Optimal transport methods for parabolic diffusion equations: the JKO scheme

Filippo Santambrogio

Abstract These lecture notes present the main results presented by the author in the course he gave in Chania in July 2024, about the use of the Jordan-Kinderlehrer-Otto scheme in the approximation of the solutions of various linear or non-linear parabolic evolution equations with a gradient structure. The convergence of the scheme to a solution is proven, and estimates on the steps of the scheme are presented. Among the applications, BV and Sobolev bounds are detailed, a strategy to obtain functional inequalities is introduced, and a stronger convergence of the scheme to the solution of the PDE is proven.

1 Introduction

Several evolution equations have a gradient structure, meaning that the evolution goes in the direction of the steepest descent of a given global quantity, and exploiting their variational structure provides deep insight into their behavior, allows for general proofs of existence, gives natural procedures to approximate their solutions, and suggests efficient points of view in order to analyze them. Speaking of evolutions of probability densities, the use of optimal transport and the study of those PDEs which have a variational structure w.r.t. the Wasserstein distance W_2 has been popularized in the last decades after the seminal paper by Jordan, Kinderlehrer and Otto, [37], who introduced a discrete-in-time variational scheme to approximate the Fokker-Planck equation. These lecture notes aim to describe the scheme they introduced, nowadays called JKO scheme, to generalize it to the case where other transport costs instead of the quadratic one are used (thus obtaining different equations at the limit), and to use more recent techniques to study the discrete sequence produced by the scheme.

Section 2 will be devoted to a general introduction to the topic of gradient flows, starting from the Euclidean case and then passing to the abstract metric theory. The

Universite Claude Bernard Lyon 1, ICJ UMR5208, CNRS, Ecole Centrale de Lyon, INSA Lyon, Université Jean Monnet, 69622 Villeurbanne, France. e-mail: santambrogio@math.univ-lyon1.fr

whole theory is extended to the case of *nonlinear gradient flows*, which correspond to the equation $x'(t) = -\nabla h^*(\nabla F(x(t)))$ instead of $x'(t) = -\nabla F(x(t))$.

Starting from Section 3 we want to focus on the part of the gradient flow theory most related to optimal transport, and we introduce the Monge-Kantorovich optimal transport probelm, the Wasserstein spaces, and then present in details the JKO scheme and the resulting PDEs, which include the Fokker-Planck and porous medium equations.

Section 4 is devoted to different techniques which can be used in order to prove the convergence of the JKO Scheme: the first requires to choose suitable interpolations of the sequence obtained in the JKO and study the relevant compactness properties in order to pass to the limit of terms, the second exploits an integral characterization of the PDE and is bases on lower semicontinuity arguments.

In Section 5 we introduce a new inequality involving the gradients of the transported densities and of their Kantorovich potentials, which is of special use when appled to the solution of one step of the JKO scheme. This inequality allows in this case to prove higher-order bounds on these solutions. BV bounds for the porous medium equation are proven, as well as bounds on generalized Fisher informations for some particular variants of the Fokker-Planck equations. Some of these estimates recover well-known bounds from the continuous PDE, some are more original in their general form.

Section, 6 provides applications to functional inequalities, focusing on the log-Sobolev inequalities. The idea is to perform a discrete counterpart of the well-known Bakry-Emery theory, using the JKO scheme instead of a continuous flow to differentiate the relevant quantities and compare them. For simplicity, only the most classical log-Sobolev inequality is considered.

Finally, Section 7 uses an improvement of the inequality presented in Section 5 in order to prove that some second-order integral quantities pass to the limit as the time step $\tau \to 0$. This allows to prove in some regime the strong $L^2_t H^2_x$ convergence of the solution of the JKO scheme for the Fokker-Planck equation, strongly improving previous results.

The content of these notes has been presented in July 2024 in the summer school Festum Pi in Chania (Crete) and, partially, in a mini-course in the Vito Volterra Meeting in June 2024 in Rome.

The whole presentation is not fully detailed, is sometimes sketchy, points to the relevant papers on which these notes are based (see in particular [49, 50, 26, 27, 20, 21, 52] for more details), and very often regularity issues which could be important for a fully rigorous analysis are left to the reader.

2 Gradient flows and variants

We present here the main ideas about the general theory of gradient flows, in the Euclidan space and then in metric spaces, also considering what we call *nonlinear gradient flows*.

2.1 The Euclidean case

We start here from the easiest framework where we can consider gradient flows, i.e. from what happens in the Euclidean space $\mathbb{R}^n$. Most of what we will say stays true in an arbitrary Hilbert space, but we will stick to the finite-dimensional case for simplicity.

Here, given a function $F : \mathbb{R}^n \to \mathbb{R}$, smooth enough, and a point $x_0 \in \mathbb{R}^n$, a gradient flow is just defined as a curve $x(t)$, with starting point at $t = 0$ given by x_0, which moves by choosing at each instant of time the direction which makes the function F decrease as much as possible. More precisely, we consider the solution of the *Cauchy Problem*

$$\begin{cases} x'(t) = -\nabla F(x(t)) & \text{for } t > 0, \\ x(0) = x_0. \end{cases} \tag{1}$$

This is a standard Cauchy problem which has a unique solution if ∇F is Lipschitz continuous, i.e. if $F \in C^{1,1}$. In some cases one could be interested in functions F which are not differentiable, and an easy example is that of convex functions. If F is convex, it could be non-differentiable, but we can replace the gradient with the subdifferential. More precisely, we can consider instead of (1), the following differential inclusion: we look for an absolutely continuous curve $x : [0,T] \to \mathbb{R}^n$ such that

$$\begin{cases} x'(t) \in -\partial F(x(t)) & \text{for a.e. } t > 0, \\ x(0) = x_0, \end{cases} \tag{2}$$

where $\partial F(x) = \{p \in \mathbb{R}^n \ : \ F(y) \geq F(x) + p \cdot (y - x) \ \text{ for all } y \in \mathbb{R}^n\}$. We refer to [47] for all the definitions and notions from convex analysis that one could need here and in the rest of the chapter.

A very interesting feature of these particular Cauchy problems which are gradient flows is their discretization in time. Actually, one can fix a small time step parameter $\tau > 0$ and look for a sequence of points $(x_k^\tau)_k$ defined through the iterated scheme, called *Minimizing Movement Scheme* (see [25, 2]),

$$x_{k+1}^\tau \in \operatorname{argmin}_x F(x) + \frac{|x - x_k^\tau|^2}{2\tau}. \tag{3}$$

Very mild assumptions on F (l.s.c. and some lower bounds, for instance $F(x) \geq C_1 - C_2|x|^2$) are sufficient to guarantee that these problems admit a solution for small τ. When F is convex the operator which associaties with the previous point x_k^τ the new point ix_{k+1}^τ, which is unique as soon as $x \mapsto F(x) + \frac{|x-x_k^\tau|^2}{2\tau}$ is strictly convex (which is the case, for instance, if F is convex, or if D^2F is bounded from below and τ is small), is called in convex optimization *proximal operator*.

We can interpret this sequence of points as the values of the curve $x(t)$ at times $t = 0, \tau, 2\tau, \dots, k\tau, \dots$. It happens that the optimality conditions of the recursive minimization exactly give a connection between these minimization problems and the equation, since we have

$$x_{k+1}^\tau \in \operatorname{argmin} F(x) + \frac{|x - x_k^\tau|^2}{2\tau} \quad \Rightarrow \quad \nabla F(x_{k+1}^\tau) + \frac{x_{k+1}^\tau - x_k^\tau}{\tau} = 0,$$

i.e.

$$\frac{x_{k+1}^\tau - x_k^\tau}{\tau} = -\nabla F(x_{k+1}^\tau).$$

This expression is exactly the discrete-time *implicit Euler scheme* for $x' = -\nabla F(x)$! (note that in the convex non-smooth case this becomes $\frac{x_{k+1}^\tau - x_k^\tau}{\tau} \in -\partial F(x_{k+1}^\tau)$).

Before going on, one can observe that it could be natural to replace the quadratic penalization with other penalizations, for instance another power of the distance. More generally, if $h : \mathbb{R}^n \to \mathbb{R}$ is a convex function, one could consider

$$x_{k+1}^\tau \in \operatorname{argmin}_x F(x) + \tau h\left(\frac{x - x_k^\tau}{\tau}\right),$$

which reduces to (3) when $h(z) = \frac{1}{2}|z|^2$. In this case the optimality conditions would read as

$$\nabla F(x_{k+1}^\tau) + \nabla h\left(\frac{x_{k+1}^\tau - x_k^\tau}{\tau}\right) = 0, \quad \text{i.e. } \frac{x_{k+1}^\tau - x_k^\tau}{\tau} = \nabla h^*\left(-\nabla F(x_{k+1}^\tau)\right)$$

(where we used the Legendre transorm h^* defined via $h^*(y) = \sup_x x \cdot y - h(x)$, which is such that $\nabla h^* = (\nabla h)^{-1}$). In this case, we would obtain a discretization of the Cauchy problem

$$\begin{cases} x'(t) = \nabla h^*(-\nabla F(x(t))) & \text{for } t > 0, \\ x(0) = x_0. \end{cases} \tag{4}$$

For simplicity we will always assume that h and h^* are even (anc actually, they will most often be radially symmetric – or powers of the norm – so that we can move the minus sign outside ∇h^*). This evolution equation can be called *nonlinear gradient flow*. In the case where $h(z) = \frac{1}{p}|z|^p$ it reads $x'(t) = -\nabla F(x(t)))^{q-1}$, where q is the conjugate exponent of p (i.e. when $\frac{1}{p} + \frac{1}{q} = 1$) and when we write $\mathbf{v}^\alpha$ for a vector $\mathbf{v}$ and an exponent $\alpha > 0$ we mean $|\mathbf{v}|^{\alpha-1}\mathbf{v}$ (if $\mathbf{v} \neq 0$, otherwise we take the value 0).

In all these cases, under some regularity assumptions on F, it is possible to prove that, for $\tau \to 0$, the sequence we found, suitably interpolated, converges to the solution of the corresponding evolution equation (see [50]).

2.2 The metric case

The iterated minimization scheme that we introduced above has another interesting feature: it even suggests how to define solutions for functions F which are only l.s.c., with no gradient at all!

Even more, a huge advantage of this discretized formulation is also that it can easily be adapted to metric spaces. Actually, if one has a metric space (X, d) and a l.s.c. function $F : X \to \mathbb{R} \cup \{+\infty\}$ (under suitable compactness assumptions to guarantee existence of the minimum), one can define (see [25, 2])

$$x^\tau_{k+1} \in \operatorname{argmin}_x F(x) + \frac{d(x, x^\tau_k)^2}{2\tau} \tag{5}$$

and study the limit as $\tau \to 0$. More generally, one can consider, for $p > 1$,

$$x^\tau_{k+1} \in \operatorname{argmin}_x F(x) + \frac{d(x, x^\tau_k)^p}{p\tau^{p-1}}, \tag{6}$$

which corresponds to the case $h(z) = \frac{1}{p}|z|^p$ previously described. We will not deal with the case of general penalizations h, since in the metric case we are only allowed to consider functions of the distance (which rouhgly corresponds to radial functions h). For simplicity, we will stick to the case of powers of the distance, but a similar analysis could be performed using $\tau h(d(x, x^\tau_k)/\tau)$.

Then, we use the piecewise constant interpolation

$$x^\tau(t) := x^\tau_k \quad \text{for every } t \in ((k-1)\tau, k\tau] \tag{7}$$

and study the limit of x^τ as $\tau \to 0$.

A first important point is to prove that a limit actually exists, i.e. the compactness of the curves $(x^\tau)_\tau$. We start from the easy estimate

$$F(x^\tau_{k+1}) + \frac{d(x^\tau_{k+1}, x^\tau_k)^p}{p\tau^{p-1}} \le F(x^\tau_k), \tag{8}$$

and

$$\sum_{k=0}^{l} d(x^\tau_{k+1}, x^\tau_k)^p \le p\tau^{p-1} \left(F(x^\tau_0) - F(x^\tau_{l+1})\right) \le C\tau^{p-1}. \tag{9}$$

The Cauchy-Schwartz inequality gives, for $t < s$, $t \in [k\tau, (k+1)\tau[$ and $s \in [l\tau, (l+1)\tau[$ (hence $|l-k| \le \frac{|t-s|}{\tau} + 1$),

$$d(x^\tau(t), x^\tau(s)) \le \sum_{j=k}^{l} d(x^\tau_{j+1}, x^\tau_j) \le \left(\sum_{j=k}^{l} d(x^\tau_{k+1}, x^\tau_k)^p\right)^{1/p} (l+1-k)^{1/q}.$$

We then observe taht we have $l+1-k \leq \frac{|t-s|}{\tau}+1$ and this allows, together with (9), obtain

$$d(x^\tau(t), x^\tau(s)) \leq \left(\sum_{j=k}^{l} d(x^\tau_{k+1}, x^\tau_k)^p\right)^{1/p} \left(\frac{|t-s|}{\tau}+1\right)^{1/q} \leq C\left(|t-s|^{1/q}+\tau^{1/q}\right).$$

This provides a Hölder bound on the curves x^τ (with a negligible error of order $\tau^{1/q}$ which disappears at the limit $\tau \to 0$), and allows to extract a uniformly converging subsequence.

It is easier to see what happens if we add some structure to the metric space (X, d), in particular if we assume that (X, d) is a *geodesic space*. This requires a short discussion about curves and geodesics in metric spaces.

Curves and geodesics in metric spaces

We recall that a curve ω is a continuous function defined on a interval, say $[0,1]$, and valued in a metric space (X, d). As it is a map between metric spaces, it is meaningful to say whether it is Lipschitz or not, but its speed $\omega'(t)$ has no meaning, unless X is a vector space. Surprisingly, it is possible to give a meaning to the modulus of the velocity, $|\omega'|(t)$.

Definition 1 If $\omega : [0,1] \to X$ is a curve valued in the metric space (X, d) we define the metric derivative of ω at time t, denoted by $|\omega'|(t)$ through

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

provided this limit exists.

In the spirit of Rademacher Theorem, it is possible to prove (see [3] or [51]) that, if $\omega : [0,1] \to X$ is Lipschitz continuous, then the metric derivative $|\omega'|(t)$ exists for a.e. t. Moreover we have, for $t_0 < t_1$,

$$d(\omega(t_0), \omega(t_1)) \leq \int_{t_0}^{t_1} |\omega'|(s)\, \mathrm{d}s.$$

The same is also true for more general curves, not only Lipschitz continuous.

Definition 2 A curve $\omega : [0,1] \to X$ is said to be *absolutely continuous* whenever there exists $g \in L^1([0,1])$ such that $d(\omega(t_0), \omega(t_1)) \leq \int_{t_0}^{t_1} g(s)\, \mathrm{d}s$ for every $t_0 < t_1$. The set of absolutely continuous curves defined on $[0,1]$ and valued in X is denoted by $\mathrm{AC}(X)$.

It is well-known that every absolutely continuous curve can be reparametrized in time (through a monotone-increasing reparametrization) and become Lipschitz continuous. The existence of the metric derivative for a.e. t is also true for $\omega \in \mathrm{AC}(X)$, via this reparametrization.

Some particular curves deserve special attention.

Definition 3 A curve $\omega : [a,b] \to X$ is said to be a constant-speed geodesic between x_0 and $x_1 \in X$ if $\omega(a) = x_0$, $\omega(b) = x_1$ and $|\omega'|(t) = d(x_0, x_1)/(b-a)$ for a.e. $t \in [a,b]$. . A space (X,d) is said to be a *geodesic space* if for every $x_0, x_1 \in X$ there exists a constant-speed geodesic between x_0 and x_1.

We can now come back to the interpolation of the points obtained through the Minimizing Movement scheme (5) and note that, if (X,d) is a geodesic space, then it is possible to replace the piecewise constant interpolation with a piecewise geodesic interpolation. This means defining a curve $x^\tau : [0,T] \to X$ such that $x^\tau(k\tau) = x_k^\tau$ and such that x^τ restricted to any interval $[k\tau, (k+1)\tau]$ is a constant-speed geodesic with speed equal to $d(x_k^\tau, x_{k+1}^\tau)/\tau$. Then, the bounds we proved above read as

$$\int_0^T |x'|(t)^p \mathrm{d}t \le p(F(x_0) - F(x(T)) \le C.$$

This proves that the curves x^τ are bounded in $W^{1,p}([0,T];X)$ and implies equicontinuity since, exactly as in the Euclidean case, one has $W^{1,p} \subset C^{0,\alpha}$ with $\alpha = 1 - 1/p = 1/q$.

The next question is how to characterize the limit curve obtained when $\tau \to 0$, and in particular how to express the fact that it is a gradient flow of the function F. Of course, one cannot try to prove the equality $x' = -\nabla F(x)$, just because neither the left-hand side nor the right-hand side have a meaning in a metric space!

If the space X, the distance d, and the functional F are explicitly known, in some cases it is possible to pass to the limit the optimality conditions of each optimization problem in the discretized scheme, and characterize the limit curves (or the limit curve) $x(t)$. It will be possible to do so in the framework of probability measures, as it will be discussed in Sectios 3 and 4, but not in general. Indeed, without a little bit of (differential) structure on the space X, it is essentially impossible to do so. Hence, if we want to develop a general theory for gradient flows in metric spaces, finer tools are needed. In particular, we need to characterize the solutions of $x' = -\nabla F(x)$ (or $x' \in -\partial F(x)$) by only using metric quantities (in particular, avoiding derivatives, gradients, and more generally vectors). The book by Ambrosio-Gigli-Savaré [4], and in particular its first part (the second being devoted to the space of probability measures) exactly aims at doing so.

The idea is to present alternative characterizations of gradient flows in the smooth Euclidean case, which can be used as a definition of gradient flow in the metric case, since all the quantities which are involved have their metric counterpart.

The first observation is the following: thanks to the definition of Legendre transform h^*, the condition $x'(t) = \nabla h^*(-\nabla F(x(t))$ is equivalent to $h(x'(t)) + h^*(-\nabla F(x(t)) = -x'(t) \cdot \nabla F(x(t))$. Since the inequality $h(x) + h^*(y) \ge x \cdot y$ is always true, we can also say that it is equivalent to the inequality

$$h(x'(t)) + h^*(-\nabla F(x(t))) \leq -x'(t) \cdot \nabla F(x(t)) = -\frac{\mathrm{d}}{\mathrm{d}t} F(x(t)),$$

and this inequality for a.e. t is also equivalent to its integral version

$$\int_0^T h(x'(t))\mathrm{d}t + \int_0^T h^*(-\nabla F(x(t)))dt + F(x(T)) \leq F(x_0).$$

In the case where $h(z) = \frac{1}{p}|z|^p$ is a power this can be written as

$$\frac{1}{p}\int_0^T |x'(t)|^p dt + \frac{1}{q}\int_0^T |\nabla F(x(t)|^q dt + F(x(T)) \leq F(x_0).$$

This condition, called EDI (*Energy Dissipation Inequality*) characterizes gradient flows (or their nonlinear counterpart) in the smooth Euclidean case. When h and h^* are radial, it can also be taken as a definiiton in the metric case, as all its terms have a meaning in the metric setting. We already saw how to define $|x'|$, let us see how to define the norm of the gradient.

Slope and modulus of the gradient.

Many definitions of the modulus of the gradient of a function F defined over a metric space are possible. The easiest possible choice is the *local Lipschitz constant*

$$|\nabla F|(x) := \limsup_{y \to x} \frac{|F(x) - F(y)|}{d(x, y)}; \tag{10}$$

another is the *descending slope* (we will often say just *slope*), which is a notion more adapted to the minimization of a function than to its maximization, and hence reasonable for lower semi-continuous functions:

$$|\nabla^- F|(x) := \limsup_{y \to x} \frac{[F(x) - F(y)]_+}{d(x, y)}$$

(note that the slope vanishes at every local minimum point).

In the general theory of Gradient Flows in metric spaces ([4]), another characterization, different from the EDI, is proposed in order to cope with uniqueness and stability results. This is only done in the case $p = 2$ (i.e. for linear gradient flows) and it is based on the following observation: if $F : \mathbb{R}^d \to \mathbb{R}$ is convex, then the inequality

$$F(y) \geq F(x) + p \cdot (y - x) \quad \text{for all } y \in \mathbb{R}^d$$

characterizes (by definition) the vectors $p \in \partial F(x)$ and, if $F \in C^1$, it is only satisfied for $p = \nabla F(x)$. We can pick a curve $x(t)$ and a point y and compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}|x(t) - y|^2 = (y - x(t)) \cdot (-x'(t)).$$

Consequently, imposing

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}|x(t)-y|^2 \le F(y) - F(x(t))$$

(for all y) will be equivalent to $-x'(t) \in -\partial F(x(t))$. This will provide a second characterization (called EVI, *Evolution Variational Inequality*) of gradient flows in a metric environment. Again, all the terms appearing in the above inequality have a metric counterpart (only squared distances and derivatives w.r.t. time appear).

If in the smooth Euclidean case the two notions (EDI and EVI) finally coincide, it is now well-known that in genreal they are quite different. In particular, it is quite easy (and we will give some hints on how to do) to prove existence of a gradient flow in the EDI sense, but in general uniqueness fails. An example can be found, for instance, in [48]. On the other hand, the EVI notion allows quite easily to prove uniqueness, but the existence of an EVI gradient flow is harder to obtain and, by the way, only holds when F satisfies some convexity assumption (more precisely: it can be proven that whenever F is such that an EVI gradient flow exists for any starting point x_0, then F is *geodesically convex*, a notion on which we wil come back later, see [24]).

2.3 Existence of a gradient flow

We can easily understand that, even if the estimate (8) is enough to provide compactness, it will never be enough to characterize the limit curve (indeed, it is satisfied by any discrete evolution where x^τ_{k+1} gives a better value than x^τ_k, without any need for optimality). Hence, we will never obtain either of the two formulations - EDI or EVI - of metric gradient flows.

In order to improve the result, we should exploit how much x^τ_{k+1} is better than x^τ_k. An idea due to De Giorgi allows to obtain the desired result, via a "variational interpolation" between the points x^τ_k and x^τ_{k+1}. In order to do so, once we fix x^τ_k, for every $\theta \in]0,1]$, we consider the problem

$$m(\theta) := \min_x \quad F(x) + \frac{d(x, x^\tau_k)^p}{p(\theta\tau)^{p-1}}$$

and call $x(\theta)$ any minimizer for this problem, and $m(\theta)$ the minimal value. It is clear that, for $\theta \to 0^+$, we have $x(\theta) \to x^\tau_k$ and $m(\theta) \to F(x^\tau_k)$, and that, for $\theta = 1$, we get back to the original problem with minimizer x^τ_{k+1}. Moreover, the function m is non-increasing and hence a.e. differentiable (actually, we can even prove that it is locally semiconcave). Its derivative $m'(\theta)$ is given by the derivative of the function $\theta \mapsto F(x) + \frac{d^p(x, x^\tau_k)}{p(\theta\tau)^{p-1}}$, computed at the optimal point $x = x(\theta)$ (the existence of $m'(\theta)$ implies that this derivative is the same at every minimal point $x(\theta)$). Hence, a quick computation shows that we have

$$m'(\theta) = -\frac{d^p(x(\theta), x_k^\tau)}{q\theta^p\tau^{p-1}}.$$

Moreover, the optimality condition for the minimization problem with $\theta > 0$ easily show that we have

$$|\nabla^- F|(x(\theta)) \le \left(\frac{d(x(\theta), x_k^\tau)}{\theta\tau}\right)^{p-1},$$

which wan also be written as $|\nabla^- F|(x(\theta))^q \le \left(\frac{d(x(\theta), x_k^\tau)}{\theta\tau}\right)^p$.

We now come back to the function m and use

$$m(0) - m(1) = -\int_0^1 m'(\theta)\,\mathrm{d}\theta$$

together with

$$-m'(\theta) = \frac{d(x(\theta), x_k^\tau)^p}{q\theta^p\tau^{p-1}} \ge \frac{\tau}{q}|\nabla^- F(x(\theta))|^q. \tag{11}$$

Hence, we get an improved version of (8):

$$F(x_{k+1}^\tau) + \frac{d(x_{k+1}^\tau, x_k^\tau)^p}{p\tau^{p-1}} \le F(x_k^\tau) - \frac{\tau}{q}\int_0^1 |\nabla^- F(x(\theta))|^q \mathrm{d}\theta.$$

If we sum up for $k = 0, 1, 2, \dots$ and then take the limit $\tau \to 0$, we can prove, for the limit curve, the inequality

$$F(x(T)) + \frac{1}{p}\int_0^T |x'|(t)^p \mathrm{d}t + \frac{1}{q}\int_0^T |\nabla^- F(x(t))|^q \mathrm{d}t \le F(x(0)), \tag{12}$$

under some suitable assumptions that we must select. In particular, we need lower-semicontinuity of F in order to handle the term $F(x_{k+1}^\tau)$ (which will become $F(x(t))$ at the limit), but we also need lower-semicontinuity of the slope $|\nabla^- F|$ in order to handle the corresponding term.

3 Introduction to OT, the Wasserstein space, and the PDE of Wasserstein gradient flows

This section will contain a gentle introduction to the topic of optimal transport and the distances it induces on the space of probability measures, moving forward the applications of these notions to those PDEs which can be expressed as a gradient flow in such a space.

3.1 Monge and Kantorovich problems

The whole theory of optimal transport was born with the following problem proposed by Monge in 1781 ([45]), that we will express in modern mathematical language (see [54, 55, 49]). Given two probability measures μ, ν on a space X, find a map $T = X \to X$ such that $T_\# \mu = \nu$ which minimizes a certain cost. The condition $T_\# \mu = \nu$ means that the image measure of μ through T is ν, where $T_\# \mu$ is the measure characterized by

$$(T_\# \mu)(A) = \mu(T^{-1}(A)) \quad \text{for every measurable set } A,$$
$$\text{or } \int_Y \phi \,\mathrm{d}\,(T_\# \mu) = \int_X \phi \circ T \,\mathrm{d}\mu \quad \text{for every measurable function } \phi.$$

The cost of a transport map T is evaluated in terms of a transport cost $c : X \times X \to \mathbb{R}$ and we minimize the quantity

$$\int_X c(x, T(x))\,\mathrm{d}\mu(x)$$

among all the maps satisfying $T_\# \mu = \nu$. This means that we have a collection of particles, distributed according to μ, that have to be moved, so that they arrange according to a new distribution ν. The cost $c(x, y)$ reprents the cost to move a unit mass from x to y. The map T describes the movement, and $T(x)$ represents the destination of the particle originally located at x.

The problem of Monge has stayed with no solution (does a minimizer exist? how to characterize it?. . .) till the progress made in the 1940s with the work by Kantorovich ([38]). For simplicity of the exposition, we will suppose that c is continuous and symmetric: $c(x, y) = c(y, x)$ and X to be a compact metric space.

The formulation proposed by Kantorovich of the problem raised by Monge is the following: consider the problem

$$(KP) \quad \min \left\{ \int_{X\times X} c \,\mathrm{d}\gamma \ : \ \gamma \in \Pi(\mu, \nu) \right\}, \tag{13}$$

where $\Pi(\mu, \nu)$ is the set of the so-called *transport plans*, i.e.

$$\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(X \times X) : \ (\pi_0)_\# \gamma = \mu,\ (\pi_1)_\# \gamma = \nu, \}$$

where π_0 and π_1 are the two projections of $X \times X$ onto its factors. These probability measures over $X \times X$ are an alternative way to describe the displacement of the particles of μ: instead of saying, for each x, which is the destination $T(x)$ of the particle originally located at x, we say for each pair (x, y) how many particles go from x to y. It is clear that this description allows for more general movements, since from a single point x particles can a priori move to different destinations y. If multiple destinations really occur, then this movement cannot be described through a map T. It can be easily checked that if $(id, T)_\# \mu$ belongs to $\Pi(\mu, \nu)$ then T pushes

μ onto ν (i.e. $T_\#\mu = \nu$) and the functional takes the form $\int c(x, T(x)) \mathrm{d}\mu(x)$, thus generalizing Monge's problem.

The minimizers for this problem are called *optimal transport plans* between μ and ν. Should γ be of the form $(id, T)_\#\mu$ for a measurable map $T : X \to X$ (i.e. when no splitting of the mass occurs), the map T would be called *optimal transport map* from μ to ν.

This generalized problem by Kantorovich is much easier to handle than the original one proposed by Monge: for instance in the Monge case we would need existence of at least a map T satisfying the constraints. This is not verified when $\mu = \delta_0$, if ν is not a single Dirac mass. On the contrary, there always exists at least a transport plan in $\Pi(\mu, \nu)$ (for instance we always have $\mu \otimes \nu \in \Pi(\mu, \nu)$). Moreover, one can state that (KP) is the relaxation of the original problem by Monge: if one considers the problem in the same setting, where the competitors are transport plans, but sets the functional at $+\infty$ on all the plans that are not of the form $(id, T)_\#\mu$, then one has a functional on $\Pi(\mu, \nu)$ whose relaxation (in the sense of the largest lower-semicontinuous functional smaller than the given one) is the functional in (KP) (see for instance Section 1.5 in [49]).

Anyway, it is important to note that an easy use of the direct method of the Calculus of Variations (i.e. taking a minimizing sequence, saying that it is compact in some topology - here it is the weak convergence of probability measures - finding a limit, and proving semicontinuity, or continuity, of the functional we minimize, so that the limit is a minimizer) proves that a minimum does exist. As a consequence, if one is interested in the problem of Monge, the question may become "does this minimizer come from a transport map T?" (note, on the contrary, that directly attacking by compactness and semicontinuity Monge's formulation is out of reach, because of the non-linearity of the constraint $T_\#\mu = \nu$, which is not closed under weak convergence).

Since the problem (KP) is a linear optimization under linear constraints, an important tool will be duality theory, which is typically used for convex problems. We will find a dual problem (DP) for (KP) and exploit the relations between dual and primal. A formal procedure to find the dual problem passes through an inf-sup exchange.

First express the constraint $\gamma \in \Pi(\mu, \nu)$ in the following way : notice that, if γ is a non-negative measure on $X \times X$, then we have

$$\sup_{\varphi,\, \psi} \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu - \int (\varphi(x) + \psi(y)) \,\mathrm{d}\gamma = \begin{cases} 0 & \text{if } \gamma \in \Pi(\mu, \nu) \\ +\infty & \text{otherwise} \end{cases}.$$

Hence, one can remove the constraints on γ by adding the previous sup, since if they are satisfied nothing has been added and if they are not one gets $+\infty$ and this will be avoided by the minimization. We may look at the problem we get and interchange the inf in γ and the sup in φ, ψ: for simplicity we will write $\varphi \oplus \psi$ for the function on $X \times X$ defined via $(x, y) \mapsto \varphi(x) + \psi(y)$ and

$$\min_{\gamma\geq 0}\int c\,\mathrm{d}\gamma+\sup_{\varphi,\psi}\left(\int\varphi\,\mathrm{d}\mu+\int\psi\,\mathrm{d}\nu-\int\varphi\oplus\psi\,\mathrm{d}\gamma\right)$$

becomes

$$\sup_{\varphi,\psi}\int\varphi\,\mathrm{d}\mu+\int\psi\,\mathrm{d}\nu+\inf_{\gamma\geq 0}\int(c-\varphi\oplus\psi)\,\mathrm{d}\gamma.$$

Obviously it is not always possible to exchange inf and sup, and the main tools to do this come from convex analysis. We refer to [49], Section 1.6.3 for a simple proof of this fact, or to [54], where the proof is based on Flenchel-Rockafellar duality (see, for instance, [30] or [51]). Anyway, we insist that in this case it is true that $\inf\sup=\sup\inf$.

Afterwards, one can re-write the inf in γ as a constraint on φ and ψ, since one has

$$\inf_{\gamma\geq 0}\int(c(x,y)-(\varphi(x)+\psi(y)))\,\mathrm{d}\gamma=\begin{cases}0 & \text{if } \varphi\oplus\psi\leq c \text{ on } X\times X\\ -\infty & \text{otherwise}\end{cases}.$$

This leads to the following dual optimization problem: given the two probabilities μ and ν and the cost function $c:X\times X\to[0,+\infty]$ we consider the problem

$$(DP)\qquad \max\left\{\int_X\varphi\,\mathrm{d}\mu+\int_X\psi\,\mathrm{d}\nu\ :\ \varphi,\psi\in C(X),\ \varphi\oplus\psi\leq c\text{ on }X\times X\right\}. \qquad (14)$$

This problem does not admit a straightforward existence result, since the class of admissible functions lacks compactness. Yet, it is possible to see that we can restrict the maximization to functions which are $c-$concave. Indeed, we can define a notion of $c-$transform similar to that of Legendre transform and use it as follows

Definition 4 Given a function $\chi:X\to\overline{\mathbb{R}}$ we define its $c-$*transform* (or $c-$conjugate function) by

$$\chi^c(y)=\inf_{x\in X}c(x,y)-\chi(x).$$

Moreover, we say that a function ψ is $c-$*concave* if there exists χ such that $\psi=\chi^c$.

It is quite easy to realize that, given a pair (ϕ,ψ) in the maximization problem (DP), one can always replace it with (ϕ,ϕ^c), and then with (ϕ^{cc},ϕ^c), and the constraints are preserved and the integrals increased. Moreover, any $c-$concave function shares the same modulus of continuity of the cost c. Hence, if c is uniformly continuous (which is always the case whenever c is continuous and X is compact), one can get a uniform modulus of continuity for suitable minimizing sequences and then prove existence for (DP), by applying Ascoli-Arzelà's Theorem. The functions (φ,ψ) realizing the maximum in (DP) are called *Kantorovich potentials* (we say that φ is the Kantorovich potential for the transport from μ to ν and that ψ is the one for the transport from ν to μ).

We look at two interesting cases. When $c(x,y)$ is equal to the distance $d(x,y)$ on the metric space X, then we can easily see that $c-$concave functions coincide with 1-Lipschitz functions and

$$\varphi \in \mathrm{Lip}_1 \Rightarrow \varphi^c = -\varphi. \tag{15}$$

Another interesting case is the case where $X = \Omega \subset \mathbb{R}^d$ and $c(x, y) = \frac{1}{2}|x - y|^2$. In this case if a function φ is $c-$concave then $x \mapsto \frac{x^2}{2} - \varphi(x)$ is a convex function (and this is an equivalence if $X = \mathbb{R}^d$).

A consequence of (15) is that, in the case where $c = d$, the duality result may be re-written as

$$\min(KP) = \max(DP) = \max_{\varphi \in \mathrm{Lip}_1} \int_X \varphi \, \mathrm{d}(\mu - \nu). \tag{16}$$

We now concentrate on the case when X is a domain $\Omega \subset \mathbb{R}^d$, and look at the existence of optimal transport maps T. We will use costs c of the form $c(x, y) = h(x - y)$ for a strictly convex function h.

The main tool is the duality result. If we have equality between the minimum of (KP) and the maximum of (DP) and both extremal values are realized, one can consider an optimal transport plan γ and a Kantorovich potential φ and write

$$\varphi(x) + \varphi^c(y) \leq c(x, y) \text{ on } X \times X \quad \text{and} \quad \varphi(x) + \varphi^c(y) = c(x, y) \text{ on } \operatorname{spt}\gamma.$$

The equality on $\operatorname{spt}\gamma$ is a consequence of the inequality which is valid everywhere and of

$$\int c \, \mathrm{d}\gamma = \int \varphi \, \mathrm{d}\mu + \int \varphi^c \, \mathrm{d}\nu = \int (\varphi(x) + \varphi^c(y)) \, \mathrm{d}\gamma,$$

which implies equality $\gamma-$a.e. These functions being continuous, the equality passes to the support of the measure. Once we have that, let us fix a point $(x_0, y_0) \in \operatorname{spt}\gamma$. One may deduce from the previous computations that

$$x \mapsto \varphi(x) - h(x - y_0) \quad \text{is maximal at } x = x_0$$

and, if φ is differentiable at x_0, one gets $\nabla\varphi(x_0) = \nabla h(x_0 - y_0)$ i.e. $y_0 = x_0 - \nabla h^*(\nabla\varphi(x_0))$. This shows that only one unique point y_0 can be such that $(x_0, y_0) \in \operatorname{spt}\gamma$, which means that γ is concentrated on a graph.

Theorem 1 *Assume $c(x, y) = h(x - y)$ for a strictly convex function h. Given μ and ν probability measures on a domain $\Omega \subset \mathbb{R}^d$ there exists an optimal transport plan γ; it is unique and of the form $(id, T)_\#\mu$, provided μ is absolutely continuous. Moreover there exists also at least a Kantorovich potential φ, and the gradient $\nabla\varphi$ is uniquely determined $\mu-$a.e. The optimal transport map T and the potential φ are linked by $T(x) = x - \nabla h^*(\nabla\varphi(x))$.*

In the case $h(z) = \frac{1}{2}|z|^2$ we have $T(x) = x - \nabla\varphi(x) = \nabla u(x)$, where $u(x) := \frac{x^2}{2} - \varphi(x)$ is a convex function.

The fact that the optimal map in the quadratic case is the gradient of a convex function is a well-known theorem, due to Brenier ([9, 10], see also [34, 32, 33, 42]). The only technical point above is the μ-a.e. differentiability of the potential φ, but φ has the same regularity of c, and Lipschitz functions are differentiable Lebesgue-a.e. (which explains the assumption that μ should be absolutely continuous).

In Theorem 1 only the part concerning the optimal map T is not symmetric in μ and ν: hence the uniqueness (up to additive constants) of the Kantorovich potential is true even if it ν (and not μ) has positive density a.e. (since one can retrieve φ from φ^c and viceversa).

3.2 The Wasserstein distances

Starting from the values of the problem (KP) we can define a set of distances over $\mathcal{P}(X)$.

We mainly consider costs of the form $c(x, y) = |x - y|^p$ in $X = \Omega \subset \mathbb{R}^d$, but the analysis can be adapted to a power of the distance in a more general metric space X. The exponent p will always be taken in $[1, +\infty)$ (we will not discuss the case $p = \infty$) in order to take advantage of the properties of the L^p norms. When Ω is unbounded we need to restrict our analysis to the following set of probabilities

$$\mathcal{P}_p(\Omega) := \left\{ \mu \in \mathcal{P}(\Omega) \; : \; \int_\Omega |x|^p \, \mathrm{d}\mu(x) < +\infty \right\}.$$

The distances that we want to consider are defined in the following way: for any $p \in [1, +\infty)$ set

$$W_p(\mu, \nu) = \big(\min\,(KP) \text{ with } c(x, y) = |x - y|^p \big)^{1/p}.$$

The quantities that we obtain in this way are called Wasserstein distances (this is the standard name nowadays, even if this choice is highly debated). They are very important in many fields of applications and they seem a natural way to describe distances between equal amounts of mass distributed on a same space.

We summarize here some properties of these distances. Most proofs can be found in [49], chapter 5, or in [54] or [4].

Theorem 2 *The quantity W_p defined above is a distance over $\mathcal{P}_p(\Omega)$.*

If X is compact, for any $p \geq 1$ the function W_p is a distance over $\mathcal{P}(X)$ and the convergence with respect to this distance is equivalent to the weak convergence of probability measures.

Without assuming compactness of X, for any $p \geq 1$ the function W_p is a distance over $\mathcal{P}_p(X)$ and, given a measure μ and a sequence $(\mu_n)_n$ in $\mathbb{W}_p(X)$, the following are equivalent:

- *$\mu_n \to \mu$ according to W_p;*
- *$\mu_n \rightharpoonup \mu$ and $\int |x|^p \mathrm{d}\mu_n \to \int |x|^p \mathrm{d}\mu$;*
- *$\int_X \phi \, \mathrm{d}\mu_n \to \int_X \phi \, \mathrm{d}\mu$ for any $\phi \in C^0(X)$ whose growth is at most of order p (i.e. there exist constants A and B depending on ϕ such that $|\phi(x)| \leq A + Bd(x, x_0)^p$ for any x).*

To prove that the convergence according to W_p is equivalent to weak convergence (in the compact case) one first establish this result for $p = 1$, through the use of the duality formula in the form (16). Then it is possible to use the inequalities between the distances W_p to extend the result to a general p. Indeed, as a consequence of Hölder (or Jensen) inequalities, the Wasserstein distances are always ordered, i.e. $W_{p_1} \le W_{p_2}$ if $p_1 \le p_2$. Reversed inequalities are possible only if Ω is bounded, and in this case we have, if set $D = \operatorname{diam}(\Omega)$, for $p_1 \le p_2$, we have

$$W_{p_1} \le W_{p_2} \le D^{1-p_1/p_2} W_{p_1}^{p_1/p_2}.$$

After this short introduction to the metric space $\mathbb{W}_p := (\mathcal{P}_p(X), W_p)$ and its topology, since we focus on the Euclidean case where the metric space X is a domain $\Omega \subset \mathbb{R}^d$, we want to study the curves valued in $\mathbb{W}_p(\Omega)$ in connections with PDEs.

The main point is to identify the absolutely continuous curves in the space $\mathbb{W}_p(\Omega)$ with solutions of the continuity equation $\partial_t \varrho_t + \nabla \cdot (\mathbf{v}_t \varrho_t) = 0$ with L^p vector fields $\mathbf{v}_t$. Moreover, we want to connect the L^p norm of $\mathbf{v}_t$ with the metric derivative $|\varrho'|(t)$.

We recall that standard considerations from fluid mechanics tell us that the continuity equation above may be interpreted as the equation ruling the evolution of the density ϱ_t of a family of particles initially distributed according to μ_0 and each of which follows the flow

$$\begin{cases} y_x'(t) = \mathbf{v}_t(y_x(t)) \\ y_x(0) = x. \end{cases}$$

The main theorem in this setting (originally proven in [4]) relates absolutely continuous curves in $\mathbb{W}_p$ with solutions of the continuity equation:

Theorem 3 *Let $(\varrho_t)_{t\in[0,1]}$ be an absolutely continuous curve in $\mathbb{W}_p(\Omega)$ (for $p > 1$ and $\Omega \subset \mathbb{R}^d$ an open domain). Then for a.e. $t \in [0,1]$ there exists a vector field $\mathbf{v}_t \in L^p(\varrho_t; \mathbb{R}^d)$ such that*

- *the continuity equation $\partial_t \varrho_t + \nabla \cdot (\mathbf{v}_t \varrho_t) = 0$ is satisfied in the sense of distributions,*
- *for a.e. t we have $||\mathbf{v}_t||_{L^p(\varrho_t)} \le |\varrho'|(t)$ (where $|\varrho'|(t)$ denotes the metric derivative at time t of the curve $t \mapsto \varrho_t$, w.r.t. the distance W_p);*

Conversely, if $(\varrho_t)_{t\in[0,1]}$ is a family of measures in $\mathcal{P}_p(\Omega)$ and for each t we have a vector field $\mathbf{v}_t \in L^p(\varrho_t; \mathbb{R}^d)$ with $\int_0^1 ||\mathbf{v}_t||_{L^p(\varrho_t)}\,\mathrm{d}t < +\infty$ solving $\partial_t \varrho_t + \nabla \cdot (\mathbf{v}_t \varrho_t) = 0$, then $(\varrho_t)_t$ is absolutely continuous in $\mathbb{W}_p(\Omega)$ and for a.e. t we have $|\varrho'|(t) \le ||\mathbf{v}_t||_{L^p(\varrho_t)}$.

Note that, as a consequence of the second part of the statement, the vector field $\mathbf{v}_t$ introduced in the first part must a posteriori satisfy $||\mathbf{v}_t||_{L^p(\varrho_t)} = |\varrho'|(t)$.

We will not give the proof of this theorem, which is quite involved. The main reference is [4] (but the reader can also find alternative proofs in Chapter 5 of [49], in the case where Ω is compact). Yet, if the reader wants an idea of the reason for this theorem to be true, it is possible to start from the case of two time steps: there are two measures ϱ_t and ϱ_{t+h} and there are several ways for moving the particles so as to reconstruct the latter from the former. It is exactly as when we look for a transport. One of these transports is optimal in the sense that it minimizes $\int |T(x)-x|^p \mathrm{d}\varrho_t(x)$ and the value of this integral equals $W_p^p(\varrho_t, \varrho_{t+h})$. If we call $\mathbf{v}_t(x)$ the "discrete velocity of the particle located at x at time t, i.e. $\mathbf{v}_t(x) = (T(x)-x)/h$, one has $||\mathbf{v}_t||_{L^p(\varrho_t)} = \frac{1}{h} W_p(\varrho_t, \varrho_{t+h})$. We can easily guess that, at least formally, the result of the previous theorem can be obtained as a limit as $h \to 0$.

Because of the role played by the L^p norm of the velocity field in the continuity equation it turns out that the following functional is extremely useful in optimal transport theory:

$$(\varrho, \mathbf{v}) \mapsto \int_0^1 \int_\Omega |\mathbf{v}_t|^p \mathrm{d}\varrho_t$$

to be computed on pairs $(\varrho, \mathbf{v})$ such that $\partial_t \varrho_t + \nabla \cdot (\mathbf{v}_t \varrho_t) = 0$. In particular, we have, for $p > 1$

$$W_p^p(\mu, \nu) = \min \left\{ \int_0^1 \int_\Omega |\mathbf{v}_t|^p \mathrm{d}\varrho_t \,\mathrm{d}t \ : \ \partial_t \varrho_t + \nabla \cdot (\mathbf{v}_t \varrho_t) = 0, \ \varrho_0 = \mu, \ \varrho_1 = \nu \right\}. \tag{17}$$

This is what is usually called *Benamou-Brenier formula* ([8]).

On the other hand, this minimization problem in the variables $(\varrho_t, \mathbf{v}_t)$ has non-linear constraints (due to the product $\mathbf{v}_t \varrho_t$) and the functional is non-convex (since $(s, z) \mapsto s|z|^p$ is not convex). Yet, it is possible to transform it into a convex problem. For this, it is sufficient to switch variables, from $(\varrho_t, \mathbf{v}_t)$ into (ϱ_t, E_t) where $E_t = \mathbf{v}_t \varrho_t$. We then define the functional

$$\mathcal{B}_p(\varrho, E) := \begin{cases} \int_0^1 \int_\Omega |\mathbf{v}_t|^p \mathrm{d}\varrho_t \,\mathrm{d}t & \text{if } E = \varrho \mathbf{v}, \\ +\infty & \text{if not.} \end{cases}$$

We will not provide details here but it is possible to prove that $\mathcal{B}_p$ is convex and lower-semicontinuous for the weak convergence as measures of its variables. Its convexity is the key point of the numerical methods (as it was first done in [8]).

Once we know about curves in their generality, it is interesting to think about geodesics. The following result is a characterization of geodesics in $W_p(\Omega)$ when Ω is a convex domain in $\mathbb{R}^d$. This procedure is also known as *McCann's displacement interpolation* (see [43]).

Theorem 4 *If $\Omega \subset \mathbb{R}^d$ is convex, then all the spaces $\mathbb{W}_p(\Omega)$ are length spaces and if μ and ν belong to $\mathbb{W}_p(\Omega)$, and γ is an optimal transport plan from μ to ν for the cost $c_p(x, y) = |x-y|^p$, then the curve*

$$\mu^\gamma(t) = (\pi_t)_\# \gamma$$

where $\pi_t : \Omega \times \Omega \to \Omega$ is given by $\pi_t(x, y) = (1-t)x + ty$, is a constant-speed geodesic from μ to ν. In the case $p > 1$ all the constant-speed geodesics are of this form, and, if μ is absolutely continuous, then there is only one geodesic and it has the form

$$\mu_t = [T_t]_\#\mu, \qquad \textit{where } T_t := (1-t)id + tT$$

where T is the optimal transport map from μ to ν. In this case, the velocity field $\mathbf{v}_t$ of the geodesic μ_t is given by $\mathbf{v}_t = (T - id) \circ (T_t)^{-1}$. In particular, for $t = 0$ we have $\mathbf{v}_0 = -\nabla h^*(\nabla\varphi)$ and for $t = 1$ we have $\mathbf{v}_1 = \nabla h^*(\nabla\psi)$, where $h(z) = \frac{1}{p}|z|^p$ and (φ, ψ) are the Kantorovich potentials in the transport from μ to ν for the cost $c(x, y) = h(x - y)$.

3.3 Minimizing movement schemes in the Wasserstein space and evolution PDEs

Thanks to all the theory which has been described so far, it is natural to study (linear or non-linear) gradient flows in the space $\mathbb{W}_p(\Omega)$ and to connect them to PDEs of the form of a continuity equation. The most convenient way to study this is to start from the time-discretized problem, i.e. to consider a sequence of iterated minimization problems:

$$\varrho^\tau_{k+1} = \mathbf{\Pi}[\varrho^\tau_k], \quad \mathbf{\Pi}[g] := \text{argmin}_\varrho\, F(\varrho) + \frac{W_p^p(\varrho, g)}{p\tau^{p-1}}. \tag{18}$$

This iterated minimization scheme is called *JKO scheme* (after Jordan, Kinderleher and Otto, [37]). Actually, this name was originally only used for the case $p = 2$ and we will refer to this scheme as W_p−JKO.

Note that we denote now the measures on Ω by the letter ϱ instead of μ or ν because we expect them to be absolutely continuous measures with nice (smooth) densities, and we want to study the PDE they solve. Note that a priori one could use the p−th power of arbitrary distances, and hence use W_q^p instead of W_p^p but in the Wasserstein space $\mathbb{W}_p$ the distance is defined as the power $1/p$ of a transport cost; only in the case $p = q$ the exponent goes away and we are lead to a minimization problem involving $F(\varrho)$ and a transport cost of the form

$$\mathcal{T}_c(\varrho, \nu) := \min\left\{\int c(x, y)\,\mathrm{d}\gamma \;:\; \gamma \in \Pi(\varrho, \nu)\right\},$$

for $\nu = \varrho^\tau_k$. By the way, if we want to be more general, we can also replace the minimization problem in (18) with

$$\mathbf{\Pi}[g] := \text{argmin}_\varrho\, F(\varrho) + \mathcal{T}_c(\varrho, g),$$

where $c(x, y) = \tau h(\frac{x-y}{\tau})$ for a strictly convex function h (we come back to (18) when $h(z) = \frac{1}{p}|z|^p$).

In the particular case of the space $\mathbb{W}_p(\Omega)$, which has some additional structure, if compared to arbitrary metric spaces, we would like to give a PDE description of the curves that we obtain as gradient flows, and this will pass through the optimality conditions of the minimization problem (18). In order to study these optimality conditions, we introduce the notion of first variation of a functional. This will be done in a very sketchy and formal way (we refer to Chapter 7 in [49] for more details).

Given a functional $G : \mathcal{P}(\Omega) \to \mathbb{R}$ we call $\frac{\delta G}{\delta \varrho}[\varrho]$, if it exists, the unique (up to additive constants) function such that $\frac{d}{d\varepsilon} G(\varrho + \varepsilon\chi)_{|\varepsilon=0} = \int \frac{\delta G}{\delta \varrho}[\varrho] d\chi$ for every perturbation χ such that, at least for $\varepsilon \in [0, \varepsilon_0]$, the measure $\varrho + \varepsilon\chi$ belongs to $\mathcal{P}(\Omega)$. The function $\frac{\delta G}{\delta \varrho}[\varrho]$ is called *first variation* of the functional G at ϱ. In order to understand this notion, the easiest possibility is to analyze some examples.

The main classes of examples are the following functionals[1]

$$F(\varrho) = \int f(x, \varrho(x)) \mathrm{d}x,$$

where $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a function which is convex and superlinear in its second variable (and the functional F is set to $+\infty$ if ϱ is not absolutely continuous w.r.t. the Lebesgue measure. In this case it is quite easy to realize that we have

$$\frac{\delta F}{\delta \varrho}[\varrho] = f'(x, \varrho),$$

where the derivative f' is also taken w.r.t. the second variable. A classical example is to consider $f(x, s) = s^m + V(x)s$ or $f(x, \varrho) = s \log s + V(x)s$, where $V : \Omega \to \mathbb{R}$ is regular enough.

It is clear that the first variation of a functional is a crucial tool to write optimality conditions for variational problems involving such a functional. In order to study the problem (18), we need to complete the picture by undestanding the first variation of functionals of the form $\varrho \mapsto \mathcal{T}_c(\varrho, \nu)$. The result is the following:

Proposition 1 *Let $c : \Omega \times \Omega \to \mathbb{R}$ be a continuous cost function. Then the functional $\varrho \mapsto \mathcal{T}_c(\varrho, \nu)$ is convex, and its subdifferential at ϱ_0 coincides with the set of Kantorovich potentials $\{\varphi \in C^0(\Omega) : \int \varphi \, \mathrm{d}\varrho_0 + \int \varphi^c \, \mathrm{d}\nu = \mathcal{T}_c(\varrho, \nu)\}$. Moreover, if there is a unique c-concave Kantorovich potential φ from ϱ_0 to ν up to additive constants, then we also have $\frac{\delta \mathcal{T}_c(\cdot, \nu)}{\delta \varrho}[\varrho_0] = \varphi$.*

Even if a complete proof of the above proposition is not totally trivial (and Chapter 7 in [49] only provides it in the case where Ω is compact), one can guess why this is true from the following considerations. Using the duality formula we have

$$\mathcal{T}_c(\varrho, \nu) = \max_{\varphi, \psi \,:\, \varphi \oplus \psi \leq c} \int_\Omega \varphi \, \mathrm{d}\varrho + \int_\Omega \psi \, \mathrm{d}\nu.$$

This expresses $\mathcal{T}_c$ as a supremum of linear functionals in ϱ and shows convexity. Standard considerations from convex analysis allow to identify the subdifferential as the set of functions φ attaining the maximum. An alternative point of view is to consider the functional $\varrho \mapsto \int \varphi \, \mathrm{d}\varrho + \int \psi \, \mathrm{d}\nu$ for fixed (φ, ψ), in which case the first

[1] Note that in some cases the functionals that we use are actually valued in $\mathbb{R} \cup \{+\infty\}$, and we restrict to a suitable class of perturbations χ which make the corresponding functional finite.

variation is of course φ; then it is easy to see that the first variation of the supremum may be obtained (in case of uniqueness) just by selecting the optimal pair (φ, ψ).

Once we know how to compute first variations, we come back to the optimality conditions for the minimization problem (18). Which are these optimality conditions? roughly speaking, we should have

$$\frac{\delta F}{\delta \varrho}[\varrho^\tau_{k+1}] + \varphi = const$$

(where the reasons for having a constant instead of 0 is the fact that, in the space of probability measures, only zero-mean densities are considered as admissible perturbations, and the first variations are always defined up to additive constants). Note that here φ is the Kantorovich potential associated with the transport from ϱ^τ_{k+1} to ϱ^τ_k (and not viceversa), and with the transport cost $c(x,y) = \tau h\left(\frac{x-y}{\tau}\right)$. It is related to the optimal transport map T through

$$\nabla h\left(\frac{x - T(x)}{\tau}\right) = \nabla\varphi(x) \quad \text{i.e.} \quad T(x) = x - \tau\nabla h^*(\nabla\varphi(x)).$$

Let us look at the consequences we can get from this optimality condition. Actually, if we combine the fact that the above sum is constant with the formula for the optimal T, we get

$$-\mathbf{v}(x) := \frac{T(x) - x}{\tau} = -\nabla h^*\big(\nabla\varphi(x)\big) = \nabla h^*\left(\nabla \frac{\delta F}{\delta \varrho}[\varrho]\right)(x), \tag{19}$$

where used the fact that we assume h^* to be even to remove the minus signs in the last expression.

Why did we denote by $-\mathbf{v}$ the ratio $\frac{T(x)-x}{\tau}$? because, as a ratio between a displacement and a time step, it has the meaning of a velocity, but since it is the displacement associated to the transport from ϱ^τ_{k+1} to ϱ^τ_k, it is better to view it rather as a backward velocity (which justifies the minus sign).

Since here we have $\mathbf{v} = -\nabla h^*\big(\nabla \frac{\delta F}{\delta \varrho}[\varrho]\big)$, this suggests that at the limit $\tau \to 0$ we will find a solution of

$$\partial_t \varrho - \nabla \cdot \left(\varrho \nabla h^*\left(\nabla\left(\frac{\delta F}{\delta \varrho}[\varrho]\right)\right)\right) = 0. \tag{20}$$

This is a PDE where the velocity field in the continuity equation depends on the density ϱ itself.

We can see many interesting examples.

Let us start from the case where h is quadratic, i.e. $h(z) = |z|^2/2$ and $\nabla h^* = id$. In this case the equation is simpler.

First, suppose $F(\varrho) = \int f(\varrho(x))\mathrm{d}x$, with $f(s) = s\log s$. In such a case we have $f'(s) = \log s + 1$ and $\nabla(f'(\varrho)) = \frac{\nabla\varrho}{\varrho}$: this means that the gradient flow equation associated with the functional F would be the *Heat Equation*

$$\partial_t \varrho - \Delta \varrho = 0.$$

Using $F(\varrho) = \int f(\varrho(x))\,\mathrm{d}x + \int V(x)\,\mathrm{d}\varrho(x)$, we would have the *Fokker-Planck Equation*

$$\partial_t \varrho - \Delta \varrho - \nabla \cdot (\varrho \nabla V) = 0.$$

Changing the function f we can obtain other forms of diffusion. For instance, if one uses

$$F(\varrho) = \frac{1}{m-1}\int \varrho^m(x)\,\mathrm{d}x + \int V(x)\varrho(x)\,\mathrm{d}x,$$

the equation we obtain is

$$\partial_t \varrho - \Delta(\varrho^m) - \nabla \cdot (\varrho \nabla V) = 0.$$

When $m > 1$ this equation is called *Porous Media Equation* (see [53] for a complete monography and [46] for its treatment with the JKO scheme) and models the diffusion of a fluid into a material whose porosity slows down the diffusion. Among the properties of this equation there is a finite-speed propagation effect, different from linear diffusion (if ϱ_0 is compactly supported the same stays true for ϱ_t, $t > 0$, while this is not the case for the heat equation). It is also interesting to consider the case $m < 1$: the function $\varrho^m - \varrho$ is concave, but the negative coefficient $1/(m-1)$ finally gives a convex function (which is, unfortunately, not superlinear at infinity, hece more difficult to handle). The PDE that we get as a gradient flow is called *Fast diffusion* equation, and has opposite properties in terms of diffusion rate than the porous medium one.

We can then consider other choices of transport cost. When $h(z) = \frac{1}{p}|z|^p$ we have $\nabla h^*(z) = z^{q-1}$. When taking $F(\varrho) = \int f(\varrho(x))\mathrm{d}x$ we obtain the following equation

$$\partial_t \varrho = \nabla \cdot (\varrho f''(\varrho)^{q-1} (\nabla \varrho)^{q-1}),$$

which can also be rewritten as

$$\partial_t \varrho = \Delta_q(\phi(\varrho)),$$

where Δ_q is the q–Laplace operator $\Delta_q u := \nabla \cdot ((\nabla u)^{q-1})$ and ϕ is defined via $\phi'(s) = s^{p-1} f''(s)$ (in this way, when f is convex ϕ is non-decreasing).

The case where we choose $f(s) = s \log s$ provides an interesting family of equations : those which are not linear, but are as linear as possible in this class, as they are one-homogeneous. In particular, among these equations we find $\partial_t \varrho = \Delta_q(\varrho^{p-1})$. The same happens when adding a potential V: $\partial_t \varrho = \nabla \cdot (\varrho(\nabla(\log \varrho + V))^{q-1})$.

3.4 Geodesic convexity in $\mathbb{W}_p$

For several reasons in the theory of gradient flows in Wasserstein spaces it would be useful to know when a functional is geodesically convex. Being geodesically convex in $\mathbb{W}_p$ means that $t \mapsto F(\mu_t)$ is a convex function whenever $t \mapsto \mu_t$ is a constant-speed geodesic. Knowing the form of geodesics in W_p, we can say that this corresponds to the convexity of $t \mapsto F((T_t)_\# \mu)$ whenever $T_t := (1-t)id + tT$ and T is the optimal transport map from a measure μ to another measure ν. This can also be extended to the case where the transport map T is optimal for a transport cost of the form $c(x, y) = h(x - y)$ when h is not necessarily a power. In this case the curve $\mu_t = (T_t)_\# \mu)$ cannot be interpreted as a geodesic, but as an optimal displacement moving μ into ν. The notion of geodesical convexity is indeed also called *displacement convexity* and was first introduced by McCann in [43]. In the framework of these notes, geodesic convexity could be used in two - strongly related - aspects: one is the abstract notion of EVI gradient flows in metric spaces (only available for linear gradient flows, i.e. $p = 2$), the other is a sharp estimate of the rate of decrease of a functional G along the gradient flow of F which can be performed along the JKO scheme when G is displacement convex.

We will consider separately two interesting examples of functionals: $\mathscr{V}(\varrho) := \int V \mathrm{d}\varrho$ and $\mathscr{F}(\varrho) := \int f(\varrho(x))\mathrm{d}x$.

It is not difficult to check that the convexity of V is enough to guarantee geodesic convexity of $\mathscr{V}$, since

$$\mathscr{V}(\mu_t) = \int V \,\mathrm{d}\big((1-t)id + tT\big)_\# \mu = \int V\big((1-t)x + tT(x)\big)\,\mathrm{d}\mu,$$

The most interesting displacement convexity result is the one for functionals depending on the density. To consider these functionals, we need some technical facts.

The starting point is the computation of the density of an image measure, via standard change-of-variable techniques: if $T : \Omega \to \Omega$ is a map smooth enough[2] and injective, and $\det(DT(x)) \neq 0$ for a.e. $x \in \Omega$, if we set $\varrho_T := T_\# \varrho$, then ϱ_T is absolutely continuous with density given by

$$\varrho_T = \frac{\varrho}{\det(DT)} \circ T^{-1}.$$

Then, using a well-known fact in linear algebra, that the determinant raised to the power $1/d$ is a concave function on the set of positive-definite symmetric $d \times d$ matrices (which can be generalized to the case where the eigenvalues are real and non-negative), we can obtain the following result due to McCann

[2] We need at least T to be countably Lipschitz, i.e. Ω may be written as a countable union of measurable sets $(\Omega_i)_{i\geq 0}$ with Ω_0 negligible and $T \llcorner \Omega_i$ Lipschitz continuous for every $i \geq 1$.

Theorem 5 *Suppose that $f(0) = 0$ and that $s \mapsto s^d f(s^{-d})$ is convex and decreasing. Suppose that Ω is convex and take $1 < p < \infty$. Then $\mathscr{F}$ is displacement convex (hence geodesically convex in $\mathbb{W}_p$ for every $p > 1$).*

Remark 1 Note that the assumption that $s \mapsto s^d f(s^{-d})$ is convex and decreasing implies that f itself is convex (the reader can check it as an exercise), a property which can be useful to establish, for instance, lower semicontinuity of $\mathscr{F}$.

Here are some examples of convex functions satisfying the assumptions of Theorem 5:

- all the power functions $f(t) = t^q$ for $q > 1$ satisfy these assumptions, since $s^d f(s^{-d}) = s^{-d(q-1)}$ is convex and decreasing;
- the entropy function $f(t) = t \log t$ also satisfies it: $s^d f(s^{-d}) = -d \log s$ is convex and decreasing;
- the function $f(t) = -t^m$ is convex for $m < 1$, and if we compute $s^d f(s^{-d}) = -t^{m(1-d)}$ we get a convex and decreasing function as soon as $1 - \frac{1}{d} \leq m < 1$. Note that in this case f is not superlinear, which requires some attention for the semicontinuity of $\mathscr{F}$.

Finally, it is interesting to observe that the geodesic convexity of higher-order functionals such as $\varrho \mapsto \int |\nabla \varrho|^p$ generally fails, or is a very delicate matter, while these functionals are among the most standard examples of convex functionals in the usual sense. See [23] for some examples of first-order geodesically convex functionals (in dimension one).

4 Convergence of the JKO scheme

Many possible proofs can be built for the convergence of the JKO scheme. In particular, one could follow the general theory developed in [4], i.e. checking all the assumptions to prove existence and uniqueness of an EVI gradient flow for the functional F in the space $\mathbb{W}_2(\Omega)$, and then characterizing the velocity field associated by Theorem 3 with the curve obtained as a gradient flow. In [4], it is proven, under suitable conditions, that such a vector field $\mathbf{v}_t$ must belong to what is defined as the *Wasserstein sub-differential* of the functional F, provided in particular that F is λ-geodescially convex. Then, the *Wasserstein sub-differential* is proven to be of the desired form (i.e. composed only of the gradient of the first variation of F, when F admits a first variation).

This approach has the advantage to use a general theory and to adapt it to the scopes of this particular setting. On the other hand, the important point when studying these PDEs is that the curves $(\varrho_t)_t$ obtained as a limit are weak solutions of the continuity equation; from this point of view, the metric notions of EDI and EVI solutions and the formalism developed in the first part of the book [4] are too general. Moreover, in the case of the space of measures, using optimal transport theory to select a suitable distance in the discrete scheme and choosing a suitable interpolation, the passage to the limit can be done by classical compactness techniques in functional analysis. Of course, there are often some difficulties in handling some non-linear terms, which are not always seen when using the theory of [4] (which is an advantage of such a general theory).

In this section we will first present the ideas to obtain convergence using compactness and weak convergence, and then an idea based on a specific adaptation of the EDI approach.

4.1 Passing to the limit the PDE

Our goal is to find a curve $(\varrho_t)_t$ which is a solution (in the distributional sense on $\mathbb{R}^d$, which is the same as on Ω, with suitable no-flux boundary conditions on $\partial\Omega$) of the PDE

$$\partial_t \varrho_t + \nabla \cdot (\varrho_t \mathbf{v}_t) = 0, \quad \text{where we require } \mathbf{v}_t = -\nabla \left(\frac{\delta F}{\delta \varrho}[\varrho_t] \right).$$

We will set $E = \varrho\mathbf{v}$ (the variable E is called *momentum*, in physics) and require at least that E is a finite vector measure over $\Omega \times [0, T]$, acting on test functions ϕ via $\langle E, \phi\rangle := \int dt \int \phi(t, x) \cdot \mathbf{v}_t \, \mathrm{d}\varrho_t$. Being a finite vector measure is equivalent to $\int_0^T ||\mathbf{v}_t||_{L^1(\varrho_t)} \mathrm{d}t < +\infty$.

Starting from the discrete scheme (18), which defines a sequence $(\varrho_k^\tau)_k$, we also define a sequence of velocities $\mathbf{v}_k^\tau = (id - \mathrm{T})/\tau$, taking as T the optimal transport from ϱ_k^τ to ϱ_{k-1}^τ. The considerations in the previous sections guarantee that we have

$$\mathbf{v}_k^\tau = \nabla h^* \left(-\nabla \left(\frac{\delta F}{\delta \varrho}[\varrho_k^\tau] \right) \right).$$

This is proven thanks to the optimality conditions in (18) and the only delicate point is to guarantee the uniqueness of the Kantorovich potential in the transport from ϱ_k^τ to ϱ_{k-1}^τ. In some cases it is possible to prove a priori that the minimizers in (18) have strictly positive density a.e. whatever is ν (this is typically true when using the functional $\mathscr{F}$ with $f'(0) = -\infty$, as it is the case for the entropy functional), which provides uniqueness up to additive constants of the potential. In Section 8.3 of [49] full details are provided for the Fokker-Planck case, and it is indeed proved that the minimizers of the JKO scheme are strictly positive a.e., in this case. When positivity of the minimizer is not available, then one can obtain the same optimality conditions by first approximating ν with strictly positive densities, and then passing to the limit, or by adding an entropy term times a small parameter ε to the functional and then considenring $\varepsilon \to 0$.

Then, we build at least two interesting curves in the space of measures:

- first we can define some piecewise constant curves, i.e. $\overline{\varrho}^\tau_t := \varrho^\tau_{k+1}$ for $t \in (k\tau, (k+1)\tau]$; associated with this curve we also define the velocities $\overline{\mathbf{v}}^\tau_t = \mathbf{v}^\tau_{k+1}$ for $t \in (k\tau, (k+1)\tau]$ and the momentum variable $\overline{E}^\tau = \overline{\varrho}^\tau \overline{\mathbf{v}}^\tau$;
- then, we can also consider the densities $\widehat{\varrho}^\tau_t$ that interpolate the discrete values $(\varrho^\tau_k)_k$ along geodesics:

$$\widehat{\varrho}^\tau_t = \left(id - (k\tau - t)\mathbf{v}^\tau_k\right)_\# \varrho^\tau_k, \quad \text{for } t \in ((k-1)\tau, k\tau[; \tag{21}$$

  the velocities $\widehat{\mathbf{v}}^\tau_t$ are defined so that $(\widehat{\varrho}^\tau, \widehat{\mathbf{v}}^\tau)$ satisfy the continuity equation, taking

$$\widehat{\mathbf{v}}^\tau_t = \mathbf{v}^\tau_t \circ \left(id - (k\tau - t)\mathbf{v}^\tau_k\right)^{-1};$$

  moreover, as before, we define: $\widehat{E}^\tau = \widehat{\varrho}^\tau \widehat{\mathbf{v}}^\tau$.

After these definitions we look for a priori bounds on the curves and the velocities that we defined. We already know that we have

$$\sum_k \tau \left(\frac{W_p(\varrho^\tau_k, \varrho^\tau_{k-1})}{\tau}\right)^p \le C, \tag{22}$$

which is the discrete version of a $W^{1,p}$ estimate in time. As for $\widehat{\varrho}^\tau_t$, it is an absolutely continuous curve in the Wasserstein space and its velocity on the time interval $[(k-1)\tau, k\tau]$ is given by the ratio $W_p(\varrho^\tau_{k-1}, \varrho^\tau_k)/\tau$. Hence, the L^p norm of its velocity on $[0,T]$ is given by

$$\int_0^T |(\widehat{\varrho}^\tau)'|^p(t)\mathrm{d}t = \sum_k \tau \left(\frac{W_p(\varrho^\tau_k, \varrho^\tau_{k-1})}{\tau}\right)^p, \tag{23}$$

and, thanks to (22), it admits a uniform bound independent of τ. In our case, thanks to results on the continuity equation and the Wasserstein metric, this metric derivative is also equal to $||\widehat{\mathbf{v}}^\tau_t||_{L^p(\widehat{\varrho}^\tau_t)}$. This gives compactness of the curves $\widehat{\varrho}^\tau$, as well as Hölder estimates (since $W^{1,p} \subset C^{0,1/q}$). The characterization of the velocities $\overline{\mathbf{v}}^\tau$ and $\widehat{\mathbf{v}}^\tau$ allows to deduce bounds on these vector fields from the bounds on $W_p(\varrho^\tau_{k-1}, \varrho^\tau_k)/\tau$.

Considering all these facts, one obtains the following situation (see also [50] or Chapter 8 in [49]):

- The norm $\int ||\overline{\mathbf{v}}^\tau_t||^p_{L^p(\overline{\varrho}^\tau_t)}\mathrm{d}t$ is τ-uniformly bounded. This quantity is equal to $\mathcal{B}_p(\overline{\varrho}^\tau, \overline{E}^\tau)$.
- In particular, the bound is valid in L^1 as well, which implies that $\overline{E}^\tau$ is bounded in the space of measures over $[0,T] \times \Omega$.
- The very same estimates are true for $\widehat{\mathbf{v}}^\tau$ and $\widehat{E}^\tau$.
- The curves $\widehat{\varrho}^\tau$ are bounded in $W^{1,p}([0,T], \mathbb{W}_p(\Omega))$ and hence compact in $C^0([0,T], \mathbb{W}_p(\Omega))$.
- Up to a subsequence, one has $\widehat{\varrho}^\tau \to \varrho$, as $\tau \to 0$, uniformly according to the W_p distance.

- From the estimate $W_p(\overline{\varrho}^\tau_t, \widehat{\varrho}^\tau_t) \le C\tau^{1/q}$ one gets that ϱ^τ converges to the same limit ϱ in the same sense.
- If we denote by E a weak limit of $\widehat{E}^\tau$, since $(\widehat{\varrho}^\tau, \widehat{E}^\tau)$ solves the continuity equation, by linearity, passing to the weak limit, also (ϱ, E) solves the same equation.
- It is possible to prove that the weak limits of $\widehat{E}^\tau$ and $\overline{E}^\tau$ are the same.
- From the semicontinuity of $\mathcal{B}_p$ and the bound $\mathcal{B}_p(\overline{\varrho}^\tau, \overline{E}^\tau) \le C$, one gets $\mathcal{B}_p(\varrho, E) < +\infty$, which means that E is absolutely continuous w.r.t. ϱ and has an L^p density, so that we have for a.e. time t a measure E_t of the form $\varrho_t \mathbf{v}_t$.
- We only need to prove that one has $\mathbf{v}_t = -\left(\nabla\left(\frac{\delta F}{\delta\varrho}[\varrho_t]\right)\right)^{q-1}$ ϱ_t-a.e. and for a.e. t. This means proving

$$\overline{E}^\tau = -\overline{\varrho}^\tau \left(\nabla\left(\frac{\delta F}{\delta\varrho}(\overline{\varrho}^\tau)\right)\right)^{q-1} \Rightarrow E = -\varrho\left(\nabla\left(\frac{\delta F}{\delta\varrho}[\varrho]\right)\right)^{q-1} \tag{24}$$

  as a limit as $\tau \to 0$. It is crucial in this step to consider the limit of $(\overline{\varrho}^\tau, \overline{E}^\tau)$ instead of $(\widehat{\varrho}^\tau, \widehat{E}^\tau)$.

This last step is the most critical one in many cases. It requires passing to the limit (in the sense of distributions) the terms involving $\frac{\delta F}{\delta\varrho}$ on a sequence $\varrho_j \rightharpoonup \varrho$ (we do not care here where this sequence comes from). We can see what happens with the main class of functionals that we introduced so far.

The easiest case is that when $p = 1$ and the functional is $\mathscr{V}$: we have

$$\varrho\nabla\left(\frac{\delta\mathscr{V}}{\delta\varrho}[\varrho]\right) = \varrho\nabla V.$$

This term is linear in ϱ and $\varrho_j \rightharpoonup \varrho$ obviously implies $\varrho_j\nabla V \rightharpoonup \varrho\nabla V$ as soon as $V \in C^1$ (so that ∇V is a continuous function, which is exactly the functional space whose dual is the space of measures). In case $\nabla V \notin C^0$ it is possible to handle this term as soon as suitable bounds on ϱ_j provide a better weak convergence.

The case of the functional $\mathscr{F}$ (still with $p = 2$) is harder. In the case of the entropy $\mathscr{F}(\varrho) = \int f(\varrho)$ with $f(s) = s \log s$ then everything works fine because, again, the corresponding term is linear:

$$\varrho\nabla\left(\frac{\delta\mathscr{F}}{\delta\varrho}[\varrho]\right) = \varrho\frac{\nabla\varrho}{\varrho} = \nabla\varrho.$$

Then, the convergence of this term in the sense of distributions is straightforward when $\varrho_j \to \varrho$. By the way, the entropy term $\mathscr{F}$ is also enough to obtain suitable bounds to handle V or W which are only Lipschitz, as in this case we need to turn the weak convergence $\varrho_j \rightharpoonup \varrho$ from the sense of measures to the L^1 sense, which is exactly possible because the superlinear bound $\int \varrho_j \log(\varrho_j) \le C$ provides equi-integrability for ϱ_j.

Yet, for other functions f, there is no more linearity, and we need stronger bounds. For instance, for $f(s) = s^m/(m-1)$, we have

$$\varrho\nabla\left(\frac{\delta\mathcal{F}}{\delta\varrho}[\varrho]\right) = \varrho\frac{m}{m-1}\nabla(\varrho^{m-1}) = \nabla(\varrho^m).$$

This means that weak convergence of ϱ_j is no more enough, but one also needs weak convergence of ϱ_j^m, which means strong convergence of ϱ_j. This can be proven when we have stronger bounds (if possible, Sobolev bounds on ϱ_j).

The main ingredient is indeed an H^1 bound in space, which comes from the fact that we have

$$\begin{aligned}\int_0^T\int_\Omega |\nabla(\varrho^{m-1/2})|\,\mathrm{d}x\mathrm{d}t &\approx \int_0^T\int_\Omega |\nabla(\varrho^{m-1})|^2\,\mathrm{d}\varrho_t\mathrm{d}t \\ &\approx \int_0^T\int_\Omega \frac{|\nabla\varphi|^2}{\tau^2}\,\mathrm{d}\varrho_t\mathrm{d}t = \int_0^T |\varrho'|(t)^2\mathrm{d}t \le C. \quad (25)\end{aligned}$$

This is only a bound on the H^1 norm in space, and does not involve time derivatives, so that we need to use the so-called *Aubin-Lions* lemma (see [5]), which interpolates between compactness in space and in time (roughly speaking: in our situation we have an L^2 bound in time with values in a strong space norm, H^1, but also a stronger bound in time, H^1, valued in the W_2 distance, which is weaker as it metrizes weak convergence; this information together can provide strong compactness).

However, so far we only considered the case $p = 2$. In this case the term E can be non-linear in ϱ but is in general linear in $\nabla\varrho$, which means that we only need to prove strong convergence of ϱ_j (and not of $\nabla\varrho_j$) in order to obtain (24). On the other hand, when $p \neq 2$ the relation between E and $\nabla\varrho$ is not any more linear, which means that we should obtain strong convergence not only of ϱ_j but also of its gradient, which is much harder (second-order Sobolev bounds would be needed, and they are not easy to obtain in the JKO scheme).

4.2 The EDI formulation

In this section we present a very different strategy; more adapted to deal with cases where the strong convergence of the densities or of their gradients is not easy to prove. It has been used, for instance, in [1, 22, 20] to deal with non-linear gradient flows in W_p, and in [29] to deal with a system in $W_2 \times W_2$.

The starting point of this section is the following observation; finding a solution of $\partial\varrho_t = \nabla\cdot(\varrho\nabla h^*(\nabla\frac{\delta F}{\delta\varrho}))$ amounts to finding a pair $(\varrho, \mathbf{v})$ such that one has at the same time the continuity equation $\partial_t\varrho + \nabla\cdot(\varrho\mathbf{v})$ and $\mathbf{v} = -\nabla h^*(\nabla\frac{\delta F}{\delta\varrho})$. Assuming that $(\varrho, \mathbf{v})$ is regular enough so that we can differentiate in time $F(\varrho_t)$ and obtain the so-called *chain rule* $\frac{\mathrm{d}}{\mathrm{d}t}F(\varrho_t) = \int\nabla\frac{\delta F}{\delta\varrho}\cdot\mathbf{v}_t\mathrm{d}\varrho_t$ for a.e. t, this is equivalent to the continuity equation together with the following inequality

$$F(\varrho_T) + \int_0^T \mathrm{d}t\int_\Omega h^*\left(\nabla\frac{\delta F}{\delta\varrho}\right)\mathrm{d}\varrho_t + \int_0^T\mathrm{d}t\int_\Omega h(\mathbf{v}_t)\mathrm{d}\varrho_t \le F(\varrho_0). \quad (26)$$

Indeed, using the same rule we can replace $F(\varrho_T)$ with $F(\varrho_0)+\int_0^T\int_\Omega \nabla\frac{\delta F}{\delta\varrho}\cdot\mathbf{v}_t\mathrm{d}\varrho_t$ and hence see that this is equivalent to

$$\int_0^T \mathrm{d}t\int_\Omega \mathrm{d}\varrho_t\left(h^*\left(\nabla\frac{\delta F}{\delta\varrho}\right)+h(\mathbf{v}_t)-\nabla\frac{\delta F}{\delta\varrho}\cdot(-\mathbf{v}_t)\right)\le 0,$$

which is equivalent to the a.e. equality $-\mathbf{v}_t=\nabla h^*(\nabla\frac{\delta F}{\delta\varrho})$.

The goal becomes now to prove that the JKO scheme converges hence to a curve $(\varrho,\mathbf{v})$ satisfying (26). We set

$$S(\varrho):=\int_\Omega \mathrm{d}\varrho\left|\nabla\frac{\delta F}{\delta\varrho}\right|^q.$$

This functional is typically an integral functional involving convex functions of the gradient of ϱ and, as such, is lower semicontinuous for the weak convergence of probability measures (see, for instance, Chapter 3 in [51]). We will assume that it is indeed l.s.c. in the cases of our interest. The, in order to obtain (26), it is enough to find a suitable interpolation ϱ_t^τ (we dot choose here a specific one) of the solutions ϱ_k^τ such that we have

$$F(\varrho_T^\tau)+\frac{1}{q}\int_0^T S(\varrho_t^\tau)\mathrm{d}t+\sum_k\tau\frac{1}{p}\left(\frac{W_p(\varrho_k^\tau,\varrho_{k+1}^\tau)}{\tau}\right)^p\le F(\varrho_0). \tag{27}$$

Indeed, by semicontinuity of F the first term would provide at the limit the term $F(\varrho_T)$, by semicontinuity of S and Fatou's lemma the second would provide $\frac{1}{q}\int_0^T S(\varrho_t)\mathrm{d}t$, and the third one canbe written in terms of the L^p norm of the velocity field of the piecewise geodesic interpolation and, by semicontinuity of the $\mathcal{B}_p$ functional, provides $\frac{1}{p}\int_0^T\mathrm{d}t\int_\Omega|\mathbf{v}_t|^p\mathrm{d}\varrho_t$.

A first natural question is whether the piecewise constant interpolation can be used to obtain (27). This would be the case if we had the following inequality involving one JKO step:

$$F(\varrho_{k+1}^\tau)+\tau\left(\frac{W_p(\varrho_k^\tau,\varrho_{k+1}^\tau)}{\tau}\right)^p\le F(\varrho_k^\tau). \tag{28}$$

Indeed, it is possible to see using the optimality conditions that we have

$$\left(\frac{x-T(x)}{\tau}\right)^{p-1}=\nabla\varphi=-\nabla\frac{\delta F}{\delta\varrho},$$

and we have, using $q=p/(p-1)$,

$$\left(\frac{W_p(\varrho_k^\tau,\varrho_{k+1}^\tau)}{\tau}\right)^p=\int\mathrm{d}\varrho_{k+1}^\tau\left|\frac{x-T(x)}{\tau}\right|^p=\int\mathrm{d}\varrho_{k+1}^\tau\left|\nabla\frac{\delta F}{\delta\varrho}\right|^q=S(\varrho_{k+1}^\tau), \tag{29}$$

which would allow to conclude writing

$$\tau\left(\frac{W_p(\varrho_k^\tau,\varrho_{k+1}^\tau)}{\tau}\right)^p=\frac{\tau}{q}S(\varrho_{k+1}^\tau)+\frac{\tau}{p}\left(\frac{W_p(\varrho_k^\tau,\varrho_{k+1}^\tau)}{\tau}\right)^p.$$

4.2.1 The case of a geodesically convex functional – the flow interchange technique

The difficult point is that the condition (28) is in general very difficult to obtain, while it would be easy to obtain

$$F(\varrho_{k+1}^\tau)+\frac{\tau}{p}\left(\frac{W_p(\varrho_k^\tau,\varrho_{k+1}^\tau)}{\tau}\right)^p\le F(\varrho_k^\tau).$$

Yet, even if this is enough to obtain some bounds on the curve ϱ^τ, it is not enough to recover the desired EDI inequality.

Nonetheless, it is possible to improve this computation when F is geodesically convex. We start from an inequality.

Proposition 2 *Given two probability densities ϱ, g and a functional G which is geodesically convex, then we have*

$$G(g)\ge G(\varrho)-\tau\int_\Omega\nabla\frac{\delta G}{\delta\varrho}[\varrho]\cdot\nabla h^*(\nabla\varphi)\mathrm{d}\varrho,$$

where φ is the Kantorovich potential in the transport from ϱ to g for the cost $c(x,y)=\tau h(\frac{x-y}{\tau})$.

In particular, if $g=\varrho_k^\tau$, $\varrho=\mathbf{\Pi}[g]$, $h(z)=|z|^p/p$ and $G=F$, we obtain (28)

Proof. Let ϱ_s be a constant speed geodesic from $\varrho_0=\varrho$ to $\varrho_1=g$. It admits a velocity field $\mathbf{v}_s$ such that $\partial_s\varrho_s+\nabla\cdot(\varrho_s\mathbf{v}_s)=0$. The geodesic convexity of G implies

$$G(g)\ge G(\varrho)+\frac{\mathrm{d}}{\mathrm{d}s}G(\varrho_s)_{|s=0}.$$

Let us compute

$$\frac{\mathrm{d}}{\mathrm{d}s}G(\varrho_s)=\int\frac{\delta G}{\delta\varrho}[\varrho_s]\partial_s\varrho=\int\nabla\frac{\delta G}{\delta\varrho}[\varrho_s]\cdot\mathbf{v}_s\varrho_s,$$

where the last equality comes from the continuity equation and integration by parts. Computing at time $s=0$ we have $\mathbf{v}_0=T-id=-\tau\nabla h^*(\nabla\varphi)$, which gives the claim.

The last part of the statement comes from the optimality conditions in the JKO scheme. This gives, using the explicit expression for ∇h^*,

$$F(\varrho_k^\tau)\ge F(\varrho_{k+1}^\tau)+\tau\int_\Omega\nabla\frac{\delta F}{\delta\varrho}[\varrho_{k+1}^\tau]\cdot\left(\nabla\frac{\delta F}{\delta\varrho}[\varrho_{k+1}^\tau]\right)^{q-1}\mathrm{d}\varrho_{k+1}^\tau,$$

which is the desired inequality. □

This kind of computation is known under the name of *flow interchange* technique and was introduced in [41]. Even if here we used it to prove a better estimate on the rate of decrease of F along the iterations of the JKO scheme for the same functional F, it could be used to obtain estimate on the rate of decrease of another functional G. In the case $p = 2$ the same estimate could also be obtained in a slightly different way: from the optimizer ϱ^τ_{k+1} start a curve ϱ_s such that at $s = 0$ it takes the value ϱ^τ_{k+1} and for $s > 0$ it follows the gradient flow in W_2 of G. This is a curve of possible competitors for the optimization problem defining ϱ^τ_{k+1}, so we should have

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(F(\varrho_s) + \frac{W_2^2(\varrho_s, \varrho^\tau_k)}{2\tau}\right)_{|s=0} \geq 0,$$

by optimality. Then, we can use the EVI formulation of the gradient flow of G (which is geodesically convex), and obtain the upper bound

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{W_2^2(\varrho_s, \varrho^\tau_k)}{2}\right)_{|s=0} \leq G(\varrho^\tau_k) - G(\varrho^\tau_{k+1}).$$

The conclusion follows by computing the derivative of F along the gradient flow of G (using as above the continuity equation satisfied by ϱ_s).

In practice, the flow interchange technique is a way to obtain in the discrete setting (the JKO scheme) the same computation that we would obtain in the continuous setting: if ϱ_t is a gradient flow of F, i.e. a solution of $\partial_t \varrho = \nabla \cdot (\varrho \nabla \frac{\delta F}{\delta \varrho})$ and we want to differentiate G we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} G(\varrho_t) = -\int \frac{\delta G}{\delta \varrho}[\varrho_t]\partial_t \varrho = \int \nabla \frac{\delta G}{\delta \varrho}[\varrho_t] \cdot \nabla \frac{\delta F}{\delta \varrho}[\varrho_t]\mathrm{d}\varrho_t.$$

A discrete analogue of this computation would be

$$G(\varrho^\tau_{k+1}) = G(\varrho^\tau_k) - \tau \int \nabla \frac{\delta G}{\delta \varrho}[\varrho^\tau_{k+1}] \cdot \nabla \frac{\delta F}{\delta \varrho}[\varrho^\tau_{k+1}]\mathrm{d}\varrho^\tau_{k+1}.$$

This is in general false, but the flow interchange technique allow to obtain it *as an inequality* as soon as G is geodesically convex.

As an other example of application of the flow interchange technique we cite the following.

Proposition 3 *Consider $F(\varrho) = \int f(\varrho(x))\mathrm{d}x$ and $G(\varrho) = \int g(\varrho(x))\mathrm{d}x$ and assume that G is geodesically convex in W_p. Consider the sequence ϱ^τ_k defined by the W_p-JKO scheme for the functional F. Then we have*

$$G(\varrho^\tau_k) \geq G(\varrho^\tau_{k+1}),$$

i.e. G decreases along the JKO scheme for F.

Proof. We apply Proposition 2 and obtain

$$G(\varrho_k^\tau) \geq G(\varrho_{k+1}^\tau) + \tau \int_\Omega \nabla g'(\varrho_{k+1}^\tau) \cdot \left(\nabla f'(\varrho_{k+1}^\tau])\right)^{q-1} \mathrm{d}\varrho_{k+1}^\tau$$

and then use $\nabla g'(\varrho) \cdot (\nabla f'(\varrho))^{q-1} = g''(\varrho) f''(\varrho)^{q-1} |\nabla \varrho|^q \geq 0$. □

4.2.2 The general case – the De Giorgi variational interpolation

In this section we want to prove that the limit of the JKO scheme solves the EDI condition also in the case when F is not geodesically convex, by choosing a better interpolation than the piecewise constant one.

The work has essentially been already done in the abstract metric setting in Section 2.3.

We define the De Giorgi variational interpolation as a curve $\hat{\varrho}_t^\tau$: for $t \in (k\tau, (k+1)\tau]$, we define $\theta = \frac{t-k\tau}{\tau} \in (0,1]$ and

$$\hat{\varrho}_t^\tau = \operatorname{argmin} \frac{W_p^p(\varrho, \varrho_k^\tau)}{p(\theta\tau)^{p-1}} + F(\varrho). \tag{30}$$

In particular for $t = (k+1)\tau$ we have $\theta = 1$ and we do retrieve $\hat{\varrho}_{k+1}^\tau = \varrho_{k+1}^\tau$, and if $t = k\tau$, $\theta = 0$ and the minimizer has to be $\varrho = \varrho_k^\tau$.

Using the computations of Section 2.3 this interpolation allows us to derive the following precursor to the EDI interpretation of our gradient flow :

$$F(\varrho_k^\tau) \geq F(\varrho_{k+1}^\tau) + \frac{\tau}{q} \int_{k\tau}^{(k+1)\tau} \frac{W_p^p(\hat{\varrho}_t^\tau, \varrho_k^\tau)}{\theta(\theta\tau)^{p-1}} \mathrm{d}t + \frac{W_p^p(\varrho_{k+1}^\tau, \varrho_k^\tau)}{p\tau^{p-1}}. \tag{31}$$

The desired inequality is obtained once we observe that the equality (29) aso holds when replacing τ with $\theta\tau$. This provides, summing over k, the desired inequality which then passes to the limit using lower semicontinuity.

We then finish this section by underlining that the techniques described here allow to obtain the inequality (26) which formally characterizes the solution of the PDE $\partial\varrho_t = \nabla \cdot (\varrho \nabla h^*(\nabla \frac{\delta F}{\delta \varrho}))$. Yet, the equivalence between the PDE and the inequality can only be established if we are able to prove the *chain rule*, i.e. that on the curve $(\varrho_t, \mathbf{v}_t)$ obtained at the limit $\tau \to 0$ the following relation holds

$$\frac{\mathrm{d}}{\mathrm{d}t} F(\varrho_t) = \int \nabla \frac{\delta F}{\delta \varrho}[\varrho_t] \cdot \mathbf{v}_t \mathrm{d}\varrho_t \quad \text{or} \quad F(\varrho_T) - F(\varrho_0) = \int_0^T \mathrm{d}t \int \nabla \frac{\delta F}{\delta \varrho}[\varrho_t] \cdot \mathbf{v}_t \mathrm{d}\varrho_t$$

This equality can easily be established at a formal level, but proving it rigorously when $(\varrho_t, \mathbf{v}_t)$ lack regularity is more delicate. In some cases it is possible to prove it by regularization (replacing ϱ and $\mathbf{v}$ by convolution with ϱ_ε and $\mathbf{v}_\varepsilon$, for which the formula holds, and passing to the limit). Another strategy discretizes the curve considering a finite family of time intervals $[t_k, t_{k+1}]$ and estimates the increments of F in terms of the velocity field $\mathbf{v}$ provided F is displacement convex. This is done in [1, 22]. An improvement is presented in [20], where it is proven that a large family of functionals of the form $\varrho \mapsto \int f(\varrho(x))\mathrm{d}x$ is approximated by functonals which are the difference of two displacement convex functionals, which allow to treat them separately. Anyway, we do not deal with the chain rule here in these notes and the results of this section require to be able to prove it by other means.

5 Higher-order estimates

Proposition 3 is an eaxample of result providing iterable estimates on the JKO scheme: it is proven that some quantities decrease from one step to the next one. In the statement of Proposition 3 the relevant quantitis are of zero order, in the sense that they do not involve derivatives of the densities ϱ^τ_k. We will develop here techniques to deal with BV and Sobolev-like estimates.

5.1 The five-gradients inequality and BV estimates

We start by presenting a general inequality, first studied in [26] which involves the gradients of two densities, of the corresponding Kantorovich potentials, and of an extra arbitrary convex function. Originally, the inequality was only stated in the case of a quadratic transport cost, and has later been extended to more general costs of the form $c(x, y) = h(x - y)$ in [18]. Extensions to the case of Riemannian manifolds are presented in [28].

Lemma 1 *Let $\Omega \subset \mathbb{R}^d$ be bounded and convex, let c be a transport cost given by $c(x, y) = h(x - y)$ for strictly convex h, let $\varrho, g \in W^{1,1}(\Omega)$ be two probability densities and $H \in C^2(\mathbb{R}^d)$ be a radially symmetric convex function. Then we have the following inequality*

$$\int_\Omega \Big(\nabla\varrho \cdot \nabla H(\nabla\varphi) + \nabla g \cdot \nabla H(\nabla\psi)\Big)\,\mathrm{d}x \geq 0, \tag{32}$$

where φ and ψ are the Kantorovich potentials in the optimal transport from ϱ to g for the cost c.

Proof. We will assume that all functions are smooth, an assumption which can be removed by aproximation.

We perform an integration by parts so that the left hand side of (32) becomes (we denote by $\mathbf{n}$ the exterior unit normal vector of the boundary $\partial\Omega$.

$$\int_{\partial\Omega} \Big(\varrho\,\nabla H(\nabla\varphi)\cdot\mathbf{n} + g\,\nabla H(\nabla\psi)\cdot\mathbf{n}\Big)\,\mathrm{d}\mathcal{H}^{d-1} - \int_\Omega \Big(\varrho\,\nabla\cdot\big[\nabla H(\nabla\varphi)\big] + g\,\nabla\cdot\big[\nabla H(\nabla\psi)\big]\Big)\,\mathrm{d}x.$$

We first look at the boundary term. By the radial symmetry of H the vector $\nabla H(z)$ is always a scalar multiple of z. Since the gradients of the Kantorovich potentials

$\nabla\varphi$ and $\nabla\psi$ calculated in boundary points are pointing outward Ω (since $T(x) = x - \nabla\varphi(x) \in \Omega$, and $S(x) = x - \nabla\psi(x) \in \Omega$) we have

$$\nabla H(\nabla\varphi(x)) \cdot \mathbf{n}(x) \geq 0 \qquad \text{and} \qquad \nabla H(\nabla\psi(x)) \cdot \mathbf{n}(x) \geq 0, \qquad \forall x \in \partial\Omega,$$

which proves that the boundary term is nonnegative.

We now write $g = T_\# \varrho$ for the optimal map T and expand the divergence, and obtain

$$\begin{aligned}\int_\Omega \Big(\varrho\, \nabla \cdot \big[\nabla H(\nabla\varphi)\big] + g\, \nabla \cdot \big[\nabla H(\nabla\psi)\big]\Big)\, \mathrm{d}x \\ = \int_\Omega \Big(\nabla \cdot \big[\nabla H(\nabla\varphi)\big] + (\nabla \cdot \big[\nabla H(\nabla\psi)\big]) \circ T\Big)\, \mathrm{d}\varrho \\ = \int_\Omega \left(\sum_{i,j} H_{ij}(\nabla\varphi)\varphi_{ij} + H_{ij}(\nabla\psi \circ T)\psi_{ij} \circ T\right) \mathrm{d}\varrho.\end{aligned}$$

We then use the fact that $\varphi \oplus \psi - c$ is maximal on pairs (x, y) of the form $y = T(x)$ in order to obtain, from the first and the second order conditions when considering perturbations of the form $x' = x + v$, $y' = y + v$, the following information (we also use the equality $c(x', y') = c(x, y)$):

$$\nabla\phi(x) + \nabla\psi(T(x)) = 0, \quad D^2\phi(x) + D^2\psi(T(x)) \leq 0,$$

where he last inequality is to be intended in the sense of symmetric matrices. We then have

$$\sum_{i,j} H_{ij}(\nabla\varphi)\varphi_{ij} + H_{ij}(\nabla\psi \circ T)\psi_{ij} \circ T = \sum_{i,j} H_{ij}(\nabla\varphi)\left(\varphi_{ij} + \psi_{ij} \circ T\right).$$

This is the trace of the product of the matrices $D^2H(\varphi)$, which is positive definite because of the convexity of H, and of $D^2\phi(x) + D^2\psi(T(x))$, which is negative definite. The trace is then nonpositive, and this proves the claim. □

In the quadratic case $h(z) = |z|^2/2$ it is also possible to improve this inequality and obtain the following.

Lemma 2 *Let $\Omega \subset \mathbb{R}^d$ be bounded and convex, let $\varrho, g \in W^{1,1}(\Omega)$ be two probability densities and $H \in C^2(\mathbb{R}^d)$ be a radially symmetric convex function. Then we have the following inequality*

$$\int_\Omega \Big(\nabla\varrho \cdot \nabla H(\nabla\varphi) + \nabla g \cdot \nabla H(\nabla\psi)\Big)\, \mathrm{d}x \geq \int_\Omega \mathrm{tr}[D^2H(\nabla\varphi)(D^2\varphi)^2(I - D^2\varphi)^{-1}]\mathrm{d}\varrho, \tag{33}$$

where φ and ψ are the Kantorovich potentials in the optimal transport from ϱ to g for the quadratic cost.

Proof. We just need to prove $D^2\phi(x) + D^2\psi(T(x)) = -(D^2\varphi(x))^2(I - D^2\varphi(x))^{-1}$ and use the previous computations. Consider the inverse map T^{-1}. We have $T^{-1}(y) = y - \nabla\psi(y)$ and, passing to the Jacobian, for $x = T^{-1}(y)$, we have $(DT(x))^{-1} = I - D^2\psi(y)$. This means $D^2\psi(y) = I - (I - D^2\varphi(x))^{-1}$. Setting $A = D^2\varphi(x)$ we now compute

$$D^2\phi(x) + D^2\psi(T(x)) = A + I - (I - A)^{-1} = -A^2(I - A)^{-1}.$$

□

By approximating $H(z) = |z|$ with $H(z) = \sqrt{\varepsilon^2 + |z|^2}$, Lemma 2 has the following important corollary, where we use the convention $\frac{z}{|z|} = 0$ for $z = 0$.

Corollary 1 *Let $\Omega \subset \mathbb{R}^d$ be a given bounded convex set and $\varrho, g \in W^{1,1}(\Omega)$ be two probability densities. Then the following inequality holds*

$$\int_\Omega \left(\nabla\varrho \cdot \frac{\nabla\varphi}{|\nabla\varphi|} + \nabla g \cdot \frac{\nabla\psi}{|\nabla\psi|}\right) \mathrm{d}x \geq 0, \tag{34}$$

where φ and ψ are the corresponding Kantorovich potentials.

This allows to obtain a first estimate on the evolution of the BV norm across the JKO scheme.

Proposition 4 *Consider $F(\varrho) = \int f(\varrho(x))\mathrm{d}x$ and the sequence ϱ^τ_k defined by the W_p-JKO scheme for the functional F. Then we have*

$$\int |\nabla\varrho^\tau_k|\mathrm{d}x \geq \int |\nabla\varrho^\tau_{k+1}|\mathrm{d}x,$$

i.e. the BV norm decreases along the JKO scheme for F.

Proof. We apply Lemma 1 and obtain

$$\int_\Omega \left(\nabla\varrho^\tau_{k+1} \cdot \frac{\nabla\varphi}{|\nabla\varphi|} + \nabla\varrho^\tau_k \cdot \frac{\nabla\psi}{|\nabla\psi|}\right) \mathrm{d}x \geq 0.$$

Yet, from $\varphi = -f'(\varrho^\tau_{k+1})$ and $f'' > 0$, we get that $\nabla\varphi$ and $\nabla\varrho^\tau_{k+1}$ are vectors with opposite directions. Hence we have

$$\int_\Omega |\nabla\varrho^\tau_{k+1}| \leq \int_\Omega \nabla\varrho^\tau_k \cdot \frac{\nabla\psi}{|\nabla\psi|}\,\mathrm{d}x \leq \int_\Omega |\nabla\varrho^\tau_k|,$$

which is the desired estimate. □

A corollary of this result about the JKO scheme is the following result on the corresponding PDE.

Corollary 2 *Consider the parabolic PDE $\partial_t\varrho = \Delta_q(\phi(\varrho))$ for an arbitrary exponent $q \in (1,\infty)$ and an arbitrary increasing nonlinearity $\phi : \mathbb{R}_+ \to \mathbb{R}$, with no-flux boundary condition on a convex domain $\Omega \subset \mathbb{R}^d$. Then, if $\varrho_0 \in BV(\Omega)$ there* *exists a solution ϱ_t such that $||\varrho_t||_{BV} \leq ||\varrho_0||_{BV}$. Moreover, if such an equation admits uniqueness of the solution for the Cauchy problem, then $t \mapsto ||\varrho_t||_{BV}$ is nonincreasing in time.*

Proof. It is enough to choose a convex function f such that $sf''(s)^{q-1} = \phi'(s)^{q-1}$ and consider this equation as a W_p-gradient flow of $\mathscr{F}$. Then, Proposition 4 provides this uniform bound on the JKO scheme. In the limit $\tau \longrightarrow 0$ this bound is preserved by lower semicontinuity. In case of uniqueness, it is possible to prove for any $t > t_0$ the inequality $||\varrho_t||_{BV} \leq ||\varrho_{t_0}||_{BV}$ by starting the JKO scheme from ϱ_{t_0}. □

5.2 Fisher information estimates

We focus here on the case of the functional $\mathscr{E}(\varrho) = \int f(\varrho(x))\mathrm{d}x + \int V \mathrm{d}\varrho$ for $f(s) + s\log s$ (which is also a relative entropy w.r.t. the density e^{-V}), and consider transport costs of the form $c(x,y) = \tau h(\frac{x-y}{\tau})$ for a strictly convex and radial function h. We now look at the operator $\mathbf{\Pi}$ defined via

$$\mathbf{\Pi}[g] := \operatorname{argmin} \mathcal{T}_c(\varrho, g) + \mathscr{E}(\varrho). \tag{35}$$

The optimality conditions characterizing the optimal ϱ are $\log\varrho + V + \varphi = c$ and it is possible to see that ϱ depends on g in a 1-homogeneous and monotone way. Moreover, $\mathbf{\Pi}$ is also an L^1 contraction (see Theorem 1.3 from [36]).

We define the following Fisher information-type functional:

$$\mathscr{I}_{V,H}(\varrho) := \int_\Omega H(\nabla(\log\varrho + V))\,\mathrm{d}\varrho.$$

Using the five gradients inequality, one can prove the following :

Lemma 3 *Let H be a radially symmetric convex function, $g \in W^{1,1}(\Omega)$ be non negative and take $\varrho = \mathbf{\Pi}[g]$. Then we have*

$$\mathscr{I}_{V,H}(g) \geq \mathscr{I}_{V,H}(\varrho).$$

Proof. We follow the proof of Proposition 5.1 from [27]. We have, from the fact that H is convex,

$$\int_\Omega H(\nabla(\log g + V))\,\mathrm{d}g \geq \int_\Omega H(\nabla\psi)\,\mathrm{d}g + \int_\Omega \nabla H(\nabla\psi)\cdot(\nabla\log g + \nabla V - \nabla\psi))\,\mathrm{d}g.$$

Using $g = T_\#\varrho$ and $\nabla\psi \circ T = -\nabla\varphi$, together with the optimality condition of the optimization problem in (35) $\nabla\varphi = -\nabla(\log\varrho + V)$, we have

$$\int_\Omega H(\nabla\psi)\,\mathrm{d}g = \int_\Omega H(\nabla\psi\circ T)\,\mathrm{d}\varrho = \int_\Omega H(-\nabla\varphi)\,\mathrm{d}\varrho = \mathscr{I}_{V,H}(\varrho),$$

and

$$\int_\Omega \nabla H(\nabla\psi)\cdot\nabla\psi\,\mathrm{d}g = -\int_\Omega \nabla H(-\nabla\varphi)\cdot\nabla\varphi\,\mathrm{d}\varrho = \int_\Omega \nabla H(-\nabla\varphi)\cdot(\nabla\varrho+\varrho\nabla V)\,\mathrm{d}x.$$

Using the five-gradients inequality, we have

$$\int_\Omega \nabla H(\nabla\psi)\cdot(\nabla\log g)\,\mathrm{d}g + \int_\Omega \nabla H(\nabla\varphi)\cdot\nabla\varrho\,\mathrm{d}x$$
$$= \int_\Omega \nabla H(\nabla\psi)\cdot\nabla g\,\mathrm{d}x + \int_\Omega \nabla H(\nabla\varphi)\cdot\nabla\varrho\,\mathrm{d}x \geq 0,$$

so that we obtain

$$\mathcal{I}_{V,H}(g) \geq \mathcal{I}_{V,H}(\varrho) + \int \nabla H(\nabla\psi)\cdot\nabla V\mathrm{d}g - \int \nabla H(\nabla\varphi)\cdot\nabla V\mathrm{d}\varrho.$$

We can then use $\int \nabla H(\nabla\psi)\cdot\nabla V\mathrm{d}g = \int \nabla H(-\nabla\varphi)\cdot\nabla V\circ T\mathrm{d}\varrho$ to obtain

$$\int \nabla H(\nabla\psi)\cdot\nabla V\mathrm{d}g - \int \nabla H(\nabla\varphi)\cdot\nabla V\mathrm{d}\varrho = \int \nabla H(-\nabla\varphi)\cdot(\nabla V\circ T - \nabla V)\mathrm{d}\varrho.$$

Using the fact that H is radial and that $T(x)-x$ is a positive scalar multiple of $-\nabla\varphi$ we obtain, thanks to the convexity of V (which implies $(y-x)\cdot(\nabla V(y)-\nabla V(x))\geq 0$), the positivity of this last integral, which gives the claim. □

When $V=0$ and $H(z)=|z|^2$, the quantity $\mathcal{I}_{V,H}$ is the classical Fisher information. We are saying that this generalized Fisher information decreases along steps of the JKO scheme. Exactly as in Corollary 2 this can be transformed into a statement about the corresponding continous-in-time PDE, but only for 1-homogeneous equations of the form

$$\partial_t\varrho = \nabla\cdot(\varrho\nabla h^*(\nabla(\log\varrho+V))). \tag{36}$$

Such a result would be a consequence of the following statement about the JKO scheme (adapted from [27] and generalized in [21]):

Theorem 6 *Let $\varrho_0\in L^1(\Omega)$. We assume V to be convex and Lipschitz continous. We have*

$$\int_\Omega H(\nabla(\log\varrho_0+V))\,\mathrm{d}\varrho_0 \geq \int_\Omega H(\nabla(\log\varrho_k^\tau+V))\,\mathrm{d}\varrho_k^\tau$$

for all τ, k, where H is a radially symmetric and convex function. In particular, the following estimates are uniform in τ and k:

1. *if $\log(\varrho_0)+V$ is L-Lipschitz, then $\log(\varrho_k^\tau)+V$ is L-Lipschitz;*
2. *for $p>1$, if $\varrho_0^{1/p}\in W^{1,p}(\Omega)$, then $(\varrho_k^\tau)^{1/p}$ is bounded in $W^{1,p}(\Omega)$;*
3. *if $\varrho_0\in BV(\Omega)$, then ϱ_k^τ is bounded in $BV(\Omega)$;*

4. *if* $\varrho_0 \in W^{1,1}(\Omega)$*, all* ϱ_k^τ *belong to a weakly compact subset of* $W^{1,1}(\Omega)$*.*

Proof. We only give the main ideas. The first statement is a direct consequence of Lemma 3, iterated along the steps of the scheme. For 1., we pick H to be the indicator of the centered ball of radius L (by indicator we mean here the function which is 0 on the ball and $+\infty$ outside it). For 2. we take $H(z) = |z|^p$ with $p > 1$, and 3. is covered by the case $p = 1$. For 4., one can use the Dunford-Pettis theorem so find a superlinear convex function H such that $H(\nabla \log(\varrho_0) + \nabla V) \in L^1(\varrho_0)$, which decreases at each step, thus providing the weak compactness. In all cases, the boundedness of ∇V allows to remove it from the integral. □

The above theorem includes the fact that Lipschitz bounds on the logarithm are preserved along the iterations of this JKO scheme. We want now to extend this to other moduli of continuity. We will use the following abstract fact. This statement (from [21], which we did not find stated in this very version elsewhere in the literature, seems interesting in itself.

Theorem 7 *Let* $\pi : L^\infty(\Omega) \to L^\infty(\Omega)$ *be an operator such that*

1. *For all* $u, v \in L^\infty(\Omega)$*, if* $u \geq v$*, then* $\pi(u) \geq \pi(v)$*.*
2. *For all* $\lambda \in \mathbb{R}$ *and* $u \in L^\infty(\Omega)$*,* $\pi(\lambda + u) = \lambda + \pi(u)$*.*
3. *For all* $u \in L^\infty(\Omega)$*, if* u *is* k*-Lipschitz, then* $\pi(u)$ *is also* k*-Lipschitz.*

Then, if u *admits a concave modulus of continuity* ω*, then* $\pi(u)$ *admits the same modulus of continuity.*

Proof. If u admits ω as a modulus of continuity, we start by approximating it with L-Lipschitz functions : for $x \in \Omega$, set

$$u_L(x) = \inf_{y \in \Omega} L|x - y| + u(y),$$

so that u_L is L-Lipschitz. Of course u_L satisfies $u_L \leq u$ (by taking $y = x$). Furthermore, from the inequality

$$u(x) - u(y) \leq \omega(|x - y|),$$

we deduce that we have

$$L|x - y| + u(y) \geq L|x - y| + u(x) - \omega(|x - y|),$$

so that, passing to the inf in y, we obtain

$$u_L(x) \geq u(x) + \alpha(L),$$

where $\alpha(L) = \inf_{\mathrm{diam}(\Omega) > r > 0} Lr - \omega(r) \leq 0$. We can therefore conclude that we have

$$u \geq u_L \geq u + \alpha(L).$$

Applying π and using its properties 1 and 2 we obtain

$$\pi(u) \geq \pi(u_L) \geq \pi(u) + \alpha(L).$$

For $x, y \in \Omega$, since $\pi(u_L)$ is also L-Lipschitz, we can write

$$\pi(u)(x) - \pi(u)(y) \leq \pi(u_L)(x) - \pi(u_L)(y) - \alpha(L) \leq L|x-y| - \alpha(L).$$

Since ω is concave, we choose $L = \omega'(|x-y|)$ (note that ω is also non-decreasing, so we can take $L \geq 0$; in case ω is not differentiable one should use the super-differential of ω, which is non-empty because ω is finite on $\mathbb{R}_+$). This implies that $r \mapsto Lr - \omega(r)$ (which is a convex function) is minimized at $r = |x-y|$. Then we obtain for this L the equality $\alpha(L) = L|x-y| - \omega(|x-y|)$, and thus

$$\pi(u)(x) - \pi(u)(y) \leq \omega(|x-y|),$$

so that $\pi(u)$ admits ω as a modulus of continuity. □

Using the above theorem, we can then prove the following:

Theorem 8 *Let $\varrho_0 \in L^1(\Omega)$ be such that $\log(\varrho_0) + C$ admits ω as a modulus of continuity, then for all k and $\tau > 0$, $\log(\varrho_k^\tau) + V$ also admits ω as a modulus of continuity. As a consequence, this also holds for the solution of* (36).

Proof. One only has to apply Theorem 7 where π is defined as follows: given u, solve one step of the JKO scheme with $g = \exp(u - V)$, call ϱ the solution, and then take $\pi(u) = \log \varrho + V$. □

Among particular examples of (36) we wite not only equations of $q-$Laplacian type $\partial_t \varrho = \Delta_q(\varrho^{p-1})$ but also the so-called *relativistic heat equations* studied for instance in [44]. in such a paper the authors generalize the construction from [1] to "relativistic" cost functions, meaning cost functions where h is convex and only finite on a ball. The main equation of interest entering in this framework is the so called Relativistic Heat Equation

$$\partial_t \varrho = \nabla \cdot \left(\varrho \, \frac{\nabla \varrho}{\sqrt{\varrho^2 + |\nabla \varrho|^2}} \right),$$

where $h(z) = 1 - \sqrt{1 - |z|^2}$.

We finish this section with some improvements of Lemma 3 in the case $H(z) = |z|^2$

Lemma 4 *Let $g \in W^{1,1}(\Omega)$ be non negative and denote by ϱ the solution of* (35). *Suppose $D^2 V \geq \lambda I$. Then we have*

$$\mathscr{I}_{V,H}(g) \geq \mathscr{I}_{V,H}(\varrho) + 2\lambda\tau \int_\Omega |\nabla(\log \varrho + V)|^2 \, \mathrm{d}\varrho$$
$$+ 2\tau \int_\Omega \mathrm{tr}\left[(D^2(\log \varrho + V))^2 (I - A)^{-1} \right] \mathrm{d}\varrho,$$

where the matrix A is $D^2\varphi$, φ being the Kantorovich potential for the quadratic cost from ϱ to g, so that we also have $A = -\tau D^2(\log\varrho + V)$.

Proof. The improvements w.r.t. Lemma 3 are obtained in two different ways. First, we estimate the term $\int \nabla H(-\nabla\varphi)\cdot(\nabla V\circ T - \nabla V)\mathrm{d}\varrho$ by using $(y-x)\cdot(\nabla V(y) - \nabla V(x)) \geq \lambda|x-y|^2$ instead of just using that that this scalar product is non-negative. Second, we use (33) in order to obtain the last line. □

6 Application to the log-Sobolev inequality

The goal of this section is to provide a proof of the classical Sobolev Inequality based on the use of the JKO scheme. Such an inequality reads

$$\int_\Omega f^2\log(f^2)\,\mathrm{d}\mu \leq \frac{1}{\lambda}\int_\Omega |\nabla f|^2\,\mathrm{d}\mu, \tag{37}$$

where we take $\mu = e^{-V}$ for some convex function $V$with $D^2V \geq \lambda I$, and assume that

$$\int_\Omega f^2 e^{-V}\,\mathrm{d}x = 1.$$

Setting $g = f^2$, this is equivalent to

$$\int_\Omega g\log g\,\mathrm{d}\mu \leq \frac{1}{2\lambda}\int_\Omega |\nabla\log g|^2\, g\,\mathrm{d}\mu.$$

Now we set $g = \varrho e^V$, so that ϱ is a probability density on Ω, and the inequality turns into

$$\int_\Omega \varrho e^V\,(\log\varrho + V)\,e^{-V}\,\mathrm{d}x = \mathscr{E}(\varrho) \leq \frac{1}{2\lambda}\mathscr{I}_{V,H}(\varrho).$$

This inequality can indeed be proved by using the so-called Bakry-Emery technique (see [6]), which consists in comparing the two quantities $\mathscr{E}$ and $\mathscr{I}_{V,H}$ (that we will write $\mathscr{I}$ for short) along the Fokker-Planck equation $\partial_t\varrho_t = \Delta\varrho_t + \nabla\cdot(\varrho\nabla V)$. Since the steady state, to which ϱ_t converges as $t\to\infty$ is exactly $\mu = e^{-V}$, and is such that $\mathscr{E}(\mu) = \mathscr{I}(\mu) = 0$, it is enough to compare the time derivative of $\mathscr{E}(\varrho_t)$ and $\mathscr{I}(\varrho_t)$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{E}(\varrho_t) = -\int_\Omega |\nabla\,(\log\varrho_t + V)|^2\,\mathrm{d}\varrho_t = -\mathscr{I}(\varrho_t).$$

Then, it is possible to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{I}(\varrho_t) \leq -2\lambda\mathscr{I}(\varrho_t).$$

A precise computation leading to this inequality will be done in Section 7, but one can also see it as a consequence of Corollary 4. Note that here the second quantity that we differentiate happens to be the derivative of the first, so that we finally differentiate twice $\mathscr{E}(\varrho_t)$, but this is not cruvial. The importnat point is being able to copare the derivatives of the desired functionals.

Thus, we have proven

$$\frac{\mathrm{d}}{\mathrm{d}t}(2\lambda\mathscr{E} - \mathscr{I})(\varrho_t) \geq 0,$$

which implies $(2\lambda\mathscr{E} - \mathscr{I})(\varrho_0) \leq \lim_{t\to\infty}(2\lambda\mathscr{E} - \mathscr{I})(\varrho_t) = 0$, which is the desired inequality (of course this computation requires to prove that the convergence as $t \to \infty$ implies the convergence of $\mathscr{E}$ and $\mathscr{I}$).

We now want to see if the same result can be obtained by looking at the evolution of $\mathscr{E}$ and $\mathscr{I}$ along the steps of the JKO scheme. This is so far just an interesting exercise, as it only re-proves a well-known result, but the technique can then be applied to other inequalities for which a continuous flow proof is not available. In particular, a work is in progress [19] about the case where V is convex but not uniformly convex (for instance $V(x) = |x|^4$).

Again, we need to prove that $2\lambda\mathscr{E} - \mathscr{I}$ increases from one step to the other. This means thate fix a measure g and call ϱ the optimizer of one JKO step starting from g. In order to mimick what is done in the continuous proof we would need to have

$$\mathscr{I}(\varrho) \leq \mathscr{I}(g) - 2\lambda\tau\mathscr{I}(\varrho), \quad \text{and} \quad \mathscr{E}(\varrho) \geq \mathscr{E}(g) - \tau\mathscr{I}(\varrho).$$

The first of these inequalities is exactly contained in Lemma 4. As fir the second, since it amounts to differentiare a zero-order quantity which is indeed geodesically convex (since we assume V to be convex), the good tool is the flow interchange technique from Proposition 2. Yet, unfortunately, what could be proven in this way is the opposite inequality

$$\mathscr{E}(\varrho) \leq \mathscr{E}(g) - \tau\mathscr{I}(\varrho).$$

It is then necessary to work differently, and what can be proven is the following.

Lemma 5 *If $\varrho = \mathbf{\Pi}[g]$ (using the functional $\mathscr{E}$ and the quadratic transport cost), then we have*

$$\mathscr{E}(\varrho) \geq \mathscr{E}(g) - \frac{\tau}{2}\mathscr{I}(\varrho) - \frac{\tau}{2}\mathscr{I}(g).$$

Proof. We use Proposition 2 with $G = \mathscr{E}$, but this time inverting the roles of g and ϱ. We obtain, exploiting the precise choice of the quadratic cost, the following inequality

$$\mathscr{E}(\varrho) \geq \mathscr{E}(g) - \tau\int_\Omega \nabla(\log g + V)\cdot\nabla\psi \,\mathrm{d}g,$$

where ψ is the Kantorovich potential from g to ϱ (and φ is the one from ϱ to g, which appears in the optimality conditions of the JKO step). We then apply Young's inequality and obtain

$$\mathscr{E}(\varrho) \geq \mathscr{E}(g) - \frac{\tau}{2}\int_\Omega |\nabla(\log g + V)|^2 \mathrm{d}g - \frac{\tau}{2}\int_\Omega |\nabla\psi|^2 \mathrm{d}g.$$

We now use $\int_\Omega |\nabla(\log g + V)|^2 \mathrm{d}g = \mathscr{I}(g)$ and

$$\int_\Omega |\nabla\psi|^2 \mathrm{d}g = \int_\Omega |\nabla\psi \circ T|^2 \mathrm{d}\varrho = \int_\Omega |\nabla\varphi|^2 \mathrm{d}\varrho = \int_\Omega |\nabla(\log\varrho + V)|^2 \mathrm{d}\varrho = \mathscr{I}(\varrho),$$

where we used the optimality conditions on ϱ. This proves the claim. □

We then obtain the following.

Proposition 5 *Given an arbitrary ϱ_0 and starting the JKO scheme from it we have, for any $N \geq 0$*

$$2\lambda\mathscr{E}(\varrho_0) - \mathscr{I}(\varrho_0) \leq 2\lambda\mathscr{E}(\varrho_N^\tau) - \mathscr{I}(\varrho_N^\tau) + \lambda\tau(\mathscr{I}(\varrho_0) - \mathscr{I}(\varrho_N^\tau)).$$

This implies

$$2\lambda\mathscr{E}(\varrho_0) - \mathscr{I}(\varrho_0) \leq 0.$$

Proof. We apply the previous lemma to $g = \varrho_k^\tau$ and $\varrho = \varrho_{k+1}^\tau$ for $k = 0, 1, \ldots, N-1$ and sum the inequalities, thus obtaining

$$\mathscr{E}(\varrho_0) \leq \mathscr{E}(\varrho_N^\tau) + \tau\sum_{k=1}^{N-1}\mathscr{I}(\varrho_k^\tau) + \frac{\tau}{2}(\mathscr{I}(\varrho_0) + \mathscr{I}(\varrho_N^\tau)).$$

Multiplying by 2λ and subtracting the inequality

$$\mathscr{I}(\varrho_0) \geq \mathscr{I}(\varrho_N^\tau) + 2\lambda\tau\sum_{k=1}^{N}\mathscr{I}(\varrho_k^\tau)$$

we obtain the desired inequality.

It is then enough to take the limit $N \to \infty$ in order to obtain

$$2\lambda\mathscr{E}(\varrho_0) - \mathscr{I}(\varrho_0) \leq \lambda\tau\mathscr{I}(\varrho_0).$$

Note that this is even simpler to justify than in the continuous setting, as the convergence $\varrho_N^\tau \rightharpoonup \mu$ can be obtained by studying the fixed points of the JKO operator, and when $\tau > 0$ is fixed the convergence is automatically strong enough to pass to the limit both $\mathscr{E}$ and $\mathscr{I}$ as one can obtain second-order bounds of the order of τ^{-1}. Then, we take the limit as $\tau \to 0$, assuming $\mathscr{I}(\varrho_0) < +\infty$ (otherwise there is nothing to prove). □

7 Strong L^2H^2 convergence of the JKO scheme

The metric interpretation of the JKO scheme, according to the theory of [4], provides convergence as $\tau \to 0$ in the same distance which is used in the JKO scheme. This means that, in the case of the linear Fokker-Planck equation, one obtains uniform (in time) convergence in W_2, which means weak-* convergence (in space) of ϱ^τ to ϱ_t. In the original paper about the JKO scheme for the Fokker-Planck equation, [37], a slightly different convergence is proven: it is proven that ϱ^τ strongly converges in $L^1([0,T]\times\Omega)$.

In this section we use the notions developed so far in order to prove that, under some conditions on the initial datum, the convergence in the case of the linear Fokker-Planck equation is actually much stronger, and holds in $L^2([0,T];H^2(\Omega))$. We will consider the case of a convex bounded domain $\Omega\subset\mathbb{R}^d$ whose boundary $\partial\Omega$ is smooth enough. We denote by $\mathbf{n}$ the exterior unit normal vector of the boundary $\partial\Omega$. We consider the Cauchy problem for the Fokker-Planck equation with no-flux boundary condition, i.e.,

$$\begin{cases}\partial_t\varrho(t,x)=\Delta\varrho(t,x)+\mathrm{div}(\varrho(t,x)\nabla V(x)), & (t,x)\in(0,T]\times\Omega,\\ \nabla\varrho(t,x)\cdot\mathbf{n}(x)+\varrho(t,x)\nabla V(x)\cdot\mathbf{n}(x)=0, & (t,x)\in[0,T]\times\partial\Omega,\\ \varrho(0,x)=\varrho_0(x), & x\in\Omega,\end{cases}\tag{38}$$

where $\varrho_0\in\mathcal{P}(\Omega)\cap L^1_+(\Omega)$.

We refer to classical texts on parabolic differential equations (see [39], [40]) for the existence, the uniqueness, and the regularity of the solution.

We now compute the derivative of the classical Fisher information along the flow. For shortness, we write u_ϱ for u_ϱ:

Lemma 6 *If Ω and V are smooth enough and ϱ is the solution of (38), then for $t>0$ we have*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{I}(\varrho_t)=-2\int_\Omega|D^2u_\varrho|^2\varrho\,dx-2\int_\Omega(\nabla u_\varrho)^T\cdot D^2V\cdot\nabla u_\varrho\varrho\,dx$$
$$+\int_{\partial\Omega}(\nabla u_\varrho)^T\cdot D^2u_\varrho\cdot\mathbf{n}\varrho\,d\mathcal{H}^{d-1}.$$

The last term in the previous formula can be re-written using the following lemma.

Lemma 7 *Suppose $\Omega=\{h<0\}$ for a smooth function $h:\mathbb{R}^d\to\mathbb{R}$ with $\nabla h\neq 0$ on $\{h=0\}$, so that the exterior normal vector at $x\in\partial\Omega$ is given by $\mathbf{n}(x)=\nabla h(x)/|\nabla h(x)|$. Let $\mathbf{v}:\Omega\to\mathbb{R}^d$ be a smooth vector field such that $\mathbf{v}\cdot\mathbf{n}=0$ on $\partial\Omega$. Then we have the following equality for every $x\in\partial\Omega$*

$$\mathbf{v}(x)^T\cdot Dv(x)\cdot\mathbf{n}(x)=-\frac{\mathbf{v}(x)^T\cdot D^2h(x)\cdot\mathbf{v}(x)}{|\nabla h(x)|}.$$

Proof. Given $x \in \partial\Omega$, we consider a smooth curve $\gamma : (-t_0, t_0) \to \partial\Omega$ with $\gamma(0) = x$ and write the equality $\mathbf{v}(\gamma(t)) \cdot \nabla h(\gamma(t)) = 0$ for every t. Differentiating w.r.t. t we obtain

$$\gamma'(t)^T \cdot D\mathbf{v}(\gamma(t)) \cdot \nabla h(\gamma(t)) + \mathbf{v}(\gamma(t))^T \cdot D^2 h(\gamma(t)) \cdot \gamma'(t) = 0.$$

We can take $t = 0$ and choose a curve with $\gamma'(0) = \mathbf{v}(x)$ since $\mathbf{v}$ is tangent to the surface $\partial\Omega$, thus obtaining

$$\mathbf{v}(x)^T \cdot D\mathbf{v}(x) \cdot \nabla h(x) = -\mathbf{v}(x)^T \cdot D^2 h(x) \cdot \mathbf{v}(x).$$

It is then enough to divide by $|\nabla h(x)|$ in order to get the claim. □

We then obtain the following formula.

Corollary 3 *Let $\Omega = \{h < 0\}$ be a bounded domain defined as the negativity set of a convex function $h \in C^2$ with $\nabla h \neq 0$ on $\{h = 0\}$, and $V \in C^2(\bar{\Omega})$. Given a strictly positive $\varrho_0 \in H^1(\Omega)$ initial datum, let ϱ be the solution of (38). We then have*

$$\begin{aligned}\mathscr{F}(\varrho_T) - \mathscr{F}(\varrho_0) = &-\int_0^T \mathrm{d}t \int_\Omega |D^2 u_\varrho|^2 \varrho \,\mathrm{d}x \\ &-\int_0^T \mathrm{d}t \int_\Omega (\nabla u_\varrho)^T \cdot D^2 V \cdot \nabla u_\varrho \varrho \,\mathrm{d}x \\ &-\int_0^T \mathrm{d}t \int_{\partial\Omega} (\nabla u_\varrho)^T \cdot D^2 h \cdot (\nabla u_\varrho) \varrho \,\mathrm{d}\mathcal{H}^{d-1}.\end{aligned}$$

We now proceed to some uniform estimates on the minimizers of the JKO scheme and on the corresponding potentials. Note that here, in order to better see the dependance on τ of these terms, we write φ for the Kantorovich potentials w.r.t. the quadratic cost *wihout the coefficient τ*.

Proposition 6 *Suppose Ω is a bounded, smooth, and uniformly convex domain, and $V \in C^2(\bar{\Omega})$. Let (ϱ_k^τ) be the sequence obtained in the JKO scheme for the Fokker-Planck equation, (φ_k, ψ_k) denote the pair of Kantorovich potentials in the transport from ϱ_{k+1}^τ to ϱ_k^τ and T_k the corresponding optimal map, i.e. $T_k(x) = x - \nabla\varphi_k(x)$. Suppose that ϱ_0 is bounded from below and above by positive constants and denote by a, b two constants such that $a \leq \log \varrho_0 + V \leq b$. Suppose moreover that we have $\varrho_0 \in C^{0,\alpha} \cap H^1$. Then we have:*

1. *For each k we have $a \leq \log(\varrho_k^\tau) + V \leq b$. In particular, all ϱ_k^τ are bounded from below and above by some uniform positive constants.*
2. *All the potentials φ_k satisfy $||id - T_k||_{L^\infty} = ||\nabla\varphi_k||_{L^\infty} \leq C\tau^{1/(d+2)}$.*
3. *If τ is small enough (depending on V and p), then the $C^{0,\alpha}$ seminorm of $\log(\varrho_k^\tau) + V$ is bounded in terms f that of $\log(\varrho_0) + V$ and the densities ϱ_k^τ are bounded in $C^{0,\alpha}$.*
4. *All the potentials φ_k belong to $C^{2+\alpha}(\bar{\Omega})$ and $||\varphi_k||_{C^{2+\alpha}(\bar{\Omega})}$ is bounded by a uniform constant.*

5. *The potentials φ_k also satisfy $||D^2\varphi_k||_{L^\infty} \leq C\tau^\beta$ for a certain exponent $\beta > 0$.*

Proof. 1. The uniform estimates on ϱ^τ can be proven as in Lemma 2.4 of [35].

2. Whenever μ, ν are two measures in a convex domain $\Omega \subset \mathbb{R}^d$ and T is the corresponding optimal map sending μ onto ν, if the density of μ is bounded from below by a constant $c_0 > 0$, then the following remarkable estimate is proven in [7] :
$$||T - id||_{L^\infty} \leq C(d, c_0) W_2(\mu, \nu)^{2/(d+2)}.$$
If we combine this with
$$W_2^2(\varrho^\tau_{k+1}, \varrho^\tau_k) \leq 2\tau \left(F(\varrho^\tau_k) - F(\varrho^\tau_{k+1})\right) \leq C\tau.$$
and the fact that the density ϱ^τ_{k+1} is bounded from below by a universal constant, we then have
$$||\nabla\varphi_k||^{2+d}_{L^\infty(\Omega)} = ||id - T_k||^{2+d}_{L^\infty(\Omega)} \leq C W_2^2(\varrho^\tau_{k+1}, \varrho^\tau_k) \leq C\tau. \tag{39}$$

3. If V is convex we already saw in theorems 6 and 8 that the modulus of continuity of $\log \varrho + V$ are preserved along iterations. If $D^V \geq \lambda I$ but λ is negative this requires a slightly differnet argument. A proof the Lipschitz bound is contained in [31] and then we need to adapt the proof of Theroem 7 to the case where the Lipschitz constant is not preserved by π but increases by a quantified factor of order $1 + \tau$.
4. The bound on $||\varphi_k||_{C^{2+\alpha}(\bar\Omega)}$ is a consequence of Caffarelli's regularity theory for the Monge-Ampère equation (see [12]-[17]), once we have proven that the densities are uniformly bounded from above, from below, and in $C^{0,\alpha}$, when the domain is uniformly convex and C^2.
5. If we apply standard interpolation arguments to $u = \nabla\varphi$ we get
$$||\nabla\varphi_k||_{C^1(\bar\Omega)} \leq c||\nabla\varphi_k||^{\frac{1}{1+\alpha}}_{C^{1+\alpha}(\bar\Omega)} ||\nabla\varphi_k||^{\frac{\alpha}{1+\alpha}}_{L^\infty(\bar\Omega)}.$$
Using (39) and the uniform bounds on $||\varphi_k||_{C^{2+\alpha}(\bar\Omega)}$ we obtain the claim with $\beta = \frac{\alpha}{(1+\alpha)(2+d)}$.

□

We now use the previous results to prove the following estimates, where $\varepsilon(\tau)$ denotes any quantity which depends only on $\bar\Omega, \varrho_0, V$ and τ such that $\varepsilon(\tau) \to 0$ as $\tau \to 0$.

Lemma 8 *Under the same assumptions as Proposition 6, if Ω is of the form $\{h < 0\}$ for a convex and smooth function h with* $\min h < 0$*, then we have*

$$\mathcal{J}(\varrho_0) \geq \mathcal{J}(\varrho_T^\tau) + \int_0^T \int_\Omega |D^2 u_{\varrho_t^\tau}|^2 \varrho_t^\tau(x)\,\mathrm{d}x\mathrm{d}t + \int_0^T \int_\Omega (\nabla u_{\varrho_t^\tau})^T \cdot D^2 V(x) \cdot \nabla u_{\varrho_t^\tau} \varrho_t^\tau(x)\,\mathrm{d}x\mathrm{d}t + \int_0^T \int_{\partial\Omega} \frac{\varrho_t^\tau(x)}{|\nabla h(x)|} (\nabla u_{\varrho_t^\tau})^T \cdot D^2 h(x) \cdot (\nabla u_{\varrho_t^\tau}) \mathrm{d}\mathcal{H}^{d-1}\mathrm{d}t + \varepsilon(\tau). \quad (40)$$

Proof. The estimate can be obtained from Lemma 4 after improving some computatons. First, we need to keep the Hessian of V instead of just using $D^V \geq \lambda$. The C^2 regularity of V is needed to use an inequality of the form

$$(x-y)\cdot(\nabla V(x) - \nabla V(y)) \geq (x-y)^T D^2 V(x)(x-y) + o(|x-y|^2).$$

Then, we apply this to $y = T(x)$ and use point 2. in Proposition 6 in order to estimate the error term with $\varepsilon(\tau)$.

We now look at the boundary integrals. This requires to use the boundary terms that we just bounded from below by 0 in the proof of the five-gradients-inequality.

Details for both arguments are presented in [52]. □

In order to conclude we need the following lemma.

Lemma 9 *Let Ω be a bounded Lipschitz domain of $\mathbb{R}^d$, $\{u_k\}_{k=1}^\infty$ be a sequence in $L^2([0,T]; H^2(\Omega))$ and bounded in $L^\infty([0,T]\times\Omega)$. If $u_k \to u$ strongly in $L^2([0,T]; H^2(\Omega))$, then, for any $f \in C^2(\mathbb{R})$, $\{f(u_k)\}_{k=1}^\infty$ converges to $f(u)$ in $L^2([0,T]; H^2(\Omega))$.*

Proof. Using our assumptions and the Gagliardo-Nirenberg inequality (see [11, section 9])

$$||\nabla v||_{L^4(\Omega)}^4 \leq C||v||_{H^2(\Omega)}^2 ||v||_{L^\infty(\Omega)}^2 \quad (41)$$

applied to $v = u_k - u$ we obtain $u_k \to u$ in $L^4([0,T]; W^{1,4}(\Omega))$. A simple computation shows

$$D^2(f(u_k)) = f'(u_k) D^2 u_k + f''(u_k)\nabla u_k \otimes \nabla u_k$$

and the L^2 convergence of this matrix-valued function to $D^2(f(u)) = f'(u)D^2 u + f''(u)\nabla u \otimes \nabla u$ is due to the following facts:

- $D^2 u_k \to D^2 u$ in $L^2([0,T]\times\Omega)$;
- $\nabla u_k \otimes \nabla u_k \to \nabla u \otimes \nabla u$ in $L^2([0,T]\times\Omega)$;
- both $f''(u_k)$ and $f'(u_k)$ converge a.e. (to $f''(u)$ and $f'(u)$, respectively) as a consequence of the convergence of u_k to u; moreover, these terms are bounded in L^∞ as a consequence of the regularity of f and of the L^∞ bound on u_k. □

We are now ready to prove the main theorem of this section.

Theorem 9 *Suppose $0 < T < +\infty$, Ω is a bounded, smooth, and convex domain. Let $V \in C^2(\bar{\Omega})$, $\varrho_0 \in H^1(\Omega) \cap C^{0,\alpha}(\Omega)$ be an initial datum bounded away from zero and infinity and ϱ be the solution of the Fokker-Planck equation (38). Then, $\varrho_t^\tau \to \varrho$ strongly in $L^2([0,T]; H^2(\Omega))$ as $\tau \to 0$.*

Proof. Lemma 8 proves in particular that $\log \varrho_t^\tau + V$ is uniformly bounded in $L^2([0,T];H^2(\Omega))$ with respect to τ. Since the space $L^2([0,T];H^2(\Omega))$ is reflexive and ϱ_t^τ converges strongly in $L^2([0,T];H^1(\Omega))$ to ϱ (use the H^2 bound together with the compactness in space and Aubin-Lions lemma), we get that $\log \varrho_t^\tau + V$ converges weakly to $\log \varrho + V$ in $L^2([0,T];H^2(\Omega))$. This also implies that $\nabla \log \varrho_t^\tau + \nabla V$ converges weakly to $\nabla \log \varrho + \nabla V$ in $L^2([0,T];L^2(\partial\Omega))$. If τ tends to zero then we use the lower semicontinuity of each term on the right hand side of (40) in order to obtain

$$\int_0^T \int_\Omega |D^2 u_{\varrho_t}|^2 \varrho_t(x)\,\mathrm{d}x\mathrm{d}t \geq \limsup_{\tau\to 0} \int_0^T \int_\Omega |D^2 u_{\varrho_t^\tau}|^2 \varrho_t^\tau(x)\,\mathrm{d}x\mathrm{d}t. \tag{42}$$

This condition and the lower bound on ϱ^τ show that $D^2 \log \varrho^\tau$ is bounded in $L^2([0,T];L^2(\Omega))$. Since $L^2([0,T];L^2(\Omega))$ is reflexive and ϱ^τ converges strongly in $L^2([0,T];H^1(\Omega))$ to ϱ, then $D^2 \log \varrho^\tau$ converges weakly to $D^2 \log \varrho$ in $L^2([0,T];L^2(\Omega))$. By adding $D^2 V$ and multiplying times $\sqrt{\varrho^\tau}$, which converges a.e. to $\sqrt{\varrho}$ and is bounded by a constant, we also have weak convergence in $L^2([0,T];L^2(\Omega))$ of $\sqrt{\varrho^\tau} D^2(\log \varrho^\tau + V)$ to $\sqrt{\varrho} D^2(\log \varrho + V)$. Yet, this convergence becomes strong because of the convergence of the norm in (42). We can then multiply times $(\varrho^\tau)^{-1/2}$ and subtract $D^2 V$ and obtain strong convergence in $L^2([0,T];H^2(\Omega))$ for $\log \varrho^\tau$ to $\log \varrho$.

We then apply Lemma 9 to obtain $\varrho^\tau \to \varrho$ in $L^2([0,T];H^2(\Omega))$. □

Acknowledgments. The author acknowledges the support of European union via the ERC AdG 101054420 EYAWKAJKOS. Moreover, discussions with many collaborators, and in particular recent discussions with Thibault Caillet and Fanch Coudreuse, were crucial for the preparation of this manuscript.

References

1. M. Agueh. Existence of solutions to degenerate parabolic equations via the Monge- Kantorovich theory. *Adv. Differ. Equ.*, 10(3):309–360, 2005.
2. L. Ambrosio, Movimenti minimizzanti, *Rend. Accad. Naz. Sci. XL Mem. Mat. Sci. Fis. Natur.* 113, 191–246, 1995.
3. L. Ambrosio and P. Tilli, *Topics on analysis in metric spaces*. Oxford Lecture Series in Mathematics and its Applications (25). Oxford University Press, Oxford, 2004.
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. D. Bakry, I. Gentil, M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der mathematischen Wissenschaften (GL, volume 348), Springer, 2014.
7. G. Bouchitté, C. Jimenez, M. Rajesh, *A new L^∞ estimate in optimal mass transport,* Proc. Am. Math. Soc. 135,3525–3535 (2007).
8. J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numer. Math.*, 84 (2000), 375–393.
9. 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.

10. Y. BRENIER, Polar factorization and monotone rearrangement of vector-valued functions, *Communications on Pure and Applied Mathematics* 44, 375–417, 1991.
11. H. BREZIS, *Functional Analysis, Sobolev Spaces and Partial Differential Equations,* Springer-Verlag New York, 2011.
12. L. A. CAFFARELLI, *Interior $W^{2,p}$ estimates for solutions of the Monge-Ampère equation,* Ann. of Math. (2) 131, 1 (1990), 135-150.
13. L. A. CAFFARELLI, *A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity,* Ann. of Math. (2) 131, 1 (1990), 129-134.
14. L. A. CAFFARELLI, *Some regularity properties of solutions of Monge-Ampère equation,* Comm. Pure Appl. Math. 44, 8-9 (1991), 965-969.
15. L. A. CAFFARELLI, *Boundary regularity of maps with convex potentials,* Comm.Pure Appl. Math. 45, 9 (1992), 1141-1151.
16. L. A. CAFFARELLI, *The regularity of mappings with a convex potential,* J. Amer. Math. Soc. 5, 1 (1992), 99-104.
17. L. A. CAFFARELLI, *Boundary regularity of maps with convex potentials II,* Ann. of Math. (2) 144, 3 (1996), 453-496.
18. T. CAILLET The five gradients inequality for non quadratic costs. *Comptes Rendus Mathématique*, 361:715–721, 2023.
19. T. CAILLET, F. COUDREUSE Paper in preparation.
20. T. CAILLET. F. SANTAMBROGIO Doubly nonlinear diffusive PDEs: new existence results via generalized Wasserstein gradient flows, *SIAM J. Math. An.* Vol. 56, Iss. 6, 7043–7073 (2024)
21. T. CAILLET. F. SANTAMBROGIO Fisher information and continuity estimates for nonlinear but 1-homogeneous diffusive PDEs (via the JKO scheme), *Bull. Hellenic Math. Soc.*, to appear.
22. C. CANCÈS, D. MATTHES, F. NABET, AND E.-M. ROTT Finite elements for Wasserstein W_p gradient flows. Preprint, 2022
23. J.-A. CARRILLO AND D. SLEPČEV Example of a displacement convex functional of first order *Calculus of Variations and Partial Differential Equations* 36 (4), 547–564, 2009.
24. S. DANERI AND G. SAVARÉ, Eulerian Calculus for the Displacement Convexity in the Wasserstein Distance. *SIAM J. Math. An.* 40, 1104-1122, 2008.
25. 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.
26. G. DE PHILIPPIS, A. R. MÉSZÁROS, F. SANTAMBROGIO AND B. VELICHKOV BV estimates in optimal transportation and applications, *Arch. Rati. Mech. An.* 219 (2), 829-860, 2016.
27. S. DI MARINO, F. SANTAMBROGIO JKO estimates in linear and non-linear Fokker-Planck equations, and Keller-Segel: L^p and Sobolev bounds *Ann. Inst. H. Poincaré Anal. Non Linéaire*, (39):1485–1517, 2022
28. S. DI MARINO, S. MURRO, AND E. RADICI The five gradients inequality on differentiable manifolds. *J. Math. Pures Appl.* (9), 187:294–328, 2024.
29. R. DUCASSE, F. SANTAMBROGIO, H. YOLDAS A cross-diffusion system obtained via (convex) relaxation in the JKO scheme *Calculus of Variations and Partial Differential Equations* volume 62, Article number: 29 (2023)
30. I. EKELAND AND R. TEMAM, *Convex Analysis and Variational Problems*, Classics in Mathematics, Society for Industrial and Applied Mathematics (1999).
31. V. FERRARI, F. SANTAMBROGIO Lipschitz estimates on the JKO scheme for the Fokker–Planck equation on bounded convex domains, *Appli. Math. Letters*, Vol 112, 2021, 106806
32. W. GANGBO, An elementary proof of the polar factorization of vector-valued functions, *Arch. Rational Mech. Anal.*, 128, 381–399, 1994.
33. W. GANGBO, The Monge mass transfer problem and its applications, *Contemp. Math.*, 226, 79–104, 1999.
34. W. GANGBO AND R. MCCANN, The geometry of optimal transportation, *Acta Math.*, 177, 113–161, 1996.
35. M. IACOBELLI, F. S. PATACCHINI, F. SANTAMBROGIO Weighted ultrafast diffusion equations: from well-posedness to long-time behaviour, *Arch. Rati. Mech. An.*, 232 (3), 1165-1206, 2019.
36. M. JACOBS, I. C. KIM, AND J. TONG The L^1-contraction principle in optimal transport. *Annali Scuola Normale Superiore - Classe di Scienze*, 23(4), 1871–1919, 2022.

37. R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation, *SIAM J. Math. Anal.* 29 (1), 1–17, 1998.
38. L. Kantorovich, On the transfer of masses. *Dokl. Acad. Nauk. USSR*, 37, 7–8, 1942.
39. O.A. Ladyzhenskaja, V. A. Solonnikov, N.N. Ural'ceva,, *Linear and quasilinear equations of parabolic type,* Nauka, Moskow 1967 (Russian). English transl.: Transl. Math. Monographs, AMS, Providence (1968).
40. A. Lunardi,*Analytic semigroups and optimal regularity in parabolic problems.* Birkhäuser Verlag, Basel (1995). xviii+424 pp. 2013 reprint of the 1995 original, Modern Birkhäuser Classics. Birkhäuser/Springer Basel AG, Basel, 1995. xviii+424 pp.
41. D. Matthes, R. J. McCann, and G. Savaré. A family of nonlinear fourth order equations of gradient flow type. *Comm. Partial Differential Equations*, 34(10-12):1352– 1397, 2009.
42. R. J. McCann, Existence and uniqueness of monotone measure preserving maps, *Duke Math. J.*, 80, 309–323, 1995.
43. R. J. McCann A convexity principle for interacting gases. *Adv. Math.* 128 (1), 153–159, 1997.
44. R. J. Mc Cann and M. Puel Constructing a relativistic heat flow by transport time steps. *Annales de l?Institut Henri Poincaré C, Analyse non linéaire*, 26(6):2539–2580, 2009
45. G. Monge, Mémoire sur la théorie des déblais et des remblais, *Histoire de l'Académie Royale des Sciences de Paris*, 666–704, 1781.
46. F. Otto, The geometry of dissipative evolution equations: The porous medium equation, *Comm. Partial Differential Equations*, 26, 101–174, 2011.
47. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
48. F. Santambrogio, Flots de gradient dans les espaces métriques et leurs applications (d'après Ambrosio-Gigli-Savaré), proceedings of the *Bourbaki Seminar*, 2013 (in French).
49. F. Santambrogio *Optimal Transport for Applied Mathematicians*, Progress in Nonlinear Differential Equations and Their Applications no 87, Birkhäuser Basel (2015).
50. F. Santambrogio {Euclidean, Metric, and Wasserstein} Gradient Flows: an overview, *Bull. Math. Sci.*, Vol 7 (1), 87–154, 2017.
51. F. Santambrogio *A course in the calculus of variations*, Springer, 2023.
52. F. Santambrogio, G. Toshpulatov Strong L^2H^2 convergence of the JKO scheme for the Fokker-Planck equation. *Arch. Rati. Mech. An.* Volume 248, article number 99, (2024)
53. J. L Vázquez, *The porous medium equation. Mathematical theory*, The Clarendon Press, Oxford University Press, (2007).
54. C. Villani *Topics in Optimal Transportation*. Graduate Studies in Mathematics, AMS, 2003.
55. C. Villani, *Optimal transport: Old and New*, Springer Verlag, 2008.