5.1**): a curve $\rho_t \subseteq \mathcal{P}_2(\mathbb{R}^d)$ solving

$$\partial_t \rho + \nabla\cdot(\rho \mathbf{v}_t) = 0$$

where $\mathbf{v}_t \in L^2(\rho_t, \mathbb{R}^d)$ satisfies

$$\int_{\mathbb{R}^d} \mathbf{v}_t \cdot \nabla\varphi \,\mathrm{d}\rho_t = \int_{\mathbb{R}^d} \Big(\Delta\varphi - B\cdot\nabla\varphi\Big)\,\mathrm{d}\rho_t, \quad \forall\, \varphi \in C_c^\infty(\mathbb{R}^d).$$

This solution, if it exists, is unique (see **Lemma 5.3**). Under sufficient condition on the initial datum, we prove existence of such solutions and find a metric characterization (see **Theorem 5.4**).

The proof of Theorem 5.4 is divided in two parts. In the first one we assume that B is Lipschitz and prove the existence of the solution via a *splitting method* argument: we build a piecewise constant curve $\rho^\tau : [0, +\infty[\to \mathcal{P}_2(\mathbb{R}^d)$, i.e.

$$\rho^\tau_t = \rho^{\tau,n} \in \mathcal{P}_2(\mathbb{R}^d), \quad \forall t \in](n-1)\tau, n\tau],\ n \in \mathbb{N}; \tag{1.22}$$

any step $\rho^{\tau,n}$ is determined by the previous one applying separately the evolutions described by the diffusion term and the drift term

$$\rho^{\tau,n-1} \xrightarrow{\text{diffusion}} \widehat{\rho}^{\tau,n} \xrightarrow{\text{drift}} \rho^{\tau,n}: \tag{1.23}$$

$\widehat{\rho}^{\tau,n}$ is the evolution, after a time interval τ, of the heat equation from the datum $\rho^{\tau,n-1}$, while $\rho^{\tau,n}$ is obtained by transporting $\widehat{\rho}^{\tau,n}$ along the trajectories solving

$$\dot{x} + B(x) = 0. \tag{1.24}$$

More precisely, since heat equation is an example of the gradient case of the Fokker-Planck equation (with $V = 0$), $\widehat{\rho}^{\tau,n}$ is the gradient flow of the entropy functional $\mathscr{H} := \mathscr{H}_0$ (1.13) from $\rho^{\tau,n-1}$, while $\rho^{\tau,n}$ is the push-forward of $\widehat{\rho}^{\tau,n}$ through the semigroup $S(\cdot)$ associated to (1.24). It is possible to show (see **Theorem 5.6**) that ρ^τ_t converges for $\tau \to 0$ in W_2 to the unique Wasserstein solution ρ^τ of (1.11).

In the second part, we generalize the existence result for a generic B by approximation with Lipschitz operators. We consider the approximation problem

$$\partial_t \rho_t - \Delta\rho_t - \nabla\cdot(\rho_t B^\varepsilon) = 0 \quad \rho_{|t=0} = \rho_0, \tag{1.25}$$

where B^ε is a Moreau-Yosida approximation of B. We prove (see **Theorem 5.11**) that the Wasserstein solution ρ^ε_t of (1.25) converges to unique Wasserstein solution of (1.11).

Plan of the Thesis In Chapter 2, we recall some useful results. In Section 2.1 we focus our attention on monotone operators, their approximations and their associated evolution equations; we are particularly interested in the special case of subgradients of convex equations. We report some basic elements of measure theory (narrow topology, tightness, push forward) in Section 2.2. Section 2.3 is about Optimal transportation theory: we recall the important duality result of Kantorovich and the definition of the Wasserstein distance: we discuss some properties of this distance, in particular in the one dimensional case. Section 2.4 is dedicated to absolutely continuous curves in $\mathcal{P}_2(\mathbb{R}^d)$: we recall their link with continuity equations and their metric characterization w.r.t. W_2. In Section 2.5, we give the definition of gradient flow in the Wasserstein space: we report some properties and consider the important example of the entropy functional and the heat equation.

Chapter 3 is dedicated to the sticky particle system (SPS) (see [36]). In the first section we recall some basic definition and notation and we state our main results. Section 3.2 collects the main properties related to the convex cone $\mathcal{K}$ of nondecreasing functions in $L^2(0,1)$ (projection, polar cone, subdifferential of the indicator function)—they provide simple but crucial tools for the analysis of the discrete SPS presented in section 3.3, which contains all the basic calculations. Section 3.4 deals with the existence, stability, and uniqueness of the solution in the Lagrangian formulation. The final steps of the proofs (mainly concerning the various limit processes) will be detailed in the last section, where we also show a new derivation of the Brenier–Grenier theorem [18] from the Lagrangian representation of the SPS.

The last two chapters are devoted to the Fokker-Planck equation. Chapter 4 is a joint work with Mark A. Peletier and Giuseppe Savaré [35], while the results of Chapter 5 are presented for the first time in this thesis.

Chapter 4 concerns the contraction property of the equation when the drift term is λ-monotone. In the first section, we state the main results of the chapter and describe the strategy of the proof. In Section 4.2, we collect some tools useful to our arguments: we present an approximation technique of the cost functional and a rescaling trick which allows to consider $\lambda = 0$ in the following arguments. Section 4.3 is devoted to show a comparison result for backward Kolmogorov equation (Theorem 4.5): this is the key of the proof of the contraction property (Theorem 4.1) contained in the last Section.

In Chapter 5 we introduce the definition of Wasserstein solutions to the Fokker-Planck equation with monotone drift term and discuss its existence. In the first Section we present the main statements. In Section 5.1 we prove the uniqueness of Wasserstein solutions. In Section 5.2 we prove the existence when operator B is Lipschitz continuous following a splitting method scheme. In the last Section we extend the existence result for general B by approximation with Lipschitz operators.

Chapter 2

Preliminaries

2.1 Maximal monotone operators and gradient flows in Hilbert spaces

Let H be a Hilbert space. We discuss the differential inclusion

$$\dot{x} + \partial\phi(x) \ni 0,$$

with $\partial\phi(x)$ is the subdifferential of a convex function $\phi : H \to \mathbb{R} \cup \{+\infty\}$. This equation is a particular case of a more general class of evolution equations associated to maximal monotone operators. In this section we recall some elements of the theory of the maximal monotone operators and derive some properties of the gradient flows in the Hilbert space. The aim is double: maximal monotone operators play an important role in the dissertation and it is interesting to compare the Hilbert gradient flows with the Wasserstein ones (see 2.5). We refer to [19] for an exhaustive treatment of the maximal monotone operator theory.

2.1.1 Convex cones in Hilbert spaces

Let us consider a convex subset $\mathcal{K}$ of H. We call $\mathcal{K}$ a *cone* if

$$z \in \mathcal{K} \Longrightarrow \alpha z \in \mathcal{K}, \quad \forall \alpha \geq 0.$$

The projection operator on a convex closed set K is well defined and is characterized by

$$g = \mathsf{P}_{\mathcal{K}}(f) \quad \Longleftrightarrow \quad g \in \mathcal{K}, \langle f - g, z - g \rangle \leq 0 \quad \forall\, z \in \mathcal{K}. \tag{2.1}$$

Lemma 2.1. *Let $\mathcal{K}$ a closed convex cone. Then $g = \mathsf{P}_{\mathcal{K}}(f)$ if and only if*

$$g \in \mathcal{K}, \quad \langle f - g, z \rangle \leq 0 \quad \forall\, z \in \mathcal{K}, \quad \langle f - g, g \rangle = 0 \tag{2.2}$$

Proof. It is easy to see that (2.2) implies (2.1); on the other hand, if $g = \mathsf{P}_{\mathcal{K}}(f)$,

$$g \in \mathcal{K}, \langle f - g, \alpha z - g \rangle \leq 0 \quad \forall\, z \in \mathcal{K}, \quad \forall\, \alpha \geq 0.$$

Choosing $\alpha = 2$ and $z = g$, $\langle f - g, g \rangle \leq 0$. Choosing $\alpha = 0$, $\langle f - g, g \rangle \geq 0$. Then $\langle f - g, g \rangle = 0$: the other inequality follows immediately. □

The *polar cone* $\mathcal{K}^\circ$ associated to a closed convex cone $\mathcal{K}$ is defined by

$$f \in \mathcal{K}^\circ \quad \Longleftrightarrow \quad \langle f, z\rangle \le 0 \quad \forall\, z \in \mathcal{K};$$

equivalently, $f \in \mathcal{K}^\circ \Longleftrightarrow \mathsf{P}_\mathcal{K}(f) = 0$. According to Lemma 2.1 we can reformulate projection in terms of the polar cone:

$$g = \mathsf{P}_\mathcal{K}(f) \Longleftrightarrow \; g \in \mathcal{K}, \;\; f-g \in \mathcal{K}^\circ, \; \langle f-g, g\rangle = 0. \tag{2.3}$$

2.1.2 Monotone operators

Consider an operator $A : Dom(A)_{\subseteq H} \rightrightarrows H$, not necessarily single-valued.

Definition 2.2. A is said monotone if

$$\forall x_1, x_2 \in Dom(A), \quad \langle y_1 - y_2, x_1 - x_2\rangle \quad \forall y_1 \in A(x_1), y_2 \in A(x_2).$$

A could be identified with its graph, i.e. $\{(x,y) \in H \times H : y \in A(x)\}$. The **set** of monotone operators is partially ordered w.r.t. the inclusion of the graphs: this justifies the definition of maximal monotone operators. Explicitly, a monotone operator $A : Dom(A) \rightrightarrows H$ is maximal if and only if for any $(x,y) \in H \times H$ s.t.

$$\langle y - \xi, x - \eta\rangle \ge 0, \forall(\eta, \xi) \text{ s.t. } \xi \in A(\eta) \Longrightarrow y \in A(x).$$

Example 2.3 (Subgradient of a convex function)**.** Let $\phi : H \to \mathbb{R} \cup \{+\infty\}$ be a convex function, i.e.

$$\phi(tx + (1-t)y) \le t\phi(x) + (1-t)\phi(y) \quad t \in (0,1), \; \forall x, y \in H.$$

The set $Dom(\phi) = \{x \in H : \phi(x) < +\infty\}$ is convex. The subgradient $\partial\phi : Dom(\phi) \rightrightarrows H$ of ϕ, defined by

$$\xi \in \partial\phi(x) \Longleftrightarrow \forall y \in H, \quad \phi(y) \ge \phi(x) + \langle \xi, y - x\rangle,$$

is a monotone operator. Furthermore, if ϕ is lower semicontinuous, $\partial\phi$ is also maximal.

In Chapter 3 we consider a particular convex function: the indicator function of a closed convex cone $\mathcal{K}$

$$I_\mathcal{K}(z) = \begin{cases} 0 & \text{if } z \in \mathcal{K}, \\ +\infty & \text{otherwise.} \end{cases}$$

The subgradient is given by

$$\xi \in \partial I_\mathcal{K}(g) \quad \Longleftrightarrow \quad \langle \xi, z - g\rangle \le 0 \quad \forall\, z \in \mathcal{K}.$$

In terms of the projection on $\mathcal{K}$

$$\xi \in \partial I_\mathcal{K}(g) \quad \Longleftrightarrow \quad g = \mathsf{P}_\mathcal{K}(\xi + g). \tag{2.4}$$

$\mathcal{K}^\circ$ and $\partial I_\mathcal{K}$ are clearly linked by $\mathcal{K}^\circ = \partial I_\mathcal{K}(0)$ and

$$\xi \in \partial I_\mathcal{K}(g) \quad \overset{(2.4)(2.3)}{\Longleftrightarrow} \quad g \in \mathcal{K}, \; \xi \in \mathcal{K}^\circ, \quad \langle \xi, g\rangle = 0.$$

Properties

A fundamental characterization of maximal monotone operators is given by the following

Proposition 2.4 ([19, Prop2.2, p.23])**.** *Let A be a monotone operator. The following three properties are equivalent:*

a) A is maximal monotone,

b) A is monotone and $\mathbf{i} + A$ is surjective,

c) $\forall \varepsilon > 0$, $(\mathbf{i} + \varepsilon A)^{-1} : H \to H$ is a contraction.

Moreover, maximality implies the closure of the graph of the operator in $\mathbb{R}^d \times \mathbb{R}^d$ [19, Prop. 2.5, p.27]. By Proposition 2.4 the *resolvent* $J_\varepsilon^A : H \to H$ of the maximal monotone operator A

$$J_\varepsilon^A(x) = (\mathbf{i} + \varepsilon A)^{-1}(x), \quad x \in H,$$

is a well defined contraction; it is also possible to check that $\lim_{\varepsilon\to 0} J_\varepsilon^A(x) = x$, $\forall x \in Dom(A)$. Since maximality ensures the convexity of $A(x)$, we can introduce the *minimal selection* $A^\circ : Dom(A) \to H$ of A: it is a monotone operator s.t.

$$A^\circ(x) \in A(x), \quad |A^\circ(x)| = \min\{|y| : y \in A(x)\}.$$

Now we could give the description of the important tool of the *Moreau-Yosida approximation* $A^\varepsilon : H \to H$ of A [19, Prop.2.6],

$$A^\varepsilon(x) := \frac{x - J_\varepsilon^A(x)}{\varepsilon}.$$

Proposition 2.5. *A^ε is maximal monotone and Lipschitz continuous with Lipschitz constant less equal than $1/\varepsilon$. Moreover, for any $x \in H$*

$$\lim_{\varepsilon\to 0} A^\varepsilon(x) = A^\circ(x), \quad |A^\varepsilon(x)| \uparrow |A^\circ(x)| \text{ for } \varepsilon \downarrow 0.$$

Let us recall a very simple (but useful) Lemma holds (see [19, p.62]):

Lemma 2.6. *Let $A : Dom(A) \to H$ be a monotone operator. If $Dom(A)$ contains a ball $\mathrm{B}_R(x_0)$ with center in x_0 and radius $R > 0$, there exists $M > 0$ s.t.*

$$|A(x)| \le \frac{1}{R} A(x) \cdot (x - x_0) + \frac{1}{R} M |x - x_0| + M. \tag{2.5}$$

Proof. For any z s.t. $|z| \le 1$,

$$\langle A(x) - A(x_0 + Rz), x - x_0 - Rz \rangle \ge 0.$$

In particular,

$$\begin{aligned} R\langle A(x), z\rangle &\le \langle A(x), x - x_0\rangle - \langle A(Rz), x - x_0\rangle + \langle A(Rz), Rz\rangle \\ &\le \langle A(x), x - x_0\rangle + |x - x_0| \sup_{\mathrm{B}_R(x_0)} |A| + R \sup_{\mathrm{B}_R(x_0)} |A| \end{aligned}$$

Choosing $z := A(x)/|A(x)|$ and $M := \sup_{\mathrm{B}_R(x_0)} |A|$, we obtain (2.5). □

2.1.3 Bounded, smooth approximations of a monotone operator in finite dimension

In Chapters 4 and 5 we will assume that $H = \mathbb{R}^d$. If $A : \mathbb{R}^d \to \mathbb{R}^d$ is a monotone operator then there exists [19, Corollary 2.1] a maximal monotone multivalued extension $\mathsf{A} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ (thus taking values in $2^{\mathbb{R}^d}$) such that $A(x) \in \mathsf{A}(x)$ for every $x \in \mathbb{R}^d$. We denote by $\mathsf{A}^\circ(x)$ the minimal selection of $\mathsf{A}(x)$. [1, Corollary 1.4] shows that the set $\mathsf{A}(x) \subset \mathbb{R}^d$ reduces to the singleton $\{A(x)\}$ $\mathscr{L}^d$-almost everywhere: in fact it satisfies

$$\begin{aligned} \mathsf{A}(x) &= \{\mathsf{A}^\circ(x)\} = \{A(x)\} \quad \text{for } \mathscr{L}^d\text{-a.e. } x \in \mathbb{R}^d, \\ \mathsf{A}(x) &= \operatorname{conv}\big\{ \lim_{n\to\infty} A(x_n) \text{ for some } x_n \to x \big\}. \end{aligned} \tag{2.6}$$

We recall the following important approximation result [27, Theorem 4.1]: we denote by U the open unit ball in $\mathbb{R}^d$.

Theorem (Fitzpatrick-Phelps)**.** *For any maximal monotone operator $\mathsf{A} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$, there exists a sequence of maximal monotone operators $\mathsf{A}_n : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ such that, for each $x \in \mathbb{R}^d$ and all n,*

$$\mathsf{A}(x) \cap n\,U \subset \mathsf{A}_n(x) \subset n\,\overline{U}, \quad \mathsf{A}_n(x) \setminus \mathsf{A}(x) \subset n\,\partial U \quad \textit{for every } x \in \mathbb{R}^d. \tag{2.7}$$

Notice that (2.7) yields in particular

$$|\mathsf{A}^\circ_n(x)| = \min\big(|\mathsf{A}^\circ(x)|, n\big) \quad \text{for every } x \in \mathbb{R}^d. \tag{2.8}$$

Theorem 2.7. *Let $\mathsf{A} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be a maximal monotone operator and $(\beta_n)_{n\in\mathbb{N}}$ a vanishing sequence of positive real numbers. There exists a sequence of smooth, globally Lipschitz, and bounded monotone operators $A_n : \mathbb{R}^d \to \mathbb{R}^d$ such that*

$$\begin{aligned} \operatorname{Lip}(A_n) \le n, \quad |A_n(x)| \le \min\big(|\mathsf{A}^\circ(x)|, n\big) + \beta_n, \\ \lim_{n\to+\infty} A_n(x) = \mathsf{A}^\circ(x) \quad \textit{for every } x \in \mathbb{R}^d. \end{aligned} \tag{2.9}$$

Proof. Let A_n be a sequence of maximal monotone operators satisfying (2.7) and let $Y_n : \mathbb{R}^d \to \mathbb{R}^d$ be the Moreau-Yosida approximation of A_n of parameter n^{-1} (see Prop. 2.5)

$$Y_n(x) := n\Big(x - (I + n^{-1}\,\mathsf{A}_n)^{-1} x\Big)$$

Note that Y_n is a n-Lipschitz monotone map satisfying

$$|Y_n(x)| \le |\mathsf{A}^\circ_n(x)| \overset{(2.8)}{=} \min\big(|\mathsf{A}^\circ(x)|, n\big) \quad \text{for every } x \in \mathbb{R}^d \tag{2.10}$$

Let us fix $x \in \mathbb{R}^d$ and let $x_n \in \mathbb{R}^d$ be the unique solution of

$$x_n + n^{-1}\,\mathsf{A}_n(x_n) \ni x \quad \text{so that} \quad Y_n(x) = n(x - x_n) \in \mathsf{A}_n(x_n).$$

If $n > |\mathsf{A}^\circ(x)|$ then (2.10) yields $Y_n(x) \notin n\,\partial U$; applying (2.7) and (2.10) again we get

$$\begin{gathered} Y_n(x) \in \mathsf{A}(x_n), \quad |Y_n(x)| \le |\mathsf{A}^\circ(x)|, \\ |x - x_n| \le n^{-1}\,|\mathsf{A}^\circ(x)| \text{ for every } n > |\mathsf{A}^\circ(x)|. \end{gathered}$$

Since the graph of A is closed, any accumulation point y of the bounded sequence $Y_n(x)$ satisfies

$$y \in \mathsf{A}(x), \quad |y| \le |\mathsf{A}^\circ(x)|.$$

We thus conclude that $\lim_{n\uparrow+\infty} Y_n(x) = \mathsf{A}^\circ(x)$ for every $x \in \mathbb{R}^d$.

To conclude the proof we need to regularize Y_n: to this aim we consider the family of mollifiers κ_η defined by

$$\kappa_\eta(x) := \eta^{-d}\kappa\big(x/\eta\big), \quad x \in \mathbb{R}^d, \ \eta > 0, \tag{2.11}$$

where $\kappa \in C^\infty_{\mathrm{c}}(\mathbb{R}^d)$, $\kappa \ge 0$ and $\int_{\mathbb{R}^d} \kappa(x)\,\mathrm{d}x = 1$; we set

$$A_n := Y_n * \kappa_\eta \quad \text{with} \quad \eta := (n\,k)^{-1}\beta_n \text{ where } k := \int_{\mathbb{R}^d} |x|\kappa(x)\,\mathrm{d}x,$$

so that

$$|A_n(x) - Y_n(x)| \le \eta\,k\,\mathrm{Lip}(Y_n) \le n\,\eta\,k \le \beta_n. \qquad \square$$

We consider now a radial smoothing:

Proposition 2.8. *Let $A_n : \mathbb{R}^d \to \mathbb{R}^d$ be smooth, Lipschitz, and bounded monotone operators satisfying* (2.9)*. For every $m \in \mathbb{N}$ there exists bounded, smooth, Lipschitz, and monotone operators $A_{n,m}$ such that*

$$\begin{gathered}\mathrm{Lip}(A_{n,m}) \le n, \quad \sup_{x\in\mathbb{R}^d} |A_{n,m}(x)| \le n + \beta_n, \\ \sup_{x\in\mathbb{R}^d} \big|\mathrm{D}\,A_{n,m}(x)\cdot x\big| \le 2m\,(n+\beta_n),\end{gathered} \tag{2.12}$$

$$\lim_{m\uparrow+\infty} A_{n,m}(x) = A_n(x) \quad \textit{for every } x \in \mathbb{R}^d. \tag{2.13}$$

Proof. We consider a family of mollifiers $\kappa_\eta = \eta^{-1}\kappa(\cdot/\eta) \in C^\infty_{\mathrm{c}}(\mathbb{R})$, where κ satisfies

$$\mathrm{supp}(\kappa) \subset [0,2], \quad 0 \le \kappa \le \kappa(1) = 1, \quad (1-x)\kappa'(x) \ge 0, \quad \int_{\mathbb{R}} \kappa(x)\,\mathrm{d}x = 1, \tag{2.14}$$

and the function $\vartheta \in C^\infty_{\mathrm{c}}(0,+\infty)$ defined by $\vartheta(r) := \kappa(-\log r)$, $r > 0$. We set

$$A_{n,m}(x) := m\int_0^{+\infty} A(rx)\vartheta(r^m)\,\frac{\mathrm{d}r}{r}$$

The change of variable $r = e^{-z}$ shows that

$$A_{n,m}(x) = m\int_{\mathbb{R}} A_n(x\,\mathrm{e}^{-z})\,\kappa(m\,z)\,\mathrm{d}z = A_n^x * \kappa_{1/m}(0),$$

where $A_n^x(z) := A_n(x\,e^z)$ for $z \in \mathbb{R}$. It is then easy to check that $|\mathrm{D}A_{n,m}| \le n$ since

$$|DA_{n,m}(x)| \le m\int_{\mathbb{R}} \big|\mathrm{D}A_n(x\,\mathrm{e}^{-z})\big|\mathrm{e}^{-z}\,\kappa(m\,z)\,\mathrm{d}z \overset{(2.9)}{\le} n\int_{\mathbb{R}} \mathrm{e}^{-y/m}\,\kappa(y)\,\mathrm{d}y \overset{(2.14)}{\le} n,$$

and $A_{n,m}$ converges pointwise to A_n as $m \uparrow +\infty$.

Concerning the second bound of (2.12) we easily have

$$\mathrm{D}\, A_{n,m}(x) \cdot x = m \int_0^{+\infty} \mathrm{D}\, A_n(rx) \cdot x \vartheta(r^m)\, \mathrm{d}r = m \int_0^{+\infty} \frac{\mathrm{d}}{\mathrm{d}r} \Big(A_n(rx) \Big) \vartheta(r^m)\, \mathrm{d}r$$
$$= -m^2 \int_0^{+\infty} A_n(rx) \tilde{\vartheta}(r^m) \frac{\mathrm{d}r}{r} \qquad \text{where } \tilde{\vartheta}(r) := r\vartheta'(r);$$

the inequality follows since by (2.14) the total variation of κ and thus of ϑ is 2, so that

$$m^2 \int_0^{+\infty} |\tilde{\vartheta}(r^m)| \frac{\mathrm{d}r}{r} = m \int_0^{+\infty} |\tilde{\vartheta}(r)| \frac{\mathrm{d}r}{r} = m \int_0^{+\infty} |\vartheta'(r)|\, \mathrm{d}r = 2m. \qquad \square$$

2.1.4 Evolution equations associated to maximal monotone operators

Consider the equation

$$\dot{x} + A(x) \ni 0, \tag{2.15}$$

where $A : Dom(A) \rightrightarrows H$ is maximal monotone.

Theorem 2.9 ([19, Th.3.1, p.54])**.** *For any $x_0 \in Dom(A)$ there exists a curve $x(t)$ from $[0, +\infty[$ to H s.t.*

1) $x(0) = x_0$ and $x(t) \in Dom(A)$;

2) $x(t)$ is Lipschitz continuous in $[0, \infty[$ and solves (2.15) $\mathscr{L}^1$-a.e.;

3) x admits the right derivative for any $t \in [0, +\infty[$ and

$$\frac{\mathrm{d}^+}{\mathrm{d}t} x(t) + A^\circ x(t) = 0, \qquad \Big| \frac{\mathrm{d}^+}{\mathrm{d}t} x(t) \Big| \le |A^\circ(x_0)|.$$

Moreover, considering two solutions $x^1(t), x^2(t)$ respectively associated to the starting points $x_0^1, x_0^2 \in Dom(A)$,

$$|x^1(t) - x^2(t)| \le |x_0^1 - x_0^2|, \quad t \in [0, +\infty[. \tag{2.16}$$

For any $t > 0$, the map $x_0 \to x(t)$ is a contraction in $Dom(A)$. Denoting with $S(t)$ the extension (by continuity) of this map on $\overline{Dom(A)}$, it can be checked that $S(t)$ is a continuous semigroup of contractions on $\overline{Dom(A)}$, i.e.

$$S(t) : \overline{Dom(A)} \to \overline{Dom(A)} \text{ is a contraction;} \tag{2.17a}$$
$$S(t)S(s) = S(t+s), \text{ for } t, s > 0; \tag{2.17b}$$
$$\lim_{t \to 0} S(t)x_0 = x_0, \text{ for any } x_0 \in \overline{Dom(A)}. \tag{2.17c}$$

From Theorem 2.9 3), we also have

$$|x_0 - S(t)x_0| \le t|A^\circ(x_0)|. \tag{2.17d}$$

$S(t)$ is called the semigroup generated by $-A$. It can be approximated via the exponential formula involving the resolvent of A: as shown in [19, Cor.4.4, p.126], for any $x \in Dom(A)$,

$$\left| \left(J^A_{\frac{t}{m}} \right)^m (x) - S(t)x \right| \le 2 \frac{t}{\sqrt{m}} |A^\circ(x)|. \tag{2.18}$$

The correspondence between semigroups and maximal monotone operators is one-to-one. Let $S(t)$ be a semigroup of contractions in a convex set C: there is a unique maximal monotone operator A satisfying the relation (see [19, Th.4.1, p.114])

$$\overline{Dom(A)} = C, \quad \lim_{t \downarrow 0} \frac{x - S(t)x}{t} = A^\circ(x), \, \forall x \in Dom(A) \tag{2.19}$$

Now consider a lower semicontinuous convex function $\phi : H \to \mathbb{R} \cup \{+\infty\}$ and the related gradient flow

$$\dot{x} + \partial\phi(x) \ni 0. \tag{2.20}$$

Since $\partial\phi$ is maximal monotone, Theorem 2.15 applies: the equation admits a solution for any initial datum $x_0 \in Dom(\partial\phi)$. Moreover (see [19, Th.3.2, p.57]),

Theorem 2.10. *Let $x(t)$ be the solution of the differential inclusion (2.20) with $x(0) = x_0 \in Dom(\partial\phi)$. The map $t \mapsto \phi(x(t))$ is convex, not increasing e Lipschitz continuous on $[0, +\infty[$ and*

$$\frac{\mathrm{d}^+}{\mathrm{d}t} \phi(x(t)) = -\left| \frac{\mathrm{d}^+}{\mathrm{d}t} x(t) \right|^2.$$

We can also notice that the solution of the gradient flow could be characterized by the Evolution Variational Inequality

$$\frac{1}{2} \frac{d}{dt}^+ |x(t) - y|^2 = \langle \dot{x}(t), x(t) - y \rangle \le \phi(y) - \phi(x(t)), \text{ for } \mathscr{L}^1\text{-a.e. } t > 0, \quad \forall y \in H. \tag{2.21}$$

2.1.5 Nonhomogeneous problems and perturbed operators

Consider the equation

$$\dot{x} + \partial\phi(x) - \lambda x \ni w, \tag{2.22}$$

where $\partial\phi : Dom(\partial\phi) \rightrightarrows H$ is the subgradient of a lower semicontinuous convex function ϕ and $\lambda \ge 0$. Then, (see [19, Theorems 3.17, 3.4, 3.5 and 3.6])

Theorem 2.11. *If $w \in L^2(0, T, H)$, for any $x_0 \in Dom(\partial\phi)$ there exists a unique curve $x(t) : [0, T] \to H$ s.t.*

1) $x(0) = x_0$ and $x(t) \in Dom(\partial\phi)$ for any Lebesgue point t of w in $]0, T[$;

2) $x(t)$ is (locally) absolutely continuous in $]0, T[$;

3) x solves (2.22) $\mathscr{L}^1$-a.e. in $[0, T]$;

4) x admits the right derivative for any Lebesgue point t of w in $]0, T[$.

We recall a consequence of the stability result [19, Prop. 3.14, p.108] for the nonhomogeneous equation (2.22):

Theorem 2.12. *Let $w^n, w \in L^2(0,T,H)$ and $x_0^n, x_0 \in Dom(\partial\phi)$ and let us consider solutions $x^n(t)$ of*

$$\dot{x^n} + \partial\phi(x^n) - \lambda x^n \ni w^n, \quad x^n(0) = x_0^n,$$

and the solution $x(t)$ of (2.22) with starting point $x(0) = x_0$. If $x_0^n \to x_0$, $w^n \to w$ in $L^2(0,T,H)$, then $x^n(t)$ converges uniformly to $x(t)$ on $[0,T]$.

2.2 Elements of measure theory

In this section we recall some basic results $\mathcal{P}(\mathbb{R}^d)$, the set of Borel probability measures on $\mathbb{R}^d$.

2.2.1 Narrow topology and tightness

We say that a sequence $\rho_n \in \mathcal{P}(\mathbb{R}^d)$ *narrowly (or weakly) converges* to $\rho_n \in \mathcal{P}(\mathbb{R}^d)$ if

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x) \mathrm{d}\rho_n(x) = \int_{\mathbb{R}^d} \varphi(x) \mathrm{d}\rho(x), \quad \forall \varphi \in C_b(\mathbb{R}^d),$$

where $C_b(\mathbb{R}^d)$ is the space of continuous bounded real functions on $\mathbb{R}^d$.

It is well known that the narrow convergence is induced by a distance on $\mathcal{P}(\mathbb{R}^d)$ and we call *narrow topology* the topology induced by this distance. In particular the compact subsets of $\mathcal{P}(\mathbb{R}^d)$ coincide with sequentially compact subsets of $\mathcal{P}(\mathbb{R}^d)$.

We also recall that if $\rho_n \in \mathcal{P}(\mathbb{R}^d)$ narrowly converges to $\rho \in \mathcal{P}(\mathbb{R}^d)$ and $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a lower semi continuous function bounded from below, then ([2, Lemma 5.1.7])

$$\liminf_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x) \mathrm{d}\rho_n(x) \geq \int_{\mathbb{R}^d} \varphi(x) \mathrm{d}\rho(x). \tag{2.23}$$

A subset $\mathcal{I} \subset \mathcal{P}(\mathbb{R}^d)$ is said to be tight if

$$\forall \varepsilon > 0 \exists K_\varepsilon \subset \mathbb{R}^d \text{ compact } : \rho(\mathbb{R}^d \setminus K_\varepsilon) < \varepsilon \quad \forall \rho \in \mathcal{I}.$$

The importance of tight sets is due to the following

Theorem 2.13 (Prokhorov)**.** *$\mathcal{I} \subset \mathcal{P}(\mathbb{R}^d)$ is tight if and only if it is relatively compact in $\mathcal{P}(\mathbb{R}^d)$.*

There exists an interesting link between narrow convergence of probability measures and Kuratowski convergence of their supports (see [2, Prop.5.1.8]):

Proposition 2.14. *If $(\rho_n) \in \mathcal{P}(\mathbb{R}^d)$ is a sequence narrowly converging to $\rho \in \mathcal{P}(\mathbb{R}^d)$ then $\operatorname{spt}\rho \subset K - \liminf_{n\to\infty} \operatorname{spt}\rho_n$, i.e.*

$$\forall x \in \operatorname{spt}\rho, \quad \exists x_n \in \operatorname{spt}\rho_n : \lim_{n\to\infty} x_n = x.$$

2.2.2 Push forward of measures

If $\rho \in \mathcal{P}(\mathbb{R}^d)$ and $\mathbf{t} : \mathbb{R}^d \to \mathbb{R}^m$ is a Borel map. the *push forward of* ρ *through* $\mathbf{t}$, denoted by $\mathbf{t}_{\#}\rho \in \mathcal{P}(\mathbb{R}^m)$, is defined as follows:

$$\mathbf{t}_{\#}\rho[E] := \rho[\mathbf{t}^{-1}(E)], \quad \forall E \in \mathscr{B}(\mathbb{R}^m),$$

where $\mathscr{B}(\mathbb{R}^d)$ is the family of Borel subset of $\mathbb{R}^m$. We will repeatedly use the change-of-variable formula

$$\int_{\mathbb{R}^d} \zeta(y)\,\mathrm{d}(\mathbf{t}_{\#}\rho)(y) = \int_{\mathbb{R}^m} \zeta(\mathbf{t}(x))\,\mathrm{d}\rho(x) \quad \text{for every Borel } \zeta : \mathbb{R}^m \to [0, +\infty].$$

We will use the following result on the convergence or push-forward measures ([2, Lemma 5.2.1]):

Lemma 2.15. *Let* $\mathbf{t}_n : \mathbb{R}^d \to \mathbb{R}^m$ *be Borel maps uniformly converging to* $\mathbf{t}$ *on compact subsets of* $\mathbb{R}^d$ *and let* $(\rho_n) \in \mathcal{P}(\mathbb{R}^d)$ *be a tight sequence narrowly converging to* ρ. *If* $\mathbf{t}$ *is continuous, then* $\mathbf{t}_{n\#}\rho_n$ *narrowly converge to* $\mathbf{t}_{\#}\rho$.

Let $\rho \in \mathcal{P}(\mathbb{R}^d)$ and consider a sequence of Borel maps $\mathbf{t}_n : \mathbb{R}^d \to \mathbb{R}^d$ converging in ρ-measure to $\mathbf{t} : \mathbb{R}^d \to \mathbb{R}^d$, i.e.

$$\lim_{n\to\infty} \rho(\{x \in \mathbb{R}^d : |\mathbf{t}_n(x) - \mathbf{t}(x)| > \delta\}) = 0 \quad \forall \delta > 0.$$

Lemma 2.16 ([2, Lemma 5.4.1])**.** *Let* $\mathbf{t}_n, \mathbf{t} : \mathbb{R}^d \to \mathbb{R}^d$ *be Borel maps and let* $\rho \in \mathcal{P}(\mathbb{R}^d)$. *Then* $\mathbf{t}_n \to \mathbf{t}$ *in* ρ*-measure if and only if* $(\mathbf{i} \times \mathbf{t}_n)_{\#}\rho$ *converges to* $(\mathbf{i} \times \mathbf{t})_{\#}\rho$ *narrowly in* $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$.

Starting from an absolute continuous measure, it is possible to determine the density of the push-forward, if exists (see [2, Lemma 5.5.3])

Lemma 2.17 (Density of the push-forward)**.** *Let* $u \in L^1(\mathbb{R}^d)$ *be a nonnegative function and* $\mathbf{t} : \mathbb{R}^d \to \mathbb{R}^d$ *be a Borel bijective map s.t.* $\mathbf{t}$ *is* $\mathscr{L}^d$*-a.e. differentiable. Then* $\mathbf{t}_{\#}(u\mathscr{L}^d) \ll \mathscr{L}^d$ *if and only if* $|\det \nabla\mathbf{t}| > 0$ $\mathscr{L}^d$*-a.e. and in this case*

$$\mathbf{t}_{\#}(u\mathscr{L}^d) = \frac{u}{|\det \nabla\mathbf{t}|} \circ \mathbf{t}^{-1}\mathscr{L}^d. \tag{2.24}$$

2.3 The Optimal transportation problem

The first version of the optimal transportation was proposed by Monge in 1781. Let two measures $\rho^1, \rho^2 \in \mathcal{P}(\mathbb{R}^d)$ and a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty]$, the Monge's formulation is given by

$$\inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, \mathbf{t}(x))\,\mathrm{d}\rho^1(x) : \mathbf{t}_{\#}\rho^1 = \rho^2 \right\}.$$

In general, this problem is ill posed: for example, if ρ^1 is a Dirac measure and ρ^2 is absolutely continuous (w.r.t Lebesgue), there is no map $\mathbf{t}$ such that $\mathbf{t}_{\#}\rho^1 = \rho^2$. In 1942, Kantorovich reformulated the problem in a relaxed version:

$$\min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x_1, x_2)\,\mathrm{d}\boldsymbol{\rho}(x_1, x_2) : \boldsymbol{\rho} \text{ is a coupling between } \rho^1 \text{ and } \rho^2 \right\} \tag{2.25}$$

in terms of couplings, i.e. measures $\boldsymbol{\rho}$ in the product space $\mathbb{R}^d \times \mathbb{R}^d$ whose marginals are ρ^1 and ρ^2 respectively, so that $\boldsymbol{\rho} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ and

$$\boldsymbol{\rho}(E \times \mathbb{R}^d) = \rho^1(E), \quad \boldsymbol{\rho}(\mathbb{R}^d \times E) = \rho^2(E) \text{ for every Borel subset } E \subset \mathbb{R}^d.$$

There exists at least a coupling (e.g., $\rho^1 \times \rho^2$) and the minimum is always attained if c is lower semicontinuous (thanks to the tightness of the set of the couplings).

The Optimal Transportation Theory has been developed starting from these problems (see for instance [2] [51] for much more on this subject). In this section we only recall a few basic facts.

In Chapter 4, we focus our attention on cost functions in the form $c(x_1, x_2) = h(|x_1 - x_2|)$, with $h : [0, +\infty) \to [0, +\infty)$ is a continuous e non-decreasing function with $h(0) = 0$:

$$\begin{aligned} \mathcal{C}_h(\rho^1, \rho^2) := \inf \Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} h(|x_1 - x_2|) \, \mathrm{d}\boldsymbol{\rho}(x_1, x_2) : \boldsymbol{\rho} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \\ \boldsymbol{\rho} \text{ is a coupling between } \rho^1 \text{ and } \rho^2 \Big\}. \end{aligned} \tag{2.26}$$

Among the possible interesting choices of h, the case $h(r) := r^p$ is associated to the family of L^p Wasserstein distances on the space $\mathcal{P}_p(\mathbb{R}^d)$ of all the probability measures with moment of order p (see §2.3.2). When h is a bounded concave function satisfying $h(r) > 0$ if $r > 0$, $d(x, y) := h(|x - y|)$ is a bounded and complete distance function on $\mathbb{R}^d$ inducing the usual euclidean topology so that $\mathcal{C}_h(\cdot, \cdot)$ is a complete metric on the space $\mathcal{P}(\mathbb{R}^d)$ whose topology coincides with the usual weak one (see, e.g., [2, Proposition 7.1.5]).

2.3.1 Kantorovich duality

The minimization problem (2.25) admits a dual formulation, due to Kantorovich. It is possible to show (see [51]):

Theorem 2.18 (Kantorovich duality)**.** *If the cost function c is lower semicontinuous, the minimum in* (2.25) *is equal to*

$$\begin{aligned} \sup \Big\{ \int_{\mathbb{R}^d} \phi^1 \, \mathrm{d}\rho^1 + \int_{\mathbb{R}^d} \phi^2 \, \mathrm{d}\rho^2 : \phi^1, \phi^2 \in C_{\mathrm{b}}(\mathbb{R}^d), \\ \phi^1(x_1) + \phi^2(x_2) \le c(x_1, x_2) \Big\}. \end{aligned} \tag{2.27}$$

For the transportation cost $\mathcal{C}_h(\rho^1, \rho^2)$ (2.26), it is possible to get a more refined version of the duality formula (2.27): under suitable conditions on h, we can consider only ϕ^1, ϕ^2 smooth and compactly supported.

Proposition 2.19. *If the cost function h is Lipschitz continuous and satisfies $\lim_{r \uparrow +\infty} h(r) = +\infty$, then*

$$\begin{aligned} \mathcal{C}_h(\rho^1, \rho^2) = \sup \Big\{ \int_{\mathbb{R}^d} \phi^1 \, \mathrm{d}\rho^1 + \int_{\mathbb{R}^d} \phi^2 \, \mathrm{d}\rho^2 : \phi^1, \phi^2 \in C^\infty_{\mathrm{c}}(\mathbb{R}^d), \\ \phi^1(x_1) + \phi^2(x_2) \le h(|x_1 - x_2|) \Big\}. \end{aligned} \tag{2.28}$$

Proof. Since Theorem 2.18 applies to $\mathcal{C}_h(\rho^1, \rho^2)$, we derive (2.28) from (2.27). Let us recall that the h-transform of a given bounded function $\zeta : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\zeta^h(x) := \inf_{y \in \mathbb{R}^d} h(|x - y|) - \zeta(y),$$

and it is a bounded and Lipschitz continuous function satisfying $\zeta(x) + \zeta^h(y) \le h(|x - y|)$.

Let us fix $c < \mathcal{C}_h(\rho^1, \rho^2)$ and admissible $\phi^1, \phi^2 \in C_{\mathrm{b}}(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d} \phi^1 \, \mathrm{d}\rho^1 + \int_{\mathbb{R}^d} \phi^2 \, \mathrm{d}\rho^2 > c. \tag{2.29}$$

By possibly replacing ϕ^2 with $(\phi^1)^h \ge \phi^2$ and ϕ^1 with $(\phi^1)^{hh} \ge \phi^1$, it is not restrictive to assume that ϕ^1, ϕ^2 are also Lipschitz continuous. Adding to ϕ^1 and subtracting from ϕ^2 a suitable constant, we can also assume that $\phi^1 \ge 0$ and $\phi^2 \le 0$.

Let us now consider mollifiers κ_η as in (2.11) and a family of cutoff functions χ_R defined by

$$\chi_R(x) := \chi(x/R) \quad x \in \mathbb{R}^d, \ R > 0,$$

where $\chi \in C^\infty_{\mathrm{c}}(\mathbb{R}^d)$ satisfies

$$0 \le \chi \le 1, \quad \chi(x) = 0 \text{ if } |x| \ge 1, \quad \chi(x) = 1 \text{ if } |x| \le 1/2.$$

We set $\phi^1_\eta := \phi^1 * \kappa_\eta$ and $\phi^2_\eta := \phi^2 * \kappa_\eta - \delta_\eta$, where

$$\delta_\eta := \sup(\phi^1 * \kappa_\eta - \phi^1)^+ + \sup(\phi^2 * \kappa_\eta - \phi^2)^+.$$

The definition of δ_η yields

$$\begin{aligned}
&\phi^1_\eta(x_1) + \phi^2_\eta(x_2) \le \\
&\qquad \le \phi^1 * \kappa_\eta(x_1) - \phi^1(x_1) + \phi^2 * \kappa_\eta(x_2) - \phi^2(x_2) - \delta_\eta + h(|x_1 - x_2|) \\
&\qquad \le h(|x_1 - x_2|).
\end{aligned}$$

Moreover, since ϕ^1, ϕ^2 are Lipschitz, ϕ^1_η and ϕ^2_η converge to ϕ^1, ϕ^2 uniformly as $\eta \downarrow 0$, so that ϕ^1_η and ϕ^2_η are a smooth admissible pair still satisfying the sign condition $\phi^1_\eta \ge 0$, $\phi^2_\eta \le 0$ and (2.29) for a sufficiently small $\eta > 0$.

Let us now choose $R_0 > 0$ such that

$$h(r) \ge \sup \phi^1_\eta \quad \text{for every } r \ge R_0$$

Setting $\phi^1_{\eta,R} := \phi^1_\eta \, \chi_R \le \phi^1_\eta$ we easily have for $R \ge R_0$

$$\inf_{x_1 \in \mathbb{R}^d} h(|x_1 - x_2|) - \phi^1_{\eta,R}(x_1) \ge 0 \quad \text{if } |x_2| \ge 2R \ge R + R_0.$$

Since $\phi^2_{\eta,4R} := \phi^2_\eta \, \chi_{4R}$ satisfies $\phi^2_{\eta,4R}(x_2) = \phi^2_\eta(x_2)$ if $|x_2| \le 2R$ and $\phi^2_{\eta,4R}(x_2) \le 0$ for every $x_2 \in \mathbb{R}^d$, it follows that $\phi^1_{\eta,R}, \phi^2_{\eta,4R}$ is an admissible couple in $C^\infty_{\mathrm{c}}(\mathbb{R}^d)$, and, for R sufficiently large, it still satisfies (2.29). □

2.3.2 Kantorovich-Rubinstein-Wasserstein distance

We fix $p \geq 1$ and denote by $\mathcal{P}_p(\mathbb{R}^d)$ the space of Borel probability measures having finite p-moments, i.e.

$$\mathcal{P}_p(\mathbb{R}^d) = \Big\{\rho \in \mathcal{P}_p(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^p \mathrm{d}\rho(x) < \infty\Big\}$$

Given $\rho^1, \rho^2 \in \mathcal{P}_p(\mathbb{R}^d)$, the so called L^p-Kantorovich-Rubinstein-Wasserstein distance $W_p(\cdot,\cdot)$ between ρ^1 and ρ^2 can be defined by

$$W_p(\rho^1,\rho^2) := \Bigg(\min\Big\{\int_{\mathbb{R}^d\times\mathbb{R}^d} \big|x_1 - x_2\big|^p \,\mathrm{d}\boldsymbol{\rho}(x_1,x_2) : \\ \boldsymbol{\rho} \text{ is a coupling between } \rho^1 \text{ and } \rho^2\Big\}\Bigg)^{\frac{1}{p}} \tag{2.30}$$

We denote with $\Gamma_o(\rho^1,\rho^2)$ the set of all *optimal couplings* realizing the minimum in (2.30). $\mathcal{P}_p(\mathbb{R}^d)$ endowed with the distance W_p is a separable and complete metric space, called *Wasserstein space.*

We recall that, given a sequence $\rho^n \in \mathcal{P}_p(\mathbb{R}^d)$ and $\rho \in \mathcal{P}_p(\mathbb{R}^d)$,

$$\lim_{n\to\infty} W_p(\rho^n,\rho) = 0 \Longleftrightarrow \begin{cases} \rho^n \text{ narrowly converges to } \rho, \\ \lim_{n\to\infty} \int_{\mathbb{R}^d} |x|^p \mathrm{d}\rho^n(x) = \int_{\mathbb{R}^d} |x|^p \mathrm{d}\rho(x). \end{cases}$$

In the case when $\rho^1 \in \mathcal{P}^r_p(\mathbb{R}^d)$, the subset of $\mathcal{P}_p(\mathbb{R}^d)$ made of absolutely continuous measures w.r.t. Lebesgue measure, it can be shown that he minimum problem has a unique solution $\boldsymbol{\rho}$, and $\boldsymbol{\rho}$ is induced by a transport map $\mathbf{t}_{\rho^1}^{\rho^2} : \mathbb{R}^d \to \mathbb{R}^d$:

$$\boldsymbol{\rho} = (\mathbf{i} \times \mathbf{t}_{\rho^1}^{\rho^2})_{\#}\rho^1 \quad (\mathbf{i}(x) = x).$$

In particular $\mathbf{t}_{\rho^1}^{\rho^2}$ is the unique solution of Monge's optimal transport problem

$$\min\Big\{\int_{\mathbb{R}^d} \big|x - \mathbf{t}(x)\big|^2 \,\mathrm{d}\rho^1(x) : \mathbf{t}_{\#}\rho = \nu\Big\}.$$

2.3.3 Wasserstein distance: the one-dimensional case

Let $\rho^1, \rho^2 \in \mathcal{P}(\mathbb{R})$. Given a convex, even, and lower semicontinuous function $\psi : \mathbb{R} \to [0,+\infty]$, we can consider the cost $c_\psi(x,y) := \psi(x-y)$, $x, y \in \mathbb{R}$, and the associated optimal mass transportation problem

$$\mathcal{C}_\psi(\rho^1,\rho^2) = \inf\left\{\int_{\mathbb{R}\times\mathbb{R}} \psi(x-y)\,\mathrm{d}\boldsymbol{\rho}(x,y) : \boldsymbol{\rho} \text{ coupling between } \rho^1 \text{ and } \rho^2\right\}. \tag{2.31}$$

In the one-dimensional case, there exists a unique optimal coupling $\boldsymbol{\rho}$ realizing the minimum of (2.30) and (2.31) (at least when the cost is finite and uniformly convex): it can be explicitly characterized by inverting the distribution functions of ρ^1, ρ^2. The cumulative distribution function M_ρ associated to a probability measure $\rho \in \mathcal{P}(\mathbb{R})$ is defined by

$$M_\rho(x) := \rho((-\infty, x]) \quad \forall x \in \mathbb{R}, \quad \text{so that } \rho = \partial_x M_\rho \quad \text{in } \mathscr{D}'(\mathbb{R}); \tag{2.32}$$

we consider its monotone rearrangement X_ρ, which is the pseudoinverse of the distribution function M_ρ

$$X_\rho(w) := \inf \{x : M_\rho(x) > w\} = \inf \{x : \rho((-\infty, x]) > w\}, \quad w \in (0,1). \tag{2.33}$$

X_ρ is a right-continuous and nondecreasing function satisfying

$$(X_\rho)_\# \lambda = \rho, \quad \lambda := \mathscr{L}^1|_{(0,1)}, \qquad \int_{\mathbb{R}} \zeta(x)\,\mathrm{d}\rho(x) = \int_0^1 \zeta(X_\rho(w))\,\mathrm{d}w$$

for every nonnegative Borel map $\zeta : \mathbb{R} \to [0,+\infty]$. In particular, $\rho \in \mathcal{P}_p(\mathbb{R})$ if and only if $X_\rho \in L^p(0,1)$. Moreover, thanks to the Hoeffding–Fréchet theorem [42, sect. 3.1], the joint map $X_{\rho^1,\rho^2}(w) := (X_{\rho^1}(w), X_{\rho^2}(w))$, $w \in (0,1)$ characterizes the optimal coupling $\boldsymbol{\rho} = \Gamma_o(\rho^1,\rho^2)$ by the formula

$$\boldsymbol{\rho} = \left(X_{\rho^1,\rho^2}\right)_\# \lambda$$

so that [24, 42, 51]

$$\begin{aligned} W_p^p(\rho^1,\rho^2) &= \int_0^1 \left|X_{\rho^1}(w) - X_{\rho^2}(w)\right|^p \mathrm{d}w, \\ \mathcal{C}_\psi(\rho^1,\rho^2) &= \int_0^1 \psi\left(X_{\rho^1}(w) - X_{\rho^2}(w)\right)\,\mathrm{d}w, \end{aligned} \tag{2.34}$$

and the map $\rho \in \mathcal{P}(\mathbb{R}) \longmapsto X_\rho$ is an isometry between $\mathcal{P}_2(\mathbb{R})$ and the convex subset $\mathcal{K}$ of $L^2(0,1)$ of (essentially) nondecreasing functions (which can be identified with their right-continuous representatives).

2.4 Absolutely continuous curves in $\mathcal{P}_2(\mathbb{R}^d)$

Definition 2.20 (Absolutely continuous curve)**.** Let $I \subset \mathbb{R}$ be an interval and let $\rho_t : I \to \mathcal{P}_2(\mathbb{R}^d)$. We say that ρ_t is absolutely continuous if there exists $m \in L^1(I)$ such that

$$W_2(\rho_t,\rho_s) \le \int_s^t m(\tau)\mathrm{d}\tau, \quad \forall s,t \in I, s \le t.$$

Any absolutely continuous curve is uniformly continuous; it is not difficult to show (see [2], [4]) that the *metric derivative*

$$|\rho'|(t) := \lim_{h\to 0} \frac{W_2(\rho(t+h),\rho(t))}{|h|}$$

exists $\mathscr{L}^1$-a.e. $t \in I$ for any absolutely continuous curve ρ_t. Furthermore, $|\rho'| \in L^1(I)$ and is the minimal m fulfilling the definition.

Theorem 2.21. *Let I be an open interval in $\mathbb{R}$, let $\rho_t : I \to \mathcal{P}_2(\mathbb{R}^d)$ be an absolutely continuous curve and let $|\rho'| \in L^1(I)$ be its metric derivative. Then there exists a Borel vector field $\mathbf{v} : (x,t) \mapsto \mathbf{v}_t(x)$ such that*

$$\mathbf{v}_t \in L^2(\mathbb{R}^d,\rho_t), \quad \|\mathbf{v}_t\|_{L^2(\mathbb{R}^d,\rho_t)} \le |\rho'|(t) \text{ for } \mathscr{L}^1\text{-a.e.} t \in I, \tag{2.35}$$

and the continuity equation

$$\partial_t \rho_t + \nabla \cdot (\mathbf{v}_t \rho) = 0 \text{ in } \mathbb{R}^d \times I$$

holds in the sense of distributions. i.e.

$$\int_I \int_{\mathbb{R}^d} (\partial_t \varphi(x,t) + \langle \mathbf{v}_t(x), \nabla_x \varphi(x,t) \rangle) \mathrm{d}\rho_t(x) \mathrm{d}t = 0 \forall \varphi \in C_c^\infty(\mathbb{R}^d \times I).$$

Moreover, for $\mathscr{L}^1$-a.e. $t \in I$ $\mathbf{v}_t$ belongs to the closure in $L^2(\mathbb{R}^d, \rho_t)$ of the subspace generated by the gradients $\nabla\varphi$ with $\varphi \in C_c^\infty(\mathbb{R}^d)$.

Conversely, if a narrowly continuous curve $\rho_t : I \to \mathcal{P}_2(\mathbb{R}^d)$ satisfies the continuity equation for some Borel velocity field $\mathbf{w}_t$ with $\|\mathbf{w}_t\|_{L^2(\mathbb{R}^d,\rho_t)} \in L^1(I)$ then ρ_t is absolutely continuous and $|\rho'|(t) \leq \|\mathbf{w}_t\|_{L^2(\mathbb{R}^d,\rho_t)}$.

In particular equality holds in (2.35).

Notice that the continuity equation involves only the action of $\mathbf{v}_t$ on $\nabla\varphi$ with $\varphi \in C_c^\infty(\mathbb{R}^d)$. Moreover, the previous Theorem shows that the minimal norm among all possible vector fields $\mathbf{w}_t$ is the metric derivative and $\mathbf{v}_t$ belongs to the L^2 closure of gradients of functions in $C_c^\infty(\mathbb{R}^d)$. These facts suggest a "canonical" choice of $\mathbf{v}_t$ and the following definition of tangent bundle to $\mathcal{P}_2(\mathbb{R}^d)$.

Definition 2.22 (Tangent bundle)**.** Let $\rho \in \mathcal{P}_2(\mathbb{R}^d)$. We define

$$\mathrm{Tan}_\rho \mathcal{P}_2(\mathbb{R}^d) = \overline{\{\nabla\varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mathbb{R}^d,\rho)}.$$

Then $\mathbf{v}_t \in \mathrm{Tan}_{\rho_t} \mathcal{P}_2(\mathbb{R}^d)$. It could be shown that, in $\mathrm{Tan}_{\rho_t} \mathcal{P}_2(\mathbb{R}^d)$, $\mathbf{v}_t$ is uniquely determined: $\mathbf{v}_t$ is called the *tangent vector* to the curve ρ_t.

Remark 2.23. If $\rho \in \mathcal{P}_2^r(\mathbb{R}^d)$, the subset of $\mathcal{P}_2(\mathbb{R}^d)$ of absolutely continuous measures, we can characterized $\mathrm{Tan}_\rho \mathcal{P}_2(\mathbb{R}^d)$ in a different way:

$$\mathrm{Tan}_\rho \mathcal{P}_2(\mathbb{R}^d) = \overline{\{\lambda(\mathbf{t}_\rho^\sigma) - \mathbf{i} : \sigma \in \mathcal{P}(\mathbb{R}^d), \lambda > 0\}}^{L^2(\mathbb{R}^d,\rho)},$$

where $\mathbf{t}_\rho^\sigma$ is the optimal transport map from ρ to σ. Furthermore, if the curve ρ_t belongs to $\mathcal{P}_2^r(\mathbb{R}^d)$, the following Proposition holds:

Proposition 2.24. *Let $\rho_t : I \to \mathcal{P}_2^r(\mathbb{R}^d)$ be an absolutely continuous curve, let $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ and let $\mathbf{v}_t \in \mathrm{Tan}_{\rho_t} \mathcal{P}_2(\mathbb{R}^d)$ be its tangent vector field. Then*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\rho_t, \sigma) = \int_{\mathbb{R}^d} \langle \mathbf{v}_t, \mathbf{i} - \mathbf{t}_{\rho_t}^\sigma \rangle \mathrm{d}\rho_t.$$

2.5 Gradient flows in $\mathcal{P}_2(\mathbb{R}^d)$

We refer to [2] for a deep treatment of th topic. Here we only recall some properties useful in the dissertation and enlighten analogies with the Hilbert gradient flows.

Let us consider a lower semicontinuous functional $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ with a proper domain

$$D(\Phi) := \{\rho \in \mathcal{P}_2(\mathbb{R}^d) : \Phi(\rho) < +\infty\} \neq \emptyset.$$

To introduce the gradient flow of Φ we need two basic definition: metric slope and subdifferential. The definition of metric slope has been introduced by De Giorgi for functionals defined in arbitrary metric space.

Definition 2.25 (Metric slope)**.** Let us consider a functional $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ and a measure $\rho \in D(\Phi)$. The metric slope of Φ at ρ is defined by

$$|\partial\Phi|(\rho) = \limsup_{\nu \to \rho} \frac{(\Phi(\rho) - \Phi(\nu))^+}{W_2(\rho, \nu)}.$$

For the subdifferential, we consider only the case of absolutely continuous measures.

Definition 2.26 (Frechet subdifferential)**.** Let us consider a functional $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ and a measure $\rho \in D(\Phi) \cap \mathcal{P}_2^r(\mathbb{R}^d)$. We say that $\boldsymbol{\xi} \in L^2(\mathbb{R}^d, \rho)$ belongs to the Frechet subdifferential $\partial\Phi(\rho)$ if

$$\Phi(\nu) - \Phi(\rho) \geq \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}(x), \mathbf{t}_\rho^\nu(x) - x \rangle \mathrm{d}\rho(x) + o(W_2(\rho, \nu)).$$

When $\boldsymbol{\xi} \in \partial\Phi(\rho)$ also satisfies

$$\Phi(\mathbf{t}_{\#}\rho) - \Phi(\rho) \geq \int_{\mathbb{R}^d} \langle \boldsymbol{\xi}(x), \mathbf{t}(x) - x \rangle \mathrm{d}\rho(x) + o(\|\mathbf{t} - \mathbf{i}\|_{L^2(\mathbb{R}^d, \rho)}),$$

$\boldsymbol{\xi}$ is said a *strong* subdifferential.

We can give the definition

Definition 2.27 (Gradient flows in $\mathcal{P}_2(\mathbb{R}^d)$)**.** We say that a (locally) absolutely continuous curve $\rho_t : I_{\subset\mathbb{R}} \to \mathbb{R}^d$ is a solution of the gradient flow equation

$$\mathbf{v}_t \in -\partial\Phi(\rho_t), \quad t > 0.$$

if, for $\mathscr{L}^1$-a.e. $t > 0$, $\rho_t \in \mathcal{P}_2^r(\mathbb{R}^d)$ and the opposite of its velocity field $\mathbf{v}_t \in \operatorname{Tan}_{\rho_t} \mathcal{P}_2(\mathbb{R}^d)$ belongs to the subdifferential of Φ at ρ_t.

In [2] is deeply examined the case of λ(-geodesically)convex functionals:

Definition 2.28 (λ-convexity along geodesics)**.** Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$. Given $\lambda \in \mathbb{R}$, we say that ϕ is λ-geodesically convex in $\mathcal{P}_2(\mathbb{R}^d)$ if for every couple $\rho^1, \rho^2 \in \mathcal{P}_2(\mathbb{R}^d)$ there exists an optimal coupling $\boldsymbol{\rho}$ between ρ^1, ρ^2 such that

$$\phi(\rho_t^{1\to2}) \leq (1-t)\phi(\rho^1) + t\phi(\rho^2) - \frac{\lambda}{2} t(1-t) W_2^2(\rho^1, \rho^2), \quad t \in [0,1],$$

where $\rho_t^{1\to2} = \big((1-t)\pi^1 + t\pi^2\big)_{\#}\boldsymbol{\rho}$, π^1, π^2 being the projections onto the first and the second coordinate in $\mathbb{R}^d \times \mathbb{R}^d$, respectively.

Example 2.29 (λ-convex functionals)**.** In this thesis, we focused our attention on two functionals

- the opposite Wasserstein distance with respect to a fixed measure ρ^0,

$$\phi^{\rho^0}(\rho) := -\frac{1}{2} W_2^2(\rho, \rho^0);$$

- the entropy functional $\mathscr{H} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$

$$\mathscr{H}[\rho] = \begin{cases} \int_{\mathbb{R}^d} u \log u dx \text{ if } \rho = u\mathscr{L}^d \in \mathcal{P}_2^r(\mathbb{R}^d), \\ +\infty \text{ otherwise.} \end{cases}$$

One can check that $\mathscr{H}$ is (0-)convex and $\phi^{\rho^0}(\rho)$ is -1-convex. Morever, $\mathscr{H}$ is lower semicontinuous even w.r.t. the narrow topology, while $\phi^{\rho^0}(\rho)$ is obviously continuous in $\mathcal{P}_2(\mathbb{R}^d)$.

In [2, Th.11.2.1], one can find main properties and characterizations of the gradient flows of λ-convex functionals. We are interested in some of them, which we collect in the following Theorem

Theorem 2.30. *Let $\Phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous λ-convex functional. If $\rho_t^i : (0, +\infty) \to \mathcal{P}_2(\mathbb{R}^d)$,$i = 1, 2$ are gradient flows satisfying $\rho_t^i \to \rho^i$ as $t \downarrow 0$ in $\mathcal{P}_2(\mathbb{R}^d)$, then*

$$W_2(\rho_t^1, \rho_t^2) \leq e^{-\lambda t} W_2(\rho^1, \rho^2) \quad \forall t > 0. \tag{2.36}$$

In particular, for any $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ there is at most one gradient flow ρ_t satisfying the initial Cauchy condition $\rho_t \to \rho$ as $t \downarrow 0$ and it is also characterized by the system of "Evolution Variational Inequalities"

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\rho_t, \sigma) + \frac{\lambda}{2} W_2^2(\rho_t, \sigma) \leq \Phi(\sigma) - \Phi(\rho_t) \text{ for } \mathscr{L}^1\text{-a.e. } t > 0, \forall \sigma \in \mathcal{P}_2(\mathbb{R}^d). \tag{2.37}$$

The map $t \mapsto \Phi(\rho_t)$ is not increasing, Lipschitz continuous on $]0, +\infty[$ and

$$\frac{\mathrm{d}^+}{\mathrm{d}t} \Phi(\rho_t) = -|\rho'|^2(t) = -|\partial\Phi|^2(\rho_t) = -|\partial^\circ \Phi(\rho_t)|^2 \tag{2.38}$$

where $\partial^\circ \Phi(\rho_t)$ is the element of minimal norm of the subgradient $\partial\Phi(\rho_t)$.

Remark 2.31. Compare Theorem 2.30 with Theorems 2.10 and 2.15: for $\lambda = 0$, (2.36) and (2.37) are analogous to (2.16) and (2.21), while (2.38) recalls (2.10).

Example 2.32. (Subdifferential calculus of the entropy). The entropy functional is a particular case of the general class of internal energy functionals, investigated in [2, Th.10.4.6].

As a consequence, we have

Theorem 2.33 (Slope and subdifferential of $\mathscr{H}$)**.** *Assume that $\mathscr{H}$ has finite slope at $\rho = u\mathscr{L}^d \in \mathcal{P}_2^r(\mathbb{R}^d)$. Then $u \in W^{1,1}(\mathbb{R}^d)$, $\nabla u = \mathbf{w} u$ for some function $\mathbf{w} \in L^2(\rho, \mathbb{R}^d)$ and*

$$\|\mathbf{w}\|_{L^2(\rho, \mathbb{R}^d)} = |\partial\mathscr{H}|(\rho) < +\infty.$$

Conversely, if $u \in W_{loc}^{1,1}(\mathbb{R}^d)$ and $\nabla u = \mathbf{w} u$ for some function $\mathbf{w} \in L^2(\rho, \mathbb{R}^d)$, then $\mathscr{H}$ has finite slope at $\rho = u\mathscr{L}^d \in \mathcal{P}_2^r(\mathbb{R}^d)$ and $\mathbf{w} = \partial^\circ \mathscr{H}(\rho)$.

Example 2.34 (Heat equation as gradient flow)**.** As shown in [31], ρ_t is solution of the gradient flow of $\mathscr{H}$ if and only if its density u_t ($\rho_t = u_t \mathscr{L}^d$) solves the heat equation. In particular, we have

$$W_2(\rho_t^1, \rho_t^2) \leq W_2(\rho^1, \rho^2) \quad \forall t > 0;$$

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\rho_t, \sigma) \leq \mathscr{H}(\sigma) - \mathscr{H}(\rho_t) \text{ for } \mathscr{L}^1\text{-a.e. } t > 0, \forall \sigma \in \mathcal{P}_2(\mathbb{R}^d);$$

$$t \mapsto \mathscr{H}(\rho_t) \text{ is not increasing.}$$

Chapter 3

Sticky particles

The sticky particles system (SPS) describes the behavior of a finite collection of particles, freely moving in the absence of forces and sticking under collision. They can be mathematically represented by a time-dependent discrete measure $\rho_t^N := \sum_{i=1}^n m_i \delta_{x_i(t)}$ concentrated in a finite set of N particles $P_i(t) := (m_i, x_i(t), v_i(t))$, $i = 1, \dots, N$ with positive mass m_i, ordered positions $x_1(t) \le x_2(t) \le \cdots \le x_{N-1}(t) \le x_N(t)$, and velocities $v_i(t)$.

Denoting by $J_i(t) := \{j : x_j(t) = x_i(t)\}$ the collection of (the indices of) the particles $P_j(t)$ coinciding with $P_i(t)$ at time t, the adhesion dynamic imposes that the sets $J_i(t)$ are nondecreasing in time, so that $v_j(t+) = v_i(t+)$ for every $j \in J_i(t)$. We can thus order in a finite and monotone sequence $0 < t_1 < t_2 < \cdots$ the collection of times when the cardinality of some $J_i(t)$ has a discontinuity (corresponding to some collision). In each open interval $[t_k, t_{k+1})$ the (right-continuous) velocities $v_i(t) = \dot{x}_i(t)$ are thus supposed to be constant and, at each collision time t_k, the conservation of mass and momentum yields the updated equation for the velocities

$$v_i(t_k+) = \frac{\sum_{j \in J_i(t_k)} m_j v_j(t_k-)}{\sum_{j \in J_i(t_k)} m_j}, \qquad i = 1, \dots N. \tag{3.1}$$

It is not difficult to check that the measures ρ^N and $(\rho v)_t^N := \sum_{i=1}^N m_i v_i(t)\, \delta_{x_i(t)}$ solve the one-dimensional pressureless Euler system

$$\begin{cases} \partial_t \rho + \partial_x(\rho\, v) = 0 \\ \partial_t(\rho\, v) + \partial_x(\rho\, v^2) = 0 \end{cases} \quad \text{in } \mathbb{R} \times (0, +\infty), \quad \rho|_{t=0} = \rho_0, \quad v|_{t=0} = v_0. \tag{3.2}$$

Starting from SPS we are able to construct an evolution semigroup in a measure-theoretic phase space, allowing mass distributions in $\mathcal{P}_2(\mathbb{R})$ and corresponding L^2-velocity fields. In the study of this semigroup an important role is played by his Lagrangian formulation: associating probability measures with their monotone rearrangements, we can employ the isometric correspondence $\mathcal{P}_2(\mathbb{R})$ and the convex cone $\mathcal{K}$ of nondecreasing functions in $L^2(0,1)$ (see §2.3.3).

In the next section we recall some basic definition and notation and we state our main results. Section 3.2 collects the main properties related to the convex cone $\mathcal{K}$ in $L^2(0,1)$ (projection, polar cone, subdifferential of the indicator function)—they provide simple but crucial tools for the analysis of the discrete

SPS presented in section 3.3, which contains all the basic calculations. Section 3.4 deals with the existence, stability, and uniqueness of the solution in the Lagrangian formulation. The final steps of the proofs (mainly concerning the various limit processes) will be detailed in the last section, where we also show a new derivation of the Brenier–Grenier theorem [18] from the Lagrangian representation of the SPS.

3.1 Main results

An explicit estimate through Wasserstein distance

We introduce the set

$$\mathcal{V}_p(\mathbb{R}) := \left\{\mu = (\rho, \rho v) \in \mathcal{P}_p(\mathbb{R}) \times \mathcal{M}(\mathbb{R}) : v \in L^p_\rho(\mathbb{R})\right\}, \quad p \in [1, +\infty),$$

where $\mathcal{M}(\mathbb{R})$ is the set of all signed Borel measures with finite total variation, the semidistances (here $\mu^i = (\rho^i, \rho^i v^i)$)

$$U^p_p(\mu^1, \mu^2) := \int_{\mathbb{R}\times\mathbb{R}} |v^1(x) - v^2(y)|^p \,\mathrm{d}\boldsymbol{\rho}(x, y) \quad \boldsymbol{\rho} = \Gamma_o(\rho^1, \rho^2) \tag{3.3}$$

$$= \int_0^1 |v^1(X_{\rho^1}(w)) - v^2(X_{\rho^2}(w))|^p \,\mathrm{d}w, \tag{3.4}$$

and the distances

$$D^p_p(\mu^1, \mu^2) := W^p_p(\rho^1, \rho^2) + U^p_p(\mu^1, \mu^2). \tag{3.5}$$

We also set

$$[\mu]^p_p := \int_{\mathbb{R}} (|x|^p + |v(x)|^p) \,\mathrm{d}\rho(x) = D^p_p(\mu, (\delta_0, 0)).$$

Proposition 3.1. *D_p is a distance in $\mathcal{V}_p(\mathbb{R})$ and $(\mathcal{V}_p(\mathbb{R}), D_p)$ is a metric (but not complete) space whose topology is stronger than the one induced by the weak convergence of measures. The collection of discrete measures*

$$\hat{\mathcal{V}}(\mathbb{R}) := \left\{\mu = \left(\sum_{i=1}^N m_i\delta_{x_i}, \sum_{i=1}^N m_i v_i \delta_{x_i}\right) : m_i > 0,\ \sum_{i=1}^N m_i = 1,\ x_i, v_i \in \mathbb{R}\right\}$$

is a dense subset of $\mathcal{V}_p(\mathbb{R})$. A sequence $\mu_n = (\rho_n, \rho_n v_n)$, $n \in \mathbb{N}$ converges to $\mu = (\rho, \rho v)$ in $\mathcal{V}_p(\mathbb{R})$, $p > 1$ if and only if (see [2, Def. 5.4.3]*)*

$$W_p(\rho_n, \rho) \to 0, \quad \rho_n v_n \rightharpoonup \rho v \quad \textit{weakly in } \mathcal{M}(\mathbb{R}), \quad \int_{\mathbb{R}} |v_n|^p \,d\rho_n \to \int_{\mathbb{R}} |v|^p \,\mathrm{d}\rho. \tag{3.6}$$

Let us denote by $\mathscr{S}_t : \hat{\mathcal{V}}(\mathbb{R}) \to \hat{\mathcal{V}}(\mathbb{R})$ the map associating to any discrete initial datum $(\rho_0, \rho_0 v_0)$ the solution $(\rho_t, \rho_t v_t)$ of the (discrete) SPS. $\mathscr{S}_t$ is a semigroup in $\hat{\mathcal{V}}(\mathbb{R})$.

Theorem 3.2 (Stability w.r.t. the initial data)**.** *Let $\mu_t^\ell = (\rho_t^\ell, \rho_t^\ell v_t^\ell) = \mathscr{S}_t[\mu_0^\ell]$, $\ell = 1, 2$ be the solutions of the (discrete) SPS with initial data $\mu_0^\ell \in \hat{\mathcal{V}}(\mathbb{R})$. Then, for every convex cost* (2.31) *and every $p \geq 1$,*

$$\mathcal{C}_\psi(\rho_t^1, \rho_t^2) \leq \int_{\mathbb{R}\times\mathbb{R}} \psi\left(x + tv^1(x) - (y + tv^2(y)\right) \mathrm{d}\boldsymbol{\rho}(x, y), \quad \boldsymbol{\rho} = \Gamma_o(\rho^1, \rho^2), \tag{3.7a}$$

$$W_p(\rho_t^1, \rho_t^2) \leq W_p(\rho_0^1, \rho_0^2) + tU_p(\mu_0^1, \mu_0^2), \tag{3.7b}$$

$$\int_0^t U_2^2(\mu_r^1, \mu_r^2)\,\mathrm{d}r \leq C(1 + t)\left([\mu^1]_2 + [\mu^2]_2\right)\left(W_2(\rho_0^1, \rho_0^2) + U_2(\mu_0^1, \mu_0^2)\right) \tag{3.7c}$$

for a suitable "universal" constant C independent of t and the data.

We say that a map $\mathscr{S} : \mathcal{V}_p(\mathbb{R}) \to \mathcal{V}_p(\mathbb{R})$ is *strongly-weakly continuous* if, for every $\mu^n, \mu \in \mathcal{V}_p(\mathbb{R})$ with $\mathscr{S}[\mu^n] = (\tilde{\rho}^n, \tilde{\rho}^n\tilde{v}^n)$, $\mathscr{S}[\mu] = (\tilde{\rho}, \tilde{\rho}\tilde{v}) \in \mathcal{V}_p(\mathbb{R})$,

$$\lim_{n\uparrow+\infty} D_p(\mu_n, \mu) = 0 \implies \lim_{n\uparrow+\infty} W_p(\tilde{\rho}^n, \tilde{\rho}) = 0, \quad \tilde{\rho}_n\tilde{v}_n \rightharpoonup \tilde{\rho}\tilde{v} \text{ weakly in } \mathcal{M}(\mathbb{R}). \tag{3.8}$$

Theorem 3.3 (the evolution semigroup in $\mathcal{V}_p(\mathbb{R})$)**.** (a) *The semigroup $\mathscr{S}_t$ can be uniquely extended by density to a right-continuous semigroup (still denoted $\mathscr{S}_t$) of strongly-weakly continuous transformations in $\mathcal{V}_p(\mathbb{R})$, $p \geq 2$, thus satisfying*

$$\mathscr{S}_{s+t}[\mu] = \mathscr{S}_s[\mathscr{S}_t[\mu]] \quad \forall\, s, t \geq 0, \qquad \lim_{t\downarrow 0} D_p(\mathscr{S}_t[\mu], \mu) = 0 \quad \forall\, \mu \in \mathcal{V}_p(\mathbb{R}). \tag{3.9}$$

$\mathscr{S}_t$ complies with the same estimates (3.7a)–(3.7c) *of Theorem* 3.2.

(b) $(\rho_t, \rho_t v_t) = \mathscr{S}_t[\mu]$, $\mu \in \mathcal{V}_2(\mathbb{R})$ *is a distributional solution of* (3.2) *satisfying the Oleinik entropy condition*

$$v_t(x_2) - v_t(x_1) \leq \frac{1}{t}(x_2 - x_1) \quad \textit{for } \rho_t\textit{-a.e. } x_1, x_2 \in \mathbb{R},\ x_1 \leq x_2. \tag{3.10}$$

(c) *If $\psi : \mathbb{R} \to \mathbb{R}$ is a convex function such that $\psi(v_0) \in L^1_{\rho_0}(\mathbb{R})$ and $(\rho_t, \rho_t v_t) = \mathscr{S}_t[\mu_0]$, then the map*

$$t \mapsto \int_{\mathbb{R}} \psi(v_t)\,\mathrm{d}\rho_t(x) \tag{3.11}$$

is nonincreasing in $[0, +\infty)$, and its jump set is contained in an at most countable set $\mathcal{T} = \mathcal{T}(\mu)$ independent of ψ.

(d) *If $\mu \in \mathcal{V}_p(\mathbb{R})$ and $\mu_t = (\rho_t, \rho_t v_t) = \mathscr{S}_t[\mu]$, $t \in [0, +\infty)$, the curve $t \mapsto \rho_t$ is Lipschitz in $\mathcal{P}_p(\mathbb{R})$ with respect to W_p, and the curve $t \mapsto \rho_t v_t$ is continuous with respect to the weak topology in $\mathcal{M}(\mathbb{R})$, right-continuous in $[0, +\infty)$ with respect to the (semi-) distance U_p, and left-continuous at each $t \in (0, +\infty) \setminus \mathcal{T}$, where $\mathcal{T}$ is the at most countable jump set of* (3.11).

(e) *Let $\mu_t^n = (\rho_t^n, \rho_t^n v_t^n) = \mathscr{S}_t[\mu^n]$ and $\mu_t = (\rho_t, \rho_t v_t) = \mathscr{S}_t[\mu]$; if μ^n converges to μ in $\mathcal{V}_p(\mathbb{R})$ as $n \uparrow +\infty$, then for every $t \in [0, +\infty)$ ρ_t^n converges to ρ_t in $\mathcal{P}_p(\mathbb{R})$ and $\rho_t^n v_t^n$ weakly converges to $\rho_t v_t$ in $\mathcal{M}(\mathbb{R})$; moreover, μ_t^n converges to $(\rho_t, \rho_t v_t) = \mathscr{S}_t[\mu]$ in $\mathcal{V}_p(\mathbb{R})$ for every $t \in [0, +\infty) \setminus \mathcal{T}(\mu)$.*

(f) *For every* $0 \le s < t$, *there exists a* ρ_s*-essentially unique monotone map* $\mathsf{x}_s^t \in L^2_{\rho_s}(\mathbb{R})$ *such that*

$$\rho_t = (\mathsf{x}_s^t)_{\#}\rho_s, \quad \lim_{h\downarrow 0} \frac{\mathsf{x}_s^{s+h} - \mathsf{i}}{h} = v_s \quad in\ L^2_{\rho_s}(\mathbb{R}), \quad \mathsf{i}(x) \equiv x, \tag{3.12}$$

and, for ρ_t*-a.e.* $y \in \mathbb{R}$,

$$v_t(y) = \int_{\mathbb{R}} v_s(x)\,\mathrm{d}\rho_y^{s\to t}(x) = (t-s)^{-1}\left(y - \int_{\mathbb{R}} \mathsf{x}_s^t(x)\,\mathrm{d}\rho_y^{s\to t}(x)\right) \tag{3.13}$$

where $\rho_y^{s\to t}$ *is the disintegration of* ρ_s *with respect to* x_s^t.

Let us recall that the disintegration $\rho_y^{s\to t}$ of ρ_s with respect to the Borel (monotone) map x_s^t is a Borel family of parametrized measures uniquely determined for ρ_t-a.e. $y \in \mathbb{R}$ such that $\rho_s = \int_{\mathbb{R}} \rho_y^{s\to t}\,\mathrm{d}\rho_t(y)$ with $\rho_y^{s\to t}((\mathsf{x}_s^t)^{-1}(y)) = 1$ (see, e.g., [2, Thm. 5.3.1]).

Note that for a fixed t the map $\mathscr{S}_t : \mathcal{V}_p(\mathbb{R}) \to \mathcal{V}_p(\mathbb{R})$ may fail to be continuous with respect to the distance D_p, at least in the momentum component ρv.

The gradient flow of the (opposite) squared Wasserstein distance

Equations (3.12) and (3.13) show an interesting connection between the semigroup $\mathscr{S}_t$ in $\mathcal{V}_2(\mathbb{R})$ and the gradient flow $\mathscr{G}_t^\sigma$ in $\mathcal{P}_2(\mathbb{R})$ of the (opposite) squared distance functional

$$\phi^\sigma(\rho) := -\frac{1}{2}W_2^2(\rho,\sigma) \quad \forall\, \rho,\sigma \in \mathcal{P}_2(\mathbb{R}).$$

Let us recall [2] that for every choice of a reference measure $\sigma \in \mathcal{P}_2(\mathbb{R})$ it is possible to define a unique continuous and 1-expansive semigroup $\mathscr{G}_\tau^\sigma : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$, $\tau \ge 0$ whose Lipschitz trajectories $\hat\rho_\tau := \mathscr{G}_\tau^\sigma(\rho)$ can be uniquely characterized by the evolution variational inequality

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\tau}W_2^2(\hat\rho_\tau,\eta) - \frac{1}{2}W_2^2(\hat\rho_\tau,\eta) \le \phi^\sigma(\eta) - \phi^\sigma(\hat\rho_\tau) \quad \forall\, \eta \in \mathcal{P}_2(\mathbb{R}). \tag{3.14}$$

The next result shows that $\mathscr{S}_t$ and $\mathscr{G}_\tau^{\rho_0}$ basically coincide up to the rescaling

$$\tau = \log t, \quad t = e^\tau, \quad \hat\rho_\tau = \rho_{e^\tau}. \tag{3.15}$$

Theorem 3.4. (Gradient flow of the Wasserstein distance and SPS). *Let* $(\rho_t, \rho_t v_t) = \mathscr{S}_t(\rho_0, \rho_0 v_0) \in \mathcal{V}_2(\mathbb{R})$ *be the semigroup solution of the SPS. The Lipschitz curve* $(\rho_t)_{t\ge 0}$ *in* $\mathcal{P}_2(\mathbb{R})$ *solves the evolution variational inequality*

$$\frac{t}{2}\frac{\mathrm{d}}{\mathrm{d}t}W_2^2(\rho_t,\eta) - \frac{1}{2}W_2^2(\rho_t,\eta) \le \phi^{\rho_0}(\eta) - \phi^{\rho_0}(\rho_t) \quad a.e.\ in\ (0,+\infty) \tag{3.16}$$

for every $\eta \in \mathcal{P}_2(\mathbb{R})$. *Equivalently, the reparametrized solutions* $\hat\rho_\tau = \rho_{e^\tau}$ *satisfy* (3.14) *with* $\sigma := \rho_0$ *and we thus get the representation formula*

$$\hat\rho_\tau = \mathscr{G}^{\rho_0}_{\tau-\delta}\hat\rho_\delta \quad or,\ equivalently, \quad \rho_t = \mathscr{G}^{\rho_0}_{\log(t/\varepsilon)}\rho_\varepsilon \quad \forall\, \tau = \log t \ge \delta = \log\varepsilon. \tag{3.17}$$

Conversely, if $t \mapsto \rho_t$ is a Lipschitz curve in $\mathcal{P}_2(\mathbb{R})$ satisfying (3.16) *and the initial velocity condition*

$$\lim_{t\downarrow 0} t^{-2} \int_{\mathbb{R}} |x + tv_0(x) - y|^2 \,\mathrm{d}\boldsymbol{\rho}_t(x,y) = 0 \quad \boldsymbol{\rho}_t = \Gamma_o(\rho_0, \rho_t), \tag{3.18}$$

then there exists a unique Borel velocity vector field $v_t \in L^2_{\rho_t}(\mathbb{R})$ such that $(\rho_t, \rho_t v_t) = \mathscr{S}_t(\rho_0, \rho_0 v_0)$. v_t is the Wasserstein velocity field of ρ_t [2, Theorem 8.4.5].

Note that (3.18) corresponds to (3.12) for $s = 0$ in the case (which a posteriori is always verified) $\boldsymbol{\rho}_t = (\mathrm{i} \times \mathsf{x}_0^t)_\# \rho_0$.

We can use (3.17) to exhibit the solution ρ_t of the SPS by a simple limit procedure.

Theorem 3.5. *Let $(\rho_t, \rho_t v_t) = \mathscr{S}_t(\rho_0, \rho_0 v_0) \in \mathcal{V}_2(\mathbb{R})$ be the solution of the SPS, and let $\tilde\rho_\varepsilon := (i + \varepsilon v_0)_\# \rho_0$, $\varepsilon > 0$. Then*

$$\rho_t = \lim_{\varepsilon\downarrow 0} \mathscr{G}^{\rho_0}_{\log(t/\varepsilon)}(\tilde\rho_\varepsilon) \quad \text{in } \mathcal{P}_2(\mathbb{R}). \tag{3.19}$$

Moreover, if for some $\varepsilon_0 > 0$ the map $i + \varepsilon_0 v_0$ is ρ_0-essentially nondecreasing, then

$$\rho_\varepsilon = \tilde\rho_\varepsilon, \quad \rho_t = \mathscr{G}^{\rho_0}_{\log(t/\varepsilon)}(\tilde\rho_\varepsilon) \quad \forall\, \varepsilon \in (0, \varepsilon_0],\ t \geq \varepsilon. \tag{3.20}$$

The evolution in Lagrangian coordinates

We conclude this section with an even more explicit formula for the evolution of the monotone rearrangement function $X(t) = X_{\rho_t}$. We denote by $I_\mathcal{K}$ the indicator (convex, lower semicontinuous) function of $\mathcal{K}$ in $L^2(0,1)$

$$I_\mathcal{K}(X) = \begin{cases} 0 & \text{if } X \in \mathcal{K}, \\ +\infty & \text{otherwise,} \end{cases} \tag{3.21}$$

with (multivalued) subdifferential $\partial I_\mathcal{K} : L^2(0,1) \to 2^{L^2(0,1)}$. We also introduce the closed subspace $\mathcal{H}_X \subset L^2(0,1)$, $X \in \mathcal{K}$, whose functions $Y \in L^2(0,1)$ are essentially constant in each open interval $(a,b) \subset (0,1)$, where X is constant. It is not difficult to check that for every $X \in \mathcal{K}$ and $Y \in L^2(0,1)$

$$Y \in \mathcal{H}_X \quad \text{iff} \quad Y = y \circ X \text{ for some Borel map } y \in L^2_\rho(\mathbb{R}),\ \rho = X_\# \lambda. \tag{3.22}$$

Theorem 3.6 (Lagrangian evolution)**.** *A curve $(\rho_t, \rho_t v_t) \in \mathcal{V}_2(\mathbb{R})$, $t \geq 0$ is the semigroup solution $\mathscr{S}_t(\rho_0, \rho_0 v_0)$ of the SPS as in Theorem* 3.3 *if and only if its monotone rearrangement $X(t) = X_{\rho_t} \in \mathcal{K} \subset L^2(0,1)$ satisfies one of the following three (equivalent) characterizations in terms of the couple $X_0 := X_{\rho_0}$ and $V_0 := v_0(X_0) \in \mathcal{H}_{X_0}$:*

I. *X is the unique strong (i.e., absolutely continuous) solution of the Cauchy problem for the subdifferential inclusion*

$$\frac{\mathrm{d}}{\mathrm{d}t} X \in -\partial I_\mathcal{K}(X) + V_0, \quad X(0) = X_0, \tag{L.I}$$

which is the gradient flow in $L^2(0,1)$ of the convex functional

$$X \mapsto I_\mathcal{K}(X) - (V_0|X), \quad X \in L^2(0,1). \tag{3.23}$$

II. *X admits the representation formula*

$$X(t) = \mathsf{P}_{\mathcal{K}}(X_0 + tV_0), \tag{L.II}$$

where $\mathsf{P}_{\mathcal{K}}$ *is the* L^2*-projection on the convex cone* $\mathcal{K} \subset L^2(0,1)$.

III. *X is the unique strong solution of the rescaled gradient flow*

$$\begin{aligned} &t\frac{\mathrm{d}}{\mathrm{d}t}X(t) \in -\partial I_{\mathcal{K}}(X(t)) + X(t) - X_0,\\ &\lim_{t\downarrow 0} t^{-1}(X(t) - X_0) = V_0 \quad \textit{in } L^2(0,1). \end{aligned} \tag{L.III}$$

In each of these cases the curve $t \mapsto X(t)$ *is Lipschitz continuous in* $L^2(0,1)$ *and right-differentiable at each time* t*; the velocity field* v_t *can be recovered by the formula*

$$V(t) = \frac{\mathrm{d}^+}{\mathrm{d}t}X(t) = v_t \circ X(t) = \mathsf{P}_{\mathcal{H}_{X(t)}}(V_0) \in \mathcal{H}_{X(t)} \quad \forall\, t \geq 0, \tag{L.a}$$

where $\mathsf{P}_{\mathcal{H}_X}$ *denotes the* L^2*-orthogonal projection on the closed subspace* $\mathcal{H}_X \subset L^2(0,1)$. *The closed subspaces* $\mathcal{H}_{X(t)}$ *are nonincreasing:*

$$\mathcal{H}_{X(t)} \subset \mathcal{H}_{X(s)} \quad \textit{if } 0 \leq s \leq t, \tag{L.b}$$

and X, V *satisfy the semigroup identities*

$$X(t) = \mathsf{P}_{\mathcal{K}}(X(s) + (t-s)V(s)), \quad V(t) = \mathsf{P}_{\mathcal{H}_{X(t)}}(V(s)) \quad \forall\, 0 \leq s \leq t. \tag{L.c}$$

This result shows that the natural evolution space for the Lagrangian sticky particles flow is

$$\mathscr{X}_p(0,1) := \{(X,V) \in L^p(0,1) \times L^p(0,1) : X \in \mathcal{K},\ V = v \circ X \in \mathcal{H}_X\}, \quad p \geq 2, \tag{3.24}$$

endowed with the product distance in $L^p(0,1) \times L^p(0,1)$. The bijective map

$$(\rho, \rho v) \in \mathcal{V}_p(\mathbb{R}) \longleftrightarrow (X,V) \in \mathscr{X}_p(0,1), \quad X = X_\rho,\ V = v \circ X_\rho$$

is, in fact, an isometry with respect to D_p of (3.5).

Corollary 3.7 (Lagrangian semigroup)**.** *For every* $p \geq 2$*, the time-dependent transformations* $\mathsf{S}_t : \mathscr{X}_p(0,1) \to \mathscr{X}_p(0,1)$*,* $t \geq 0$*, which map a couple* $(X_0, V_0) \in \mathscr{X}_p(0,1)$ *into the couple* $(X(t), V(t)) = \mathsf{S}_t(X_0, V_0) \in \mathscr{X}_p(0,1)$*, where* X *is the solution of (one of the equivalent)* (L.I), (L.II), (L.III) *and* $V = \frac{\mathrm{d}^+}{\mathrm{d}t}X$ *as in* (L.a)*, define a right-continuous semigroup in* $\mathscr{X}_p(0,1)$*, satisfying*

$$\begin{aligned} (X(t), V(t)) = \mathsf{S}_t(X_0, V_0) \quad &\Longleftrightarrow \quad (\rho_t, \rho_t v_t) = \mathscr{S}_t(\rho_0, \rho_0 v_0),\\ \textit{where} \qquad &\rho_t = (X(t))_\# \lambda, \quad V(t) = v_t \circ X(t). \end{aligned} \tag{3.25}$$

Remark 3.8 (rescaling)**.** Up to the rescaling $\tau = \log t$, $\hat{X}(\tau) = X(e^\tau)$, (L.III) is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\hat{X}(t) \in -\partial I_{\mathcal{K}}(\hat{X}(t)) + \hat{X}(t) - X_0. \tag{3.26}$$

We shall show (see Theorem 3.10) that $\mathsf{P}_{\mathcal{K}}$ is a contraction in every $L^p(0,1)$ so that (L.II) provides a simple and sharp way to estimate $X(t)$ in terms of the initial data corresponding to (3.7b). Applying a general result of [44, 45], one can obtain (3.7c) from the representation (L.I).

Let us finally remark that the Wasserstein gradient flow of Theorem 3.4 is equivalent to (L.III)–(3.26). It is sufficient to introduce the functional Φ^σ

$$\Phi^\sigma(X) := -\frac{1}{2}\|X - X_\sigma\|_{L^2(0,1)} + I_{\mathcal{K}}(X), \quad X \in L^2(0,1), \tag{3.27}$$

which is related to ϕ^σ by

$$\phi^\sigma(\rho) = \Phi^\sigma(X_\rho) \quad \forall\, \rho \in \mathcal{P}_2(\mathbb{R})$$

and is a smooth quadratic perturbation of the convex and lower semicontinuous indicator functional $I_{\mathcal{K}}$; since

$$\partial\Phi^\sigma(X) = \partial I_{\mathcal{K}}(X) - (X - X_\sigma), \tag{3.28}$$

(3.26) is the subdifferential formulation in $L^2(0,1)$ of the gradient flow of Φ^{ρ_0}, whose metric characterization [2] yields (3.14) thanks to the isometry $\rho \leftrightarrow X_\rho$ between $\mathcal{P}_2(\mathbb{R})$ and $\mathcal{K}$.

Remark 3.9 (minimal Lagrangian description)**.** One can use (as in [48, 5]) the initial measure $\rho_0 \in \mathcal{P}(\mathbb{R})$ as a reference for the Lagrangian evolution, thus representing ρ_t as $\mathsf{x}(t)_\#\rho_0$ for the optimal monotone map $\mathsf{x}(t) = \mathsf{x}_0^t \in L^2_{\rho_0}(\mathbb{R})$ according to Theorem 3.3(f). We can therefore introduce the convex set $\mathcal{K}(\rho_0)$ of essentially nonincreasing Borel maps in the Hilbert space $L^2_{\rho_0}(\mathbb{R})$ and we have the corresponding formulae for the evolution in $L^2_{\rho_0}(\mathbb{R})$ ($\mathsf{i} : \mathbb{R} \to \mathbb{R}$ denotes the identity map):

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{x}(t) \in -\partial I_{\mathcal{K}(\rho_0)}(\mathsf{x}(t)) + v_0, \quad \mathsf{x}(0) = \mathsf{i}, \tag{L.I$'$}$$

$$\mathsf{x}(t) = \mathsf{P}_{\mathcal{K}(\rho_0)}(\mathsf{i} + tv_0), \quad \mathsf{i}(x) = x, \tag{L.II$'$}$$

$$t\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{x}(t) \in -\partial I_{\mathcal{K}(\rho_0)}(\mathsf{x}(t)) + \mathsf{x}(t) - \mathsf{i}, \tag{L.III$'$}$$

to be completed with the expression for the velocity $\frac{\mathrm{d}^+}{\mathrm{d}t}\mathsf{x}(t) = \mathsf{v}(t) = v_t \circ \mathsf{x}(t)$. All these relations could be easily deduced by Theorem 3.6, since the correspondence $\mathsf{x} \leftrightarrow X = \mathsf{x} \circ X_0$ is an isometry between $L^2_{\rho_0}(\mathbb{R})$ and the closed subspace $\mathcal{H}_{X_0}$ of $L^2(0,1)$. On the other hand, it is easier to deal with the convex set $\mathcal{K}$ in the space $L^2(0,1)$ with the uniform Lebesgue measure as a reference. The description provided by Theorem 3.6 is more general, since it allows us to compare solutions arising from different initial data.

3.2 Main properties of $\mathcal{K}$

In this section we will study the properties of the convex set $\mathcal{K}$ of nondecreasing functions in $L^2(0,1)$, in particular the $L^2(0,1)$-projection operator $\mathsf{P}_{\mathcal{K}}$ and the subdifferential of the indicator function $I_{\mathcal{K}}$ (3.21). Denoting by $(\cdot|\cdot)$ (resp., $\|\cdot\|$)

the usual scalar product (resp., the induced norm) in $L^2(0,1)$, since $\mathcal{K}$ is a convex cone, $\mathsf{P}_\mathcal{K}$ can be characterized by

$$g = \mathsf{P}_\mathcal{K}(f) \iff g \in \mathcal{K}, \quad (f-g|z-g) \le 0 \quad \forall\, z \in \mathcal{K}, \tag{3.29}$$
$$\iff g \in \mathcal{K}, \quad (f-g|z) \le 0 \quad \forall\, z \in \mathcal{K}, \quad (f-g|g) = 0. \tag{3.30}$$

The next result provides a useful characterization of $\mathsf{P}_\mathcal{K}(f)$ in terms of the convex envelope of the primitive of f. Recall that the convex envelope of a given continuous function $F : [0,1] \to \mathbb{R}$ is defined as

$$F^{**}(w) := \sup\left\{a + bw : a, b \in \mathbb{R},\ a + bv \le F(v) \quad \forall\, v \in [0,1]\right\}, \quad w \in [0,1], \tag{3.31}$$

and it is the greatest bounded, (lower semi-) continuous, and convex function G satisfying $G \le F$ in $[0,1]$; it is therefore right- and left-differentiable at every point $t \in (0,1)$ and its right derivative $g := \frac{\mathrm{d}^+}{\mathrm{d}w} F^{**}$ is nondecreasing and right-continuous.

Theorem 3.10 (projection on $\mathcal{K}$)**.** *Let $f \in L^2(0,1)$ and let $F(w) = \int_0^w f(s)ds$ be its primitive. Then*

$$\mathsf{P}_\mathcal{K}(f) = g = \frac{\mathrm{d}^+}{\mathrm{d}w} F^{**},$$

where F^{**} is the convex envelope of F defined by (3.31)*. Moreover, for every convex lower semicontinuous function $\psi : \mathbb{R} \to (-\infty, +\infty]$ and every $f, h \in L^2(0,1)$, we have*

$$\int_\mathbb{R} \psi\left(\mathsf{P}_\mathcal{K}(f)\right) \mathrm{d}w \le \int_\mathbb{R} \psi(f)\,\mathrm{d}w, \quad \int_\mathbb{R} \psi\left(\mathsf{P}_\mathcal{K}(f) - \mathsf{P}_\mathcal{K}(h)\right) \mathrm{d}w \le \int_\mathbb{R} \psi(f-h)\,\mathrm{d}w. \tag{3.32}$$

In particular, $\mathsf{P}_\mathcal{K}$ is a contraction in every space $L^p(0,1)$, $p \in [1, +\infty]$:

$$\|\mathsf{P}_\mathcal{K}(f) - \mathsf{P}_\mathcal{K}(h)\|_{L^p(0,1)} \le \|f-g\|_{L^p(0,1)} \quad \forall\, f, h \in L^p(0,1). \tag{3.33}$$

We split the proof into several steps. Here is a preliminary lemma.

Lemma 3.11. *For every $f \in L^2(0,1)$, F^{**} is continuous in $[0,1]$, locally Lipschitz in $(0,1)$, and coincides with F at $w = 0$ and $w = 1$. If $f \in L^\infty(0,1)$, then F and F^{**} are Lipschitz continuous in the closed interval $[0,1]$.*

Proof. Let us first assume $f \in L^\infty(0,1)$, and let L be the Lipschitz constant of F; then

$$F(0) - Lw \le F(w), \quad F(1) + L(w-1) \le F(w) \quad \forall\, w \in [0,1]$$

so that $F^{**}(0) = F(0)$, $F^{**}(1) = F(1)$, and

$$F(0) - Lw \le F^{**}(w), \quad F(1) + L(w-1) \le F^{**}(w) \quad \forall\, w \in [0,1]. \tag{3.34}$$

Therefore the right derivative g of F^{**} satisfies

$$-L \le g(0) \le g(w) \le \frac{\mathrm{d}^-}{\mathrm{d}w} F^{**}(1) \le L$$

so that F^{**} is a Lipschitz function.

In the general case when $f \in L^2(0,1)$, we can approximate its (absolutely continuous) primitive F by an increasing sequence of Lipschitz functions F_n uniformly converging to F, e.g., by setting

$$F_n(w) = \inf_{v\in[0,1]} F(v) + n|v-w|.$$

Thus F_n^{**} is an increasing sequence of Lipschitz functions satisfying $F_n^{**}(w) = F_n(w)$ at $w = 0,1$ and pointwise converging to some lower semicontinuous convex function G as $n \uparrow +\infty$ with

$$G(w) \le F^{**}(w) \le F(w) \quad \forall\, w \in [0,1]. \tag{3.35}$$

On the other hand, for $w = 0,1$ we have $G(w) = \lim_{n\uparrow+\infty} F_n(w) = F(w)$ so that $F^{**}(w) = F(w)$. □

Let us now consider the set

$$\Lambda = \{w \in [0,1] : (F - F^{**})(w) > 0\}\,. \tag{3.36}$$

Since Λ is open and does not contain 0 and 1, it is the disjoint union of an (at most countable) collection $\mathcal{O}$ of open intervals.

Lemma 3.12. *If $(a,b) \in \mathcal{O}$ is a connected component of Λ, then for every $w = (1-\theta)a + \theta b$, $\theta \in [0,1]$*

$$\begin{aligned} F^{**}((1-\theta)a + \theta b) = (1-\theta)F(a) + \theta F(b) \quad \forall\, \theta \in [0,1], \\ F(a) = F^{**}(a),\ F(b) = F^{**}(b). \end{aligned} \tag{3.37}$$

Proof. Since $a, b \notin \Lambda$, one has $F(a) = F^{**}(a)$ and $F(b) = F^{**}(b)$. Let $\bar w \in [a,b]$ be a minimizer of the continuous function

$$w \mapsto F(w) - L(w), \quad L(w) := F(a) + (w-a)\frac{F(b)-F(a)}{b-a}$$

so that $F(w) \ge F(\bar w) + L(w) - L(\bar w)$ for every $w \in [a,b]$. The continuous function

$$G(w) := \begin{cases} F^{**}(w) & \text{if } w \notin [a,b], \\ \max\left(F^{**}(w), F(\bar w) + L(w) - L(\bar w)\right) & \text{if } w \in [a,b] \end{cases}$$

provides a convex lower bound of F and therefore $G(w) \le F^{**}(w)$ for every $w \in [0,1]$. Since $G(\bar w) = F(\bar w)$, we deduce that $\bar w \notin \Lambda$ and therefore $\bar w$ coincides with a or b and the inequality $G(w) \le F^{**}(w)$ yields $F^{**}((1-\theta)a + \theta b) \ge (1-\theta)F(a) + \theta F(b)$. The opposite inequality is a consequence of the convexity of F^{**}. □

The next lemma contains the crucial inequality we need to characterize $\mathsf{P}_{\mathcal{K}}$.

Lemma 3.13. *Let $\psi \in C^1(\mathbb{R})$ be a convex function. For every $f \in L^2(0,1)$ and $z \in \mathcal{K}$ with $g := (F^{**})'$, if $(f-g)\psi'(z-g) \in L^1(0,1)$, we have*

$$\begin{aligned} \int_0^1 (f(w) - g(w))\, \psi'\left(z(w) - g(w)\right) \mathrm{d}w \le 0 \le \\ \le \int_0^1 (f(w) - g(w))\, \psi'\left(g(w) - z(w)\right) \mathrm{d}w. \end{aligned} \tag{3.38}$$

Proof. We decompose $[0,1]$ in the disjoint union of the open intervals $(a,b) \in \mathcal{O}$ covering Λ (see (3.36)) and of $[0,1] \setminus \Lambda$; note that $F(w) = F^{**}(w)$ on $[0,1] \setminus \Lambda$, and so $f(w) = g(w)$ a.e. in $[0,1] \setminus \Lambda$ (recall that F^{**} is locally Lipschitz). In each $(a,b) \in \mathcal{O}$, F^{**} is linear, g is constant, and the function $w \mapsto \psi'(z(w) - g)$ is bounded and nondecreasing; thus, its distributional derivative is a nonnegative finite measure $\gamma_{a,b}$. Since $F = F^{**}$ in $\{a, b\}$, we have

$$\int_0^1 (f-g)\,\psi'(z-g)\,\mathrm{d}w = \int_\Lambda (f-g)\psi'(z-g)\,\mathrm{d}w + \int_{[0,1]\setminus\Lambda} (f-g)\psi'(z-g)\,\mathrm{d}w$$
$$= \sum_{(a,b)\in\mathcal{O}} \int_a^b (f-g)\psi'(z-g)\,\mathrm{d}w = - \sum_{(a,b)\in\mathcal{O}} \int_a^b (F(w) - F^{**}(w))\,d\gamma_{a,b}(w) \le 0.$$

The second inequality of (3.38) can be obtained simply by considering the convex function $\tilde\psi(r) := \psi(-r)$. ☐

End of the proof of Theorem 3.10. Concerning the projection in $L^2(0,1)$, (3.29) follows (3.38) by choosing $\psi(r) := \frac{1}{2}r^2$.

In order to prove (3.32), a standard approximation of ψ by the increasing sequence of its Moreau–Yosida approximations $\psi_n(r) := \min_{s\in\mathbb{R}} \psi(s) + \frac{n}{2}|s-r|^2$ shows that it is not restrictive to assume ψ convex, C^1, and at most quadratically growing as $|r| \to \infty$. We can then apply the standard convexity inequality $\psi(s) - \psi(r) \ge \psi'(r)(s-r)$ and Lemma 3.13, obtaining

$$\int_{\mathbb{R}} \left(\psi\left(f-h\right) - \psi\left(\mathsf{P}_{\mathcal{K}}(f) - \mathsf{P}_{\mathcal{K}}(h)\right)\right)\,\mathrm{d}w$$
$$\ge \int_{\mathbb{R}} \psi'\left(\mathsf{P}_{\mathcal{K}}(f) - \mathsf{P}_{\mathcal{K}}(h)\right)\left((f - \mathsf{P}_{\mathcal{K}}(f)) - (h - \mathsf{P}_{\mathcal{K}}(h))\right)\,\mathrm{d}w \overset{(3.38)}{\ge} 0.$$

The first inequality of (3.32) is a particular case of the second one, with $h = \mathsf{P}_{\mathcal{K}}(h) = 0$.

The following result is a simple consequence of Theorem 3.10. Let us first introduce for a given $f \in L^2(0,1)$ the open set $\Omega_f \subset (0,1)$, where f is locally constant

$$\Omega_f := \{w \in (0,1) : f \text{ is essentially constant in a neighborhood of } w\}\,. \quad (3.39)$$

Equivalently Ω_f is the complement of the support of the distributional derivative of f.

Corollary 3.14. *Let $f \in L^2(0,1)$ and $g = \mathsf{P}_{\mathcal{K}}(f)$. Then*

$$\Omega_f \subset \Omega_g. \quad (3.40)$$

Proof. Note that $\Lambda \subset \Omega_g$ (Λ has been defined by (3.36)); if $w \in \Omega_f \setminus \Lambda$, then $F(w) = F^{**}(w)$, so that any linear part of the graph of F in an open interval containing w should locally coincide with F^{**}; it follows that $F^{**} = F$ in a neighborhood of w so that $w \in \Omega_g$. ☐

Definition 3.15 (the polar cone and the subdifferential of $I_{\mathcal{K}}$)**.** We denote by $\mathcal{K}^\circ$ the polar cone of $\mathcal{K}$, defined by

$$f \in \mathcal{K}^\circ \quad \Longleftrightarrow \quad (f|z) \le 0 \quad \forall\, z \in \mathcal{K} \quad \Longleftrightarrow \quad \mathsf{P}_{\mathcal{K}}(f) = 0. \quad (3.41)$$

The subdifferential $\partial I_{\mathcal{K}}(g)$ of the indicator function of $\mathcal{K}$ (see (3.27)) at some function $g \in \mathcal{K}$ is the subset of $L^2(0,1)$ characterized by

$$\xi \in \partial I_{\mathcal{K}}(g) \quad \Longleftrightarrow \quad (\xi|z-g) \le 0 \quad \forall\, z \in \mathcal{K}. \tag{3.42}$$

Remark 3.16. $\mathcal{K}^\circ$ and $\partial I_{\mathcal{K}}$ are clearly linked by $\mathcal{K}^\circ = \partial I_{\mathcal{K}}(0)$ and

$$\xi \in \partial I_{\mathcal{K}}(g) \quad \Longleftrightarrow \quad \xi \in \mathcal{K}^\circ, \quad (\xi|g) = 0. \tag{3.43}$$

$\mathcal{K}^\circ$ provides an equivalent reformulation of (3.29), since

$$g = \mathsf{P}_{\mathcal{K}}(f) \iff g \in \mathcal{K}, \quad f-g \in \mathcal{K}^\circ, \ (f-g|g) = 0 \iff f-g \in \partial I_{\mathcal{K}}(g). \tag{3.44}$$

If Ω is an open subset of $(0,1)$, we denote by $\mathcal{N}_\Omega$ the convex cone

$$\mathcal{N}_\Omega := \big\{F \in C^0([0,1]) : F \ge 0 \text{ in } [0,1], \quad F = 0 \text{ in } [0,1]\setminus\Omega\big\}\,.$$

We can give a useful characterization of $\mathcal{K}^\circ$ in terms of the cone $\mathcal{N} := \mathcal{N}_{(0,1)}$.

Proposition 3.17 (a characterization of the polar cone $\mathcal{K}^\circ$)**.** *A function f belongs to the polar cone $\mathcal{K}^\circ$ if and only if its primitive $F(w) := \int_0^w f(s)\,\mathrm{d}s$ belongs to $\mathcal{N}$.*

Proof. If $F \in \mathcal{N}$, then one easily gets for every $z \in \mathcal{K} \cap C^1([0,1])$

$$(f|z) = \int_0^1 F'(w)\,z(w)\,\mathrm{d}w = -\int_0^1 F(w)\,z'(w)\,\mathrm{d}w \le 0,$$

since $F, z' \ge 0$, $F(0) = F(1) = 0$.

Let us now assume that $f \in \mathcal{K}^\circ$; for every continuous and nonnegative function $z \ge 0$ and $c \in \mathbb{R}$ we replace $Z(w) = \int_0^w z(s)\,\mathrm{d}s - c$. Since $Z \in \mathcal{K}$ we have

$$0 \ge (f|Z) = \int_0^1 f(w)\,Z(w)\,\mathrm{d}w = -\int_0^1 F(w)\,z(w)\,\mathrm{d}w + F(1)(Z(1)-c).$$

Since c, z are arbitrary, we conclude that $F \in \mathcal{N}$. □

The last result of this section concerns a precise characterization of $\partial I_{\mathcal{K}}$. Let us first define for $f \in L^2(0,1)$ the closed subspace $\mathcal{H}_f \subset L^2(0,1)$, defined as

$$\begin{aligned}\mathcal{H}_f := \big\{h \in L^2(0,1) : h &\text{ is essentially constant}\\ &\text{in each connected component of } \Omega_f\big\}.\end{aligned} \tag{3.45}$$

We denote by $\mathsf{P}_{\mathcal{H}_f}$ the orthogonal L^2-projection on $\mathcal{H}_f$. It is easy to check that

$$\begin{aligned}&\mathsf{P}_{\mathcal{H}_g}(f) = f \text{ a.e. in } (0,1)\setminus\Omega_g,\\ &\mathsf{P}_{\mathcal{H}_g}(f) \equiv ⨍_\alpha^\beta f(w)\,\mathrm{d}w \text{ a.e. in every connected component } (\alpha,\beta) \subset \Omega_g.\end{aligned} \tag{3.46}$$

Moreover, denoting by F the primitive function of f,

$$\text{if } F \in \mathcal{N}_{\Omega_g}, \text{ then } f \text{ is orthogonal to } \mathcal{H}_g, \tag{3.47}$$

since $f = F'$ vanishes a.e. outside Ω_g and for every connected component (α,β) of Ω_g we have $\int_\alpha^\beta f(w)\,\mathrm{d}w = F(\beta) - F(\alpha) = 0$.

Theorem 3.18. (The subdifferential of $I_{\mathcal{K}}$). *Let $g \in \mathcal{K}$, $\xi \in L^2(0,1)$, and $\Xi(w) := \int_0^w \xi(s)\,\mathrm{d}s$. Then we have*

$$\xi \in \partial I_{\mathcal{K}}(g) \quad \Longleftrightarrow \quad \Xi \in \mathcal{N}_{\Omega_g}. \tag{3.48}$$

In particular, if $\xi \in \partial I_{\mathcal{K}}(g)$, then

$$\begin{cases} \xi = 0 \ \text{a.e. in } [0,1] \setminus \Omega_g, \\ \displaystyle\int_\alpha^\beta \xi(w)\,\mathrm{d}w = 0 \quad \text{for every connected component } (\alpha,\beta) \text{ of } \Omega_g \end{cases} \tag{3.49}$$

so that ξ is orthogonal to $\mathcal{H}_g$ and we have by (3.44) *and* (3.40)

$$g = \mathsf{P}_{\mathcal{K}}(f) \quad \Rightarrow \quad g = \mathsf{P}_{\mathcal{H}_g}(f), \quad \mathcal{H}_g \subset \mathcal{H}_f. \tag{3.50}$$

Proof. The left implication in (3.48) is immediate, since $\Xi \in \mathcal{N}_{\Omega_g}$ implies $\Xi \in \mathcal{N}$ and therefore $\xi \in \mathcal{K}^\circ$ by Proposition 3.17; moreover, ξ is orthogonal to $\mathcal{H}_g$ by (3.47) and therefore it is also orthogonal to $g \in \mathcal{H}_g$, so that $\xi \in \partial I_{\mathcal{K}}(g)$ by (3.43).

Conversely, if $\xi \in \partial I_{\mathcal{K}}(g)$, then $\Xi \in \mathcal{N}$ by (3.43) and Proposition 3.17. Moreover, denoting by $\gamma = g'$ the nonnegative Radon measure associated to the distributional derivative of g in $(0,1)$, the next Lemma 3.19 yields

$$0 \overset{(3.43)}{=} \int_0^1 \xi(w)g(w)\,\mathrm{d}w \overset{(3.51)}{=} -\int_0^1 \Xi(w)\,\mathrm{d}\gamma(w),$$

which shows that $\Xi(w) = 0$ on the support of γ and yields $\Xi \in \mathcal{N}_{\Omega_g}$. □

Lemma 3.19. *Let $g \in \mathcal{K}$ and $\xi \in \mathcal{K}^\circ$ with (nonnegative) primitive $\Xi \in \mathcal{N}$. If $\gamma = g'$ is the nonnegative Radon measure associated to the distributional derivative of g in $(0,1)$, then $\Xi \in L^1(\gamma)$ and*

$$\int_0^1 g(w)\xi(w)\,\mathrm{d}w = -\int_0^1 \Xi(w)\,\mathrm{d}\gamma(w). \tag{3.51}$$

Proof. Since γ is a nonnegative Radon measure in $(0,1)$ but is not necessarily finite, we need an approximation argument to justify (3.51). Let $\varphi_n \in C_0^\infty(0,1)$ be an increasing sequence of nonnegative functions such that $\lim_{n\uparrow+\infty} \varphi_n(w) = 1$, $|\varphi_n'| \le 2n$, and $\varphi_n(w) \equiv 1$ for $1/n \le w \le 1-1/n$. We have

$$\int_0^1 g\xi\varphi_n\,\mathrm{d}w = -\int_0^1 \Xi\varphi_n\,\mathrm{d}\gamma - \int_0^1 \Xi g\varphi_n'\,\mathrm{d}w \tag{3.52a}$$

$$= -\int_0^1 \Xi\varphi_n\,\mathrm{d}\gamma - \int_0^{1/n} \Xi g\varphi_n'\,\mathrm{d}w - \int_{1-1/n}^1 \Xi g\varphi_n'\,\mathrm{d}w. \tag{3.52b}$$

Applying Hardy's inequality, we get

$$\left|\int_0^{1/n} \Xi g\varphi_n'\,\mathrm{d}w\right| \le 2n\|w^{-1}\Xi\|_{L^2(0,1/n)}\|wg\|_{L^2(0,1/n)} \le 2C\|\xi\|_{L^2(0,1)}\|g\|_{L^2(0,1/n)}$$

so that the integral vanishes as $n \uparrow +\infty$. A similar argument holds for the last integral of (3.52b). Passing to the limit in (3.52a), (3.52b) as $n \uparrow +\infty$ and using the Lebesgue dominated (since $g\xi \in L^1(0,1)$) or monotone (since $\Xi \ge 0$ and φ_n is increasing) convergence theorem, we conclude. □

The last lemma of this section provides a useful example concerning a class of elements in $\partial I_{\mathcal{K}}(g)$.

Lemma 3.20 (an example of minimal selection in $\partial I_{\mathcal{K}}$)**.** *If $g, h \in \mathcal{K}$, then*

$$\xi_h := \mathsf{P}_{\mathcal{H}_g}(h) - h \in \partial I_{\mathcal{K}}(g).$$

Moreover,

$$\|z - h - \xi_h\|_{L^2(0,1)} \le \|z - h - \xi\|_{L^2(0,1)} \quad \forall\, \xi \in \partial I_{\mathcal{K}}(g), \quad z \in \mathcal{H}_g. \tag{3.53}$$

In particular,

$$\text{if } z \in \mathcal{H}_g, \text{ then } \quad \|z\|_{L^2(0,1)} \le \|z - \xi\|_{L^2(0,1)} \quad \forall\, \xi \in \partial I_K(g). \tag{3.54}$$

Proof. Since $h - \mathsf{P}_{\mathcal{H}_g}(h)$ is orthogonal to $\mathcal{H}_g$ (thus, in particular, to g), by (3.43) we have to check that $\xi_h \in \mathcal{K}^\circ$ by applying Proposition 3.17. By (3.46), $\xi_h = 0$ a.e. in $(0,1) \setminus \Omega_g$, so that the primitive Ξ_h of ξ_h satisfies

$$\Xi_h(w) = \int_{\Omega_g \cap (0,w)} \xi_h(s)\,\mathrm{d}s.$$

The thesis then follows if we show that for every connected component (α, β) of Ω_g we have $\Xi_h(\alpha) = \Xi_h(\beta) = 0 = \min_{[\alpha,\beta]} \Xi_h$. Since the characteristic function $\chi_{(0,\alpha)}$ of $(0,\alpha)$ belongs to $\mathcal{H}_g$, we have

$$\Xi_h(\alpha) = \int_0^\alpha \xi_h(w)\,\mathrm{d}w = (h - \mathsf{P}_{\mathcal{H}_g}(h) | \chi_{(0,\alpha)}) = 0.$$

A similar argument shows that $\Xi_h(\beta) = 0$. Moreover, for $w \in (\alpha, \beta)$ we have

$$\Xi_h(w) = \int_\alpha^w \xi_h(s)\,\mathrm{d}s \stackrel{(3.46)}{=} (w - \alpha) \fint_\alpha^\beta h(w)\,\mathrm{d}w - \int_\alpha^w h(w)\,\mathrm{d}w,$$

which shows that Ξ_h is concave and therefore nonnegative in (α, β).

Equation (3.53) follows immediately by observing that $\xi, \xi_h \in (\mathcal{H}_g)^\perp$ and $z - h - \xi_h$ belongs to $\mathcal{H}_g$ and, therefore, it is the orthogonal projection of $z - h$ and of $z - h - \xi$ onto $\mathcal{H}_g$. □

3.3 The Lagrangian formulation of the discrete sticky particle system

In this section we shall show that the discrete SPS satisfies the three characterizations of Theorem 3.6 and we prove Theorem 3.2.

Notation 3.21. Let us recapitulate our basic notation and definitions.

1. $P_i(t) = (m_i, x_i(t), v_i(t))$, $i \in I = \{1, \ldots, N\}$, $t \ge 0$ is a solution of the discrete SPS.
2. The positions of the particles are ordered: $x_1(t) \le x_2(t) \le \cdots \le x_N(t)$.
3. The sets $J_i(t) := \{j \in I : x_j(t) = x_i(t)\}$ are nondecreasing with respect to time. They correspond to a single particle of mass $\sum_{j \in J_i(t)} m_j$.

4. At each time t we pick up the collection of minimal indices

$$I(t) := \{\min J_i(t) : i = 1, \dots, N\} = \big\{i_1(t) < \dots < i_{N(t)}(t)\big\} \subset I$$

so that each $J_i(t)$ is of the form $\{j \in I : i_k(t) \le j < i_{k+1}(t)\}$ for some k and $(J_i(t))_{i \in I(t)}$ is a partition of I.

5. We denote by $0 < t_1 < t_2 < \cdots t_h < \cdots < t_{H-1}$ the (finite) sequence of times at which the cardinality of some $J_i(t)$ has an increasing jump; setting $t_0 = 0$ and $t_H = +\infty$, $\{[t_h, t_{h+1})\}_{h=0}^H$ is the associate partition of the positive real line with step sizes $\delta_h := t_h - t_{h-1}$.

6. The functions x_i are globally continuous and linear on each interval $[t_h, t_{h+1})$, with piecewise constant, right-continuous derivatives $v_i(t)$ satisfying (3.1). Each set $J_i(t)$ and $I(t)$ is also constant in each interval $[t_h, t_{h+1})$.

Let $\rho_t = \sum_{i \in I} m_i \delta_{x_i(t)}$ be the measure induced by the discrete SPS. In order to explicitly write the function $X(t) := X_{\rho_t}$ we consider the subdivision of $[0,1]$ given by

$$w_0 = 0 < w_1 < \dots < w_N = 1, \quad w_i = w_{i-1} + m_i = \sum_{j=1}^{i} m_j, \qquad i \in I.$$

We also set

$$W_i := [w_{i-1}, w_i), \quad W_i(t) = \bigcup_{j \in J_i(t)} W_j, \qquad i \in I,$$

and we note that

$$X(t) = \sum_{i=1}^{N} x_i(t) \mathbb{1}_{W_i}, \quad \frac{\mathrm{d}^+}{\mathrm{d}t} X(t) = V(t) = \sum_{i=1}^{N} v_i(t) \mathbb{1}_{W_i}. \tag{3.55}$$

The main result of this section is contained in the following theorem.

Theorem 3.22. (Lagrangian formulation of the discrete SPS). *The couple* (X, V)*, defined by* (3.55)*, satisfies the equations* (L.I)–(L.III) *and the properties* (L.a)–(L.c) *of Theorem* 3.6*. In particular, it defines a semigroup* S_t *in the discrete subspace*

$$\hat{\mathcal{X}} := \Big\{(X, V) \in \mathcal{X}_p(0,1) : X = \sum_{i=1}^{N} x_i \mathbb{1}_{W_i} \\ \text{for a finite interval partition } (W_i)_{i=1}^N \text{ of } [0,1)\Big\}. \tag{3.56}$$

Proof. We split the proof into various steps.

The collection $(W_i(t))_{i \in I(t)}$ is a partition of $[0,1)$. In $L^2(0,1)$ we introduce the decreasing family of finite-dimensional spaces $\mathcal{H}(t)$ whose elements are piecewise constant on each interval $W_i(t)$, $i \in I(t)$. Note that, by the very definitions of $\Omega_{X(t)}$ and $\mathcal{H}_{X(t)}$, (3.39) and (3.45),

$$\Omega_{X(t)} = (0,1) \setminus \{w_i : i \in I(t)\}, \quad \mathcal{H}(t) = \mathcal{H}_{X(t)}. \tag{3.57}$$

Besides (3.55), the crucial features describing the evolution of $X(t)$ are

$$X(t) \in \mathcal{K} \cap \mathcal{H}(t), \quad V(t) \in \mathcal{H}(t), \qquad (3.58)$$
$$\mathcal{H}(t), V(t) \text{ are constant in each time interval } [t_h, t_{h+1})$$

(we set $\mathcal{H}_h := \mathcal{H}(t)$, $V_h := V(t)$ if $t \in [t_h, t_{h+1})$) and the update rule for the velocity (3.1): $V(t_h)$ is constant in each interval $W_i(t_h) = \cup_{j\in J_i(t_h)} W_j$ and its value is given by

$$V(t_h+)|_{W_i(t_h)} = \frac{\sum_{j\in J_i(t_h)} m_j v_j(t_{h-1})}{\sum_{j\in J_i(t_h)} m_i} = \big(\mathscr{L}^1(W_i(t_h))\big)^{-1} \int_{W_i(t_h)} V(t_{h-1})\,\mathrm{d}w,$$

so that by (3.46)

$$V_h = \mathsf{P}_{\mathcal{H}_h}(V_{h-1}) = \mathsf{P}_{\mathcal{H}_h}(V_0) \quad \text{since} \quad \mathcal{H}_0 \supset \mathcal{H}_1 \supset \mathcal{H}_2 \supset \cdots \mathcal{H}_h, \qquad (3.59)$$

which yields (L.a) and (L.b). The next lemma shows (L.III).

Lemma 3.23. *Let $\tilde X(t) := X_0 + tV_0$ be associated to the free system $\tilde P_i = (m_i, \tilde x_i, \tilde v_i)$ given by $\tilde x_i(t) = x_i(0) + tv_i(0)$, $\tilde v_i(t) \equiv \tilde v_i = v_i(0)$. Then*

$$X(t) = \mathsf{P}_{\mathcal{H}(t)}(\tilde X(t)) = \mathsf{P}_{\mathcal{H}(t)}(X_0 + tV_0), \qquad (3.60)$$

$$t\frac{\mathrm{d}^+}{\mathrm{d}t} X(t) = tV(t) = X(t) - X_0 - \Xi(t)$$

for $\Xi(t) := -X_0 + \mathsf{P}_{\mathcal{H}(t)}(X_0) \in \partial I_{\mathcal{K}}(X(t))$.

Proof. Suppose that $t \in [t_h, t_{h+1})$; since $X(t) \in \mathcal{H}(t) = \mathcal{H}_h \subset \mathcal{H}(r)$ and $V(r) = \mathsf{P}_{\mathcal{H}(r)}(V_0)$ for $0 \le r \le t$ by (3.59), we have by the linearity of $\mathsf{P}_{\mathcal{H}(r)}$

$$\begin{aligned} X(t) &= X_0 + \int_0^t V(r)\,\mathrm{d}r \overset{(3.58)}{=} \mathsf{P}_{\mathcal{H}(t)}\Big(X_0 + \int_0^t V(r)\,\mathrm{d}r\Big) \\ &\overset{(3.59)}{=} \mathsf{P}_{\mathcal{H}(t)}(X_0) + \int_0^t \mathsf{P}_{\mathcal{H}(t)}(\mathsf{P}_{\mathcal{H}(r)}(V_0))\,\mathrm{d}r = \mathsf{P}_{\mathcal{H}(t)}(X_0) + \int_0^t \mathsf{P}_{\mathcal{H}(t)}(V_0)\,\mathrm{d}r \\ &= \mathsf{P}_{\mathcal{H}(t)}\Big(X_0 + \int_0^t V_0\,\mathrm{d}r\Big) = \mathsf{P}_{\mathcal{H}(t)}(\tilde X(t)). \end{aligned}$$

From (3.60) we have

$$t\frac{\mathrm{d}^+}{\mathrm{d}t} X(t) = tV(t) \overset{(3.59)}{=} \mathsf{P}_{\mathcal{H}(t)}(tV_0) \overset{(3.60)}{=} X(t) - \mathsf{P}_{\mathcal{H}(t)}(X_0) = X(t) - X_0 - \Xi(t),$$

where $\Xi(t) = \mathsf{P}_{\mathcal{H}(t)}(X_0) - X_0$; since $X_0 \in \mathcal{K}$ and $\mathcal{H}(t) = \mathcal{H}_{X(t)}$, by Lemma 3.20 we conclude that $\Xi(t) \in \partial I_{\mathcal{K}}(X(t))$. □

We now conclude the proof of (L.I) and (L.II); note that (L.c) follows directly from (L.II) and (L.a) via the semigroup property of $\mathscr{S}_t$ in $\hat{\mathcal{V}}(\mathbb{R})$.

Lemma 3.24. *Under the same notation and assumptions as before, we have*

$$U(t) := V_0 - V(t) = V_0 - \mathsf{P}_{\mathcal{H}(t)}(V_0) \in \partial I_{\mathcal{K}}(X(t)), \quad X(t) = \mathsf{P}_{\mathcal{K}}(X_0 + tV_0). \quad (3.61)$$

Proof. Since $(U(t)|X(t)) = 0$ $(X(t) \in \mathcal{H}(t))$, the first inclusion of (3.61) is equivalent to

$$U(t) = V_0 - V(t) \in \mathcal{K}^\circ \quad \forall\, t \geq 0 \tag{3.62}$$

by (3.43). It is not restrictive to assume that $t = t_h$ and $V(t) = V_h$ for some $h \in \{1, \ldots, H-1\}$. Since $\mathcal{K}^\circ$ is a cone and $V_0 - V_h$ can be decomposed into the sum

$$V_0 - V_h = \sum_{k=0}^{h-1} (V_k - V_{k+1}),$$

it is sufficient to prove that $V_k - V_{k+1} \in \mathcal{K}^\circ$ or, equivalently, that $\delta_{k+1}(V_k - V_{k+1}) \in \mathcal{K}^\circ$. Since $V_{k+1} = \mathsf{P}_{\mathcal{H}_{k+1}}(V_k)$ we obtain

$$\begin{aligned}\delta_{k+1}(V_k - V_{k+1}) &= \delta_{k+1}V_k - \mathsf{P}_{\mathcal{H}_{k+1}}(\delta_{k+1}V_k) = (X_{k+1} - X_k) - \mathsf{P}_{\mathcal{H}_{k+1}}(X_{k+1} - X_k)\\ &= X_{k+1} - \mathsf{P}_{\mathcal{H}_{k+1}}(X_{k+1}) + \mathsf{P}_{\mathcal{H}_{k+1}}(X_k) - X_k = \mathsf{P}_{\mathcal{H}_{k+1}}(X_k) - X_k \in \mathcal{K}^\circ\end{aligned}$$

by Lemma 3.20 and (3.43).

The second identity of (3.61) follows now by a similar argument by checking the conditions of (3.44). Since $X(t) \in \mathcal{K}$ and $(\tilde{X}(t) - X(t)|X(t)) = 0$ by (3.60), it is sufficient to show that $\tilde{X}(t) - X(t) \in \mathcal{K}^\circ$. On the other hand,

$$\tilde{X}(t) - X(t) = tV_0 - \int_0^t V(r)\,\mathrm{d}r = \int_0^t (V_0 - V(r))\,\mathrm{d}r = \int_0^t U(r)\,\mathrm{d}r$$

and (3.62) shows that $U(r) \in \mathcal{K}^\circ$ for every $r \geq 0$. Since $\mathcal{K}^\circ$ is a cone, we conclude. □

Proof of Theorem 3.2. Let us now consider two discrete Lagrangian solutions $(X^\ell(t), V^\ell(t)) = \mathsf{S}_t(X_0^\ell, V_0^\ell) \in \hat{\mathcal{X}}$, $\ell = 1, 2$. Equations (3.32), (3.33), and (L.II) immediately yield the estimates

$$\int_0^1 \psi\left(X^1(t) - X^2(t)\right)\mathrm{d}w \leq \int_0^1 \psi\left(X_0^1 - X_0^2 + t(V_0^1 - V_0^2)\right)\mathrm{d}w, \tag{3.63}$$

$$\|X^1(t) - X^2(t)\|_{L^p(0,1)} \leq \|X_0^1 - X_0^2\|_{L^p(0,1)} + t\|V_0^1 - V_0^2\|_{L^p(0,1)}, \tag{3.64}$$

which are equivalent to (3.7a) and (3.7b). Equation (L.I) yields [44, Thm. 3], [45, Thm. 1.2]

$$\int_0^t \|V^1 - V^2\|^2\,\mathrm{d}r \leq C(1+t)\left(\sum_{\ell=1,2} \|X_0^\ell\| + \|V_0^\ell\|\right)\left(\|X_0^1 - X_0^2\| + \|V_0^1 - V_0^2\|\right) \tag{3.65}$$

(here $\|\cdot\|$ denotes the norm of $L^2(0,1)$), which is equivalent to (3.7c). □

3.4 Stability and uniqueness of Lagrangian solutions

Our first result concerns the stability of Lagrangian solutions to (L.I)–(L.III) of Theorem 3.6 (in particular, it applies to those obtained by the discrete SPS in $\hat{\mathcal{X}}$).

Lemma 3.25. *Let $X^n, V^n := \frac{\mathrm{d}^+}{\mathrm{d}t}X^n$ be curves satisfying the equations* (L.I)–(L.III) *and the properties* (L.a)–(L.c) *stated in Theorem* 3.6 *with respect to initial data $X_0^n, V_0^n = v_0^n(X_0^n)$ converging to $X_0, V_0 = v_0(X_0)$ in $L^p(0,1)$, $p \geq 2$. Then we have the following.*

(a) *$X^n(t)$ converges to $X(t)$ in $L^p(0,1)$ uniformly in each compact interval; X is Lipschitz continuous with values in $L^p(0,1)$.*

(b) *The Lipschitz curve X is right-differentiable at each point t, with right-continuous derivative $V(t)$, and it satisfies* (L.I)–(L.III) *and* (L.a)–(L.c) *of Theorem* 3.6*.*

(c) *V^n strongly converges to V in $L^2(0,T;L^2(0,1))$ for every $T > 0$.*

(d) *The curve X is differentiable in $L^p(0,1)$ and V is continuous at each point of $(0,+\infty)\setminus\mathcal{T}$, where $\mathcal{T}$ is the jump set of the nonincreasing map $t \mapsto \|V(t)\|_{L^2(0,1)}$.*

(e) *If $\bar{V}$ is any weak accumulation point of $V^n(t)$ in $L^p(0,1)$, then $\mathsf{P}_{\mathcal{H}_{X(t)}}(\bar{V}) = V(t)$.*

(f) *$V^n(t) \to V(t)$ in $L^p(0,1)$ for every $t \in [0,+\infty)\setminus\mathcal{T}$.*

Proof. (a) The proof of (a) is an immediate consequence of (L.II) and (3.33), which also shows that X^n is uniformly Lipschitz continuous with values in $L^p(0,1)$ and Lipschitz constant bounded by $\|V_0^n\|_{L^p(0,1)}$. The convergence is therefore uniform in each compact interval and the limit function X satisfies the same Lipschitz bound with constant $\|V_0\|_{L^p(0,1)}$.

(b), (c) Standard stability results for gradient flows in Hilbert spaces [19] show that X solves (L.I) and (L.III); in particular X is right-differentiable in $L^2(0,1)$ at each $t \geq 0$, with $L^2(0,1)$ right derivative $V(t)$ which is right-continuous. Equation (3.65) shows that V is the limit of V^n in $L^2(0,T;L^2(0,1))$ for every $T > 0$ (this proves point (c)): in particular, up to the extraction of a suitable subsequence n_k, we can find an $\mathscr{L}^1$-negligible set $N \subset (0,+\infty)$ such that $V^{n_k}(t) \to V(t)$ in $L^2(0,1)$ for every $t \in [0,+\infty)\setminus N$ as $k \uparrow +\infty$. Passing to the limit in (L.c) and in (L.I), we obtain that

$$X(t) = \mathsf{P}_{\mathcal{K}}(X(s) + (t-s)V(s)), \qquad \frac{\mathrm{d}^+}{\mathrm{d}t}X(t) = V(t) \in -\partial I_{\mathcal{K}}(X(t)) + V(s) \tag{3.66}$$

for every $s \in [0,+\infty)\setminus N$ and $t \geq s$. Since V is right-continuous, (3.66) eventually holds for every $0 \leq s \leq t$.

The projection formula of (3.66) shows that for every $h \geq 0$

$$\|X(t+h) - X(t)\|_{L^p(0,1)} \leq h\|V(t)\|_{L^p(0,1)} \leq h\|V(s)\|_{L^p(0,1)} \quad \forall\, 0 \leq s \leq t, \tag{3.67}$$

and, more generally,

$$\int_0^1 \psi\left(h^{-1}(X(t+h) - X(t))\right) \mathrm{d}w \leq \int_0^1 \psi(V_s)\,\mathrm{d}w \leq \int_0^1 \psi(V_0)\,\mathrm{d}w \quad \forall\, 0 \leq s \leq t \tag{3.68}$$

for every convex nonnegative function $\psi : \mathbb{R} \to \mathbb{R}$. Equation (3.67) and the right-differentiability of X in $L^2(0,1)$ yields that $V(t)$ is also the right derivative of X in $L^p(0,1)$, its L^p-norm is not increasing, and, by (3.68), the family V_s is uniformly p-integrable (by the Dunford–Pettis criterion, it is sufficient to choose a convex function ψ with $\psi(r)/|r|^p \to +\infty$ as $|r| \to +\infty$ and $\psi \circ V_0 \in L^1(0,1)$; see, e.g., [43, Lem. 3.7]).

From (L.III) we deduce that $tV(t) = X(t) - X_0 - \Xi(t)$, where $\Xi(t)$ is characterized by

$$\Xi(t) \in \partial I_{\mathcal{K}}(X(t)), \quad \|X(t) - X_0 - \Xi(t)\| \le \|X(t) - X_0 - \xi\| \quad \forall \xi \in \partial I_{\mathcal{K}}(X(t)). \tag{3.69}$$

Applying Lemma 3.20 with $g := X(t)$ and $h := X_0$, we obtain $\Xi(t) = \mathsf{P}_{\mathcal{H}_{X(t)}}(X_0) - X_0$ and, therefore,

$$tV(t) = X(t) - \mathsf{P}_{\mathcal{H}_{X(t)}}(X_0), \quad V(t) \in \mathcal{H}_{X(t)}. \tag{3.70}$$

It follows by (3.50) and (3.66) that $\mathcal{H}_{X(s)} \supset \mathcal{H}_{X(t)}$ if $0 \le s \le t$; moreover, by (3.22), there exists a Borel map $v_t \in L^p_{\rho_t}(\mathbb{R})$ such that

$$V(t) = v_t \circ X(t), \quad V(t) = \mathsf{P}_{\mathcal{H}_{X(t)}}(V(s)) \quad \forall\, 0 \le s \le t, \tag{3.71}$$

where the last identity follows from the second formula of (3.66) and the fact that $V(t)$ belongs to $\mathcal{H}_{X(t)}$, whereas $\partial I_{\mathcal{K}}(X(t))$ is orthogonal to $\mathcal{H}_{X(t)}$.

(d) Let $\mathcal{T}$ be the jump set of the L^2-norm of $V(t)$; we show that V is left-continuous at every $\bar t \in (0,+\infty) \setminus \mathcal{T}$ (this also yields the left-differentiability of X at $\bar t$). Equation (L.I) provides the minimal selection characterization of V at every time $t \ge 0$:

$$V(t) \in V_0 - \partial I_{\mathcal{K}}(X(t)), \quad \|V(t)\|_{L^2(0,1)} \le \|V_0 - \xi\|_{L^2(0,1)} \quad \forall \xi \in \partial I_{\mathcal{K}}(X(t)). \tag{3.72}$$

Take an arbitrary increasing sequence $t_n \uparrow \bar t$ such that $V(t_n) \rightharpoonup \bar V$ in $L^p(0,1)$. Since the graph of $\partial I_{\mathcal{K}}$ is strongly-weakly closed in $L^2(0,1)$, we have $\bar V \in V_0 - \partial I_K(X(\bar t))$. Passing to the limit in (3.72) we obtain

$$\begin{aligned}\|\bar V\|_{L^2(0,1)} &\le \lim_{n\to\infty} \|V(t_n)\|_{L^2(0,1)}\\ &= \|V(\bar t)\|_{L^2(0,1)} \le \|V_0 - \bar\xi\|_{L^2(0,1)} \quad \forall \bar\xi \in \partial I_{\mathcal{K}}(X(\bar t)).\end{aligned}$$

Since $\partial I_K(X(\bar t))$ is a closed convex set, it follows that $\bar V = V(\bar t)$ and that the convergence is strong in $L^2(0,1)$ and therefore also in $L^p(0,1)$, since $V(t_n)$ is uniformly p-integrable.

(e) Let n_k be an arbitrary subsequence such that $V^{n_k}(t) \rightharpoonup \bar V$ in $L^p(0,1)$. Passing to the limit in the inclusion $V^n(t) \in V_0^n - \partial I_{\mathcal{K}}(X^n(t))$, we obtain $\bar V \in V_0 - \partial I_{\mathcal{K}}(X(t))$. By Theorem 3.18 any element in $\partial I_{\mathcal{K}}(X(t))$ is orthogonal to $\mathcal{H}_{X(t)}$ so that

$$\mathsf{P}_{\mathcal{H}_{X(t)}}(\bar V) = \mathsf{P}_{\mathcal{H}_{X(t)}}(V_0) \overset{\text{(L.a)}}{=} V(t).$$

(f) Now let $t \in (0,+\infty)\setminus\mathcal{T}$, and let $n_k, \bar V$ be as in the previous point (e). Up to the extraction of a further subsequence (still denoted by n_k), there exists a dense set $S \subset (0,+\infty)$ such that $V^{n_k}(s) \to V(s)$ for every $s \in S$ so that $\forall\, s \in S$, $s < t$

$$\|\bar V\|_{L^2(0,1)} \le \limsup_{k\uparrow+\infty} \|V^{n_k}(t)\|_{L^2(0,1)} \le \limsup_{k\uparrow+\infty} \|V^{n_k}(s)\|_{L^2(0,1)} = \|V(s)\|_{L^2(0,1)}.$$

Since t is a continuity point for V, we obtain by (3.72)

$$\|\bar{V}\|_{L^2(0,1)} \le \|V(t)\|_{L^2(0,1)} \le \|V_0 - \xi\|_{L^2(0,1)} \quad \forall \xi \in \partial I_{\mathcal{K}}(X(t)), \tag{3.73}$$

which yields $\bar{V} = V(t)$, $\limsup_{k\uparrow+\infty} \|V^{n_k}(t)\|_{L^2(0,1)} \le \|V(t)\|_{L^2(0,1)}$ and the strong convergence of $V^n(t)$ to $V(t)$ in $L^2(0,1)$. The strong convergence in $L^p(0,1)$ follows by the uniform p-integrability estimate (3.68). □

Corollary 3.26 (existence of the Lagrangian semigroup)**.** *For every initial data* $(X_0, V_0) \in \mathcal{X}_2(0,1)$*, there exists a unique Lipschitz curve* X *in* $L^2(0,1)$ *satisfying the equations* (L.I)–(L.III) *and the properties* (L.a)–(L.c) *stated in Theorem* 3.6*. Setting* $V(t) := \frac{\mathrm{d}^+}{\mathrm{d}t}X(t)$*, the map* $\mathsf{S}_t : (X_0, V_0) \mapsto (X(t), V(t))$ *defines a right-continuous semigroup in each space* $\mathcal{X}_p(0,1)$*,* $p \ge 2$*.*

Proof. It is sufficient to approximate $(X_0, V_0) \in \mathcal{X}_p(0,1)$ by a sequence $(X_0^n, V_0^n) \in \hat{\mathcal{X}}$ of initial data arising from finite discrete distributions of space and velocities in $\hat{\mathcal{V}}(\mathbb{R})$ and to apply the previous lemma. □

Corollary 3.27 (equivalent characterizations)**.** *Let* $(X_0, V_0) \in \mathcal{X}_2(0,1)$ *be given initial data. If* X *is a solution of one of the equations* (L.I)*,* (L.II)*,* (L.III)*, then it satisfies all the formulations* (L.I)–(L.III) *and the properties* (L.a)–(L.c) *stated in Theorem* 3.6*.*

Proof. The thesis is obvious in the case of (L.I) and (L.II), whose solution is unique, and should coincide with the Lagrangian evolution provided by Corollary 3.26.

Let us now assume that X is a Lipschitz curve solving (L.III), let $\tilde{X}$ be the Lagrangian solution given by the previous Corollary 3.26 with initial data X_0, V_0, and let us set $V_0^n := n(X(n^{-1}) - X_0)$, $X^n(t) := \mathsf{P}_{\mathcal{K}}(X_0 + tV_0^n)$. $X^n(t)$ is thus a Lagrangian flow satisfying (L.I)–(L.III) with respect to the initial data X_0, V_0^n; in particular,

$$t\frac{\mathrm{d}}{\mathrm{d}t}X^n(t) \in -\partial I_{\mathcal{K}}(X^n(t)) + X^n(t) - X_0, \quad X^n(n^{-1}) = X(n^{-1}) \tag{3.74}$$

so that $X^n(t) = X(t)$ for $t \ge n^{-1}$. On the other hand, the stability Lemma 3.25 yields

$$\|X^n(t) - \tilde{X}(t)\| \le t\|V_0^n - V_0\| = t\|n(X(n^{-1}) - X_0) - V_0\| \overset{\text{(L.III)}}{\to} 0 \quad \text{as } n \uparrow +\infty \tag{3.75}$$

so that $X = \tilde{X}$. □

3.5 The continuous sticky particle system in Eulerian coordinates

In this section we conclude the proofs of the various theorems of section 3.1.

Proof of Proposition 3.1. Starting from (3.4) it is immediate to check that D_p is a metric on $\mathcal{V}_p(\mathbb{R})$. Let us check the equivalence characterization (3.6): assuming first that $D_p(\mu_n, \mu) \to 0$, we obviously have $W_p(\rho_n, \rho) \to 0$; since

$X_n = X_{\rho_n} \to X = X_\rho$ and $v_n(X_n) \to v(X)$ in $L^p(0,1)$ as $n \uparrow +\infty$, for a continuous and bounded test function $\zeta : \mathbb{R} \to \mathbb{R}$ we easily get

$$\begin{aligned}\lim_{n\uparrow+\infty} \int_{\mathbb{R}} \zeta(x) v_n(x)\,\mathrm{d}\rho_n(x) &= \lim_{n\uparrow+\infty} \int_0^1 \zeta(X_n(w)) v_n(X_n(w))\,\mathrm{d}w \\ &= \int_0^1 \zeta(X(w)) v(X(w))\,\mathrm{d}w = \int_{\mathbb{R}} \zeta(x) v(x)\,\mathrm{d}\rho(x),\end{aligned} \tag{3.76}$$

showing that $\rho_n v_n \rightharpoonup \rho v$ and

$$\begin{aligned}\lim_{n\uparrow+\infty} \int_{\mathbb{R}} |v_n(x)|^p\,\mathrm{d}\rho_n(x) &= \lim_{n\uparrow+\infty} \int_0^1 |v_n(X_n(w))|^p\,\mathrm{d}w \\ &= \int_0^1 |v(X(w))|^p\,\mathrm{d}w = \int_{\mathbb{R}} |v(x)|^p\,\mathrm{d}\rho(x).\end{aligned}$$

The converse implication is a particular case of [2, Thm. 5.4.4]; here is a simplified argument. If (3.6) holds, then one gets the strong convergence of X_n to X in $L^p(0,1)$; since $V_n := v_n \circ X_n$ is bounded in $L^p(0,1)$ up to the extraction of a suitable subsequence, one has $V_n \rightharpoonup V$ in $L^p(0,1)$ and, arguing as in (3.76),

$$\int_0^1 \zeta(X(w)) V(w)\,\mathrm{d}w = \int_0^1 \zeta(X(w)) v(X(w))\,\mathrm{d}w \quad \forall\,\zeta \in C_b(\mathbb{R}). \tag{3.77}$$

Note that a function in $L^p(0,1)$ of the form $b\circ X$ for some Borel map $b : \mathbb{R} \to \mathbb{R}$ belongs to $\mathcal{H}_X$; a simple approximation argument shows that the set $\{\zeta \circ X : \zeta \in C_b(\mathbb{R})\}$ is dense in $\mathcal{H}_X$ so that (3.77) yields

$$v \circ X = \mathsf{P}_{\mathcal{H}_X} V.$$

On the other hand, the last limit property stated in (3.6) yields

$$\|V\|_{L^p(0,1)} \le \lim_{n\uparrow+\infty} \|V_n\|_{L^p(0,1)} = \|v\circ X\|_{L^p(0,1)} = \|\mathsf{P}_{\mathcal{H}_X}(V)\|_{L^p(0,1)} \le \|V\|_{L^p(0,1)} \tag{3.78}$$

so that $v \circ X$ should coincide with V, which is also the strong limit of V_n in $L^p(0,1)$.

Let us finally consider the density of $\hat{\mathcal{V}}$: if $(\rho, \rho v) \in \mathcal{V}_p(\mathbb{R})$, we can first approximate v in $L^p_\rho(\mathbb{R})$ by a sequence of bounded and continuous functions $v_n \in C_b(\mathbb{R})$. We can then find a sequence $\rho^N = \sum_{j=1}^N m_{j,N}\delta_{x_{j,N}}$, $N \in \mathbb{N}$ such that $\rho^N \to \rho$ in $\mathcal{P}_p(\mathbb{R})$. It is then easy to check that $v_n\rho^N \rightharpoonup v_n\rho$ as $N \uparrow +\infty$ according to (3.6).

Proof of Theorem 3.3. (a) The extension of the semigroup $\mathscr{S}$ is not difficult, using the estimates of Theorem 3.2 and the density of $\hat{\mathcal{V}}(\mathbb{R})$ in $\mathcal{V}_p(\mathbb{R})$, but it is not completely trivial since the space $\mathcal{V}_p(\mathbb{R})$ is not complete and (3.7c) and (3.65) do not provide a pointwise continuous dependence of the velocity on the initial data. Therefore, we will use the equivalence stated in Theorem 3.6 (which we already proved at the level of discrete data in Theorem 3.22) and the Lagrangian stability result of Lemma 3.25. It is clear that the only possible extension of $\mathscr{S}_t$ to $\mathcal{V}_p(\mathbb{R})$ is given by formula (3.25). Since S_t is a semigroup in

$\mathcal{X}_p(0,1)$ satisfying $\lim_{t\downarrow 0}\mathsf{S}_t(X_0,V_0) = (X_0,V_0)$ strongly in $L^p(0,1)^2$, $\mathscr{S}_t$ satisfies (3.9).

In order to check that $\mathscr{S}_t$ is strongly-weakly continuous, we take a sequence $\mu_t^n = (\rho_t^n, \rho_t^n v_t^n) = \mathscr{S}_t[\mu_0^n] \in \hat{\mathcal{V}}$, with μ_0^n converging to $\mu = (\rho, \rho v) \in \mathcal{V}_p(\mathbb{R})$ with respect to D_p, and we consider the associated maps $(X^n(t), V^n(t)) = \mathsf{S}_t(X_0^n, V_0^n)$. By Lemma 3.25(f), for every weakly converging sequence $V^{n_k} \rightharpoonup \bar{V}$ in $L^p(0,1)$ and every test function $\zeta \in C_b^0(\mathbb{R})$ we have

$$\begin{aligned}\int_{\mathbb{R}} \zeta\, v_t^{n_k}\,\mathrm{d}\rho_t^{n_k} &= \int_0^1 \zeta(X^{n_k}(t))v_t^{n_k}(X^{n_k}(t))\,\mathrm{d}w \overset{\text{(L.a)}}{=} \int_0^1 \zeta(X_t^{n_k})V^{n_k}(t)\,\mathrm{d}w \\ &\xrightarrow{k\uparrow+\infty} \int_0^1 \zeta(X(t))\bar{V}\,\mathrm{d}w \overset{\text{Lemma 3.25(e)}}{=} \int_0^1 \zeta(X(t))V(t)\,\mathrm{d}w \overset{\text{(L.a)}}{=} \int_{\mathbb{R}} \zeta\, v_t\,\mathrm{d}\rho_t,\end{aligned}$$

where we used the fact that $\zeta(X^n(t)) \to \zeta(X(t))$ strongly in $L^p(0,1)$.

(b) It is immediate to check that $(\rho, \rho v) = \mathscr{S}(\rho_0, \rho_0 v_0)$ is a distributional solution of (3.2), since in Lagrangian coordinates the continuity equation reads

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} \zeta(x)\,\mathrm{d}\rho_t(x) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 \zeta(X(t))\,\mathrm{d}w \overset{\text{(L.a)}}{=} \int_0^1 \zeta'(X(t))V(t)\,\mathrm{d}w \\ &\overset{\text{(L.a)}}{=} \int_0^1 \zeta'(X(t))v_t(X(t))\,\mathrm{d}w = \int_{\mathbb{R}} \zeta'(x)v_t(x)\,\mathrm{d}\rho_t(x),\end{aligned}$$

and the momentum equation similarly becomes

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} \zeta(x)v_t(x)\,\mathrm{d}\rho_t(x) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 \zeta(X(t))V(t)\,\mathrm{d}w \overset{\text{(3.22) (L.a)}}{=} \frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 \zeta(X(t))V_0\,\mathrm{d}w \\ &\overset{\text{(L.a)}}{=} \int_0^1 \zeta'(X(t))V(t)V_0\,\mathrm{d}w = \int_0^1 \zeta'(X(t))v_t(X(t))V_0\,\mathrm{d}w \\ &\overset{\text{(3.22) (L.a)}}{=} \int_0^1 \zeta'(X(t))v_t^2(X(t))\,\mathrm{d}w = \int_0^1 \zeta'(x)v_t^2(x)\,\mathrm{d}\rho_t(x).\end{aligned}$$

Oleinik entropy condition (3.10) follows easily by (3.70), by observing that $\mathsf{P}_{\mathcal{H}_{X(t)}}(X_0)$ is a nondecreasing map, $V(t) = v_t(X(t))$, and $\rho_t = (X(t))_{\#}\lambda$.

(c) The proof of (c) follows from (3.68).

(d) The proof is equivalent to point (d) of Lemma 3.25; concerning the left continuity of $\rho_t v_t$ in the weak topology, we fix an arbitrary bounded Lipschitz test function $\zeta : \mathbb{R} \to \mathbb{R}$, and we observe that

$$\lim_{s\uparrow t}\int_{\mathbb{R}} \zeta(x)v_s(x)\,\mathrm{d}\rho_s(x) = \lim_{s\uparrow t}\int_0^1 \zeta(X(s))V(s)\,\mathrm{d}w = \lim_{s\uparrow t}\int_0^1 \zeta(X(t))V(s)\,\mathrm{d}w$$

since $X(s) \to X(t)$ in $L^2(0,1)$ as $s \uparrow t$. On the other hand, since $\zeta\circ X(t) \in \mathcal{H}_{X(t)}$, we have

$$\begin{aligned}\int_0^1 \zeta(X(t))V(s)\,\mathrm{d}w &= \int_0^1 \zeta(X(t))V(t)\,\mathrm{d}w \\ &= \int_0^1 \zeta(X(t))v_t(X(t))\,\mathrm{d}w = \int_{\mathbb{R}} \zeta(x)v_t(x)\,\mathrm{d}\rho_t(x).\end{aligned}$$

(e) The proof of (e) has already been discussed in point (a), except for the convergence at $t \in (0,+\infty)\setminus\mathcal{T}$, which follows from Lemma 3.25(f).

(f) Equation (3.12) follows by the projection representation (3.66) and Corollary 3.14. The limit in (3.12) can be obtained in the Lagrangian coordinate:

$$\lim_{h\downarrow 0}\int_{\mathbb{R}}\left|h^{-1}(\mathsf{x}_s^{s+h}-\mathsf{i})-v_s\right|^2\mathrm{d}\rho_s=\lim_{h\downarrow 0}\int_0^1\left|h^{-1}(X(s+h)-X(s))-V(s)\right|^2\mathrm{d}w=0,$$

since $t\mapsto X(t)$ is right-differentiable. Equation (3.13) is an immediate consequence of (3.70), which yields

$$(t-s)V(t)=X(t)-\mathsf{P}_{\mathcal{H}_{X(t)}}(X(s))\quad\forall\,0\le s<t.$$

Proof of Theorem 3.6. The proof now follows by applying Lemma 3.25 and its Corollaries 3.26 and 3.27.

Proof of Theorem 3.4. Equation (3.16) follows from a simple calculation starting from (L.III): we introduce Z, the monotone rearrangement of the measure $\eta\in\mathcal{P}_2(\mathbb{R})$ and we observe that $W_2^2(\rho_t,\eta)=\|X(t)-Z\|^2$ (we refer to the usual notation for (X,V) and we denote by $\|\cdot\|$ the norm in $L^2(0,1)$). We get for some $\Xi(t)\in\partial I_{\mathcal{K}}(X(t))$

$$\begin{aligned}\frac{t}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}W_2^2(\rho_t,\eta)&=\frac{t}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\|X(t)-Z\|^2=t(\dot X(t)|X(t)-Z)\\&\overset{\text{(L.III)}}{=}(X(t)-X_0-\Xi(t)|X(t)-Z)\overset{(3.42)}{\le}(X(t)-X_0|X(t)-Z)\\&=\frac12\|X(t)-Z\|^2-\frac12\|Z-X_0\|^2+\frac12\|X(t)-X_0\|^2\\&=\frac12W_2^2(\rho_t,\eta)-\phi^{\rho_0}(\rho_t)+\phi^{\rho_0}(\eta).\end{aligned}$$

Let us now consider the converse implication: if ρ_t satisfies (3.16), then $X(t)=X_{\rho_t}$ satisfies (see (3.27))

$$\frac{t}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|X(t)-Z\|^2-\frac12\|X(t)-Z\|^2\le\Phi^{\rho_0}(Z)-\Phi^{\rho_0}(X(t))\quad\forall\,Z\in\mathcal{K},\tag{3.79}$$

which is the equivalent metric formulation [2] of the differential inclusion (L.III).

Since $\boldsymbol{\rho}_t=(X_0,X(t))_\#\lambda$, (3.18) yields

$$\lim_{t\downarrow 0}t^{-2}\int_0^1|X_0+tV_0-X(t)|^2\,\mathrm{d}w=0,\tag{3.80}$$

i.e., $X(t)$ also satisfies the initial limit condition of (L.III). Therefore, setting $V:=\frac{\mathrm{d}}{\mathrm{d}t}X=v\circ X$, by Corollary 3.27 the couple $(X(t),V(t))$ coincides with the Lagrangian flow $\mathsf{S}_t(X_0,V_0)$ so that $(\rho_t,\rho_tv_t)=\mathscr{S}_t(\rho_0,\rho_0v_0)$.

Proof of Theorem 3.5. Let us first note that when $\mathsf{i}+\varepsilon_0v_0$ is ρ_0-essentially nondecreasing, (3.20) follows directly from (3.17), since the collision-free motion $\rho_t=(\mathsf{i}+tv_0)_\#\rho_0$ for $t\in[0,\varepsilon_0)$ is a solution of the SPS.

Let us now consider the general case, setting $\tilde\rho_{\varepsilon,t}:=\mathscr{G}^{\rho_0}_{\log(t/\varepsilon)}(\tilde\rho_\varepsilon)$. For every $\varepsilon>0$ let us consider the convex set of bounded Lipschitz functions

$$BL(\varepsilon):=\left\{u\in C^{0,1}(\mathbb{R}):\sup|u|\le\varepsilon^{-1},\mathrm{Lip}(u)\le(2\varepsilon)^{-1}\right\},$$

and let $u_\varepsilon \in BL(\varepsilon)$ be a minimizer of

$$m_\varepsilon = \min_{u\in BL(\varepsilon)} \|v_0 - u\| = \|v_0 - u_\varepsilon\|. \tag{3.81}$$

By standard approximation results, $\lim_{\varepsilon\downarrow 0} m_\varepsilon = 0$ so that u_ε converges to v_0.

By the definition of $BL(\varepsilon)$ the map $\mathsf{i} + \varepsilon u_\varepsilon$ is monotone and, therefore, it is the optimal map pushing ρ to $\hat\rho_\varepsilon = (\mathsf{i} + \varepsilon u_\varepsilon)_\#\rho_0$. The sticky particle solution $(\hat\rho_{\varepsilon,t}, \hat\rho_{\varepsilon,t}\hat v_{\varepsilon,t}) := \mathscr{S}_t(\hat\rho_0, \hat\rho_0 u_\varepsilon)$ admits the representation (see (3.17))

$$\hat\rho_{\varepsilon,t} = \mathscr{G}^{\rho_0}_{\log(t/\varepsilon)}(\hat\rho_\varepsilon)$$

so that, by the exponential rate of expansion of $\mathscr{G}$ we get

$$\begin{aligned} W_2(\hat\rho_{\varepsilon,t}, \tilde\rho_{\varepsilon,t}) &\le \exp\left(\log(t/\varepsilon)\right) W_2(\hat\rho_\varepsilon, \tilde\rho_\varepsilon) = \frac{t}{\varepsilon} W_2(\hat\rho_\varepsilon, \tilde\rho_\varepsilon) \\ &\le t\|v_0 - u_\varepsilon\|_{L^2_{\rho_0}(\mathbb{R})} \overset{(3.81)}{=} t m_\varepsilon. \end{aligned} \tag{3.82}$$

On the other hand, if $(\rho_t, \rho_t v_t) = \mathscr{S}_t(\rho_0, \rho_0 v_0)$, (3.7b) yields

$$W_2(\hat\rho_{\varepsilon,t}, \rho_t) \le t\|v_0 - u_\varepsilon\|_{L^2_{\rho_0}(\mathbb{R})} = t m_\varepsilon \quad \text{so that} \quad W_2(\rho_t, \tilde\rho_{\varepsilon,t}) \le 2 m_\varepsilon t, \tag{3.83}$$

and concludes the proof of (3.19).

We conclude this section by showing that the representation convergence theorem of Brenier and Grenier [18] can be easily deduced by our result, in particular by formula (L.II) of Theorem 3.6.

Theorem 3.28 (Brenier–Grenier theorem)**.** *Let $v_0 \in C^0(\mathbb{R})$, let ρ_0^N, $N \in \mathbb{N}$ be a sequence of discrete probability measures supported in a fixed compact interval $[-R, R]$ and weakly converging to ρ_0 in $\mathcal{P}(\mathbb{R})$, and let ρ_t^N be the solution of the discrete SPS with initial data $(\rho_0^N, v_0\rho_0^N)$. For every $t \ge 0$, ρ_t^N weakly converge to a probability measure ρ_t, whose distribution function $M_t(x) := \rho_t((-\infty, x])$, $t \ge 0$ is the unique entropy solution of*

$$\partial_t M + \partial_x(A(M)) = 0, \quad M(0) = M_0, \tag{3.84}$$

where the flux function $A : [0,1] \to \mathbb{R}$ is defined by

$$A(w) := \int_0^w V_0(r)\,\mathrm{d}r, \quad \textit{where} \quad V_0 := v_0 \circ X_0, \quad X_0 := X_{\rho_0}. \tag{3.85}$$

Proof. The convergence part follows by Theorem 3.3 and we can represent $X_t := X_{\rho_t}$ by the formula $X_t = \mathsf{P}_{\mathcal{K}}(X_0 + tV_0)$ of Theorem 3.6. Introducing the convex primitive functions $F_t(w) := \int_0^w X_t(r)\,\mathrm{d}r$, Theorem 3.10 yields

$$F_t = (F_0 + tA)^{**} \quad \text{so that} \quad (F_t)^* = (F_0 + tA)^*. \tag{3.86}$$

On the other hand, since the derivative X_t of F_t is the pseudoinverse of M_t (2.33), a standard duality result shows that $(F_t)^* = G_t$, where $G_t(x) = \int_{-\infty}^x M_t(y)\,\mathrm{d}y$, so that

$$G_t = (F_0 + tA)^* = (G_0^* + tA)^*. \tag{3.87}$$

It was already observed in [18, sect. 4] that (3.87) provides the second Hopf formula [6] for the viscosity solution of the Hamilton–Jacobi equation

$$\partial_t G + A(\partial_x G) = 0 \quad \text{in } \mathbb{R} \times (0, +\infty), \tag{3.88}$$

and therefore the derivative $M_t = \partial_x G_t$ is the entropy solution of (3.84). □

Chapter 4

Contraction property of the Fokker-Planck equation.

In this chapter we discuss a contraction property result for nonnegative measure-valued solutions to the Fokker-Planck equation

$$\partial_t\rho - \Delta\rho - \nabla\cdot(\rho B) = 0, \quad \rho_{|t=0} = \rho_0, \tag{4.1}$$

where $B : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel λ-monotone operator, $\lambda \in \mathbb{R}$, i.e.

$$\langle B(x) - B(y), x - y\rangle \geq \lambda \left|x - y\right|^2 \quad \text{for every } x, y \in \mathbb{R}^d.$$

More precisely, we consider weakly continuous families of probability measures $(\rho_t)_{t\geq 0} \subset \mathcal{P}(\mathbb{R}^d)$ satisfying the equation (4.1) in the sense of distributions

$$\int_0^{+\infty}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\,\mathrm{d}\rho_t\mathrm{d}t = 0 \quad \forall\,\zeta \in C^\infty_{\mathrm{c}}(\mathbb{R}^d\times(0,+\infty)), \tag{4.2}$$

with the initial datum ρ_0.

In the first section, we state the main results of the chapter and describe our strategy for the proof. In section 4.2, we collect some tools useful to our arguments: we present an approximation technique of the cost functional and a rescaling trick which allows to consider $\lambda = 0$ in the following arguments. Section 4.3 is devoted to show a comparison result for backward Kolmogorov equation (see Theorem 4.5): this is the key of the proof of the contraction property (see Theorem 4.1) contained in the last Section.

4.1 Main result

Let us first recall the general cost functional

$$\begin{aligned}\mathcal{C}_h(\rho^1,\rho^2) := \inf\Big\{\int_{\mathbb{R}^d\times\mathbb{R}^d} h(|x_1 - x_2|)\,\mathrm{d}\boldsymbol{\rho}(x_1,x_2) : \boldsymbol{\rho}\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d),\\ \boldsymbol{\rho}\text{ is a coupling between }\rho^1\text{ and }\rho^2\Big\}.\end{aligned} \tag{4.3}$$

Throughout this chapter we assume that

$h : [0, +\infty) \to [0, +\infty)$ is a continuous and **non-decreasing** function with $h(0) = 0$.

Since we are not assuming any homogeneity on the general cost function h, its rescaled versions

$$h_s(r) := h(r\,\mathrm{e}^s) \quad s \in \mathbb{R},\ r \geq 0$$

will be useful. Let us now state our main results:

Theorem 4.1. *If ρ^1, ρ^2 are two distributional solutions to (4.2) satisfying the summability condition*

$$\int_{t_0}^{t_1} \int_{\mathbb{R}^d} |B(x) - \lambda\, x| \,\mathrm{d}\rho_t(x)\,\mathrm{d}t < +\infty \quad \text{for every } 0 < t_0 < t_1 < +\infty, \tag{4.4}$$

then they satisfy

$$\mathcal{C}_{h_{\lambda t}}(\rho^1_t, \rho^2_t) \leq \mathcal{C}_h(\rho^1_0, \rho^2_0) \quad \text{for every } t \geq 0. \tag{4.5}$$

In particular, if $\rho^1_0 = \rho^2_0$ then ρ^1 and ρ^2 coincide for every time $t \geq 0$.

Let us make explicit some consequences of (4.5) according to the different signs of λ and the behaviour of h near 0 and $+\infty$:

Corollary 4.2. *Let ρ^1, ρ^2 be two distributional solutions to (4.2) satisfying (4.4).*

a) If B is monotone, i.e. $\lambda \geq 0$, then

$$\mathcal{C}_h(\rho^1_t, \rho^2_t) \leq \mathcal{C}_h(\rho^1_0, \rho^2_0).$$

b) If B is λ-monotone with $\lambda > 0$ and h satisfies for some exponent $p > 0$

$$h(\alpha\, r) \geq \alpha^p h(r) \quad \text{for every } \alpha \geq 1 \text{ and } r \geq 0 \tag{4.6}$$

then

$$\mathcal{C}_h(\rho^1_t, \rho^2_t) \leq \mathrm{e}^{-p\lambda\, t}\, \mathcal{C}_h(\rho^1_0, \rho^2_0).$$

c) If B is λ-monotone with $\lambda < 0$ and h satisfies for some exponent $p > 0$

$$h(\alpha\, r) \geq \alpha^p h(r) \quad \text{for every } \alpha \leq 1 \text{ and } r \geq 0$$

then

$$\mathcal{C}_h(\rho^1_t, \rho^2_t) \leq \mathrm{e}^{-p\lambda\, t}\, \mathcal{C}_h(\rho^1_0, \rho^2_0).$$

In the particular case of the Wasserstein distance W_p, $p \geq 1$, we have

$$W_p(\rho^1_t, \rho^2_t) \leq \mathrm{e}^{-\lambda t}\, W_p(\rho^1_0, \rho^2_0). \tag{4.7}$$

Theorem 4.1 has a simple application to invariant measures $\rho_\infty \in \mathcal{P}(\mathbb{R}^d)$, which are stationary solutions of (4.2) and therefore satisfy

$$\int_{\mathbb{R}^d} \Big(\Delta\zeta - B \cdot \nabla\zeta\Big)\,\mathrm{d}\rho_\infty = 0 \quad \forall\, \zeta \in C^\infty_{\mathrm{c}}(\mathbb{R}^d). \tag{4.8}$$

Corollary 4.3 (Strongly monotone operators and invariant measures)**.** *Let us suppose that B is strongly monotone, i.e. $\lambda > 0$. Then equation (4.8) has at most one solution $\rho_\infty \in \mathcal{P}(\mathbb{R}^d)$ satisfying the integrability condition*

$$\int_{\mathbb{R}^d} |Bx - \lambda x| \,\mathrm{d}\rho_\infty(x) < \infty. \tag{4.9}$$

For each solution ρ_t to (4.2)-(4.4) and each cost h satisfying (4.6) we have

$$\mathcal{C}_h(\rho_t, \rho_\infty) \le \mathrm{e}^{-p\lambda\,(t-t_0)} \mathcal{C}_h(\rho_{t_0}, \rho_\infty).$$

Note that in the case $\lambda > 0$ condition (4.9) is weaker than $B \in L^1(\rho_\infty; \mathbb{R}^d)$.

Remark 4.4 (An equivalent formulation of the contraction estimate)**.** We can give an equivalent version of (4.5) by keeping fixed the cost but rescaling the measures. In fact, we can associate to the solutions ρ^1, ρ^2 of (4.2) their rescaled versions $\tilde\rho^1, \tilde\rho^2$ defined by

$$\tilde\rho^j(E) := \rho^j(\mathrm{e}^{-\lambda t} E) \quad \text{for every Borel set } E \subset \mathbb{R}^d,\ j = 1, 2.$$

Then $\tilde\rho^j$ is the push-forward of ρ^j through the map $x \mapsto \mathrm{e}^{\lambda t} x$ and satisfies the change-of-variables formula

$$\int_{\mathbb{R}^d} \zeta(y) \,\mathrm{d}\tilde\rho^j(y) = \int_{\mathbb{R}^d} \zeta(\mathrm{e}^{\lambda t} x) \,\mathrm{d}\rho^j(x) \quad \text{for every } \zeta \in C_\mathrm{b}(\mathbb{R}^d). \tag{4.10}$$

Inequality (4.5) is then equivalent to

$$\mathcal{C}_h(\tilde\rho^1_t, \tilde\rho^2_t) \le \mathcal{C}_h(\rho^1_0, \rho^2_0) \quad \text{for every } t > 0.$$

Strategy of the proof: Kantorovich duality and a variable-doubling technique. In order to prove Theorem 4.1 we develop a new strategy, generalizing [40]. It relies on the dual Kantorovich formulation (see §2.3.1) of the transportation cost (4.3):

$$\begin{aligned} \mathcal{C}_h(\rho^1, \rho^2) = \sup\Big\{ \int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\rho^1 + \int_{\mathbb{R}^d} \phi^2 \,\mathrm{d}\rho^2 : \\ \phi^1, \phi^2 \in C_\mathrm{b}(\mathbb{R}^d),\ \phi^1(x_1) + \phi^2(x_2) \le h(|x_1 - x_2|) \Big\}. \end{aligned} \tag{4.11}$$

This formula reduces the estimate of the cost $\mathcal{C}_h(\rho^1_T, \rho^2_T)$ of two solutions of (4.1) at a certain final time T to the estimate of

$$\Sigma(\phi^1, \phi^2; T) := \int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\rho^1_T + \int_{\mathbb{R}^d} \phi^2 \,\mathrm{d}\rho^2_T \tag{4.12}$$

for an arbitrary pair of functions ϕ^1, ϕ^2 satisfying the constraint

$$\phi^1(x_1) + \phi^2(x_2) \le h(|x_1 - x_2|) \quad \text{for every } x_1, x_2 \in \mathbb{R}^d. \tag{4.13}$$

Assuming for the sake of simplicity that B is monotone, bounded and smooth, we can obtain an estimate of $\Sigma(\phi^1, \phi^2; T)$ by solving the final-value problem for the adjoint equation

$$\partial_t \phi^i + \Delta\phi^i - B \cdot \nabla\phi^i = 0 \quad \text{in } \mathbb{R}^d \times (0, T), \quad \phi^i(\cdot, T) := \phi^i \tag{4.14}$$

since the distributional formulation (4.2) yields

$$\Sigma(\phi^1_T, \phi^2_T; T) = \Sigma(\phi^1_0, \phi^2_0; 0)$$

The following crucial result, based on a “variable-doubling technique”, provides the final step, showing that ϕ^1_0, ϕ^2_0 still satisfy the constraint (4.13) so that $\Sigma(\phi^1_0, \phi^2_0; 0) \le \mathcal{C}_h(\rho^1_0, \rho^2_0)$.

Theorem 4.5. *If $\phi^1, \phi^2 \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [0,T])$ are solutions of (4.14) in the case when B is monotone, bounded and smooth, such that*

$$\phi^1(x_1, T) + \phi^2(x_2, T) \le h(|x_1 - x_2|) \quad \forall\, x_1, x_2 \in \mathbb{R}^d,$$

then

$$\phi^1(x_1, 0) + \phi^2(x_2, 0) \le h(|x_1 - x_2|) \quad \forall\, x_1, x_2 \in \mathbb{R}^d.$$

Remark 4.6. While we prove Theorem 4.5 for bounded and smooth drifts B, and solutions $\phi^{1,2} \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [0,T])$, the property clearly carries over to any pointwise limit of such solutions. We therefore expect it to hold for a much larger class of monotone drifts B and solutions.

4.2 Regularization and rescaling

In this section we collect some preliminary and technical results which will turn to be useful in the sequel.

4.2.1 Regularization of the cost function.

In this section we shall show that it is sufficient to consider nonnegative, Lipschitz, and unbounded costs (as those considered in Proposition 2.19) in the proof of Theorem 4.1.

Lemma 4.7. *If (4.5) holds for every nonnegative Lipschitz and nondecreasing cost function h with $\lim_{r\uparrow+\infty} h(r) = +\infty$, then it holds for every continuous and nondecreasing cost h.*

Proof. We first prove that it is sufficient to consider nonnegative Lipschitz costs; in a second step, we deal with the asymptotic requirement.
Step 1: *h Lipschitz.* Adding a suitable constant we can assume that $h(r) \ge h(0) = 0$. We can then approximate h from below by the increasing sequence of nonnegative Lipschitz functions

$$h^n(r) := \inf_{s\ge 0} h(s) + n|r - s|$$

which satisfies

$$0 = h^n(0) \le h^n(r) \le h(r), \qquad \lim_{n\uparrow+\infty} h^n(r) = h(r) \quad \forall\, r \ge 0,$$

the convergence being uniform on each compact interval of $[0, +\infty)$. Applying Lemma 4.8 below we find

$$\mathcal{C}_{h_{\lambda\,t}}(\rho^1_t, \rho^2_t) \overset{(4.16)}{=} \lim_{n\uparrow+\infty} \mathcal{C}_{h^n_{\lambda\,t}}(\rho^1_t, \rho^2_t) \overset{(4.5)}{\le} \liminf_{n\uparrow+\infty} \mathcal{C}_{h^n}(\rho^1_0, \rho^2_0) \overset{(4.16)}{=} \mathcal{C}_h(\rho^1_0, \rho^2_0).$$

Step 2: $\lim_{r\uparrow+\infty} h(r) = +\infty$. Let us set $\rho_0 := \rho_0^1 + \rho_0^2$, let us introduce the function

$$m(r) := \rho_0(\mathbb{R}^d \setminus r\,U), \quad U := \big\{x \in \mathbb{R}^d : |x| < 1\big\},$$

and let us consider a sequence r_n in $[0, +\infty)$ such that

$$r_0 := 0, \quad r_1 := 1, \quad r_{n+1} - r_n \geq r_n - r_{n-1} \quad \text{and } m(r_{n+1}) \leq 2^{-n}.$$

It is easy to check that r_n is a diverging increasing sequence; if g is the piecewise linear function satisfying $g(r_n) = n$, i.e.

$$g(r) := n + \frac{r - r_n}{r_{n+1} - r_n} \quad \text{if } r \in [r_n, r_{n+1}],$$

then g is Lipschitz continuous, increasing, unbounded, concave, it satisfies $g(0) = 0$ and

$$\begin{aligned}
G &:= \int_{\mathbb{R}^d} g(|x|)\,\mathrm{d}\rho_0(x) = \int_{\mathbb{R}^d} \Big(\int_0^{|x|} g'(r)\,\mathrm{d}r\Big)\,\mathrm{d}\rho_0(x) \\
&= \int_{\mathbb{R}^d} \Big(\int_0^{+\infty} g'(r)\mathbb{1}_{r\leq|x|}\,\mathrm{d}r\Big)\,\mathrm{d}\rho_0(x) = \int_0^{\infty} g'(r)\,m(r)\,\mathrm{d}r \\
&= \sum_{n=1}^{+\infty} \frac{1}{r_n - r_{n-1}} \int_{r_{n-1}}^{r_n} m(r)\,\mathrm{d}r \leq \sum_{n=0}^{+\infty} m(r_n) < +\infty.
\end{aligned}$$

We can thus consider the perturbed cost

$$h^\varepsilon(r) := h(r) + \varepsilon\,g(r)$$

which is Lipschitz, increasing, unbounded. Since g is concave, increasing, and $g(0) = 0$, we have

$$g(|x_1 - x_2|) \leq g(|x_1| + |x_2|) \leq g(|x_1|) + g(|x_2|) \quad \text{for every } x_1, x_2 \in \mathbb{R}^d, \tag{4.15}$$

so that if $\boldsymbol{\rho}_0$ is an optimal coupling between ρ_0^1 and ρ_0^2 for the cost h (we can assume that the initial cost is finite), then

$$\begin{aligned}
\mathcal{C}_h(\rho_0^1, \rho_0^2) &\leq \mathcal{C}_{h^\varepsilon}(\rho_0^1, \rho_0^2) \leq \mathcal{C}_h(\rho_0^1, \rho_0^2) + \varepsilon \int_{\mathbb{R}^d\times\mathbb{R}^d} g(|x_1 - x_2|)\,\mathrm{d}\boldsymbol{\rho}_0(x_1, x_2) \\
&\overset{(4.15)}{\leq} \mathcal{C}_h(\rho_0^1, \rho_0^2) + \varepsilon \int_{\mathbb{R}^d\times\mathbb{R}^d} \Big(g(|x_1|) + g(|x_2|)\Big)\,\mathrm{d}\boldsymbol{\rho}_0(x_1, x_2) = \mathcal{C}_h(\rho_0^1, \rho_0^2) + \varepsilon G.
\end{aligned}$$

Therefore, if Theorem 4.1 holds for h^ε we have

$$\mathcal{C}_h(\rho_t^1, \rho_t^2) \leq \mathcal{C}_{h^\varepsilon}(\rho_t^1, \rho_t^2) \leq \mathcal{C}_{h^\varepsilon}(\rho_0^1, \rho_0^2) \leq \mathcal{C}_h(\rho_0^1, \rho_0^2) + \varepsilon G.$$

Passing to the limit as $\varepsilon \downarrow 0$ we conclude. □

The following result provides a variant of well known stability properties of transportation costs (see [46, Theorem 3], [52, Theorem 5.20]) and holds the much more general setting of optimal transportation in Radon metric spaces [2, Chapter 6].

Lemma 4.8. (Lower semicontinuity of the cost functional w.r.t. local uniform convergence of h). *Let $h : [0, +\infty) \to [0, +\infty)$ be a continuous cost function and let $h^n : [0, +\infty) \to [0, +\infty)$ be a sequence of lower semicontinuous functions converging to h locally uniformly in $[0, +\infty)$. For every couple $\rho^1, \rho^2 \in \mathcal{P}(\mathbb{R}^d)$ we have*

$$\liminf_{n\uparrow+\infty} \mathcal{C}_{h^n}(\rho^1, \rho^2) \geq \mathcal{C}_h(\rho^1, \rho^2).$$

In particular, if $h^n \leq h$ for every $n \in \mathbb{N}$ then

$$\lim_{n\to+\infty} \mathcal{C}_{h^n}(\rho^1, \rho^2) = \mathcal{C}_h(\rho^1, \rho^2). \tag{4.16}$$

Proof. Let us set $H^n(x_1, x_2) := h^n(|x_1 - x_2|)$ and observe that H^n converges to $H(x_1, x_2) := h^n(|x_1 - x_2|)$ uniformly on compact sets of $\mathbb{R}^d \times \mathbb{R}^d$. If $\boldsymbol{\rho}_n \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is an optimal coupling between ρ^1, ρ^2 with respect to the cost h^n then

$$\mathcal{C}_{h^n}(\rho^1, \rho^2) = \int_{[0,+\infty)} z \,\mathrm{d}\rho_n(z), \quad \text{where } \rho_n = (H^n)_\# \boldsymbol{\rho}_n.$$

Since the marginals of $\boldsymbol{\rho}_n$ are fixed, the sequence $(\boldsymbol{\rho}_n)_{n\in\mathbb{N}}$ is tight and up to the extraction of a suitable subsequence (still denoted by $\boldsymbol{\rho}_n$) we can suppose that $\boldsymbol{\rho}_n$ converge to to some limit coupling $\boldsymbol{\rho}$ between ρ^1, ρ^2 in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. Since ρ_n weakly converges to $\rho = H_\# \boldsymbol{\rho}$ by [2, Lemma 5.2.1], standard lower semicontinuity of integrals with nonnegative continuous integrands [2, Lemma 5.1.7] yields

$$\liminf_{n\to+\infty} \int_{[0,+\infty)} z \,\mathrm{d}\rho_n(z) \geq \int_{[0,+\infty)} z \,\mathrm{d}\rho(z) =$$

$$= \int_{\mathbb{R}^d\times\mathbb{R}^d} H(x_1, x_2) \,\mathrm{d}\boldsymbol{\rho}(x_1, x_2) \geq \mathcal{C}_h(\rho^1, \rho^2). \qquad \square$$

4.2.2 λ-monotonicity and rescaling

We show here a simple rescaling argument (inspired by [23], where the rescaling technique has been applied to a wide class of diffusion equations), which is useful to deduce the estimates in the general λ-monotone case to the simpler case of a monotone operator.

We therefore assume that $\lambda \neq 0$, and we introduce the time rescaling functions

$$\mathsf{s}(t) := \int_0^t \mathrm{e}^{2\lambda r} \,\mathrm{d}r = \frac{1}{2\lambda}\big(\mathrm{e}^{2\lambda t} - 1\big), \quad \mathsf{t}(s) := \frac{1}{2\lambda} \log(1 + 2\lambda s) \quad s \in [0, S_\infty) \tag{4.17}$$

where

$$S_\infty := \begin{cases} +\infty & \text{if } \lambda > 0, \\ -1/(2\lambda) & \text{if } \lambda < 0; \end{cases} \qquad \text{notice that } \mathsf{t}\big(\mathsf{s}(t)\big) = t. \tag{4.18}$$

We associate to a family of probability measures ρ_t, $t \in [0, T]$, their rescaled versions σ_s, $s \in [0, S_\infty)$, defined by

$$\sigma_s(E) := \rho_{\mathsf{t}(s)}(\mathrm{e}^{-\lambda \mathsf{t}(s)} E) \quad \text{for every Borel set } E \subset \mathbb{R}^d. \tag{4.19}$$

If $B : \mathbb{R}^d \to \mathbb{R}^d$ is a λ-monotone Borel map we set $A := B - \lambda I$ and

$$\tilde{B}(y,s) := \mathrm{e}^{-\lambda \mathsf{t}(s)} B(\mathrm{e}^{-\lambda \mathsf{t}(s)}\, y), \quad \tilde{A}(y,s) = \mathrm{e}^{-\lambda \mathsf{t}(s)} A(\mathrm{e}^{-\lambda \mathsf{t}(s)}\, y) \tag{4.20}$$

for $y \in \mathbb{R}^d$, $s \in \mathbb{R}$. Notice that if B is λ-monotone, then A and $\tilde{A}(\cdot, s)$, $s \in [0, S_\infty)$, are monotone.

Proposition 4.9. *A continuous family $\rho_t \in \mathcal{P}(\mathbb{R}^d)$ is a distributional solution of (4.2) if and only if the rescaled measures σ_s defined by (4.19) and (4.17) satisfy*

$$\int_0^{S_\infty} \int_{\mathbb{R}^d} \Big(\partial_s \varphi + \Delta\varphi - \tilde{A}(\cdot, s)\cdot\nabla\varphi\Big)\, \mathrm{d}\sigma_s\, \mathrm{d}s = 0 \quad \forall\, \varphi \in C^\infty_{\mathrm{c}}(\mathbb{R}^d \times (0, S_\infty)). \tag{4.21}$$

If ρ satisfies (4.4) then

$$\int_{s_0}^{s_1} \int_{\mathbb{R}^d} |\tilde{A}(x,s)|\, \mathrm{d}\sigma_s\, \mathrm{d}s < +\infty \quad \textit{for every } 0 < s_0 < s_1 < S_\infty, \tag{4.22}$$

and in this case σ satisfies

$$\int_{\mathbb{R}^d} \varphi(\cdot, s_1)\, \mathrm{d}\sigma_{s_1} - \int_{\mathbb{R}^d} \varphi(\cdot, s_0)\, \mathrm{d}\sigma_{s_0} = \int_{s_0}^{s_1} \int_{\mathbb{R}^d} \Big(\partial_s \varphi + \Delta\varphi - \tilde{A}(y,s)\cdot\nabla\varphi\Big)\, \mathrm{d}\rho_s\, \mathrm{d}s. \tag{4.23}$$

for every test function $\varphi \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [s_0, s_1])$ with bounded first and second derivatives.

Proof. We introduce the change of variable map $\mathsf{X}(x,t) := (\mathrm{e}^{\lambda t} x, \mathsf{s}(t))$ and for a given smooth function $\varphi \in C^\infty_{\mathrm{c}}(\mathbb{R}^d \times (0, S_\infty))$ we set $\zeta(x,t) := \varphi(\mathrm{e}^{\lambda \mathsf{s}(t)}, \mathsf{s}(t)) = \varphi \circ \mathsf{X}$. Denoting by $(y,s) \in \mathbb{R}^d \times [0, S_\infty)$ the new variables, easy calculations show that in $\mathbb{R}^d \times (0, +\infty)$ we have

$$\partial_t \zeta = \mathsf{s}'\Big(\partial_s\varphi + \lambda \mathrm{e}^{-2\lambda t}\nabla_y\varphi\cdot y\Big)\circ \mathsf{X}, \qquad \nabla_x\zeta = \mathrm{e}^{\lambda t}\,\big(\nabla_y\varphi\circ\mathsf{X}\big),$$
$$\Delta_x\zeta = \mathrm{e}^{2\lambda t}\,\big(\Delta_y\varphi\circ\mathsf{X}\big), \qquad B\cdot\nabla_x\zeta = \mathrm{e}^{2\lambda t}\,\Big(\big(\tilde{B}(y,s)\cdot\nabla_y\varphi\big)\circ\mathsf{X}\Big),$$

where we used the fact that $B = \mathrm{e}^{\lambda t}\big(\tilde{B}\circ\mathsf{X}\big)$. In particular we have

$$\partial_t\zeta - B\cdot\nabla_x\zeta = \mathsf{s}'\Big(\partial_s\varphi - \tilde{A}(y,s)\cdot\nabla_y\varphi\Big)\circ\mathsf{X}$$

We thus get

$$\begin{aligned}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta_x\zeta - B\cdot\nabla_x\zeta\Big)\,\mathrm{d}\rho_t &= \mathsf{s}'(t)\int_{\mathbb{R}^d}\Big(\partial_s\varphi + \Delta_y\varphi - \tilde{A}(y,s)\cdot\nabla_y\varphi\Big)\circ\mathsf{X}\,\mathrm{d}\rho_t\\ &= \mathsf{s}'(t)\int_{\mathbb{R}^d}\Big(\partial_s\varphi + \Delta_y\varphi - \tilde{A}(y,s)\cdot\nabla_y\varphi\Big)\,\mathrm{d}\sigma_{\mathsf{s}(t)}\end{aligned}$$

since $\sigma_{\mathsf{s}(t)}(E) = \rho_t(\mathrm{e}^{-\lambda t}\, E)$ for every Borel set $E \subset \mathbb{R}^d$. Eventually we obtain

$$\int_0^{+\infty}\int_{\mathbb{R}^d}\Big(\partial_t\zeta + \Delta_x\zeta - B\cdot\nabla_x\zeta\Big)\,\mathrm{d}\rho_t\,\mathrm{d}t = \int_0^{S_\infty}\int_{\mathbb{R}^d}\Big(\partial_s\varphi + \Delta_y\varphi - \tilde{A}(y,s)\cdot\nabla_y\varphi\Big)\,\mathrm{d}\sigma_s\,\mathrm{d}s$$

(4.22) follows by a simple application of the change of variable formula (4.10), since for every $t > 0$

$$\int_{\mathbb{R}^d} \big|\tilde{A}(y,s)\big| \,\mathrm{d}\sigma_s(y) \overset{(4.20)}{=} \mathrm{e}^{-\lambda \mathsf{t}(s)} \int_{\mathbb{R}^d} \big|A(\mathrm{e}^{-\lambda \mathsf{t}(s)} y)\big| \,\mathrm{d}\sigma_s(y)$$
$$\overset{(4.19)}{=} \mathrm{e}^{-\lambda \mathsf{t}(s)} \int_{\mathbb{R}^d} \big|A(x)\big| \,\mathrm{d}\rho_{\mathsf{t}(s)}(x) = \mathrm{e}^{-\lambda \mathsf{t}(s)} \int_{\mathbb{R}^d} \big|B(x) - \lambda x\big| \,\mathrm{d}\rho_{\mathsf{t}(s)}(x).$$

Since $\mathsf{t}'(s) = \mathrm{e}^{-\lambda \mathsf{t}(s)}$ we eventually get for $t_i = \mathsf{t}(s_i)$

$$\int_{s_0}^{s_1} \int_{\mathbb{R}^d} |\tilde{A}(x,s)| \,\mathrm{d}\sigma_s \,\mathrm{d}s = \int_{s_0}^{s_1} \Big(\int_{\mathbb{R}^d} \big|B(x) - \lambda x\big| \,\mathrm{d}\rho_{\mathsf{t}(s)}(x) \Big) \mathsf{t}'(s) \,\mathrm{d}s$$
$$= \int_{t_0}^{t_1} \int_{\mathbb{R}^d} \big|B(x) - \lambda x\big| \,\mathrm{d}\rho_t(x) \,\mathrm{d}t \overset{(4.4)}{<} +\infty.$$

(4.23) follows from (4.21) when φ belongs to $C^\infty_{\mathrm{c}}(\mathbb{R}^d \times [s_0, s_1])$. If $\varphi \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [s_0, s_1])$ via a standard convolution and truncation argument we find an approximation sequence $\varphi_k \in C^\infty_{\mathrm{c}}(\mathbb{R}^d \times [s_0, s_1])$ such that $\varphi_k, \partial_t \varphi_k, \nabla \varphi_k, \Delta \varphi_k$ remains uniformly bounded and converge pointwise to $\varphi, \partial_t \varphi, \nabla \varphi, \Delta \varphi$ respectively. By (4.22) we can apply the Lebesgue Dominated Convergence theorem to pass to the limit in (4.23) written for φ_k, thus obtaining the same identity for φ. □

We conclude this section by a simple remark combining the regularization technique of § 2.1.3 and the time rescaling (4.20).

Lemma 4.10. *Let $A := B - \lambda I$ be a monotone operator, let us consider a sequence $A_{n,m}$, $n, m \in \mathbb{N}$, of smooth monotone operators given by Theorem 2.7 and Proposition 2.8, and let us set*

$$\tilde{A}_{n,m}(y,s) := \mathrm{e}^{-\lambda \mathsf{t}(s)} A_{n,m}(\mathrm{e}^{-\lambda \mathsf{t}(s)} y) \quad y \in \mathbb{R}^d,\ s \in [0, S_\infty) \tag{4.24}$$

defined as in (4.20), (4.17). *Then $\tilde{A}_{n,m}$ are Lipschitz in $\mathbb{R}^d \times [0, S]$ for every $S \in [0, S_\infty)$.*

Proof. We just have to check that $|\partial_s \tilde{A}_{n,m}(\cdot, s)|$ is uniformly bounded in $\mathbb{R}^d \times [0, S]$: since $\mathsf{t}'(s) = \mathrm{e}^{-2\lambda \mathsf{t}(s)}$ a simple calculation yields

$$\partial_s \tilde{A}_{n,m}(y,s) = -\lambda \mathrm{e}^{-3\lambda \mathsf{t}(s)} A_{n,m}(\mathrm{e}^{-\lambda \mathsf{t}(s)} y) - \lambda \mathrm{e}^{-3\lambda \mathsf{t}(s)} \mathrm{D} A_{n,m}(\mathrm{e}^{-\lambda \mathsf{t}(s)} y) \cdot y$$
$$= -\lambda \mathrm{e}^{-2\lambda \mathsf{t}(s)} \tilde{A}_{n,m}(y,s) - \lambda \mathrm{e}^{-\lambda \mathsf{t}(s)} \tilde{Q}_{n,m}(y,s)$$

where $Q_{n,m}(x) := \mathrm{D} A_{n,m}(x) \cdot x, \quad x \in \mathbb{R}^d$.

Since $\mathrm{e}^{-\lambda \mathsf{t}(s)}$ is uniformly bounded with all its derivative in each compact interval $[0, S]$, $S < \infty$, (2.12) show that $Q_{n,m}$ is bounded and therefore $\tilde{A}_{n,m}$ is Lipschitz with respect to s. □

4.3 A comparison result for the backward equation

In this section we give the proof of Theorem 4.5 in a slightly more general form, in order to be applied to (a suitably regularized version of) the rescaled formulation considered in Proposition 4.9.

Let us suppose that $\tilde{A} : (y, s) \in \mathbb{R}^d \times [0, S_\infty) \to \tilde{A}(y, s) \in \mathbb{R}^d$ is a smooth vector field satisfying

$$\sup_{\mathbb{R}^d \times [0,S]} |\tilde{A}_s| + |\partial_s \tilde{A}| + |\mathrm{D}\tilde{A}| < +\infty \quad \text{for every } S \in [0, S_\infty), \tag{4.25}$$

$$\tilde{A}(\cdot, s) \text{ is monotone for every } s \in [0, S_\infty). \tag{4.26}$$

We denote by $\mathscr{L}[\cdot]$ the differential operator defined by

$$\mathscr{L}[\varphi](y, s) := \Delta_y \varphi(y, s) - \tilde{A}(y, s) \cdot \nabla_y \varphi(y, s) \quad \varphi(\cdot, s) \in C^2(\mathbb{R}^d), \tag{4.27}$$

for $y \in \mathbb{R}^d$, $s \in [0, S_\infty)$. Thanks to (4.25) and (4.26), we can apply the existence result [50, Theorem 3.2.1] and for every $S \in [0, S_\infty)$ and $\phi \in C^\infty_{\mathrm{c}}(\mathbb{R}^d)$ we can find a solution $\varphi \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [0, S))$ of the backward evolution equation

$$\partial_s \varphi + \mathscr{L}[\varphi] = 0 \quad \text{in } \mathbb{R}^d \times [0, S], \quad \varphi(\cdot, S) = \phi(\cdot). \tag{4.28}$$

We have

Theorem 4.11. *Let $h : [0, +\infty) \to \mathbb{R}$ be a continuous and non-decreasing function. Let $\varphi^1, \varphi^2 \in C^{2,1}_{\mathrm{b}}(\mathbb{R}^d \times [0, S])$ be solutions of the "backward" inequality*

$$\partial_s \varphi + \mathscr{L}[\varphi] \geq 0 \quad \textit{in } \mathbb{R}^d \times [0, S] \tag{4.29}$$

such that

$$\varphi^1(y_1, S) + \varphi^2(y_2, S) \leq h(|y_1 - y_2|) \quad \textit{for every } y_1, y_2 \in \mathbb{R}^d.$$

Then

$$\varphi^1(y_1, 0) + \varphi^2(y_2, 0) \leq h(|y_1 - y_2|) \quad \textit{for every } y_1, y_2 \in \mathbb{R}^d.$$

Proof. By approximating h from above, it is not restrictive to assume that $h \in C^1[0, +\infty)$ with $h'(0) = 0$; in particular the map $H(y_1, y_2) := h(|y_1 - y_2|)$ is of class C^1 in $\mathbb{R}^d \times \mathbb{R}^d$ and satisfies

$$\nabla_{y_1} H(y_1, y_2) = -\nabla_{y_2} H(y_1, y_2) = g(y_1, y_2)(y_1 - y_2), \tag{4.30}$$

where

$$0 \leq g(y_1, y_2) = g(y_2, y_1) := \begin{cases} \frac{h'(|y_1 - y_2|)}{|y_1 - y_2|} & \text{if } y_1 \neq y_2, \\ 0 & \text{if } y_1 = y_2. \end{cases} \tag{4.31}$$

The argument combines a variable-doubling technique and a classical variant of the maximum principle. Let us first show that if φ^1, φ^2 satisfy the *strict* inequality

$$\partial_s \varphi^j + \mathscr{L}[\varphi^j] > 0 \quad \text{in } \mathbb{R}^d \times [0, S), \quad j = 1, 2. \tag{4.32}$$

then the function

$$f(y_1, y_2, s) := \varphi^1(y_1, s) + \varphi^2(y_2, s) - H(y_1, y_2)$$

cannot attains a (local) maximum in a point $(\bar{y}_1, \bar{y}_2, \bar{s})$ with $\bar{s} < S$. We argue by contradiction and we suppose that $(\bar{y}_1, \bar{y}_2, \bar{s})$ is a local maximizer of f with $\bar{s} < S$; we thus have

$$\partial_s f(\bar{y}_1, \bar{y}_2, \bar{s}) \leq 0, \quad \nabla_{y_1} f(\bar{y}_1, \bar{y}_2, \bar{s}) = 0, \quad \nabla_{y_2} f(\bar{y}_1, \bar{y}_2, \bar{s}) = 0;$$

so that

$$\partial_t \varphi^1(\bar y_1, \bar s) + \partial_t \varphi^2(\bar y_2, \bar s) \le 0 \tag{4.33}$$

$$\nabla_{y_1} \varphi^1(\bar y_1, \bar s) = \nabla_{y_1} H(\bar y_1, \bar y_2) \overset{(4.30)}{=} g(\bar y_1, \bar y_2)(y_1 - y_2)$$
$$\nabla_{y_2} \varphi^2(\bar y_2, \bar s) = \nabla_{y_2} H(\bar y_1, \bar y_2) \overset{(4.30)}{=} g(\bar y_1, \bar y_2)(y_2 - y_1).$$

It follows that

$$\begin{aligned}\tilde A(\bar y_1, \bar s) \cdot \nabla_{y_1} \varphi^1(\bar y_1, \bar s) &+ \tilde A(\bar y_2, \bar s) \cdot \nabla_{y_2} \varphi^2(\bar y_2, \bar s) \\ &= g(\bar y_1, \bar y_2)\big(\tilde A(\bar y_1, \bar s) - \tilde A(\bar y_2, \bar s)\big) \cdot (\bar y_1 - \bar y_2) \overset{(4.31)}{\ge} 0\end{aligned} \tag{4.34}$$

On the other hand, since $H(\bar y_1 + z, \bar y_2 + z) = H(\bar y_1, \bar y_2)$, the function

$$\mathbb{R}^d \ni z \mapsto \varphi^1(\bar y_1 + z, \bar s) + \varphi^2(\bar y_2 + z, \bar s) - H(\bar y_1, \bar y_2) = f(\bar y_1 + z, \bar y_2 + z, \bar s)$$

has a local maximum at $z = 0$ so that

$$\Delta_{y_1} \varphi^1(\bar y_1, \bar s) + \Delta_{y_2} \varphi^2(\bar y_2, \bar s) \le 0. \tag{4.35}$$

Combining (4.33),(4.34), and (4.35) we obtain

$$(\partial_s \varphi^1 + \mathscr{L}[\varphi^1])(\bar y_1, \bar s) + (\partial_s \varphi^2 + \mathscr{L}[\varphi^2])(\bar y_2, \bar s) \le 0,$$

which contradicts (4.32).

Suppose now that φ^1, φ^2 satisfy the inequality (4.29) and let us set for $\varepsilon, \delta > 0$

$$\varphi^j_{\varepsilon,\delta}(y_j, s) := \varphi^j(y_j, s) - \delta(S - s) - \varepsilon \mathrm{e}^{-s}|y_j|^2 \quad j = 1, 2.$$

We easily get

$$\begin{aligned}\partial_s \varphi^j_{\varepsilon,\delta} &= \partial_s \varphi^j + \delta + \varepsilon \mathrm{e}^{-s}|y_j|^2 \\ \mathscr{L}[\varphi^j_{\varepsilon,\delta}] &= \mathscr{L}[\varphi^j] - \mathrm{e}^{-s}\big(d\varepsilon + 2\varepsilon \tilde A(y_j, s) \cdot y_j\big) \\ \partial_s \varphi^j_{\varepsilon,\delta} + \mathscr{L}[\varphi^j_{\varepsilon,\delta}] &\ge \delta + \varepsilon \mathrm{e}^{-s}\big(|y_j|^2 - d - C_n|y_j|\big),\end{aligned}$$

where $C_n = \sup_{y,s} |\tilde A_n(y, s)| < +\infty$.

It follows that for every $\delta > 0$ there exists a coefficient $\varepsilon > 0$ sufficiently small such that $\varphi^1_{\varepsilon,\delta}, \varphi^2_{\varepsilon,\delta}$ satisfy (4.32). On the other hand, the continuous function $f_{\varepsilon,\delta} : \mathbb{R}^d \times \mathbb{R}^d \times [0, S] \to \mathbb{R}$ defined by

$$(y_1, y_2, s) \mapsto f_{\varepsilon,\delta}(y_1, y_2, s) := \varphi^1_{\varepsilon,\delta}(y_1, s) + \varphi^2_{\varepsilon,\delta}(y_2, s) - h(|y_1 - y_2|)$$

attains its maximum at some point $(\bar y_1, \bar y_2, \bar s) \in \mathbb{R}^d \times \mathbb{R}^d \times [0, S]$; by the previous argument, we conclude that $\bar s = S$ and therefore for every $y_1, y_2 \in \mathbb{R}^d$

$$\begin{aligned}\varphi^1_{\varepsilon,\delta}(y_1, 0) + \varphi^2_{\varepsilon,\delta}(y_2, 0) - h(|y_1 - y_2|) &\le f_{\varepsilon,\delta}(\bar y_1, \bar y_2, S) \le \\ &\le \varphi^1(\bar y_1, S) + \varphi^2(\bar y_2, S) - h(|\bar y_1 - \bar y_2|) \le 0.\end{aligned}$$

Passing to the limit as $\varepsilon, \delta \downarrow 0$ we conclude. □

We conclude this section by recalling two well known estimates:

Lemma 4.12 (Uniform estimates). *Let $\varphi \in C^{2,1}_{\rm b}(\mathbb{R}^d \times [0,S]) \cap C^\infty(\mathbb{R}^d \times (0,S))$ be the solution of* (4.28)*. Then*

$$\sup_{\mathbb{R}^d \times [0,S]} |\varphi| \le \sup_{\mathbb{R}^d} |\phi|, \qquad \sup_{\mathbb{R}^d \times [0,S]} |\nabla\varphi| \le \sup_{\mathbb{R}^d} |\nabla\phi|. \tag{4.36}$$

Proof. The first inequality is direct application of the maximum principle (see e.g. [50, Theorem 3.1.1]. By differentiating the equation with respect to y we obtain

$$\partial_s \mathrm{D}\varphi + \mathscr{L}[\mathrm{D}\varphi] - \mathrm{D}\tilde{A}\,\mathrm{D}\varphi = 0$$

and then

$$\frac{1}{2}\partial_t |\mathrm{D}\varphi|^2 + \frac{1}{2}\mathscr{L}[|\mathrm{D}\varphi|^2] - \mathrm{D}\tilde{A}\mathrm{D}\varphi \cdot \mathrm{D}\varphi - |\mathrm{D}^2\varphi|^2 = 0.$$

Since $\tilde{A}$ is monotone the quadratic form associated to $\mathrm{D}\tilde{A}$ is nonnegative and therefore

$$\partial_t |\mathrm{D}\varphi|^2 + \mathscr{L}[|\mathrm{D}\varphi|^2] \ge 0.$$

A further application of the maximum principle yields (4.36). □

4.4 Proof of Theorem 4.1

We split the proof in various steps. Just to fix some notation, we consider a family $A_{n,m}$ of smooth, bounded, Lipschitz, and monotone operators approximating $A := B - \lambda I$ as in Proposition 2.8 and their rescaled version $\tilde{A}_{n,m}$ defined by (4.24). $\mathscr{L}_{n,m}[\cdot]$ are the associated differential operators

$$\mathscr{L}_{n,m}[\varphi](y,s) := \Delta_y \varphi(y,s) - \tilde{A}_{n,m}(y,s) \cdot \nabla_y \varphi(y,s) \quad \varphi(\cdot,s) \in C^2(\mathbb{R}^d), \tag{4.37}$$

for $y \in \mathbb{R}^d$, $s \in [0, S_\infty)$, as in (4.27). Lemma 4.10 show that $\tilde{A}_{n,m}$ satisfies (4.25).

Step 1: *reduction to the monotone case $\lambda = 0$.* When $\lambda \neq 0$ we apply the rescaling argument of section 4.2.2: we thus introduce the time rescaling $\mathsf{t}(s)$ defined by (4.17) and the corresponding measures $\sigma^i_s = \tilde{\rho}^i_{\mathsf{t}(s)}$ as in (4.19), which satisfy (4.22) and (4.23) for the rescaled operators $\tilde{A}$ of (4.20). Taking into account Remark 4.4 and the fact that $\sigma^i_s = \tilde{\rho}^i_{\mathsf{t}(s)}$, the thesis follows if we show that

$$\mathcal{C}_h(\sigma^1_s, \sigma^2_s) \le \mathcal{C}_h(\sigma^1_0, \sigma^2_0) \quad \text{for every } s \in [0, S_\infty), \tag{4.38}$$

(see (4.18) for the definition of S_∞).

Step 2: *If*

$$\mathcal{C}_h(\sigma^1_{s_1}, \sigma^2_{s_1}) \le \mathcal{C}_h(\sigma^1_{s_0}, \sigma^2_{s_0}) \quad \textit{for every } 0 < s_0 < s_1 < S_\infty, \tag{4.39}$$

then (4.38) *holds.* When h is bounded, (4.39) implies (4.38) by taking a simple limit as $s_0 \downarrow 0$ and using the fact that the map $(\sigma^1, \sigma^2) \mapsto \mathcal{C}_h(\sigma^1, \sigma^2)$ is continuous with respect to weak convergence in $\mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$. If (4.38) holds for every bounded Lipschitz cost, then it holds for every continuous and nondecreasing cost by Lemma 4.7.

Step 3: We claim the following:

Let $\phi^1, \phi^2 \in C^\infty_{\rm c}(\mathbb{R}^d)$ be satisfying the constraint $\phi^1(y_1) + \phi^2(y_2) \le h(|y_1 - y_2|)$ Then

$$\int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\sigma^1_{s_1} + \int_{\mathbb{R}^d} \phi^2 \sigma^2_{s_1} \le \mathcal{C}_h(\sigma^1_{s_0}, \sigma^2_{s_0}) + \ell\, K_{n,m} \tag{4.40}$$

where $\ell := \sup_{\mathbb{R}^d} |\nabla\phi^1| + \sup_{\mathbb{R}^d} |\nabla\phi^2|$ and

$$K_{n,m} := \int_{s_0}^{s_1} \int_{\mathbb{R}^d} |\tilde{A}_{n,m} - \tilde{A}| \,\mathrm{d}\sigma^1_s \,\mathrm{d}s + \int_{s_0}^{s_1} \int_{\mathbb{R}^d} |\tilde{A}_{n,m} - \tilde{A}| \,\mathrm{d}\sigma^2_s \,\mathrm{d}s.$$

Indeed, applying [50, Thm 3.2.1] we can introduce the solutions $\varphi^1_{n,m}, \varphi^2_{n,m} \in C^{2,1}_{\rm b}(\mathbb{R}^d \times [s_0, s_1])$ of the backward equations

$$\partial_s \varphi^j_{n,m} + \mathscr{L}_{n,m}[\varphi^j] = 0 \quad \text{in } \mathbb{R}^d \times [s_0, s_1], \quad \varphi^j_{n,m}(\cdot, s_1) = \phi^j(\cdot) \quad \text{in } \mathbb{R}^d.$$

Identity (4.23) shows that, for $j = 1, 2$,

$$\begin{aligned}\int_{\mathbb{R}^d} \phi^j(\cdot) \,\mathrm{d}\sigma^j_{s_1} - \int_{\mathbb{R}^d} \varphi^j_{n,m}(\cdot, s_0) \,\mathrm{d}\sigma^j_{s_0} &= \int_{s_0}^{s_1} \int_{\mathbb{R}^d} \big(\tilde{A}_{n,m} - \tilde{A}\big) \cdot \nabla\varphi^j_{n,m} \,\mathrm{d}\sigma^j_s \,\mathrm{d}s \\ &\overset{(4.36)}{\le} \ell \int_{s_0}^{s_1} \int_{\mathbb{R}^d} \big|\tilde{A}_{n,m} - \tilde{A}\big| \,\mathrm{d}\sigma^j_s \,\mathrm{d}s\end{aligned}$$

Summing up the these equation for $j = 1, 2$ we obtain

$$\int_{\mathbb{R}^d} \phi^1 \,\mathrm{d}\sigma^1_{s_1} + \int_{\mathbb{R}^d} \phi^2 \,\mathrm{d}\sigma^2_{s_1} \le \int_{\mathbb{R}^d} \varphi^1_{n,m}(\cdot, s_0) \,\mathrm{d}\sigma^1_{s_0} + \int_{\mathbb{R}^d} \varphi^2_{n,m}(\cdot, s_0) \,\mathrm{d}\sigma^2_{s_0} + \ell K_{n,m}$$

Theorem 4.11 yields $\varphi^1_{n,m}(y_1, s_0) + \varphi^2_{n,m}(y_2, s_0) \le h(|y_1 - y_2|)$ which implies (4.40).

<u>Step 4:</u>

$$\limsup_{n\uparrow+\infty} \Big(\limsup_{m\uparrow+\infty} K_{n,m} \Big) = 0. \tag{4.41}$$

Let us first notice that setting $t_i := \mathsf{t}(s_i)$ and recalling that $\mathsf{t}'(s) = \mathrm{e}^{-\lambda \mathsf{t}(s)}$ we have

$$\begin{aligned}\int_{s_0}^{s_1} \int_{\mathbb{R}^d} |\tilde{A}_{n,m} - \tilde{A}| \,\mathrm{d}\sigma^1_s \,\mathrm{d}s &= \int_{\mathsf{s}(t_0)}^{\mathsf{s}(t_1)} \mathsf{t}'(s) \int_{\mathbb{R}^d} |A_{n,m} - A| \,\mathrm{d}\rho^i_{\mathsf{t}(s)} \,\mathrm{d}s \\ &= \int_{t_0}^{t_1} \int_{\mathbb{R}^d} |A_{n,m} - A| \,\mathrm{d}\rho^i_t \,\mathrm{d}t\end{aligned}$$

so that

$$K_{n,m} = K^1_{n,m} + K^2_{n,m}, \quad K^j_{n,m} := \int_{t_0}^{t_1} \int_{\mathbb{R}^d} |A_{n,m} - A| \,\mathrm{d}\rho^j_t \,\mathrm{d}t \quad j = 1, 2.$$

We can estimate $K^j_{n,m}$ by

$$K^j_{n,m} \le \int_{t_0}^{t_1} \int_{\mathbb{R}^d} |A_{n,m} - A_n| \,\mathrm{d}\rho^j_t \,\mathrm{d}t + \int_{t_0}^{t_1} \int_{\mathbb{R}^d} |A_n - A| \,\mathrm{d}\rho^j_t \,\mathrm{d}t,$$

observing that by (2.12), (2.13), and the Lebesgue Dominated Convergence Theorem we get

$$\lim_{m\uparrow+\infty} K^j_{n,m} = \int_{t_0}^{t_1}\int_{\mathbb{R}^d} |A_n - A|\,\mathrm{d}\rho^j_t\,\mathrm{d}t.$$

Since $|A_n(x)| \le |\mathsf{A}^\circ(x)| \le |A(x)| = |B(x) - \lambda x|$ for every $x \in \mathbb{R}^d$, the integrability assumption (4.4), a further application of the Lebesgue Theorem, and (2.9) yield

$$\lim_{n\uparrow+\infty}\Big(\lim_{m\uparrow+\infty} K_{n,m}\Big) = \int_{t_0}^{t_1}\int_{\mathbb{R}^d} |\mathsf{A}^\circ - A|\,\mathrm{d}\rho^j_t\,\mathrm{d}t. \tag{4.42}$$

This last integrand is 0 if A coincides with the minimal selection of A, in particular when A is continuous. In the general case, the regularity result of [11] shows that $\rho^j_t \ll \mathscr{L}^d$ for $\mathscr{L}^1$ a.e. $t \in (0,+\infty)$ and (2.6) says that $\mathsf{A}^\circ = A$ $\mathscr{L}^d$-a.e. in $\mathbb{R}^d$; therefore the last integral of (4.42) vanishes and we get (4.41).
Step 5: *conclusion.*

Thanks to (4.41), passing to the limit in (4.40) we obtain

$$\int_{\mathbb{R}^d} \phi^1\,\mathrm{d}\sigma^1_{s_1} + \int_{\mathbb{R}^d} \phi^2\,\mathrm{d}\sigma^2_{s_1} \le \mathcal{C}_h(\sigma^1_{s_0}, \sigma^2_{s_0}).$$

Taking the supremum with respect to $\phi^1, \phi^2 \in C^\infty_{\mathrm{c}}(\mathbb{R}^d)$ and recalling Proposition 2.19 we obtain (4.39).

Remark 4.13. As it appears from the final argument of the previous step 4, in the case when $A = B - \lambda I$ is the minimal selection A° of A (in particular when B is continuous), we do not need to invoke the regularity result of [11] to conclude our proof.

Proof of Corollary 4.2. For (a), it is sufficient to observe that $e^{\lambda t} \ge 1$; this implies $h(r) \le h_{\lambda t}(r)$ and so

$$\mathcal{C}_h(\rho^1_t, \rho^2_t) \le \mathcal{C}_{h_{\lambda t}}(\rho^1_t, \rho^2_t) \overset{(4.5)}{\le} \mathcal{C}_h(\rho^1_0, \rho^2_0).$$

Similarly, for (a) and (b)

$$e^{p\lambda t}\mathcal{C}_h(\rho^1_t, \rho^2_t) \le \mathcal{C}_{h_{\lambda t}}(\rho^1_t, \rho^2_t) \overset{(4.5)}{\le} \mathcal{C}_h(\rho^1_0, \rho^2_0).$$

We conclude recalling that

$$W_p(\rho^1, \rho^2) = \mathcal{C}_h(\rho^1, \rho^2)^{1/p} \text{ with } h(r) = |r|^p$$

and applying (a) and (b). □

Chapter 5

Wasserstein solutions to the Fokker-Planck equation

This chapter is dedicated to the existence of curves in $\mathcal{P}_2(\mathbb{R}^d)$ (distributionally) satisfying
$$\partial_t \rho_t - \Delta \rho_t - \nabla\cdot(\rho_t B) = 0 \text{ in } \mathbb{R}^d \times (0, +\infty) \tag{5.1}$$
where B is a monotone operator with domain $\mathbb{R}^d$. We will provide sufficient conditions to find such solutions. In order to simplify the calculation, we also assume that $B(0) = 0$.

We propose the definition of Wasserstein solution of (5.1) and give the precise statement of the existence. In Section 5.1 we prove the uniqueness of the Wasserstein solution by showing that such solution satisfies the summability condition (4.4). In Section 5.2 we prove the existence when the operator B is Lipschitz continuous by a splitting method approximation scheme. In the last Section we extend the existence result for general B by approximation with Lipschitz operators.

Definition 5.1 (Wasserstein solutions)**.** We call $\rho_t : [0, +\infty[\to \mathcal{P}_2(\mathbb{R}^d)$ a Wasserstein solution of (5.1) if ρ_t is a distributional solution of the continuity equation
$$\partial_t \rho_t + \nabla\cdot(\mathbf{v}_t \rho_t) = 0,$$
where $\mathbf{v}_t : [0, +\infty[\to \mathbb{R}^d$ is a Borel vector field such that
$$\mathbf{v}_t \in L^2(\mathbb{R}^d, \rho_t), \quad \|\mathbf{v}_t\|_{L^2(\mathbb{R}^d,\rho_t)} \in L^1_{loc}([0, +\infty[)$$
satisfying
$$\int_{\mathbb{R}^d} \mathbf{v}_t \cdot \nabla\varphi \, \mathrm{d}\rho_t = \int_{\mathbb{R}^d} \Big(\Delta\varphi - B \cdot \nabla\varphi\Big) \mathrm{d}\rho_t, \quad \forall\, \varphi \in C^\infty_c(\mathbb{R}^d). \tag{5.2}$$

Remark 5.2. The previous definition is equivalent to (see §2.4)

- $\rho_t : [0, +\infty[\to \mathcal{P}_2(\mathbb{R}^d)$ is absolutely continuous;
- ρ_t solves
$$\int_0^{+\infty} \int_{\mathbb{R}^d} \Big(\partial_t\zeta + \Delta\zeta - B\cdot\nabla\zeta\Big)\, \mathrm{d}\rho_t \mathrm{d}t = 0 \quad \forall\, \zeta \in C^\infty_c(\mathbb{R}^d \times (0, +\infty)). \tag{5.3}$$

In particular a Wasserstein solution is a distributional solution to (5.1).

Lemma 5.3 (Uniqueness)**.** *A Wasserstein solution satisfies the summability condition* (4.4)*, the contraction estimate* (4.5)*, and the quadratic moment estimate*

$$\int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\rho_{t_1} + \int_{t_0}^{t_1} \mathrm{d}s \int_{\mathbb{R}^d} \Big(B(x)\cdot x\Big)\,\mathrm{d}\rho_s = \int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\rho_{t_0} + d(t_1 - t_0), \tag{5.4}$$

for every $0 \le t_0 < t_1 < +\infty$*.*

We can prove the existence of this kind of solution under suitable conditions on the initial datum: this solution is also Lipschitz continuous w.r.t. the L^2-Wasserstein distance and has a metric characterization in terms of operator B and the entropy functional $\mathscr{H} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}\cup\{+\infty\}$

$$\mathscr{H}[\rho] = \begin{cases} \int_{\mathbb{R}^d} u \log u dx \text{ if } \rho = u\mathscr{L}^d \in \mathcal{P}_2^r(\mathbb{R}^d),\\ +\infty \text{ otherwise.}\end{cases}$$

The precise statement is

Theorem 5.4. *For* $\rho_0 \in \mathcal{P}_2^r(\mathbb{R}^d)$ *such that*

$$B \in L^2(\mathbb{R}^d, \rho_0), \quad u \in W^{1,1}_{loc}(\mathbb{R}^d), \quad \tfrac{\nabla u}{u} \in L^2(\mathbb{R}^d,\rho_0) \quad (\rho_0 = u\mathscr{L}^d), \tag{5.5}$$

there exists a unique Wasserstein solution $\rho_t \subset \mathcal{P}_2^r(\mathbb{R}^d)$ *of* (5.1) *in the sense of Definition 5.1 s.t.* $\lim_{t\to 0}\rho_t = \rho_0$*. Moreover,*

1. *(equicontinuity) for every* $t, s \ge 0$*,*

$$W_2(\rho_t,\rho_s) \le |t-s|\Big(\|B\|_{L^2(\mathbb{R}^d,\rho_0)} + \Big\|\tfrac{\nabla u}{u}\Big\|_{L^2(\mathbb{R}^d,\rho_0)}\Big); \tag{5.6}$$

2. *(integral evolution inequality) for every* $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ *with a compact support,*

$$\begin{aligned} \frac{1}{2}W_2^2(\rho_{t_2},\sigma) + \int_{t_1}^{t_2}\mathrm{d}s \int_{\mathbb{R}^d} \langle B\circ \mathbf{t}_{\rho_s}^{\sigma}, \mathbf{i} - \mathbf{t}_{\rho_s}^{\sigma}\rangle \mathrm{d}\rho_s \le \\ \le \frac{1}{2}W_2^2(\rho_{t_1},\sigma) + \int_{t_1}^{t_2} \big(\mathscr{H}[\sigma] - \mathscr{H}[\rho_s]\big)\mathrm{d}s. \end{aligned} \tag{5.7}$$

where $\mathbf{t}_{\rho_s}^{\sigma}$ *is the unique optimal transport map from* ρ_s *to* σ *(see §2.3.2).*

Moreover, if $|B(x)| \le C_0 + C_1|x|$*, for every* $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ *,*

$$\begin{aligned} \frac{1}{2}W_2^2(\rho_{t_2},\sigma) + \int_{t_1}^{t_2}\mathrm{d}s \int_{\mathbb{R}^d} \langle B, \mathbf{i} - \mathbf{t}_{\rho_s}^{\sigma}\rangle \mathrm{d}\rho_s \le \\ \le \frac{1}{2}W_2^2(\rho_{t_1},\sigma) + \int_{t_1}^{t_2} \big(\mathscr{H}[\sigma] - \mathscr{H}[\rho_s]\big)\mathrm{d}s. \end{aligned} \tag{5.8}$$

Remark 5.5. In the sublinear case, condition $B \in L^2(\mathbb{R}^d,\rho)$ is always verified. Notice that (5.8) implies (5.7) by monotonicity of B:

$$\langle B\circ \mathbf{t}_{\rho_s}^{\sigma}, \mathbf{i} - \mathbf{t}_{\rho_s}^{\sigma}\rangle \le \langle B, \mathbf{i} - \mathbf{t}_{\rho_s}^{\sigma}\rangle.$$

Moreover, in the sublinear case, there are no restrictions on σ.

5.1 Uniqueness of the Wasserstein solution

Proof of Lemma 5.3. Let us check that a Wasserstein solution satisfies the condition (4.4) for $\lambda = 0$. Since B is a monotone operator and $Dom(B) = \mathbb{R}^d$, there exists M, $R > 0$ s.t. $\forall x \in \mathbb{R}^d$,

$$|B(x)| \leq \frac{1}{R} B(x) \cdot x + \frac{1}{R} M|x| + M$$

(it follows from Lemma 2.6, choosing $x_0 = 0$ and an arbitrary R). It is enough to control the three terms of the right hand side: the check is easy for the last two ones, recalling that, from the absolutely continuity of the curve,

$$\|x\|^2_{L^2(\rho_t, \mathbb{R}^d)} = \int_{\mathbb{R}^d} |x|^2 \, \mathrm{d}\rho_t \in L^\infty([0, +\infty]).$$

It remains only to check for every $0 < t_0 < t_1 < +\infty$,

$$\int_{t_0}^{t_1} \int_{\mathbb{R}^d} \Big(B(x) \cdot x \Big) \, \mathrm{d}\rho_t(x) \, \mathrm{d}t < +\infty. \tag{5.9}$$

Let us consider a bounded function $\zeta(r) \in C^\infty(\mathbb{R})$ such that

$$|\zeta(r)| \leq r, \quad \zeta(r) = r \text{ in } [-1, 1], \quad \zeta'(r) \geq 0, \quad \zeta'(r) \in C^\infty_{\mathrm{c}}(\mathbb{R}).$$

and let us define $\zeta_n(r) = n\zeta(r/n)$. Thus the sequence $\varphi_n(x) := \zeta_n(|x|^2/2) \in C^\infty_{\mathrm{b}}(\mathbb{R}^d)$ is such that, for $n \to \infty$,

$$\varphi_n(x) \to |x|^2/2, \quad \nabla\varphi_n(x) \to x, \quad \Delta\varphi_n(x) \to \mathrm{d}, \quad \forall x \in \mathbb{R}^d.$$

Denoting $M_1 := \max|\zeta'|$ and $M_2 := \max|\zeta''|$,

$$\varphi_n(x) \leq |x|^2/2, \quad |\nabla\varphi_n(x)| \leq M_1|x|, \quad |\Delta\varphi_n(x)| \leq M_1\mathrm{d} + \tfrac{M_2}{n}|x|^2 \; \forall x \in \mathbb{R}^d :$$

applying the Dominated Convergence theorem,

$$|\varphi_n(x) - |x|^2/2| \xrightarrow{L^1(\rho_t)} 0, \quad |\nabla\varphi_n(x) - x| \xrightarrow{L^2(\rho_t)} 0, \quad |\Delta\varphi_n(x) - \mathrm{d}| \xrightarrow{L^1(\rho_t)} 0.$$

Since $\varphi_n \in H^2(\rho_t)$, there exists a sequence $\psi^k_n \in C^\infty_{\mathrm{c}}(\mathbb{R}^d)$ s.t. $\psi^k_n \xrightarrow{H^2(\rho_t)} \varphi_n$:

$$\begin{aligned}\int_{\mathbb{R}^d} \Big(\Delta\varphi_n - \mathbf{v}_t \cdot \nabla\varphi_n \Big) \, \mathrm{d}\rho_t &= \lim_{k\to\infty} \int_{\mathbb{R}^d} \Big(\Delta\psi^k_n - \mathbf{v}_t \cdot \nabla\psi^k_n \Big) \, \mathrm{d}\rho_t \\ &\overset{(5.2)}{=} \lim_{k\to\infty} \int_{\mathbb{R}^d} \Big(B(x) \cdot \nabla\psi^k_n(x) \Big) \, \mathrm{d}\rho_t(x) = \int_{\mathbb{R}^d} \Big(B(x) \cdot \nabla\varphi_n(x) \Big) \, \mathrm{d}\rho_t(x).\end{aligned} \tag{5.10}$$

Recalling that $B(x) \cdot \nabla\varphi_n(x) = \zeta'_n(|x|^2/2) B(x) \cdot x \geq 0$, from Fatou's Lemma

$$\int_{\mathbb{R}^d} \Big(B(x) \cdot x \Big) \, \mathrm{d}\rho_t(x) \leq \liminf_{n\to\infty} \int_{\mathbb{R}^d} \Big(B(x) \cdot \nabla\varphi_n(x) \Big) \, \mathrm{d}\rho_t(x).$$

According to (5.10),

$$\begin{aligned}\liminf_{n\to\infty} \int_{\mathbb{R}^d} \Big(B(x) \cdot \nabla\varphi_n(x) \Big) \, \mathrm{d}\rho_t(x) &= \lim_{n\to\infty} \int_{\mathbb{R}^d} \Big(\Delta\varphi_n - \mathbf{v}_t \cdot \nabla\varphi_n \Big) \, \mathrm{d}\rho_t \\ &= \mathrm{d} + \int_{\mathbb{R}^d} \Big(\mathbf{v}_t \cdot x \Big) \, \mathrm{d}\rho_t.\end{aligned}$$

Summing up,

$$\int_{\mathbb{R}^d} \Big(B(x)\cdot x\Big)\,\mathrm{d}\rho_t(x) \le \mathrm{d} + \int_{\mathbb{R}^d} \Big(\mathbf{v}_t\cdot x\Big)\,\mathrm{d}\rho_t. \tag{5.11}$$

The right-hand side of the inequality is in $L^1_{loc}([0,+\infty])$: (5.9) is verified. Since (4.4) holds, Theorem 4.1 ensures the contraction property for Wasserstein solutions.

Now we check (5.4). Since ρ_t is a weak solution of (5.1), for $t_2, t_1 > 0$

$$\int_{\mathbb{R}^d} \varphi_n\,\mathrm{d}\rho_{t_1} - \int_{\mathbb{R}^d} \varphi_n\,\mathrm{d}\rho_{t_0} = \int_{t_0}^{t_1} \mathrm{d}s \int_{\mathbb{R}^d} \Big(\Delta\varphi_n - B(x)\cdot\nabla\varphi_n(x)\Big)\,\mathrm{d}\rho_s. \tag{5.12}$$

Since $|B(x)\cdot\nabla\varphi_n(x)| \le M_1 B(x)\cdot x$, the Dominated Convergence theorem applies: letting n go to ∞, we obtain (5.4). □

5.2 The splitting method scheme

Throughout this section, we assume the additional condition on B:

$$B \text{ is Lipschitz continuous } (\Rightarrow B \text{ is maximal monotone }). \tag{5.13}$$

We consider a piecewise constant curve $\rho^\tau : [0,+\infty[\to \mathcal{P}_2(\mathbb{R}^d)$, $(\tau > 0)$ i.e.

$$\rho^{\tau,0} = \rho_0, \quad \rho^\tau_t = \rho^{\tau,n} \in \mathcal{P}_2(\mathbb{R}^d), \quad \forall t \in](n-1)\tau, n\tau],\ n \in \mathbb{N} \tag{5.14}$$

which approximates the solution of (5.1) for $\tau \to 0$. This construction is based on a *splitting method*: any value $\rho^{\tau,n}$ is determined by the previous one in two steps:

$$\widehat{\rho}^{\tau,n} = \widehat{\mathcal{J}_\tau}\rho^{\tau,n-1}, \quad \rho^{\tau,n} = S(\tau)_\# \widehat{\rho}^{\tau,n}. \tag{5.15}$$

The operator $\widehat{\mathcal{J}_\tau}$ is defined by

$$\begin{aligned} \widehat{\mathcal{J}_\tau} : \mathcal{P}_2(\mathbb{R}^d) &\longrightarrow \mathcal{P}_2(\mathbb{R}^d) \\ \nu &\longmapsto \widehat{\nu}(\tau) \end{aligned} \tag{5.16}$$

where $\widehat{\nu}(\tau)$ is the evolution after a time interval via the gradient flow of entropy functional $\mathscr{H}$ from ν (see Example 2.32) and

$$S(t)_{t>0} \text{ is the semigroup generated by } -B \text{ (see §2.1.4).} \tag{5.17}$$

We prove

Theorem 5.6. *Let us assume that B is Lipschitz continuous. For any $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ satisfying condition* (5.5)*, the discrete curve ρ^τ_t defined by* (5.14), (5.15) *(locally) uniformly converges w.r.t. W_2 as $\tau \to 0$ to the unique Wasserstein solution $\rho_t \in \mathcal{P}^r_2(\mathbb{R}^d)$ of* (5.1) *in the sense of Definition 5.1 with initial datum ρ_0. It satisfies inequalities* (5.6) *and* (5.8)*; moreover for every $t \ge 0$,*

$$\mathscr{H}[\rho_t] \le \mathscr{H}[\rho_0] + d\,\mathrm{Lip}(B)t, \tag{5.18}$$

where $\mathrm{Lip}(B)$ *is the Lipschitz constant of the operator B.*

We split the proof in various steps. After noticing some properties of ρ^τ_t, we prove the convergence of the sequence as $\tau \to 0$. Finally, we show that the limit curve is a Wasserstein solution of (5.1).

The discrete curve ρ_t^τ

Let us recall that $\widehat{\mathcal{J}_\tau}$ satisfies the inequalities

$$W_2^2(\widehat{\mathcal{J}_\tau}\nu,\sigma)-W_2^2(\nu,\sigma)\le 2\tau\Big(\mathscr{H}[\sigma]-\mathscr{H}[\widehat{\mathcal{J}_\tau}\nu]\Big)\qquad \text{for any } \sigma\in\mathcal{P}_2(\mathbb{R}^d),\ (5.19)$$

$$\mathscr{H}[\nu]\ge\mathscr{H}[\widehat{\mathcal{J}_\tau}\nu].\tag{5.20}$$

Furthermore,

$$m_2^2(\widehat{\mathcal{J}_\tau}\nu)\le m_2^2(\nu)+2\tau d\qquad \text{(bound of the quadratic moment)};\quad (5.21a)$$

$$W_2(\widehat{\mathcal{J}_\tau}\nu_1,\widehat{\mathcal{J}_\tau}\nu_2)\le W_2(\nu_1,\nu_2)\qquad \text{(contraction property)}.\quad (5.21b)$$

The semigroup $S(t)$ satisfies (2.17a), (2.17b),(2.17d) (see §2.1.4). The interaction between $S(t)$ and $\mathscr{H}$ is ruled by the following

Lemma 5.7. *For any absolute continuous measure $\nu\in\mathcal{P}_2(\mathrm{R^d})$ holds*

$$\mathscr{H}[S(\tau)_\#\nu]\le\mathscr{H}[\nu]+d\ \mathrm{Lip}(B)\tau.\tag{5.22}$$

Proof. The strategy is to prove a similar result involving the resolvent J_τ^B of B and to get (5.22) using the resolvent property

$$|(J^B_{\frac{\tau}{m}})^m x_0-S(\tau)x_0|\le 2\tfrac{\tau}{\sqrt{m}}|B(x_0)|.$$

We recall that $J_\tau^B(x)=(\mathbf{i}+\tau B)^{-1}$. Since B is Lipschitz continuous, J_τ^B is bilipschitz and then $\nu\ll\mathscr{L}^d\Longrightarrow J^B_{\tau\ \#}\nu\ll\mathscr{L}^d$. We fix $h\in(0,1)$ and claim

$$\mathscr{H}[J^B_{h\ \#}\nu]\le\mathscr{H}[\nu]+d\ \mathrm{Lip}(B)h+C_{d,\mathrm{Lip}(B)}h^2.\tag{5.23}$$

By denoting $J^B_{\tau\ \#}\nu=vdx$, $\nu=udx$, we have

$$\begin{aligned}
\mathscr{H}[J^B_{h\ \#}\nu]-\mathscr{H}[\nu]&=\int v\log v dx-\int u\log u dx\\
&=\int u[\log v(J_h^B(x))-\log u]dx\le\int u\frac{1}{u}[v(J_h^B(x))-u]dx=\\
&=\int\Big[v(J_h^B(x))-u\Big]dx\overset{(2.24)}{=}\int\Big[\frac{u(x)}{|det\nabla J_h^B(x)|}-u\Big]dx=\\
&=\int\Big[\frac{1}{|det\nabla J_h^B(x)|}-1\Big]udx=\int\Big[|det\nabla J_h^B(x)|^{-1}-1\Big]udx=\\
&=\int\Big[|det\nabla J_h^{B^{-1}}(J_h^B(x))|-1\Big]udx=\int\Big[|det\nabla J_h^{B^{-1}}(x)|-1\Big]vdx=\\
&=\int\Big[|det\nabla(\mathbf{i}+hB(x))|-1\Big]vdx=\int\Big[|det(I+h\nabla B(x))|-1\Big]vdx\le\\
&\overset{h\in(0,1)}{\le}d\ \mathrm{Lip}(B)h+C_{d,\mathrm{Lip}(B)}h^2\approx O(h).
\end{aligned}$$

On the other hand,

$$(J^B_{\frac{\tau}{m}})^m_{\#}\nu\longrightarrow S(\tau)_\#\nu\ \text{in}\ \mathcal{P}_2(\mathrm{R^d})\ \text{per}\ m\to+\infty;\tag{5.24}$$

in fact,

$$W_2^2((J^B_{\frac{\tau}{m}})^m_{\#}\nu, S(\tau)_{\#}\nu) \le \int_{\mathbb{R}^d} |(J^B_{\frac{\tau}{m}})^m x - S(\tau)x|^2 d\nu(x)$$
$$\overset{(2.18)}{\le} 4\frac{\tau^2}{m}\int_{\mathbb{R}^d} |Bx|^2 d\nu(x) \to 0.$$

For any m sufficiently big, $\tau/m < 1$: applying (5.23) recursively

$$\begin{aligned}\mathscr{H}[(J^B_{\frac{\tau}{m}})^m{}_{\#}\nu] &\le \mathscr{H}[\nu] + m\Big[d\ \mathrm{Lip}(B)\tfrac{\tau}{m} + C_{d,\mathrm{Lip}(B)}(\tfrac{\tau}{m})^2\Big]\\ &= \mathscr{H}[\nu] + d\ \mathrm{Lip}(B)\tau + \tfrac{1}{m}C_{d,\mathrm{Lip}(B)}\tau^2.\end{aligned}$$

From (5.24) and the lower semicontinuity of $\mathscr{H}$,

$$\liminf_{m\to\infty} \mathscr{H}[(J^B_{\frac{\tau}{m}})^m{}_{\#}\nu] \ge \mathscr{H}[S(\tau)_{\#}\nu].$$

Finally,

$$\mathscr{H}[S(\tau)_{\#}\nu] \le \liminf_{m\to\infty} \mathscr{H}[(J^B_{\frac{\tau}{m}})^m{}_{\#}\nu] \le \mathscr{H}[\nu] + d\ \mathrm{Lip}(B)\tau.$$

□

Proposition 5.8. *Let ρ_0 satisfy* (5.5) *and consider the discrete curves ρ^τ defined by* (5.14)*,*(5.15) *and $\widehat{\rho}^\tau : [0,+\infty[\to \mathbb{R}^d$, defined by:*

$$\widehat{\rho}^{\tau,0} = \rho_0 \quad \widehat{\rho}^\tau_t = \widehat{\rho}^{\tau,n}, \quad \forall t \in](n-1)\tau, n\tau],\ n \in \mathbb{N}. \tag{5.25}$$

We have:

1. *("τ-equicontinuity") for $\varepsilon > 0$ and any sufficient small τ,*
$$W_2(\rho^\tau_t, \rho^\tau_s) \le (|t-s|+\tau)\Big(\|B\|_{L^2(\mathbb{R}^d,\rho_0)} + \Big\|\frac{\nabla u}{u}\Big\|_{L^2(\mathbb{R}^d,\rho_0)} + \varepsilon\Big); \tag{5.26}$$

2. *(entropy bound) for every $t \ge 0$,*
$$\mathscr{H}[\rho^\tau_t] \le \mathscr{H}[\rho_0] + d\ \mathrm{Lip}(B)(t+\tau); \tag{5.27}$$

3. *(discrete evolution inequality)*
$$\begin{aligned}&\frac{1}{2}\Big[W_2^2(\rho^{\tau,n},\sigma) - W_2^2(\rho^{\tau,n-1},\sigma)\Big] \le\\ &\quad\le -\tau\int_{\mathbb{R}^d}\langle \mathcal{B}_\tau, \tfrac{S(\tau)+\mathbf{i}}{2} - \mathbf{t}^\sigma_{\widehat{\rho}^{\tau,n}}\rangle \mathrm{d}\widehat{\rho}^{\tau,n} + \tau\Big(\mathscr{H}[\sigma] - \mathscr{H}[\widehat{\rho}^{\tau,n}]\Big).\end{aligned} \tag{5.28}$$

   where $\mathcal{B}_\tau : \mathbb{R}^d \to \mathbb{R}^d$ is the maximal monotone operator $\mathcal{B}_\tau = \frac{\mathbf{i}-S(\tau)}{\tau}$.

Proof. To verify (5.26), we start by considering $t \ge s$, $t \in](n-1)\tau, n\tau]$, $s \in ((k-1)\tau, k\tau]$.

$$W_2(\rho^\tau_t, \rho^\tau_s) = W_2(\rho^{\tau,n}, \rho^{\tau,k}) \le \sum_{j=k+1}^n W_2(\rho^{\tau,j}, \rho^{\tau,j-1}).$$

Thanks to the contraction properties (5.21b) and (2.17a), we have

$$W_2(\rho_t^\tau, \rho_s^\tau) \le (n-k) W_2(\rho^{\tau,1}, \rho_0) \le \frac{|t-s|+\tau}{\tau} W_2(\rho^{\tau,1}, \rho_0). \tag{5.29}$$

It is enough to estimate the last term of (5.29): let us observe

$$\frac{1}{\tau} W_2(\rho^{\tau,1}, \rho_0) \le \frac{1}{\tau}\Big[W_2(\rho^{\tau,1}, \widehat{\rho}^{\tau,1}) + W_2(\widehat{\rho}^{\tau,1}, \rho_0)\Big];$$

about the first term

$$W_2(\rho^{\tau,1}, \widehat{\rho}^{\tau,1}) \le \Big(\int_{\mathbb{R}^d} |S(\tau) - \mathbf{i}|^2 \mathrm{d}\widehat{\rho}^{\tau,1}\Big)^{\frac{1}{2}} \le \tau \Big(\int_{\mathbb{R}^d} |B|^2 \mathrm{d}\widehat{\rho}^{\tau,1}\Big)^{\frac{1}{2}};$$

since $\lim_{\tau\to 0} W_2(\widehat{\rho}^{\tau,1}, \rho_0) = 0$,

$$\lim_{\tau\to 0} \int_{\mathbb{R}^d} |x|^2 \mathrm{d}\widehat{\rho}^{\tau,1} = \int_{\mathbb{R}^d} |x|^2 \mathrm{d}\rho_0;$$

recalling that growth of B is almost linear,

$$\lim_{\tau\to 0} \int_{\mathbb{R}^d} |B|^2 \mathrm{d}\widehat{\rho}^{\tau,1} = \int_{\mathbb{R}^d} |B|^2 \mathrm{d}\rho_0.$$

In particular,

$$\limsup_{\tau\to 0} \frac{1}{\tau} W_2(\rho^{\tau,1}, \widehat{\rho}^{\tau,1}) \le \lim_{\tau\to 0} \Big(\int_{\mathbb{R}^d} |B|^2 \mathrm{d}\widehat{\rho}^{\tau,1}\Big)^{\frac{1}{2}} = \|B\|_{L^2(\mathbb{R}^d, \rho_0)}.$$

On the other hand, thanks to the interpretation of the heat equation as a gradient flow and (2.38),

$$\lim_{\tau\to 0} \frac{W_2(\widehat{\rho}^{\tau,1}, \rho_0)}{\tau} = \Big\|\tfrac{\nabla u}{u}\Big\|_{L^2(\mathbb{R}^d, \rho_0)}.$$

Hence

$$\limsup_{\tau\to 0} \frac{W_2(\rho^{\tau,1}, \rho_0)}{\tau} \le \|B\|_{L^2(\mathbb{R}^d, \rho_0)} + \Big\|\tfrac{\nabla u}{u}\Big\|_{L^2(\mathbb{R}^d, \rho_0)}. \tag{5.30}$$

For (5.27), it is enough to compute, for $t \in](n-1)\tau, n\tau]$

$$\mathscr{H}[\rho_t^\tau] = \mathscr{H}[\rho^{\tau,n}] \le \mathscr{H}[\widehat{\rho}^{\tau,n}] + d\ \mathrm{Lip}(B)\tau \le \mathscr{H}[\rho^{\tau,n-1}] + d\ \mathrm{Lip}(B)\tau.$$

Applying this recursively, we obtain

$$\mathscr{H}[\rho_t^\tau] \le \mathscr{H}[\rho^{\tau,0}] + d\ \mathrm{Lip}(B) n\tau \le \mathscr{H}[\rho_0] + d\ \mathrm{Lip}(B)(t+\tau).$$

We conclude (5.28) by observing that

$$\begin{aligned} W_2^2(\rho^{\tau,n}, \sigma) - W_2^2(\rho^{\tau,n-1}, \sigma) = W_2^2(S(\tau)_\# \widehat{\rho}^{\tau,n}, \sigma) - W_2^2(\widehat{\rho}^{\tau,n}, \sigma) \\ + W_2^2(\widehat{\rho}^{\tau,n}, \sigma) - W_2^2(\rho^{\tau,n-1}, \sigma); \end{aligned} \tag{5.31}$$

in fact, one can notice

$$
\begin{aligned}
W_2^2(S(\tau)_\# \widehat{\rho}^{\tau,n},\sigma) - W_2^2(\widehat{\rho}^{\tau,n},\sigma) &\le \int_{\mathbb{R}^d} |S(\tau) - \mathbf{t}^{\sigma}_{\widehat{\rho}^{\tau,n}}|^2 \mathrm{d}\widehat{\rho}^{\tau,n} - \int_{\mathbb{R}^d} |\mathbf{i} - \mathbf{t}^{\sigma}_{\widehat{\rho}^{\tau,n}}|^2 \mathrm{d}\widehat{\rho}^{\tau,n} \\
&= \int_{\mathbb{R}^d} \langle S(\tau) - \mathbf{i}, S(\tau) + \mathbf{i} - 2\mathbf{t}^{\sigma}_{\widehat{\rho}^{\tau,n}} \rangle \mathrm{d}\widehat{\rho}^{\tau,n} \\
&= 2\tau \int_{\mathbb{R}^d} \langle \tfrac{S(\tau)-\mathbf{i}}{\tau}, \tfrac{S(\tau)+\mathbf{i}}{2} - \mathbf{t}^{\sigma}_{\widehat{\rho}^{\tau,n}} \rangle \mathrm{d}\widehat{\rho}^{\tau,n};
\end{aligned}
$$

while, from (5.19),

$$
W_2^2(\widehat{\rho}^{\tau,n},\sigma) - W_2^2(\rho^{\tau,n-1},\sigma) \le 2\tau\Big(\mathscr{H}[\sigma] - \mathscr{H}[\widehat{\rho}^{\tau,n}]\Big).
$$

□

Remark 5.9. Inequality (5.28) could be interpreted as

$$
\begin{aligned}
&\frac{1}{2}\Big[W_2^2(\rho^\tau_{n\tau},\sigma) - W_2^2(\rho^\tau_{(n-1)\tau},\sigma)\Big] \le \\
&\qquad \le - \int_{(n-1)\tau}^{n\tau} \mathrm{d}s \int_{\mathbb{R}^d} \langle \mathcal{B}_\tau, \tfrac{S(\tau)+\mathbf{i}}{2} - \mathbf{t}^{\sigma}_{\widehat{\rho}^\tau_s} \rangle \mathrm{d}\widehat{\rho}^\tau_s + \int_{(n-1)\tau}^{n\tau} \big(\mathscr{H}[\sigma] - \mathscr{H}[\widehat{\rho}^\tau_s]\big)\mathrm{d}s.
\end{aligned}
$$

This means, for $t_2 \ge t_1 \ge 0$, $t_2 = m\tau$ and $t_1 = n\tau$,

$$
\begin{aligned}
&\frac{1}{2}\Big[W_2^2(\rho^\tau_{t_2},\sigma) - W_2^2(\rho^\tau_{t_1},\sigma)\Big] \le \\
&\qquad \le - \int_{t_1}^{t_2} \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle \mathcal{B}_\tau, \tfrac{S(\tau)+\mathbf{i}}{2} - \mathbf{t}^{\sigma}_{\widehat{\rho}^\tau_s} \Big\rangle \mathrm{d}\widehat{\rho}^\tau_s + \int_{t_1}^{t_2} \big(\mathscr{H}[\sigma] - \mathscr{H}[\widehat{\rho}^\tau_s]\big)\mathrm{d}s.
\end{aligned}
\tag{5.32}
$$

The convergence of the scheme

Theorem 5.10. *For ρ_0 satisfying (5.5), consider the curves $\rho^\tau, \widehat{\rho}^\tau : [0,+\infty[\to \mathbb{R}^d$ described in (5.14),(5.15),(5.25). Choosing $\tau_m = \tau/2^m$, we have*

$\rho^{\tau_m}, \widehat{\rho}^{\tau_m}$ (locally) uniformly converge (to the same limit) in W_2.

Denoting with $\bar{\rho}_t$ this limit, it satisfies for $t, s \in [0,+\infty[$,

$$
W_2(\bar{\rho}_t,\bar{\rho}_s) \le |t-s|\Big(\|B\|_{L^2(\mathbb{R}^d,\rho_0)} + \Big\|\tfrac{\nabla u}{u}\Big\|_{L^2(\mathbb{R}^d,\rho_0)}\Big), \tag{5.33}
$$

$$
\mathscr{H}[\bar{\rho}_t] \le \mathscr{H}[\rho_0] + d\ \mathrm{Lip}(B)t. \tag{5.34}
$$

Proof. For $\mu_0, \nu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ satisfying (5.5), we construct (as in (5.14),(5.15)) two curves $\mu^\tau, \nu^{\frac{\tau}{2}} : [0,+\infty[\to \mathcal{P}_2(\mathbb{R}^d)$ respectively with initial data μ_0, ν_0 and time steps τ, $\tau/2$. From (2.17a) and (2.17b),

$$
W_2^2(\mu^{\tau,n},\nu^{\frac{\tau}{2},m}) = W_2^2(S(\tau)_\#\widehat{\mu}^{\tau,n}, S(\tfrac{\tau}{2})_\#\widehat{\nu}^{\frac{\tau}{2},m}) \le W_2^2(S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n}, \widehat{\nu}^{\frac{\tau}{2},m}) \tag{5.35}
$$

On the other hand, from (5.19) we obtain for μ^τ and any $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$

$$
W_2^2(\sigma,\widehat{\mu}^{\tau,n}) = W_2^2(\sigma,\widehat{\mathcal{J}}_\tau\mu^{\tau,n-1}) \le 2\tau\Big(\mathscr{H}[\sigma] - \mathscr{H}[\widehat{\mu}^{\tau,n}]\Big) + W_2^2(\sigma,\mu^{\tau,n-1}); \tag{5.36}
$$

respectively, for $\nu^{\frac{\tau}{2}}$

$$W_2^2(\sigma,\widehat{\nu}^{\frac{\tau}{2},m}) = W_2^2(\sigma,\widehat{\mathcal{J}_{\frac{\tau}{2}}}\nu^{\frac{\tau}{2},m-1}) \leq \tau\Big(\mathscr{H}[\sigma]-\mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},m}]\Big) + W_2^2(\sigma,\nu^{\frac{\tau}{2},m-1}). \tag{5.37}$$

Then

$$\begin{aligned} W_2^2(\mu^{\tau,n},\nu^{\frac{\tau}{2},2n}) &\overset{(5.35)}{\leq} W_2^2(S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n},\widehat{\nu}^{\frac{\tau}{2},2n}) \leq \\ &\overset{(5.37)}{\leq} \tau\Big(\mathscr{H}[S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n}]-\mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n}]\Big) + W_2^2(S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n},\nu^{\frac{\tau}{2},2n-1}). \end{aligned}$$

Recalling that

$$\begin{aligned} W_2^2(S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n},\nu^{\frac{\tau}{2},2n-1}) &= W_2^2(S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n},S(\tfrac{\tau}{2})_\#\widehat{\nu}^{\frac{\tau}{2},2n-1}) \\ &\overset{(2.17a)}{\leq} W_2^2(\widehat{\mu}^{\tau,n},\widehat{\nu}^{\frac{\tau}{2},2n-1}), \end{aligned}$$

we have

$$\begin{aligned} W_2^2(\mu^{\tau,n},\nu^{\frac{\tau}{2},2n}) &\leq \tau\Big(\mathscr{H}[S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n}]-\mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n}]\Big) + W_2^2(\widehat{\mu}^{\tau,n},\widehat{\nu}^{\frac{\tau}{2},2n-1}) \\ &\overset{(5.37)}{\leq} \tau\Big(\mathscr{H}[S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n}]-\mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n}]+\mathscr{H}[\widehat{\mu}^{\tau,n}]-\mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n-1}]\Big) \\ &\quad + W_2^2(\widehat{\mu}^{\tau,n},\nu^{\frac{\tau}{2},2n-2}), \end{aligned}$$

applying (5.37) for $m = 2n-1$ and $\sigma = \widehat{\mu}^{\tau,n}$. As observed in (5.22) and (5.20),

$$\begin{aligned} &\mathscr{H}[S(\tfrac{\tau}{2})_\#\widehat{\mu}^{\tau,n}] \overset{(5.22)}{\leq} \mathscr{H}[\widehat{\mu}^{\tau,n}] + d\ \mathrm{Lip}(B)\tfrac{\tau}{2}, \\ &\mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n}] \overset{(5.20)}{\leq} \mathscr{H}[\nu^{\frac{\tau}{2},2n-1}] \overset{(5.22)}{\leq} \mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n-1}] + d\ \mathrm{Lip}(B)\tfrac{\tau}{2}; \end{aligned}$$

hence

$$W_2^2(\mu^{\tau,n},\nu^{\frac{\tau}{2},2n}) \leq 2\tau\Big(\mathscr{H}[\widehat{\mu}^{\tau,n}]-\mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n}]\Big)+W_2^2(\widehat{\mu}^{\tau,n},\nu^{\frac{\tau}{2},2n-2})+d\ \mathrm{Lip}(B)\tau^2.$$

Applying (5.36) for $\sigma = \nu^{\frac{\tau}{2},2n}$,

$$\begin{aligned} &W_2^2(\mu^{\tau,n},\nu^{\frac{\tau}{2},2n}) \leq \\ &\quad\leq 2\tau\Big(\mathscr{H}[\nu^{\frac{\tau}{2},2n-2}]-\mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n}]\Big) + W_2^2(\mu^{\tau,n-1},\nu^{\frac{\tau}{2},2n-2}) + d\ \mathrm{Lip}(B)\tau^2. \end{aligned}$$

Since $\mathscr{H}[\nu^{\frac{\tau}{2},2n}] \leq \mathscr{H}[\widehat{\nu}^{\frac{\tau}{2},2n}] + d\ \mathrm{Lip}(B)\frac{\tau}{2}$,

$$\begin{aligned} &W_2^2(\mu^{\tau,n},\nu^{\frac{\tau}{2},2n}) \leq \\ &\quad\leq 2\tau\Big(\mathscr{H}[\nu^{\frac{\tau}{2},2n-2}]-\mathscr{H}[\nu^{\frac{\tau}{2},2n}]\Big) + W_2^2(\mu^{\tau,n-1},\nu^{\frac{\tau}{2},2n-2}) + 2d\ \mathrm{Lip}(B)\tau^2. \end{aligned}$$

We have finally showed that, for the time instant $t = n\tau$,

$$\begin{aligned} W_2^2(\mu^\tau_{n\tau},\nu^{\frac{\tau}{2}}_{n\tau}) \leq 2\tau\Big(\mathscr{H}[\nu^{\frac{\tau}{2}}_{(n-1)\tau}]-\mathscr{H}[\nu^{\frac{\tau}{2}}_{n\tau}]\Big)+ \\ + W_2^2(\mu^\tau_{(n-1)\tau},\nu^{\frac{\tau}{2}}_{(n-1)\tau}) + 2d\ \mathrm{Lip}(B)\tau^2. \end{aligned} \tag{5.38}$$

Recursively applying (5.38), we get

$$W_2^2(\mu_{n\tau}^{\tau}, \nu_{n\tau}^{\frac{\tau}{2}}) \;\le\; 2\tau\Big(\mathscr{H}[\nu_0^{\frac{\tau}{2}}] - \mathscr{H}[\nu_{n\tau}^{\frac{\tau}{2}}]\Big) + W_2^2(\mu_0^{\tau}, \nu_0^{\frac{\tau}{2}}) + 2\tau d\ \mathrm{Lip}(B)(n\tau).$$

Setting $\mu_0 = \nu_0 = \rho_0$,

$$W_2^2(\rho_{n\tau}^{\tau}, \rho_{n\tau}^{\frac{\tau}{2}}) \le 2\tau\Big(\mathscr{H}[\rho_0] - \mathscr{H}[\rho_{n\tau}^{\frac{\tau}{2}}]\Big) + 2\tau d\ \mathrm{Lip}(B)(n\tau). \tag{5.39}$$

As a consequence, for all $t \in [0, +\infty[$ that are integer multiples of $\tau/2^m$,

$$W_2^2(\rho_t^{\frac{\tau}{2^m}}, \rho_t^{\frac{\tau}{2^{m+1}}}) \le 2\tau\Big(\mathscr{H}[\rho_0] + d\ \mathrm{Lip}(B)T\Big) \tag{5.40}$$

with $T > t$ sufficiently large; so that

$$W_2(\rho_t^{\frac{\tau}{2^m}}, \rho_t^{\frac{\tau}{2^k}}) \le \sum_{i=m}^{k-1} 2^{-i/2}\sqrt{2\tau}\sqrt{\mathscr{H}[\rho_0] + d\ \mathrm{Lip}(B)T} \tag{5.41}$$

for all $k > m \ge j$ for all $t \in [0, +\infty[$ that are integer multiples of $\tau/2^j$. For any such t (and therefore on a dense set of times) the sequence $\{\rho^{\frac{\tau}{2^m}}(t)\}$ has the Cauchy property and converges in $\mathcal{P}_2(\mathbb{R}^d)$ to some limit, that we shall denote by $\bar\rho_t$.

Using the discrete estimate (5.26) we obtain convergence for all times, as well as the uniform continuity (5.33) of $t \mapsto \bar\rho_t$. For (5.34), it is enough to pass to the limit in (5.27).

It remains to control that $\widehat{\rho}^{\frac{\tau}{2^m}}$ converges to $\bar\rho$. But for T sufficiently large

$$W_2^2(\rho^{\frac{\tau}{2^m},n-1}, \widehat{\rho}^{\frac{\tau}{2^m},n}) \le 2\frac{\tau}{2^m}\Big(\mathscr{H}[\rho^{\frac{\tau}{2^m},n-1}]\Big) \le 2\frac{\tau}{2^m}\Big(\mathscr{H}[\rho_0] + d\ \mathrm{Lip}(B)T\Big), \tag{5.42}$$

that yields

$$\lim_{m\to\infty} W_2^2(\rho_t^{\frac{\tau}{2^m}}, \widehat{\rho}_t^{\frac{\tau}{2^m}}) = 0, \quad t \in [0, +\infty[.$$

□

The proof of Theorem 5.6

The limit of the splitting method scheme found in Theorem 5.10 is the candidate for the solution claimed in Theorem 5.6 .

Proof of Theorem 5.6. The limit $\bar\rho_t$ of the sequence $\rho_t^{\tau_m}$ already satisfies (5.6) and (5.18). We now focus our attention on (5.8) and show how (5.6) and (5.8) imply that $\bar\rho_t$ solves (5.1) in the sense of Definition 5.1. We know (see Prop.5.8 and the following Remark) that, for $t_2 \ge t_1 \ge 0$, $t_1 = n\tau_k$, $t_1 = m\tau_k$,

$$\frac{1}{2}\Big[W_2^2(\rho_{t_2}^{\tau_m}, \sigma) - W_2^2(\rho_{t_1}^{\tau_m}, \sigma)\Big] \le$$
$$\le -\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\Big\langle \mathcal{B}_{\tau_m}, \tfrac{S(\tau_m)+\mathbf{i}}{2} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma}\Big\rangle \mathrm{d}\widehat{\rho}_s^{\tau_m} + \int_{t_1}^{t_2}\Big(\mathscr{H}[\sigma] - \mathscr{H}[\widehat{\rho}_s^{\tau_m}]\Big)\mathrm{d}s.$$

We want to pass to the limit in the previous inequality for $m \to \infty$ to obtain (5.8). In particular, we must be careful to the two terms of the right hand side of the inequality. About the first one, we note

$$\int_{\mathbb{R}^d} \Big\langle \mathcal{B}_{\tau_m}, \tfrac{S(\tau_m)+\mathbf{i}}{2} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma} \Big\rangle \mathrm{d}\widehat{\rho}_s^{\tau_m} = \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \mathcal{B}_{\tau_m}(x), \tfrac{S(\tau_m)x+x}{2} - y \rangle \mathrm{d}\widehat{\pi}_s^{m,\sigma}(x,y), \tag{5.43}$$

with

$$\widehat{\pi}_s^{m,\sigma} := (\mathbf{i}\times \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma})_{\#}\widehat{\rho}_s^{\tau_m},$$

the unique *optimal transference plan* between $\widehat{\rho}_s^{\tau_m}$ and σ. Recalling that

$$\widehat{\rho}_s^{\tau_m} \to \bar{\rho}_s \text{ in } (\mathcal{P}_2(\mathbb{R}^d), W_2),$$

from [2, Prop.7.1.3],

$$\widehat{\pi}_s^{m,\sigma} \to \pi_s^{\sigma} := (\mathbf{i}\times \mathbf{t}_{\bar{\rho}_s}^{\sigma})_{\#}\rho_s \text{ in } (\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d), W_2).$$

Moreover, the map $\mathbf{r}_\tau : \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}^d\times\mathbb{R}^d$

$$\mathbf{r}_{\tau_m}(x,y) = \Big(\mathcal{B}_{\tau_m}(x), \tfrac{S(\tau_m)x+x}{2} - y\Big)$$

converges (uniformly on the compact sets) for $m\to\infty$ to the continuous map $\mathbf{r} : \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}^d\times\mathbb{R}^d$,

$$\mathbf{r}(x,y) = (B(x), x-y).$$

Then thanks to the Lemma 2.15, we argue that

$$\begin{aligned}
&\int_{\mathbb{R}^d} \Big\langle \mathcal{B}_{\tau_m}, \tfrac{S(\tau_m)+\mathbf{i}}{2} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma} \Big\rangle \mathrm{d}\widehat{\rho}_s^{\tau_m} = \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle x,y\rangle \mathrm{d}\mathbf{r}_{\tau_m\#}\widehat{\pi}_s^{m,\sigma}(x,y)\\
&\quad \xrightarrow[m\to\infty]{} \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle x,y\rangle \mathrm{d}\mathbf{r}_{\#}\pi_s^{\sigma}(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle B(x), x-y\rangle \mathrm{d}\pi_s^{\sigma}(x,y)\\
&\quad = \int_{\mathbb{R}^d} \langle B, \mathbf{i} - \mathbf{t}_{\bar{\rho}_s}^{\sigma}\rangle \mathrm{d}\bar{\rho}_s.
\end{aligned}$$

On the other hand, we can observe

$$\begin{aligned}
&\int_{\mathbb{R}^d} \Big\langle \mathcal{B}_{\tau_m}, \tfrac{S(\tau_m)+\mathbf{i}}{2} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma} \Big\rangle \mathrm{d}\widehat{\rho}_s^{\tau_m} = \int_{\mathbb{R}^d} \Big\langle \mathcal{B}_{\tau_m}, -\tfrac{1}{2}\mathcal{B}_{\tau_m} + \mathbf{i} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma} \Big\rangle \mathrm{d}\widehat{\rho}_s^{\tau_m}\\
&\quad \le \int_{\mathbb{R}^d} |\mathcal{B}_{\tau_m}|^2 + \Big\langle \mathcal{B}_{\tau_m}, \mathbf{i} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma} \Big\rangle \mathrm{d}\widehat{\rho}_s^{\tau_m} \le \|\mathcal{B}_{\tau_m}\|^2 + \|\mathcal{B}_{\tau_m}\|\|\mathbf{i} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma}\|\\
&\quad \le d^2\,\mathrm{Lip}(B)^2\|x\|^2 + d\,\mathrm{Lip}(B)\|x\|\|\mathbf{i} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma}\|\\
&\quad \le d^2\,\mathrm{Lip}(B)^2 W_2^2(\widehat{\rho}_s^{\tau_m}, \delta_0) + d\,\mathrm{Lip}(B) W_2(\widehat{\rho}_s^{\tau_m}, \delta_0) W_2(\widehat{\rho}_s^{\tau_m}, \sigma)
\end{aligned} \tag{5.44}$$

In according to (5.6), the last term is integrable. Thus, we can apply the Dominated Convergence Theorem and argue

$$\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d} \Big\langle \mathcal{B}_{\tau_m}, \tfrac{S(\tau_m)+\mathbf{i}}{2} - \mathbf{t}_{\widehat{\rho}_s^{\tau_m}}^{\sigma} \Big\rangle \mathrm{d}\widehat{\rho}_s^{\tau_m} \xrightarrow[m\to\infty]{} \int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d} \langle B, \mathbf{i} - \mathbf{t}_{\bar{\rho}_s}^{\sigma}\rangle \mathrm{d}\bar{\rho}_s.$$

About the other term, we recall that $\mathscr{H}$ is lower semicontinuous with respect to narrow convergence:

$$\liminf_{m\to+\infty} \mathscr{H}[\widehat{\rho}_s^{\tau_m}] = \mathscr{H}[\bar{\rho}_s];$$

applying Fatou's Lemma,

$$\limsup_{m\to\infty} \int_{t_1}^{t_2} \big(\mathscr{H}[\sigma] - \mathscr{H}[\widehat{\rho}_s^{\tau_m}]\big)\mathrm{d}s \le \int_{t_1}^{t_2} \big(\mathscr{H}[\sigma] - \mathscr{H}[\bar{\rho}_s]\big)\mathrm{d}s.$$

This concludes the proof of (5.8).

We now can verify that $\bar{\rho}_t$ satisfies Definition 5.1. Since $\bar{\rho}_t$ is an absolutely continuous curve (see Section 2.4), there exists a vector field $\mathbf{v}_t$, such that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\bar{\rho}_t, \sigma) = \int_{\mathbb{R}^d} \langle \mathbf{v}_t, \mathbf{i} - \mathbf{t}_{\bar{\rho}_t}^{\sigma} \rangle \mathrm{d}\bar{\rho}_t. \tag{5.45}$$

Moreover, we can derive from (5.8)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\bar{\rho}_t, \sigma) \le -\int_{\mathbb{R}^d} \langle B, \mathbf{i} - \mathbf{t}_{\bar{\rho}_t}^{\sigma} \rangle \mathrm{d}\bar{\rho}_t + \mathscr{H}[\sigma] - \mathscr{H}[\bar{\rho}_t]. \tag{5.46}$$

Comparing (5.46) and (5.45), we obtain

$$\int_{\mathbb{R}^d} \langle \mathbf{v}_t + B, \mathbf{i} - \mathbf{t}_{\bar{\rho}_t}^{\sigma} \rangle \mathrm{d}\bar{\rho}_t \le \mathscr{H}[\sigma] - \mathscr{H}[\bar{\rho}_t];$$

this means $\mathbf{v}_t + B \in -\partial\mathscr{H}[\bar{\rho}_t]$. Since $\bar{\rho}_t = u_t \mathrm{d}\mathscr{L}^d$ and $\partial\mathscr{H}[\bar{\rho}_t]$ is not empty, Theorem 2.33 ensures that

$$u_t \in W^{1,1}_{loc}(\mathbb{R}^d), \quad \tfrac{\nabla u_t}{u_t} \in L^2(\mathbb{R}^d, \bar{\rho}_t) \quad \partial^\circ \mathscr{H}[\bar{\rho}_t] = \Big\{\tfrac{\nabla u_t}{u_t}\Big\}.$$

Hence

$$\int_{\mathbb{R}^d} \langle \mathbf{v}_t + B, \mathbf{i} - \mathbf{t}_{\bar{\rho}_t}^{\sigma} \rangle \mathrm{d}\bar{\rho}_t = -\int_{\mathbb{R}^d} \langle \tfrac{\nabla u_t}{u_t}, \mathbf{i} - \mathbf{t}_{\bar{\rho}_t}^{\sigma} \rangle u_t \mathrm{d}\mathscr{L}^d. \tag{5.47}$$

From Remark 2.23

$$\overline{\{\lambda(\mathbf{i} - \mathbf{t}_{\bar{\rho}_t}^{\sigma}) : \sigma \in \mathcal{P}(\mathbb{R}^d), \lambda > 0\}}^{L^2(\mathbb{R}^d, \bar{\rho}_t)} = \overline{\{\nabla\varphi : \varphi \in C_{\mathrm{c}}^\infty(\mathbb{R}^d)\}}^{L^2(\mathbb{R}^d, \bar{\rho}_t)};$$

then

$$\int_{\mathbb{R}^d} \langle \mathbf{v}_t + B, \nabla\varphi \rangle \mathrm{d}\bar{\rho}_t = -\int_{\mathbb{R}^d} \langle \nabla u_t, \nabla\varphi \rangle \mathrm{d}\mathscr{L}^d \quad \forall \varphi \in C_{\mathrm{c}}^\infty(\mathbb{R}^d);$$

integrating by parts the right-hand side,

$$\int_{\mathbb{R}^d} \langle \mathbf{v}_t + B, \nabla\varphi \rangle \mathrm{d}\bar{\rho}_t = \int_{\mathbb{R}^d} \Delta\varphi \; u_t \mathrm{d}\mathscr{L}^d.$$

$\bar{\rho}_t$ is a solution in the sense of Definition 5.1. □

5.3 Convergence of the regularized problem

B could be extended to a maximal monotone operator B. From [1, Corollary 1.4],

$$B = \mathsf{B} = \mathsf{B}^\circ \quad \mathscr{L}^d\text{-a.e.}$$

Denoting with B^ε the Moreau-Yosida approximation of B:

$$\lim_{\varepsilon\to 0} B^\varepsilon(x_0) = B(x_0), \quad |B^\varepsilon(x_0)| \uparrow |B(x_0)| \text{ for } \varepsilon \downarrow 0, \text{ for } \mathscr{L}^d\text{-a.e. } x_0; \tag{5.48}$$

B^ε is a Lipschitz continuous maximal monotone operator and verifies $B^\varepsilon(0) = 0$. When condition (5.5) is satisfied, Theorem 5.6 applies to the problem

$$\partial_t\rho_t - \Delta\rho_t - \nabla\cdot(\rho_t B^\varepsilon) = 0 \text{ in } \mathbb{R}^d \times (0, +\infty), \quad \lim_{t\to 0}\rho_t = \rho_0 : \tag{5.49}$$

there exists a (unique) solution ρ^ε to (5.49) in the sense of Definition 5.1. We prove

Theorem 5.11. *Let us consider a family of Moreau-Yosida approximations B^ε satisfying* (5.48) *and a probability measure $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ satisfying condition* (5.5)*. The solution ρ^ε of equation* (5.49) *(locally) uniformly converges w.r.t. W_2 as $\varepsilon \to 0$ to the unique solution $\rho_t \in \mathcal{P}_2^r(\mathbb{R}^d)$ of* (5.1) *in the sense of Definition 5.1 with initial datum ρ_0. Moreover, ρ_t satisfies* (5.6)*.*

Convergence of ρ_t^ε

By Theorem 5.6, ρ^ε has the following properties

$$W_2(\rho_t^\varepsilon, \rho_s^\varepsilon) \le |t-s|\Big(\|B^\varepsilon\|_{L^2(\mathbb{R}^d,\rho_0)} + \Big\|\tfrac{\nabla u}{u}\Big\|_{L^2(\mathbb{R}^d,\rho_0)}\Big); \tag{5.50}$$

$$\int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\rho_{t_2}^\varepsilon + \int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\Big(B(x)\cdot x\Big)\,\mathrm{d}\rho_s^\varepsilon = \mathrm{d}(t_2 - t_1) + \int_{\mathbb{R}^d}|x|^2\,\mathrm{d}\rho_{t_1}^\varepsilon; \tag{5.51}$$

$$\begin{aligned}\frac{1}{2}W_2^2(\rho_{t_2}^\varepsilon, \sigma) + \int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B^\varepsilon, \mathbf{i} - \mathbf{t}_{\rho_s^\varepsilon}^\sigma\rangle\mathrm{d}\rho_s^\varepsilon \le \\ \le \frac{1}{2}W_2^2(\rho_{t_1}^\varepsilon, \sigma) + \int_{t_1}^{t_2}\big(\mathscr{H}[\sigma] - \mathscr{H}[\rho_s^\varepsilon]\big)\mathrm{d}s.\end{aligned} \tag{5.52}$$

Thus we can show

Proposition 5.12. *Under the assumption* (5.5)*, the set*

$$\{\rho^\varepsilon : [0, +\infty[\to \mathcal{P}_2(\mathbb{R}^d) :\ \varepsilon > 0, \rho_t^\varepsilon \textit{solution of } (5.49)\} \tag{5.53}$$

is relatively compact w.r.t. (local) uniform convergence in W_2. Any limit curve $\bar\rho_t$ belongs to $\mathcal{P}_2^r(\mathbb{R}^d)$ for $\mathscr{L}^1$-a.e. t, the curve is (narrowly) continuous, and satisfies

$$m_2^2(\bar\rho_t) \le \mathrm{d}t + m_2^2(\rho_0).$$

Proof. First of all, we observe that for any t, $\{\rho_t^\varepsilon : \varepsilon > 0\}$ is relatively compact in the narrow topology, since this set is bounded in $\mathcal{P}_2(\mathbb{R}^d)$ (and therefore it is tight), as follows from

$$m_2^2(\rho_t^\varepsilon) \overset{(5.51)}{\leq} \mathrm{d}t + m_2^2(\rho_0^\varepsilon) = \mathrm{d}t + m_2^2(\rho_0).$$

On the other hand, from (5.48) we have $\|B^\varepsilon\|_{L^2(\mathbb{R}^d,\rho_0)} \leq \|B\|_{L^2(\mathbb{R}^d,\rho_0)}$. Then

$$W_2(\rho_t^\varepsilon, \rho_s^\varepsilon) \leq |t-s| \Big(\|B\|_{L^2(\mathbb{R}^d,\rho_0)} + \Big\| \tfrac{\nabla u}{u} \Big\|_{L^2(\mathbb{R}^d,\rho_0)} \Big), \tag{5.54}$$

and the set is equicontinuous in the Wasserstein metric. Since Wasserstein metric topology is stronger than the narrow topology, the set is equicontinuous in the narrow topology. Applying the Ascoli-Arzelá theorem, we conclude the relative compactness of the set in $C^0(0,T;\mathcal{P}_2(\mathbb{R}^d))$ for every $T > 0$.

Let $\bar{\rho}_t$ be a limit curve for the set: there exists a vanishing sequence ε_n s.t. $\rho_t^{\varepsilon_n} \to \bar{\rho}_t$ narrowly. We have

$$\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\bar{\rho}_t \leq \liminf_{\varepsilon_n \to 0} \int_{\mathbb{R}^d} |x|^2 \mathrm{d}\rho^\varepsilon(t) \leq \mathrm{d}t + m_2^2(\rho_0).$$

From (5.52), we have

$$\int_{t_1}^{t_2} \mathscr{H}[\rho_s^\varepsilon]\mathrm{d}s \leq \int_{t_1}^{t_2} \mathrm{d}s \int_{\mathbb{R}^d} \langle B^\varepsilon, \mathbf{t}_{\rho_s^\varepsilon}^\sigma - \mathbf{i} \rangle \mathrm{d}\rho_s^\varepsilon + \int_{t_1}^{t_2} \mathscr{H}[\sigma]\mathrm{d}s + \frac{1}{2} W_2^2(\rho_{t_1}^\varepsilon, \sigma).$$

We can choose σ s.t. its support is contained in $\mathrm{B}_1(0)$: so $|\mathbf{t}_{\rho_s^\varepsilon}^\sigma(x)| \leq 1$, ρ_s^ε-a.e.; by the monotonicity of B^ε,

$$\begin{aligned} \langle B^\varepsilon(x), \mathbf{t}_{\rho_s^\varepsilon}^\sigma(x) - x \rangle &\leq \langle B^\varepsilon \circ \mathbf{t}_{\rho_s^\varepsilon}^\sigma, \mathbf{t}_{\rho_s^\varepsilon}^\sigma(x) - x \rangle \\ &\leq \sup_{y \in \mathrm{B}_1(0)} |B^\varepsilon(y)| |\mathbf{t}_{\rho_s^\varepsilon}^\sigma(x) - x| \leq \sup_{y \in \mathrm{B}_1(0)} |B(y)| |\mathbf{t}_{\rho_s^\varepsilon}^\sigma(x) - x|. \end{aligned}$$

We obtain

$$\int_{t_1}^{t_2} \mathscr{H}[\rho_s^\varepsilon]\mathrm{d}s \leq \sup_{\mathrm{B}_1(0)} |B| \int_{t_1}^{t_2} \mathrm{d}s \int_{\mathbb{R}^d} |\mathbf{t}_{\rho_s^\varepsilon}^\sigma(x) - x| \mathrm{d}\rho_s^\varepsilon + \int_{t_1}^{t_2} \mathscr{H}[\sigma]\mathrm{d}s + \frac{1}{2} W_2^2(\rho_{t_1}^\varepsilon, \sigma). \tag{5.55}$$

Observing that

$$\int_{\mathbb{R}^d} |\mathbf{t}_{\rho_s^\varepsilon}^\sigma(x) - x| \mathrm{d}\rho_s^\varepsilon \leq \|\mathbf{t}_{\rho_s^\varepsilon}^\sigma(x) - x\|_{L^2(\rho_s^\varepsilon, \mathbb{R}^d)} = W_2(\rho_s^\varepsilon, \sigma),$$

$$W_2^2(\rho_s^\varepsilon, \sigma) \leq m_2^2(\rho_s^\varepsilon) + m_2^2(\sigma) \leq \mathrm{d}s + m_2^2(\rho_0) + m_2^2(\sigma),$$

the right hand side of (5.55) could be estimated independently from ε. By Fatou's Lemma

$$\int_{t_1}^{t_2} \mathscr{H}[\bar{\rho}_s]\mathrm{d}s \leq \liminf_{\varepsilon_n \to 0} \int_{t_1}^{t_2} \mathscr{H}[\rho_s^{\varepsilon_n}]\mathrm{d}s < \infty :$$

for $\mathscr{L}^1$-almost every t, $\mathscr{H}[\bar{\rho}_t]$ is finite and $\bar{\rho}_t$ is absolutely continuous. □

The proof of Theorem 5.11

The idea is to show the limit curve found in the previous Proposition solves (5.1). We need a preliminary result

Lemma 5.13. *Let $B : \mathbb{R}^d \to \mathbb{R}^d$ be a monotone operator and consider a sequence $\{\nu_k\} \subset \mathcal{P}(\mathbb{R}^d)$ narrowly converging to an absolutely continuous measure ν. Then, for any vanishing sequence ε_k,*

$$(B^{\varepsilon_k} \times \mathbf{i})_{\#}\nu_k \to (B \times \mathbf{i})_{\#}\nu \textit{ narrowly in } \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d). \tag{5.56}$$

Proof. We firstly show the tightness of sequence $\boldsymbol{\nu}_k = (B^{\varepsilon_k} \times \mathbf{i})_{\#}\nu_k$, i.e.

$$\forall \delta > 0, \quad \exists\, \boldsymbol{K}_\delta \Subset \mathbb{R}^d \times \mathbb{R}^d \text{ t.c. } \boldsymbol{\nu}_k(\boldsymbol{K}_\delta) > 1 - \delta \text{ for all } k.$$

Fix δ: Since ν_k is tight,

$$\exists\, K_\delta \Subset \mathbb{R}^d \text{ t.c. } \nu_k(K_\delta) > 1 - \delta \text{ for all } k.$$

We can assume $K_\delta = \overline{\mathrm{B}_\eta(0)}$, for η sufficiently large. Being B the maximal monotone extension of B, since $B(0) \ni 0$ $|J^{\mathsf{B}}_{\varepsilon_k}(x)| \leq |x|$:

$$x \in K_\delta \Longrightarrow J^{\mathsf{B}}_{\varepsilon_k}(x) \in K_\delta.$$

Let us define $\boldsymbol{K}_\delta = \mathsf{B}(K_\delta) \times K_\delta$: it is closed and bounded set and for any k,

$$\begin{aligned}(B^{\varepsilon_k} \times \mathbf{i})(K_\delta) = \{(B^{\varepsilon_k}(x), x) : x \in K_\delta\} \subseteq \{(\mathsf{B}(J^{\mathsf{B}}_{\varepsilon_k}(x)), x) : x \in K_\delta\} \subseteq \\ \subseteq \{(\mathsf{B}(y), x) : x, y \in K_\delta\} = \boldsymbol{K}_\delta;\end{aligned}$$

it follows that

$$K_\delta \subseteq (B^{\varepsilon_k} \times \mathbf{i})^{-1}\boldsymbol{K}_\delta.$$

As immediate consequence,

$$\boldsymbol{\nu}_k(\boldsymbol{K}_\delta) = \nu_k((B^{\varepsilon_k} \times \mathbf{i})^{-1}\boldsymbol{K}_\delta) \geq \nu_k(K_\delta) > 1 - \delta.$$

The arbitrariness of δ yields to the tightness of sequence $\boldsymbol{\nu}_k$. We can extract a converging subsequence $\boldsymbol{\nu}_j$: let $\boldsymbol{\nu}$ its limit. Denoting with $\pi_2 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ the projection on the second component, let us observe that

$$\pi_{2\#}\boldsymbol{\nu}_k = \nu_k,\ \forall \varepsilon > 0 \Longrightarrow \pi_{2\#}\boldsymbol{\nu} = \nu. \tag{5.57}$$

From Proposition 2.14,

$$\forall \mathbf{x} \in \operatorname{supp} \boldsymbol{\nu}, \quad \exists\, \mathbf{x}_j \in \operatorname{spt} \boldsymbol{\nu}_j : \lim_{j \to \infty} \mathbf{x}_j = \mathbf{x};$$

since $\operatorname{spt} \boldsymbol{\nu}_j \subseteq (B^{\varepsilon_j} \times \mathbf{i})(\mathbb{R}^d)$, $\mathbf{x}_j = (B^{\varepsilon_j}(x_j), x_j)$ for some $x_j \in \mathbb{R}^d$. We denote $x = \lim_{j\to\infty} x_j$. Since $B^{\varepsilon_j}(x_j) \in \mathsf{B}(J^{\mathsf{B}}_{\varepsilon_j}(x_j))$ and

$$|x_j - J^{\mathsf{B}}_{\varepsilon_j}(x_j)| \leq \varepsilon_j |B^{\varepsilon_j}(x_j)| \leq \varepsilon_j \sup_{|y| \leq \sup_j |x_j|} |B(y)|, \tag{5.58}$$

we get $\mathbf{x} = \lim_{j\to\infty}(B^{\varepsilon_j}(x_j), x_j) \in \{(y, x) : y \in \mathsf{B}(x)\}$. This yields

$$\operatorname{spt} \boldsymbol{\nu} \subseteq \{(y, x) : x \in \mathbb{R}^d,\ \ y \in \mathsf{B}(x)\}. \tag{5.59}$$

From (5.57), (5.59) and the absolute continuity of ν, we obtain

$$\boldsymbol{\nu} = (\mathsf{B} \times \mathbf{i})_{\#}\nu = (B \times \mathbf{i})_{\#}\nu;$$

$\boldsymbol{\nu}$ does not depend from the subsequence $\boldsymbol{\nu}_j$ and any other converging subsequence of $\boldsymbol{\nu}_k$ tends to the same limit $(B \times \mathbf{i})_{\#}\nu$: this means $\boldsymbol{\nu}_k$ is converging and (5.56) follows. □

Proof of Theorem 5.11. 1. We show that any limit curve $\bar{\rho}_t$ of the set (5.53) is a (distributional) solution of (5.1). By definition of limit curve, there exists a vanishing sequence ε_k such that $\rho_t^{\varepsilon_k} \to \bar{\rho}_t$ narrowly: for any k, ρ^{ε_k} solves

$$\int_0^{+\infty} \int_{\mathbb{R}^d} \Big(\partial_t \zeta + \Delta\zeta - B^{\varepsilon_n} \cdot \nabla\zeta\Big)\,\mathrm{d}\rho_t^{\varepsilon_n}\,\mathrm{d}t = 0 \quad \forall\, \zeta \in C_{\mathrm{c}}^{\infty}(\mathbb{R}^d \times (0,+\infty)); \tag{5.60}$$

we want to obtain (5.3) passing to the limit in (5.60). This passage to the limit is immediate if we verify

$$\int_0^{+\infty} \mathrm{d}t \int_{\mathbb{R}^d} \Big(B^{\varepsilon_n} \cdot \nabla\zeta\Big)\,\mathrm{d}\rho^{\varepsilon_n} \longrightarrow \int_0^{+\infty} \mathrm{d}t \int_{\mathbb{R}^d} \Big(B \cdot \nabla\zeta\Big)\,\mathrm{d}\bar{\rho}_t. \tag{5.61}$$

Assuming that $\operatorname{spt}\zeta \subseteq \mathrm{B}_\eta(0) \times [t_1, t_2]$, let us consider $X \in C_{\mathrm{c}}^{\infty}(\mathbb{R}^d)$ such that

$$X(y) = y \text{ in } \mathrm{B}_M(0), \text{ where } M = \sup_{\mathrm{B}_\eta(0)} |B(x)|.$$

Denoting $Z(y,x;t) = X(y) \cdot \nabla\zeta(x,t) \in C_{\mathrm{c}}^{\infty}(\mathbb{R}^d \times \mathbb{R}^d)$, we know from Lemma 5.13

$$\begin{aligned}\int_{\mathbb{R}^d} \Big(B^{\varepsilon_k} \cdot \nabla\zeta\Big)\,\mathrm{d}\rho_t^{\varepsilon_k} &= \int_{\mathbb{R}^d \times \mathbb{R}^d} Z(y,x;t)\mathrm{d}(B^{\varepsilon_k} \times \mathbf{i})_{\#}\rho_t^{\varepsilon_k} \\ &\xrightarrow{n\to\infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} Z(y,x;t)\mathrm{d}(B \times \mathbf{i})_{\#}\bar{\rho}_t = \int_{\mathbb{R}^d} \Big(B \cdot \nabla\zeta\Big)\,\mathrm{d}\bar{\rho}_t.\end{aligned}$$

In addition,

$$\int_{\mathbb{R}^d} \Big(B^{\varepsilon_k} \cdot \nabla\zeta\Big)\,\mathrm{d}\rho_t^{\varepsilon_k} \le M \|\nabla\zeta\|_{\infty}, \quad t \in [t_1, t_2]:$$

Dominated Convergence Theorem applies and yields to (5.61).

2. Any limit curve $\bar{\rho}_t$ also satisfies the quadratic moment estimate (5.4): we immediately obtain it if we check that

$$\text{the map } t \in [0,+\infty[\mapsto \int_{\mathbb{R}^d} \Big(B(x) \cdot x\Big)\,\mathrm{d}\bar{\rho}_t \text{ belongs to } L^1_{loc}([0,+\infty[). \tag{5.62}$$

In fact, since $\bar{\rho}_t$ is a distributional solution, it verifies (5.12): thanks to (5.62), we can pass to the limit via Dominated Convergence theorem and get the estimate (5.4). To check (5.62), we recall from Lemma 5.13,

$$\begin{aligned}\liminf_{k\to\infty} \int_{\mathbb{R}^d} \Big(B^{\varepsilon_k}(x) \cdot x\Big)\,\mathrm{d}\rho_t^{\varepsilon_k} &\overset{B(0)=0}{=} \liminf_{k\to\infty} \int_{\mathbb{R}^d} \Big(B^{\varepsilon_k}(x) \cdot x\Big)^+ \mathrm{d}\rho_t^{\varepsilon_k} \\ &= \liminf_{k\to\infty} \int_{\mathbb{R}^d} \Big(y \cdot x\Big)^+ \mathrm{d}(B^{\varepsilon_k} \times \mathbf{i})_{\#}\rho_t^{\varepsilon_k} \ge \int_{\mathbb{R}^d} \Big(y \cdot x\Big)^+ \mathrm{d}(B \times \mathbf{i})_{\#}\bar{\rho}_t \\ &= \int_{\mathbb{R}^d} \Big(B(x) \cdot x\Big)^+ \mathrm{d}\bar{\rho}_t \overset{B(0)=0}{=} \int_{\mathbb{R}^d} \Big(B(x) \cdot x\Big)\,\mathrm{d}\bar{\rho}_t.\end{aligned}$$

By Fatou's Lemma

$$\int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \Big(B(x)\cdot x\Big)\,\mathrm{d}\bar\rho_s \le \liminf_{k\to\infty} \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \Big(B^{\varepsilon_k}(x)\cdot x\Big)\,\mathrm{d}\rho_s^{\varepsilon_n} \le \mathrm{d}t + \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\rho_0.$$

3. At this point, we can show that ρ^{ε_k} converges to $\bar\rho_t$ not only narrowly but also in the Wasserstein metric. It is enough to check

$$\limsup_{k\to\infty} \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\rho_t^{\varepsilon_k} \le \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\bar\rho_t. \tag{5.63}$$

Comparing (5.4) and (5.51), we get

$$\begin{aligned}\lim_{k\to\infty} \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\rho_t^{\varepsilon_k} + \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \Big(B^{\varepsilon_k}(x)\cdot x\Big)\,\mathrm{d}\rho_s^{\varepsilon_k} &= \mathrm{d}t + \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\rho_0 \\ &= \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\bar\rho_t + \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \Big(B(x)\cdot x\Big)\,\mathrm{d}\bar\rho_s.\end{aligned} \tag{5.64}$$

We recall that for two real sequences a_n, b_n,

$$\liminf(a_n+b_n) \le \limsup a_n + \liminf b_n \le \limsup(a_n+b_n)$$

and observe from (5.64) that , in this case, $\exists \lim(a_n+b_n)$: then

$$\begin{aligned}\limsup_{n\to\infty} \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\rho_t^{\varepsilon_n} + \liminf_{n\to\infty} \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \Big(B^{\varepsilon_n}(x)\cdot x\Big)\,\mathrm{d}\rho_s^{\varepsilon_n} &= \\ = \int_{\mathbb{R}^d} |x|^2\,\mathrm{d}\bar\rho_t + \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \Big(B(x)\cdot x\Big)\,\mathrm{d}\bar\rho_s.\end{aligned}$$

from (5.3), we argue (5.63).

4. Since $\bar\rho_t$ is also the limit of $\rho_t^{\varepsilon_k}$ in the Wasserstein metric, we can pass to the limit in (5.54) to obtain (5.6); $\bar\rho_t$ is an absolutely continuous curve and a distributional solution of (5.1): as already noticed in the proof this implies $\bar\rho_t$ is a solution of (5.1) in the sense of the Definition 5.1. Thanks to the uniqueness of the solution, $\bar\rho_t$ is the unique limit point: $\rho_t^\varepsilon \to \bar\rho_t$ for $\varepsilon \to 0$. □

Proof of Theorem 5.4

In according to Theorem 5.11, the proof of Theorem 5.4 is complete if we show metric characterizations (5.8) and (5.7).

Proof of Theorem 5.4. 1. We firstly assume that B is sublinear and show (5.8). Let us assume that σ is absolutely continuous (otherwise it is trivial). As shown in 5.11, ρ_t is the limit of ρ_t^ε, solution of the approximation problem. It is enough to pass to the limit in (5.52) to obtain (5.8). We can easily get to the limit for terms concerning Wasserstein distance. About the entropy contribution we already know that

$$\limsup_{k\to\infty} \int_{t_1}^{t_2} \big(\mathscr{H}[\sigma] - \mathscr{H}[\rho_s^{\varepsilon_k}]\big)\mathrm{d}s \le \int_{t_1}^{t_2} \big(\mathscr{H}[\sigma] - \mathscr{H}[\rho_s]\big)\mathrm{d}s.$$

To conclude we need only to compute that

$$\lim_{\varepsilon\to 0}\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B^\varepsilon,\mathbf{i}-\mathbf{t}^{\sigma}_{\rho^\varepsilon_s}\rangle\mathrm{d}\rho^\varepsilon_s=\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B,\mathbf{i}-\mathbf{t}^{\sigma}_{\rho_s}\rangle\mathrm{d}\rho_s;$$

Since σ is absolutely continuous, it is equivalent to show

$$\lim_{\varepsilon\to 0}\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B^\varepsilon\circ\mathbf{t}^{\rho^\varepsilon_s}_{\sigma},\mathbf{t}^{\rho^\varepsilon_s}_{\sigma}-\mathbf{i}\rangle\mathrm{d}\sigma=\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B\circ\mathbf{t}^{\rho_s}_{\sigma},\mathbf{t}^{\rho_s}_{\sigma}-\mathbf{i}\rangle\mathrm{d}\sigma \qquad (5.65)$$

where $\mathbf{t}^{\rho^\varepsilon_s}_{\sigma}$, $\mathbf{t}^{\rho_s}_{\sigma}$ are the optimal transport maps from σ to ρ^ε_s, ρ_s respectively. From the convergence of ρ^ε_s to ρ_s in W_2, we have (see [2, Proposition 7.1.3, Lemma 5.1.7, Lemma 5.4.1])

$$\mathbf{t}^{\rho^\varepsilon_s}_{\sigma}\to\mathbf{t}^{\rho_s}_{\sigma}\ \text{ in } L^2(\mathbb{R}^d,\sigma);$$

up to subsequences, $\mathbf{t}^{\rho^\varepsilon_s}_{\sigma}\to\mathbf{t}^{\rho_s}_{\sigma}$ σ-almost everywhere. Using the argument of (5.58), we can show that

$$B^\varepsilon\circ\mathbf{t}^{\rho^\varepsilon_s}_{\sigma}\to B\circ\mathbf{t}^{\rho_s}_{\sigma}\ \ \sigma\text{-almost everywhere.}$$

Since $|B^\varepsilon(x)|\le|B(x)|\le C_0+C_1|x|$,

$$B^\varepsilon\circ\mathbf{t}^{\rho^\varepsilon_s}_{\sigma},B\circ\mathbf{t}^{\rho_s}_{\sigma}\in L^2(\mathbb{R}^d,\sigma),$$
$$\|B^\varepsilon\circ\mathbf{t}^{\rho^\varepsilon_s}_{\sigma}\|_{L^2(\mathbb{R}^d,\sigma)},\|B\circ\mathbf{t}^{\rho_s}_{\sigma}\|_{L^2(\mathbb{R}^d,\sigma)}\le M\int_{\mathbb{R}^d}|x|^2\mathrm{d}\rho_s,$$

for a sufficient large $M>0$. Thus it is possible to prove that

$$B^\varepsilon\circ\mathbf{t}^{\rho^\varepsilon_s}_{\sigma}\to B\circ\mathbf{t}^{\rho_s}_{\sigma}\ \text{ in } L^2(\mathbb{R}^d,\sigma)$$

by appropriately adapting the argument of the Lebesgue Dominated Convergence Theorem. It is verified that

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d}\langle B^\varepsilon\circ\mathbf{t}^{\rho^\varepsilon_s}_{\sigma},\mathbf{t}^{\rho^\varepsilon_s}_{\sigma}-\mathbf{i}\rangle\mathrm{d}\sigma=\int_{\mathbb{R}^d}\langle B\circ\mathbf{t}^{\rho_s}_{\sigma},\mathbf{t}^{\rho_s}_{\sigma}-\mathbf{i}\rangle\mathrm{d}\sigma.$$

(5.65) immediately follows by the Lebesgue Dominated Convergence Theorem.

2. We end with the general inequality (5.7) for $\sigma\in\mathcal{P}_2(\mathbb{R}^d)$ with a compact support. Let us assume that σ is absolutely continuous and $\operatorname{spt}\sigma\subseteq \mathrm{B}_\eta(0)$. From (5.52), by the monotonicity of B^ε,

$$\frac{1}{2}\Big[W_2^2(\rho^\varepsilon_{t_2},\sigma)-W_2^2(\rho^\varepsilon_{t_1},\sigma)\Big]\le$$
$$\le\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B^\varepsilon\circ\mathbf{t}^{\sigma}_{\rho^\varepsilon_s},\mathbf{t}^{\sigma}_{\rho^\varepsilon_s}-\mathbf{i}\rangle\mathrm{d}\rho^\varepsilon_s+\int_{t_1}^{t_2}\big(\mathscr{H}[\sigma]-\mathscr{H}[\rho^\varepsilon_s]\big)\mathrm{d}s.$$

We need again to pass to the limit. With respect to the previous case the only difference is to compute

$$\lim_{\varepsilon\to 0}\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B^\varepsilon\circ\mathbf{t}^{\sigma}_{\rho^\varepsilon_s},\mathbf{t}^{\sigma}_{\rho^\varepsilon_s}-\mathbf{i}\rangle\mathrm{d}\rho^\varepsilon_s=\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B\circ\mathbf{t}^{\sigma}_{\rho_s},\mathbf{t}^{\sigma}_{\rho_s}-\mathbf{i}\rangle\mathrm{d}\rho_s.$$

Since σ is absolutely continuous, it is equivalent to show

$$\lim_{\varepsilon\to 0}\int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B^\varepsilon, \mathbf{i}-\mathbf{t}_\sigma^{\rho_s^\varepsilon}\rangle \mathrm{d}\sigma = \int_{t_1}^{t_2}\mathrm{d}s\int_{\mathbb{R}^d}\langle B, \mathbf{i}-\mathbf{t}_\sigma^{\rho_s}\rangle \mathrm{d}\sigma. \tag{5.66}$$

Observing that

$$|\langle B^\varepsilon, \mathbf{i}-\mathbf{t}_\sigma^{\rho_s^\varepsilon}\rangle| \leq C \sup_{\mathrm{B}_\eta(0)} |B|\, |\mathbf{i}-\mathbf{t}_\sigma^{\rho_s}| \quad \sigma\text{-a.e. },$$

we can apply the Lebesgue Dominated Convergence Theorem and obtain

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^d}\langle B^\varepsilon, \mathbf{i}-\mathbf{t}_\sigma^{\rho_s^\varepsilon}\rangle \mathrm{d}\sigma = \int_{\mathbb{R}^d}\langle B, \mathbf{i}-\mathbf{t}_\sigma^{\rho_s}\rangle \mathrm{d}\sigma.$$

We get (5.66) by noticing

$$\Big|\int_{\mathbb{R}^d}\langle B^\varepsilon, \mathbf{i}-\mathbf{t}_\sigma^{\rho_s^\varepsilon}\rangle \mathrm{d}\sigma\Big| \leq C \sup_{\mathrm{B}_\eta(0)} |B|\, W_2(\rho_s,\sigma), \qquad \text{for } \mathscr{L}^1\text{-a.e. } t$$

and applying the Lebesgue Dominated Convergence Theorem again. □

Bibliography

[1] G. Alberti and L. Ambrosio, *A geometrical approach to monotone functions in* $\mathbf{R}^n$, Math. Z., 230 (1999), pp. 259–316.

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Lectures Math. ETH Zürich, Birkhäuser-Verlag, Basel, 2008.

[3] L. Ambrosio, G. Savaré, and L. Zambotti, *Existence and stability for Fokker-Planck equations with log-concave reference measure.*, Probab. Theory Relat. Fields, 145 (2009), pp. 517-564.

[4] L. Ambrosio and P. Tilli, *Topics on analysis in metric spaces*, Oxford Lecture Series in Mathematics and its Applications 25. Oxford: Oxford University Press, 2004.

[5] A. A. Andrievsky, S. N. Gurbatov, and A. N. Sobolevsky, *Ballistic aggregation in symmetric and nonsymmetric flows*, J. Exp. Theor. Phys., 104 (2007), pp. 887–896.

[6] M. Bardi and L. C. Evans, *On Hopf's formulas for solutions of Hamilton-Jacobi equations*, Nonlinear Anal., 8 (1984), pp. 1373–1381.

[7] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[8] F. Bolley, Y. Brenier, and G. Loeper, *Contractive metrics for scalar conservation laws*, J. Hyperbolic Differ. Equ., 2 (2005), pp. 91–107.

[9] V.I. Bogachev, G. Da Prato, M. Röckner, and W. Stannat, *Uniqueness of solutions to weak parabolic equations for measures*, Bull. Lond. Math. Soc., 39 (2007), pp. 631–640.

[10] V.I. Bogachev, G. Da Prato, and M. Röckner, *On parabolic equations for measures.*, Commun. Partial Differ. Equations, 33 (2008), pp. 397–418.

[11] V.I. Bogachev, N.V. Krylov, and M. Röckner, *On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions.*, Commun. Partial Differ. Equations, 26 (2001), pp. 2037–2080.

[12] F. Bouchut, *Advances in Kinetic Theory and Computing*, Ser. Adv. Math. Appl. Sci. 22, World Scientific, River Edge, NJ, 1994.

[13] F. Bouchut and F. James, *Équations de transport unidimensionnelles à coefficients discontinus*, C. R. Acad. Sci. Paris Sér. I Math., 320 (1995), pp. 1097–1102.

[14] L. Boudin, *A solution with bounded expansion rate to the model of viscous pressureless gases*, SIAM J. Math. Anal., 32 (2000), pp. 172–193.

[15] Y. Brenier, *Order preserving vibrating strings and applications to electrodynamics and magnetohydrodynamics*, Methods Appl. Anal., 11 (2004), pp. 515–532.

[16] Y. Brenier, *Le système de Born-Infeld élargi: des ondes aux particules et aux cordes*, Sémin. Équ. Dériv. Partielles VII, École Polytech., Palaiseau, France, 2005, 17 pp.

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

[18] Y. Brenier and E. Grenier, *Sticky particles and scalar conservation laws*, SIAM J. Numer. Anal., 35 (1998), pp. 2317–2328.

[19] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Math. Stud. 5, North–Holland, Amsterdam, 1973.

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

[21] J.A. Carrillo, S. Lisini, G. Savaré, and D. Slepcev, *Nonlinear mobility continuity equations and generalized displacement convexity*, J. Funct. Anal., 258 (2010), pp. 1273-1309.

[22] 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 (2006), pp. 217–263.

[23] 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 (2000), pp. 113–142.

[24] G. Dall'Aglio, *Sugli estremi dei momenti delle funzioni di ripartizione doppia*, Ann. Scuola Norm. Sup. Pisa (3), 10 (1956), pp. 35–74.

[25] S. Daneri and G. Savaré, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM J. Math. Anal., 40 (2008), pp. 1104–1122.

[26] W. E, Y. G. Rykov, and Y. G. Sinai, *Generalized variational principles, global weak solutions and behavior with random initial data for systems of conservation laws arising in adhesion particle dynamics*, Comm. Math. Phys., 177 (1996), pp. 349–380.

[27] S. Fitzpatrick and R.R. Phelps, *Bounded approximants to monotone operators on Banach spaces.*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 9 (1992), pp. 573–595.

[28] W. Gangbo, T. Nguyen, and A. Tudorascu, *Euler-Poisson systems as action-minimizing paths in the Wasserstein space*, Arch. Ration. Mech. Anal., 192 (2009), pp. 419–452.

[29] E. Grenier, *Existence globale pour le système des gaz sans pression*, C. R. Acad. Sci. Paris Sér. I Math., 321 (1995), pp. 171–174.

[30] F. Huang and Z. Wang, *Well posedness for pressureless flow*, Comm. Math. Phys., 222 (2001), pp. 117–146.

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

[32] A. Martin and J. Piasecki, *One dimensional ballistic aggregation: Rigorous long-time estimates*, J. Stat. Phys., 76 (1994).

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

[34] O. Moutsinga, *Convex hulls, sticky particle dynamics and pressure-less gas system*, Ann. Math. Blaise Pascal, 15 (2008), pp. 57–80.

[35] L. Natile, M. A. Peletier, and G. Savaré, *Contraction of general transportation costs along solutions to Fokker-Planck equations with monotone drifts.*, J. Math. Pures Appl. (to appear).

[36] L. Natile and G. Savaré, *A Wasserstein approach to the one-dimensional sticky particle system.*, SIAM J. Math. Anal., 41 (2009), pp. 1340–1365.

[37] T. Nguyen and A. Tudorascu, *Pressureless Euler/Euler–Poisson systems via adhesion dynamics and scalar conservation laws*, SIAM J. Math. Anal., 40 (2008), pp. 754–775.

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

[39] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal., 37 (2005), pp. 1227–1255 (electronic).

[40] J. W. Portegies and M. A. Peletier, *Well-posedness of a parabolic moving-boundary problem in the setting of Wasserstein gradient flows*, arXiv:0812.1269, (2008).

[41] F. Poupaud and M. Rascle, *Measure solutions to the linear transport equations with nonsmooth coefficients*, Comm. Partial Differential Equations, 22 (1997), pp. 337–358.

[42] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems. Vol.* I. *Theory*, Probab. Appl. (N. Y.), Springer-Verlag, New York, 1998.

[43] R. Rossi and G. Savaré, *Tightness, integral equicontinuity and compactness for evolution problems in Banach spaces*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5), 2 (2003), pp. 395–431.

[44] G. Savaré, *Approximation and regularity of evolution variational inequalities*, Rend. Accad. Naz. Sci. XL Mem. Mat. (5), 17 (1993), pp. 83–111.

[45] G. Savaré, *Weak solutions and maximal regularity for abstract evolution inequalities*, Adv. Math. Sci. Appl., 6 (1996), pp. 377–418.

[46] W. Schachermayer and J. Teichmann, *Characterization of optimal transport plans for the Monge-Kantorovich problem*, Proc. Amer. Math. Soc., 137 (2009), pp. 519–529.

[47] M. Sever, *An existence theorem in the large for zero-pressure gas dynamics*, Differential Integral Equations, 14 (2001), pp. 1077–1092.

[48] A. I. Shnirel′man, *On the principle of the shortest way in the dynamics of systems with constraints,* in Application of Topology in Modern Analysis, Voronezh. Gos. Univ., Voronezh, 1985, pp. 124–137 (in Russian); in Global Analysis—Studies and Applications, II, Lecture Notes in Math. 1214, Springer-Verlag, Berlin, 1986, pp. 117–130 (in English).

[49] A. N. Sobolevskiĭ, *The small viscosity method for a one-dimensional system of equations of gas dynamic type without pressure*, Dokl. Akad. Nauk, 356 (1997), pp. 310–312.

[50] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes. Reprint of the 2nd correted printing (1997).*, Classics in Mathematics. Berlin: Springer. xii, 338 p., 2006.

[51] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, AMS, Providence, RI, 2003.

[52] C. Villani, *Optimal transport. Old and new*, vol. 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag,Berlin, 2009.

[53] G. Wolansky, *Dynamics of a system of sticking particles of finite size on the line*, Nonlinearity, 20 (2007), pp. 2175–2189.

[54] YA. B. Zel′dovich, *Gravitational instability: An approximate theory for large density perturbations*, Astro. Astrophys., 5 (1970), pp. 84–89.