ESAIM: PROCEEDINGS AND SURVEYS, Vol. ?, 2018, 1-10
Editors: Will be set by the publisher

CROWD MOTION AND POPULATION DYNAMICS UNDER DENSITY CONSTRAINTS

Filippo Santambrogio[1]

Abstract. This is a survey about the theory of density-constrained evolutions in the Wasserstein space developed by B. Maury, the author, and their collaborators as a model for crowd motion. Connections with microscopic models and other PDEs are presented, as well as several time-discretization schemes based on variational techniques, together with the main theorems guaranteeing their convergence as a tool to prove existence results. Then, a section is devoted to the uniqueness question, and a last one to different numerical methods inspired by optimal transport.

Résumé. Ceci est un article de review sur la théorie des équations d'évolution sous contraintes de densité dans l'espace de Wasserstein developpée par B. Maury, l'auteur, et leurs collaborateurs, comme modèle pour le mouvement de foules. Les interactions avec les modèles microscopiques et d'autres EDP sont présentées, tout comme plusieurs schémas discrets en temps basés sur des techniques variationnelles, ainsi que les théorèmes principaux qui en garantissent la convergence, comme outil pour prouver des résultats d'existence. Ensuite, une section est dédiée à la question de l'unicité, et une dernière aux méthodes numériques inspirées par le transport optimal.

Introduction

Modeling the behavior of human crowds when individuals are an obstacle for each other is a natural issue in applied mathematics, connected in a broad sense to many subdisciplines including, for instance, game theory (when individuals are considered to be rational agents) and fluid mechanics (when they are described as the particles of a fluid). Applications of such a modeling challenge can be crucial in particular in emergency scenarios, such as the evacuation of an area in case of danger, or in potentially critical situations, whenever huge crowds gather together (as in big sport, political, or religious meetings).

Many models have been studied, each with the goal of handling a particular sub-brick of the whole modeling construction. We cite for instance the studies by Helbing [26] or Hughes [27] which are now widely cited.

The present paper is concerned with a sort of meta-model introduced by B. Maury and his collaborators approximately ten years ago, first in a microscopic setting and then in a macroscopic framework in collaboration with the author. Calling it meta-model instead of model means that it is not really concerned with the behavioral choices of the individuals but only with a rigorous mathematical handling of the contacts between them: the motion of the crowd is described through a spontaneous velocity field u (which can be determined in exogenous or endogenous ways, and can take into account exterior obstacles, individual preferences, strategical choices...), which stands for the velocity each agent would follow if the others were not present, and is then transformed

[1] Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay cedex, FRANCE; filippo.santambrogio@math.u-psud.fr

into the actual velocity field v which drives the evolution; the difference between v and u is due to the contacts between individuals.

The paper will be essentially devoted to the macroscopic setting, which has been developed as a continuous counterpart of the microscopic one via tools coming from optimal transport. The work leading to the first paper on the subject, [33], started more than ten years ago, and many developments have occurred since then.

After a brief introduction we will see how the problem can be formulated into a class of evolution PDEs under density constraints, and discuss its connections with other evolution equations. The questions to be studied are the classical ones in PDEs: existence, approximation, uniqueness and numerical simulations, and the present paper is a survey on these issues.

The idea to write such a survey comes from the need to clarify and summarize the evolution of the theory since [33]. For instance, we now understand better uniqueness issues, thanks in particular to [22], but we are also able to formulate clear conjectures on this matter. In what concerns existence and approximation, [33] was only concerned with the gradient-flow case, which is an important setting but not the only reasonable one, and we will see how the gradient structure simplifies the setting but the PDE can also be studied without it. Looking back at [33], the reader will find how much effort had been put in dealing with the mass which exits the domain (in the evacuation setting: the mass going out of the door), which did not appear any more in the subsequent literature. This was essentially due to a certain fear of dealing with non-convex domains (a remainder of the geodesic convexity condition usually required for gradient flows in [4]). This suggested that it was better not to extend a domain beyond a door, and use instead subtle mathematical tricks to handle possible concentrations of the mass on part of the boundary. One of the goals of this survey is also to provide a better reference for the readers, so that they do not have to refer to [33] and get confused with the exit problem, and to optimize the assumptions in the main theorems. Finally, numerical methods have evolved since [33] and we will present both the stochastic projection idea used there and the methods which are the state-of-the-art now for Wasserstein gradient flows in the optimal transport community (see also [47]).

1. Micro and macro models for crowd motions with constraints

In this section, starting from the microscopic setting, we will describe the models presented in [33–36] to deal with crowd evolution involving constraints due to contacts between the individuals.

We consider a population of particles in a domain $\Omega \subset \mathbb{R}^d$, and suppose that each particle, representing an individual, has its own *spontaneous velocity* u (which could a priori depend on time, position, on the particle label itself, on the interaction with the others... this depends on the exact model one wants to consider), which represents the velocity it would follow if alone. Yet, particles are described as rigid disks that cannot overlap, and the presence of the others has an influence on the feasible velocities. The actual velocity cannot always be u, in particular when u tends to concentrate (which is the case every time many individuals in a large area try to exit through a small opening, for instance). We will call v the actual velocity of each particle, and the main assumption of the model is the following: v will be the projection of u onto the set of feasible velocities, which depends on the configuration of the particles. More precisely, $v = P_{\mathrm{adm}(q)}[u]$, where q is the particle configuration, $\mathrm{adm}(q)$ is the set of velocities that do not induce (for an infinitesimal time) overlapping starting from the configuration q, and $P_{\mathrm{adm}(q)}$ is the projection (in the high-dimensional Euclidean space $\mathbb{R}^{dN}$, N being the number of particles) onto this set.

For simplicity, we will suppose that every particle is a disk with a same radius R and denote by q_i the position of the center of the i-thparticle. In this case we define the admissible set K of configurations through

$$K := \{q = (q_i)_i \in \Omega^N \ : \ |q_i - q_j| \geq 2R \ \text{ for all } i \neq j\}.$$

The set of admissible velocities is easily seen to be the following convex cone:

$$\mathrm{adm}(q) = \{v = (v_i)_i \in \mathbb{R}^{dN} \ : \ (v_i - v_j) \cdot (q_i - q_j) \geq 0 \ \text{ for all } (i,j) \text{ with } |q_i - q_j| = 2R\}.$$

The evolution equation which models the motion of q will then be the ODE

$$q'(t) = P_{\mathrm{adm}(q(t))}[u_t] \tag{1.1}$$

(with $q(0)$ given). Equation (1.1), not easy from a mathematical point of view, was studied by Maury and Venel in [35, 36] in the framework of differential inclusions and of the so-called *Moreau Sweeping process* (see, for instance, [42]). Indeed, (1.1) can be written as the differential inclusion

$$q'(t) \in u_t - N_K(q(t))$$

where N_K is the normal cone to the set K:

$$N_K(q_0) = \{v \,:\, q_1 \in K \Rightarrow v \cdot (q_1 - q_0) \leq o(|q_1 - q_0|)\}.$$

It is also possible to discretize in time this equation via the so-called *catching-up algorithm* as follows

$$\tilde{q}^\tau_{n+1} = q^\tau_n + \tau u_{n\tau}, \quad q^\tau_{n+1} = P_K[\tilde{q}^\tau_{n+1}],$$

(for a small time step $\tau > 0$ which will tend to 0), where P_K is the projection onto the set K (note the difference between the projection in the space of velocities onto $\mathrm{adm}(q)$ and the projection in the configuration space onto K). The same algorithm is also called *prediction-correction algorithm* or, in a less specific way, *splitting* algorithm.

It is important here to notice that K, even if not convex in Ω^N, is as at least *prox-regular* (the projection on K is well defined on a neighborhood of K), which makes the above differential inclusion and the above discrete scheme well-defined.

All in all, we exactly have a perturbed sweeping process with exterior forcing u. The mathematical and numerical analysis, with efficient numerical simulations with impressively many particles, has been done in [35, 36].

We are now interested in the simplest continuous counterpart of this microscopic model. We do not claim at all that this is any kind of homogenized limit of the microscopic case, but we only propose it as an easy re-formulation in a density setting. In this case

- the particles population will be described by a probability density $\varrho \in \mathcal{P}(\Omega)$ (a measure, normalized to unit mass),
- the constraint becomes a maximal density constraint $\varrho \leq 1$ (we define the set $K = \{\varrho \in \mathcal{P}(\Omega) \,:\, \varrho \leq 1\}$, which means that measures in K are absolutely continuous and their density is smaller than 1),
- the set of admissible velocities will be described by the sign of the divergence on the saturated region $\{\varrho = 1\}$: $\mathrm{adm}(\varrho) = \left\{v : \Omega \to \mathbb{R}^d \,:\, \nabla \cdot v \geq 0 \ \text{ on } \{\varrho = 1\}\right\}$ (in a sense to be made precise, because we are considering divergences of non-smooth vector fields),
- the projection P in the space of velocities will be either the projection in $L^2(\mathcal{L}^d)$ (where $\mathcal{L}^d$ denotes the d-dimensional Lebesgue measure) or in $L^2(\varrho)$, but this turns out to be the same, since the only relevant zone is $\{\varrho = 1\}$,
- the motion of the crowd will be described by the continuity equation

$$\partial_t \varrho_t + \nabla \cdot \left(\varrho_t \left(P_{\mathrm{adm}(\varrho_t)}[u_t]\right)\right) = 0, \tag{1.2}$$

  with no-flux boubdary conditions on $\partial\Omega$.

The main difficulties in studying Equation (1.2) come from the lack of regularity: the vector field $v = P_{\mathrm{adm}(\varrho_t)}[u_t]$ is neither smooth itself (since it is obtained as an L^2 projection, and may only be expected to be L^2 a priori), nor it depends continuously on ϱ (it is very sensitive to small changes in the values of ϱ: passing from a density 1 to a density $1 - \varepsilon$ completely modifies the saturated zone, and hence the admissible set of velocities and the projection onto it).

Before introducing an approximation scheme for this equation, we want to make precise the definitions above, in particular in what concerns the divergence. Indeed, it is more convenient to describe $\mathrm{adm}(\varrho)$ by duality:

$$\mathrm{adm}(\varrho) = \left\{ v \in L^2(\varrho) \ : \ \int v \cdot \nabla p \leq 0 \quad \forall p \in H^1(\Omega) \ : \ p \geq 0,\, p(1-\varrho) = 0 \right\}.$$

In this way we characterize $v = P_{\mathrm{adm}(\varrho)}(u)$ through

$$u = v + \nabla p, \quad v \in \mathrm{adm}(\varrho), \quad \int v \cdot \nabla p = 0, \quad p \in \mathrm{press}(\varrho) := \{p \in H^1(\Omega),\, p \geq 0,\, p(1-\varrho) = 0\},$$

where $\mathrm{press}(\varrho)$ is the space of functions p used as test functions in the dual definition of $\mathrm{adm}(\varrho)$. They play the role of pressures affecting the movement. The two cones $\nabla\mathrm{press}(\varrho)$ (defined as the set of gradients of elements of $\mathrm{press}(\varrho)$) and $\mathrm{adm}(\varrho)$ are in duality for the L^2 scalar product (i.e. one is defined as the set of vectors with a negative scalar product with all the elements of the other). This allows for the orthogonal decomposition $u_t = v_t + \nabla p_t$, and gives an alternative expression of Equation (1.2), i.e.

$$\begin{cases} \partial_t \varrho_t + \nabla \cdot \big(\varrho_t (u_t - \nabla p_t)\big) = 0, & \text{in } (0,T)\times\Omega \\ \varrho_t (u_t - \nabla p_t)\cdot \mathbf{n} = 0 & \text{on } (0,T)\times\partial\Omega \\ 0 \leq \varrho \leq 1,\, p \geq 0,\, p(1-\varrho) = 0. \end{cases} \tag{1.3}$$

Note (details are in [21], for instance), that the orthogonality condition $\int (u_t - \nabla p_t)\cdot \nabla p_t = 0$ has disappeared, since it is guaranteed, for a.e. t, by the other conditions. Finally, the term $\nabla\cdot(\varrho_t \nabla p_t))$ can be replaced by Δp_t since anyway the pressure p_t and its gradient are non-zero only on the saturated region $\{\varrho_t = 1\}$.

It is interesting to see that the above equation strongly recalls the density dynamics formulation of the Hele-Shaw flow. In order to present it in few words, consider an equation of the form

$$\partial_t \varrho_t - \Delta p_t = G,$$

where the term G possibly stands for reaction terms, and p and ϱ are bound by the condition $p_t \in H(\varrho_t)$ where H is a monotone graph. When H is given by $H(s) = s^m$ we have a *porous-medium equation* (a form of non-linear diffusion, see [48]), and when $H = \partial I_{[0,1]}$ (hence $H(s) = [0,+\infty[$ for $s = 1$, $H(s) = \{0\}$ for $0 < s < 1$), the pressure $p \geq 0$ is arbitrary, but satisfies $p(1-\varrho) = 0$, exactly as in (1.3). We refer, for instance, to [11,16] and to the more recent [44] for these equations.

In particular, when $G = \varrho\tilde{G}$, with $\tilde{G} \geq 0$, and $\varrho_0 = \mathbb{1}_{\Omega_0}$ is a patch (i.e. no intermediate density between 0 and 1 is allowed), the evolution is of the form $\varrho_t = \mathbb{1}_{\Omega_t}$ with Ω_t evolving with normal velocity

$$v_t = -\frac{\partial p_t}{\partial \mathbf{n}}, \quad -\Delta p_t = G \text{ in } \Omega_t,\ p_t = 0 \text{ on } \partial\Omega_t.$$

This free-boundary formulation of the Hele-Shaw flow, usually studied in terms of viscosity solutions, could be more familiar for some readers (see, for instance, [30,31])

Our equation (1.3) has the same form, but we replace reaction terms with an advection term. This kind of terms happens to be much less studied in the Hele-Shaw literature and lets different difficulties appear, in particular when studying the uniqueness of the solution (see Section 3). Moreover, it seems that the approach with optimal transport (which will be the key to provide the existence of a solution in the next section) has not yet been exploited in the Hele-Shaw community.

2. Existence and approximation: the role of optimal transport

We will present in this section an iterative scheme, together with some variants, which provide time-discrete approximations of the solution of (1.3).

First, we fix a time step $\tau > 0$: we look for a sequence $(\varrho_n^\tau)_n$ where ϱ_n^τ will represent the measure ϱ_t at time $t = n\tau$. To pass from ϱ_n^τ to ϱ_{n+1}^τ, we define a splitting scheme in the space of measures

$$\tilde{\varrho}_{n+1}^\tau = (id + \tau u_{n\tau})_\# \varrho_n^\tau \; ; \quad \varrho_{n+1}^\tau = P_K[\tilde{\varrho}_{n+1}^\tau]. \tag{2.1}$$

Here $_\#$ denotes the push-forward (or image) measure, and the important point is to choose the metric in the projection P_K. The vector field $u_{n\tau}$ is the evaluation of the time-dependent vector field u_t at time $t = n\tau$. The key point is using the W_2 projection (where W_2 is the quadratic Wasserstein distance, induced by optimal transport, see below), instead of L^2 or other Hilbertian projections. Indeed, only the distance W_2 corresponds to the L^2 projection of velocity fields and of (Lagrangian) positions.

2.1. **Optimal transport and Wasserstein distances**

For the whole theory of optimal transport we refer to [49] or [46]. Here we only give the most basic ingredients that we need, with no proof. If two probabilities $\mu, \nu \in \mathcal{P}(\Omega)$ are given on a compact domain, the Monge problem, [41], for the quadratic cost reads

$$\inf \left\{ \int \frac{1}{2}|x - T(x)|^2 \mathrm{d}\mu \; : \; T : \Omega \to \Omega, \; T_\# \mu = \nu \right\}. \tag{2.2}$$

The formulation given by Kantorovich, [29], is a relaxation of the above optimization problem

$$\inf \left\{ \int \frac{1}{2}|x - y|^2 \mathrm{d}\gamma \; : \; \gamma \in \mathcal{P}(\Omega \times \Omega), \, (\pi_x)_\# \gamma = \mu, (\pi_y)_\# \gamma = \nu \right\}, \tag{2.3}$$

where π_x and π_y are the two projections on $\Omega \times \Omega$ onto its factors (hence, the admissible measures γ, called *transport plans*, are those with prescribed marginals μ and ν). Since this second problem is convex, it is possible to find a dual (this is the main contribution by Kantorovich), which reads

$$\sup \left\{ \int \phi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu \; : \; \phi(x) + \psi(y) \leq \frac{1}{2}|x - y|^2 \right\}. \tag{2.4}$$

It is possible to prove that both the primal and the dual Kantorovich problems admit solutions and that, whenever $\mu \ll \mathcal{L}^d$, then the optimal γ in the Kantorovich problem is of the form $\gamma = (id, T)_\# \mu$ (it is concentrated on the graph of a map $T : \Omega \to \Omega$), and this map T, which is optimal for the Monge problem, is related to the optimal pair (ϕ, ψ) in the dual problem via $T(x) = x - \nabla\phi(x)$. The optimal ϕ is called *Kantorovich potential*, is Lipschitz continuous (which guarantees the differentiability a.e.), and is such that $u(x) = |x|^2/2 - \phi(x)$ is convex. In particular the optimal transport map is of the form $T = \nabla u$, i.e. it is the gradient of a convex function. This is a well-known result by Brenier [12, 13].

Moreover, one can use the values of the above problems (which all coincide, at least when $\mu \ll \mathcal{L}^d$) in order to define a distance on probability measures. More precisely, we define

$$W_2(\mu, \nu) := \sqrt{\inf \left\{ \int |x - y|^2 \mathrm{d}\gamma \; : \; \gamma \in \mathcal{P}(\Omega^2), \, (\pi_x)_\# \gamma = \mu, (\pi_y)_\# \gamma = \nu \right\}}.$$

Hence, the values in (2.2),(2.3),(2.4), which depend on μ and ν, are equal to $\frac{1}{2}W_2^2(\mu, \nu)$. Then, it is possible to prove that W_2, called *Wasserstein distance* is a distance on $\mathcal{P}(\Omega)$ which metrizes the weak-* convergence of probabilities (on compact domains; see Chapter 5 in [46]), and use this distance to define projections.

A final important property that we need about Wasserstein distances and Kantorovich potentials is the following (see Chapter 7 in [46]): if ν is fixed and one defines $\varrho_\varepsilon := (1-\varepsilon)\varrho + \varepsilon\tilde\varrho$, then we have

$$\frac{d}{d\varepsilon}\left(\frac{1}{2}W_2^2(\varrho_\varepsilon,\nu)\right)_{|\varepsilon=0} = \int \phi\, d(\tilde\varrho - \varrho),$$

where ϕ is the Kantorovich potential in the transport from ϱ to ν, provided it is unique up to additive constants. This means that ϕ is the first variation of the squared Wasserstein distance.

2.2. **Wasserstein projections**

The projection problem, which is a crucial tool in all the analysis of this paper, reads as follows: fix a measure $\nu \in \mathcal{P}(\Omega)$ and solve

$$\min\left\{\frac{1}{2}W_2^2(\varrho,\nu)\ :\ \varrho\in K\right\} \quad = \quad \min_{\varrho\le 1}\left\{\sup_{\phi,\psi}\int \phi\,\mathrm{d}\varrho + \int\psi\,\mathrm{d}\nu\right\}.$$

The right-hand side of the above inequality shows that this is a convex problem, and it is possible to characterize its solution via the optimality condition. As usual, we can say that, if $\bar\varrho$ is a solution of the above projection problem, then $\bar\varrho$ must also minimize, under the same constraints, the linearization (around $\varrho=\bar\varrho$) of the functional $\varrho\mapsto\frac{1}{2}W_2^2(\varrho,\nu)$. Hence, it is possible to prove (see [33]) that there exists a function $\bar\phi$, which is a Kantorovich potential in the transport from $\bar\varrho$ to ν, such that $\bar\varrho$ also solves

$$\min\left\{\int\bar\phi\,\mathrm{d}\varrho\ :\ \varrho\le 1\right\}.$$

The solutions of the above problem are easy to characterize: there should exist a constant ℓ such that

$$\bar\varrho = \begin{cases} 1 & \text{on } \bar\phi<\ell,\\ 0 & \text{on } \bar\phi>\ell,\\ \in[0,1] & \text{on } \bar\phi=\ell\end{cases}$$

In particular, we can define $p:=(\ell-\phi)_+$ and observe that we have $p\in\mathrm{press}(\bar\varrho)$ and, passing to gradients, we have $\bar\varrho-$a.e. $\nabla p = -\nabla\phi = T(x)-x$, where $T(x)=x+\nabla p(x)$ is the optimal transport from $\bar\varrho$ to ν.

It is useful to understand why this projection allows, when applied to the setting of the splitting scheme described so far, to recover the desired PDE, i.e. why we chose the distance W_2 in the projection step.

Indeed, in this case T transports ϱ^τ_{n+1} to $\tilde\varrho^\tau_{n+1}$. We first note that we have

$$||\nabla p||_{L^2(\varrho^\tau_{n+1})} = W_2(\varrho^\tau_{n+1},\tilde\varrho^\tau_{n+1}) \le W_2(\varrho^\tau_n,\tilde\varrho^\tau_{n+1}) \le \tau||u_{n\tau}||_{L^2(\varrho^\tau_n)}.$$

This suggest to scale the pressure (we call it now τp) and get the situation in Figure 1. Formally, we have $(id+\tau u_{n\tau})^{-1}(id+\tau\nabla p) = id - \tau(u_{(n+1)\tau}-\nabla p)+o(\tau)$ provided u is regular enough. This allows to get, in the limit $\tau\to 0$, the vector field $v_t = P_{\mathrm{adm}(\varrho_t)}[u_t]$ and get a solution of the PDE.

Before going on, we want to summarize some properties of the projection operator P_K. Its properties are essential for proving convergence. To be more general, we will describe the properties of the operator $P_{K(f)}$, defined via $P_{K(f)}[\nu] := \mathrm{argmin}\{W_2(\varrho,\nu), \varrho\le f\}$. Here is what we know about $P_{K(f)}$:

- $W_2^2(\cdot,\nu)$ is strictly convex as soon as $\nu\ll\mathcal{L}^d$ (see Chapter 7 in [46]). This provides uniqueness of the projection, and hence continuity of the projection operator (for arbitrary f, if ν is absolutely continuous).
- Uniqueness actually holds for every ν, in the case $f\ll\mathcal{L}^d$ (see [20]) as a consequence of the fact that the projection must be of the form $\varrho = f\mathbb{1}_A + \nu\mathbb{1}_{A^c}$.

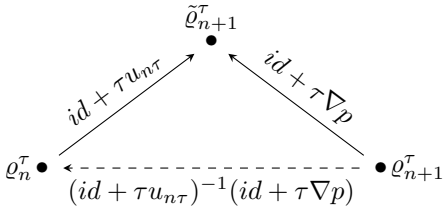

FIGURE 1. The splitting scheme

- For $f = 1$, when Ω is convex, the geodesic convexity of $\{\varrho : \varrho \leq 1\}$ (w.r.t. Wasserstein geodesics, see Section 3) also gives uniqueness, and Hölder continuity w.r.t. W_2 (see [33], but in particular Proposition 2.3.4 in [45]).
- Lipschitz continuity of $P_{K(f)}$ is an open question, even for $f = 1\ldots$
- The projection preserves ordering (in order to give a meaning to this statement, one needs to extend $P_{K(f)}$ to measures with arbitrary finite mass, and not only fwor probabilities) and decreases the L^1 distance between densities (these facts are folklore among specialists, but a precise written reference is unfortunately missing; they are true for every $f \ll \mathcal{L}^d$).
- Estimates (order 0): for every convex U, $\varrho \mapsto \int f(\varrho(x))\mathrm{d}x$ decreases under projection, in the case $f = 1$ (see [40]).
- Estimates (order 1): the BV norm decreases under projection in the case $f = 1$ (see [20]; a corresponding estimate, involving the BV norm of f also exists for non-constant f, but reads: $||P_{K(f)}[\nu]||_{BV} \leq ||\nu||_{BV} + 2||f||_{BV}$, and the constant 2 is sharp). On the other hand, $W^{1,p}$ norms do not decrease under projection, and it can happen that $\varrho \notin W^{1,p}$ while $\nu \in W^{1,p}$ (as in the example illustrated in Figure 2, where the projection $\varrho = P_K[\nu]$ has jumps which were absent in ν).

2.3. **Convergence of the scheme and diffusive variants**

We summarize here some convergence statements for the splitting scheme. In all these statements, what we do is the following: we take an initial measure $\varrho_0 \in \mathcal{P}(\Omega)$ and a time horizon $T < +\infty$; we iterate the splitting scheme defining the two sequences $\tilde{\varrho}^\tau_n$ and ϱ^τ_n as soon as $n\tau \leq T$; then, we consider the piecewise constant interpolation $\varrho^\tau_t = \varrho^\tau_{n+1}$ for $t \in]n\tau, (n+1)\tau]$. The goal is to prove that up to subsequences, as $\tau \to 0$, the curves $t \mapsto \varrho^\tau_t$ uniformly converge on $[0, T]$ (as curves valued in the Wasserstein space with distance W_2) to a curve $t \mapsto \varrho_t$ which solves the desired equation. The proofs proposed in [21, 33, 40, 46] actually use several interpolation sequences, but this is just a technical tool and we will not enter here into the details of the proof. Also, the convergence as $\tau \to 0$ is essentially a weak convergence (since we mentioned convergence for W_2) and this requires a special attention for the convergence of non-linear terms, involving in particular the pressure. The non-linear term $\nabla \cdot (\varrho\nabla p)$ in the equation has been transformed, as we said in Section 1, into an easier

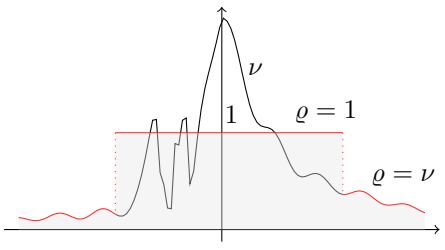

FIGURE 2. An example of projection

term Δp, but the non-linearity is preserved in the constraint $p(1-\varrho)=0$. In order to handle it, the main tool exploits H^1 bounds in x on the pressure and a variant of the so-called Aubin-Lions lemma (see [5]), which allows to combine the bounds in time and space. For the precise proofs in the framework of density-constrained equations, we refer to Step 3 of Section 3.2 in [33], Lemma 3.5 in [40], and Lemma 4.6 in [21]. A similar argument is also performed in Lemma 1 in [14].

We first start from the statement related to the scheme we described so far.

Theorem 2.1. *Given $\varrho_0 \in K$ and a time-dependent velocity field, define the splitting scheme according to* (2.1)*. Suppose that u is continuous in (t,x), and C^1 in x, uniformly in t. Then, the scheme converges up to a subsequence to a solution of* (1.3)*.*

This statement is given in [45] (Theorem 2.3.6), with two extra assumptions. First, Ω was required to be convex, which is not really necessary (it was used to guarantee the uniqueness of the projection, but this can also be obtained in other ways, as we saw above). Then, there was the assumption that $(id+\tau u)(\Omega)\subset\Omega$ for small τ (which amounts to a condition on u on $\partial\Omega$). This was necessary so that the first step in the scheme does not go out of Ω, but is not crucial neither. We can define the projection step also for measures not supported on Ω, but on $\Omega'\supset\Omega$, and project on the set of those which belong to K and are concentrated on Ω. Also note that, in this case, close to the regions where u points outwards there will be instantaneous creation of a saturated region $\{\varrho=1\}$.

We now observe that taking $\tilde\varrho^\tau_{n+1}=(id+\tau u_{n\tau})_\#\varrho^\tau_n$ is just a possible choice. When u is not regular enough or depends on ϱ there are better options, that we will describe in a more general framework. We will consider a possible volatily in the equation, by replacing the first equation in (1.3) with

$$\partial_t\varrho_t+\nabla\cdot(\varrho_t u_t)-\sigma\Delta\varrho_t-\Delta p_t=0 \tag{2.5}$$

where $\sigma\geq 0$ is a volatility. Again, we consider a no-flux boundary condition on $\partial\Omega$, which reads in this case $(\varrho_t u_t-\nabla p_t-\sigma\nabla\varrho_t)\cdot\mathbf{n}=0$.

In this case, we can define a splitting scheme by first taking the solution of the Fokker-Planck equation on the interval $(n\tau,(n+1)\tau)$:

$$\begin{cases}\partial_s\varrho_s+\nabla\cdot(u_s\varrho_s)-\sigma\Delta\varrho_s=0, & \text{in } (n\tau,(n+1)\tau)\times\Omega\\ (\varrho_t u_t-\sigma\nabla\varrho_t)\cdot\mathbf{n}=0, & \text{on } \partial\Omega\\ \varrho_{n\tau}=\varrho^\tau_n;\end{cases} \tag{2.6}$$

and define $\tilde\varrho^\tau_{n+1}=\varrho_{(n+1)\tau}$ (this means that we first follow the Fokker-Planck equation ignoring the density constraint, we do it over an interval of time of length τ, and that this is the measure we project). Note that in this case, due to definition of the first step via the solution of the PDE and the effect of the diffusion, the behavior of u on $\partial\Omega$ is not relevant.

In [40] a detailed proof of convergence is presented in the case $\sigma>0$, thus obtaining:

Theorem 2.2. *Given $\varrho_0\in K$ and a time-dependent vector field $u\in L^\infty$, suppose $\sigma>0$ and define the splitting scheme as above. Then, the scheme converges up to a subsequence to a solution of* (2.5)*.*

It is interesting to see that this schemes converges under much weaker assumptions on u than what required in Theorem 2.1. This is partially due to the effect of the diffusion, and partially to better properties of the scheme. It has not been proven rigorously, but we expect the same scheme to converge even with $\sigma=0$ provided u satisfies the assumptions for the equation (2.6) to be well-posed (in the sense of admitting existence and stability properties). It should at least work whenever u satisfies the assumptions of the DiPerna-Lions (or Ambrosio) theory (i.e. $u\in W^{1,1}$ or $u\in BV$ and some lower L^∞ bounds on $\nabla\cdot u$, see [3, 23]), but we will not develop here the details.

2.4. **The gradient flow case**

An interesting framework, which is actually the original one studied in [33], where the scheme can be made easier and the results require weaker assumptions on u occurrs when u has a suitable gradient structure. In this case it is possible to do the two steps of the splitting algorithm at once, thanks to the theory of gradient flows.

First, let us recall in few words the general theory about gradient flows (the interested reader can look at [4] or [47]). Consider the simple ODE.

$$x'(t) = -\nabla F(x(t))$$

(which follows the steepest descent lines of a function $F : \mathbb{R}^n \to \mathbb{R}$). We can discretize in time such an equation by solving

$$x^\tau_{n+1} \in \mathrm{argmin}_x \, F(x) + \frac{1}{2\tau}|x - x^\tau_n|^2, \quad \text{for } \tau > 0 \text{ fixed.}$$

The optimal x^τ_{n+1} satisfies

$$\frac{x^\tau_{n+1} - x^\tau_n}{\tau} + \nabla F(x^\tau_{n+1}) = 0$$

which corresponds to the implicit Euler scheme for $x' = -\nabla F(x)$, the solution being found as a limit $\tau \to 0$. The important point (see [2,19]) is that this formulation may easily be adapted to a general metric space (X, d), by replacing the Euclidean distance $|x-x^\tau_n|$ with the distance d in X (as it was introduced in [2,19], under the name *minimizing movements*). In particular, we can look at what happens whenever we take $(X, d) = (\mathcal{P}(\Omega), W_2)$.

Let F be a functional over $(\mathcal{P}(\Omega), W_2)$, and let us follow the so-called JKO scheme (see [28]):

$$\varrho^\tau_{n+1} \in \mathrm{argmin}_\varrho \, F(\varrho) + \frac{W_2^2(\varrho, \varrho^\tau_n)}{2\tau}$$

As in the Euclidean case above, in order to find the equation solved by the limit curve we need to write down the optimality conditions at each time step of the discrete scheme. Let us define, for given F, the first variation of F as the function $\delta F/\delta\varrho$ such that, setting again $\varrho_\varepsilon := (1-\varepsilon)\varrho + \varepsilon\tilde\varrho$, we have

$$\frac{d}{d\varepsilon}\left(F(\varrho_\varepsilon)\right)_{|\varepsilon=0} = \int \frac{\delta F}{\delta \varrho}\, d(\tilde\varrho - \varrho)$$

(see Chapter 7 in [46] for precise definitions). Then, using also the considerations about the first variation of the Wasserstein distance, we can say that the optimality conditions of the above minimization problem read

$$\frac{\delta F}{\delta \varrho}(\varrho^\tau_{n+1}) + \frac{\phi}{\tau} = const.$$

This implies that, if we define a vector field v via $v(x) := (x - T(x))/\tau$, we have

$$v(x) = \frac{x - T(x)}{\tau} = \frac{\nabla\phi(x)}{\tau} = -\nabla\big(\frac{\delta F}{\delta\varrho}(\varrho)\big). \tag{2.7}$$

Here, the vector field v represents a discrete velocity (in the sense of a ratio displacement / time step), and we can guess that, at the limit $\tau \to 0$, the limit curve will solve a continuity equation of the form $\partial_t\varrho + \nabla\cdot(\varrho v) = 0$, with v given as in (2.7), i.e.

$$\partial_t\varrho - \nabla\cdot\left(\varrho\,\nabla\Big(\frac{\delta F}{\delta\varrho}(\varrho)\Big)\right) = 0.$$

An example is given by $F(\varrho) = \int f(\varrho(x))\mathrm{d}x$ (for U convex and superlinear: a functional which is set to $+\infty$ when ϱ is not absolutely continuous). Then $\frac{\delta F}{\delta\varrho}(\varrho) = f'(\varrho)$ and the equation becomes the second-order non-linear diffusion PDE

$$\partial_t\varrho - \nabla\cdot\big(\varrho\,\nabla f'(\varrho)\big) = 0.$$

For instance, for $f(t) = t\log t$ we get $\nabla f'(\varrho) = \frac{\nabla\varrho}{\varrho}$, which gives the heat equation $\partial_t\varrho - \Delta\varrho = 0$. Using $f(t) = t^m/(m-1)$ we get $f'(\varrho) = \frac{m}{m-1}\varrho^{m-1}$ and $\nabla f'(\varrho) = m\varrho^{m-2}\nabla\varrho$, so that the equation that we obtain is the porous medium equation $\partial_t\varrho - \Delta(\varrho^m) = 0$ (see [43] for the gradient-flow interpretation of this equation).

For $F(\varrho) = \int V(x)\mathrm{d}\varrho$ we get $\frac{\delta F}{\delta\varrho}(\varrho) = V$. We can obtain the Fokker-Planck equation in the case $F(\varrho) = \int V(x)\mathrm{d}\varrho + \int \varrho\log\varrho\ldots$

In the case $u = -\nabla V$, the equation (1.3) is the gradient flow of the functional

$$F(\varrho) = \begin{cases} \int V(x)\mathrm{d}\varrho & \text{if } \varrho \in K, \\ +\infty & \text{if not.} \end{cases} \tag{2.8}$$

By the way, this functional is the limit as $m \to \infty$ of $\int(\frac{1}{m-1}\varrho(x)^m + V(x)\varrho(x))\mathrm{d}x$, which shows that (1.3) can be seen as the limit of the porous medium equation when the exponent tends to ∞ (as in [44] and in [1], where explicit rate of convergence are also presented). For the diffusive variant, just add $\sigma\int\varrho(x)\log\varrho(x)\mathrm{d}x$.

The result proven in [33] can be summarized as follows:

Theorem 2.3. *Given $\varrho_0 \in K$ and $u = -\nabla V$, define $(\varrho_n^\tau)_n$ according to the JKO scheme with the choice of F in* (2.8)*. Suppose that $V \in W^{1,1}(\Omega)$. Then, the scheme converges up to a subsequence to a solution of* (1.3)*.*

Note that the original statement in [33] required many useless assumptions, in particular convexity of the domain Ω and λ-convexity of V. Both these assumptions were necessary to guarantee λ-convexity (in the geodesic sense) of the functional and ensure, for instance, uniqueness of the minimizer in the JKO scheme (but this uniqueness is not crucial, and can also be proven differently, using standard convexity). The assumption $V \in W^{1,1}(\Omega)$ is needed in order to handle the terms $\varrho\nabla V$ in the equation (we have $\varrho \in L^\infty$ and $\nabla V \in L^1$), and this is all we need, with the same proof stategy as in [33]. We also insist on the fact that, thanks to the gradient structure of u, the regularity we require on u itself is very weak, since it is only supposed to be L^1. Finally, also note that it would not be difficult to handle the case of a time-dependent gradient vector field $u_t = -\nabla V_t$, provided the dependence in t is nice ($t \mapsto V_t$ needs to be Lipschitz for the L^1 norm).

Also, we want to stress that the gradient flow case is indeed a very natural one for modeling purposes. Indeed, the first case which has been considered was exactly the following: $u = -\nabla d(\cdot,\Gamma)$, where $\Gamma \subset \partial\Omega$ is a subset of the boundary standing for the exit. On the other hand, the λ-convexity assumption on V needed to be removed if one wanted to consider V defined as the *geodesic distance* to the exit in a non-convex domain Ω (in case of obstacles inside the evacuation area).

2.5. **Moving domains**

We finish this section by briefly addressing a variant, presented in [21] and inspired by the Moreau sweeping process (see [42]). We will consider the sweeping of a probability measure ϱ_t in a moving convex set $\Omega(t)$, with the constraint $\varrho_t \le 1$. Considering a large domain Ω' which contains all the $\Omega(t)$, the set of probability measures we are interested in is now

$$K(t) = \{\varrho \in \mathcal{P}(\Omega')\ ,\ \ \mathrm{spt}\,\varrho \subset \Omega(t)\ :\ \varrho \le 1\}. \tag{2.9}$$

By the word "sweeping", in the spirit of Moreau, we mean that the particles of ϱ_t try to stay at rest, and only move in order to satisfy the constraint. These particles are either pushed by the boundary of the moving domain $\Omega(t)$ (as it happens in the standard sweeping process) or by the other particles (in particular, those which are interposed between them and the boundary).

Given a density $\varrho \in \mathcal{P}(\Omega(t))$ such that $\varrho \le 1$, we can heuristically describe the set of "admissible" velocities $\mathrm{adm}(\varrho, C)$, by requiring $\nabla\cdot v \ge 0$ on the set $\{\varrho = 1\}$ (in order to preserve the density constraint) and taking care of the fact that, on the boundary, the inward normal velocity of the particles must be at least that of the boundary: this sums up as $v\cdot\mathbf{n} \le v_{\partial\Omega}\cdot\mathbf{n}$ where $v_{\partial\Omega}$ is the boundary velocity (we will denote $V = v_{\partial\Omega}\cdot\mathbf{n}$). As

usual, this will be expressed by duality. We consider the following formal computation, for $p \in \mathrm{press}(\varrho)$:

$$\int_{\Omega(t)} v \cdot \nabla p \, \mathrm{d}x = -\int_{\Omega(t)} (\nabla \cdot v) p \, \mathrm{d}x + \int_{\partial\Omega(t)} p(v \cdot \mathbf{n}) \leq \int_{\partial l\Omega(t)} p \, V;$$

this leads us to the following definition:

$$\mathrm{adm}(\varrho, t) = \left\{ v \in L^2(\Omega, \varrho) \ : \ \int_{\Omega(t)} v \cdot \nabla p \, \mathrm{d}x \leq \int_{\partial\Omega(t)} p \, V \quad \text{for all } p \in \mathrm{press}(\varrho) \right\},$$

where the dependence of $\mathrm{adm}(\varrho, t)$ in t includes both the domain $\Omega(t)$ and its normal velocity V (which is a function defined on $\partial\Omega(t)$). Now, the evolution equation we want to solve is a continuity equation where the velocity has to be admissible, and have minimal L^2 norm (it is the projection of the 0 velocity field)

$$\begin{cases} \partial_t \varrho_t + \nabla \cdot (\varrho_t P_{\mathrm{adm}(\varrho_t, t)}[0]) = 0, \\ \mathrm{spt}(\varrho_t) \subset \Omega(t), \varrho_t \leq 1. \end{cases} \tag{2.10}$$

Using the dual characterization of the cone $\mathrm{adm}(\varrho_t, t)$, it is possible to translate this system into

$$\begin{cases} \partial_t \varrho_t = \Delta p_t, \\ 0 \leq \varrho_t \leq 1, \, p_t \geq 0, \, p_t(1 - \varrho_t) = 0, \\ \mathrm{spt}(\varrho_t) \subset \Omega(t). \end{cases} \tag{2.11}$$

Note that the very last condition ($\mathrm{spt}(\varrho_t) \subset \Omega(t)$) includes the fact that no mass can exit the boundary $\partial\Omega(t)$, which, considering that this boundary is itself moving, means $-\nabla p_t \cdot \mathbf{n} \leq -V$ on $\partial\Omega(t)$.

The discrete scheme used in [21] to find a solution of (2.10) is inspired from [42]: given $\varrho_n^\tau \in \mathcal{P}(\Omega')$, define

$$\varrho_{n+1}^\tau := P_{K((n+1)\tau)}[\varrho_n^\tau].$$

The convergence result is not surprising:

Theorem 2.4. *Given $\varrho_0 \in K(0)$ define $(\varrho_n^\tau)_n$ according to the above projection scheme. Suppose that all the sets $\Omega(t)$ are convex, that $\inf_{t\in[0,T]} |\Omega(t)| > 1$, and that $t \mapsto \Omega(t)$ is Lipschitz continuous for the Hausdorff distance. Then, the scheme converges up to a subsequence to a solution of* (2.11).

3. Few words about uniqueness

An easy computation in the Euclidean gradient flow $x' = -\nabla F(x)$ shows that uniqueness of the solution is guaranteed as soon as F is semi-convex, i.e. for some $\lambda \in \mathbb{R}$ the function $F(x) - \lambda|x|^2/2$ is convex (which is the same, for smooth functions, of $D^2F \geq \lambda I \ldots$ we say in this case that F is λ-convex). Indeed, take $x^1(t)$ and $x^2(t)$ two different solutions of $x' = -\nabla F(x)$, set $E(t) = |x^1(t) - x^2(t)|^2$, and compute

$$E'(t) = 2\big(x^1(t) - x^2(t)\big) \cdot \big((x^1)'(t) - (x^2)'(t)\big) = -2\big(x^1(t) - x^2(t)\big) \cdot \big(\nabla F(x^1(t)) - \nabla F(x^2(t))\big) \leq -2\lambda E(t),$$

and apply a Gronwall lemma to deduce $|x^1(t) - x^2(t)| \leq e^{-\lambda t}|x^1(0) - x^2(0)|$.

A very similar result is available in the general theory of gradient flows in metric spaces, [4]. Without entering into detais, what is needed is geodesical semi-convexity. In the case of the metric space W_2, this means that for some $\lambda \in \mathbb{R}$ the function $s \mapsto g(s) := F(\varrho_s)$ satisfies $g'' \geq \lambda W_2^2(\varrho_0, \varrho_1)$ whenever ϱ_s is a constant-speed geodesic connecting ϱ_0 to ϱ_1. This condition is quite easy to check, since in W_2 we know the geodesic curves (see, for

instance, chapter 5 in [46]): the geodesic connecting μ to ν has the form $\varrho_s = (id - s\nabla\phi)_{\#}\varrho_0 = ((1-s)id + sT)_{\#}\varrho_0)$ where T is the optimal transport map from μ to ν.

Also, for many functionals F it is well-known whether they are or not geodesically (semi)-convex (see, for instance, Chapter 7 in [46]). The functional $\varrho \mapsto \int V d\varrho$ is λ-geodesically convex if and only if the function V itself is λ-convex; for the functional $\varrho \mapsto \int f(\varrho(x))\mathrm{d}x$ the situation is trickier, and the sharp conditions have been found by McCann (see [37]). Such a functional is geodesically convex (i.e., for $\lambda = 0$) if and only if f is such that $s \mapsto s^d f(s^{-d})$ is convex and decreasing, which is the case in many examples, including $f(s) = s^m$ for $m > 1$, $f(s) = s \log s \ldots$. In particular, this means that the set of densities with $\int \varrho^m \leq C$ is geodesically convex, and, for $m \to \infty$, the same is true for the set K of densities bounded above by 1.

The general technique of [4] allows to obtain the estimate

$$\frac{d}{dt} W_2^2(\varrho_t^1, \varrho_t^2) \leq -2\lambda W_2^2(\varrho_t^1, \varrho_t^2),$$

whenever the curves ϱ_t^i are two gradient flows of F, and F is λ-geodesically convex. Using the fact that the constraint $\varrho \in K$ is geodesically convex, this applies to the case $u = -\nabla V$, provided $D^2 V \geq \lambda I$.

In order to consider more general velocity fields u (i.e., of non-gradient type), we can try to directly differentiate the Wasserstein distance between two solutions of (1.3). This requires to use the formula for differentiating the W_2 distance between two curves (see Chapter 5 in [46]).

Take two solutions (ϱ_t^i, p_t^i) and let ϕ and ψ be the Kantorovich potentials between ϱ_t^1 and ϱ_t^2, with $T(x) = x - \nabla\phi(x)$. We can compute

$$\begin{aligned} \frac{d}{dt}\frac{1}{2} W_2^2(\varrho_t^1, \varrho_t^2) &= \int \varrho_t^1 \nabla\phi \cdot (u - \nabla p_t^1) + \int \varrho_t^2 \nabla\psi \cdot (u - \nabla p_t^2) \\ &= \int (x - T(x)) \cdot (u(x) - u(T(x))\mathrm{d}\varrho_t^1 + \int p_t^1 \Delta\phi + p_t^2 \Delta\psi. \end{aligned}$$

If u satisfies $(u(x) - u(y)) \cdot (x - y) \leq -\lambda |x - y|^2$ (which is the case if $u = -\nabla V$ and $D^2 V \geq \lambda I$, but also if u is $(-\lambda)$-Lispchitz), the first term gives $-\lambda W_2^2(\varrho_t^1, \varrho_t^2)$; for the second, we have

$$p_t^1 > 0 \Rightarrow \varrho_t^1 = 1, \varrho_t^2 \leq 1 \Rightarrow \det(I - D^2\phi) = \det(\mathrm{d}t) \geq 1 \Rightarrow \Delta\phi \leq 0,$$

and this rest is negative (the term with p_t^2 can be dealt with in a similar manner). This computation (made precise in [22]) allows to prove the following result.

Theorem 3.1. *Suppose that u satisfies $(u(x) - u(y)) \cdot (x - y) \leq -\lambda |x - y|^2$ and let (ϱ^1, p^1) and (ϱ^2, p^2) be two solutions of (1.3). Then we have*

$$\frac{d}{dt} W_2^2(\varrho_t^1, \varrho_t^2) \leq -2\lambda W_2^2(\varrho_t^1, \varrho_t^2).$$

In particular, this gives $W_2(\varrho_t^1, \varrho_t^2) \leq W_2(\varrho_0^1, \varrho_0^2)$ and uniqueness of the solution for fixed initial datum.

This applies as soon as u satisfies some mild regularity conditions, which essentially suggest that it has to be Lipschitz (or "one-sided Lipschitz", i.e. a form of monotonicity). Yet, for many applications it is useful to consider vector fields u which do not satisfy this assumption, as it happens when $u = -\nabla V$ but V is not semi-convex, which is the case when obstacles are present in the domain.

Considering that, in the above theorem, the pressure term in (1.3) (or the density constraint) only improves the contractivity that we would have with u only, a natural question is whether we can say that uniqueness holds as soon as u is such to guarantee uniqueness for $\partial_t \varrho + \nabla \cdot (\varrho u) = 0$. This uniqueness is usually guaranteed by the DiPerna-Lions (or Ambrosio) theory for $u \in W^{1,1}$ (or $u \in BV$): see [3, 23]).

Yet the technique to prove uniqueness in the DiPerna-Lions theory is not at all based on W_2 estimates. It recalls much more an estimate for $\int |\varrho_t^1 - \varrho_t^2|$, i.e. an L^1-contraction result. The question then becomes whether it is possible to combine these two techniques, or directly prove that we have an L^1 contraction.

Note that this is what is usually done for the Hele-Shaw flow, where uniqueness comes from estimates on the adjoint equation (by the way, the technique based on the differentiation of the W_2 distance seems to be essentially unknown in the Hele-Shaw community, and uniqueness results in the presence of non-trivial drifts are not standard).

However, uniqueness for (1.3) with non-smooth vector fields u is so far only a conjecture, and the only result in this direction is the one obtained in the second part of [22], about the diffusive case ($\sigma > 0$ in (2.5)):

Theorem 3.2. *Given $u \in L^\infty$ and $\sigma > 0$, consider (ϱ^1, p^1) and (ϱ^2, p^2) two solutions of (2.5). Then we have*

$$\frac{d}{dt}||\varrho_t^1 - \varrho_t^2||_{L^1} \leq 0.$$

In particular, this gives $||\varrho_t^1 - \varrho_t^2||_{L^1} \leq ||\varrho_0^1 - \varrho_0^2||_{L^1}$, and uniqueness of the solution for fixed initial datum.

Without diffusion the same question is open, even if the discrete steps in the splitting scheme are indeed an L^1 contraction (but, to pass this to the limit on arbitrary solutions, one already needs uniqueness).

4. Numerical methods

The approximation results of Section 2 naturally suggest that they could be used to develop numerical methods to approximate the solutions of the corresponding equations. This requires numerical solvers for one of the two (strictly related) variational problems

$$\min\left\{F(\varrho) + \frac{1}{2}W_2^2(\varrho,\nu) \ : \ \varrho \in \mathcal{P}(\Omega)\right\} \tag{4.1}$$

$$\text{or} \quad \min\left\{\frac{1}{2}W_2^2(\varrho,\nu) \ : \ \varrho \in K\right\}. \tag{4.2}$$

Of course, Problem (4.2) is a particular (but singular) case of the first one (choosing $F = I_K$, this notation standing for the indicator function in convex analysis, which vanishes on K and takes value $+\infty$ outside K) and corresponds to the projection step, while Problem (4.1) appears in the JKO scheme in the gradient case. Note that we have

$$I_K(\varrho) = \int I_{[0,1]}(\varrho(x))\mathrm{d}x.$$

Using suitable smoth functions $f : \mathbb{R}_+ \to [0, +\infty]$ which approximate $I_{[0,1]}$ it is also possible to approximate the second problem via standard (non-singular) issues of the first one.

We present in this section first some numerical methods which have been recently proposed to handle gradient flows in W_2 via their variational JKO scheme, and then a numerical method which is ad-hoc for Problem (4.2).

4.1. Numerical methods for the JKO scheme

We will present three methods. One, essentially taken from [9], is based on the Benamou-Brenier formula first introduced in [6] as a numerical tool for optimal transport (also see [7]). This method is well-suited for the case where the energy $F(\varrho)$ is a convex function of ϱ and can be easily dualized (i.e. F^* is explicit and easy to handle). The second method (essentially developed by M. Cuturi and collaborators, see [8,17,18]) comes from a clever approximation of the linear programming formulation of optimal transport in the discrete case (while the Benamou-Brenier formula best fits the continuous setting), and is easy to perform whenever F is separable (i.e. of the form $\int f(x, \varrho(x))\mathrm{d}x$). The third method (essentially adapted from the techniques used by Q. Mérigot for semidiscrete optimal transport see [32,38]) translates the problem into an optimization problem in the class of convex functions; it is well suited for the case where F is geodesically convex, which means that $\int f(\varrho(x))\,\mathrm{d}x$ is ony admissible if f satisfies McCann's condition (see Section 3).

Luckily, the three above methods are all able to handle the case $F(\varrho) = \int f(\varrho(x)\mathrm{d}x + \int V(x)\varrho(x)\mathrm{d}x$, with possible assumptions on f and V, but these assumptions are satisfied for $f = I_{[0,1]}$ and for many of its approximations. They can be used for Fokker-Planck and porous medium equations, or to approximate the gradient-flow version of Equation (1.3) (or to perform a projection step in the non-gradient case).

Augmented Lagrangian methods. Let us recall the basis of the Benamou-Brenier method. This method is based on the equality

$$\frac{1}{2}W_2^2(\mu,\nu) = \min\left\{\int_0^1\int_\Omega \frac{1}{2}\varrho_t|v_t|^2\mathrm{d}x\mathrm{d}t \ : \ \partial_t\varrho_t + \nabla\cdot(\varrho_t v_t) = 0, \varrho_0 = \mu, \varrho_1 = \nu\right\}.$$

Using $E = \varrho v$ as a variable instead of v, we can see (and this was the main idea in [6] to produce a numerical method) that this problem is convex, since we have

$$\int_0^1\int_\Omega \frac{1}{2}\varrho_t|v_t|^2\mathrm{d}x\mathrm{d}t = \sup\left\{\int\int a\mathrm{d}\varrho + \int\int b\cdot\mathrm{d}E \ : \ (a,b) : [0,1]\times\Omega \to K_2\right\}$$

where $K_2 = \{(a,b)\in\mathbb{R}\times\mathbb{R}^d \ : \ a + \frac{1}{2}|b|^2 \leq 0\}$. The continuity equation constraint can also be written as a sup penalization, by adding

$$\sup_{\phi\in C^1([0,1]\times\Omega)} -\int\int \partial_t\phi\,\mathrm{d}\varrho - \int\int \nabla\phi\cdot\mathrm{d}E + \int\phi_1\mathrm{d}\nu - \int\phi_0\mathrm{d}\mu,$$

a sup which is 0 if the constraint is satisfied, $+\infty$ if not. We then express everything in the space-time formalism, writing $\nabla_{t,x}\phi$ for $(\partial_t\phi,\nabla\phi)$, $\mathfrak{m}$ for (ϱ,E) and A for (a,b). We also set $G(\phi) := \int\phi_1\mathrm{d}\nu - \int\phi_0\mathrm{d}\mu$, thus getting

$$\min_{\mathfrak{m}}\sup_{A,\phi} \quad \mathfrak{m}\cdot(A - \nabla_{t,x}\phi) - I_{K_2}(A) + G(\phi),$$

where the scalar product is here the L^2 scalar product, but becomes a standard Euclidean scalar product as soon as one discretizes (in time-space).

The problem can now be seen as the search for a saddle-point of the Lagrangian

$$L(\mathfrak{m},(A,\phi)) := \mathfrak{m}\cdot(A - \nabla_{t,x}\phi) - I_{K_2}(A) + G(\phi).$$

We look for a pair $(\mathfrak{m},(A,\phi))$ (A and ϕ play together the role of the second variable) where $\mathfrak{m}$ minimizes for fixed (A,ϕ) and (A,ϕ) maximizes for fixed $\mathfrak{m}$. The idea of the *Augmented Lagrangian method* (see [25]) is that saddle points are the same if we subtract in the min/max expression the quantity $\frac{r}{2}||A - \nabla_{t,x}\phi||^2$, which vanishes, together with its first-order expansion, at optimal solutions. We then obtain a new saddle point problem

$$\min_{\mathfrak{m}}\max_{A,\phi} \qquad \mathfrak{m}\cdot(A - \nabla_{t,x}\phi) - I_{K_2}(A) + G(\phi) - \frac{r}{2}||A - \nabla_{t,x}\phi||^2$$

(again, the last term is an L^2 norm in time and space), which is then solved by iteratively repeating three steps: for fixed A and $\mathfrak{m}$, finding the optimal ϕ (which amounts to minimizing a quadratic functional in calculus of variations, i.e. solving a Poisson equation in space-time with prescribed Neumann boundary conditions, coming from the term G); then for fixed ϕ and $\mathfrak{m}$ find the optimal A (which amounts to a pointwise minimization problem, in this case a projection on the convex set K_2); finally update $\mathfrak{m}$ by going in the direction of the gradient descent, i.e. replacing $\mathfrak{m}$ with $\mathfrak{m} - r(A - \nabla_{t,x}\phi)$ (it is convenient to choose the parameter of the gradient descent to be equal to that of the Augmented Lagrangian).

This is what is done in the case where the initial and final measures are fixed. In a JKO scheme, at every step, one is fixed (say, μ), but the other is not, and a penalization on the final ϱ_1 is added, of the form $\tau F(\varrho_1)$.

Inspired from the considerations above, the saddle point below allows to treat the problem

$$\min_{\varrho_1} \quad \frac{1}{2}W_2^2(\varrho_1,\mu)+\int f(\varrho_1(x))\mathrm{d}x+\int V\mathrm{d}\varrho_1$$

by formulating it as

$$\min_{\mathfrak{m},\varrho_1}\max_{A,\phi,\lambda} \quad \int\int \mathfrak{m}\cdot(A-\nabla_{t,x}\phi)+\int \varrho_1\cdot(\phi_1+\lambda+V)-\int\int I_{K_2}(A)-\int\phi_0\mathrm{d}\mu-\int f^*(\lambda(x))\mathrm{d}x$$
$$-\frac{r}{2}\int\int|A-\nabla_{t,x}\phi|^2-\frac{r}{2}\int|\phi_1+\lambda+V|^2.$$

The role of the variable λ is to be dual to ϱ_1, which allows to express $f(\varrho_1)$ as $\sup_\lambda\ \varrho_1\lambda-f^*(\lambda)$. To find a solution to this saddle-point problem, an iterative procedure is also used, as above.

For the applications to gradient flows, a small time-step $\tau>0$ has to be fixed, and this scheme has to be done for each n, using $\mu=\varrho_n^\tau$ and setting ϱ_{n+1}^τ equal to the optimizer ϱ_1 and the functions f and V must include the scale factor τ. The time-space $[0,1]\times\Omega$ has to be discretized but the evolution in time is infinitesimal (due to the small time scale τ), which allows to discretize the fictitious time-interval $[0,1]$ using less than 10 time steps for each n...

Finally, in order to apply to density-constrained evolutions, one has to take $f=I_{[0,1]}$, or approximate it (via $f(s)=s^m/(m-1)$ for $m\to\infty$, for example). The interested reader can consult [9] for details, examples and simulations.

Entropic regularization. Another, very recent (see [8, 17]), but easy to explain and efficient, approach to optimal transport is based on its linear programming formulation in the discrete case (when the measures are finitely atomic, or have been approximated by atomic measures). As before, we first present the procedure in the case of two fixed measures. We suppose $\mu=\sum_i a_i\delta_{x_i}$ and $\nu=\sum_j b_j\delta_{y_j}$, and we set $c_{ij}:=c(x_i,y_j)$, where c is the transport cost we use (in the applications to W_2 gradient flows, $c(x,y)=|x-y|^2$). We consider the following approximation of Kantorovich's linear programming problem: fix $\varepsilon>0$ and look at

$$\min\left\{\sum_{i,j}\left(c_{ij}\gamma_{ij}+\varepsilon\gamma_{ij}\log(\gamma_{ij})\right)\ :\ \gamma_{ij}\geq 0,\ \sum_i\gamma_{ij}=b_j,\ \sum_j\gamma_{ij}=a_i\right\}. \tag{4.3}$$

It is clear that, for $\varepsilon\to 0$, the above minimization problem converges to the standard transport problem for the cost c. The objective function in (4.3) can be re-written using

$$c_{ij}\gamma_{ij}+\varepsilon\gamma_{ij}\log(\gamma_{ij})=\varepsilon\gamma_{ij}\log\left(\frac{\gamma_{ij}}{\eta_{ij}}\right)\quad\sum_{i,j}\left(c_{ij}\gamma_{ij}+\varepsilon\gamma_{ij}\log(\gamma_{ij})\right)=\varepsilon\mathrm{KL}(\gamma|\eta),$$

where $\eta_{ij}=e^{-c_{ij}/\varepsilon}$ and KL denotes the so-called *Kullback-Leibler* divergence (a relative entropy):

$$\mathrm{KL}(\gamma|\eta):=\sum_{i,j}h\left(\frac{\gamma_{ij}}{\eta_{ij}}\right)\eta_{ij}\quad\text{for } h(t)=\begin{cases}t\log t & \text{if } t\geq 0,\\ +\infty & \text{if } t<0.\end{cases}$$

Hence, the above minimization problem reads as the projection, for the Kullback-Leibler divergence, of the point $\eta\in\mathbb{R}^{I\times J}$ (where I and J stand for the cardiality of the sets of indices i and j, respectively) on the set of constraints $C=C^x\cap C^y$, where $C^x=\{\gamma\in\mathbb{R}^{I\times J}\ :\ \sum_j\gamma_{ij}=a_i\}$ and $C^y=\{\gamma\in\mathbb{R}^{I\times J}\ :\ \sum_i\gamma_{ij}=b_j\}$. The positivity constraint on γ is now included in the definition of the entropy function h.

The Kullback-Leibler divergence is not a distance, but shares some of the properties of distances. In particular it can be proven that the projection onto the intersection of two linear subspaces such as C^x and C^y can be

obtained by alternate projections, i.e. defining

$$\gamma^{2k+1} = \operatorname{argmin}\left\{\sum_{i,j} h\left(\frac{\gamma_{ij}}{\gamma_{ij}^{2k}}\right)\gamma_{ij}^{2k} \;:\; \sum_i \gamma_{ij} = b_j \text{ for all } j\right\}, \tag{4.4}$$

$$\gamma^{2k+2} = \operatorname{argmin}\left\{\sum_{i,j} h\left(\frac{\gamma_{ij}}{\gamma_{ij}^{2k+1}}\right)\gamma_{ij}^{2k+1} \;:\; \sum_j \gamma_{ij} = a_i \text{ for all } i\right\}, \tag{4.5}$$

with $\gamma^0 = \eta$, and looking at the limit as $k \to \infty$. Since the above projections on the two subspaces individually can be computed explicitly, this provides a very efficient method to approximate optimal transport plans. It is based on a sequence of explicit computations, easily parallelizable, and not too costly in terms of memory storage, since one can prove that at every step the current γ^{2k} is of the form $\gamma_{ij}^{2k} = \eta_{ij} p_i q_j$, which allows to stock $I + J$ data instead of $I \times J$.

When, instead of fixing μ and ν, the measure $\mu = \varrho_k$ is fixed and the other is penalized via a functional F, as it happens for a JKO problem, a slight variant of the above scheme allows to find the solution. Indeed, in order to solve (4.1) for $F(\varrho) = \int f(\varrho(x))\mathrm{d}x + \int V(x)\varrho(x)\mathrm{d}x$ it is enough to include the function V inside the function c, and then to replace (4.4) with

$$\gamma^{2k+1} = \operatorname{argmin}\left\{\sum_{i,j} h\left(\frac{\gamma_{ij}}{\gamma_{ij}^{2k}}\right)\gamma_{ij}^{2k} + \sum_j f\left(\sum_i \gamma_{ij}\right)\right\}$$

(see [18]). According to the function f, this problem can be solved explicitely or not, but in any case it allows for parallelization (the dependence in j is separable) and the solution is again of the form $\gamma_{ij}^{2k} = \eta_{ij} p_i q_j$. In particular, for $f = I_{[0,1]}$, which is of interest for the density-constrained problems, the solution of the above problem is given by

$$\gamma_{ij}^{2k+1} = \frac{\gamma_{ij}^{2k}}{\max\{1, \sum_i \gamma_{ij}^{2k}\}}.$$

Optimization among convex functions and computational geometry. It is clear that the optimization problem

$$\min_{\varrho} \quad \frac{1}{2} W_2^2(\varrho, \mu) + F(\varrho)$$

can be formulated in terms of transport maps and, taking advantage of Brenier's theorem, we can recast it as

$$\min_{u \text{ convex} :\ \nabla u \in \Omega} \quad \frac{1}{2}\int_\Omega |\nabla u(x) - x|^2 \mathrm{d}\mu(x) + F((\nabla u)_\# \mu). \tag{4.6}$$

Hence, we are facing a calculus of variations problem in the class of convex functions. These are non-trivial problems, as the literature (see [15, 39]) has shown, even in the case of functionals only involving u and ∇u. Here instead, if we suppose that F is of the form $F(\varrho) := \int f(\varrho(x))\mathrm{d}x + \int V \mathrm{d}\varrho$, the first part of this expression involves the Monge-Ampère operator $\det(D^2 u)$, which appears in the computation of the density ϱ of the image measure $(\nabla u)_\# \mu$.

We will roughly present here the method proposed in [10]. The main idea is to suppose that μ is a discrete measure of atomic type, i.e. of the form $\sum_j a_j \delta_{x_j}$. We look for a function u defined on its support $S := \{x_j\}_j$. We define at each point $x \in S$ the subdifferential

$$\partial u(x) := \{p \in \mathbb{R}^d \,:\, u(x) + p \cdot (y - x) \le u(y) \text{ for all } y \in S\}.$$

We say that u is convex if $\partial u(x) \neq \emptyset$ for all $x \in S$.

There are two difficulties in understanding the minimization problem (4.6) when μ is discrete: first, in order to define $(\nabla u)_{\#}\mu$ we need to define a gradient $\nabla u(x)$ at each point of $x \in S$, and if we manage to do it, $(\nabla u)_{\#}\mu$ will be an atomic measure; second, one part of our functional, namely $\int f(\varrho(x))\mathrm{d}x$, is never finite on atomic measures. Yet, we need to define the minimizer as an atomic measure, since we need to iterate the variational scheme, and use the minimizer as a starting point for the next JKO step.

First, let us define how to handle the part of the functional which requires absolutely continuous measures. We need to define a reasonable notion of $\varrho := (\nabla u)_{\#}\mu$ when μ is atomic, in such a way that we can compute functionals of the form $\int f(\varrho(x))\mathrm{d}x$. To do this, we take all the mass a_j contained in the point $x_j \in S$ and we spread it uniformly on the whole subdifferential $\partial u(x_j)$. This means that we define a measure that we will denote by $\varrho := (\partial u)_{\#}\mu$, given by

$$(\partial u)_{\#}\mu := \sum_j \frac{a_j}{|A_j|}\mathcal{L}^d \llcorner A_j,$$

where $A_j := \partial u(x_j) \cap \Omega$ (the intersection with Ω is done in order to take care of the constraint $\nabla u \in \Omega$).

Computing $\int f(\varrho(x))\mathrm{d}x$ hence gives

$$\int f(\varrho(x))\mathrm{d}x = \sum_j |A_j| f\left(\frac{a_j}{|A_j|}\right). \tag{4.7}$$

It is possible to prove that, thanks to the Brunn-Minkowski inequality that the above expression is convex in u as soon as f satisfies McCann's condition (see Section 3).

In order to associate with every u a discrete version of $(\nabla u)_{\#}\mu$, we need to introduce an extra variable, due to the fact that the gradient is not uniquely defined but we only have subgradients. The choice which is described in [10] is the following: define the optimization problem over the set of pairs $(u, P) : S \to \mathbb{R} \times \mathbb{R}^d$ where $P(x) \in \partial u(x)$ for every $x \in S$. Use u to define $(\partial u)_{\#}\mu$ (the "diffuse representative" of $(\nabla u)_{\#}\mu$), and $P_{\#}\mu$ as the "atomic representative" of $(\nabla u)_{\#}\mu$. The functional we minimize will then be

$$\tilde{F}(u, P) := \sum_j \left(\frac{1}{2}|P(x_j) - x_j|^2 + V(P(x_j))\right) a_j + |A_j| f\left(\frac{a_j}{|A_j|}\right).$$

Given $\varrho_n^\tau = \mu$, one finds the optimal (u, P) and then sets $\varrho_{n+1}^\tau := P_{\#}\mu$.

The above minimization problem can be seen to be convex in (u, P) whenever $x \mapsto |x|^2/2 + V(x)$ is convex and f satisfies McCann's condition; the constraints $P(x_j) \in A_j := \partial u(x_j) \cap \Omega$ is convex in (u, P) if Ω is convex. Note that incorporating the scale factor τ in the functional F means that convexity in P is guaranteed as soon as V is semi-convex and τ is small (the quadratic term coming will always overwhelm possible concavity of the V term). The delicate point is how to compute the subdifferentials $\partial u(x_j)$, and optimize them (i.e. compute derivatives of the relevant quantity w.r.t. u).

This is now possible, and in a very efficient way, thanks to tools from computational geometry. Indeed, in this context, subdifferentials are exactly a particular case of what are called Laguerre cells. Laguerre cells, similar to Voronoi cells, are very useful in optimal transport: given a set of values ψ_j, they are for the cells

$$W_j := \left\{x \in \Omega \,:\, \frac{1}{2}|x - x_j|^2 + \psi_j \leq \frac{1}{2}|x - x_{j'}|^2 + \psi_{j'} \text{ for all } j'\right\}.$$

This means that we look at points which are closer to x_j than to the other points $x_{j'}$, up to a correction given by the values ψ_j. It is not difficult to see that also in this case cells are convex polyhedra. Also, it can be easily seen that the Laguerre cells corresponding to $\psi_j := u(x_j) - \frac{1}{2}|x_j|^2$ are nothing but the subdifferentials of u (possibly intersected with Ω).

Handling Laguerre cells from the computer point of view has for long been difficult, but it is now state-of-the-art in computational geometry, and it is possible to compute very easily their volumes, as well as the derivatives

of their volumes (which depend on the measures of each faces) w.r.t. the values ψ_j. The reader can have a look at [32, 38] but also to Section 6.4.2 in [46].

We conlude the description of this method by noting that, compared to the previous ones, this is essentially a Lagrangian approach to the JKO scheme, as we follow the motion of each particle, while in the other two we looked at the evolution of the density on a fixed grid, or at the evolution of the masses at fixed atom points.

4.2. **A stochastic method for the projection on the density constraint**

We move now to a different numerical method, the first used for the continuous models of crowd motion under density constraints which are at the core of this survey. This method is only concerned with the projection problem (4.2), which means that, even in the gradient flow case, it has to be preceded by a transport step, where $\tilde{\varrho}^\tau_{n+1}$ is computed via more standard tools (finite volumes,...). This step has to be followed by a less standard one, i.e. the projection onto K.

The scheme proposed in [33, 34] to approximate the projection onto K with respect to the Wasserstein distance is based on the following - extremely heuristic - considerations. Consider a density $\nu \notin K$ such that the violation of the constraint $\nu \leq 1$ occurrs on a domain $\omega \subset \Omega$, but is small. This means that there exists a non-zero density $g \geq 0$ supported on $\overline{\omega}$ with $\nu = \mathbb{1}_\omega + \varepsilon g$ on ω. The projection of ν onto K will necessarily be identically 1 in ω. Let us denote by εv the displacement field which corresponds to the optimal transport between ν and $P_K[\nu]$. For ε small, the displacement v will be small far from ω. From $(id + \varepsilon v)_\# (\nu) = P_K[\nu]$, looking at what happens on ω and using that the projection is identically 1 in ω, one has

$$\frac{1 + \varepsilon g}{\det(I + \varepsilon Dv)} = 1,$$

so that, at the first order in ε,

$$\nabla \cdot \varepsilon v = \varepsilon g.$$

As $id + \varepsilon v$ is an optimal map, v is a gradient : $v = -\nabla p$, where p approximately verifies the Poisson problem

$$-\Delta p = g$$

with Dirichlet boundary conditions on $\partial\omega$.

Since we are considering a violation of the density constraint on a set ω and supposing that, after the projection, the density will be saturated on ω itself, we are only interested in finding the mass that will exit ω when the displacement is given by εv. This means, at a first order approximation, we only need to estimate the flux of density through the boundary $\partial\omega$, i.e. $v \cdot \mathbf{n} = -\partial p/\partial \mathbf{n}$. Now consider the stochastic interpretation of the Poisson equation $-\Delta p = g$ (see, for instance, classical texts such as [24]).

Suppose that g is a probability density on ω, and consider a random variable $X \in \omega$ distributed according to g. Then, consider a Brownian motion stemming from X, i.e. $t \mapsto X + B_t$, where B is a Brownian motion and X and B are independent. Set $T := \inf\{t : X + B_t \notin \omega\}$ and $Y = X + B_T$. This is a randome variable standing for the first exit point of $X + B_t$ from ω. The random variable Y takes values in $\partial\omega$ and is known to follow the probability law which has a density $-\partial p/\partial \mathbf{n}$ with respect to the $\mathcal{H}^{d-1}$-dimensional measure on $\partial\omega$.

The idea is therefore to redistribute the exceeding mass of the saturated zone in the following way (see figure 4.2): discretize the domain on a regular grid and, for each saturated cell, a random walk is started, transporting the exceding mass $(\varrho - 1)_+$. When this random walk encounters a non-saturated cell, it gets rid of as much mass as it can, and continues as long as the transported mass is not fully distributed. When all the saturated cells have been treated, the obtained density ϱ_{n+1} is admissible. As it should represent a good approximation of what obtained after stopping $X + B_t$ when meeting the non-saturated zone (up to the fact that we use a random wall on a discrete grid instead of a Brownian motion), the saturated zone should have moved with a normal speed close to $-\partial p/\partial \mathbf{n}$, which was the desired motion.

It is important to insist that no convergence result has been proven on this stochastic algorithm, but the results obtained in [34] are convincing. Also, the same method has been applied in [21] to the case of moving

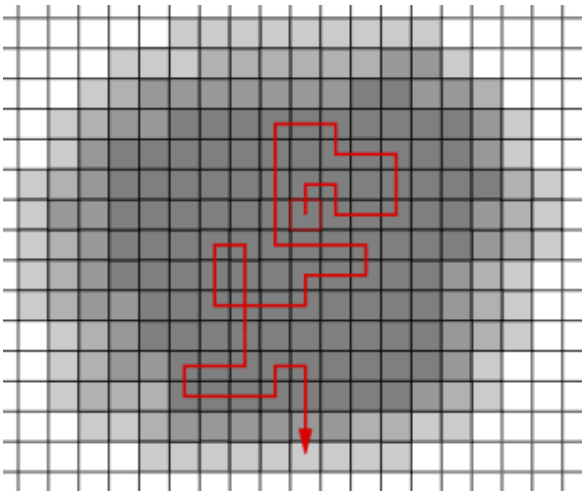

FIGURE 3. A random walk realization of the projection onto K

domains. This is not a variational algorithm, but its pertinence is justified by the heuristics that we have described so far: for small violations of the constraint $\varrho \leq 1$ the projection (a transport optimization problem) can be well approximated by the linearization of the Monge-Ampère equation (this is also what we see in the expression of the equation (1.3), and can also be interpreted in terms of the approximation of the W_2 distance via the H^{-1} distance, see Section 5.5.2 in [46]); then, the stochastic interpretation of such an equation provides an efficient way to mimick the motion of the mass in the projection.

REFERENCES

[1] D. ALEXANDER, I. KIM, Y. YAO, Quasi-static evolution and congested crowd motion, *Nonlinearity* 27 (4), 823–858, 2014.

[2] L. AMBROSIO, Movimenti minimizzanti, *Rend. Accad. Naz. Sci. XL Mem. Mat. Sci. Fis. Natur.* 113, 191–246, 1995.

[3] L. AMBROSIO, Transport equation and Cauchy problem for BV vector fields, *Inv. Math.* 158 (2004), 227-260.

[4] L. AMBROSIO, N. GIGLI AND G. SAVARÉ, *Gradient flows in metric spaces and in the spaces of probability measures.* Lectures in Mathematics, ETH Zurich, Birkhäuser, 2005.

[5] J.-P. AUBIN, Un théorème de compacité. (French). *C. R. Acad. Sci. Paris* 256. pp. 5042–5044

[6] J.-D. BENAMOU AND Y. BRENIER, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numer. Math.*, 84 (2000), 375–393.

[7] J.-D. BENAMOU AND G. CARLIER, Augmented Lagrangian Methods for Transport Optimization, Mean Field Games and Degenerate Elliptic Equations, *J. Opt. Theor. Appl.*,167 (1), pp 1–26, 2015.

[8] J.-D. BENAMOU, G. CARLIER, M. CUTURI, L. NENNA, G. PEYRÉ Iterative Bregman projections for regularized transportation problems *SIAM Journal on Scientific Computing* 37 (2), 1111-1138, 2015

[9] J.-D. BENAMOU, G. CARLIER AND M. LABORDE An augmented Lagrangian approach to Wasserstein gradient flows and applications, *ESAIM: Proc.* 54, 1–17, 2016.

[10] J.-D. BENAMOU, G. CARLIER, Q. MÉRIGOT AND É. OUDET Discretization of functionals involving the Monge-Ampère operator, *Numerische Mathematik*, 134(3), 611–636, 2016.

[11] PH. BÉNILAN, L. BOCCARDO, AND M.A. HERRERO, On the limit of solutions of $u_t = \Delta u^m$ as $m \to \infty$, in *Some topics in Nonlinear PDEs*, Torino ,1989.

[12] Y. BRENIER, Décomposition polaire et réarrangement monotone des champs de vecteurs. (French) *C. R. Acad. Sci. Paris Sér. I Math.* 305 (19), 805–808, 1987.

[13] Y. BRENIER, Polar factorization and monotone rearrangement of vector-valued functions, *Communications on Pure and Applied Mathematics* 44, 375–417, 1991.

[14] J.-A. CARRILLO AND F. SANTAMBROGIO L^∞ estimates for the JKO scheme in parabolic-elliptic Keller-Segel systems, *Quarterly of Applied Mathematics*, to appear.

[15] P. CHONÉ AND H. LE MEUR, Non-convergence result for conformal approximation of variational problems subject to a convexity constraint, *Numer. Funct. Anal. Optim.* 5-6 (22), 529–547, 2001.

[16] M.G. Crandall, Nonlinear Semigroups and Evolution Governed by Accretive Operators. In Proceeding of Symposium in Pure Mat., Part I (F. Browder, ed.) A.M.S., Providence 1986, 305–338.
[17] M. Cuturi Sinkhorn Distances: Lightspeed Computation of Optimal Transport, *Advances in Neural Information Processing Systems,* 2292-2300, 2013
[18] M. Cuturi, G. Peyré Computational Optimal Transport, preprint
[19] E. De Giorgi, New problems on minimizing movements, *Boundary Value Problems for PDE and Applications, C. Baiocchi and J. L. Lions eds.* (Masson, 1993) 81–98.
[20] G. De Philippis, A. R. Mészáros, F. Santambrogio, B. Velichkov, BV estimates in optimal transportation and applications, *Arch. Rat. Mech. An.* 219 (2), 829-860, 2016.
[21] S. Di Marino, B. Maury and F. Santambrogio Measure sweeping processes, *J. Conv. An.*, 23 (1), 567–601, 2016.
[22] S. Di Marino and A. R. Mészáros Uniqueness issues for evolution equations with density constraints, *Math. Models Methods Appl. Sci.*, 26 (09), 1761–1783(2016)
[23] R. J. DiPerna, P.-L. Lions, Ordinary differential equations, transport theory and Sobolev spaces, *Invent. Math.* 98 (1989), 511-547.
[24] J. L. Doob, *Classical Potential Theory and Its Probabilistic Counterpart*, Springer, 1984 (2001).
[25] M. Fortin and R. Glowinski, *Augmented Lagrangian methods, Applications to the Numerical Solution of Boundary-Value Problems*, North-Holland (1983).
[26] D. Helbing, *A fluid dynamic model for the movement of pedestrians,* Compl. Syst. 6, 391-415, 1992.
[27] R. L. Hughes, *A continuum theory for the flow of pedestrian,* Transp. Res. B, 36 (2002), 507-535.
[28] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation, *SIAM J. Math. Anal.* 29 (1), 1–17, 1998.
[29] L. Kantorovich, On the transfer of masses. *Dokl. Acad. Nauk. USSR*, 37, 7–8, 1942.
[30] I. Kim, Regularity of the one-phase Hele-Shaw problem, *J. Diff. Equations*, 223, 161–184, 2006.
[31] I. Kim Regularity and speed of the Hele-Shaw flow *Viscosity Solution Theory of Differential Equations and its Developments*, RIMS Kokyuroku (Proceedings), 1428, 19–27, 2005.
[32] J. Kitagawa, Q. Mérigot and B. Thibert, Convergence of a Newton algorithm for semi-discrete optimal transport, preprint, 2016.
[33] B. Maury, A. Roudneff-Chupin and F. Santambrogio A macroscopic crowd motion model of gradient flow type, *Math. Models and Methods in Appl. Sciences* 20 (10), 1787–1821, 2010.
[34] B. Maury, A. Roudneff-Chupin, F. Santambrogio and J. Venel, Handling congestion in crowd motion modeling, *Net. Het. Media* 6 (3), 485–519, 2011.
[35] B. Maury and J. Venel, A mathematical framework for a crowd motion model, *C. R. Acad. Sci. Paris, Ser. I* 346 (2008), 1245–1250.
[36] B. Maury and J. Venel, A discrete contact model for crowd motion *ESAIM: Math. Mod. Num. An.* 45 (1), 145–168, 2011.
[37] R. J. McCann A convexity principle for interacting gases. *Adv. Math.* 128 (1), 153–159, 1997.
[38] Q. Mérigot A multiscale approach to optimal transport. *Computer Graphics Forum* 30 (5) 1583–1592, 2011.
[39] Q. Mérigot and É. Oudet Handling convexity-like constraints in variational problems *SIAM Journal on Numerical Analysis,* 52 (5), 2466–2487, 2014
[40] A. R. Mészáros and F. Santambrogio Advection-diffusion equations with density constraints, *Analysis & PDE* 9-3 (2016), 615–644.
[41] G. Monge, Mémoire sur la théorie des déblais et des remblais, *Histoire de l'Académie Royale des Sciences de Paris*, 666–704, 1781.
[42] J.-J. Moreau, Evolution problem associated with a moving convex set in a Hilbert space, *J. Differential Equations* 26, 346–374, 1977.
[43] F. Otto, The geometry of dissipative evolution equations: The porous medium equation, *Comm. Partial Differential Equations*, 26, 101–174, 2011.
[44] B. Perthame, F. Quiros and J.-L. Vazquez The Hele-Shaw asymptotics for mechanical models of tumor growth *Arch. Rati. Mech. An.*, 212 (1), 93–127, 2014.
[45] A. Roudneff-Chupin, *Modélisation macroscopique de mouvements de foule* (French), PhD Thesis, Université Paris-Sud, 2011.
[46] F. Santambrogio *Optimal Transport for Applied Mathematicians*, Progress in Nonlinear Differential Equations and Their Applications no 87, Birkhäuser Basel (2015).
[47] F. Santambrogio {Euclidean, Metric, and Wasserstein} Gradient Flows: an overview, *Bull. Math. Sci.*, 7 (1), 87–154, 2017.
[48] J. L. Vázquez, *The porous medium equation. Mathematical theory*, The Clarendon Press, Oxford University Press, (2007).
[49] C. Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics, AMS, 2003.