[1]

STEEPEST DESCENT FLOWS AND APPLICATIONS TO SPACES OF PROBABILITY MEASURES

LUIGI AMBROSIO – LECTURE NOTES, SANTANDER (SPAIN), JULY 2004

1. Introduction

These notes are the outcome of a series of seminars held in Santander (Spain) in July 2004, concerning some of the results contained in the book [AGS04]. The first part of these lecture notes is devoted to the study of gradient flows in a general metric setting, while the second part is concerned with the metric space of probability measures endowed with the Wasserstein distance. In the first part (section 2) I start by recalling the main concepts of the classical theory of gradient flows in Hilbert spaces. Then, I introduce the definition of steepest descent flow, which generalizes the idea of gradient flow to a purely metric setting. To this aim, I make use of the concepts of metric derivative and local slope (see also [DGMT80, Amb95]). Then, I deduce some useful properties typically satisfied by functionals which are convex along geodesics. Such properties also involve the notion of upper gradient (see [HK98]). These properties ensure the convergence of the implicit Euler time discretization scheme. I also establish the link between our metric framework and the well known results of the classical theory on Hilbert spaces. Finally, I recall the conditions needed to have uniqueness and error estimates of the approximating scheme.

In the second part (section 3) I study the differentiable structure of the Wasserstein space to give an equivalent concept of gradient flow. The Wasserstein distance arises as a useful tool to study the asymptotic behaviour of solutions to certain PDEs enjoying a gradient flow structure with respect to it. Such gradient flow structure is closely related to some logarithmic Sobolev inequalities, which ensure convergence towards the asymptotic states. This point of view has been widely developed in the recent years (see [Tos96, Tos97, JKO98, CT00, Ott01, OV00, DPD02, Agu02, Agu03, CMV] among the others.). The Wasserstein distance could seem very abstract, but in fact it is linked to a very intuitive problem: *the mass transportation problem* which can be formulated in terms of a pile of sand and a hole. The problem is how to transport the pile into the hole with the least possible cost. I start with some notation and definitions concerning probability theory, in order to give the definition of Wasserstein distance. Then, I study the properties and the geodesics of the Wasserstein space, recovering in this framework the concepts of λ-convexity along geodesics. Finally, I focus on the structure of $(\mathcal{P}_2(X), W_2)$ (in some case presenting results for the general space $(\mathcal{P}_p(X), W_p)$). I prove that $(\mathcal{P}_2(X), W_2)$ is a positively curved space, but squared Wasserstein distance is not 2-convex. This shows the necessity to extend the concept to λ-convexity by taking into account the so-called generalized geodesics. I give a list of λ-convex functionals. Finally, I relate this part to the previous one using gradient flows in the Wasserstein space to study evolution PDEs.

[1]Notes taken and further elaborated by María J. Cáceres (Departamento de Matemática Aplicada, Universidad de Granada, 18071 Granada, Spain, email: caceresg@ugr.es) and Marco Di Francesco (Sezione di Matematica per l'Ingegneria, Università dell'Aquila, Piazzale Pontieri, Monteluco di Roio, I67100 L'Aquila (Italy), email: difrance@univaq.it).

All the results presented in these notes are contained, with detailed proofs and greater generality, in [AGS04].

Last but not least, I would like to thank warmly María J. Cáceres and Marco di Francesco for the extensive and nice work they did in typing these notes.

2. Gradient flows and steepest descent flows.

2.1. The classical theory. Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space with norm $||\cdot||$. Let ϕ be a convex, lower semi-continuous functional defined on a dense domain $Dom(\phi) \subset H$. The *subdifferential* of ϕ at a point $u \in Dom(\phi)$ is the set $\partial\phi(u) \subset H$ defined by

$$\partial\phi(u) = \{p \in H : \phi(v) \geq \phi(u) + \langle p, v-u\rangle \ \forall v \in Dom(\phi)\}.$$

Under the above assumptions on the functional ϕ, the subdifferential $\partial\phi(u)$ at some point u is a closed convex subset of H which may be interpreted as the set of all possible 'slopes' of affine hyperplanes touching the graph of ϕ from below at the point u. The set valued mapping $\partial\phi : H \longrightarrow 2^H$ is called the subdifferential of ϕ. We next recall the following standard definition.

Definition 2.1. *Let $0 \leq T_1 < T_2$. A function $u : [T_1, T_2] \longrightarrow H$ is said to be* absolutely continuous *on $[T_1, T_2]$ ($u \in AC\,([T_1, T_2]; H)$) if and only if, for any $\varepsilon > 0$ there exists $\delta > 0$ such that, given any finite set of disjoint subintervals $\{(a_k, b_k)\} \subset [T_1, T_2]$, $k = 1, \dots n$, with $\sum_{k=1}^n (b_k - a_k) < \delta$, the inequality*

$$\sum_{k=1}^n \|u(b_k) - u(a_k)\| < \varepsilon$$

is satisfied. A function $u : (0, +\infty) \longrightarrow H$ is called locally absolutely continuous *on $(0, +\infty)$ ($u \in AC_{loc}\,((0, +\infty); H)$) if and only if $u \in AC\,([T_1, T_2]; H)$ for all $0 < T_1 < T_2$.*

We recall that locally absolutely continuous functions u with values in a Hilbert space are differentiable almost everywhere, as stated in the following proposition (see for instance [Amb95]).

Proposition 2.1. *Let $u \in AC_{loc}\,((0, +\infty); H)$. Then, for $\mathcal{L}^1$–almost all $t \in [0, +\infty)$, the limit*

$$u'(t) = \lim_{h \to 0} \frac{u(t+h) - u(t)}{h}$$

exists in the strong topology of H.

We then recall the classical definition of gradient flow on a Hilbert space.

Definition 2.2. *A function $u \in AC_{loc}\,((0, +\infty); H)$ is a* gradient flow *of the convex, lower semi-continuous functional ϕ if and only if the differential inclusion*

$$u'(t) \in -\partial\phi\,(u(t)) \tag{2.1}$$

is satisfied $\mathcal{L}^1$– almost everywhere with respect to t.

The inclusion (2.1) is usually coupled with the initial condition

$$\lim_{t \downarrow 0} u(t) = u_0 \tag{2.2}$$

and typically the initial datum u_0 is chosen in $Dom(\phi)$, or at least in its closure. The study of the Cauchy problem (2.1)–(2.2) is a classical topic. It is well known that the subdifferential mapping $\partial\phi : H \longrightarrow 2^H$ may be seen as a possibly multivalued nonlinear operator on H, which is easily seen to be *maximal monotone* (cfr. [Bré73]). The theory of existence, uniqueness and regularity of solutions has been developed in [CP72, CL71, Bré73]. In particular, the unique solution $u(t)$ to (2.1)–(2.2) is 'well approximated' on the time interval $[0, T)$ by the solution to the *implicit Euler scheme*

$$\begin{cases} u_N^0 = u_0 \\ \dfrac{u_N^{n+1} - u_N^n}{h} \in -\partial\phi\left(u_N^{n+1}\right), \quad Nh = T, \quad n = 0, \dots, N-1, \end{cases} \tag{2.3}$$

for large values of N. More precisely, one may consider the piecewise linear interpolation $\overline{u}_N(t)$ satisfying

$$\overline{u}_N(nh) = u_N^n, \quad n = 0, \dots, N-1,$$

and show that $\overline{u}_N$ converges (in a suitable way) to the unique solution to (2.1)–(2.2) as $N \to \infty$. Such technique was introduced in [CL71] in order to recover an approximation formula for nonlinear semigroups *generated* by more general monotone operators on Banach spaces. We also mention here that the approximating scheme (2.3) is sometimes replaced by the *variational* formulation

$$\begin{cases} u_N^0 = u_0 \\ u_N^{n+1} \ \text{ minimizes } \ \phi(v) + \dfrac{1}{2h}\|v - u_N^n\|^2, \quad Nh = T, \quad n = 0, \dots, N-1. \end{cases} \tag{2.4}$$

We remark that the second expression in (2.3) is nothing else than the usual Euler equation associated to the functional $\phi(v) + \frac{1}{2h}\|v - u_N^n\|^2$.

2.2. **Steepest descent flows on metric spaces.** Our goal is to rephrase the differential viewpoint of the classical theory into a purely *metric* framework. To perform this task, De Giorgi (see [DGMT80] and, in connection with Euler's scheme, [DG93]) proposed a metric formulation of the gradient flow that we will call *steepest descent flow.* We present De Giorgi's ideas following the presentation of [AGS04], that involves the natural concepts of *metric derivative* and *local slope.* In what follows we shall work on a complete metric space $(\mathcal{S}, d)$.

Definition 2.3. *Let $u : (0, +\infty) \to \mathcal{S}$ be a curve on $\mathcal{S}$.*

(i) *u is said to be* absolutely continuous *in $(a, b) \subset (0, +\infty)$ ($u \in AC((a, b); \mathcal{S})$) if there exists $g \in L^1((a, b); \mathbb{R})$ such that*

$$d(u(s), u(t)) \le \int_s^t g(\tau)d\tau, \qquad a < s < t < b. \tag{2.5}$$

(ii) *u is said to be* locally absolutely continuous *($u \in AC_{loc}((0, +\infty); \mathcal{S})$) if $u \in AC((a, b); \mathcal{S})$ for any $0 < a < b < \infty$.*

(iii) *The* variation *of u on the interval $[a, b] \subset (0, +\infty)$ is*

$$\mathrm{Var}_a^b(u) = \sup\left\{\sum_{i=1}^{n-1} d(u(t_i), u(t_{i-1})) \mid a \le t_1 < \ldots < t_n \le b\right\}.$$

The following theorem is a sort of generalization of Rademacher's theorem for absolutely continuous curves on a complete metric space.

Theorem 2.1. *For any $u \in AC_{loc}((0,+\infty);\mathcal{S})$ the limit*

$$|u'|(t) := \lim_{s \to t} \frac{d(u(s),u(t))}{|s-t|} \tag{2.6}$$

exists for $\mathcal{L}^1$–almost any $t \in [0,+\infty)$.

Proof. Let $T > 0$ be fixed. Let us choose a dense sequence $\{x_n\}_{n=0}^{\infty}$ in $[0,T]$, and let us define, for any $n \in \mathbb{N}$, the real valued function $\psi_n(t) = d(u(t),x_n)$. The assumptions on u and the triangular inequality imply

$$|\psi_n(t) - \psi_n(s)| \le d(u(s),u(t)) \le \int_s^t g(\tau)d\tau.$$

Therefore, the functions ψ_n are absolutely continuous, and hence differentiable $\mathcal{L}^1$–almost everywhere. We can then define the function

$$m(t) = \sup_n |\psi_n'(t)|$$

for almost all $t \in [0,T]$. Moreover, since almost all $t \in [0,T]$ are Lebesgue points for the L^1 function g, we have $|\psi_n'(t)| \le g(t)$ for almost all t and for any $n \in \mathbb{N}$, and therefore $m \in L^1(a,b;\mathbb{R})$. We next prove that the limit in (2.6) exists and $m(t) = |u'|(t)$ $\mathcal{L}^1$–almost everywhere. For any $n \in \mathbb{N}$, the triangular inequality yields

$$\liminf_{h \to 0} \frac{d(u(t+h),u(t))}{|h|} \ge \lim_{h \to 0} \frac{|\psi_n(t+h) - \psi_n(t)|}{|h|} = |\psi_n'(t)|, \quad \text{for } \mathcal{L}^1\text{–almost any } t \in [0,T].$$

The definition of m then implies the inequality

$$\liminf_{h \to 0} \frac{d(u(t+h),u(t))}{|h|} \ge m(t).$$

Now, the triangular inequality and the density of the sequence $\{x_n\}_{n\in\mathbb{N}}$ imply

$$d(u(s),u(t)) = \sup_{n\in\mathbb{N}} |d(u(s),x_n) - d(x_n,u(t))|\, leq \sup_{n\in\mathbb{N}} \int_s^t |\psi_n'(\tau)|d\tau \le \int_s^t m(\tau)d\tau,$$

for any $s,t \in [0,T]$, $s \le t$. Hence, at any Lebesgue point for the function m we have

$$\limsup_{h \to 0} \frac{d(u(t+h),u(t))}{|h|} \le m(t),$$

and the proof is complete. □

Definition 2.4. *The limit*

$$|u'|(t) = \lim_{s \to t} \frac{d(u(s),u(t))}{|s-t|} \tag{2.7}$$

is called metric derivative *of the curve u at the point t.*

From the above computations it is easily seen that, if u is an absolutely continuous curve, then its metric derivative $|u'|$ is the minimal g satisfying estimate (2.5). Moreover (see [AT04] for the details) one can prove the identity

$$\mathrm{Var}_a^b(u) = \int_a^b |u'|(\tau)d\tau.$$

Let us introduce the concept of local slope, which will, in some sense, replace the subdifferential in our reformulation of gradient flow.

Definition 2.5. *The* local slope *of a functional ϕ at a point $u \in Dom(\phi)$ is 0 if u is an isolated point of $\mathcal{S}$, otherwise is given by*

$$\left|\partial^- \phi\right|(u) := \limsup_{v \to u} \frac{[\phi(u) - \phi(v)]^+}{d(u,v)}.$$

We remark that the local slope of ϕ at u equals zero if u is a minimal point of ϕ.

So far we have devoted our attention to generalizing the ingredients of the gradient flow formulation, in such a way that they make sense in a purely metric framework. Next, we describe the heuristic idea (due to De Giorgi) which leads us to the new definition of gradient flow. Suppose we are given a Gateaux differentiable functional ϕ on a Hilbert space H. Suppose $u(t)$ is a classical solution of the gradient flow equation

$$u'(t) = -\nabla\phi(u). \tag{2.8}$$

By taking the modulus in both sides we obtain the scalar equation

$$\left\|u'(t)\right\| = \left\|\nabla\phi(u)\right\|, \tag{2.9}$$

which may possibly make sense in a metric framework if we make use of the metric derivative and the local slope. Clearly, such step produces a loss of information, since the two vectors $u'(t)$ and $-\nabla\phi(u(t))$ need not to be parallel in order to satisfy (2.9). However, we can recover this information by looking at the time derivative of $\phi(u(t))$. Indeed, $u'(t)$ and $-\nabla\phi(u(t))$ have same direction if and only if

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

On the other hand, (2.9) holds if and only if

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

Hence, (2.8) can be equivalently rewritten as

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

Finally, the Young inequality trivially implies that (2.8) is equivalent to

$$\frac{d}{dt}\phi(u(t)) \leq -\frac{1}{2}\|u'(t)\|^2 - \frac{1}{2}\|\nabla\phi(u(t))\|^2. \tag{2.10}$$

The equivalence between (2.10) and (2.8) allows us to generalize the notion of gradient flow as follows.

Definition 2.6 (Steepest descent flow). *Let $(\mathcal{S}, d)$ be a complete metric space. A locally absolutely continuous curve $u : (0, +\infty) \to \mathcal{S}$ is a* steepest descent flow *for the functional ϕ (or a* curve of maximal slope *for ϕ) if $\phi \circ u$ is $\mathcal{L}^1$–almost equal to a non–increasing map ψ and*

$$\psi'(t) \leq -\frac{1}{2}|u'|^2(t) - \frac{1}{2}|\partial^-\phi|^2(u(t)), \tag{2.11}$$

where $|u'|(t)$ is the metric derivative and $|\partial^-\phi|(u)$ is the local slope.
We say that $\overline{u}$ is the starting point of the curve u if $\lim_{t\to 0} u(t) = \overline{u}$.

2.3. **The geodesically convex case and the notion of upper gradient.** The existence and uniqueness of solutions to the steepest descent flow of a given functional with given starting point can be achieved in many different ways, depending on the specific structure of the functional and the ambient metric space. Here we focus on the case when the functional satisfies the so called λ–*convexity along geodesics*. We shall be more precise about the hypothesis required for the metric space $\mathcal{S}$ and for the functional ϕ in the sequel. For the present moment, $(\mathcal{S}, d)$ denotes a complete metric space, while ϕ is simply a functional on $\mathcal{S}$.

Definition 2.7 (Geodesics in metric spaces). *Let x, y be points in $\mathcal{S}$. Let us denote by Γ_x^y the set of all absolutely continuous curves $\gamma : [a,b] \to \mathcal{S}$ such that $\gamma(a) = x$ and $\gamma(b) = y$. A curve $u \in \Gamma_x^y$ is a* geodesic *connecting x and y if*

$$\mathrm{Var}_a^b(u) = \min_{\gamma \in \Gamma_x^y} \mathrm{Var}_a^b(\gamma).$$

We recall a basic fact about geodesics on metric spaces in the following theorem (see [AT04] for the proof).

Theorem 2.2 (Reparameterization). *Let $\gamma : [a,b] \to \mathcal{S}$ be absolutely continuous and let $L = Var_a^b(\gamma)$ be its variation on $[a,b]$. Then there exists a Lipschitz curve $\widetilde{\gamma} : [0,L] \to \mathcal{S}$ such that $|\widetilde{\gamma}'| \equiv 1$ a.e. in $[0,L]$ and $\gamma([a,b]) = \widetilde{\gamma}([0,L])$.*

We now impose the first significant assumption on the metric space $\mathcal{S}$, namely, we require that for any two points of the metric space there exists an absolutely continuous curve (a geodesic) connecting the points, whose length equals the distance between them.

Definition 2.8 (Length space). *A complete metric space $(\mathcal{S}, d)$ is called a* length space *if, for any $x, y \in \mathcal{S}$, there exists an absolutely continuous curve $u : [a,b] \to \mathcal{S}$ connecting x and y such that*

$$\mathrm{Var}_a^b(u) = d(x,y).$$

The condition above can also be written as $\int_a^b |u'|\,dt = d(x,y)$. Notice that the definition of length space could also be given in a weaker form, by saying that the infimum of $\mathrm{Var}_a^b(u)$ among all admissible u is equal to $d(x,y)$. The two definitions are easily seen to be equivalent if bounded and closed subsets of $\mathcal{S}$ are compact.

As a simple consequence of Theorem 2.2, we have the following

Corollary 2.1. *Let $(\mathcal{S}, d)$ be a length space. Then, for any $x, y \in \mathcal{S}$, there exists a geodesic $u : [0,1] \to \mathcal{S}$ connecting x and y such that*

$$d(u(s), u(t)) = (t-s)d(x,y), \quad \textit{for all } s,t \in [0,1] \quad s \le t. \tag{2.12}$$

The curve u provided by the above corollary is called a *constant speed geodesic* connecting x and y, the reason of that definition being the identity

$$|u'|(t) \equiv d(x,y), \quad \text{for all } \ t \in [0,1].$$

The following concept was introduced independently (and for different purposes) by Jost [Jos97], Mayer [May98] and McCann [McC97].

Definition 2.9 (λ–convexity along geodesics). *We say that a functional ϕ on $\mathcal{S}$ is λ–convex along geodesics if*

$$\phi(\gamma(t)) \le t\phi(\gamma(1)) + (1-t)\phi(\gamma(0)) - \frac{\lambda}{2}t(1-t)d^2(\gamma(0),\gamma(1)) \tag{2.13}$$

for every constant speed geodesic $\gamma : [0,1] \to \mathcal{S}$.

The following definition is due to Heinonen and Koskela (see [HK98]) and it constitutes a sort of "weak" formulation of the modulus of the derivative.

Definition 2.10 (Upper gradient). *A function $g : \mathcal{S} \to \mathbb{R}$ is an* upper gradient *for the functional ϕ if, for any $T \ge 0$ and for any absolutely continuous curve $v \in AC\,(0,T;\mathcal{S})$ such that*

$$g(v(\cdot))|v'|(\cdot) \in L^1\,(0,T;\mathcal{S}) \tag{2.14}$$

the following inequality holds

$$|\phi(v(0)) - \phi(v(T))| \le \int_0^T g(v(\tau))|v'|(\tau)d\tau. \tag{2.15}$$

Theorem 2.3. *Let ϕ be a lower semi–continuous functional on the length space $(\mathcal{S}, d)$ which is λ–convex along constant speed geodesics. Then, the local slope $|\partial^-\phi|$ is an upper gradient for ϕ.*

Proof. To simplify the proof we assume $\lambda \ge 0$. The proof of the general case can be obtained by means of simple modifications.

Step 1. We first prove that the local slope $|\partial^-\phi|$ is actually (due to the convexity assumption) a *global* slope, i. e.

$$\limsup_{v\to u} \frac{[\phi(u)-\phi(v)]^+}{d(u,v)} = \sup_{v\neq u} \frac{[\phi(u)-\phi(v)]^+}{d(u,v)} =: I\phi(u) \tag{2.16}$$

Of course, the inequality $|\partial^-\phi(u)| \le I\phi(u)$ is obvious. Then, let $v \neq u$ and let us choose a constant speed geodesic $v(t)$ connecting u to v. By the convexity assumption (2.13), thanks to (2.12), we easily obtain the inequality

$$\frac{\phi(u)-\phi(v(t))}{d(v(t),u)} \ge \frac{\phi(u)-\phi(v)}{d(u,v)}.$$

By taking the positive part in the above inequality and by letting t tend to zero, we get

$$\frac{[\phi(u)-\phi(v)]^+}{d(u,v)} \le \limsup_{v\to u} \frac{[\phi(u)-\phi(v)]^+}{d(u,v)} \quad \text{for all } \; v \in \mathcal{S},$$

and the inequality $|\partial^-\phi(u)| \ge I\phi(u)$ is proven.

Step 2. We next prove that $I\phi(u)$ is an upper gradient for ϕ (without any convexity assumption). For simplicity we prove relation (2.15) in definition (2.10) for all absolutely continuous curves $v : [0,1] \to \mathcal{S}$ satisfying (2.14) (with g replaced by $I\phi$) and such that $|v'|(t) \equiv 1$. Hence, condition (2.14) reduces to $I\phi(v(\cdot)) \in L^1\,(0,T;\mathcal{S})$. The general case then follows after a reparameterization (see Theorem 2.2). By definition of $I\phi$ we have, for $0 \le s \le t \le 1$,

$$\begin{aligned} |\phi(v(s)) - \phi(v(t))| &\le \max\{I\phi(v(s)), I\phi(v(t))\}d(v(s),v(t)) \le \max\{I\phi(v(s)), I\phi(v(t))\}\mathrm{Var}_s^t(v) \\ &= \max\{I\phi(v(s)), I\phi(v(t))\}\int_s^t |v'|(\tau)d\tau = \max\{I\phi(v(s)), I\phi(v(t))\}(t-s). \end{aligned} \tag{2.17}$$

Now, the above condition implies that the function $\phi \circ v$ belongs to the Sobolev space $W^{1,1}(0,1;\mathbb{R})$. Indeed, by a difference quotients argument one can prove that the signed measure $(\phi \circ v)'$ (in the sense of distributions) is absolutely continuous w.r.t. $\mathcal{L}^1$, and that its total variation is less than $2\|(I\phi)\circ v\|_{L^1(0,1)}$ (see Lemma 1.2.6 in [AGS04]). Moreover, since $\phi \circ v$ is lower semicontinuous, and since

$$\limsup_{\varepsilon\to 0} \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} (\phi \circ v(s+r) - \phi \circ v(s))dr \le \limsup_{\varepsilon\to 0} \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} I\phi \circ v(s+r)|r|dr$$
$$\le \limsup_{\varepsilon\to 0} \frac{1}{2}\int_{-\varepsilon}^{\varepsilon} I\phi \circ v(s+r)dr = 0,$$

we obtain that the function $\phi \circ v$ is the continuous representative in its equivalence class in $W^{1,1}$, hence it is absolutely continuous. A similar argument proves also continuity at $t = 0$ and $t = T$. Finally, from relation (2.17) one easily deduces that $(\phi \circ v)' \le (I\phi) \circ v$ almost everywhere, which implies the desired estimate

$$|\phi(v(0)) - \phi(v(T))| \le \int_0^T I\phi(v(\tau))d\tau.$$

□

2.4. **Consistency with the classical theory and uniqueness.**

Theorem 2.4. *Let $(H, (\cdot,\cdot))$ be a Hilbert space. Then, steepest descent flows and gradient flows for λ–convex functionals coincide.*

Proof.

STEP 1. Let us fix $u \in H$. We start by proving the following *fundamental* relation,

$$|\partial^-\phi|(u) = \min\{\|p\| \mid p \in \partial\phi(u)\}, \tag{2.18}$$

linking the "differential" viewpoint (i.e. the subdifferential) to the "variational" one (i.e. the slope). By the definition of the subdifferential $\partial\phi(u)$, we easily get the inequality

$$\frac{\phi(u) - \phi(v)}{\|u - v\|} \le \frac{\langle p, u - v\rangle}{\|u - v\|}, \quad \text{for all } \ p \in \partial\phi(u), \ \ v \in H, \ v \ne u.$$

Hence, by taking the positive parts and the supremum over $v \ne u$, we obtain

$$\sup_{v\ne u} \frac{[\phi(u) - \phi(v)]_+}{\|u - v\|} \le \sup_{v\ne u} \frac{\langle p, u - v\rangle}{\|u - v\|} = \|p\|, \quad p \in \partial\phi(u).$$

Since ϕ is convex, in view of (2.16), we have

$$|\partial^-\phi|(u) \le \min\{\|p\| \mid p \in \partial\phi(u)\}.$$

To prove the opposite inequality, we need to find an element $p \in \partial\phi(u)$ such that

$$\|p\| \le |\partial^-\phi|(u) = I\phi(u).$$

By definition of global slope, we have

$$|\partial^-\phi|(u) = I\phi(u) \ge \frac{[\phi(u) - \phi(v)]_+}{\|u - v\|} \ge \phi(u) - \phi(v)\|u - v\|,$$

which implies

$$-|\partial^-\phi|(u)\|u - v\| \le \phi(v) - \phi(u), \quad \text{for all } v \in H.$$

Thus, the two convex subsets of $H \times \mathbb{R}$

$$\mathcal{A} = \{(v,r) \in H \times \mathbb{R} \mid r \geq \phi(v) - \phi(u)\}, \qquad \mathcal{B} = \{(v,r) \in H \times \mathbb{R} \mid r < -I\phi(u)\|u-v\|\}$$

are disjoint. Since $\mathcal{B}$ is open, we can apply the first geometric version of Hahn–Banach theorem, which provides the existence of a $p \in H$ and of a real constant α such that

$$-|\partial^- \phi|(u)\|u-v\| \leq \langle p, u-v\rangle + \alpha \leq \phi(v) - \phi(u), \quad \text{for all } v \in H. \tag{2.19}$$

Taking $v = u$ we get $\alpha = 0$. The first inequality in (2.19) implies

$$|\partial^- \phi|(u) \geq \sup_{v \neq u} \frac{\langle p, v-u\rangle}{\|u-v\|} = \|p\|,$$

while the second inequality in (2.19) means that $p \in \partial\phi(u)$. Therefore, (2.18) is proven.

Step 2. Let $u : [0,T] \to H$ be a gradient flow for ϕ. This means that u is differentiable $\mathcal{L}^1$–almost everywhere and

$$u'(t) \in -\partial\phi(u), \quad \text{at almost any } t \in [0,T].$$

Moreover, from the classical theory of gradient flows (see [Bré73]), one can prove that $\phi \circ u$ is absolutely continuous (even locally Lipschitz), and hence differentiable almost everywhere. By taking the limit in the difference quotient $\frac{\phi(u(t+h)) - \phi(u(t))}{h}$ as $h \to 0^+$ and $h \to 0^-$ respectively, thanks to the definition of subdifferential, we have

$$\frac{d}{dt}\phi \circ u(t) = -\|u'(t)\|^2$$

at any differentiability point of $\phi \circ u$. But, since $u'(t) \in -\partial\phi(u(t))$, then we have $\|u'(t)\| \geq |\partial^- \phi|(u(t))$. Therefore,

$$\frac{d}{dt}\phi \circ u(t) \leq -\frac{1}{2}\|u'(t)\|^2 - \frac{1}{2}|\partial^- \phi|^2(u(t)),$$

which is the definition of steepest descent flow.

Conversely, let $u : [0,T] \to H$ be a steepest descent flow for ϕ. From (2.18) we know that there exists a $p_0(t) \in -\partial\phi(u(t))$ such that

$$\frac{d}{dt}\phi \circ u(t) \leq -\frac{1}{2}\|u'(t)\|^2 - \frac{1}{2}\|p_0(t)\|^2.$$

As above, since $\phi \circ u$ is absolutely continuous, we have for almost every t

$$\frac{d}{dt}\phi \circ u(t) = \langle p_0(t), u'(t)\rangle,$$

and therefore

$$\langle p_0(t), u'(t)\rangle \leq -\frac{1}{2}\|u'(t)\|^2 - \frac{1}{2}\|p_0(t)\|^2,$$

which implies $p_0(t) = -u'(t)$ and the proof is complete. □

The assertion in the above theorem can be easily extended to λ–convex functionals. Indeed, in this case geodesics coincide with straight lines, and condition (2.13) reads

$$\phi(tu + (1-t)v) \leq t\phi(u) + (1-t)\phi(v) - \frac{\lambda}{2}t(1-t)\|u-v\|^2, \quad u, v \in H, \quad t \in (0,1).$$

Hence, λ–convexity for nonnegative λ implies standard convexity, while in general one can consider the auxiliary functional

$$\psi(u) = \phi(u) - \frac{\lambda}{2}\|u\|^2$$

which turns out to be convex in the standard sense. This fact is a trivial consequence of the parallelogram identity in Hilbert spaces, which implies the relation

$$\|tu + (1-t)v\|^2 = t\|u\|^2 + (1-t)\|v\|^2 - t(1-t)\|u-v\|^2.$$

Notice also that the relation above implies that $\|\cdot - u\|^2/2$ is 1–convex for any $u \in H$.

Our next purpose is to show that gradient flows of λ–convex functionals in Hilbert spaces are unique. To this aim we first prove the following lemma.

Lemma 2.1. *Let ϕ be a λ–convex functional on H. Let $u, v \in Dom(\phi)$, $\xi \in \partial\phi(u)$, $\eta \in \partial\phi(v)$. Then*

$$\langle \xi - \eta \rangle \geq \lambda \|u - v\|^2.$$

Proof.

As a trivial consequence of the definition of subdifferential, one can prove that

$$\langle \xi - \eta, u - v \rangle \geq 0 \tag{2.20}$$

whenever $\xi \in \partial\phi(u)$, $\eta \in \partial\phi(v)$ and ϕ is convex in the standard sense. In the general λ–convex case, we consider again the auxiliary convex functional $\psi(u) = \phi(u) - \frac{\lambda}{2}\|u\|^2$. We observe that $\xi + \lambda u \in \partial\psi(u)$ and $\eta + \lambda v \in \partial\psi(v)$. Hence, we conclude the proof by applying (2.20) to the convex functional ψ. □

Theorem 2.5. *Let ϕ be a λ–convex functional on H. Then, there exists at most one solution to the Cauchy problem*

$$u'(t) \in -\partial\phi\left(u(t)\right), \quad \lim_{t\downarrow 0} u(t) = \bar{u}.$$

Proof.

Let u_1, u_2 be two gradient flows for ϕ. Then, we can write

$$\left(\frac{1}{2}\|u_1 - u_2\|^2\right)' = \langle u_1 - u_2, u_1' - u_2' \rangle$$

and from Lemma 2.1 we obtain

$$\left(\frac{1}{2}\|u_1 - u_2\|^2\right)' \leq -\lambda \|u_1 - u_2\|^2,$$

which implies

$$\|u_1(t) - u_2(t)\| \leq \lim_{s\downarrow 0} e^{-\lambda s}\|u_1(s) - u_2(s)\| = 0,$$

concluding the proof. □

2.5. **The time step approximation.** From now on we shall work with a functional ϕ defined on a complete length space $(\mathcal{S}, d)$ and satisfying the following assumptions:

(i) ϕ is lower semi–continuous,
(ii) ϕ coercive, i.e., for any $t \in \mathbb{R}$ the sub–level $\{x \in \mathcal{S} \mid \phi(x) \leq t\}$ is compact,
(iii) ϕ λ–convex along geodesics for some $\lambda \in \mathbb{R}$.

Remark 2.1 (A more general coercitivity condition)**.** The coercivity assumption (iii) has been chosen to simplify the exposition, but it turns out to be too restrictive for many applications (for example, the case when $(\mathcal{S}, d)$ is the Wasserstein space of probability measures, discussed in the next sections). So, as usual in Functional Analysis, one can assume instead that there exists a topology σ with the following properties:

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

$$(u_n, v_n) \overset{\sigma}{\rightharpoonup} (u, v) \quad \Longrightarrow \quad \liminf_{n \to \infty} d(u_n, v_n) \geq d(u, v). \tag{2.21}$$

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

$$\sup_{n,m} d(u_n, u_m) < +\infty, \quad u_n \overset{\sigma}{\rightharpoonup} u \quad \Longrightarrow \quad \liminf_{n \to \infty} \phi(u_n) \geq \phi(u). \tag{2.22}$$

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

$$m^* := \inf_{v \in \mathcal{S}} \phi(v) + \frac{1}{2\tau} d^2(v, u_*) > -\infty. \tag{2.23}$$

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

$$\begin{gathered} \text{every sequence } (u_n) \subset \mathcal{S} \text{ with } \sup_n \phi(u_n) < +\infty, \quad \sup_{n,m} d(u_n, u_m) < +\infty \\ \text{admits a } \sigma\text{-convergent subsequence.} \end{gathered} \tag{2.24}$$

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

$$|\partial\phi|(u) = \inf\Big\{ \liminf_{n \to +\infty} |\partial\phi|(u_n) \ : \ u_n \overset{\sigma}{\rightharpoonup} u, \ \sup_n \{d(u_n, u), \phi(u_n)\} < +\infty \Big\}. \tag{2.25}$$

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. In the case of the Wasserstein space of probability measures the topology σ is just the topology induced by a weak convergence, in the duality either with continuous and compactly supported functions, or in the duality with continuous and bounded functions (see Theorem 3.2).

Next we introduce the main ingredients of our time step approximation. Given $\tau > 0$, we define, for a fixed $U \in \mathcal{S}$, the modified functional

$$\Phi(\tau, U, V) := \frac{1}{2\tau} d(U, V) + \phi(V).$$

Let us fix the initial point of the gradient flow $u_0 \in \mathcal{S}$. In the spirit of the variational formulation (2.4) of the classical gradient flow, we define the recursive scheme

$$\begin{cases} U_\tau^0 = u_0, \\ \text{Given } U_\tau^1, \ldots, U_\tau^{n-1}, \text{ find } U_\tau^n \in \mathcal{S} \text{ such that} \\ \Phi(\tau, U_\tau^{n-1}, U_\tau^n) \leq \Phi(\tau, U_\tau^{n-1}, V) \text{ for all } V \in \mathcal{S}. \end{cases} \tag{2.26}$$

The (possibly multivalued) operator which provides all the solutions U_τ^n of the variational problem (2.26) for a given U_τ^{n-1} is called *resolvent operator* and it is denoted by $J_\tau[\cdot]$. More precisely we have

$$J_\tau[U] = \text{argmin } \Phi(\tau, U, \cdot).$$

Hence, a sequence $\{U_\tau^n\}_n$ solves the recursive scheme (2.26) if and only if $U_\tau^n \in J_\tau[U_\tau^{n-1}]$ for all $n \geq 1$. We observe that the assumptions previously required for the functional ϕ guarantee the existence of minimizers in (2.26). In particular, there exists at least one sequence $\{U_\tau^n\}_n$ solving the scheme. In the sequel we shall make use of several continuous versions of the sequence $\{U_\tau^n\}_n$; for the present moment we define, for any $\tau > 0$, the piecewise constant interpolation $\overline{u}_\tau(t) = U_\tau^n$ as $t \in ((n-1)\tau, n\tau)$.

Next we define, for positive τ, the *Moreau–Yosida approximation* of ϕ

$$\phi_\tau(u) = \inf \left\{ \Phi(\tau, u, v) \mid v \in \mathcal{S} \right\}.$$

The resolvent operator and the Moreau–Yosida approximation are variational reformulations of analogous ingredients of the classical theory, namely the resolvent operator $(\mathbb{I} + \tau\partial\phi)^{-1}$ and the Yosida transformation $A_\tau(u) = \frac{u - J_\tau[u]}{\tau}$ respectively. In the classical Hilbert case the following convergence result holds (see [BA89, Ru96, NSV00] and Chapter 4 of [AGS04], in a much more general context).

Theorem 2.6 (Error estimate in the classical case). *Let $u(t)$ be the unique gradient flow of the convex, lower semi–continuous and coercive functional ϕ defined on the Hilbert space H, with initial point u_0. Then*

$$\|\overline{u}_\tau(t) - u(t)\|^2 \leq \tau(\phi(u_0) - \inf \phi).$$

We are now ready to state our main theorem.

Theorem 2.7. *Let $\mathcal{S}$ be a complete length space. Let ϕ be a functional on $\mathcal{S}$ satisfying assumptions (i)–(iii) stated above. Let U_τ^n be a sequence solving the recursive scheme (2.26) and let $\overline{u}_\tau$ be the corresponding piecewise constant curve defined above. Then, there exist a sequence $\tau_n \to 0$ and an absolutely continuous function u such that $\overline{u}_{\tau_n} \to u$ locally uniformly in $[0, +\infty)$. Moreover,*

(a) *$u \in Lip_{loc}([0, +\infty); \mathcal{S})$ and there exist the right metric derivative $|u'_+|(t)$ for any $t \geq 0$ and the right derivative $(\phi \circ u)'_+$ for any $t \geq 0$.*
(b) *$(\phi \circ u)'_+(t) = -|u'_+|^2(t) = -|\partial^- \phi|^2(u(t))$ for any $t \geq 0$.*

In particular, u is a steepest descent flow for the functional ϕ and the following energy identity holds,

$$\frac{1}{2}\int_s^t |u'|^2(\sigma)d\sigma + \frac{1}{2}\int_s^t |\partial^- \phi(u(\sigma))|^2 d\sigma = \phi(u(s)) - \phi(u(t)). \tag{2.27}$$

The right metric derivative in the above theorem is trivially defined by taking the right limit $s \downarrow t$ in (2.7).

2.6. **Existence of curves of maximal slope.** The aim of this subsection is to give a sketch of the proof of Theorem 2.7, at least for the convergence part. To perform this task we first prove some compactness of the family $\{\overline{u}_\tau\}_{\tau>0}$. Then, we shall prove that the limit point is a curve of maximal slope of the functional ϕ by means of an energy inequality. We start with the following proposition.

Proposition 2.2 (Compactness). *There exist a sequence $\tau_n \to 0$ and a curve*

$$u \in C^{0,\frac{1}{2}}([0,+\infty);\mathcal{S})$$

such that $\overline{u}_{\tau_n} \to u$ uniformly on compact intervals.

Proof.
By definition of the Moreau–Yosida approximation ϕ_τ we have the simple estimate

$$\phi_\tau(U_\tau^n) = \min_{V\in\mathcal{S}}\left[\frac{1}{2\tau}d^2(V,U_\tau^n)+\phi(V)\right] \le \frac{1}{2\tau}d^2(U_\tau^n,U_\tau^n)+\phi(U_\tau^n) = \phi(U_\tau^n).$$

Therefore, we obtain the following discrete energy inequality

$$\frac{1}{2\tau}d^2(U_\tau^{n+1},U_\tau^n) = \phi_\tau(U_\tau^n) - \phi(U_\tau^{n+1}) \le \phi(U_\tau^n) - \phi(U_\tau^{n+1}),$$

which implies

$$\sum_{n=0}^{\infty}\frac{1}{2\tau}d^2(U_\tau^{n+1},U_\tau^n) \le \phi(u_0). \tag{2.28}$$

The above inequality (2.28) provides a uniform bound for the curves $\{\overline{u}_\tau\}_{\tau<1}$ on compact intervals $[0,T]$. To see this, let $0<\tau<1$, let $t\in((n-1)\tau,n\tau)$, we have $\overline{u}_\tau(t) = U_\tau^n$ and consequently

$$d^2(\overline{u}_\tau(t),u_0) \le n\sum_{k=0}^{n} d^2(U_\tau^{k+1},U_\tau^k) \le 2(t+\tau)\phi(u_0) \le 2(T+1)\phi(u_0).$$

Moreover, for positive times $t_1<t_2$ with $t_1\in((m-1)\tau,m\tau]$, $t_2\in((n-1)\tau,n\tau]$, estimate (2.28) implies

$$d^2(\overline{u}_\tau(t_1),\overline{u}_\tau(t_2)) \le (n-m)\sum_{k=m}^{n-1} d^2(U_\tau^{k+1},U_\tau^k) \le 2(n-m)\tau\phi(u_0) \le 2(t_2-t_1+\tau)\phi(u_0).$$

Hence, we have

$$\limsup_{\tau\to 0} d^2(\overline{u}_\tau(t_1),\overline{u}_\tau(t_2)) \le 2(t_2-t_1)\phi(u_0)$$

uniformly with respect to t_1, t_2. A slight modification of Arzelà's Theorem yields compactness of the family of curves $\{\overline{u}_\tau(\cdot)\}_\tau$, together with the fact that any limit curve is Hölder continuous. □

We now aim to prove the consistency of the recursive scheme (2.26). Let $\{\overline{u}_{\tau_i}\}_i$ be the sequence provided by the above proposition. We wish to prove that u is the desired curve of maximal slope. We first prove the following discrete energy inequality.

Proposition 2.3 (Discrete energy inequality).

$$\frac{1}{2\tau}d^2\left(U_\tau^n,U_\tau^{n+1}\right)+\frac{1}{2}\int_0^\tau \frac{1}{r^2}d^2\left(U_\tau^r,U_\tau^n\right)dr \le \phi\left(U_\tau^n\right)-\phi\left(U_\tau^{n+1}\right), \tag{2.29}$$

where $U_\tau^r \in J_r(U_\tau^n)$, for all $r\in(0,\tau)$.

Proof.

We start with the simple identity

$$\phi(U_\tau^n) - \phi_\tau(U_\tau^n) = \phi(U_\tau^n) - \phi(U_\tau^{n+1}) - \frac{1}{2\tau} d^2(U_\tau^n, U_\tau^{n+1}). \tag{2.30}$$

Next, we prove that for any fixed $w \in \mathcal{S}$ the function $\tau \mapsto \phi_\tau(w)$ is monotone increasing as $\tau \searrow 0$ and $\phi(w) = \lim_{\tau\searrow 0} \phi_\tau(w)$. To see this, for $u_\tau \in J_\tau[w]$ we have

$$\phi_\tau(w) = \phi(u_\tau) + \frac{1}{2\tau} d^2(w, u_\tau).$$

Hence, since $\phi_\tau(w) \le \phi(w)$, we deduce $d(u_\tau, w) \to 0$ as $t \searrow 0$. Then, the lower semicontinuity of ϕ implies

$$\phi(w) \le \liminf_{\tau\searrow 0} \phi(u_\tau) \le \liminf_{\tau\searrow 0} \Phi(\tau, w, u_\tau) = \liminf_{\tau\searrow 0} \phi_\tau(w) \le \limsup_{\tau\searrow 0} \phi_\tau(w) \le \phi(w).$$

The monotonicity of $\tau \mapsto \phi_\tau(w)$ is a trivial consequence of the definition of ϕ_τ. A consequence of the above facts is the inequality

$$\phi(w) - \phi_\tau(w) \ge -\int_0^\tau \frac{d}{dr} \phi_r(w) dr. \tag{2.31}$$

Now, since $\Phi(r+h, w, u_{r+h}) \le \Phi(r+h, w, u_r)$, we have

$$\begin{aligned} &\phi_{r+h}(w) - \phi_r(w) \le \Phi(r+h, w, u_r) - \Phi(r, w, u_r) \\ &= \phi(u_r) + \frac{d^2(u_r, w)}{2(r+h)} - \phi(u_r) - \frac{d^2(u_r, w)}{2r} = -\frac{h}{2r(r+h)} d^2(u_r, w), \quad r, r+h > 0. \end{aligned}$$

Therefore, we deduce

$$\frac{d}{dr} \phi_r(w) = -\frac{1}{2r^2} d^2(u_r, w) \tag{2.32}$$

at any differentiability point of $s \mapsto \phi_s(w)$. Finally, (2.32) and (2.31) with $w = U_\tau^n$ imply

$$\phi(U_\tau^n) - \phi_\tau(U_\tau^n) \ge \frac{1}{2} \int_0^\tau \frac{1}{r^2} d^2(U_\tau^r, U_\tau^n) dr,$$

which we can put into (2.30) to obtain the desired estimate (2.29).

□

One can also check (see [AGS04]) that $r \mapsto \phi_r(w)$ is locally Lipschitz, hence the argument above gives that actually equality holds in (2.29) (but only the inequality $\le$ will play a role in the sequel). The term U_τ^r for $r \in (0, \tau)$ in (2.29) is one the main tools of the present theory, and it deserves a definition.

Definition 2.11 (De Giorgi variational interpolation). *Let $\{U_\tau^n\}$ be a solution of the variational scheme (2.26). We denote by $\widetilde{U}_\tau(\cdot) : [0, +\infty) \to \mathcal{S}$ any interpolation of the discrete values U_τ^n satisfying*

$$\widetilde{U}_\tau(t) \in J_\delta[U_\tau^{n-1}] \qquad \textit{if } t = (n-1)\tau + \delta, \quad 0 < \delta < \tau.$$

The above interpolation need not be continuous, it is only right continuous at points t such that t is an integer multiple of τ. In the sequel we shall also make use of the following continuous interpolation.

Definition 2.12. *The continuous interpolation $\overline{U}_\tau(\cdot) : [0,+\infty) \to \mathcal{S}$ is determined by the conditions*

$$|\overline{U}'_\tau|(t) = \frac{d(U^n_\tau, U^{n-1}_\tau)}{\tau} \qquad as \quad t \in [(n-1)\tau, n\tau).$$

The existence of a (absolutely) continuous interpolation is ensured by the assumption that the ambient metric space is a length space: it suffices to interpolate between U^n_τ and U^{n-1}_τ with a constant speed geodesic (notice however that the argument used in [AGS04] does not need the length space assumption).

We shall now modify the previous discrete energy inequality (2.29) by taking into account the local slope $|\partial^-\phi|$ of the functional ϕ.

Proposition 2.4 (Slope estimate)**.** *Let $w \in \mathcal{S}$ and let $u \in J_\tau[w]$. We have*

$$|\partial^-\phi|(u) \leq \frac{d(u,w)}{\tau}. \tag{2.33}$$

Proof.

Since $u \in J_\tau[w]$, we have

$$\frac{1}{2\tau}d^2(u,w) + \phi(u) \leq \frac{1}{2\tau}d^2(v,w) + \phi(v), \quad \text{for all } v \in \mathcal{S}.$$

Hence, the triangular inequality implies

$$\begin{aligned}\frac{\phi(u)-\phi(v)}{d(u,v)} &\leq \frac{\left(d^2(v,w) - d^2(u,w)\right)}{2\tau d(u,v)} \leq \frac{\left(d(v,w)-d(u,w)\right)\left(d(v,w)+d(u,w)\right)}{2\tau d(u,v)} \\ &\leq \frac{d(u,v)\left(d(v,w)+d(u,w)\right)}{2\tau d(u,v)} = \frac{d(v,w)+d(u,w)}{2\tau}.\end{aligned}$$

Finally, by taking the lim sup as $v \to u$ we obtain the desired inequality (2.33). $\square$

In view of the above proposition and thanks to definition 2.11, we can rewrite the discrete energy inequality (2.29) in the following improved version

$$\frac{1}{2\tau}d^2\left(U^n_\tau, U^{n+1}_\tau\right) + \frac{1}{2}\int_0^\tau |\partial^-\phi|^2(\widetilde{U}_\tau((n-1)\tau + r))dr \leq \phi\left(U^n_\tau\right) - \phi\left(U^{n+1}_\tau\right). \tag{2.34}$$

Now, let $\overline{U}_\tau(t)$ be the continuous interpolation defined before. Let $t \geq \tau$. After summation of the inequalities (2.34) over $n \in \left\{1, \ldots, \left[\frac{t}{\tau}\right]\right\}$ and by changing variable under the integrals in (2.34) we obtain

$$\frac{1}{2}\int_0^{t-\tau} |\overline{U}'_\tau|(r)dr + \frac{1}{2}\int_0^{t-\tau} \left|\partial^-\phi\right|^2(\widetilde{U}_\tau(r))dr + \phi(\overline{U}_\tau(\tau[t/\tau])) \leq \phi(u_0)$$

Finally, thanks to lower semi–continuity of the local slope, we can put $\tau = \tau_i$ and pass to the limit as $i \to \infty$ in the above estimate. Using the fact that all interpolations converge to the same limit (by the $C^{0,1/2}$ estimate used in the proof of the discrete compactness) we obtain the following steepest descent condition in integral form,

$$\frac{1}{2}\int_0^t |u'|^2(r)dr + \frac{1}{2}\int_0^t |\partial^-\phi|^2(u(r))dr + \phi(u(t)) \leq \phi(u_0).$$

Then, we can use the upper gradient property to obtain

$$\frac{1}{2}\int_0^t |u'|^2(r)dr + \frac{1}{2}\int_0^t |\partial^-\phi|^2(u(r))dr \leq \int_0^t |u'(r)||\partial^-\phi|(u(r))\,dr.$$

By Young inequality, this gives that $|u'| = |\partial^- \phi(u)|$ a.e. in $(0,t)$, and the previous two inequalities are equalities. As a consequence we obtain that $t \mapsto \phi(u(t))$ is absolutely continuous, with derivative given a.e. by $-|u'|^2(t)$.

Finally, the proof of the other properties follows by suitable monotonicity properties of the slope along discrete trajectories: see Chapter 3 of [AGS04] for details.

2.7. **Uniqueness and error estimate.** Some more properties of the steepest descent flow such as uniqueness and explicit error estimates for the recursive scheme (2.26) can be obtained by requiring some properties for the distance d. It is already known (see [May98]) that steepest descent flows are unique provided the functional $d^2(\cdot, u)$ is 2–convex along geodesics. Such assumption is equivalent to the so called *Alexandroff NPC condition*, which translates into non positivity of sectional curvature in case $\mathcal{S}$ is a Riemannian manifold. Unfortunately, this condition is not satisfied by the space of probability measures $\mathcal{P}_2(X)$ which we shall define later on.

We consider the following generalization of the NPC condition.

Assumption. For any $u, v_0, v_1 \in \mathcal{S}$, there exists a curve v_t connecting v_0 to v_1 such that $\phi(v_t) + \frac{1}{2\tau} d^2(u, v_t)$ is $\left(\lambda + \frac{1}{\tau}\right)$–convex as a function of t.

We observe that v_t needs not be a geodesic in the above assumption.

Theorem 2.8. *Let ϕ be a lower semicontinuous and coercive functional on $\mathcal{S}$. Under the above additional assumption, steepest descent flows are unique and satisfy the following properties:*

(1) *If the starting point u_0 belongs in $\overline{D(\phi)}$, then the recursive scheme (2.26) converges with distance error estimate of order $O(\tau^2)$.*
(2) *The semigroup $u(t)$ is λ–contractive, i.e. $d(u_1(t), u_2(t)) \le e^{\lambda t} d(u_1(0), u_2(0))$ for any two steepest descent flows u_1, u_2.*
(3) *The following evolution variational inequality holds: for any $v \in \mathcal{S}$ we have*

$$\frac{1}{2}\frac{d}{dt} d^2(u(t), v) + \frac{\lambda}{2} d^2(u(t), v) \le \phi(v) - \phi(u(t)) \quad a.e.\ in\ (0, +\infty). \tag{2.35}$$

We refer to [AGS04] for the proof of the above theorem. We only observe that the inequality (2.35) easily implies uniqueness and λ–contractivity of the semigroup by the following *formal* argument:

$$\begin{aligned}\frac{d}{dt}\frac{1}{2} d^2(u(t), v(t)) &= \frac{d}{ds}\frac{1}{2} d^2(u(s), u(t))|_{s=t} + \frac{d}{ds}\frac{1}{2} d^2(u(t), u(s))|_{s=t} \\ &\le -\lambda d^2(v(t), u(t)) + \phi(v(t)) - \phi(u(t)) + \phi(u(t)) - \phi(v(t)) = -\lambda d^2(v(t), u(t)).\end{aligned}$$

This argument can be made rigorous either with Kruzhkov's method of doubling of variables (based on distributional inequalities) of working with pointwise derivatives (see Lemma 4.3.4 in [AGS04]).

3. Optimal transport and Wasserstein distance

The second part of these lecture notes is devoted to the study of the Wasserstein space of probability measures in a separable Hilbert space. The Wasserstein distance arises as a powerful tool to analyze the asymptotic behaviour of solutions of PDE's of diffusion type, to obtain new general proofs of geometric and functional inequalities and to characterize solutions of shape optimization problems.

We refer to [Vil03] for a fairly complete picture of the applications of this theory (and to [AGS04] for the optimal transport problem and the applications to evolution PDE's), quoting here only the papers more relevant for this presentation.

Jordan, Kinderlehrer, Otto [JKO98] showed in their seminal paper that the linear Fokker-Planck can be interpreted as the "gradient flow" with respect to the Wasserstein distance between probability measures. Later on, Otto [Ott01] generalized this approach to the porous medium equation.

The Wasserstein distance appears in the mass transportation problem, which has a very intuitive formulation: we consider a pile of sand and a hole and we want to completely fill up the hole with the sand (both, pile and hole, have the same volume). Of course, we have a cost for transporting any unit of mass from the point x to the point y, we will call it $c(x,y)$. Then, the problem is *how to do the transportation with minimal cost ?* In a more general setup, the pile and the hole can be modelled by Borel probability measures μ, ν in some complete and separable metric space X (in the following we will mostly consider a Hilbert space H). In this way, $d\mu(x)$ could be interpreted as the amount of sand located at point x and $d\nu(y)$ as the amount of sand that we have to transfer to position y.

In order to understand our problem we have to interpret mathematically what "way of transportation" means . We will consider "transport plans", which will be probability measures π on $X \times X$ having μ as first marginal and ν as second marginal. In this way, we can understand $d\pi(x,y)$ as the amount of sand transferred from the point x to the point y. Therefore, our problem can be formulated as follows:

$$\text{Minimize} \quad \left\{ \int_{X\times X} c(x,y)\ d\pi(x,y) \quad \text{among all transport plans } \pi \right\}.$$

This is the known Kantorovich's optimal transportation problem.

The original mass transportation problem is due to Monge. Monge's problem is the same as Kantorovich's, but with the additional requirement that no mass is split, i.e, to every point x a unique destination y is associated (in Kantorovich's problem the possibility of splitting is considered, therefore Kantorovich's problem is a relaxed version of Monge's).

3.1. **Preliminary notation and definitions.** We denote by $\mathcal{P}(X)$ the space of probability measures μ in X, where X is a Borel subset of a separable Hilbert space $(H, \|\cdot\|)$

Definition 3.1 (Push forward). *Let μ be a probability measure on X and let $t : X \to X$ be a Borel map. The* push forward *$t_{\#}\mu \in \mathcal{P}(X)$ of μ through t is defined by $t_{\#}\mu(E) := \mu(t^{-1}(E))$ for any Borel subset $E \subset X$. More generally,*

$$\int_X f(t(x))\, d\mu(x) = \int_X f(y)\, dt_{\#}\mu(y)$$

for every bounded or positive Borel function f.

We will be interested in probability measures on X that are related by means of a push forward:

$$\mu,\ \nu \in \mathcal{P}(X) \quad \text{and} \quad t : X \to X \quad \text{such that} \quad t_{\#}\mu = \nu.$$

The map t is called the *transport map* between the probability measures μ and ν.

Remark 3.1.

(a) *Notice that this "class of transport maps" might be empty, for instance, if we consider $\mu = \delta_x$ and $\nu \neq \delta_y$ $\forall y \in X$.*

(b) *Another difficulty is the fact that the set* $\{t : X \to X : \ t_{\#}\mu = \nu\}$ *is not weakly closed.*
(c) *If* $X = \mathbb{R}^n$ *let us consider* $\mu = f\mathcal{L}^n$, $\nu = g\mathcal{L}^n$. *Then*

$$t_{\#}\mu = \nu \quad \textit{if and only if} \quad \int_X h(t(x))f(x)\,dx = \int_X h(y)g(y)\,dy$$

for every bounded or positive Borel function h. *If* $t \in C^1(X)$ *is one to one this is true if and only if (by the change of variables formula)* $|\det \nabla t(x)|\, g(t(x)) = f(x) \quad a.e.\ x.$

In the following definition we propose another way to link two different probability measures on X.

Definition 3.2 (Transport plan). *Given two measures* μ *and* ν *of* $\mathcal{P}(X)$ *the* set of transport plans *between them is defined by*

$$\Gamma(\mu,\nu) := \{\gamma \in \mathcal{P}(X \times X) : \ \pi^1_{\#}\gamma = \mu\,,\ \pi^2_{\#}\gamma = \nu\}$$

where $\pi^i : X \times X \to X\,,\ i = 1,2$ *are the projections onto the first and second coordinate:* $\pi^1(x,y) = x,\ \pi^2(x,y) = y.$

$\pi^1_{\#}\gamma\,,\ \pi^2_{\#}\gamma$ *are called* marginals of γ. *Therefore, transport plans are those having marginals* μ *and* ν.

We observe that this set is always not empty, since for any μ and ν in $\mathcal{P}(X)$, $\mu \otimes \nu$ belongs to $\Gamma(\mu,\nu)$. Moreover, we remark that the condition $\gamma \in \Gamma(\mu,\nu)$ is equivalent to:

$$\gamma(A \times X) = \mu(A),\ \gamma(X \times B) = \nu(B),\ \forall A,\ B \subset X \text{ Borel}.$$

In this way, $\gamma(A \times B)$ corresponds to the amount of mass initially in A sent to B.

Remark 3.2. *Transport plans and transport maps are related in the following way (see for instance* [Amb03, Vil03]*):*

(a) *Any transport map* t *(between* μ *and* ν*) induces a transport plan* γ *defined by*

$$(Id \times t)_{\#}\mu, \ \textit{where}\ (Id \times t)(x) = (x, t(x)).$$

(Id *is the identity map on* X*).*
(b) *Conversely, if* γ *is a transport plan concentrated on a* γ*-measurable graph in* $X \times X$, *then* γ *admits the form* $(Id \times t)_{\#}\mu$ *for some* μ*-measurable map* t.

In order to define the Wasserstein distance we introduce the following notation:

$$\mathcal{P}_p(X) = \left\{\mu \in \mathcal{P}(X) \,/ \int_X |x|^p \ d\mu(x) < \infty\right\},$$

which denotes the space of probability measures in X with finite moment of order p (finiteness of moments is always true if X is bounded).

The Wasserstein distance of order p is defined on the family of measures with finite p-th moment, with $p \in [1,\infty)$:

Definition 3.3 (Wasserstein distance in $\mathcal{P}_p(X)$). *For any probability measures* μ, ν *in* $\mathcal{P}_p(X)$ *the* Wasserstein distance *of order* p *between them is defined by*

$$W_p(\mu,\nu) := \min\left\{\left(\int_{X\times X} |y-x|^p \ d\gamma(x,y)\right)^{1/p} \ : \ \gamma \in \Gamma(\mu,\nu)\right\}.$$

Using the fact that $c(x,y) = |y-x|^p$ is lower semicontinuous it can be proved that the minimum is always attained (see, for instance, [Amb03, Vil03]). We denote by $\Gamma_0(\mu,\nu)$ the set of optimal plans, i.e., the subset of $\Gamma(\mu,\nu)$ where the minimum is attained

$$\Gamma_0(\mu,\nu) = \left\{ \gamma \in \Gamma(\mu,\nu) : \int_{X\times X} |y-x|^p \; d\gamma(x,y) = W_p^p(\mu,\nu) \right\}.$$

The following Lemma shows that W_p is indeed a distance.

Lemma 3.1. *W_p is a distance in $\mathcal{P}_p(X)$ for $p \in [1,\infty)$.*

Proof. We will focus in the triangle inequality, since the other properties are an easy consequence of the fact that $|x-y|$ is a distance in $X \times X$. To prove the triangle inequality we use the following classical lemma.

Lemma 3.2. *Let μ_1, μ_2, μ_3 be three probability measures in X_1, X_2, X_3 resp., and let $\gamma_{12} \in \Gamma(\mu_1,\mu_2)$ and $\gamma_{23} \in \Gamma(\mu_2,\mu_1)$ be two transport plans. Then there exists $\gamma \in P\,(X_1 \times X_2 \times X_3)$ with marginals γ_{12} on $X_1 \times X_2$ and γ_{23} on $X_2 \times X_3$.*

Using this lemma (considering $X = X_1 = X_2 = X_3$) we can prove the triangle inequality: let us consider μ_1, μ_2, $\mu_3 \in P(X)$, γ_{12} optimal plan between μ_1 and μ_2 and γ_{23} optimal plan between μ_2 and μ_3. Thus, Dudley's lemma gives $\gamma \in P(X \times X \times X)$ with marginals γ_{12} and γ_{23}. In this way, since $\gamma_{13} := \pi^{13}_{\#}\gamma \in \Gamma(\mu^1,\mu^3)$, we obtain

$$\begin{aligned}
W_p(\mu_1,\mu_3) &\leq \left(\int_{X\times X} |x_1 - x_3|^p \; d\gamma_{13}(x_1,x_3) \right)^{1/p} \\
&= \left(\int_{X\times X\times X} |x_1 - x_3|^p \; d\gamma(x_1,x_2,x_3) \right)^{1/p} =: \|x_1 - x_3\|_{L^p(\gamma)} \\
&\leq \|x_1 - x_2\|_{L^p(\gamma)} + \|x_2 - x_3\|_{L^p(\gamma)} = \|x_1 - x_2\|_{L^p(\gamma_{12})} + \|x_2 - x_3\|_{L^p(\gamma_{23})} \\
&= W_p(\mu_1,\mu_2) + W_p(\mu_2,\mu_3).
\end{aligned}$$

(We recall $\pi^{13} : X^3 \to X^2$, $\pi^{13}(x_1,x_2,x_3) = (x_1,x_3)$). □

Remark 3.3 (Kantorovich's and Monge's problem)**.** *Taking into account the relation between transport plans and maps, we deduce the following relation between Kantorovich's and Monge's problems*

$$W_p(\mu,\nu) \leq \inf \left\{ \left(\int_X |t(x) - x|^p \; d\mu(x) \right)^{1/p} \; : \; t : X \to X, \; t_{\#}\mu = \nu \right\}.$$

The inequality is indeed an equality under suitable assumptions on X and μ, see Theorem 3.3 below.

3.2. **Properties of** $(\mathcal{P}_p(X), W_p)$**.** The aim of this section is to show the main properties of $(\mathcal{P}_p(X), W_p)$ in a list of results without proofs. We refer to Chapters 6 and 7 of [AGS04] for proofs, more detailed statements and full bibliographical informations.

Theorem 3.1. *$(\mathcal{P}_p(X), W_p)$ is a separable metric space which is complete if and only if X is closed. Moreover $(\mathcal{P}_p(X), W_p)$ is compact if and only if X is compact.*

In the special case $X = \mathbb{R}$ the lack of compactness of the space is a consequence of a canonical isometry between $\mathcal{P}_2(X)$ and $L^2(0,1)$, built by means of the distribution function $F_\mu(t) := \mu((-\infty, t]) \quad t \in \mathbb{R}$. Indeed, we have

$$W_p^p(\mu,\nu) = \int_0^1 |F_\mu^{-1}(s) - F_\nu^{-1}(s)|^p \, ds,$$

where for any measure $\mu \in \mathcal{P}(\mathbb{R})$ the inverse of F_μ is defined as

$$F_\mu^{-1}(s) := \sup\{x \in \mathbb{R} : \quad F_\mu(x) \le s\} \qquad s \in [0,1].$$

Theorem 3.2. *Given a sequence $\{\mu_n\} \subset \mathcal{P}_p(X)$ and $\mu \in \mathcal{P}_p(X)$ it holds:*

$$\lim_{n\to\infty} W_p(\mu_n, \mu) = 0 \iff \begin{cases} \{\mu_n\} \text{ weakly converges to } \mu \text{ in } \mathcal{P}(X) \\ \\ \lim_{n\to\infty} \int_X |x|^p \, d\mu_n(x) = \int_X |x|^p \, d\mu(x). \end{cases}$$

where the weak convergence is with respect to the duality with continuous and bounded functions ($C_b(X)$), namely,

$$\lim_{n\to\infty} \int_X f(x) \, d\mu_n(x) = \int_X f(x) \, d\mu(x) \qquad \forall f \in C_b(X).$$

Before stating the next results we must recall some preliminary definitions.

Definition 3.4 (Gaussian measures). *Given a separable Banach space, X, with dual X' and $\mu \in \mathcal{P}(X)$, we say that μ is a* nondegenerate Gaussian probability measure *in X if for any $f \in X'$ the image measure $f_\#\mu \in \mathcal{P}(\mathbb{R})$ is a nondegenerate Gaussian measure, i.e. there exist $m = m(f) \in \mathbb{R}$ and $\sigma = \sigma(f) > 0$ such that*

$$\mu(\{x \in X : -\infty < f(x) < a\}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{a} e^{-|t-m|^2/2\sigma^2} \, dt \quad \forall a \in \mathbb{R}.$$

A set $B \in \mathcal{B}(X)$ (where $\mathcal{B}(X)$ denotes the Borel σ-algebra) is a Gaussian null set *if $\mu(B) = 0$ for any nondegenerate Gaussian measure μ in X.*

We denote by $\mathcal{P}_p^r(X)$ the set of probability measures with finite moment of order p which vanish on any Gaussian null set:

$$\mathcal{P}_p^r(X) := \{\mu \in \mathcal{P}_p(X) : \ \mu(B) = 0 \quad \forall \text{ Gaussian null set } B\}.$$

There is a large literature on the existence of optimal transport maps and on necessary and sufficient optimality conditions for maps or plans (recent works in this direction are due to Ambrosio, Brenier, Caffarelli, Gangbo, McCann, Kirchheim, Pratelli, Rachev-Rüschendorf, Feyel-Üstunel, Sudakov, Trudinger, Wang, but the list is not exhaustive: we refer to [AGS04, Vil03] and to [AKP04] for a detailed bibliographical information).

The following result is due to Brenier [Bre87, Bre91] and Rachev-Rüschendorf [1] ($p = 2$) and Gangbo-McCann [GMC] ($p > 1$) in the finite dimensional case. The infinite-dimensional version has been considered in [AGS04].

Theorem 3.3. *(Existence and uniqueness of the optimal map) If $\mu \in \mathcal{P}_p^r(X)$ and either $\dim(H)$ is finite or ν has a bounded support then there is a unique optimal plan γ and this plan is induced by a map T, that is in particular the unique optimal transport map. Moreover, if $p = 2$, T is the (Gateaux) gradient of a l.s.c. convex function φ.*

The following result provides a (weak) regularity of the optimal transport map which works not only in the case $p = 2$ (where regularity follows by the differentiability properties of gradients of convex maps), but also in the case $p > 1$. It has been proved in Theorem 6.2.7 of [AGS04].

Theorem 3.4. *(Regularity of the optimal map) Assume* $\dim(H) < \infty$ *and let* μ, T *as in the previous theorem. Then* T *is approximately differentiable at* μ*-a.e. point and the approximate differential is nonnegative and with nonnegative eigenvalues.*

We recall that T is said to be approximately differentiable at x if there is a linear map L (the differential) such that all sets

$$\{y \in B_r(x) : \ |T(y) - T(x) - \langle L, y - x\rangle| > \varepsilon|y - x|\} \qquad \varepsilon > 0$$

have 0 density at x. Approximate differentiability μ-a.e. on a Borel set A implies (see for instance 3.1.8 of [FE67]) the existence of $\{T_n\} \in Lip(X)$ such that $\mu(A \cap \{T \neq T_n\})$ tends to 0, when n goes to infinity. In turn, this property implies the validity of the area formula for approximately differentiable maps, as discussed in §5.5 of [AGS04] (this plays a role in the sequel).

As far as higher regularity of T is concerned, Caffarelli proved that if $p = 2$, $H = X = \mathbb{R}^n$, μ, $\nu \in C^{k,\alpha}$ and $supp\,\nu = \mathbb{R}^n$, then $T \in C^{k+2,\beta}$, but very little appears to be known when $p \neq 2$.

3.3. **Geodesics in $\mathcal{P}_p(X)$.** In this section we will relate constant speed geodesics in $\mathcal{P}_p(X)$ with optimal plans for X a separable Hilbert space. We recall in the framework of the Wasserstein space (see Corollary 2.1 for definition in a general metric space) that a curve $\mu_t \in \mathcal{P}_p(X)$, $t \in [0,1]$ is a constant speed geodesic if

$$W_p(\mu_s, \mu_t) = (t - s)W_p(\mu_0, \mu_1) \quad \forall\ 0 \leq s \leq t \leq 1.$$

Theorem 3.5 (Characterization of constant speed geodesics). *For any* $\gamma \in \Gamma_0(\mu, \nu)$ *the curve*

$$\mu_t := (\pi_t)_{\#}\gamma \quad \textit{with} \quad \pi_t := (1 - t)\,\pi_0 + t\,\pi_1 \quad t \in [0,1]$$

is a constant speed geodesic joining μ *and* ν*, where* π_0, π_1 *are the projections on the first and second variables respectively. Conversely, any geodesic connecting* μ *and* ν *has this representation for a suitable optimal plan* γ'*.*

Proof. The fact that linear interpolation at the level of transport plans or maps produces geodesics is somehow implicit in many papers on this subject, [McC97], [BB00], [Ott01]. A general proof is given in Chapter 7 of [AGS04], pointing also out the one to one correspondence between geodesics and optimal plans. Here we describe briefly how one can recover an optimal plan from a geodesic μ_t: one fixes $t \in (0,1)$ and proves that $\Gamma_0(\mu_t, \mu_0)$ and $\Gamma_0(\mu_t, \mu_1)$ contain a unique element (and actually are induced by transport maps). Then one defines

$$\gamma' = \gamma^{t1} \circ \gamma^{0t},$$

where γ^{t1} is the unique optimal plan between μ_t and μ_1 and γ^{0t} is the unique optimal plan between μ_0 and μ_t. This "composition" of plans is possible (and canonical, in this case, because the plans in question are induced by transports) using the same construction based on Lemma 3.1 used to prove the triangle inequality. □

This result shows that the definition of λ-convexity along geodesics given in the previous section (Definition 2.13) can be rewritten as follows (see [McC97, CMV, AGS04]):

Definition 3.5 (λ-Convexity along geodesics). *Let X be a separable Hilbert space, $\lambda \in \mathbb{R}$ and let $\phi : \mathcal{P}_p(X) \to (-\infty, +\infty]$ be a functional. We say that ϕ is* λ-(displacement) convex along geodesics *if for every $\mu, \nu \in Dom(\phi)$ there exists $\gamma \in \Gamma_0(\mu, \nu)$ such that*

$$\phi(\mu_t) \leq t\, \phi(\nu) + (1-t)\, \phi(\mu) - \frac{\lambda}{2} t\,(1-t)\, W^2(\mu, \nu),$$

where $t \in [0,1]$ and μ_t is defined in Theorem 3.5.

3.4. **Structure of $(\mathcal{P}_2(X), W_2)$.** To conclude these notes we will focus in the structure of the probability space $(\mathcal{P}_2(X), W_2)$, for X convex subset of a separable Hilbert space H.

Let us recall the definition of positively curved spaces in the sense of Aleksandrov [Ale51]

Definition 3.6 (Positively/Non Positively curved (PC/NPC) space). *A metric space (S, d) is* positively curved *if for any $\bar{\mu} \in S$ we have*

$$d^2(\mu_t, \bar{\mu}) \geq (1-t)\, d^2(\mu_0, \bar{\mu}) + t\, d^2(\mu_1, \bar{\mu}) - t\,(1-t)\, d^2(\mu_0, \mu_1),$$

where $\mu_0, \mu_1 \in S$ and μ_t is any constant speed geodesic joining μ_0 and μ_1. The space is called non positively curved *is the opposite inequality holds:*

$$d^2(\mu_t, \bar{\mu}) \leq (1-t)\, d^2(\mu_0, \bar{\mu}) + t\, d^2(\mu_1, \bar{\mu}) - t\,(1-t)\, d^2(\mu_0, \mu_1).$$

Since the previous inequalities are, in fact, identities when the metric space is a Hilbert space, the condition of being positively curved can be understood as a comparison property for triangles. In this sense, we can speak about triangles in PC-space, where the edges are described by constant speed geodesics joining the vertices.

Otto's differential calculus in $\mathcal{P}_2(X)$ shows in a formal way [Ott01] that this is a PC-space. Using the Aleksandrov metric formulation, the following result has been proved in [AGS04].

Theorem 3.6 ($(\mathcal{P}_2(X), W_2)$ is a PC-space). *For each $\bar{\mu}, \mu_0, \mu_1 \in \mathcal{P}(X)$*

$$W_2^2(\mu_t, \bar{\mu}) \geq t\, W_2^2(\mu_1, \bar{\mu}) + (1-t)\, W_2^2(\mu_0, \bar{\mu}) - t\,(1-t)\, W_2^2(\mu_0, \mu_1) \tag{3.1}$$

along any constant speed geodesic μ_t connecting μ_0 and μ_1.
Therefore, $(\mathcal{P}_2(X), W_2)$ is a PC-space.

Proof. Using an extension of Dudley's lemma (Proposition 7.3.1 of [AGS04]) we can find $\gamma_t \in \mathcal{P}(X^3)$ having marginals μ_0, μ_1 and μ_t such that

$$(\pi^1, \pi^2)_{\#}\, \gamma_t \in \Gamma_0(\mu_0, \mu_1), \quad ((1-t)\pi^1 + t\,\pi^2, \pi^3)_{\#}\, \gamma_t \in \Gamma_0(\mu_t, \bar{\mu}),$$

therefore

$$W_2^2(\mu_t, \bar{\mu}) = \int_{X^3} |x_3 - (1-t)x_1 - t\, x_2|^2 \; d\gamma_t(x_1, x_2, x_3).$$

(We recall $\pi^i : X^3 \to X$ is the projection onto the i-coordinate, $i = 1, 2, 3$) Considering the following Hilbertian identity

$$|x_3 - (1-t)x_1 - t\, x_2|^2 = (1-t)\, |x_1 - x_3|^2 + t\, |x_2 - x_3|^2 - t\,(1-t)|x_2 - x_1|^2$$

we obtain

$$\begin{aligned} W_2^2(\mu_t, \bar{\mu}) &= (1-t) \int_{X^3} |x_1 - x_3|^2 \; d\gamma_t(x_1, x_2, x_3) + t \int_{X^3} |x_2 - x_3|^2 \; d\gamma_t(x_1, x_2, x_3) \\ &- t(1-t) \int_{X^3} |x_2 - x_1|^2 \; d\gamma_t(x_1, x_2, x_3) \end{aligned}$$

and then, since $\int_{X^3} |x_2 - x_1|^2 \, d\gamma_t(x_1, x_2, x_3) = W_2^2(\mu_0, \mu_1)$, we prove

$$W_2^2(\mu_t, \bar\mu) \geq (1-t) W_2^2(\mu_0, \bar\mu) + t\, W_2^2(\mu_1, \bar\mu) - t(1-t)\, W_2^2(\mu_0, \mu_1).$$

□

Remark 3.4. *In general the inequality (3.1) is strict. We show it with the following example*

$$\mu_0 = \frac{1}{2}\left(\delta_{(1,1)} + \delta_{(5,3)}\right), \ \mu_1 = \frac{1}{2}\left(\delta_{(-1,1)} + \delta_{(-5,3)}\right), \ \bar\mu = \frac{1}{2}\left(\delta_{(0,0)} + \delta_{(0,-4)}\right),$$

since in this case $W_2^2(\mu_0, \mu_1) = 40$, $W_2^2(\mu_0, \bar\mu) = W_2^2(\mu_1, \bar\mu) = 30$, *the constant speed geodesic connecting* μ_0 *and* μ_1 *is* $\mu_t := \frac{1}{2}\left(\delta_{(1-6t, 1+2t)} + \delta_{(5-6t, 3-2t)}\right)$ *and therefore for* $t = 1/2$ $W_2^2(\mu_{1/2}, \bar\mu) = 24 > \frac{30}{2} + \frac{30}{2} - \frac{40}{4}$.

The PC property and the previous remark prove that in dimension greater than 1, the squared Wasserstein distance is *not* 2-convex along constant speed geodesics. But, 2-convexity of the squared distance is a quite essential property as we observed in Section 2.7. In this sense, it is natural to look for a different choice of the connecting curve. Could it be possible to consider other curves that guarantee 2-convexity? The answer is affirmative and they will be called "generalized geodesics".

Definition 3.7 (Generalized geodesics)**.** *For* $\bar\mu \in \mathcal{P}_2^r(X)$ *and* $\mu_0, \mu_1 \in \mathcal{P}_2(X)$ *we define the* generalized geodesic *joining* μ_0 *and* μ_1 *(with base* $\bar\mu$*) as follows*

$$\mu_t = (T_{\bar\mu}^{\mu_0} + t\,(T_{\bar\mu}^{\mu_1} - T_{\bar\mu}^{\mu_0}))_\# \bar\mu \quad t \in [0,1]$$

where $T_{\bar\mu}^{\mu_0}$ *is an optimal map between* $\bar\mu$ *and* μ_0 *and* $T_{\bar\mu}^{\mu_1}$ *is an optimal map between* $\bar\mu$ *and* μ_1.

Remark 3.5. *Through a measure in* $\sigma \in \mathcal{P}(X^3)$ *whose projection on the first and second variable belongs to* $\Gamma_0(\bar\mu, \mu_0)$ *and whose projection on the first and the third variable belongs* $\Gamma_0(\bar\mu, \mu_1)$, *generalized geodesics can also be defined (for a base point* $\bar\mu \in \mathcal{P}_2(X)$*) setting*

$$\mu_t := \left((1-t)\pi^2 + t\pi^1\right)_\# \sigma.$$

A generalized geodesic satisfies the curvature inequality (the reverse of (3.1)):

$$W^2(\mu_t, \bar\mu) \leq (1-t)\, W^2(\mu_0, \bar\mu) + t\, W^2(\mu_1, \bar\mu) - t\,(1-t)\, W^2(\mu_0, \mu_1), \tag{3.2}$$

since

$$\begin{aligned} W^2(\mu_t, \bar\mu) &\leq \int_X |(1-t) T_{\bar\mu}^{\mu_0}(x) + t\, T_{\bar\mu}^{\mu_1}(x) - x|^2 \, d\bar\mu(x) \\ &\leq \int_X (1-t)|T_{\bar\mu}^{\mu_0}(x) - x|^2 + t\,|T_{\bar\mu}^{\mu_1}(x) - x|^2 - t\,(1-t)|T_{\bar\mu}^{\mu_0}(x) - T_{\bar\mu}^{\mu_1}(x)|^2 \, d\bar\mu(x) \\ &\leq (1-t)\, W^2(\mu_0, \bar\mu) + t\, W^2(\mu_1, \bar\mu) - t\,(1-t)\, W^2(\mu_0, \mu_1). \end{aligned}$$

The last inequality follows by the fact that $\gamma := \left(T_{\bar\mu}^{\mu_0}, T_{\bar\mu}^{\mu_1}\right)_\# \bar\mu$ is an admissible plan between μ_0 and μ_1.

Definition 3.8 (λ-convexity along generalized geodesics)**.** *A functional* $\phi : \mathcal{P}_2(X) \to (-\infty, +\infty]$ *is* λ-convex along generalized geodesics *if for every* $\bar\mu, \mu_0, \mu_1 \in Dom(\phi) \subset \mathcal{P}_2(X)$ *there exists a generalized geodesic* μ_t *connecting* μ_0 *and* μ_1 *induced by a plan* $\gamma \in \Gamma(\bar\mu, \mu_0, \mu_1)$ *such that*

$$\phi(\mu_t) \leq (1-t)\, \phi(\mu_0) + t\, \phi(\mu_1) - \frac{\lambda}{2} t(1-t)\, W_\gamma^2(\mu_0, \mu_1),$$

where

$$W_\gamma^2(\mu_0,\mu_1) := \int_{X^3} |x_1 - x_0|^2 \, d\gamma(x, x_0, x_1) \left(\geq W_2^2(\mu_0,\mu_1)\right).$$

Therefore, as a first example, we obtain using (3.2) that $W_2(\cdot,\bar\mu)$ is 2-convex along generalized geodesics (of course not all, but those with base point $\bar\mu$).

3.4.1. *Basic examples of geodesically convex funtionals.* We now list the main examples of λ-convex functionals that we will use at the end of these notes to study PDE's as gradient flows with respect to the Wasserstein metric (see [CMV, Vil03] and Chapter 9 of [AGS04]).

Example 3.1 $(-W_2^2(\cdot,\bar\mu))$. *By the PC condition the functional* $-W_2^2(\cdot,\bar\mu)$ *is (-2)-convex.*

Example 3.2 (Potential energy). *For a potential function* $V : X \to (-\infty, +\infty]$ *whose negative part grows at most with order* p *we define in* $\mathcal{P}_p(X)$ *the potential energy functional as*

$$\mathcal{V}(\mu) = \int_X V(x) \; d\mu(x).$$

The potential energy $\mathcal{V}$ *is* λ*-convex if and only if* V *is* λ*-convex.*

Example 3.3 (Interaction energy). *For* $W : X \to (-\infty, +\infty]$ *whose negative part grows at most with order* p *we define in* $\mathcal{P}_p(X)$ *the interaction energy functional as*

$$\mathcal{W}(\mu) = \int_{X\times X} W(x-y) \; d\mu(x)\, d\mu(y).$$

The interaction energy functional $\mathcal{W}$ *is convex if* W *is convex.*

Example 3.4 (Internal energy). *Assume that* $X \subset H = \mathbb{R}^n$ *is convex and let* $F : [0,\infty) \to (-\infty, +\infty]$ *be such that* $s \mapsto s^n F(s^{-n})$ *is convex, non increasing in* $(0,\infty)$ *and* $F(0) = 0$. *Assume also that* F *has a more than linear growth at infinity (see* [AGS04] *to see how this condition can be removed). We define the internal energy functional*

$$\mathcal{J}(\mu) = \int_{\mathbb{R}^n} F(\rho) \; dx \quad \text{for } \mu = \rho\,\mathcal{L}^n$$

and $\mathcal{J}(\mu) = +\infty$ *if* μ *is not absolutely continuous with respect to the Lebesgue measure. Then, the internal energy* $\mathcal{J}$ *is convex along geodesics in* $\mathcal{P}_p(X)$.

As examples of functions F *we can consider:*

- *(a)* $F(t) = \dfrac{t^m}{m-1}$ *with* $m \geq 1 - \dfrac{1}{n}$ *(this restriction for* m *comes from the convexity of* $s \mapsto s^n F(s^{-n})$*); notice that in the case* $m < 1$ *the function* F *is not superlinear, hence the natural extension of* $\mathcal{J}$ *to the whole space of measures, by lower semicontinuous relaxation, can be finite also for measures not abolutely continuous with respect to Lebesgue measure, see again* [AGS04] *for details.*
- *(b) For the limit case in the previous example* $m \to 1$ *we obtain* $F(t) = t \log t$ *(the entropy functional).*

Given two regular measures $\mu_i = \rho_i\,\mathcal{L}^n \in Dom(\mathcal{J}) \subset \mathcal{P}_p^r(\mathbb{R}^n)$, $i = 0,1$ and T, the optimal transport map between μ_0 and μ_1 for p-Wasserstein distance such that $T_\#\mu_0 = \mu_1$, we can consider $T_t := Id + t(T - Id)$ and therefore T_t is an optimal transport map between μ_0 and $\mu_t := T_{t\#}\mu_0$ for any $t \in [0,1]$. Then, the approximate differentiability of T_t yields the validity of the area formula and therefore an explicit representation of μ_t (see Proposition 9.3.9 of [AGS04]):

$$\mu_t = \rho_t \mathcal{L}^n \quad \text{with } \rho_t = \frac{\rho_0}{J\,T_t} \circ T_t^{-1},$$

where $J\,T_t$ is the Jacobian of T_t. Therefore

$$\mathcal{J}(\mu_t) = \int_{\mathbb{R}^n} F\left(\rho_t(y)\right)\, dy = \int_{\mathbb{R}^n} F\left(\frac{\rho_0(x)}{J\,T_t(x)}\right) JT_t(x)\, dx.$$

Next we give an infinite-dimensional example.

Example 3.5 (Relative entropy). *For any μ, γ Borel probability measures on a separable Hilbert space X we define the relative entropy of μ with respect to γ as*

$$\mathcal{H}(\mu,\gamma) := \begin{cases} \displaystyle\int_X \frac{d\mu}{d\gamma} \log\left(\frac{d\mu}{d\gamma}\right) d\gamma & \text{if } \mu \ll \gamma, \\ +\infty & \text{otherwise} \end{cases}$$

If we consider the convex function

$$F(t) := \begin{cases} t\,(\log t - 1) + 1 & t > 0, \\ 1 & t = 0, \\ +\infty & t < 0, \end{cases}$$

we obtain

$$\mathcal{H}(\mu,\gamma) = \int_X F\left(\frac{d\mu}{d\gamma}\right)\, d\gamma.$$

And, by Jensen's inequality we observe $\mathcal{H}(\mu,\gamma) \geq 0$ and $\mathcal{H}(\mu,\gamma) = 0$ if and only if $\mu = \gamma$.

Then, it turns out that for each pair of probability measures μ_0, $\mu_1 \in \mathcal{P}_p(X)$ the functional $\mathcal{H}(\cdot,\gamma)$ is convex along the interpolating curve $((1-t)\,\pi^1 + t\,\pi^2)_{\#}\mu$ for $\mu \in \Gamma(\mu_0,\mu_1)$ if and only if γ is log-concave, i.e.

$$\log\gamma(tA + (1-t)B) \geq t\log\gamma(A) + (1-t)\log\gamma(B) \quad t \in [0,1],$$

for any pair of open sets A, B. In the finite-dimensional case this condition is equivalent to

$$\gamma = e^{-V} \cdot \mathcal{H}^k|_{aff(supp\,\gamma)}\,, \quad \text{where } k = dim(supp\,\gamma)$$

for some V convex and l.s.c. (here $aff(supp\,\gamma)$ denotes the affine envelope of $supp\,\gamma$).

The relative entropy functional can be defined for more general convex functions F, see [AGS04] for details.

We will conclude by using these functionals to study evolution PDEs.

4. Gradient flows in the space of probability measures

The previous theory will be applied in the last part of this paper to study PDEs of diffusion type. Jordan, Kinderlehrer and Otto [JKO98] and then the latter author alone [Ott01] showed that many interesting and classical parabolic equations can be interpreted as gradient flows of functionals with respect to the Wassertein distance between probability measures.

In the space-time open cylinder $\mathbb{R}^n \times (0,+\infty)$ we look for *nonnegative solutions* $\rho : \mathbb{R}^n \times (0,+\infty)$ of a parabolic equation of the type

$$\partial_t\rho - \nabla\cdot\left(\rho\nabla\left(\frac{\delta\mathcal{F}}{\delta\rho}\right)\right) = 0 \quad \text{in } \mathbb{R}^n \times (0,+\infty), \tag{4.1}$$

where

$$\frac{\delta\mathcal{F}(\rho)}{\delta\rho} = -\nabla\cdot F_p(x,\rho,\nabla\rho) + F_z(x,\rho,\nabla\rho).$$

is the first variation of a typical integral functional

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^n} F(x,\rho(x),\nabla\rho(x))\,dx \tag{4.2}$$

associated to a (smooth) Lagrangian $F = F(x,z,p) : \mathbb{R}^n \times [0,+\infty) \times \mathbb{R}^n \to \mathbb{R}$.

Observe that (4.1) has the following structure:

$$\partial_t\rho + \nabla\cdot(\rho v) = 0 \qquad \textit{(continuity equation)}, \tag{4.3a}$$

$$\rho v = \rho\nabla\psi \qquad \textit{(gradient condition)}, \tag{4.3b}$$

$$\psi = -\frac{\delta\mathcal{F}(\rho)}{\delta\rho} \qquad \textit{(nonlinear relation)}. \tag{4.3c}$$

Observe that in the case when F depends only on $z=\rho$ then we have

$$\frac{\delta\mathcal{F}(\rho)}{\delta\rho} = F_z(\rho), \quad \rho\nabla F_z(x,\rho) = \nabla L_F(\rho), \quad L_F(z) := zF'(z) - F(z). \tag{4.4}$$

Since we look for *nonnegative solutions* having (constant, by (4.3a), normalized) *finite mass*

$$\rho(x,t) \geq 0, \qquad \int_{\mathbb{R}^n} \rho(x,t)\,dx = 1 \quad \forall t \geq 0, \tag{4.5}$$

and *finite p-th momentum*

$$\int_{\mathbb{R}^n} |x|^p\rho(x,t)\,dx < +\infty \quad \forall\, t \geq 0, \tag{4.6}$$

we can

$$\text{identify } \rho \text{ with the measures } \mu_t := \rho(\cdot,t)\cdot\mathcal{L}^n, \tag{4.7}$$

and we consider $\mathcal{F}$ as a functional defined in $\mathcal{P}_p(\mathbb{R}^n)$.

We will interpret (4.3b) as a *tangency* condition for the velocity field v, and the nonlinear coupling (4.3c) as a *subdifferential inclusion.*

At this level of generality the equivalence between the system (4.3a,b,c) and the gradient flow of $\mathcal{F}$ is known only for smooth solutions (which, by the way, may not exist); nevertheless, the point of view of gradient flow in the Wasserstein spaces, introduced by Otto, still presents some interesting features, whose role should be discussed in each concrete case:

a) The gradient flow formulation suggests a general variational scheme, discussed in the first part of these notes, to approximate the solution of (4.3a,b,c): proving its convergence is interesting both from the theoretical and the numerical point of view.
b) The variational scheme exhibits solutions which are *a priori* nonnegative, even if the equation does not satisfies any maximum principle as in the fourth order case [Ott98, GTS04].
c) Working in Wasserstein spaces allows for weak assumptions on the data: initial values which are general measures (as for fundamental solutions, in the linear cases) fit quite naturally in this framework.
d) The gradient flow structure suggests new contraction and energy estimates, which may be useful to study the asymptotic behaviour of solutions, see for instance [CMV], or to prove uniqueness under weak assumptions on the data.

e) The interplay with the theory of Optimal Transportation provides a novel point of view to get new functional inequalities with sharp constants.
f) The variational structure provides an important tool in the study of the dependence of solutions from perturbation of the functional.
g) The setting in space of measures is particularly well suited when one considers evolution equations in infinite dimensions and tries to "pass to the limit" as the dimension n goes to ∞.

Let us consider for instance the linear Fokker-Planck equation (physical constants are considered equal to 1)

$$\partial_t \rho = \Delta \rho + \nabla \cdot (\rho \nabla V), \tag{4.8}$$

where V is a given potential and ρ is a distribution function depending on time, $t \in (0, +\infty)$ and position, $x \in \mathbb{R}^n$. Writing the equation in the form

$$\partial_t \rho = \nabla \cdot (\rho(\nabla \log \rho + \nabla V))$$

we immediately recognize it as the (formal) gradient flow of the perturbed entropy functional

$$\mathcal{F}(\rho) = \int_{\mathbb{R}^n} \rho(x) \log \rho(x) + \int \rho(x) V(x)\, dx \left(= \int_{\mathbb{R}^n} \rho(x) \log \left(\frac{\rho(x)}{e^{-V(x)}} \right) dx = \mathcal{H}(\rho, e^{-V}) \right).$$

Notice also that the expression inside parentheses represents the functional as a *relative* entropy functional with respect to the reference measure $\gamma = e^{-V}\mathcal{L}^n$ (the steady state of the FP equation). This second expression paves the way to the analysis and to the gradient flow interpretation of the PDE in *infinite* dimensions, where there is no analogue of the Lebesgue measure but still an analogue of γ.

Let us to consider the following more complicated example. Consider the following convex functional (see [CMV, AGS04])

$$\phi(\mu) := \int_{\mathbb{R}^n} F(\rho)(x)\, dx + \int_{\mathbb{R}^n} V(x)\, d\mu + \frac{1}{2} \int_{\mathbb{R}^n \times \mathbb{R}^n} W\, d\mu \otimes \mu, \quad \text{for } \mu = \rho\, \mathcal{L}^n, \tag{4.9}$$

which is the sum of internal, potential and interaction energy (see the previous subsection and [AGS04] for assumptions on F, V and W) and $\phi(\mu) = +\infty$ otherwise.

The gradient flow in $\mathcal{P}_p(\mathbb{R}^n)$ produces the *nonlinear and nonlocal* PDE of diffusion type

$$\partial_t \rho = \nabla \cdot \left(\rho\, J_q \left(\frac{\nabla L_F(\rho)}{\rho} + \nabla V + (\nabla W) * \rho \right) \right), \tag{4.10}$$

where

$$J_q : L^q \to L^p, \quad p = q', \qquad J_q(v) = |v|^{q-2} v,\ J_q(0) = 0$$

and

$$L_F(t) = t\, F'(t) - F(t).$$

In order to understand the power of the gradient flow formulation and the suitability of working in Wasserstein spaces we need to know the differential structure of $\mathcal{P}_p(X)$ and "subdifferential calculus" in $\mathcal{P}_p(X)$. Besides the initial work of Otto, this problem has been investigated independently and with slightly different motivations by Carrillo-McCann-Villani in [CMV] and by Ambrosio-Gigli-Savaré [AGS04].

To conclude, we describe briefly along the lines of [AGS04] the differentiable structure of $\mathcal{P}_p(H)$ and the (sub)-differential calculus in this space, showing how (4.3b) and (4.3c) can be made rigorous.

Let us first introduce a good class of test functions in infinite dimensions that reduces to the standard one in the finite-dimensional case.

Definition 4.1 (Smooth cylindrical functions). *A function $\varphi : H \to \mathbb{R}$ is* cylindrical *if there exists $\psi \in C_c^\infty(\mathbb{R}^n)$ such that $\varphi = \psi \circ \pi$, with $\pi \in \Pi_n(H)$, where we denote by $\Pi_n(H)$ the space of all maps $\pi : H \to \mathbb{R}^n$ of the form*

$$\pi(x) = (\langle x, e_1\rangle, ..., \langle x, e_n\rangle) \quad x \in X$$

where $\{e_1, ..., e_n\}$ is any orthonormal family of vectors in H.

We denote by $Cyl(H)$ the family of cylindrical functions on H. If $I = (a, b)$ (open interval) the space $Cyl(H \times I)$ is defined analogously considering functions $\psi \in C_c^\infty(\mathbb{R}^n \times I)$ and functions $\varphi(x, t) = \psi(\pi(x), t)$.

The following theorem is the fundamental bridge between the metric and the differentiable viewpoint, as it links a metric concept (the absolutely continuous curves with values in the Wasserstein space) to a differential one, namely the continuity equation.

Theorem 4.1 (Absolutely continuous curves in $\mathcal{P}_p(H)$ and the continuity equation). *Let $I \subset \mathbb{R}$ be an open interval, let $\mu_t : I \to \mathcal{P}_p(X)$ be an absolutely continous curve and let $|\mu'| \in L^1(I)$ be its metric derivative (see Theorem 2.1). Then there exists $\Psi_t \in L^p(\mu_t, X)$ such that the continuity equation*

$$\partial_t \mu_t + \nabla \cdot (\Psi_t \mu_t) = 0 \tag{4.11}$$

holds in the distribution sense in $I \times H$, i.e.

$$\int_I \int_H (\partial_t \varphi(x,t) + \langle \Psi_t(x), \nabla_x \varphi(x,t)\rangle) \; d\mu_t(x) \; dt = 0 \qquad \varphi \in Cyl(H \times I)$$

and

$$\|\Psi_t\|_{L^p(\mu_t)} \leq |\mu_t'| \quad \text{for} \quad \mathcal{L}^1 - a.e. \; t \in I. \tag{4.12}$$

Conversely, if $\mu_t : I \to \mathcal{P}_p(X)$ satisfies the continuity equation (4.11) for some Borel velocity field Ψ_t with $\int_0^T \|\Psi_t\|_{L^p(\mu_t)} \, dt < \infty$, then $\mu_t \in AC([0,T], \mathcal{P}_p(H))$ and $|\mu_t'| \leq \|\Psi_t\|_{L^p(\mu_t)}$ $\mathcal{L}^1-$ a.e. $t \in I$.

For a given curve μ_t, there is no uniqueness of the vector fields Ψ_t satisfying the continuity equation (4.11): choosing vector fields w_t such that $\nabla \cdot (w_t \mu_t) = 0$ the vectors $\Psi_t + w_t$ still satisfy (4.11). In view of the previous theorem it is natural to consider as velocity vectors those Ψ_t's for which equation (4.11) holds and the $L^p(\mu_t)$ norm is minimal, i.e.

$$\|\Psi_t\|_{L^p(\mu_t, H)} = |\mu'|(t) \quad \text{for}\, \mathcal{L}^1 - a.e. \; t \in I \tag{4.13}$$

hold. A duality argument shows that the minimizer $\bar{v}$ of the L^p norm of $v + w$ among all w such that $w\mu$ is divergence-free is characterized by the property of being in the $L^p(\mu_t)$ closure of $J_q(\nabla\varphi)$ among all cylindrical functions φ (this is a nonlinear version of the standard Helmholtz decomposition of vector fields). These remarks lead to the following definition.

Definition 4.2 (Tangent bundle). *Let $\mu \in \mathcal{P}_2(H)$. We define*

$$Tan_\mu \mathcal{P}_2(H) := \overline{\{\nabla\varphi \; : \; \varphi \; cylindrical \; in \; H\}}_{L^2(\mu)}.$$

In general let $\mu \in \mathcal{P}_p(H)$. We define

$$Tan_\mu \mathcal{P}_p(H) := \overline{\{J_q\,(\nabla\varphi) \; : \; \varphi \; cylindrical \; in \; H\}}_{L^p(\mu)}.$$

As a consequence Theorem 4.1 gives that *for any absolutely continuous curve μ_t with values in $\mathcal{P}_p(H)$ there exists a uniquely determined $\mathcal{L}^1$-a.e. tangent vector field Ψ_t. It is characterized by the validity of the continuity equation and by $|\mu_t'| = \|\Psi_t\|_{L^p(\mu_t)}$ for $\mathcal{L}^1 - a.e.\ t \in I$.*

Another consequence of Theorem 4.1 is the following formula, found independently (in the case $p = 2$ and $\dim H < \infty$) by Benamou and Brenier [BB00] and Otto [Ott01]:

Theorem 4.2 (Benamou-Brenier formula). *Given $\mu_0,\, \mu_1 \in \mathcal{P}_p(H)$ it holds*

$$W_p^2(\mu_0,\, \mu_1) = \inf \int_0^1 \|\Psi_t\|^2_{L^p(\mu_t)}\, dt, \tag{4.14}$$

where the infimum is taken among all absolutely continous curves $\mu^t : [0,1] \to \mathcal{P}_p(H)$ such that $\mu^i = \mu_i$ for $i = 0, 1$ and Ψ_t are given by the equations (4.11) and (4.13).

According to the definition of tangent space we gave, the formula can be interpreted (as formally Otto did in [Ott01]) by saying that the Wasserstein distance is the (Riemannian when $p = 2$, Finsler when $p > 1$) distance induced by the natural L^p norm on the tangent bundle.

At this point, the path to follow consists of the following steps (see [AGS04] for full details):

(a) *Define the subdifferential of a λ–convex functional ϕ at $\mu \in \mathcal{P}_p(H)$.* In the case when $\mu \in \mathcal{P}_r^p(H)$ this is easy to do as follows: a vector $v \in Tan_\mu \mathcal{P}_p(H)$ belongs to $\partial\phi(\mu)$ if

$$\phi(\nu) \geq \phi(\mu) + \int_H \langle v(x), T_\mu^\nu(x) - x\rangle\, d\mu(x) + \frac{\lambda}{2} W_p^2(\mu, \nu) \qquad \forall \nu \in \mathcal{P}^p(H),$$

where T_μ^ν denotes the unique optimal map between μ and ν. In the general case this definition can be adapted using plans instead of maps.

(b) *Define the concept of gradient flow.* It is simply a locally absolutely continuous map μ_t whose tangent velocity field v_t satisfies $J_p(v_t) \in -\partial\phi(\mu_t)$ for $\mathcal{L}^1$–a.e. t (recall that J_p is the identity in the standard case $p = 2$).

(c) Prove that the "differential" definition in (b) is equivalent to the metric one of De Giorgi. This is achieved in Theorem 11.1.3 of [AGS04] using an analogue of the identity (2.18) already used in the Hilbertian context.

(d) In the case $p = 2$ and for λ-convex functionals we obtain not only uniqueness of the gradient flow, but also the contractivity property

$$W_2(\mu_t, \nu_t) \leq e^{\lambda t} W_2(\lim_{s\downarrow 0} \mu_s, \lim_{s\downarrow 0} \nu_s). \tag{4.15}$$

The proof is based, as in the classical case, on the monotonicity property of the subdifferential and on a precise formula for the derivative of $t \mapsto W_2^2(\sigma_t, \sigma)$, with $t \mapsto \sigma_t$ absolutely continuous and σ fixed, given for almost every t by

$$2 \int_H \langle w_t(x), y - x\rangle\, d\gamma_t(x, y), \tag{4.16}$$

where w_t is the tangent velocity field to $t \mapsto \sigma_t$ and γ_t is *any* optimal plan between σ_t and σ. We apply it to compute the derivatives

$$\frac{d}{ds} W_2^2(\mu_s, \nu_t) \qquad \text{and} \qquad \frac{d}{ds} W_2^2(\mu_t, \nu_s)$$

and then, by the doubling of variables argument, to compute the derivative of $t \mapsto W_2^2(\mu_t, \nu_t)$. Denoting by v_t and w_t the velocity fields relative to μ_t and ν_t respectively, in the simpler case when μ_t and ν_t are absolutely continuous we get

$$\begin{aligned}\frac{d}{dt}W_2^2(\mu_t,\nu_t) &= \frac{d}{ds}W_2^2(\mu_s,\nu_t)\big|_{s=t}+\frac{d}{ds}W_2^2(\mu_t,\nu_s)\big|_{s=t}\\ &= 2\int_H\langle v_t(x),T_{\mu_t}^{\nu_t}(x)-x\rangle\,d\mu_t(x)+2\int_H\langle w_t(x),T_{\nu_t}^{\mu_t}(x)-x\rangle\,d\nu_t(x)\\ &\leq 2\lambda W_2^2(\mu_t,\nu_t),\end{aligned}$$

whence by integration we obtain (4.15). Notice that this scheme is independent of the functional (just λ-convexity is needed) and applies also in the infinite dimensional case and, with minor variants, for singular measures (notice indeed that (4.16) does not require any absolute continuity property and therefore just a more general definition of subdifferential is needed). See also [CMV] for related and independent results.

Thanks to step (c), all the results of the metric theory can be applied, see in particular Remark 2.1 for the coercitivity issue. In Chapter 11 of [AGS04] we explore also other, more differential in spirit, ways to pass to the limit in the discrete scheme that don't use the variational interpretation discussed above (see [JKO98, Agu02, Agu03]). However, it is our belief that only the variational scheme produces the more general convergence results, without restriction on the measures (absolutely continuous or not, in finite and infinite dimensions) and also, in some cases [GTS04], for non-convex functionals.

References

[Agu02] M. Agueh *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, PhD Thesis, Georgia Institute of Technology, 2002.

[Agu03] M. Agueh, *Asymptotic behavior for doubly degenerate parabolic equations*, C. R. Math. Acad. Sci. Paris, Ser. I 337, (2003), 331-336.

[AGS04] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the spaces of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.

[Ale51] A. D. Aleksandrov, *A theorem on triangles in a metric space and some of its applications*, Trudy Mat. Inst. Steklov **38** (1951), 5–23, Izdat. Akad. Nauk SSSR, Moscow.

[Amb95] L. Ambrosio, *Minimizing movements*, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5) **19** (1995), 191–246.

[Amb03] ———, *Lectures notes on optimal transport problem*, Mathematical aspects of evolving interfaces, CIME summer school in Madeira (Pt) (P. Colli and J. Rodrigues, eds.), vol. 1812, Springer, 2003, 1–52.

[AT04] L. Ambrosio and P. Tilli, *Topics on analysis in metric spaces*, Oxford Lecture Series in Mathematics and its Applications, vol. 25, Oxford University Press, Oxford, 2004.

[AKP04] L. Ambrosio, B. Kirchheim, and A.Pratelli, *Existence of optimal transport maps for crystalline norms*, Duke Math. J., **125** (2004), 207–241.

[BA89] C. Baiocchi, *Discretization of evolution variational inequalities*, in Partial Differential Equations an the calculus of variations, Vol I. F. Colombini, A. Marino, L. Modica and S. Spagnolo eds., Birkhäuser, Boston, 1989, 59–92.

[BB00] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), no. 3, 375–393.

[Bre87] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs.* C.R. Acad. Sci. Paris, Série I, **305** (1987),805–808.

[Bre91] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. 44, (1991), 375–417.

[Bré73] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973.

[CL71] M. G. Crandall and T. M. Liggett, *Generation of semi-groups of nonlinear transformations on general Banach spaces*, Amer. J. Math. **93** (1971), 265–298.

[CMV] J.A. Carrillo, R.J. McCann, and C. Villani, *Contraction in the 2-wasserstein metric length space and thermalization of granular media.* To appear in Archive for Rational Mech. Anal.

[CP72] M. G. Crandall and A. Pazy, *Nonlinear evolution equations in Banach spaces*, Israel J. Math. **11** (1972), 57–94.

[CT00] J. A. Carrillo and G. Toscani, *Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity*, Indiana Univ. Math. J. **49** (2000), no. 1, 113–142.

[DG93] E. De Giorgi, *New problems on minimizing movements*, Boundary value problems for partial differential equations and applications, RMA Res. Notes Appl. Math., vol. 29, Masson, Paris, 1993, pp. 81–98.

[DGMT80] E. De Giorgi, A. Marino, and M. 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.

[DPD02] Manuel Del Pino and Jean Dolbeault, *Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions*, J. Math. Pures Appl. (9) **81** (2002), no. 9, 847–875.

[FE67] H. Federer, *Geometric Measure Theory*, Springer Verlag, 1967.

[GMC] W. Gangbo and R. McCann, *The geometry of optimal transportation.* Acta Math., **177** (1996), 113–161.

[GTS04] U. Gianazza, G. Toscani, and G. Savaré, *A fourth-order parabolic equation and Wasserstein distance*, Preprint IMATI-CNR, Pavia, 2004.

[HK98] J. Heinonen and P. Koskela, *Quasiconformal maps in metric spaces with controlled geometry*, Acta Math. **181** (1998), no. 1, 1–61.

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17.

[Jos97] J. Jost, *Nonpositive curvature: geometric and analytic aspects*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1997.

[May98] U F. Mayer, *Gradient flows on nonpositively curved metric spaces and harmonic maps*, Comm. Anal. Geom. **6** (1998), no. 2, 199–253.

[McC97] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), 153–179.

[McC99] R.J. McCann, *Polar factorization of maps on Riemannian manifolds.*, Geom. Funct. Anal. **11** (2001), 589–608.

[NSV00] R. Nochetto, G. Savaré and C. Verdi, *A posteriori error estimates for variable time-step discretizations of nonlinear evolution equations*, Comm. Pure Appl. Math. **53** (2000), 525–589.

[Ru96] J. Rulla, *Error analysis for implicit approximations to solutions to Cauchy problems*, SIAM J. Numer Anal **33** (1996), 68–87.

[Ott98] F. Otto, *Lubrication approximation with prescribed nonzero contact angle*, Comm. PDE, **23** (1998), 2077–2164.

[Ott01] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[OV00] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), no. 2, 361–400.

[1] R. Rachev and L.Rüschendorf, *A characterization of random variables with minimum L^2 distance*, J. Multivariate Anal. **32** (1990), 48–54 and corrigendum, **34** (1990), 156.

[Tos96] G. Toscani, *Kinetic approach to the asymptotic behaviour of the solution to diffusion equations*, Rend. Mat. Appl. (7) **16** (1996), no. 2, 329–346.

[Tos97] G. Toscani, *Sur l'inégalité logarithmique de Sobolev*, C. R. Acad. Sci. Paris Sér. I Math. **324** (1997), no. 6, 689–694.

[Vil03] Cédric Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.