GRADIENT FLOWS WITH METRIC AND DIFFERENTIABLE STRUCTURES, AND APPLICATIONS TO THE WASSERSTEIN SPACE

L.AMBROSIO, N.GIGLI, AND G.SAVARÉ

ABSTRACT. In this paper we summarize some of the main results of a forthcoming book on this topic, where we examine in detail the theory of curves of maximal slope in a general metric setting, following some ideas introduced in [11, 5], and study in detail the case of the Wasserstein space of probability measures. In the first part we derive new general conditions ensuring convergence of the implicit time discretization scheme to a curve of maximal slope, the uniqueness, and the error estimates. In the second part we study in detail the differentiable structure of the Wasserstein space, to which the metric theory applies, and use this structure to give also an equivalent concept of gradient flow. Our analysis includes measures in infinite-dimensional Hilbert spaces and it does not require any absolute continuity assumption on the measures involved.

1. Introduction

In this paper we summarize the main results of the forthcoming book [3], devoted to the theory of gradient flows in a general metric setting and in the framework of the space of probability measures endowed with the L^2 Wasserstein metric. The presentation here reflects the structure of the book, with two parts that can be read almost independently of each other: in the first one we study gradient flows in metric spaces (calling them curves of maximal slope, since only the slope makes sense in this general setting), introduced following the ideas in [11] (see also [2]). We find general conditions ensuring the convergence of the implicit time discretization scheme to a curve of maximal slope. These conditions are typically satisfied if the energy functional is convex along geodesics. We also find a new convexity condition, which takes into account both the behaviour of the energy functional and of the metric, ensuring both uniqueness of curves of maximal slope (for a given initial datum) and full convergence, with explicit error estimates, of the implicit time discretization scheme (here we follow mostly the ideas in [5, 20]). These findings extend some convergence results known in Alexandroff Non Positively Curved metric spaces (see [18], [16]), but apply also to some Positively Curved metric spaces, as the Wasserstein space of probability measures.

In the second part we describe the Wasserstein space of probability measures in a separable Hilbert space X and its differentiable structure, recovering in a more general framework the formal calculus introduced in [21] and the representation of the Wasserstein distance as a "Riemannian" metric in [6]. One of the main features of our presentation is that we don't assume that X is finite-dimensional or that all measures μ under consideration are absolutely continuous with respect to a given measure (e.g. the Lebesgue measure in finite dimensions). Our derivation of the "Riemannian" structure (i.e. the tangent bundle and the metric on it) stems from a differential characterization of all absolutely continuous curves with values in the Wasserstein space. After the construction of the tangent bundle we examine in detail the differentiability properties along a rectifiable curve of the Wasserstein distance from a given point, and show the infinitesimal behaviour of the distance between nearby points on a rectifiable curve. In this framework also the notions of subdifferential and of gradient flows can be estabilished (see also [9], where analogous concepts appear in the context of the so-called Riemannian length spaces) and it turns out that gradient flows coincide with curves of maximal slope. This leads in many situations to the existence and the uniqueness of gradient flows (see also [9] for related results, mostly concerning uniqueness and rate of convergence as $t \to \infty$), to the convergence of the implicit time discretization scheme (see also [15], [8], [1]) and, under a slightly stronger convexity condition, to the error estimates for the scheme.

Due to obvious space constraints, the references quoted here are only a small part of the large literature available on these topics, and which will be represented in more detail in the book.

2. Gradient flows in metric spaces

2.1. **Basic notions.** Let us begin with some basic notions which make sense in any complete metric space $(\mathscr{S}, d)$ (see [2, 4] for a more detailed treatment of this topic).

Definition 2.1. *Let $v : (a,b) \to \mathscr{S}$ be a curve; we say that v belongs to $AC^p(a,b;\mathscr{S})$, for $p \in [1,+\infty]$, if there exists $m \in L^p(a,b)$ such that*

$$d(v(s), v(t)) \le \int_s^t m(r)\, dr \qquad \forall\, a < s \le t < b. \tag{2.1}$$

In the case $p = 1$ we are dealing with absolutely continuous *curves and we will denote the corresponding space simply with $AC(a,b;\mathscr{S})$.*

Theorem 2.2 (Metric derivative)**.** *Let $p \in [1,+\infty]$. Then for any $v \in AC^p(a,b;\mathscr{S})$ the limit*

$$|v'|(t) := \lim_{s\to t} \frac{d(v(s), v(t))}{|s-t|} \tag{2.2}$$

exists for $\mathscr{L}^1$-a.e. $t \in (a,b)$. Moreover the function $t \mapsto |v'|(t)$ belongs to $L^p(a,b)$, it is an admissible integrand for the right hand side of (2.1)*, and it is minimal in the following sense:*

$$|v'|(t) \le m(t) \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (a,b), \quad \text{for each function } m \text{ satisfying (2.1).} \tag{2.3}$$

Definition 2.3 (Local slope)**.** *The local slope of a functional ϕ at a point $v \in D(\phi)$ is 0 if v is isolated point of $\mathscr{S}$, otherwise it is given by*

$$|\partial\phi|(v) := \limsup_{w\to v} \frac{\big(\phi(v) - \phi(w)\big)^+}{d(v,w)}. \tag{2.4}$$

We will denote by $D(\partial\phi)$ the set of $v \in \mathscr{S}$ such that $|\partial\phi|(v) < +\infty$.

With these definitions in mind, we can formulate as in [11] (see also [11, 17, 2] the concept of gradient flow in a general metric setting.

Definition 2.4 (Curves of maximal slope). *We say that a locally absolutely continuous map $u:(a,b)\to\mathscr{S}$ is a* curve of maximal slope *for the functional ϕ if $\phi\circ u$ is $\mathscr{L}^1$-a.e. equal to a non-increasing map φ and*

$$\varphi'(t)\le -\frac{1}{2}|u'|^2(t)-\frac{1}{2}|\partial\phi|^2(u(t))\quad \textit{for } \mathscr{L}^1\textit{-a.e. } t\in(a,b). \tag{2.5}$$

We say that $\tilde u$ is the starting point of the curve u if $\lim_{t\to a}u(t)=\tilde u$.

To illustrate the heuristic ideas behind the previous definition, let us start with the classical setting of a gradient flow

$$u'(t)=-\nabla\phi\,(u(t)) \tag{2.6}$$

in a Hilbert space. If we take the modulus in both sides we have the equation $|u'|(t)=|\nabla\phi(u(t))|$ which makes sense in a metric setting, interpreting the left hand side as the metric derivative and the right hand side as the local slope. However, in passing from (2.6) to a scalar equation we clearly have a loss of information. This information can be retained by looking at the derivative of the energy:

$$\frac{d}{dt}\phi\,(u(t))=\langle u'(t),\nabla\phi\,(u(t))\rangle=-|u'(t)||\nabla\phi\,(u(t))\,|=-\frac{1}{2}|u'|^2(t)-\frac{1}{2}|\nabla\phi\,(u(t))\,|^2.$$

The second equality holds iff u' and $-\nabla\phi(u)$ are parallel and the third equality holds iff $|u'|$ and $|\nabla\phi(u)|$ are equal, so that we can rewrite (2.6) as

$$\frac{1}{2}|u'|^2(t)+\frac{1}{2}|\nabla\phi\,(u(t))\,|^2=-\frac{d}{dt}\phi\,(u(t))\,.$$

This argument shows that the metric formulation is consistent with the classical Hilbertian framework.

2.2. **Implicit time discretization.** The basic and classical construction of curves of maximal slope is based on a time discretization, building discrete solutions depending from a time step τ and passing to the limit as τ goes to 0.

The discrete solutions are built through a variational approximation of the problem and a recursive argument in the following way: we fix the parameter $\tau>0$ and the starting point U_τ^0, then we choose U_τ^{n+1} in the set

$$J_\tau[U_\tau^n]:=\operatorname{argmin}\Big\{\phi(\cdot)+\frac{1}{2\tau}d^2(U_\tau^n,\cdot)\Big\}. \tag{2.7}$$

Definition 2.5 (Discrete solutions). *The discrete solution is the piecewise constant function*

$$\bar U_\tau(t):=U_\tau^n\qquad if\ t\in[n\tau,(n+1)\tau).$$

Since we wish to show convergence of the discrete solutions to a limit curve, we have to look for some compactness property of the metric space (or of the sublevels of ϕ). Of course we may ask that this compacteness is with respect to the topology induced by the distance d; however, since this assumption is too strong in some applications, we believe it is more convenient to look for convergence w.r.t. a weaker auxiliary topology σ. So here and in the sequel we assume that there exists a topology σ with the following properties.

2.1a: ***Weak topology.*** σ is an Hausdorff topology on $\mathscr{S}$ compatible with d in the sense that σ is weaker than the topology induced by d and d is sequentially σ-lower semicontinuous:

$$(u_n, v_n) \overset{\sigma}{\rightharpoonup} (u, v) \quad \Longrightarrow \quad \liminf_{n \to \infty} d(u_n, v_n) \geq d(u, v). \tag{2.8}$$

De Giorgi introduced in [10] the notion of limits of discrete solutions in an abstract framework (where more general perturbations than the square of the distance are considered), and named them *minimizing movements*. Let us recall his definition.

Definition 2.6 (Minimizing movements). *For a given functional ϕ and an initial datum $u_0 \in \mathscr{S}$ we say that a curve $u : [0, +\infty) \to \mathscr{S}$ is a* mimimizing movement *for ϕ starting from u_0 if for every sufficiently small $\tau > 0$ there exists a discrete solution $\bar{U}_\tau$ defined as in (2.5) such that*

$$\lim_{\tau \downarrow 0} \phi(U_\tau^0) = \phi(u_0), \quad \lim_{\tau \downarrow 0} d(U_\tau^0, u_0) = 0, \quad \bar{U}_\tau(t) \overset{\sigma}{\rightharpoonup} u(t) \quad \forall\, t \in [0, +\infty). \tag{2.9}$$

We denote by $MM(\phi; u_0)$ the collection of all the minimizing movements for ϕ starting from u_0.
Analogously, we say that a curve $u : [0, +\infty) \to \mathscr{S}$ is a generalized mimimizing movement *for ϕ starting from u_0 if there exists a sequence $\tau_k \downarrow 0$ and a corresponding sequence of discrete solutions $\bar{U}_{\tau_k}$ defined as in (2.5) such that*

$$\lim_{k \to \infty} \phi(U_{\tau_k}^0) = \phi(u_0), \quad \lim_{k \to \infty} d(U_{\tau_k}^0, u_0) = 0, \quad \bar{U}_{\tau_k}(t) \overset{\sigma}{\rightharpoonup} u(t) \quad \forall\, t \in [0, +\infty). \tag{2.10}$$

We denote by $GMM(\phi; u_0)$ the collection of all the generalized minimizing movements for ϕ starting from u_0.

In order to be sure that discrete solutions exist and that minimizing movements (or the generalized ones) are curves of maximal slope, we need to make some assumptions on ϕ, and here are the main ones.

2.1b: ***Lower semicontinuity.*** We suppose that ϕ is sequentially σ-lower semicontinuous on d-bounded sets

$$\sup_{n,m} d(u_n, u_m) < +\infty, \quad u_n \overset{\sigma}{\rightharpoonup} u \quad \Longrightarrow \quad \liminf_{n \to \infty} \phi(u_n) \geq \phi(u). \tag{2.11a}$$

2.1c: ***Coercivity.*** There exist $\tau_* > 0$ and $u_* \in \mathscr{S}$ such that

$$m^* := \inf_{v \in \mathscr{S}} \phi(v) + \frac{1}{2\tau} d^2(v, u_*) > -\infty. \tag{2.11b}$$

2.1d: ***Compactness.*** Every d-bounded set contained in a sublevel of ϕ is relatively σ-sequentially compact: i.e.,

$$\begin{gathered}\text{every sequence } (u_n) \subset \mathscr{S} \text{ with } \sup_n \phi(u_n) < +\infty, \quad \sup_{n,m} d(u_n, u_m) < +\infty \\ \text{admits a } \sigma\text{-convergent subsequence.}\end{gathered} \tag{2.11c}$$

2.1e: ***Semicontinuity of the slope.*** The slope satisfies the following equation

$$|\partial\phi|(u) = \inf \Big\{ \liminf_{n \to +\infty} |\partial\phi|(u_n) \ : \ u_n \overset{\sigma}{\rightharpoonup} u, \ \sup_n \{d(u_n, u), \phi(u_n)\} < +\infty \Big\}. \tag{2.11d}$$

Basically the assumptions b, c ensure the existence of discrete solutions, d is needed to find a limit curve and e to is needed to show that this limit curve is of maximal slope.

Proposition 2.7 (Convergence of the implicit time discretization scheme). *Let us suppose that the assumptions of* 2.1*a, b, c, d hold and let be given a sequence* $(\tau_n) \downarrow 0$ *and a corresponding family of initial data* $\{U^0_{\tau_n}\}$ *satisfying*

$$U^0_{\tau_n} \overset{\sigma}{\rightharpoonup} u_0, \quad \phi(U^0_{\tau_n}) \to \phi(u_0) < +\infty \quad \textit{as } n \to +\infty, \quad \sup_n d(U^0_{\tau_n}, u_0) < +\infty. \tag{2.12}$$

Then there exist a subsequence, not relabeled, and a limit curve $u \in AC^2_{\rm loc}([0,+\infty);\mathscr{S})$ *such that*

$$\bar U_{\tau_n}(t) \overset{\sigma}{\rightharpoonup} u(t) \qquad \forall\, t \in [0,+\infty). \tag{2.13}$$

In particular $u \in GMM(\phi; u_0)$*, which is a nonempty subset of* $AC^2_{\rm loc}([0,+\infty);\mathscr{S})$.

Theorem 2.8 (Limit curves are of maximal slope). *Assume that conditions* 2.1*a, b, c, e are fullfilled and that* ϕ *satisfies the continuity property*

$$\sup_{n\in\mathbb{N}} \Big\{ |\partial\phi|(v_n), d(v_n, v_0), \phi(v_n) \Big\} < +\infty, \quad v_n \overset{\sigma}{\rightharpoonup} v \quad \Longrightarrow \quad \phi(v_n) \to \phi(v). \tag{2.14}$$

Then every $u \in GMM(\phi; u_0)$ *is a curve of maximal slope for* ϕ.

Some stronger results (i.e. the energy identity and the convergence of various discrete quantities to their continuous counterparts) can be given replacing the continuity condition (2.14) by the assumption that $|\partial\phi|$ is an upper gradient of ϕ in the sense of Heinonen and Koskela (see [14] and (2.15) below).

Theorem 2.9. *Assume that conditions a, b, c, e hold and that* $|\partial\phi|$ *has the following property: for every absolutely continuous curve* $v : [0,1] \to \mathscr{S}$ *the function* $|\partial\phi| \circ v$ *is Borel and*

$$|\phi(v(t)) - \phi(v(s))| \le \int_t^s |\partial\phi|(v(r))|v'(r)|\, dr \quad \forall 0 < s \le t < 1. \tag{2.15}$$

Then every curve $u \in GMM(\phi; u_0)$ *is a curve of maximal slope for* ϕ *and* u *satisfies the energy identity*

$$\frac{1}{2}\int_0^T |u'|^2(t)\, dt + \frac{1}{2}\int_0^T |\partial\phi|^2(u(t))\, dt + \phi(u(T)) = \phi(u_0) \qquad \forall T > 0. \tag{2.16}$$

Moreover, if $\{\bar U_{\tau_n}\}_{n\in\mathbb{N}}$ *is a sequence of discrete solutions satisfying* (2.12) *and* (2.13)*, we have*

$$\lim_{n\to\infty} \phi(\bar U_{\tau_n}(t)) = \phi(u(t)) \quad \forall\, t \in [0,+\infty), \tag{2.17}$$

$$\lim_{n\to\infty} |\partial\phi|(\bar U_{\tau_n}) = \lim_{n\to\infty} |U'_{\tau_n}| = |u'| = |\partial\phi|(u) \quad \textit{in } L^2_{\rm loc}([0,+\infty)), \tag{2.18}$$

where $|U'_\tau|$ *is defined as*

$$|U'_\tau|(t) = \frac{d(U^{n-1}_\tau, U^n_\tau)}{\tau} \quad \textit{if } t \in ((n-1)\tau, n\tau) \tag{2.19}$$

and $|\partial^-\phi|$ *is the right hand side in* (2.11d).

2.3. **The geodesically convex case.** The abstract conditions given above are fulfilled in the case when the functional is convex along constant speed geodesics, according to the definition below. This case is relevant for many applications, see for instance [16, 19, 21, 22] and the examples mentioned in the following section.

Definition 2.10 (Convexity along curves). *A functional $\phi : \mathscr{S} \to (-\infty + \infty]$ is said to be convex along the curve $\gamma : [0,1] \to \mathscr{S}$ if*

$$\phi(\gamma_t) \le (1-t)\phi(\gamma_0) + t\phi(\gamma_1) \tag{2.20}$$

Definition 2.11 (Constant speed geodesics and length spaces). *A curve $\gamma : [0,1] \to \mathscr{S}$ is a* constant speed geodesic *connecting two points $v_0, v_1 \in \mathscr{S}$ if $\gamma_i = v_i$, $i = 0, 1$, and*

$$d(\gamma_s, \gamma_t) = (t-s)d(v_0, v_1) \qquad \forall s, t \in [0,1], \;\; s \le t. \tag{2.21}$$

A metric space $\mathscr{S}$ is said to be a length space *if for every couple of point $x, y \in \mathscr{S}$ there exists at least one (not necessarily unique) geodesic connecting them.*

Definition 2.12 (Geodesically convex functionals). *A functional $\phi : \mathscr{S} \to (-\infty, +\infty]$ on a length space $\mathscr{S}$ is* geodesically convex *if for every $x, y \in \mathscr{S}$ there exists a constant speed geodesic connecting them along which ϕ is convex.*

For geodesically convex functionals the local slope admits the following simple representation:

$$|\partial\phi|(v) = \sup_{w \ne v} \left(\frac{\phi(v) - \phi(w)}{d(v,w)} \right)^+, \tag{2.22}$$

from which one can show that $|\partial\phi|$ has the upper gradient property stated in Theorem 2.9 and that 2.1e holds with σ equal to the topology induced by d. Then, one obtains the following result.

Theorem 2.13. *Suppose that ϕ is geodesically convex and that all the assumptions 2.1 hold. Then every u_0 such that $\phi(u_0) < +\infty$ is the starting point of a curve of maximal slope for ϕ and all the conclusions of Theorem 2.9 hold.*

Moreover the convexity property ensures some pointwise differentiability properties, first observed by Brezis [7] in the Hilbertian setting. We collect them in the following theorem.

Theorem 2.14 (Pointwise properties). *Let us suppose that assumptions 2.1a, b hold and that ϕ is geodesically convex. If $\phi(u_0) < +\infty$ then each element $u \in GMM(\phi; u_0)$ is locally Lipschitz in $(0, +\infty)$ and satisfies the following properties:*

(i) *The right metric derivative*

$$|u'_+|(t) := \lim_{s \downarrow t} \frac{d(u(s), u(t))}{s - t} \tag{2.23}$$

exists and $|\partial\phi|(u(t)) < +\infty$ for all $t > 0$.

(ii) *The map $t \mapsto \phi(u(t))$ is convex and the map $t \mapsto |\partial\phi|(u(t))$ is non-increasing and right continuous.*

(iii) *The equation*

$$\frac{d}{dt_+}\phi(u(t)) = -|\partial\phi|^2(u(t)) = -|u'_+|^2(t) = -|\partial\phi|(u(t))\,|u'_+|(t) \tag{2.24}$$

is satisfied at every $t \in (0, +\infty)$.

Even though the geodesic convexity ensures *existence* of curves of maximal slope and several useful properties, including the energy identity, *uniqueness* of curves of maximal slope for a given initial datum is an open problem. This problem is open even in a linear framework, for instance when $\mathscr{S}$ is a reflexive Banach space, and seems to be related to the lack of an Hilbertian structure, even on small scales. We discuss this problem in the next subsection.

2.4. **A new kind of convexity assumption.** Explicit error estimates for the implicit time discretization scheme and uniqueness of curves of maximal slope can be proved if we ask convexity not only for the functional, but even for the distance.

Assumption 2.15. *We suppose that for every choice of* $w, v_0, v_1 \in \mathscr{S}$ *there exists a curve* $\gamma : [0,1] \to \mathscr{S}$ *such that* $\gamma_i = v_i$, $i = 0, 1$, ϕ *is convex along the curve and the following estimate holds*

$$d^2(w, \gamma_t) \le (1-t)d^2(w, v_0) + td^2(w, v_1) - t(1-t)d^2(v_0, v_1) \tag{2.25}$$

This assumption is in some sense stronger than the geodesic convexity, since we are requiring a uniform convexity condition also on the distance; on the other hand it is weaker, since the curve γ along which this has to happen need not be a geodesic. In fact, in the applications described in the next section, this degree of freedom in the choice of the curve is extremely useful. Notice also that condition (2.25) along geodesics is the definition of Non Positively Curved (NPC) metric spaces in the sense of Alexandroff (for Riemannian metric spaces the condition is equivalent to non positivity of all sectional curvatures). In the setting of NPC metric spaces the uniqueness of gradient flows was proved in [18] (see also [16]), but our more general result follows a different path, strongly related to the ideas in [5, 20].

Moreover we identify now the topology σ with the one induced by the distance d and drop the compactness assumption 2.1d: indeed, Assumption (2.15) guarantees both the existence of a minimizer for the discrete problem 2.7 and the convergence of the discrete solutions to a curve of maximal slope. The main results are contained in the following theorem, in which we use the notation $\phi_\tau(u)$ for the function $\inf \phi(\cdot) + d^2(\cdot, u)/2\tau$.

Theorem 2.16. *Let us assume that* ϕ *is a proper, lower semicontinuous, coercive functional (according to* (2.1)*c) and that Assumption 2.15 is satisfied. Then*

i) Convergence and exponential formula: *for each* $u_0 \in \overline{D(\phi)}$ *there exists a unique element* $u = S[u_0]$ *in* $MM(\Phi; u_0)$ *which therefore can be expressed through the exponential formula*

$$u(t) = S[u_0](t) = \lim_{n\to\infty} U^n_{t/n}(u_0). \tag{2.26}$$

ii) Regularizing effect: *u is a locally Lipschitz curve of maximal slope with $u(t) \in D(\partial\phi) \subset D(\phi)$ for $t > 0$; in particular the following a priori bounds hold:*

$$\phi(u(t)) \le \phi_t(u_0), \quad |\partial\phi|^2(u(t)) \le |\partial\phi|^2(v) + \frac{1}{t^2}d^2(v, u_0) \quad \forall\, v \in D(\phi). \tag{2.27}$$

iii) Uniqueness and evolution variational inequalities: *u is the unique solution of the evolution variational inequality*

$$\frac{1}{2}\frac{d}{dt}d^2(u(t), v) + \phi(u(t)) \le \phi(v) \quad \mathscr{L}^1\text{-a.e. in } (0, +\infty), \quad \forall\, v \in D(\phi). \tag{2.28}$$

iv) Contraction semigroup: *The map $t \mapsto S[u_0](t)$ is a non expansive semigroup, i.e.*

$$d\big(S[u_0](t), S[v_0](t)\big) \le d(u_0, v_0) \quad \forall\, u_0, v_0 \in \overline{D(\phi)}. \tag{2.29}$$

v) Optimal a priori estimate: *if $u_0 \in D(\phi)$ then*

$$d^2\Big(S[u_0](t), U^n_{t/n}(u_0)\Big) \le \frac{t}{n}\Big(\phi(u_0) - \phi_{t/n}(u_0)\Big) \le \frac{t^2}{2n^2}|\partial\phi|^2(u_0). \tag{2.30}$$

3. Gradient flows on the space of probability measures

3.1. **Basic definitions and notation.** This section is devoted to the study of the differential structure of the space of probability measures on a separable Hilbert space X (in the following denoted by $\mathscr{P}(X)$), endowed with the Wasserstein distance W_2. Although most of the results stated here are true even in the general case $p > 1$, see [3], in this paper for the sake of simplicity we will state them only for $p = 2$.

Let us first recall the basic definitions and properties.

Definition 3.1 (Transport of measures). *Let μ be a probability measure on X and let $r : X \to X$ be a Borel map. The push forward $r_\#\mu \in \mathscr{P}(X)$ of μ through r is defined by*

$$r_\#\mu(B) := \mu(r^{-1}(B)) \qquad \text{for any Borel subset } B \subset X. \tag{3.1}$$

More generally, the integral w.r.t. μ and the integral w.r.t. $r_\#\mu$ are related by

$$\int_X f(r(x))\,d\mu(x) = \int_X f(y)\,dr_\#\mu(y) \tag{3.2}$$

for every bounded (or positive) Borel function f.

Definition 3.2 (Transport plans). *Given two measures μ^1, $\mu^2 \in \mathscr{P}(X)$, the set of* transport plans *between them is:*

$$\Gamma(\mu^1, \mu^2) := \Big\{\boldsymbol{\mu} \in \mathscr{P}(X \times X) : \pi^1_\#\boldsymbol{\mu} = \mu^1,\ \pi^2_\#\boldsymbol{\mu} = \mu^2\Big\}, \tag{3.3}$$

where $\pi^i : X \times X \to X$, $i = 1, 2$, are the projections onto the first and onto the second coordinate; notice that this set is always non empty since it contains at least $\mu^1 \times \mu^2$.

The family of transport plans includes in some sense the family of transport maps: indeed any transport map r induces a plan $\boldsymbol{\mu}$ defined by $(Id \times r)_\#\boldsymbol{\mu}$, where $(Id \times r)(x) = (x, r(x))$.

The Wasserstein distance is defined on the family of measures in $\mathscr{P}(X)$ whose second moment is finite, i.e.:

$$\mathscr{P}_2(X) := \Big\{\mu \in \mathscr{P}(X) : \int_X |x|^2 d\mu(x) < +\infty\Big\}. \tag{3.4}$$

Definition 3.3 (The optimal transportation problem). *Given μ^1, $\mu^2 \in \mathscr{P}_2(X)$ their Wasserstein distance is defined by*

$$W_2^2(\mu^1,\mu^2) := \min\Big\{ \int_{X^2} |x_1 - x_2|^2 \, d\boldsymbol{\mu}(x_1,x_2) : \ \boldsymbol{\mu} \in \Gamma(\mu^1,\mu^2)\Big\}. \tag{3.5}$$

It is not hard to show that the minimum is always attained and that the function defined above is a distance: we will denote by $\Gamma_o(\mu^1,\mu^2)$ the subset of $\Gamma(\mu^1,\mu^2)$ where the minimum is attained, i.e.

$$\boldsymbol{\mu} \in \Gamma_o(\mu^1,\mu^2) \iff \int_{X\times X} |x_1 - x_2|^2 \, d\boldsymbol{\mu}(x_1,x_2) = W_2^2(\mu^1,\mu^2). \tag{3.6}$$

Note that $\Gamma_o(\mu^1,\mu^2)$ is a closed convex subset of $\Gamma(\mu^1,\mu^2)$. In the case when μ^1 vanishes on any Gaussian null set (in the case $X = \mathbb{R}^d$ this condition is equivalent to the absolute continuity of μ^1 with respect to $\mathscr{L}^d$) we can say much more thanks to the following theorem, due in the finite dimensional case to Knott–Smith and Brenier (see also [12] for an analogous result in Wiener spaces and [13] for a sharper result in finite dimensions).

Theorem 3.4 (Existence and uniqueness of the optimal transport map). *Let μ^1, $\mu^2 \in \mathscr{P}_2(X)$ and suppose that μ^1 vanishes on any Gaussian null set. Then $\Gamma_o(\mu^1,\mu^2)$ contains only one element $\boldsymbol{\mu}$, moreover $\boldsymbol{\mu}$ is induced by a Borel map r and r is the Gateaux gradient of a convex function.*

We will denote by $\mathscr{P}_2^r(X)$ the subset of $\mathscr{P}_2(X)$ made by all measures vanishing on any Gaussian null set and, for every $\mu \in \mathscr{P}_2^r(X)$, we will denote by T_μ^σ the map given by the previous theorem; note that $(T_\mu^\sigma)_\#\mu = \sigma$ and $\int_X |T_\mu^\sigma - Id|^2 \, d\mu = W_2^2(\mu,\sigma)$.

Finally we recall the basic properties of convergence and compactness in $(\mathscr{P}_2(X), W_2)$; remember that a subset $\mathcal{K}$ of $\mathscr{P}(X)$ is said to be *tight* if

$$\lim_{R\to+\infty} \sup_{\mu\in\mathcal{K}} \mu(X \setminus B_R) = 0, \tag{3.7}$$

and *2-uniformly integrable* if

$$\lim_{R\to+\infty} \sup_{\mu\in\mathcal{K}} \int_{X\setminus B_R} |x|^2 d\mu(x) = 0. \tag{3.8}$$

Theorem 3.5 (Compactness and convergence). *$(\mathscr{P}_2(X), W_2)$ is a complete and separable metric space. Moreover, a set $\mathcal{K} \subset \mathscr{P}_2(X)$ is relatively compact if and only if it is tight and 2-uniformly integrable. Finally, given a sequence $(\mu_n) \subset \mathscr{P}_2(X)$ and $\mu \in \mathscr{P}_2(X)$ it holds:*

$$\lim_{n\to\infty} W_2(\mu_n,\mu) = 0 \iff \begin{cases} \mu_n \text{ weakly converge to } \mu, \\ \{\mu_n\} \text{ is 2-uniformly integrable,} \end{cases} \tag{3.9}$$

where the weak convergence is w.r.t. the duality with continuous and bounded functions.

3.2. **The Riemannian structure of** $(\mathscr{P}_2(X), W_2)$**.** In this section we analyze the "Riemannian" structure of $(\mathscr{P}_2(X), W_2)$, developing a first order calculus which makes rigorous the approach pursued in [21] and is consistent with the Benamou–Brenier formula (3.15). Our starting point is to derive the differentiable structure of the space (i.e. the tangent bundle and the metric on it) starting from a characterization of the absolutely continuous curves, defined only through the metric of the space; notice that this viewpoint can also be used in a classical setting, e.g. a finite-dimensional Riemannian manifold embedded in a Euclidean space.

In the statement of the following theorem we use the class of *cylindrical* test functions in X, of the form $\varphi \circ \pi$ where $\varphi \in C_c^\infty(\mathbb{R}^d)$ and $\pi : X \to \mathbb{R}^d$ is the coordinate map induced by d orthonormal vectors in X. If $I \subset \mathbb{R}$ is an open interval, the class of cylindrical test functions in $X \times I$ is defined analogously, using functions $\varphi \in C_c^\infty(\mathbb{R}^d \times I)$.

Theorem 3.6 (Absolutely continuous curves and the continuity equation)**.** *Let $I \subset \mathbb{R}$ be an open interval, let $\mu_t : I \to \mathscr{P}_2(X)$ be an absolutely continuous curve and let $|\mu'| \in L^1(I)$ be its metric derivative. Then there exist Borel vector fields $v_t(x)$ such that*

$$v_t \in L^2(\mu_t, X), \quad \|v_t\|_{L^2(\mu_t)} \le |\mu'|(t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I \tag{3.10}$$

and the continuity equation

$$\dot\mu_t + \nabla \cdot (v_t \mu_t) = 0 \qquad \text{in } X \times (0,1) \tag{3.11}$$

holds in the sense of distributions, i.e.

$$\int_0^1 \int_X \Big(\partial_t \phi(x,t) + \langle v_t(x), \nabla_x \phi(x,t)\rangle\Big)\, d\mu_t(x)\, dt = 0 \tag{3.12}$$

for any cylindrical function φ in $X \times I$. Morever, for $\mathscr{L}^1$-a.e. $t \in I$ v_t belongs to the closure in $L^2(\mu_t, X)$ of the subspace generated by the gradients of cylindrical functions in X.

Conversely, if $\mu_t : I \to \mathscr{P}_2(X)$ satisfies the continuity equation for some Borel velocity field v_t with $\|v_t\|_{L^2(\mu_t)} \in L^1(I)$ then μ_t is an absolutely continuous curve and $|\mu'|(t) \le \|v_t\|_{L^2(\mu_t)}$ for $\mathscr{L}^1$-a.e. $t \in I$.

Obviously for a given curve μ_t there is no uniqueness for the vector fields v_t satisfying the continuity equation (3.11): choosing vector fields w_t such that $\nabla \cdot (w_t \mu_t) = 0$ the vectors $v_t + w_t$ still satify (3.11). However the following facts suggest a canonical choice of v_t:

- the vectors v_t act only on $\nabla\varphi$, with φ cylindrical in X;
- the second implication shows that the norm of v_t in $L^2(\mu_t, X)$ is always greater than the metric derivative $|\mu'|(t)$ for $\mathscr{L}^1$-a.e. $t \in I$;
- the first implication shows that it is possible to have
$$||v_t||_{L^2(\mu_t, X)} = |\mu'|(t) \qquad \text{for } \mathscr{L}^1\text{-a.e. } t \in I; \tag{3.13}$$
- the linearity of (3.11) and the strict convexity of the L^2 norm imply the uniqueness of the vectors satisfying (3.13).

So it is natural to consider as velocity vectors those v_t's for which both (3.11) and (3.13) hold and to give the following definition.

Definition 3.7 (Tangent bundle). *Let $\mu \in \mathscr{P}_2(X)$. We define*

$$\mathrm{Tan}_\mu \mathscr{P}_2(X) := \overline{\{\nabla\varphi : \ \varphi \ \textit{cylindrical in } X\}}_{L^2(\mu)}.$$

Remark 3.8. *As we said before, vectors in $\mathrm{Tan}_\mu \mathscr{P}_2(X) \subset L^2(\mu, X)$ can be characterized even by the following variational principle: $v \in \mathrm{Tan}_\mu \mathscr{P}_2(X)$ iff*

$$\|v + w\|_{L^2(\mu)} \geq \|v\|_{L^2(\mu)} \quad \forall w \in L^2(\mu; X) \ \textit{such that} \ \nabla \cdot (w\mu) = 0. \tag{3.14}$$

The tangent space, being a closed subspace of $L^2(\mu, X)$, is endowed with a natural inner product $\langle \cdot, \cdot \rangle_\mu$; however the space $(\mathscr{P}_2(X), W_2)$ is *not* an infinite dimensional Riemannian manifold (not even in the Euclidean case $X = \mathbb{R}^d$ or for measures μ with a smooth density) since it is not possible to define an exponential map from a neighbourhood of the origin in $\mathrm{Tan}_\mu \mathscr{P}_2(X)$ into $\mathscr{P}_2(X)$ which is an homeomorfism. This is due to the fact that the characterization of geodesics in the space gives

$$\exp_\mu(w) := (Id + w)_\# \mu$$

provided $x \mapsto x + w(x)$ is the gradient of a convex function. On the other hand, there are elements $v \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^d)$ whose second derivatives are unbounded from below, therefore $x \mapsto x + hv(x)$ is not an optimal transport map for any $h > 0$.

In spite of this difference we keep the terminology *Riemannian structure* since there are a lot of analogies with the Riemannian case (these analogies lead in [9] to the terminology of *Riemannian length space*): the most relevant one is that indeed the Wasserstein metric *is* the Riemannian metric induced by the tangent bundle and its inner product defined above. This fact was noticed independently by Benamou and Brenier [6] and Otto [21], in the case when $X = \mathbb{R}^d$ and all measures under consideration are absolutely continuous with respect to $\mathscr{L}^d$.

Theorem 3.9 (Benamou–Brenier formula). *Given μ^0, $\mu^1 \in \mathscr{P}_2(X)$ it holds*

$$W_2^2(\mu^0, \mu^1) = \inf \int_0^1 ||v_t||^2_{L^2(\mu_t)} dt, \tag{3.15}$$

where the infimum is taken among all absolutely continuous curves $\mu_t : [0,1] \to \mathscr{P}_2(X)$ such that $\mu_i = \mu^i$ for $i = 0, 1$ and v_t are given by the equations (3.11) and (3.13).

The following theorem shows how to recover the tangent vectors v_t through the infinitesimal behaviour of optimal transport plans along the curve. It is interesting to note that in the limit we recover a plan $(Id \times v_t)_\# \mu_t$ induced by a map even when μ_t is not necessarily regular.

Theorem 3.10 (Optimal plans along a.c. curves). *Let $\mu_t : I \to \mathscr{P}_2(X)$ be an absolutely continuous curve and let $v_t \in \mathrm{Tan}_{\mu_t} \mathscr{P}_2(X)$ be its velocity vector. Then, for $\mathscr{L}^1$-a.e. $t \in I$ the following property holds: for any choice of $\boldsymbol{\mu}_h \in \Gamma_0(\mu_t, \mu_{t+h})$ we have*

$$\lim_{h \to 0} \left(\pi^1, \frac{1}{h}(\pi^2 - \pi^1)\right)_\# \boldsymbol{\mu}_h = (Id \times v_t)_\# \mu_t \qquad \textit{in } \mathscr{P}_2(X \times X) \tag{3.16}$$

and

$$\lim_{h \to 0} \frac{W_2(\mu_{t+h}, (Id + hv_t)_\# \mu_t)}{|h|} = 0. \tag{3.17}$$

In particular, for $\mathscr{L}^1$-a.e. $t \in I$ such that $\mu_t \in \mathscr{P}_2^r(X)$ we have

$$\lim_{h\to 0} \frac{1}{h}(T_{\mu_t}^{\mu_{t+h}} - Id) = v_t \qquad \text{in } L^2(\mu_t, X). \tag{3.18}$$

The notions and theorems just introduced lead in a natural way to the concept of differentiability; all definitions and results given below can be adapted to any measure (see [3]) but, for the sake of simplicity, we will state them only for regular ones. The technical problem which arises without this assumption is that in general an optimal transport plan is not induced by a map.

Definition 3.11 (Differentiability at regular measures)**.** *We say that $\phi : \mathscr{P}_2(X) \to \mathscr{P}_2(X)$ is differentiable at $\mu \in \mathscr{P}_2^r(X)$ if there exists $v \in \mathrm{Tan}_\mu \mathscr{P}_2(X)$ such that*

$$\lim_{\mu'\to\mu} \frac{\phi(\mu') - \phi(\mu) - \langle v, T_\mu^{\mu'} - Id\rangle_\mu}{W_2(\mu', \mu)} = 0.$$

The vector v if exists is unique, and it will be called differential of ϕ at μ.

An interesting and useful fact is that for any $\sigma \in \mathscr{P}_2(X)$ the function $\mu \to W_2^2(\mu, \sigma)$ is differentiable at any $\mu \in \mathscr{P}_2^r(X)$.

Theorem 3.12 (Differential of the squared distance)**.** *The function $\mu \mapsto W^2(\mu, \sigma)$ is differentiable at any $\mu \in \mathscr{P}_2^r(X)$. Its differential is $v := 2(Id - T_\mu^\sigma)$.*

Note that by Theorem 3.10 it follows that if μ_t is an absolutely continuous curve and ϕ a function which is differentiable at μ_t for $\mathscr{L}^1$-a.e. t, then the map $t \mapsto \phi(\mu_t)$ is $\mathscr{L}^1$-a.e. differentiable and its derivative is $\langle w_t, v_t\rangle_{\mu_t}$, where w_t is the differential of ϕ at μ_t and v_t is the velocity vector of the curve. In particular for any curve $t \mapsto \mu_t \in \mathscr{P}_2^r(X)$ and any $\sigma \in \mathscr{P}_2(X)$ the derivative of the function $t \mapsto W_2^2(\mu_t, \sigma)$ exists $\mathscr{L}^1$-a.e. and is equal to $2\langle Id - T_{\mu_t}^\sigma, v_t\rangle_{\mu_t}$. One can show, using again Theorem 3.10, that this differentiability property is still true for maps $t \mapsto \mu_t \in \mathscr{P}_2(X)$ and, in this case, the derivative involves optimal transport plans instead of optimal transport maps.

3.3. **Subdifferential and gradient flow.** Following the same ideas of the previous section it seems that the most natural definition of subdifferential $\partial^-\phi(\mu)$ of the function ϕ at a point $\mu \in \mathscr{P}_2^r(X)$ is

$$v \in \partial^-\phi(\mu) \quad \text{iff} \quad \phi(\sigma) \geq \phi(\mu) + \langle v, T_\mu^\sigma - Id\rangle_\mu \quad \forall \sigma \in \mathscr{P}_2(X). \tag{3.19}$$

However, for technical reasons, in order to obtain stronger pointwise properties of the subdifferential operator (for istance the closure) we need to extend this notion.

In the following definition we need to consider multiple plans, i.e. probability measures in $X \times X \times X$; we denote by $\pi^{i,j} : X^3 \to X^2$ the canonical projections, for $1 \leq i, j \leq 3$.

Definition 3.13 (Subdifferential)**.** *Let $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ and let $\mu \in D(\phi)$. A plan $\boldsymbol{\mu} \in \mathscr{P}_2(X \times X)$ belongs to the subdifferential of ϕ at μ, denoted by $\partial^-\phi(\mu)$ if*

(i) $\pi^1_\# \boldsymbol{\mu} = \mu$;

(ii) *for any* $\sigma \in \mathscr{P}_2(X)$ *there exists* $\boldsymbol{\gamma} \in \mathscr{P}_2(X \times X \times X)$ *such that* $\pi^{1,2}_{\#}\boldsymbol{\gamma} = \boldsymbol{\mu}$ *and* $\pi^{1,3}_{\#}\boldsymbol{\gamma} \in \Gamma_0(\mu, \sigma)$, *satisfying*

$$\phi(\sigma) \geq \phi(\mu) + \int_{X^3} \langle x_2, x_3 - x_1 \rangle \, d\boldsymbol{\gamma}. \tag{3.20}$$

Note that even if $\mu \in \mathscr{P}_2^r(X)$ the previous definition is finer than the first one in (3.19), since it may happen that a measure $\boldsymbol{\mu}$ in the subdifferential is not induced by a map; the formula (3.20) reduces to the one in (3.19) if μ is regular and $\boldsymbol{\mu} = (Id \times v)_{\#}\mu$.

Remark 3.14. *The set* $\partial^- \phi(\mu)$ *is a closed convex subset of* $\mathscr{P}_2(X \times X)$.

As we said before one can show that the subdifferential just defined is a closed operator, in the sense specified by the next proposition.

Proposition 3.15 (Closure of the subdifferential operator). *Let* $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ *be a lower semicontinuous functional and assume that* $(\mu_h) \subset D(\phi)$ *converge to* $\mu \in D(\phi)$ *in* $\mathscr{P}_2(X)$ *and* $\boldsymbol{\mu}_h \in \partial^- \phi(\mu_h)$ *narrowly converge to* $\boldsymbol{\mu}$, *with* $\int |x_2|^2 d\boldsymbol{\mu}_h$ *bounded. Then* $\boldsymbol{\mu} \in \partial^- \phi(\mu)$.

Now we turn to the definition of gradient flow, based on the *differentiable* structure of the Wasserstein space; we will soon show that in the case of geodesically convex functionals this notion coincides with the purely *metric* one of curve of maximal slope.

Definition 3.16 (Gradient flow). *We say that a map* $\mu_t \in AC^2_{\mathrm{loc}}((0, +\infty); \mathscr{P}_2(X))$ *is a solution of the gradient flow equation*

$$\dot{\mu}_t \in -\partial^- \phi(\mu_t) \tag{3.21}$$

if for $\mathscr{L}^1$*-a.e.* $t > 0$ *its velocity vector* $v_t \in \mathrm{Tan}_{\mu_t}\mathscr{P}_2(X)$ *satisfies* $(Id \times v_t)_{\#}\mu \in \partial^- \phi(\mu_t)$.

The key ingredient for the equivalence we stated before is the following lemma.

Lemma 3.17 (Metric slope and subdifferential). *Let* $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ *be a geodesically convex and l.s.c. functional satisfying*

$$\phi(\mu) \geq -aW_2^2(\mu, \bar{\mu}) + b \qquad \forall \mu \in \mathscr{P}_2(X) \tag{3.22}$$

for some $a > 0$, $b \in \mathbb{R}$ *and* $\bar{\mu} \in \mathscr{P}_2(X)$. *Then*

$$|\partial \phi|^2(\mu) = \min \left\{ \int_{X \times X} |x_2|^2 \, d\boldsymbol{\mu} : \ \boldsymbol{\mu} \in \partial^- \phi(\mu) \right\} \qquad \forall \mu \in D(\phi),$$

with the convention $\min \emptyset = +\infty$.

Theorem 3.18 (Curves of maximal slope coincide with gradient flows). *Let* $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ *be a geodesically convex and l.s.c. functional satisfying* (3.22). *Then* $\mu_t : (0, +\infty) \to \mathscr{P}_2(X)$ *is a curve of maximal slope according to Definition 2.4 iff* μ_t *is a gradient flow.*

Because of this equivalence, in order to show the existence of gradient flows we can apply the Theorems 2.8, 2.9, 2.13, 2.16 of the previous section; in this case it is convenient to choose as weak topology the one induced by the narrow convergence (i.e. in the duality with continuous and bounded functions in X). But one can also use the differentiable structure of the space and pass to the limit in a kind of discrete subdifferential inequality, in the same spirit of [15] (but the energy arguments involved in the metric theory seem to be more general, handling for instance also singular measures or the case when $p \neq 2$).

As we already said, uniqueness of curves of maximal slope is still an open problem under the only geodesic convexity assumption. In this case, however, one can use the differentiable structure of the Wasserstein space, and in particular the differentiability properties of the distance recalled above, to show a kind of contraction property leading to uniqueness.

Theorem 3.19 (Uniqueness of gradient flows). *If $\phi : \mathscr{P}_2(X) \to (-\infty, +\infty]$ is a geodesically convex functional then for any $\mu_0 \in \mathscr{P}_2(X)$ there is at most one gradient flow $\mu_t : (0, +\infty) \to \mathscr{P}_2(X)$ satisfying $\lim_{t\downarrow 0} \mu_t = \mu_0$.*

Finally, it is important to note that, quite surprisingly, for a large class of geodesically convex functionals considered in the literature (the *internal* energy, the *potential* energy and the *interaction* energy considered in [19]), even the Assumption 2.15 is satisfied and therefore not only uniqueness of the gradient flow, but also error estimates for the implicit time discretization scheme are available, according to Theorem 2.16 (also, existence does not require anymore any compactness property of the sublevels of ϕ). In fact, even though the geodesics of the space $\mathscr{P}_2(X)$ do not satisfy the inequality (2.25) (actually it was first noticed in [21] that always the opposite inequality holds!), it is possible to interpolate with different curves along which a geodesically convex functional is (often, but not always) convex and (2.25) holds. Given μ_0, $\mu_1 \in \mathscr{P}_2(X)$, the curves we are talking about are

$$\mu_t := \big((1-t)T_\mu^{\mu_0} + tT_\mu^{\mu_1}\big)_{\#}\mu \tag{3.23}$$

in the case when $\mu \in \mathscr{P}_2^r(X)$ (the general definition requires multiple transport plans, as in the definition of the subdifferential).

References

[1] Martial Agueh, *Existence of solutions to degenerate parabolic equations via the monge-kantorovich theory*, Arch. Rat. Mech. Anal., to appear (2002).

[2] Luigi Ambrosio, *Minimizing movements*, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5) **19** (1995), 191–246. MR **97c:**49044

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the wasserstein space of probability measures*, Birkhäuser, 2004.

[4] Luigi Ambrosio and Paolo Tilli, *Selected topics on "analysis in metric spaces"*, Scuola Normale Superiore, Pisa, 2000.

[5] Claudio Baiocchi, *Discretization of evolution variational inequalities*, Partial differential equations and the calculus of variations, Vol. I (F. Colombini, A. Marino, L. Modica, and S. Spagnolo, eds.), Birkhäuser Boston, Boston, MA, 1989, pp. 59–92. MR **90m:**49002

[6] Jean-David Benamou and Yann Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), no. 3, 375–393. MR **2000m:**65111

[7] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973, North-Holland Mathematics Studies, No. 5. Notas de Matemática (50). MR 50 #1060

[8] E.A. Carlen and Wilfrid Gangbo, *Constrained steepest descent in the 2-wasserstein metric*, Ann. Math., forthcoming (2000).

[9] José A. Carrillo, Robert J. McCann, and Cedric Villani, *Contraction in the 2-wasserstein metric length space and thermalization of granular media*, To appear.
[10] Ennio De Giorgi, *New problems on minimizing movements*, Boundary Value Problems for PDE and Applications (Claudio Baiocchi and Jacques Louis Lions, eds.), Masson, 1993, pp. 81–98.
[11] Ennio De Giorgi, Antonio Marino, and Mario Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **68** (1980), no. 3, 180–187. MR **83m:**49052
[12] Denis Feyel and Ali Süleyman Üstünel, *Measure transport on Wiener space and the Girsanov theorem*, C. R. Math. Acad. Sci. Paris **334** (2002), no. 11, 1025–1028. MR 1 913 729
[13] Wilfrid Gangbo and Robert J. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), no. 2, 113–161. MR **98e:**49102
[14] Juha Heinonen and Pekka Koskela, *Quasiconformal maps in metric spaces with controlled geometry*, Acta Math. **181** (1998), no. 1, 1–61. MR **99j:**30025
[15] Richard Jordan, David Kinderlehrer, and Felix Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17 (electronic). MR **2000b:**35258
[16] Jürgen Jost, *Nonpositive curvature: geometric and analytic aspects*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1997. MR **98g:**53070
[17] Antonio Marino, Claudio Saccon, and Mario Tosques, *Curves of maximal slope and parabolic variational inequalities on nonconvex constraints*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **16** (1989), no. 2, 281–330. MR **91d:**58035
[18] Uwe F. Mayer, *Gradient flows on nonpositively curved metric spaces and harmonic maps*, Comm. Anal. Geom. **6** (1998), no. 2, 199–253. MR **99m:**58067
[19] Robert J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), no. 1, 153–179. MR **98e:**82003
[20] Ricardo H. Nochetto, Giuseppe Savaré, and Claudio Verdi, *A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations*, Comm. Pure Appl. Math. **53** (2000), no. 5, 525–589. MR 1 737 503
[21] Felix Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174. MR **2002j:**35180
[22] Cedric Villani, *Topics in optimal transportation*, AMS, 2003.

(L.Ambrosio) Scuola Normale Superiore, Piazza dei Cavalieri, 56100 Pisa (Italy)
E-mail address: `l.ambrosio@sns.it`

(N.Gigli) Scuola Normale Superiore, Piazza dei Cavalieri, 56100 Pisa (Italy)
E-mail address: `n.gigli@sns.it`

(G.Savaré) *Dipartimento di Matematica, Universitá di Pavia, Via Ferrata, 27100 Pavia (Italy)*
E-mail address: `savare@imati.cnr.it`