An overview on calculus and heat flow in metric measure spaces and spaces with Riemannian curvature bounded from below

Luigi Ambrosio *

January 7, 2012

Contents

*Scuola Normale Superiore, Pisa. email: l.ambrosio@sns.it

1 Introduction and overview

In these notes I will illustrate the main results contained in the recent two papers [4] and [5], written in collaboration with N.Gigli and G.Savaré. I will basically present results and proofs in some detail for [4], and present only the main results of [5]. These notes follow to a large extent the presentation given in Montreal, in July 2011.

Acknowledgement. The author acknowledges the support of the ERC ADG GeMeThNES.

1.1 Some by now "classical" results

Let us consider in $\mathbb{R}^n$ the heat equation (with the notation $u_t(x) = u(t,x)$)

$$\frac{d}{dt}u_t = \Delta u_t.$$

Classically, this PDE can be viewed as the gradient flow of the energy

$$\mathrm{Dir}(u) := \frac{1}{2}\int_{\mathbb{R}^n} |\nabla u|^2\,dx \qquad (+\infty \text{ if } u \notin H^1(\mathbb{R}^n))$$

in the Hilbert space $H = L^2(\mathbb{R}^n)$. Indeed, formally $t \mapsto u_t$ solves the ODE $u' = -\nabla\mathrm{Dir}(u)$ in H because

$$\text{Dir "differentiable" at } u \qquad \Longleftrightarrow \qquad -\Delta u \in L^2,\ \nabla\mathrm{Dir}(u) = -\Delta u$$

The precise meaning of "differentiability" is provided by convex analysis and it will be specified later on in rigorous terms.

In 1998, Jordan, Kinderlehrer and Otto proved [16] that the same equation arises as gradient flow of the *entropy* functional

$$\mathrm{Ent}(\rho\mathscr{L}^n) := \int_{\mathbb{R}^n} \rho\log\rho\,dx \qquad (+\infty \text{ if } \mu \text{ is not a.c. w.r.t. } \mathscr{L}^n)$$

in the space $\mathscr{P}_2(\mathbb{R}^n)$ of Borel probability measures in $\mathbb{R}^n$ with finite quadratic moments, with respect to Wasserstein distance W_2 (I am denoting here by $\mathscr{L}^n$ the Lebesgue measure in $\mathbb{R}^n$). Recall that W_2^2 is defined by the minimum transportation cost, in the Kantorovich formulation, using $c(x,y) = d^2(x,y)$ as cost function, namely

$$W_2^2(\mu,\nu) := \min\left\{\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2\,d\boldsymbol{\gamma}(x,y) :\ (\pi_1)_\sharp\boldsymbol{\gamma} = \mu,\ (\pi_2)_\sharp\boldsymbol{\gamma} = \nu\right\}.$$

Here and in the sequel I will adopt the standard push forward notation: any $f : X \to Y$ Borel induces a map $f_\sharp : \mathscr{P}(X) \to \mathscr{P}(Y)$ given by

$$f_\sharp\mu(B) := \mu\big(f^{-1}(B)\big) \qquad \forall B \in \mathscr{B}(Y).$$

The proof of this equivalence, and the reasons for it, can be found at many levels:

(1) By the so-called Otto calculus [21], i.e. formally viewing $\mathscr{P}(X)$ as an infinite-dimensional Riemannian manifold and computing with this structure the gradient flow of Ent.

(2) Proving that the implicit time discretization scheme (the so-called Euler scheme), traditionally used for the time discrete approximation of gradient flows, when done with energy $E = \mathrm{Ent}$ and distance $d = W_2$, does converge to the heat equation. Recall that this scheme involves a time step τ and the recursive minimization of

$$y \mapsto E(y) + \frac{1}{2\tau} d^2(y, x)$$

to provide a discrete (piecewise constant in time) solution to the gradient flow. This is the strategy pursued in [16].

(3) Giving a rigorous meaning to what "gradient flow of Ent in $\mathscr{P}(X)$ w.r.t. W_2 means", and check that solutions of this gradient flow are solutions to the heat equation. Then, having proved that W_2 gradient flows are contained in L^2 gradient flows, suffices to apply standard uniqueness results for $\frac{d}{dt} u_t = \Delta u_t$.

The last strategy, used in [3], is more abstract, but still uses to some extent the differentiable structure of $\mathbf{R}^n$. The question is: are there deeper reasons for this equivalence? This is motivated also by a long series of papers where the JKO result is extended to more general classes of metric spaces: Riemannian manifolds [11], Finsler spaces [20], Wiener spaces [12] (a class of infinite-dimensional Gaussian spaces), Alexandrov spaces [14], etc.

1.2 Metric measure spaces

Let us consider a *metric measure space* $(X, d, \mathfrak{m})$, with $\mathfrak{m} \in \mathscr{P}(X)$. In this framework it is still possible to define a "Dirichlet energy", that we call Cheeger functional:

$$\mathsf{Ch}(f) := \frac{1}{2} \inf \left\{ \liminf_{n \to \infty} \int_X |\nabla f_n|^2 \, d\mathfrak{m} : \; f_n \in \mathrm{Lip}(X), \; \int_X |f_n - f|^2 \, d\mathfrak{m} \to 0 \right\},$$

where

$$|\nabla g|(x) := \limsup_{y \to x} \frac{|g(y) - g(x)|}{d(y, x)}$$

is the *slope* (also called local Lipschitz constant). Our terminology is motivated by Cheeger's seminal paper [8], where a similar relaxation procedure is considered. Cheeger considered arbitrary functions f_n in the approximation procedure, and upper gradients g_n of them in place of the slopes; for this reason our functional is a priori larger than the original functional in [8]. A nontrivial fact, a consequence of the identification theorem of weak gradients (see (8.1)), is that the two functionals coincide.

Also, one can consider the so-called *relative entropy functional* $\mathrm{Ent}_{\mathfrak{m}} : \mathscr{P}(X) \to [0, \infty]$

$$\mathrm{Ent}_{\mathfrak{m}}(\rho \mathfrak{m}) := \int_X \rho \log \rho \, d\mathfrak{m} \qquad (+\infty \text{ if } \mu \text{ is not a.c. w.r.t. } \mathfrak{m}).$$

The basic result is that the equivalence between L^2-gradient flow of Ch and W_2-gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ *always* holds, if the latter is properly understood. But, without additional assumptions on the space, both objects can be trivial, as the following simple example shows.

Example 1.1 (Triviality of Ch) Let $X = [0,1]$, d the Euclidean distance, $\mathfrak{m} = \sum_{n\geq 1} 2^{-n}\delta_{q_n}$, where $\{q_n\}_{n\geq 1}$ is an enumeration of $[0,1]\cap\mathbb{Q}$. Let $A_n \supset \mathbb{Q}\cap X$ be open sets with $\mathscr{L}^1(A_n)\to 0$ and

$$\chi_n(t) := \int_0^t \big(1-\chi_{A_n}(s)\big)\,ds \qquad t\in[0,1].$$

Then $f\circ\chi_n \to f$ in $L^2(X,\mathfrak{m})$ for all $f\in \mathrm{Lip}(X)$ and $f\circ\chi_n$ is locally constant in $\mathbb{Q}\cap X$ hence

$$\mathsf{Ch}(f) = 0 \qquad \forall f\in \mathrm{Lip}(X).$$

It follows that $\mathsf{Ch}\equiv 0$ in $L^2(X,\mathfrak{m})$.

1.3 Identification of weak gradients

A closely related question, relevant in particular for the second paper, is the identification of weak gradients. The first one, that we call *relaxed* gradient $|\nabla f|_*$, is the object that provides integral representation to Ch:

$$\mathsf{Ch}(f) = \frac{1}{2}\int_X |\nabla f|_*^2\,d\mathfrak{m} \qquad \forall f\in D(\mathsf{Ch}).$$

It has all the natural properties (locality, chain rules, etc.) a weak gradient should have, see Theorem 3.2 and (3.1) below. This gradient is useful when doing "vertical" variations $\epsilon\mapsto f+\epsilon g$ (i.e. in the *dependent* variable). On the other hand, when computing variations of the entropy, the "horizontal" variations $\epsilon\to f(\gamma_\epsilon)$ (i.e. in the *independent* variable) are necessary. These are related to another weak gradient $|\nabla f|_w$, defined as follows.

We consider the so-called *weak* upper gradient property by requiring

$$|f(\gamma_1)-f(\gamma_0)| \leq \int_\gamma G$$

along "almost all" curves γ in $AC^2([0,1];X)$. Then, we define $|\nabla f|_w$ as the weak upper gradient G with smallest $L^2(X,\mathfrak{m})$ norm. This definition crucially depends on the notion of null set of curves, that we shall specify later on.

The remarkable fact is that these two gradients *always* coincide (and, of course, maybe both trivial without extra assumptions). The proof of this identification uses ideas from optimal transportation, as lifting of solutions to the heat flow to probability measures in $AC^2\big([0,1];X\big)$ and the energy dissipation rate of $\mathrm{Ent}_{\mathfrak{m}}$ along the L^2 gradient flow of Ch. I think that this identification result, which a priori has nothing to do with entropy and optimal transportation, is a nice illustration of the power of the optimal transport theory.

1.4 Why gradients are not trivial in Lott-Sturm-Villani spaces

In these spaces one imposes convexity along W_2 geodesics of $\mathrm{Ent}_{\mathfrak{m}}$ (the so-called $CD(0,\infty)$ condition) or of functionals

$$\rho\mathfrak{m} \mapsto -\int_X \rho^{1-1/N}\,d\mathfrak{m}$$

(the $CD(0,N)$ condition). In this case the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ is not trivial, and since it coincides with the L^2 gradient flow of Ch, also the latter is not trivial. As a consequence, pathological situations as those described in Example 1.1 cannot occur.

Notice that, formally, the energy dissipation rate is

$$\begin{aligned}\frac{d}{dt}\int_X \rho_t \log \rho_t \, d\mathfrak{m} &= \int_X \log \rho_t \Delta \rho_t \, d\mathfrak{m} = -\int_{\{\rho_t>0\}} \frac{|\nabla \rho_t|^2}{\rho_t} \, d\mathfrak{m} \\ &= -4\int_X |\nabla \sqrt{\rho_t}|^2 \, d\mathfrak{m}.\end{aligned}$$

We shall develop a calculus that makes this result meaningful and rigorous.

The standing assumption on the metric measure structure in these notes are that (X, d) is a compact metric space and $\mathfrak{m} \in \mathscr{P}(X)$. Notice that the results in [4] require neither global or local compactness assumption on (X, d) nor finiteness of $\mathfrak{m}$, and are therefore appropriate to deal with infinite-dimensional spaces. Those of [5], instead, have been established for the moment under the additional assumption that $\mathfrak{m} \in \mathscr{P}(X)$. Good prerequisites needed for the reading of these notes are the basic facts of optimal transport theory, see [26], [2] and [3].

Acknowledgement. Work written within the research project ERC ADG GeMeThNES.

2 Hopf-Lax formula and Hamilton-Jacobi semigroup

Given a function $f : X \to \mathbb{R}$ bounded from below, we define

$$Q_t f(x) := \inf_{y\in X} f(y) + \frac{1}{2t} d^2(x, y) \qquad \text{(Hopf-Lax formula)}$$

Theorem 2.1 *Assume that f bounded and lower semicontinuous. It holds:*

(1) $Q_t f(x) \uparrow f(x)$ as $t \downarrow 0$;

(2) $Q_t(Q_s f(x)) \geq Q_{t+s} f(x)$, with equality if (X, d) is geodesic;

(3) $\frac{d^+}{dt} Q_t f(x) + \frac{1}{2}|\nabla Q_t f(x)|^2 \leq 0$;

(4) $Q_t f(x)$ restricted to $(\epsilon, \infty) \times X$ is Lipschitz for all $\epsilon > 0$.

Sketch of proof.
(1) It follows by the lower semicontinuity of f, which ensures also that minimizers do exist.
(2) It follows by

$$\begin{aligned} &\inf_y \left(\inf_z f(z) + \frac{1}{2s} d^2(z, y) \right) + \frac{1}{2t} d^2(x, y) \\ =\ &\inf_z \inf_y \left(\frac{1}{2s} d^2(z, y) + \frac{1}{2t} d^2(x, y) \right) + f(z) \\ \geq\ &\inf_z \frac{1}{2(s+t)} d^2(x, z) + f(z), \end{aligned} \tag{2.1}$$

noticing that the last inequality is an equality in geodesic spaces.
In order prove (3), we set

$$\begin{cases} D_f^+(x, t) := \max\{d(x, y) : \ y \text{ minimizer}\} \\ D_f^-(x, t) := \min\{d(x, y) : \ y \text{ minimizer}\}. \end{cases}$$

Since any limit of minimizers is a minimizer, D_f^+ is upper semicontinuous, while D_f^- is lower semicontinuous.

In addition $D_f^+(x,s) \geq D_f^-(x,s) \geq D_f^+(x,t)$ if $0 < t < s$. Indeed,

$$f(x_t) + \frac{d^2(x_t,x)}{2t} \leq f(x_s) + \frac{d^2(x_s,x)}{2t}$$
$$f(x_s) + \frac{d^2(x_s,x)}{2s} \leq f(x_t) + \frac{d^2(x_t,x)}{2s}.$$

Adding up and using $\frac{1}{t} > \frac{1}{s}$ we deduce that $D_f^-(x,s) \geq D_f^+(x,t)$.

It follows that, given x, $D_f^+(x,t) = D_f^-(x,t)$ with at most countably many exceptions. We prove first that $\frac{d^\pm}{dt} Q_t f(x) = -[D_f^\pm(x,t)]^2/(2t^2)$. Choosing x_t at maximum distance and x_s at minimum distance yields

$$\begin{aligned} Q_s f(x) - Q_t f(x) &\leq \frac{1}{2s} d^2(x,x_t) + f(x_t) - f(x_t) - \frac{1}{2t} d^2(x,x_t) \\ &= \frac{(D_f^+(x,t))^2}{2} \left(\frac{1}{s} - \frac{1}{t}\right) \end{aligned}$$

$$\begin{aligned} Q_s f(x) - Q_t f(x) &\geq \frac{1}{2s} d^2(x,x_s) + f(x_s) - f(x_s) - \frac{1}{2t} d^2(x,x_s) \\ &= \frac{(D_f^-(x,s))^2}{2} \left(\frac{1}{s} - \frac{1}{t}\right) \end{aligned}$$

One can then use semicontinuity of $D_f^\pm$ and monotonicity $D_f^+ \geq D_f^-$ to conclude.

To conclude, suffices to show that $|\nabla Q_t f|(x) \leq D_f^+(x,t)/t$. The same trick used before, now for variations in space, yields:

$$\begin{aligned} Q_t f(x) - Q_t f(y) &\leq \frac{1}{2t} d^2(x,z) + f(z) - f(z) - \frac{1}{2t} d^2(z,y) \\ &\leq d(x,y) \left(\frac{D_f^-(y,t)}{t} + \frac{d(x,y)}{2t}\right) \end{aligned}$$

and we can use the upper semicontinuity of D_f^+ to conclude. □

If y is kept fixed and we let $x \to y$ we obtain the sharper inequality

$$|\nabla^+ Q_t f|(y) \leq \frac{D_f^-(y,t)}{t}, \tag{2.2}$$

where the *ascending slope* $|\nabla^+ f|$ is defined by

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

2.1 Hamilton-Jacobi and optimal transportation

Why the Hopf-Lax formula and the Hamilton-Jacobi equation are relevant in the theory of optimal transport?

c-transform. Given a cost function $c : X \times Y \to \mathbb{R}$, the c-transforms $\varphi^c : Y \to \mathbb{R} \cup \{-\infty\}$, $\psi^c : X \to \mathbb{R} \cup \{-\infty\}$ are defined by

$$\varphi^c(y) := \inf_{x \in X} c(x,y) - \varphi(x), \qquad \psi^c(x) := \inf_{y \in Y} c(x,y) - \psi(y).$$

Notice the analogy with convex analysis: $\psi^c = (-\psi)^*$ if X is Hilbert and $c(x,y) = \langle x, y \rangle$. The relation with the HL formula is also obvious:

$$\psi^c = Q_1(-\psi).$$

Then, we say that $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is *c-concave* if $\varphi = \psi^c$ for some $\psi : Y \to \mathbb{R} \cup \{-\infty\}$. As in convex analysis, $\varphi \mapsto \varphi^c$ is an involution in the class of c-concave functions: $(\varphi^c)^c = \varphi$.

Definition 2.2 *We say that a c-concave function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is a Kantorovich potential relative to (μ, ν) if it satisfies*

$$\varphi(x) + \varphi^c(y) = c(x,y) \qquad \text{for } \boldsymbol{\gamma}\text{-a.e. } (x,y) \tag{2.3}$$

for any optimal plan $\boldsymbol{\gamma}$ from μ to ν.

Proposition 2.3 *If φ is a Kantorovich potential from μ to ν it holds:*

$$|\nabla^+ \varphi|(x) \le d(x,y) \qquad \text{for } \boldsymbol{\gamma}\text{-a.e. } (x,y)$$

for any optimal plan $\boldsymbol{\gamma}$ from μ to ν. In particular $\int |\nabla^+ \varphi|^2 \, d\mu \le W_2^2(\mu, \nu)$.

Proof. Since $\varphi = (\varphi^c)^c$ we may write $\varphi = Q_1(-\varphi^c)$. Combining this with the optimality condition (2.3) and with (2.2) gives

$$|\nabla^+ \varphi(x)| = |\nabla^+ Q_1(-\varphi^c)|(x) \le D^-_{-\varphi^c}(x,1) \le d(x,y) \quad \boldsymbol{\gamma}\text{-a.e. in } X \times X.$$

□

2.2 The classical Brenier theorem and its metric counterpart

In the Euclidean case $c(x,y) = |x-y|^2/2$, if φ is differentiable at x and

$$\varphi(x) + \varphi^c(y) = \frac{1}{2}|x-y|^2$$

one can differentiate at x and obtain $\nabla \varphi(x) = (x - y)$, which tells us that y is uniquely determined by x and

$$|\nabla^+ \varphi|(x) = d(x,y). \tag{2.4}$$

We say that a *metric* Brenier theorem holds if (2.4) holds $\boldsymbol{\gamma}$-a.e. for any optimal plan $\boldsymbol{\gamma}$, so that in particular $W_2^2(\mu,\nu) = \int |\nabla^+ \varphi|^2 \, d\mu$. The following simple example shows that this equality ma fail in general, see Theorem 9.14 for a positive result.

Example 2.4 $X = [0,1]$, $\mu = \delta_0$, $\mu_t = t^{-1} \chi_{[0,t]} \mathscr{L}^1$. In this case

$$\varphi(x) = \frac{x^2}{2} - x, \qquad \int |\nabla^+ \varphi|^2 \, d\mu_0 = 0$$

while $W_2^2(\mu_0, \mu_1) = \frac{1}{3}$.

2.3 Optimal transport and Kantorovich potentials in geodesic spaces

If (X, d) is Polish (i.e. complete and separable) and geodesic we may formulate the optimal transport problem in terms of *geodesic plans*, namely probability measures $\boldsymbol{\pi}$ concentrated in the Polish space $\mathrm{Geo}(X)$ of constant speed geodesics:

$$\min\left\{\int d^2(\gamma_0,\gamma_1)\,d\boldsymbol{\pi}(\gamma):\ (e_0)_\sharp\boldsymbol{\pi}=\mu,\ (e_1)_\sharp\boldsymbol{\pi}=\nu\right\}. \tag{2.5}$$

Here $e_t : C([0,1];X) \to X$ are the evaluation maps, namely $e_t(\gamma) = \gamma_t$. The relation with the classical optimal plans $\boldsymbol{\gamma}$ of Kantorovich theory is that if $\boldsymbol{\pi}$ is a minimizer in (2.5), then $(e_0, e_1)_\sharp\boldsymbol{\pi}$ is an optimal plan, and that any optimal $\boldsymbol{\gamma}$ admits a (possibly nonunique) "lifting" $\boldsymbol{\pi}$, i.e. $(e_0, e_1)_\sharp\boldsymbol{\pi} = \boldsymbol{\gamma}$. The nice fact is that constant speed geodesics are in 1-1 correspondence with optimal geodesic plans:

Theorem 2.5 *Any constant speed geodesic μ_t in $\mathscr{P}(X)$ can be represented as $(e_t)_\sharp\boldsymbol{\pi}$ for a suitable optimal geodesic plan $\boldsymbol{\pi}$. Conversely, any optimal geodesic plan $\boldsymbol{\pi}$ induces a constant speed geodesic $(e_t)_\sharp\boldsymbol{\pi}$.*

The following fundamental result provides a deeper connection between geodesics and the Hopf-Lax formula, see the seminal paper [7] and [26] for much more on this subject.

Theorem 2.6 *Let μ_t, $t \in [0,1]$ be a constant speed geodesic and let φ be a Kantorovich potential relative to μ_0, μ_1. Then, for all $t \in (0,1]$, $\varphi_t := Q_t(-\varphi^c)$ is a Kantorovich potential, relative to the scaled cost $c_t := c/t$, from μ_{1-t} to μ_1.*

Sketch of proof. It is obvious that $\varphi_t + \varphi \le c_t$. The key implication is

$$\varphi(\gamma_0)+\varphi^c(\gamma_1)=c(\gamma_0,\gamma_1)\quad\text{implies}\quad\varphi_t(\gamma_{1-t})+\varphi^c(\gamma_1)=c_t(\gamma_{1-t},\gamma_1). \tag{2.6}$$

Hence, if $\boldsymbol{\pi}$ is an optimal geodesic plan, $\varphi+\varphi^c = c$ $\boldsymbol{\pi}$-a.e. implies $\varphi_t+\varphi^c = c/t$ $\boldsymbol{\pi}_t$-a.e., where

$$\boldsymbol{\pi}_t := (\gamma_{1-t},\gamma_1)_\sharp\boldsymbol{\pi}$$

is an optimal geodesic plan from μ_{1-t} to μ_1. The implication (2.6) is not difficult to prove, and related to the fact that characteristic lines for the Hamilton-Jacobi equation are geodesics, see also (2.1). □

3 Cheeger's energy and relaxed gradients

Let us recall the definition of Ch we already mentioned in the introduction:

$$\mathsf{Ch}(f) := \frac{1}{2}\inf\left\{\liminf_{h\to\infty}\int_X |\nabla f_h|^2\,d\mathfrak{m}:\ f_h\in \mathrm{Lip}(X),\ \int_X |f_h-f|^2\,d\mathfrak{m}\to 0\right\}.$$

By construction $\mathsf{Ch} : L^2(X,\mathfrak{m}) \to [0,\infty]$ is lower semicontinuous, and it is easily seen to be convex. Can we provide an integral representation to it?

Relaxed slope: $G \in L^2(X,\mathfrak{m})$ is a relaxed slope of f if G bounds from above a function in

$$\left\{\text{weak } L^2 \text{ limit points of } |\nabla f_n|,\ f_n\in\mathrm{Lip}(X),\ \|f_n-f\|_2\to 0\right\}$$

or equivalently in

$$\big\{ \textit{strong } L^2 \text{ limit points of } G_n \geq |\nabla f_n|,\ f_n \in \mathrm{Lip}(X),\ \|f_n - f\|_2 \to 0 \big\}\,.$$

The equivalence between the two characterizations of relaxed slopes follows by Mazur's lemma: the first characterization is useful to show that f has a relaxed slope iff $\mathsf{Ch}(f) < \infty$, while the second one is useful to perform diagonal arguments and to show that the collection of relaxed slopes is a convex closed set, possibly empty. This motivates the next definition.

Definition 3.1 (Minimal relaxed slope) *We call* minimal relaxed slope*, and denote by* $|\nabla f|_*$*, the function with smallest* $L^2(X, \mathfrak{m})$ *norm among relaxed slopes.*

Theorem 3.2 *Let* $f \in D(\mathsf{Ch})$*. Then:*

(1) $\mathsf{Ch}(f) = \frac{1}{2}\int |\nabla f|_*^2\, d\mathfrak{m}$*;*

(2) if G_1*,* G_2 *are relaxed slopes, so is* $\min\{G_1, G_2\}$*;*

(3) $|\nabla f|_* \leq G$ $\mathfrak{m}$*-a.e. for any relaxed slope* G*;*

(4) $g = f$ $\mathfrak{m}$*-a.e. on a Borel set* B *implies* $|\nabla f|_* = |\nabla g|_*$ $\mathfrak{m}$*-a.e. on* B*.*

Calculus rules. If $N \subset \mathbb{R}$ is Lebesgue negligible, then $|\nabla f|_* = 0$ a.e. in $f^{-1}(N)$. In addition, we have the (weak) chain rule

$$|\nabla \phi(f)|_* \leq |\phi'(f)||\nabla f|_* \qquad \text{with equality if } \phi' \geq 0. \tag{3.1}$$

Sketch of proof.
(1) Any weak limit point of $|\nabla f_n|$ yields a relaxed slope, hence $\mathsf{Ch}(f) \geq \frac{1}{2}\int |\nabla f|_*^2\, d\mathfrak{m}$. Writing $G \leq |\nabla f|_*$ as the strong limit of $G_n \geq |\nabla f_n|$ we have

$$\int |\nabla f|_*^2\, d\mathfrak{m} \geq \int G^2\, d\mathfrak{m} \geq \liminf_n \int |\nabla f_n|^2\, d\mathfrak{m} \geq 2\mathsf{Ch}(f).$$

(2) By approximation, suffices to show that $\chi_{X\setminus B} G_1 + \chi_B G_2$ is a relaxed slope if B is closed. Set $\rho(x) = \mathrm{dist}(x, B)$, $\chi_r(x) = \min\{1, r^{-1}\rho\}$, so that $\chi_r \uparrow \chi_{X\setminus B}$ as $r \downarrow 0$, and pass to the limit in

$$|\nabla(\chi_r f_{n,1} + (1-\chi_r) f_{n,2})| \leq \chi_r |\nabla f_{n,1}| + (1-\chi_r)|\nabla f_{n,2}| + \mathrm{Lip}(\chi_r)|f_{n,1} - f_{n,2}|.$$

(3) Just take $\tilde{G} := \min\{|\nabla f|_*, G\}$. Its L^2 norm is strictly smaller than $\||\nabla f|_*\|_2$ if the set $\{|\nabla f|_* > G\}$ has positive $\mathfrak{m}$-measure. □

3.1 Heat flow and Laplacian

Let's start with some reminders on the classical theory of gradient flows of convex and l.s.c. functionals $F : H \to \mathbb{R} \cup \{+\infty\}$ in a Hilbert space H.

Subdifferential ∂F**.** It is the multivalued map defined by

$$\partial F(x) := \{p \in H :\ F(x) + \langle p, y - x\rangle \leq F(y)\ \ \forall y \in H\}$$

for all $x \in D(F) := \{F < \infty\}$. The set $\partial F(x)$ is closed and convex. The *gradient* $\nabla F(x)$ is the element with minimal norm in $\partial F(x)$.

Definition 3.3 (Gradient flow) *It is a locally absolutely continuous map $x : (0, \infty) \to H$ satisfying*

$$-x'(t) \in \partial F\big(x(t)\big) \qquad \text{for a.e. } t > 0.$$

In addition, we say that $x(t)$ starts from $\bar{x}$ if $\lim_{t \downarrow 0} x(t) = \bar{x}$.

Theorem 3.4 (Existence and uniqueness) *For all $\bar{x} \in \overline{D(F)}$ there exists a unique gradient flow starting from $\bar{x}$ and the induced semigroup*

$$S_t : [0, \infty) \times \overline{D(F)} \to \overline{D(F)}$$

is contractive. In addition, we have the regularizing effects:

(1) $S_t\bar{x} \in D(\partial F) \subset D(F)$ for all $t > 0$ and

$$F(S_t\bar{x}) \leq \inf_{v \in D(F)} F(v) + \frac{1}{2t} d^2(v, \bar{x});$$

(2) $\frac{d^+}{dt} S_t\bar{x} = -\nabla F(S_t\bar{x})$ for all $t > 0$;

(3) $t \mapsto |\nabla F|^2(S_t\bar{x})$ is nonincreasing, so that $S_t\bar{x}$ is Lipschitz in (ϵ, ∞) for all $\epsilon > 0$;

(4) $\frac{d^+}{dt} F(S_t\bar{x}) = -|\nabla F|^2(S_t\bar{x}) = -|\frac{d^+}{dt} S_t\bar{x}|^2$ for all $t > 0$.

According to these results, we may choose $H = L^2(X, \mathfrak{m})$ and $F = \mathsf{Ch}$ and define

$$-\Delta f := \text{the element with minimal } L^2\text{-norm of } \partial\mathsf{Ch}(f)$$

so that (by the density of $D(\mathsf{Ch}) \supset \mathrm{Lip}(X)$ in $L^2(X, \mathfrak{m})$) we obtain a L^2 *heat flow* $\mathbf{h}_t f$ solving (the derivative being understood in $L^2(X, \mathfrak{m})$)

$$\frac{d}{dt}\mathbf{h}_t\bar{f} = \Delta\mathbf{h}_t\bar{f}$$

starting from any initial condition $\bar{f} \in L^2(X, \mathfrak{m})$.
Remarks. (1) $\Delta = \Delta_{d,\mathfrak{m}}$. Even in the classical situations, $\Delta f = \mathrm{div}(\nabla f)$, where ∇f depends on the metric (to associate a vector ∇f to df) while div depends on the volume form $\mathfrak{m}$, via the adjoint formula

$$\int g \,\mathrm{div} F \, d\mathfrak{m} = -\int \langle \nabla g, F \rangle \, d\mathfrak{m}.$$

(2) Δ need not to be linear in this context! Take $X = \mathbb{R}^2$ with the L^∞ norm, to get

$$\mathsf{Ch}(f) = \frac{1}{2}\int \left(\left|\frac{\partial f}{\partial x}\right| + \left|\frac{\partial f}{\partial y}\right|\right)^2 dxdy.$$

Nowithstanding this potential lack of linearity, a reasonable calculus can be developed:

$$-\int g\Delta f \, d\mathfrak{m} \leq \int |\nabla f|_* |\nabla g|_* \, d\mathfrak{m}, \tag{3.2}$$

$$-\int \phi(f)\Delta f \, d\mathfrak{m} = \int \phi'(f)|\nabla f|_*^2 \, d\mathfrak{m}. \tag{3.3}$$

The inequality (3.2) follows by

$$\mathsf{Ch}(f) - \epsilon \int g\Delta f \, d\mathfrak{m} \le \frac{1}{2}\int |\nabla(f+\epsilon g)|_*^2 \, d\mathfrak{m}$$

noticing that $|\nabla(f+\epsilon g)|_* \le |\nabla f|_* + \epsilon|\nabla g|_*$ for $\epsilon > 0$, so that

$$-\epsilon \int g\Delta f \, d\mathfrak{m} \le \epsilon \int |\nabla f|_* |\nabla g|_* \, d\mathfrak{m} + \frac{\epsilon^2}{2}\int |\nabla g|_*^2 \, d\mathfrak{m}.$$

The proof of (3.3) is based on the chain rule (3.1).

Proposition 3.5 (Properties of the heat flow) *The following properties hold:*

(1) Homogeneity: $\mathbf{h}_t(\lambda f) = \lambda \mathbf{h}_t f \ \forall \lambda \in \mathbb{R}$*;*

(2) Comparison principle: if $f \le g$*, then* $\mathbf{h}_t f \le \mathbf{h}_t g$ *for all* $t \ge 0$*;*

(3) Energy dissipation: if $J \subset \mathbb{R}$ *is an interval,* $f_t : X \to J$ *and* $\mathrm{e} : J \to \mathbb{R}$ *is convex and locally* $C^{1,1}$*, then*

$$\int e(\mathbf{h}_t f) \, d\mathfrak{m} = \int \mathrm{e}(f) \, d\mathfrak{m} - \int_0^t \int \mathrm{e}''(\mathbf{h}_s f)|\nabla \mathbf{h}_s f|_*^2 \, d\mathfrak{m} ds.$$

(4) Mass preservation: $\int \mathbf{h}_t f \, d\mathfrak{m} = \int f \, d\mathfrak{m}$ *for all* $t \ge 0$*.*

Strictly speaking, (3) does not cover the most interesting case, the case of the entropy $\mathrm{e}(z) = z\log z$ when $\mathbf{h}_t f \ge 0$ and $J = [0,\infty)$:

$$\int \mathbf{h}_t f \log \mathbf{h}_t f \, d\mathfrak{m} = \int f \log f \, d\mathfrak{m} - \int_0^t \int_{\{\mathbf{h}_s f > 0\}} \frac{|\nabla \mathbf{h}_s f|_*^2}{\mathbf{h}_s f} \, d\mathfrak{m} ds. \tag{3.4}$$

It can be recovered by the approximation $f \mapsto \max\{f, \epsilon\}$, $\epsilon \downarrow 0$ (this is possible thanks to the fact that $\mathfrak{m}(X) < \infty$, the general case is much more delicate, see [4]).
Sketch of proof.
(1) Since Ch is 2-homogeneous, one can prove that $\partial \mathsf{Ch}(\lambda f) = \lambda \partial \mathsf{Ch}(f)$, hence f_t is a gradient flow iff λf_t is. Uniqueness then gives the identity.
(2) Since Δ is not linear, the standard argument (take $f_t - g_t$ and use that it is a gradient flow) does not apply. We appeal to the Euler scheme and prove that $f \le g$ and

$$f_\tau \text{ minimizer of } \frac{1}{2\tau}\|\cdot - f\|_2^2 + \mathsf{Ch}(\cdot), \quad g_\tau \text{ minimizer of } \frac{1}{2\tau}\|\cdot - g\|_2^2 + \mathsf{Ch}(\cdot),$$

(here τ is the time step) implies $f_\tau \le g_\tau$.
(3) By monotone approximation, $e \in C^{1,1}(\mathbb{R})$. Then

$$\frac{d}{dt}\int e(\mathbf{h}_t f) \, d\mathfrak{m} = \int e'(\mathbf{h}_t f)\Delta \mathbf{h}_t f \, d\mathfrak{m} = -\int e''(\mathbf{h}_t f)|\nabla \mathbf{h}_t f|_*^2 \, d\mathfrak{m}.$$

(4) Suffices to use the function identically equal to 1 in the PDE. □

3.2 Absolutely continuous functions and metric speed

A curve $\gamma : [0,1] \to X$ is said to be absolutely continuous if

$$d(\gamma_t, \gamma_s) \le \int_t^s f(r)\,dr \qquad \forall [t,s] \subset [0,1] \tag{3.5}$$

for some $f \in L^1(0,1)$.

If γ is absolutely continuous, the *metric speed* $|\dot\gamma| : [0,1] \to [0,\infty]$ is defined by

$$|\dot\gamma_t| := \lim_{h\to 0} \frac{d(\gamma_{t+h}, \gamma_t)}{|h|}.$$

It is possible to prove that the limit exists for a.e. t, that $|\dot\gamma| \in L^1(0,1)$, and that it is the minimal L^1 function for which the bound (3.5) holds.

3.3 Kuwada's lemma

This lemma, taken from [14] provides one of the two key connections between the "Eulerian" or "vertical" viewpoint implicit in the theory of relaxed gradients and the "Lagrangian" of "horizontal" viewpoint of the theory of optimal transportation.

Lemma 3.6 *Let $f_0 \in L^2(X, \mathfrak{m})$ a probability density, $f_t = \mathbf{h}_t f_0$. Then the curve $\mu_t := f_t \mathfrak{m}$ is absolutely continuous in $\mathscr{P}(X)$ and*

$$|\dot\mu_t|^2 \le \int_{\{f_t>0\}} \frac{|\nabla f_t|_*^2}{f_t}\,d\mathfrak{m} \qquad \text{for a.e. } t > 0.$$

It is also convenient to introduce the Fisher information functional, defined on $\{\rho : \ \rho \ge 0,\ \sqrt{\rho} \in D(\mathsf{Ch})\}$, as follows:

$$F(\rho) := 4 \int |\nabla \sqrt{\rho}|_*^2\,d\mathfrak{m} = \int_{\{\rho>0\}} \frac{|\nabla\rho|_*^2}{\rho}\,d\mathfrak{m}$$

(the last equality follows by chain rule).

Sketch of proof of Kuwada lemma. We prove an integral version of the lemma, namely

$$W_2^2(\mu_t, \mu_s) \le \ell \int_t^s F(f_r)\,dr$$

with $0 \le s < t < \infty$ and $\ell := (s-t)$. By Kantorovich's duality formula, suffices to show

$$\int -\varphi\,d\mu_t + \int Q_1\varphi\,d\mu_s \le \frac{\ell}{2} \int_t^s F(f_r)\,dr,$$

where φ runs in the class of bounded continuous functions. Replacing φ by $Q_\epsilon\varphi$ and letting $\epsilon \downarrow 0$ we can assume that $Q_t\varphi$ is Lipschitz in $[0,1] \times X$.

Now we set $g(r) := \int Q_r\varphi\,d\mu_{t+\ell r}$, so that $\int \varphi\,d\mu_t = g(0)$ and $\int Q_1\varphi\,d\mu_s = g(1)$, and we write the inequality as

$$\int_0^1 g'(r)\,dr \le \frac{\ell}{2} \int_t^s F(f_r)\,dr.$$

Using the HJ subsolution property of $Q_r\varphi$ and the "integration by parts" we get

$$\begin{aligned} g'(r) &= \int (\frac{d}{dr}Q_r\varphi) f_{t+\ell r}\, d\mathfrak{m} + \ell \int Q_r\varphi \Delta f_{t+\ell r}\, d\mathfrak{m} \\ &\leq -\frac{1}{2}\int |\nabla Q_r\varphi|_*^2 f_{t+\ell r}\, d\mathfrak{m} + \ell \int |\nabla Q_r\varphi|_* \sqrt{f_{t+\ell r}} \frac{|\nabla f_{t+\ell r}|_*}{\sqrt{f_{t+\ell r}}}\, d\mathfrak{m}. \end{aligned}$$

Eventually the Young inequality gives

$$g'(r) \leq \frac{\ell^2}{2} F(f_{t+\ell r})$$

and an integration in $(0,1)$ with respect to r gives the result. □

4 W_2-gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$

Since the ambient space $\mathscr{P}(X)$ is not linear (at least if we take the viewpoint of optimal transportation), what do we mean by gradient flow?

Key idea. (De Giorgi) Encode the *system* $x'(t) = -\nabla F(x(t))$ in a *single* differential inequality, by looking at the rate of energy dissipation:

$$\text{(DG)} \qquad \frac{d}{dt}F(x(t)) \leq -\frac{1}{2}|\nabla F|^2(x(t)) - \frac{1}{2}|x'(t)|^2.$$

Indeed, in a sufficiently smooth setting, along *any* curve $y(t)$, we have

$$\begin{aligned} &\frac{d}{dt}F(y(t)) = \langle \nabla F(y(t)), y'(t)\rangle \\ \geq\ & -|\nabla F(y(t))||y'(t)| \quad (= \text{iff } -y'(t) \text{ is parallel to } \nabla F(y(t))) \\ \geq\ & -\frac{1}{2}|\nabla F|^2(y(t)) - \frac{1}{2}|y'(t)|^2 \quad (= \text{iff } |\nabla F|(y(t)) = |y'(t)|). \end{aligned}$$

All terms in (DG) make sense in a metric space (X,d): $|x'|$ can be replaced by the metric derivative and $|\nabla F|$ by the *descending slope* $|\nabla^- F|$, so that the speed is 0 at minimum points. By looking at integral versions of this optimal dissipation rate we can write down an energy dissipation inequality and an energy dissipation identity:

$$\text{(EDI)} \qquad F(x(t)) + \int_0^t \frac{1}{2}|x'(r)|^2 + \frac{1}{2}|\nabla^- F|^2(x(r))\, dr \leq F(x(0)) \quad \forall t \geq 0.$$

$$\text{(EDE)} \qquad F(x(t)) + \int_0^t \frac{1}{2}|x'(r)|^2 + \frac{1}{2}|\nabla^- F|^2(x(r))\, dr = F(x(0)) \quad \forall t \geq 0.$$

5 Properties of the slope of K-convex functions

The following lemma provides very useful properties of the descending slope of K-convex and l.s.c. functions.

Lemma 5.1 *If F is K-convex, l.s.c in a geodesic metric space, we have the upper gradient property*

$$F(y(0)) \leq F(y(t)) + \int_0^t |y'_r||\nabla^- F|(y(r))\,dr \tag{5.1}$$

along any absolutely continuous curve $y : [0,t] \to X$. As a consequence, (EDE) and (EDI) are equivalent for F, $|x'(t)| = |\nabla^- F|(x(t))$ for a.e. $t > 0$, $t \mapsto F(x(t))$ is locally a.c. in $(0,\infty)$, with derivative equal to $-|\nabla^- F|^2(x(t))$.

Sketch of proof. For simplicity, assume $K = 0$. In this case, using monotonicity of difference quotients one has the *duality formula for the descending slope*:

$$|\nabla^- F|(x) = \sup_{y \neq x} \frac{[F(x) - F(y)]^+}{d(x,y)}. \tag{5.2}$$

It implies at once that $x \mapsto |\nabla^- F|(x)$ is l.s.c. in X, and provides the *one-sided* (because no modulus is present) and *local* (because the factor in front of the distance is not constant) Lipschitz property:

$$F(x) - F(y) \leq |\nabla^- F|(x) d(x,y) \qquad \forall x \in D(F). \tag{5.3}$$

A real analysis lemma (see [3, Lemma 1.2.6]) then shows that (5.3) implies the upper gradient property.

Now, if we have an (EDI) solution, we can bound $F(x(0))$ from above using (5.1) to get

$$\int_0^t \frac{1}{2}|x'(r)|^2 + \frac{1}{2}|\nabla^- F|^2(x(r))\,dr \leq \int_0^t |x'(r)||\nabla^- F|(x(r))|\,dr.$$

Since t is arbitrary, this implies that $|x'| = |\nabla^- F|(x)$ a.e. in $(0,\infty)$. □

These results apply of course to the $\mathrm{Ent}_{\mathfrak{m}}$ in $\mathscr{P}(X)$, under the $CD(K,\infty)$ assumption, and provide lower semicontinuity of $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$, the upper gradient property of $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ and the equivalence of the (EDE) and (EDI) formulations.

6 Fisher bounds squared slope from above

We have seen that the energy dissipation rate of $\mathrm{Ent}_{\mathfrak{m}}$ along the L^2 heat flow is given by the Fisher information functional. It is natural to related this functional to energy dissipation of $\mathrm{Ent}_{\mathfrak{m}}$ seen from the Wasserstein viewpoint.

Proposition 6.1 *In a $CD(K,\infty)$ space $(X,d,\mathfrak{m})$, assume that $\rho \in L^1(X,\mathfrak{m})$ is a probability density with $\sqrt{\rho} \in D(\mathsf{Ch})$. Then*

$$|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\rho\mathfrak{m}) \leq \int_{\{\rho>0\}} \frac{|\nabla \rho|_*^2}{\rho}\,d\mathfrak{m} \left(= 4 \int |\nabla \sqrt{\rho}|_*^2\,d\mathfrak{m} \right).$$

Notice that it is precisely this inequality that prevents, in $CD(K,\infty)$ spaces, triviality of the theory!

Sketch of proof. By approximation (recall that Ch is defined by approximation with Lipschitz functions and that $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ is l.s.c.) we can assume that $\sqrt{\rho} \in \mathrm{Lip}(X)$. By truncation, we can also assume that $c^{-1} \geq \sqrt{\rho} \geq c > 0$, so that $\log \rho \in \mathrm{Lip}(X)$.

Let us consider another density η and an optimal plan $\boldsymbol{\pi}_\eta$ from ρ to η. Then, following [26, Theorem 20.1], we can estimate:

$$\begin{aligned}
\mathrm{Ent}_{\mathfrak{m}}(\rho\mathfrak{m}) - \mathrm{Ent}_{\mathfrak{m}}(\eta\mathfrak{m}) &\le \int \log\rho(\rho-\eta)\,d\mathfrak{m} = \int \log\rho(x) - \log\rho(y)\,d\boldsymbol{\pi}_\eta \\
&\le \int \big(|\nabla^- \log\rho|(x) + \omega_x(y)\big) d(x,y)\,d\boldsymbol{\pi}_\eta(x,y) \\
&\le W_2(\eta\mathfrak{m},\rho\mathfrak{m})\Big(\int (|\nabla^- \log\rho|(x) + \omega_x(y))^2\,d\boldsymbol{\pi}_\eta\Big)^{1/2}
\end{aligned}$$

where $\omega_x(y)$ is a uniformly bounded modulus of continuity with $\omega_x(x) = 0$. Dividing both sides by $W_2(\eta\mathfrak{m},\rho\mathfrak{m})$ and letting $\eta\mathfrak{m} \to \rho\mathfrak{m}$ gives the result, by the weak convergence of $\boldsymbol{\pi}_\eta$ to the identity plan $\boldsymbol{\pi}_\rho$, concentrated on the diagonal (since the first marginal is fixed, the limit works even though $|\nabla^- \log\rho|$ is discontinuous). □

7 Identification of gradient flows

Coming back to the notation used in the introduction, where we dealt with the JKO result for entropy and the heat equation, the conventional strategy goes as follows:

$$\begin{cases} \{\text{ gradient flow of Ent}\} \subset \{\text{ gradient flow of Dir }\} \\ \\ \text{Uniqueness of gradient flow of Dir} \end{cases} \qquad \Longrightarrow \qquad = \text{holds}$$

The new strategy, initiated in [14], proves instead

$$\begin{cases} \{\text{ gradient flow of } \mathsf{Ch}\} \subset \{\text{ gradient flow of } \mathrm{Ent}_{\mathfrak{m}} \} \\ \\ \text{Uniqueness of gradient flow of } \mathrm{Ent}_{\mathfrak{m}} \end{cases} \qquad \Longrightarrow \qquad = \text{holds}$$

The new strategy is feasible thanks to the recent uniqueness result proved by Gigli in [13] for the W_2 gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$. This result is surprising, because *no* contractivity property of W_2 can be expected at this level of generality [25], not even in Finsler (non Riemannian) spaces.

We want to show that any L^2 heat flow $f_t := \mathbf{h}_t f_0$ (with f_0 probability density) is a W_2-gradient flow with $\mu_t := f_t\mathfrak{m}$, i.e.

$$\int f_t \log f_t\,d\mathfrak{m} + \int_0^t \frac{1}{2}|\dot\mu_r|^2 + \frac{1}{2}|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\mu_r)\,dr \le \int f_0\log f_0\,d\mathfrak{m}.$$

Indeed, Kuwada's Lemma 3.6 and Proposition 6.1 give for almost every $t > 0$

$$|\dot\mu_t|^2 \le \int_{\{f_t>0\}} \frac{|\nabla f_t|_*^2}{f_t}\,d\mathfrak{m}, \qquad |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(f_t\mathfrak{m}) \le \int_{\{f_t>0\}} \frac{|\nabla f_t|_*^2}{f_t}\,d\mathfrak{m}, \tag{7.1}$$

while the Hilbertian energy dissipation gives

$$\frac{d}{dt}\int f_t\log f_t\,d\mathfrak{m} = -\int_{\{f_t>0\}} \frac{|\nabla f_t|_*^2}{f_t}\,d\mathfrak{m} \qquad \text{for a.e. } t>0.$$

Coming now to uniqueness, the key contribution of [13] is the proof of convexity (in the usual sense) of $\mu \mapsto |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu)$. Once we know this, if we have two EDI solutions μ_t^1, μ_t^2, both starting from $\bar\mu$, namely

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t^i) + \int_0^t \frac{1}{2}|\dot\mu_s^i|^2 + \frac{1}{2}|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_s^i)\,ds \le \mathrm{Ent}_{\mathfrak{m}}(\bar\mu) \qquad i=1,2$$

we can combine them into $\mu_t := (\mu_t^1 + \mu_t^2)/2$ to get (using convexity of $\mathrm{Ent}_{\mathfrak{m}}$ and of the squared metric derivative as well)

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t) + \int_0^t \frac{1}{2}|\dot\mu_s|^2 + \frac{1}{2}|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_s)\,ds \le \mathrm{Ent}_{\mathfrak{m}}(\bar\mu).$$

By the upper gradient property the inequality has to be an equality, and this can happen only if $2\mathrm{Ent}_{\mathfrak{m}}(\mu_t) = \mathrm{Ent}_{\mathfrak{m}}(\mu_t^1) + \mathrm{Ent}_{\mathfrak{m}}(\mu_t^2)$. Strict convexity of the entropy then gives the result.

Another byproduct of the inclusion of L^2 gradient flows of Ch into W_2-gradient flows of $\mathrm{Ent}_{\mathfrak{m}}$ is that all inequalities (see (7.1) in particular) should be equalities, so that the energy dissipation rates are equal a.e. in $(0,\infty)$:

$$|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(f_t\mathfrak{m}) = \int_{\{f_t>0\}} \frac{|\nabla f_t|_*^2}{f_t}\,d\mathfrak{m} \qquad \text{for a.e. } t>0.$$

By letting $t \downarrow 0$, this can be used to show that Fisher *coincides* with slope:

$$|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(f\mathfrak{m}) = \int_{\{f>0\}} \frac{|\nabla f|_*^2}{f}\,d\mathfrak{m}.$$

8 Weak gradients and their identification

Let's start from the Euclidean case. We discuss only the case $W^{1,2}$, although all $W^{1,p}$ spaces $1<p<\infty$ (and even the $W^{1,1}$ and BV spaces) could be treated, see [6]. The two standard definitions of Sobolev spaces are of W type (weak derivatives)

$$W^{1,2}(\mathbb{R}^n) := \left\{ u \in L^2(\mathbb{R}^n) :\ \frac{\partial u}{\partial x_i} \in L^2(\mathbb{R}^n),\ 1 \le i \le n \right\}$$

and of H type (strong derivatives)

$$H^{1,2}(\mathbb{R}^n) := \left\{ \text{completion of } C^\infty \cap W^{1,2} \text{ for the } W^{1,2} \text{ norm} \right\}.$$

The celebrated "$H=W$" theorem by Meyers-Serrin in 1960 provides equivalence of the two definitions, even in any open domain. Another less known approach goes back to a paper by B.Levi [17] in 1906. Levi was looking for a function space where the minimization of the Dirichlet energy in a planar domain with given boundary conditions could find a solution, I will adopt his definition to n space dimensions, denoting for any $i=1,\ldots,n$ by x_i the i-th variable and by x_i' the block of the remaining $(n-1)$ variables.

Definition 8.1 (Beppo Levi space) *Let $u : \mathbb{R}^n \to \mathbb{R}$. We say that $u \in BL^{1,2}(\mathbb{R}^n)$ if:*

(a) for $i=1,\ldots,n$ and for $\mathscr{L}^{n-1}$-a.e. $x_i' \in \mathbb{R}^{n-1}$, $u(x_i',\cdot)$ is absolutely continuous in $\mathbb{R}$;

(b) $\sum_{i=1}^{n} \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} |\frac{\partial u}{\partial x_i}|^2 \, dx_i dx_i' < \infty$.

Using Fubini's theorem we can still define a gradient in BL spaces. However, a drawback of this approach is that it is not clear whether the BL property is frame-indifferent or not. It is actually true, but to realize this one has to pass through the equivalence with the H and W definitions, that we are now going to discuss. Actually, another way to bypass this objection would be to consider in some sense all or, more precisely, almost all curves (not only geodesics), this is exactly the strategy pursued in the metric theory. For a proof of this result, see for instance [1] or [15].

Theorem 8.2 *$BL^{1,2}(\mathbb{R}^n) \subset W^{1,2}(\mathbb{R}^n)$. In addition, any $u \in W^{1,2}(\mathbb{R}^n)$ has a version (for instance $\limsup_\epsilon u * \rho_\epsilon$) in $BL^{1,2}(\mathbb{R}^n)$.*

In metric spaces, the W definition seems to be difficult to adapt. On the other hand, the H definition roughly corresponds to Cheeger's energy Ch (with $\mathrm{Lip}(X)$ playing the role of C^∞), while Levi's definition corresponds to Shanmugalingam's notion [22] of Newtonian space $N^{1,2}(X, d, \mathfrak{m})$, that now I am going to illustrate. Actually in [4] a different notion of gradient is used, a priori smaller than the gradient of [22]; however, to simplify the exposition, I will just confine myself to the relaxed gradient and the gradient of [22].

Definition 8.3 *Given $\Gamma \subset AC([0,1]; X)$ we define*

$$\mathrm{Mod}_2(\Gamma) := \inf \left\{ \int g^2 \, d\mathfrak{m} : \; g : X \to [0, \infty] \text{ Borel, } \int_\gamma g \geq 1 \text{ for all } \gamma \in \Gamma \right\}.$$

Here the curvilinear integral is defined, as in the theory of upper gradients, using the metric derivative, namely $\int_\gamma g := \int_0^1 g(\gamma_s) |\dot{\gamma}_s| \, ds$.

We now define $N^{1,2}(X, d, \mathfrak{m})$ by

$$\left\{ f : X \to \mathbb{R} : \; |\int_{\partial\gamma} f| \leq \int_\gamma G \text{ for } \mathrm{Mod}_2\text{-a.e. } \gamma, \text{ for some } G \in L^2(X, \mathfrak{m}) \right\}$$

and $|\nabla f|_S \in L^2(X, \mathfrak{m})$ as the function G with smallest L^2 norm.

Lemma 8.4 (Absolute continuity lemma) *Any $f \in N^{1,2}(X, d, \mathfrak{m})$ is absolutely continuous along Mod_2-a.e. curve γ.*

Proof. Let Γ be the set of curves γ where the u.g. property with $|\nabla f|_S$ does not hold,

$$\Gamma_1 := \left\{ \gamma : \; \gamma \supset \gamma' \in \Gamma \right\},$$

$$\Gamma_2 := \left\{ \gamma : \; \int_\gamma |\nabla f|_S = \infty \right\}.$$

Now, Γ_1 is Mod_2-negligible because Γ is (any g admissible for Γ is admissible for Γ'), Γ_2 is Mod_2-negiglible by the "Markov" inequality

$$\mathrm{Mod}_2\big(\Gamma_2 \cap \{ \int_\gamma |\nabla f|_S \geq n \}\big) \leq \frac{1}{n^2} \int |\nabla f|_S^2 \, d\mathfrak{m} \to 0.$$

If $\gamma \notin (\Gamma_1 \cup \Gamma_2)$ we have

$$\left|\int_{\partial\gamma'} f\right| \le \int_{\gamma'} |\nabla f|_S \le \int_{\gamma} |\nabla f|_S < \infty \qquad \forall \gamma' \subset \gamma$$

which yields immediately the absolute continuity property of $t \mapsto f(\gamma_t)$.

The gradient $|\nabla f|_S$ has pointwise minimality properties analogous to $|\nabla f|_*$ (see Theorem 3.2(3)), in particular if G satisfies the weak upper gradient property

$$\left|\int_{\partial\gamma} f\right| \le \int_{\gamma} G \qquad \text{for } \mathrm{Mod}_2\text{-a.e. curve } \gamma$$

then $|\nabla f|_S \le G$ $\mathfrak{m}$-a.e. in X. □

Are the gradients $|\nabla f|_*$, $|\nabla f|_S$ equal? While the first gradient is relevant in connection with the L^2 heat flow and the "vertical" derivative, the second one is relevant in connection with the derivative of $\mathrm{Ent}_{\mathfrak{m}}$ and the "horizontal" derivative.

If we assume doubling & Poincaré (with the S-gradient in the right hand side), then we can approximate any $f \in N^{1,2}(X,d,\mathfrak{m})$, see for instance [8], in the strong norm and even in the Lusin sense by Lipschitz maps f_n. This leads to the equality of gradients.

The strategy is to consider the maximal function

$$M(x) := \sup_{r>0} \frac{\int_{B_r(x)} |\nabla f|_S(y)\, d\mathfrak{m}(y)}{\mathfrak{m}(B_r(x))}$$

and to prove that $f|_{\{M \le n\}}$ is Cn-Lipschitz. Defining f_n as a Lipschitz extension of $f|_{\{M \le n\}}$, locality of gradients and

$$\mathfrak{m}\big(\{M > n\}\big) = o(\frac{1}{n^2})$$

provide the result.

With "optimal transportation tools" we can provide in [4], see also [6], the equivalence of gradients and the density in energy of Lipschitz maps without doubling & Poincaré. This requires an approximation by Lipschitz functions f_n in "energy", namely

$$\limsup_{n\to\infty} \int |\nabla f_n|^2\, d\mathfrak{m} \le \int |\nabla f|_S^2\, d\mathfrak{m}, \qquad \int |f_n - f|^2\, d\mathfrak{m} \to 0. \tag{8.1}$$

By uniform convexity, this provides also

$$\lim_{n\to\infty} \int \big||\nabla f_n| - |\nabla f|_S\big|^2\, d\mathfrak{m} = 0.$$

Notice that, as soon as we know that the Sobolev spaces are reflexive, we can use Mazur's lemma (i.e. take convex combinations) to improve the approximation from weak to strong (while, without doubling & Poincaré, the Lusin approximation seems really to be out of reach).

In order to prove (8.1) we need, besides Kuwada's lemma, three more auxiliary results.

Lemma 8.5 *If $\boldsymbol{\eta} \in \mathscr{P}(C([0,1];X))$ concentrated on $AC^2([0,1];X)$ has uniformly bounded time marginals, i.e. $(e_t)_\sharp \boldsymbol{\eta} \le C(\boldsymbol{\eta})\mathfrak{m}$ for all $t \in [0,1]$, then*

$$[\boldsymbol{\eta}(\Gamma)]^2 \le C(\boldsymbol{\eta}) \left(\int \int_0^1 |\dot\gamma_s|^2\, ds\, d\boldsymbol{\eta}(\gamma\right) \mathrm{Mod}_2(\Gamma) \qquad \forall \Gamma \subset AC^2([0,1];X).$$

In particular $\mathrm{Mod}_2(\Gamma) = 0$ *implies* $\boldsymbol{\eta}(\Gamma) = 0$.

Proof. If g is admissible for Γ we have

$$[\boldsymbol{\eta}(\Gamma)]^2 \leq \left(\int_0^1 \int g(\gamma_s)|\dot\gamma_s|\,ds\,d\boldsymbol{\eta}(\gamma)\right)^2.$$

Then, it suffices to apply Hölder and to minimize w.r.t. g. □

Proposition 8.6 (Superposition principle [3], [18]) *Let $(\mu_t)_{t\in[0,T]} \subset \mathscr{P}(X)$ be absolutely continuous with L^2-integrable metric derivative. Then there exists $\boldsymbol{\eta} \in \mathscr{P}(C([0,T];X))$ concentrated on $AC^2([0,T];X)$ and satisfying*

(1) $\mu_t = (e_t)_\sharp\boldsymbol{\eta}$ for all $t \in [0,T]$;

(2) $|\dot\mu_t|^2 = \int |\dot\gamma_t|^2\,d\eta(\gamma)$ for a.e. $t \in (0,T)$.

Lemma 8.7 (Stability of weak upper gradients [22], [15]) *If $f_n \to f$ in $L^2(X,\mathfrak{m})$, $G_n \to G$ weakly in $L^2(X,\mathfrak{m})$ and*

$$\left|\int_{\partial\gamma} f_n\right| \leq \int_\gamma G_n \qquad \text{for } \mathrm{Mod}_2\text{-a.e. curve } \gamma,$$

then there is a version $\tilde f$ of f satisfying

$$\left|\int_{\partial\gamma} \tilde f\right| \leq \int_\gamma G \qquad \text{for } \mathrm{Mod}_2\text{-a.e. curve } \gamma.$$

Using this lemma with f_n equal to the optimal sequence in the definition of Ch and $G_n = |\nabla f_n|$, weakly convergent to $|\nabla f|_*$, we obtain

$$|\nabla \tilde f|_S \leq G = |\nabla f|_* \qquad \mathfrak{m}\text{-a.e. in } X.$$

The proof of the converse inequality is constructive: we need Lipschitz functions f_n satisfying $f_n \to f$ in $L^2(X,\mathfrak{m})$ and

$$\limsup_{n\to\infty} \int |\nabla f_n|^2\,d\mathfrak{m} \leq \int |\nabla f|_S^2\,d\mathfrak{m}.$$

By a diagonal argument, suffices to find $f_n \in D(\mathsf{Ch})$ satisfying $\limsup_n \int |\nabla f_n|_*^2\,d\mathfrak{m} \leq \int |\nabla f|_S^2\,d\mathfrak{m}$. By a truncation argument we assume that, $0 < c \leq f \leq c^{-1} < \infty$ and by homogeneity $\int f^2\,d\mathfrak{m} = 1$. We set $k = f^2$, $k_t = \mathbf{h}_t k$, $\mu_t = k_t\mathfrak{m} \in \mathscr{P}(X)$, $\boldsymbol{\eta}$ given by the superposition principle. Then we argue as in Proposition 6.1, this time using the S-gradient in place of the relaxed slope:

$$\begin{aligned}
&\int k\log k - k_t \log k_t\,d\mathfrak{m} \\
\leq\ & \int \log k(k-k_t)\,d\mathfrak{m} = \int \log k(\gamma_0) - \log k(\gamma_t)\,d\boldsymbol{\eta}(\gamma) \\
\leq\ & \int\int_0^t |\nabla \log k|_S(\gamma_s)|\dot\gamma_s|ds d\boldsymbol{\eta}(\gamma) \\
\leq\ & \left(\int_0^t\int |\nabla \log k|_S^2(\gamma_s) d\eta ds\right)^{1/2}\left(\int_0^t\int |\dot\gamma_s|^2\,d\boldsymbol{\eta}(\gamma)ds\right)^{1/2} \\
\leq\ & \frac{1}{2}\int_0^t\int |\nabla \log k|_S^2 k_s d\mathfrak{m} ds + \frac{1}{2}\int |\dot\mu_s|^2\,ds.
\end{aligned}$$

By the Kuwada lemma we get

$$\begin{aligned}&\int k\log k - k_t\log k_t\,d\mathfrak{m}\\ &\le \frac{1}{2}\int_0^t\int|\nabla\log k|_S^2 k_s d\mathfrak{m}ds+\frac{1}{2}\int_0^t\int_{\{k_s>0\}}\frac{|\nabla k_s|_*^2}{k_s}\,d\mathfrak{m}ds.\end{aligned}$$

The entropy dissipation formula (3.4) then gives

$$\int_0^t\int_{\{k_s>0\}}\frac{|\nabla k_s|_*^2}{k_s}\,d\mathfrak{m}ds\le\int_0^t\int|\nabla\log k|_S^2 k_s d\mathfrak{m}ds,$$

so that the identity $|\nabla\log k|_S=|\nabla k|_S/k=2|\nabla f|_S/f$ we get

$$\frac{4}{t}\int_0^t \mathsf{Ch}(\sqrt{k_s})\,ds\le\frac{4}{t}\int_0^t\int\frac{|\nabla f|_S^2}{f^2}k_s\,d\mathfrak{m}ds.$$

Letting $t\downarrow 0$ and using the w^*-convergence in $L^\infty(X,\mathfrak{m})$ of k_s to $k=f^2$ gives the result.

9 Riemannian Ricci lower bounds

As noticed by Cordero Erausquin, Sturm and Villani, all Minkowski spaces ($\mathbb{R}^n$ endowed with the Lebesgue measure and any norm $\|\cdot\|$) satisfy the $CD(0,n)$, and therefore the $CD(0,\infty)$ condition. On the other hand, Cheeger and Colding ruled out in [9] the possibility to obtain these spaces as limits of Riemannian manifolds with uniform lower bounds on Ricci curvature and uniform upper bounds on volume.

In [5] we tried to give an answer to the following question: is there a more restrictive notion, still *stable* and *strongly consistent* with the Riemannian case, that rules out Minkowski (non Hilbert) spaces? Recall that the $CD(K,\infty)$ condition is stable [23, 24, 19], meaning that measured Gromov-Hausdorff limits of $CD(K,\infty)$ spaces are $CD(K,\infty)$, and strongly consistent, meaning that a Riemannian manifold M endowed with Riemannian distance $d=d_M$ and the volume measure $\mathfrak{m}=\mathrm{vol}_M$ is $CD(K,\infty)$ iff $\mathrm{Ric}_M\ge KI$. So, a positive answer to this question would provide more insight, among other things, on the closure of Riemannian manifolds under uniform Ricci lower bounds.

This led us to the definition of spaces with *Riemannian* Ricci lower bounds. We have 3 equivalent definitions (and their equivalence is far from being trivial), summarized below, and this class of spaces provides a positive answer to the question I raised.

Definition 9.1 ($RCD(K,\infty)$ spaces) *We say that $(X,d,\mathfrak{m})$ has* Riemannian Ricci curvature *bounded from below by $K\in\mathbb{R}$, and write $RCD(K,\infty)$, if one of the following equivalent conditions hold:*

- *(i) $(X,d,\mathfrak{m})$ is a strong $CD(K,\infty)$ space and the L^2 gradient flow $\mathbf{h}_t$ of Ch is linear;*
- *(ii) $(X,d,\mathfrak{m})$ is a strong $CD(K,\infty)$ space and the W_2 gradient flow $\mathbf{H}_t$ of $\mathrm{Ent}_{\mathfrak{m}}$ is additive (i.e. convex and concave) on $\mathscr{P}(X)$;*
- *(iii) for all $\mu\in\mathscr{P}(X)$ with $\operatorname{supp}\mu\subset\operatorname{supp}\mathfrak{m}$, $\mathbf{H}_t\mu$ is a gradient flow in the EVI_K sense.*

I will illustrate later on what strong $CD(K,\infty)$ and EVI_K mean, in the next subsections I will instead list some properties of this class of spaces.

9.1 Stability under Gromov-Hausdorff limits of $RCD(K,\infty)$ spaces

We say that two metric measure spaces $(X, d_X, \mathfrak{m}_X)$ and $(Y, d_Y, \mathfrak{m}_Y)$ are *isomorphic* if there exists a bijective isometry $f : \operatorname{supp} \mathfrak{m}_X \to \operatorname{supp} \mathfrak{m}_Y$ such that $f_\sharp \mathfrak{m}_X = \mathfrak{m}_Y$. We will denote by $\mathbb{X}$ the set of isomorphism classes of metric measure spaces that we will consider (as we said at the beginning we confine ourselves to compact metric spaces (X, d) and probability reference measures $\mathfrak{m}$):

$$\mathbb{X} := \Big\{(X, d, \mathfrak{m}) \ : \ (X, d) \text{ is compact and } \mathfrak{m} \in \mathscr{P}(X)\Big\}.$$

Definition 9.2 *Given two metric measure spaces $(X, d_X, \mathfrak{m}_X)$, $(Y, d_Y, \mathfrak{m}_Y)$, we consider the product space $(X \times Y, d_{XY})$, where*

$$d_{XY}\big((x_1, y_1), (x_2, y_2)\big) := \sqrt{d_X^2(x_1, x_2) + d_Y^2(y_1, y_2)}.$$

We say that $(\mathbf{d}, \boldsymbol{\gamma})$ is an admissible coupling if:

(a) $\mathbf{d}$ *is a pseudo distance on $X \sqcup Y$ which coincides with d_X (resp. d_Y) when restricted to $\operatorname{supp} \mathfrak{m}_X \times \operatorname{supp} \mathfrak{m}_X$ (resp. $\operatorname{supp} \mathfrak{m}_Y \times \operatorname{supp} \mathfrak{m}_Y$) and $\mathbf{d}|_{X\times Y} : X \times Y \to [0, \infty)$ is Borel.*

(b) $\boldsymbol{\gamma}$ *is a Borel measure on $X \times Y$, $\pi^X_\sharp \boldsymbol{\gamma} = \mathfrak{m}_X$ and $\pi^Y_\sharp \boldsymbol{\gamma} = \mathfrak{m}_Y$.*

It is not hard to see that the set of admissible couplings is always non empty. The *cost* $C(\mathbf{d}, \boldsymbol{\gamma})$ of a coupling is given by

$$C(\mathbf{d}, \boldsymbol{\gamma}) := \int \mathbf{d}^2(x, y)\, \boldsymbol{\gamma}(x, y).$$

In analogy to the definition of W_2, Sturm's distance $\mathbb{D}$ is then defined as

$$\mathbb{D}^2\big((X, d_X, \mathfrak{m}_X), (Y, d_Y, \mathfrak{m}_Y)\big) := \inf C(\mathbf{d}, \boldsymbol{\gamma}),$$

the infimum being taken among all couplings $(\mathbf{d}, \boldsymbol{\gamma})$ of $(X, d_X, \mathfrak{m}_X)$ and $(Y, d_Y, \mathfrak{m}_Y)$.

Theorem 9.3 *Let $(X_n, d_n, \mathfrak{m}_n) \in \mathbb{X}$ be $RCD(K,\infty)$ spaces. If $\mathbb{D}\big((X_n, d_n, \mathfrak{m}_n), (X, d, \mathfrak{m})\big) \to 0$, then $(X, d, \mathfrak{m})$ is a $RCD(K,\infty)$ space as well.*

In the proof of this result it is the EVI_K formulation that plays a decisive role.

9.2 Tensorization of $RCD(K,\infty)$ spaces

Remember [24] that the product of two non-branching $CD(K,\infty)$ spaces is still $CD(K,\infty)$, and it is still open the validity of the implication if the non-branching assumption is removed. Here non-branching means that the map $(e_0, e_t) : \mathrm{Geo}(X) \to X^2$ is injective for all $t > 0$; in other words, geodesics can't split. The following result proves that the same property persists for the stronger $RCD(K,\infty)$ notion.

Theorem 9.4 *If $(X, d_X, \mathfrak{m}_X)$ and $(Y, d_Y, \mathfrak{m}_Y)$ are $RCD(K,\infty)$ and non-branching, so is $(X \times Y, \sqrt{d_X^2 + d_Y^2}, \mathfrak{m}_X \times \mathfrak{m}_Y)$.*

In the proof of this result we failed to prove directly tensorization of the EVI_K, so we rely on the above mentioned Sturm's result saying that the product is $CD(K,\infty)$. Since the nonbranching property tensorizes, and since it is not difficult to show that

$$CD(K,\infty) + \text{nonbranching} \quad \Longrightarrow \quad \text{strong } CD(K,\infty)$$

the product is also strong $CD(K,\infty)$.

Finally, we have to prove that linearity of the heat flow tensorizes as well. This is equivalent to say that

$$\mathsf{Ch}^{X\times Y}(f) = \int \mathsf{Ch}^X(f^y)\, d\mathfrak{m}_Y(y) + \int \mathsf{Ch}^Y(f^x)\, d\mathfrak{m}_X(x)$$

or equivalently that (with $f^x(y) = f(x,y) = f^y(x)$)

$$|\nabla f|_*^2(x,y) = |\nabla f^x|_*^2(y) + |\nabla f^y|_*^2(x) \qquad \mathfrak{m}_X \times \mathfrak{m}_Y\text{-a.e. in } X \times Y.$$

The proof of the inequality $|\nabla f|_*^2(x,y) \le |\nabla f^x|_*^2(y) + |\nabla f^y|_*^2(x)$ is based on the following calculus lemma and on a smoothing argument with the product semigroup $\mathbf{h}_t^X \times \mathbf{h}_t^Y$.

Lemma 9.5 *If $f : X \times Y \to \mathbb{R}$ is Lipschitz and $\gamma = (\gamma^X, \gamma^Y) : [0,1] \to X \times Y$ is absolutely continuous, then, for a.e. t, $|\frac{d}{dt}(f\circ\gamma)(t)|$ is bounded from above by*

$$\limsup_{h\downarrow 0} \frac{|f(\gamma^X_{t-h}, \gamma^Y_t) - f(\gamma^X_t, \gamma^Y_t)|}{h} + \limsup_{h\downarrow 0} \frac{|f(\gamma^X_t, \gamma^Y_{t+h}) - f(\gamma^X_t, \gamma^Y_t)|}{h}.$$

The proof of the converse inequality $|\nabla f|_*^2(x,y) \ge |\nabla f^x|_*^2(y) + |\nabla f^y|_*^2(x)$ is much more delicate. It relies indeed in a version of Kuwada's lemma in product spaces, when we consider the product semigroup $\mathbf{h}_t^X \times \mathbf{h}_t^Y$, correspondent to the gradient flow of $\frac{1}{2}\int |\nabla f^x|_*^2(y) + |\nabla f^y|_*^2(x)\, d\mathfrak{m}_X \times \mathfrak{m}_Y(x,y)$:

Proposition 9.6 (Kuwada's lemma in product spaces) *Let $f \in L^2(X\times Y, \mathfrak{m}_X \times \mathfrak{m}_Y)$ be a probability density and let f_t be the evolution of f under the product semigroup, $\mu_t = f_t \mathfrak{m}_X \times \mathfrak{m}_Y$. Then for a.e. $t > 0$ it holds*

$$|\dot\mu_t|^2 \le \int_{\{f_t>0\}} \frac{|\nabla f_t^x|_*^2(y) + |\nabla f_t^y|_*^2(x)}{f_t(x,y)}\, d\mathfrak{m}_X \times \mathfrak{m}_Y(x,y).$$

Once we know this, the machinery of identification theorems provides identification of the two gradients.

9.3 The heat flow in $RCD(K,\infty)$ spaces

The identification between the L^2 gradient flow $\mathbf{h}_t$ and the W_2 gradient flow $\mathbf{H}_t$ allows to pick the best properties from each of them: for instance, the symmetry of the transition probabilities $\theta_t : X \times X \to [0,\infty)$, defined by $\mathbf{H}_t\delta_x := \theta_t(x,\cdot)\mathfrak{m}$, comes from the fact that $\mathbf{h}_t$ is L^2-selfadjoint, while the contractivity properties of $\mathbf{h}_t$ in spaces different from $L^2(X,\mathfrak{m})$ follow from those of $\mathbf{H}_t$.

Proposition 9.7 (Properties of the heat flow) *It holds:*

(1) The pointwise formula $\tilde{\mathbf{h}}_t f(x) := \int f \, d\mathbf{H}_t \delta_x$ *provides a version of* $\mathbf{h}_t f$ *and an extension of* $\mathbf{h}_t$ *to a contraction semigroup in all* $L^p(X, \mathfrak{m})$ *spaces.*

(2) $\tilde{\mathbf{h}}_t$ *leaves* $\mathrm{Lip}(\operatorname{supp} \mathfrak{m})$ *invariant and, by the contractivity estimate* $W_2(\mathbf{H}_t \delta x, \mathbf{H}_t \delta_y) \leq e^{-Kt} d(x, y)$, $\mathrm{Lip}(\tilde{\mathbf{h}}_t f) \leq e^{-Kt} \mathrm{Lip}(f)$*. Furthermore,* $\tilde{\mathbf{h}}_t$ *maps* $L^\infty(X, \mathfrak{m})$ *in* $C_b(\operatorname{supp} \mathfrak{m})$*.*

(3) The Bakry-Emery estimate holds:

$$|\nabla(\mathbf{h}_t f)|_*^2 \leq e^{-2Kt} \mathbf{h}_t |\nabla f|_*^2 \qquad \mathfrak{m}\text{-a.e. in } X. \tag{9.1}$$

9.4 $RCD(K, \infty)$ spaces and Dirichlet forms

Since Ch is a quadratic form in $RCD(K, \infty)$ spaces, the analysis of the connection with Fukushima's theory of Dirichlet forms is useful and mandatory. Let

$$\mathcal{E}(u, v) := \frac{1}{4} \big(\mathsf{Ch}(u + v) - \mathsf{Ch}(u - v) \big)$$

be the symmetric bilinear form associated to Ch. It is a Dirichlet form (i.e. closable and Markovian) because Ch is $L^2(X, \mathfrak{m})$-lower semicontinuous and decreases, by chain rule, under left composition with 1-Lipschitz maps.

In the theory of Dirichlet forms, two objects are naturally defined, namely the *local energy measure*

$$[u](\varphi) := \mathcal{E}(u, u\varphi) - \mathcal{E}(\frac{u^2}{2}, \varphi)$$

and the induced distance

$$d_{\mathcal{E}}(x, y) := \sup \{ |\psi(x) - \psi(y)| : \ [\psi] \leq \mathfrak{m} \} .$$

We proved that in the class of $RCD(K, \infty)$ spaces these objects coincide with the natural ones.

Theorem 9.8 *In a* $RCD(K, \infty)$ *space* $(X, d, \mathfrak{m})$ *the local energy measure* $[u]$ *coincides with* $|\nabla u|_*^2 \mathfrak{m}$ *and the induced distance* $d_{\mathcal{E}}$ *coincides with* d*.*

The proof involves the construction of a symmetric bilinear form

$$(u, v) \in \big[D(\mathsf{Ch}) \big]^2 \mapsto \nabla u \cdot \nabla v \in L^1(X, \mathfrak{m})$$

satisfying the Leibniz rule and providing integral representation to $\mathcal{E}$, namely $\mathcal{E}(u, v) = \int \nabla u \cdot \nabla v \, d\mathfrak{m}$.

In addition, since $\mathcal{E}$ is also strongly local, the theory of Dirichlet forms can be applied as a black box to obtain a unique (in law) *Brownian motion* in $(\operatorname{supp} \mathfrak{m}, d, \mathfrak{m})$, i.e. a Markov process $\boldsymbol{X}_t$ with continuous sample paths satisfying

$$\mathbf{P}(\boldsymbol{X}_t | \boldsymbol{X}_0 = x) = \mathbf{H}_t \delta_x \qquad \forall x \in \operatorname{supp} \mathfrak{m}, \ t \geq 0.$$

9.5 Strong $CD(0,\infty)$ condition and EVI_K

In a geodesic metric space, convexity of an energy F can be asked along *some* geodesic connecting a given pair of points, or along *all* geodesics connecting a given pair of points. It is well known and easy to check that the first choice (call it *convexity*) leads to a stable condition, under Gromov-Hausdorff limits, while the second choice (call it *strong convexity*) leads in general to an unstable condition.

The definition given below represents a sort of compromise: we ask for a distinguished geodesic in $\mathscr{P}(X)$ (i.e. an optimal geodesic plan $\boldsymbol{\pi}$), but we ask that the same property persists for all weighted geodesic plans $h\boldsymbol{\pi}$, $h \in C_b(\mathrm{Geo}(X))$.

Definition 9.9 (Strong $CD(K,\infty)$) *We say that $(X, d, \mathfrak{m})$ is a* strong $CD(0,\infty)$ *space if for all μ_0, $\mu_1 \in \mathscr{P}(X)$ with finite entropy there exists an optimal geodesic plan $\boldsymbol{\pi}$ between them satisfying*

$$\begin{aligned}\mathrm{Ent}_{\mathfrak{m}}\big((e_t)_\sharp(h\boldsymbol{\pi})\big) &\leq (1-t)\mathrm{Ent}_{\mathfrak{m}}\big((e_0)_\sharp(h\boldsymbol{\pi})\big) + t\mathrm{Ent}_{\mathfrak{m}}\big((e_1)_\sharp(h\boldsymbol{\pi})\big) \\ &- \frac{K}{2}t(1-t)W_2^2\big((e_0)_\sharp(h\boldsymbol{\pi}),(e_1)_\sharp(h\boldsymbol{\pi})\big)\end{aligned}$$

for all $t \in [0,1]$, $h \in C_b(\mathrm{Geo}(X))$ nonnegative, $\int h\,d\boldsymbol{\pi} = 1$.

Question: is there a condition stronger than strong convexity and stable? The answer is yes, it is the existence of EVI_K gradient flows.

If H is Hilbert and $F : H \to \mathbb{R}\cup\{+\infty\}$ is K-convex and l.s.c., we can write the differential inclusion $-x'(t) \in \partial F(x(t))$ for a.e. $t > 0$ as follows:

$$\forall y \in D(F),\ \langle -x'(t), y - x(t)\rangle + F(x(t)) + \frac{K}{2}|x(t)-y|^2 \leq F(y) \qquad \text{for a.e. } t > 0.$$

Equivalently

$$\forall y \in D(F),\ \frac{d}{dt}\frac{1}{2}|x(t)-y|^2 + F(x(t)) + \frac{K}{2}|x(t)-y|^2 \leq F(y) \qquad \text{for a.e. } t > 0.$$

Definition 9.10 *In a metric space (E,d), a locally absolutely continuous curve $u : (0,\infty) \to E$ is an EVI_K solution to the gradient flow of $F : X \to \mathbb{R}\cup\{+\infty\}$ if for all $v \in D(F)$ it holds*

$$\frac{d}{dt}\frac{1}{2}d^2(u(t),v) + F(u(t)) + \frac{K}{2}d^2(u(t),v) \leq F(v) \qquad \textit{for a.e. } t > 0.$$

This formulation of gradient flows is equivalent in Hilbert spaces, but in general *stronger* than the one based on energy dissipation, see [2]. A remarkable result, proved by Daneri and Savaré in [10], is that existence of a sufficiently rich family of EVI_K gradient flows implies K-convexity.

Theorem 9.11 *In a geodesic space (X,d), assume that EVI_K gradient flows exist starting from any $\bar{x} \in \overline{D(F)}$. Then, F is K-convex along any geodesic contained in $\overline{D(F)}$.*

Sketch of proof. We assume $K = 0$ and give a formal proof. Let $\gamma : [0,1] \to X$ be a constant speed geodesic contained in $\overline{D(F)}$ and $t \in (0,1)$, x_s the EVI solution starting from

γ_t. With no loss of generality, $\gamma_0, \gamma_1 \in D(F)$. The EVI property gives (we assume that the derivative exists at $s = 0$)

$$\frac{d^+}{ds}\frac{1}{2}d^2(x_s, \gamma_0)\Big|_{s=0} \le F(\gamma_0) - F(x_0), \quad \frac{d^+}{ds}\frac{1}{2}d^2(x_s, \gamma_1)\Big|_{s=0} \le F(\gamma_1) - F(x_0).$$

Multiply the first one by $(1-t)$, the second one by t, and use the fact that $x_0 = \gamma_t$, to get

$$(1-t)F(\gamma_0) + tF(\gamma_1) - F(\gamma_t) \ge \frac{1}{2}\frac{d^+}{ds}(1-t)d^2(\gamma_0, x_s) + td^2(x_s, \gamma_1)\Big|_{s=0}.$$

On the other hand, Young and triangle inequality easily imply that

$$(1-t)d^2(x,z) + td^2(z,y) \ge t(1-t)d^2(x,y) \qquad \forall x,\, y,\, z \in X,$$

with equality (only) if $d(x,z) = td(x,y)$, $d(z,y) = (1-t)d(x,y)$.

Hence, choosing $x = \gamma_0$, $y = \gamma_1$, the quantity

$$s \mapsto (1-t)d^2(\gamma_0, x_s) + td^2(x_s, \gamma_1)$$

is minimal at $s = 0$ (recall that $x_0 = \gamma_t$), so that its right derivative is nonnegative. □

Having now defined EVI_K, we can now give a sketchy proof of the stability; as in the proof of the analogous result for $CD(K,\infty)$ spaces, the crucial property is the joint lower semicontinuity property of $(\mu, \mathfrak{m}) \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu)$. The latter is a direct consequence of the duality formula

$$\mathrm{Ent}_{\mathfrak{m}}(\sigma) = \sup_{f \in C_b} \int f\, d\sigma - \int F^*(f)\, d\mathfrak{m}$$

where $F(z) = e^{z-1}$ is the transform of $z \log z$ (set to $+\infty$ for $z < 0$).

Sketch of proof of stability of $RCD(K,\infty)$**.** Let $(X_n, d_n, \mathfrak{m}_n) \to (X, d, \mathfrak{m})$ w.r.t. the distance $\mathbb{D}$, $(X_n, d_n, \mathfrak{m}_n)$ $RCD(K,\infty)$. Thanks to an embedding theorem and the invariance in the isomorphism class we can (possibly extracting a subsequence) assume that both $X_n = \operatorname{supp} \mathfrak{m}_n$ and $X = \operatorname{supp} \mathfrak{m}$ are contained in a fixed metric space (Y, d_Y) and that $d_n = d_Y|_{X_n \times X_n}$, $d = d_Y|_{X \times X}$.

Given an initial condition $\bar\mu = \rho\mathfrak{m}$ with $\rho \in L^\infty(X, \mathfrak{m})$, we want to build an EVI solution starting from $\bar\mu$ in the space $(X, d, \mathfrak{m})$. We find $\bar\mu_n := \rho_n \mathfrak{m}_n$, with $\|\rho_n\|_\infty \le \|\rho\|_\infty$, weakly convergent to $\bar\mu$, and the EVI solutions μ_t^n starting from $\bar\mu_n$:

$$\forall \nu,\ \frac{d}{dt}\frac{1}{2}W_2^2(\mu_t^n, \nu) + \mathrm{Ent}_{\mathfrak{m}_n}(\mu_t^n) \le \mathrm{Ent}_{\mathfrak{m}_n}(\nu) \qquad \text{for a.e. } t > 0.$$

By standard tightness estimates in space and equi-continuity estimates in time we can assume that $\mu_t^n \to \mu_t$ in $\mathscr{P}(X)$ for all $t \ge 0$, so that μ_t is our candidate EVI solution.

In order to check this, for any $\nu \in D(\mathrm{Ent}_{\mathfrak{m}})$ we can find $\nu_n \in D(\mathrm{Ent}_{\mathfrak{m}_n})$ convergent to ν in $\mathscr{P}(X)$ and satisfying $\mathrm{Ent}_{\mathfrak{m}_n}(\nu_n) \to \mathrm{Ent}_{\mathfrak{m}}(\nu)$.

In this way the right hand sides in the EVI converge and also the time derivatives (in the sense of distributions). It remains to prove that

$$\liminf_{n\to\infty} \mathrm{Ent}_{\mathfrak{m}_n}(\mu_t^n) \ge \mathrm{Ent}_{\mathfrak{m}}(\mu_t) \qquad \forall t \ge 0$$

and this follows, as we said, by the joint lower semicontinuity of the entropy.

9.6 Some auxiliary results

In this section I just state some auxiliary results needed to prove the equivalence of conditions (i), (ii), (iii) in Definition 9.1.

Lemma 9.12 (Derivative of squared Wasserstein distance) *Let* $(X, d, \mathfrak{m})$ *be a* $CD(K,\infty)$ *space,* $\mu = \rho\mathfrak{m} \in \mathscr{P}(X)$ *such that* $0 < c \leq \rho \leq C < \infty$ *and set* $\mu_t := \mathbf{H}_t(\mu) = \rho_t\mathfrak{m}$*. Let* $\nu = \sigma\mathfrak{m} \in \mathscr{P}(X)$ *and let* φ_t *be a Kantorovich potential relative to* (μ_t, ν)*. Then for a.e.* $t > 0$ *it holds*

$$\frac{d}{dt}\frac{1}{2}W_2^2(\rho_t\mathfrak{m}, \sigma\mathfrak{m}) \leq \frac{\mathsf{Ch}(\rho_t - \epsilon\varphi_t) - \mathsf{Ch}(\rho_t)}{\epsilon} \qquad \forall \epsilon > 0.$$

Lemma 9.13 (Derivative of the entropy along a geodesic) *Let* $(X, d, \mathfrak{m}) \in \mathbb{X}$ *be a strong* $CD(K,\infty)$ *space and let* ρ*,* σ *be bounded probability densities. Assume that* σ *has bounded support and that* $\rho \geq c > 0$*. Then there exists an optimal geodesic plan* $\boldsymbol{\pi}$ *from* $\rho\mathfrak{m}$ *to* $\nu := \sigma\mathfrak{m}$ *satisfying*

$$\frac{\mathsf{Ch}(\varphi) - \mathsf{Ch}(\varphi + \epsilon\rho)}{\epsilon} \leq \lim_{s\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\sigma_s\mathfrak{m}) - \mathrm{Ent}_{\mathfrak{m}}(\rho\mathfrak{m})}{s} \qquad \forall \epsilon > 0,$$

where $(e_s)_\sharp\boldsymbol{\pi} = \sigma_s\mathfrak{m}$ *and* φ *is any Kantorovich potential relative to* $(\rho\mathfrak{m}, \nu)$*.*

Theorem 9.14 (Metric Brenier theorem for strong $CD(K,\infty)$ spaces) *Let* $(X, d, \mathfrak{m}) \in \mathbb{X}$ *be a strong* $CD(K,\infty)$ *space,* $\mu = \rho\mathfrak{m} \in \mathscr{P}(X)$ *with* $c^{-1} \geq \rho \geq c > 0$ *and* $\nu \in \mathscr{P}(X)$ *with bounded support and density. Then there exist an optimal geodesic plan* $\boldsymbol{\pi}$ *and* $L : X \to [0,\infty)$ *satisfying*

$$L(\gamma_0) = d(\gamma_0, \gamma_1) \qquad \text{for } \boldsymbol{\pi}\text{-a.e. } \gamma \in \mathrm{Geo}(X).$$

Furthermore,

$$L(x) = |\nabla\varphi|_*(x) = |\nabla^+\varphi|(x) \qquad \text{for } \mu\text{-a.e. } x \in X,$$

where φ *is any Kantorovich potential relative to* (μ, ν)*.*

10 Open problems and perspectives

(1) It would be interesting to examine the impact of the additional axiom, i.e. linearity of the heat flow, at the level of the finite-dimensional theory, i.e. $CD(K,N)$ with $N < \infty$. Or to understand, at least in the case $K = 0$, where the Reny entropy is available, the role of the EVI formulation

(2) At least formally, one can use this calculus to write down a differential $CD(K,N)$ inequality

$$\Delta\frac{|\nabla f|_*^2}{2} \geq \langle \nabla\Delta f, \nabla f\rangle + \frac{(\Delta f)^2}{N} + K|\nabla f|_*^2,$$

and try to investigate its relation with the existing theories. Also, in this differential perspective, one might try to reverse the implication from the Bakry-Emery condition (9.1) to $(R)CD(K,\infty)$.

(3) In presence of doubling & Poincaré, Cheeger's theory applies and provides, in a suitable and very weak sense, local coordinates and a tangent bundle. The relations with the calculus described in these lectures are still not completely understood.

(4) What about the behaviour on small scales of $RCD(K, \infty)$ spaces? The question makes sense, if one adds a doubling condition. The natural conjecture is that tangent metric spaces, in the measured Gromov-Hausdorff sense, are Euclidean. This has been proved by Cheeger-Colding, but for limits of Riemannian manifolds.

References

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, Clarendon Press, Oxford, 2000.

[2] L. Ambrosio and N. Gigli, *User's guide to optimal transport theory*, To appear in the CIME Lecture Notes in Mathematics, B.Piccoli and F.Poupaud Eds., (2011).

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[4] ——, *Calculus and heat flows in metric measure spaces with Ricci curvature bounded from below*, Submitted paper, arXiv:1106.2090, (2011).

[5] ——, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, Submitted paper, (2011).

[6] L. Ambrosio, N. Gigli, and G. Savaré, *Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces*, preprint, 2011.

[7] P. Bernard and B. Buffoni, *Optimal mass transportation and Mather theory*, J. Eur. Math. Soc. (JEMS), 9 (2007), pp. 85–127.

[8] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal., 9 (1999), pp. 428–517.

[9] J. Cheeger and T. Colding, *On the structure of spaces with ricci curvature bounded below i*, J. Diff. Geom., 46 (1997), pp. 37–74.

[10] S. Daneri and G. Savaré, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM J. Math. Anal., 40 (2008), pp. 1104–1122.

[11] M. Erbar, *The heat equation on manifolds as a gradient flow in the Wasserstein space*, Annales de l'Institut Henri Poincaré - Probabilités et Statistiques, 46 (2010), pp. 1–23.

[12] S. Fang, J. Shao, and K.-T. Sturm, *Wasserstein space over the Wiener space*, Probab. Theory Related Fields, 146 (2010), pp. 535–565.

[13] N. Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations, 39 (2010), pp. 101–120.

[14] N. Gigli, K. Kuwada, and S. Ohta, *Heat flow on Alexandrov spaces*, Submitted paper, (2010).

[15] J. Heinonen, *Nonsmooth calculus*, Bull. Amer. Math. Soc., 44 (2007), pp. 163–232.

[16] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck Equation*, SIAM Journal on Mathematical Analysis, 29 (1998), pp. 1–17.

[17] B. Levi, *Sul principio di Dirichlet*, Rendiconti del Circolo Matematico di Palermo, (1906), pp. 293–359.

[18] S. Lisini, *Characterization of absolutely continuous curves in Wasserstein spaces*, Calc. Var. Partial Differential Equations, 28 (2007), pp. 85–120.

[19] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2), 169 (2009), pp. 903–991.

[20] S.-I. Ohta and K.-T. Sturm, *Heat flow on Finsler manifolds*, Comm. Pure Appl. Math., 62 (2009), pp. 1386–1433.

[21] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[22] N. Shanmugalingam, *Newtonian spaces: an extension of Sobolev spaces to metric measure spaces*, Rev. Mat. Iberoamericana, 16 (2000), pp. 243–279.

[23] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math., 196 (2006), pp. 65–131.

[24] ——, *On the geometry of metric measure spaces. II*, Acta Math., 196 (2006), pp. 133–177.

[25] K.-T. Sturm and S.-i. Ohta, *Non-contraction of heat flow on Minkowski spaces*, (2010).

[26] C. Villani, *Optimal transport. Old and new*, vol. 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.