Heat flow and calculus on metric measure spaces with Ricci curvature bounded below - the compact case

Luigi Ambrosio * Nicola Gigli † Giuseppe Savaré ‡

October 4, 2012

Contents

*Scuola Normale Superiore, Pisa. email: l.ambrosio@sns.it
†Nice University. email: nicola.gigli@unice.fr
‡Università di Pavia. email: giuseppe.savare@unipv.it

1 Introduction

Aim of these notes is to provide a quick overview of the main results contained in [?] and [?] in the simplified case of compact metric spaces (X, d) endowed with a reference probability measure $\mathfrak{m}$. The idea is to give the interested reader the possibility to get as quickly as possible the key ideas behind the proofs of our recent results, neglecting all the problems that appear in a more general framework (as a matter of fact, no compactness is made in [?, ?] and finiteness of $\mathfrak{m}$ is assumed only in [?]). We want to underline that passing from compact spaces to complete and separable ones (and even to a more general framework which includes the so-called Wiener space) is not just a technical problem, meaning that several concepts need to be properly adapted in order to achieve such generality. Anyway, for the purpose of these notes, all the metric spaces (X, d) under consideration are compact, and all the reference measures $\mathfrak{m}$ are Borel probability ones. Hence, in particular, the discussion here is by no means exhaustive, as both the key statements and the auxiliary lemmas are stated in this simplified case.

Apart some very basic concept about optimal transport, Wasserstein distance and gradient flows, this paper pretends to be self-contained. All the concepts that we need are recalled in the preliminary section, whose proofs can be found, for instance, in the first three chapters of [?] (for an overview on the theory of gradient flows, see also [?], and for a much broader discussion on optimal transport, see the monograph by Villani [?]). For completeness reasons, we included in our discussion some results coming from previous contributions which are potentially less known, in particular: the (sketch of the) proof by Lisini [?] of the characterization of absolutely continuous curves w.r.t. the Wasserstein distance (Proposition 4.18), and the proof of uniqueness of the gradient flow of the relative entropy w.r.t. the Wasserstein distance on spaces with Ricci curvature bounded below in the sense of Lott-Sturm-Villani ($CD(K, \infty)$ spaces in short) given by the second author in [?] (Theorem 5.7).

In summary, the main results that we present here are the following.

(1) The proof of the fact that the Hopf-Lax formula produces subolutions of the Hamilton-Jacobi equation, and solutions on geodesic spaces (Theorem 3.5 and Theorem 3.6).

(2) A new approach to the theory of Sobolev spaces over metric measure spaces, which leads in particular to the proof that *Lipschitz functions are always dense in energy in* $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ (Theorem 4.23).

(3) The already alluded to proof of uniqueness of the gradient flow w.r.t. W_2 of the relative entropy in $CD(K, \infty)$ spaces (Theorem 5.7).

(4) The identification of the L^2-gradient flow of the natural "Dirichlet energy" and the W_2-gradient flow of the relative entropy in $CD(K, \infty)$ spaces (see also [?] for the Alexandrov case, a paper to which our paper [?] owes a lot).

(5) A metric version of Brenier's theorem valid in spaces having Ricci curvature bounded from below in a sense slightly stronger than the one proposed by Lott-Sturm-Villani. What our result says, is that if this curvature assumption holds (Definition 7.1) and μ, ν are absolutely continuous w.r.t. $\mathfrak{m}$, then "the distance traveled is uniquely determined by the starting point", i.e. there exists a map $D : X \to \mathbb{R}$ such that for any optimal plan $\boldsymbol{\gamma}$ it holds $\mathsf{d}(x, y) = D(x)$ for $\boldsymbol{\gamma}$-a.e. (x, y). Moreover, this map is nothing but

the weak gradient (according to the theory illustrated in Section 4) of any Kantorovich potential. See Theorem 7.3.

(6) A key lemma (Lemma 8.2 concerning "horizontal" and "vertical" differentiation which allows to compare the derivative of the squared Wasserstein distance from a along the heat flow with the derivative of the relative entropy along a geodesic.

(7) A new definition of Ricci bound from below for metric measure spaces which, while still being stable w.r.t. measured Gromov-Hausdorff convergence, rules out Finsler geometries (Theorem 9.1 and the discussion thereafter).

[1]

Acknowledgement. The authors acknowledge the support of the ERC ADG GeMeThNES. The authors also thank A.Mondino for his careful reading of a preliminary version of this manuscript.

2 Preliminary notions

2.1 Absolutely continuous curves and slopes

Let (X, d) be a metric space, $J \subset \mathbb{R}$ an interval with nonempty interior and $J \ni t \mapsto x_t \in X$. We say that x_t is *absolutely continuous* if

$$\mathsf{d}(x_s, x_t) \leq \int_t^s g(r)\, \mathrm{d}r, \qquad \forall s,\, t \in J,\ t < s$$

for some $g \in L^1(J)$. It turns out that, if x_t is absolutely continuous, there is a minimal function g with this property, called *metric speed* and given for a.e. $t \in J$ by

$$|\dot{x}_t| = \lim_{s \to t} \frac{\mathsf{d}(x_s, x_t)}{|s - t|}.$$

See [?, Theorem 1.1.2] for the simple proof. Notice that the absolute continuity properties of the integral ensure that absolutely continuous functions can be extended by continuity to the closure of their domain.

Given $f : X \to \mathbb{R} \cup \{\pm\infty\}$, we define *slope* (also called local Lipschitz constant) at points x where $f(x) \in \mathbb{R}$ by

$$|\nabla f|(x) := \overline{\lim_{y \to x}} \frac{|f(y) - f(x)|}{\mathsf{d}(y, x)}.$$

We shall also need the one-sided counterparts of the local Lipschitz constant, called respectively *descending slope* and *ascending slope*:

$$\begin{aligned} |\nabla^- f|(x) &:= \overline{\lim_{y \to x}} \frac{[f(y) - f(x)]^-}{\mathsf{d}(y, x)}, \\ |\nabla^+ f|(x) &:= \overline{\lim_{y \to x}} \frac{[f(y) - f(x)]^+}{\mathsf{d}(y, x)}, \end{aligned} \tag{2.1}$$

[1] Qui poi inserirei una dedica a Magenes, penso che qui Giuseppe potrebbe fare molto meglio di noi due. Altrimenti ci provo io

where $[\cdot]^+$ and $[\cdot]^-$ denote respectively the positive and negative part.

It is not difficult to see that for f Lipschitz the two slopes and the local Lipschitz constant are upper gradients according to [**?**], namely

$$|f(\gamma_1) - f(\gamma_0)| \le \int_\gamma |\nabla^\pm f|$$

for any absolutely continuous curve $\gamma : [0,1] \to X$.

Also, for $f, g : X \to \mathbb{R}$ Lipschitz it clearly holds

$$|\nabla(\alpha f + \beta g)| \le |\alpha||\nabla f| + |\beta||\nabla g|, \qquad \forall \alpha, \beta \in \mathbb{R}; \tag{2.2a}$$

$$|\nabla(fg)| \le |f||\nabla g| + |g||\nabla f|. \tag{2.2b}$$

2.2 The space $(\mathscr{P}(X), W_2)$

Let (X, d) be a compact metric space. The set $\mathscr{P}(X)$ consists of all Borel probability measures on X. Given $\mu, \nu \in \mathscr{P}(X)$, we define the Wasserstein distance $W_2(\mu, \nu)$ between them as

$$W_2^2(\mu, \nu) := \min \int \mathsf{d}^2(x, y) \, \mathrm{d}\boldsymbol{\gamma}(x, y),$$

where the infimum is taken among all Borel probability measures $\boldsymbol{\gamma}$ on X^2 such that

$$\pi^1_\# \boldsymbol{\gamma} = \mu,$$
$$\pi^2_\# \boldsymbol{\gamma} = \nu.$$

Such measures are called admissible plans for the couple (μ, ν), and the set of admissible plans will be denoted by $\textsc{Adm}(\mu, \nu)$. A plan $\boldsymbol{\gamma}$ which realizes the minimum is called optimal, and we write $\boldsymbol{\gamma} \in \textsc{Opt}(\mu, \nu)$. From the linearity of the admissibility condition we get that the squared Wasserstein distance is convex, i.e.:

$$W_2^2\big((1-\lambda)\mu_1 + \lambda\nu_1, (1-\lambda)\mu_2 + \lambda\nu_2\big) \le (1-\lambda) W_2^2(\mu_1, \nu_1) + \lambda W_2^2(\mu_2, \nu_2). \tag{2.3}$$

It is also well known that the Wasserstein distance metrizes the weak convergence of measures (see e.g. Theorem 2.7 in [**?**]).

An equivalent definition of W_2 comes from the dual formulation of the transport problem:

$$\frac{1}{2} W_2^2(\mu, \nu) = \sup_\psi \int \psi \, \mathrm{d}\mu + \int \psi^c \, \mathrm{d}\nu, \tag{2.4}$$

the supremum being taken among all Lipschitz functions ψ, where the c-transform in this formula is defined by

$$\psi^c(y) := \inf_{x \in X} \frac{\mathsf{d}^2(x, y)}{2} - \psi(x).$$

A function $\psi : X \to \mathbb{R}$ is said to be c-concave if $\psi = \phi^c$ for some $\phi : X \to \mathbb{R}$. The maximum in (2.4) is always achieved by a c-concave function, and we will call any such function ψ a Kantorovich potential. We shall also use the fact that c-concave functions satisfy

$$\psi^{cc} = \psi. \tag{2.5}$$

The (graph of the) c-superdifferential $\partial^c\psi$ of a c-concave function ψ is the subset of X^2 defined by

$$\partial^c\psi := \Big\{(x,y) \ : \ \psi(x)+\psi^c(y) = \frac{\mathsf{d}^2(x,y)}{2}\Big\},$$

and the c-superdifferential $\partial^c\psi(x)$ at x is the set of y's such that $(x,y)\in\partial^c\psi$. A consequence of the compactness of X is that any c-concave function ψ is Lipschitz and that the set $\partial^c\psi(x)$ is non empty for any $x\in X$.

The relation between optimal plans and Kantorovich potentials is the fact that a plan $\boldsymbol{\gamma}$ is optimal, relative to its marginals, if and only if $\mathrm{supp}(\boldsymbol{\gamma})\subset\partial^c\psi$ for some c-concave function ψ.

We will denote by $C([0,1],X)$ the space of continuous curves on $[0,1]$ with values in X endowed with the sup norm. The set $AC^2([0,1],X)\subset C([0,1],X)$ consists of all absolutely continuous curves γ such that $\int_0^1|\dot\gamma_t|^2\,\mathrm{d}t<\infty$: it is easily seen to be equal to the countable union of the closed sets $\{\gamma : \int_0^1|\dot\gamma_t|^2\,\mathrm{d}t\le n\}$, and thus it is a Borel subset of $C([0,1],X)$. The *evaluation maps* $\mathrm{e}_t : C([0,1],X)\to X$ are defined by

$$\mathrm{e}_t(\gamma) := \gamma_t,$$

and are clearly 1-Lipschitz.

We say that (X,d) is a geodesic space if for any x, $y\in X$ there exists a curve (γ_t) on $[0,1]$ such that $\gamma_0=x$, $\gamma_1=y$ and $\mathsf{d}(\gamma_t,\gamma_s)=|t-s|\mathsf{d}(x,y)$ for all s, $t\in[0,1]$. Such a curve is called constant speed geodesic, or simply geodesic. The space of all geodesics endowed with the sup distance will be denoted by Geo(X). If (X,d) is geodesic, then so is $(\mathscr{P}(X),W_2)$, and in this case a curve (μ_t) is a constant speed geodesic from μ_0 to μ_1 if and only if there exists a measure $\boldsymbol{\pi}\in\mathscr{P}(C([0,1],X))$ concentrated on Geo(X) such that $(\mathrm{e}_t)_\sharp\boldsymbol{\pi}=\mu_t$ for all $t\in[0,1]$ and $(\mathrm{e}_0,\mathrm{e}_1)_\sharp\in\textsc{Opt}(\mu_0,\mu_1)$. We will denote the set of such measures, called optimal geodesic plans, by $\mathrm{GeoOpt}(\mu_0,\mu_1)$.

2.3 Geodesically convex functionals and their gradient flows

Given a geodesic space (Y,d_Y) (in the following this will always be the Wasserstein space built over a geodesic space (X,d)), a functional $E:Y\to\mathbb{R}\cup\{+\infty\}$ is said $K-$geodesically convex (or simply K-convex) if for any y_0, $y_1\in Y$ there exists a constant speed geodesic $\gamma:[0,1]\to Y$ such that $\gamma_0=y_0$, $\gamma_1=y_1$ and

$$E(\gamma_t)\le(1-t)E(y_0)+tE(y_1)-\frac{K}{2}t(1-t)\mathsf{d}_Y^2(y_0,y_1),\qquad\forall t\in[0,1].$$

We will denote by $D(E)$ the domain of E i.e. $D(E):=\{y : E(y)<\infty\}$.

An easy consequence of the K-convexity is the fact that the descending slope defined in (2.1) can de computed as a sup, rather than as a limsup:

$$|\nabla^-E|(y)=\sup_{z\neq y}\left(\frac{E(y)-E(z)}{\mathsf{d}_Y(y,z)}+\frac{K}{2}\mathsf{d}_Y(y,z)\right)^+. \tag{2.6}$$

What we want to discuss here is the definition of gradient flow of a K-convex functional. There are essentially two different ways of giving such a notion in a metric setting. The first one, which we call Energy Dissipation Equality (EDE), ensures existence for any K-convex and lower semicontinuous functional (under suitable compactness assumptions), the second one, which we call Evolution Variation Inequality (EVI), ensures uniqueness and K-contrativity of the flow. However, the price we pay for these stronger properties is that existence results for EVI solutions hold under much more restrictive assumptions.

It is important to distinguish the two notions. The EDE one is the "correct one" to be used in a general metric context, because it ensures existence for any initial datum in the domain of the functional. However, typically gradient flows in the EDE sense are not unique: this is the reason of the analysis made in Section 5, which ensures that for the special case of the entropy functional uniqueness is indeed true.

EVI gradient flows are in particular gradient flows in the EDE sense (see Proposition 2.5), ensure uniqueness, K-contractivity and provide strong a priori regularizing effects. Heuristically speaking, existence of gradient flows in the EVI sense depends also on properties of the distance, rather than on properties of the functional only. A more or less correct way of thinking at this is: gradient flows in the EVI sense exist if and only if the distance is Hilbertian on small scales. For instance, if the underlying metric space is an Hilbert space, then the two notions coincide.

Now recall that one of our goals here is to study the gradient flow of the relative entropy i n spaces with Ricci curvature bounded below (Definition 5.1), and recall that Finsler geometries are included in this setting (see page 926 of [**?**]). Thus, in general we must deal with the EDE notion of gradient flow. The EVI one will come into play in Section 9, where we use it to identify those spaces with Ricci curvature bounded below which are more 'Riemannian like'. **Note**: later on we will refer to gradient flows in the EDE sense simply as "gradient flows", keeping the distinguished notation EVI-gradient flows for those in the EVI sense.

2.3.1 Energy Dissipation Equality

An important property of K-geodesically convex and lower semicontinuous functionals (see Corollary 2.4.10 of [**?**]) is that the descending slope is an upper gradient, that is: for any absolutely continuous curve $y_t : J \subset \mathbb{R} \to D(E)$ it holds

$$|E(y_t) - E(y_s)| \le \int_t^s |\dot{y}_r| |\nabla^- E|(y_r) \,\mathrm{d}r, \qquad \forall t \le s. \tag{2.7}$$

An application of Young inequality gives that

$$E(y_t) \le E(y_s) + \frac{1}{2}\int_t^s |\dot{y}_r|^2 \,\mathrm{d}r + \frac{1}{2}\int_t^s |\nabla^- E|^2(y_r) \,\mathrm{d}r, \qquad \forall t \le s. \tag{2.8}$$

This inequality motivates the following definition:

Definition 2.1 (Energy Dissipation Equality definition of gradient flow) *Let E be a K-convex and lower semicontinuous functional and let $y_0 \in D(E)$. We say that a continuous curve $[0,\infty) \ni t \mapsto y_t$ is a gradient flow for the E in the EDE sense (or simply a gradient flow) if it is locally absolutely continuous in $(0,\infty)$, it takes values in the domain of E and it holds*

$$E(y_t) = E(y_s) + \frac{1}{2}\int_t^s |\dot{y}_r|^2 \,\mathrm{d}r + \frac{1}{2}\int_t^s |\nabla^- E|^2(y_r) \,\mathrm{d}r, \qquad \forall t \le s. \tag{2.9}$$

Notice that due to (2.8) the equality (2.9) is equivalent to

$$E(y_0) \ge E(y_s) + \frac{1}{2}\int_0^s |\dot{y}_r|^2 \,\mathrm{d}r + \frac{1}{2}\int_0^s |\nabla^- E|^2(y_r) \,\mathrm{d}r, \qquad \forall s > 0. \tag{2.10}$$

Indeed, if (2.10) holds, then (2.9) holds with $t = 0$, and then by linearity (2.9) holds in general.

It is not hard to check that if $E : \mathbb{R}^d \to \mathbb{R}$ is a smooth functional, then a curve $y_t : J \to \mathbb{R}^d$ is a gradient flow according to the previous definition if and only if it satisfies

$$y_t' = -\nabla E(y_t), \qquad \forall t \in J,$$

so that the metric definition reduces to the classical one when specialized to Euclidean spaces.

The following theorem has been proved in [**?**] (Corollary 2.4.11):

Theorem 2.2 (Existence of gradient flows in the EDE sense) *Let (Y, d_Y) be a compact geodesic space and let $E : Y \to \mathbb{R} \cup \{+\infty\}$ be a K-convex and lower semicontinuous functional. Then every $y_0 \in D(E)$ is the starting point of a gradient flow in the EDE sense of E.*

It is important to stress the fact that in general gradient flows in the EDE sense are *not* unique. A simple example is $Y := \mathbb{R}^2$ endowed with the L^∞ norm, and E defined by $E(x, y) := x$. It is immediate to see that E is 0-convex and that for any point (x_0, y_0) there exist uncountably many gradient flows in the EDE starting from it, for instance all curves $(x_0 - t, y(t))$ with $|y'(t)| \le 1$ and $y(0) = y_0$.

2.3.2 Evolution Variational Inequality

To see where the EVI notion comes from, notice that for a K-convex and smooth function f on $\mathbb{R}^d$ it holds $x_t' = -\nabla f(x)$ for any $t \ge 0$ if and only if

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{|x_t - y|^2}{2} + \frac{K}{2}|x_t - y|^2 + f(x_t) \le f(y), \qquad \forall y \in \mathbb{R}^d,\ \forall t \ge 0. \tag{2.11}$$

This equivalence is true because K-convexity ensures that $v = -\nabla f(x)$ if and only

$$\langle v, x - y\rangle + \frac{K}{2}|x - y|^2 + f(x) \le f(y), \qquad \forall y \in \mathbb{R}^d.$$

Inequality (2.11) can be written in a metric context in several ways, which we collect in the following statement (we omit the easy proof).

Proposition 2.3 (Evolution Variational Inequality: equivalent statements) *Let (Y, d_Y) be a compact geodesic space*[2]*, $E : Y \to [0, \infty]$ a lower semicontinuous functional. Then the following properties are equivalent.*

(i) For any $z \in E$ it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathsf{d}_Y^2(y_t, z)}{2} + \frac{K}{2}\mathsf{d}_Y^2(y_t, z) + E(y_t) \le E(z), \qquad \text{for a.e. } t \in (0, \infty).$$

(ii) For any $z \in E$ it holds

$$\frac{\mathsf{d}_Y^2(y_s, z) - \mathsf{d}_Y^2(y_t, z)}{2} + \frac{K}{2}\int_t^s \mathsf{d}_Y^2(y_r, z)\,\mathrm{d}r + \int_t^s E(y_r)\,\mathrm{d}r \le (s-t)E(z), \qquad \forall 0 < t < s < \infty.$$

[2]Qui metterei Y completo e separabile, queste ipotesi mi sembrano misleading, in fondo e' Y e non X e questo non contraddice le nostre standing assumptions su X

(iii) *There exists a set $A \subset D(E)$ dense in energy (i.e., for any $z \in D(E)$ there exists $(z_n) \subset A$ converging to z such that $E(z_n) \to E(z)$) such that for any $z \in A$ it holds*

$$\overline{\lim_{h\downarrow 0}} \frac{\mathsf{d}_Y^2(y_{t+h}, z) - \mathsf{d}_Y^2(y_t, z)}{2} + \frac{K}{2}\mathsf{d}_Y^2(y_t, z) + E(y_t) \leq E(z), \qquad \forall t \in (0, \infty).$$

Definition 2.4 (Evolution Variational Inequality definition of gradient flow) *We say that a curve (y_t) is a gradient flow of E in the EVI sense relative to $K \in \mathbb{R}$ (in short,* EVI_K*-gradient flow), if any of the above equivalent properties are true. We say that y_t starts from y_0 if $y_t \to y_0$ as $t \downarrow 0$.*

This definition of gradient flow is stronger than the one discussed in the previous section, because of the following result proved by Savaré in [**?**] (see also Proposition 3.6 of [**?**]), which we state without proof.

Proposition 2.5 (EVI implies EDE) *Let (Y, d_Y) be a compact*[3] *space, $K \in \mathbb{R}$, $E : Y \to [0, \infty]$ a lower semicontinuous functional and $y_t : (0, \infty) \to D(E)$ a locally absolutely continuous curve. Assume that y_t is an* EVI_K*-gradient flow for E. Then* (2.9) *holds for any $0 < t < s$.*

Remark 2.6 (Contractivity) It can be proved that if (y_t) and (z_t) are gradient flows in the EVI_K sense of the l.s.c. functional E, then

$$\mathsf{d}_Y(y_t, z_t) \leq e^{-Kt}\mathsf{d}_Y(y_0, z_0), \qquad \forall t \geq 0.$$

In particular, gradient flows in the EVI sense are unique. This contractivity property, used in conjunction with *(ii)* of Proposition 2.3, guarantees that if existence of gradient flows in the EVI sense is known for initial data lying in some subset $S \subset Y$, then it is also known for initial data in the closure $\overline{S}$ of S. ■

We also point out the following geometric consequence of the EVI, proven in [**?**].

Proposition 2.7 *Let $E : Y \to [0, \infty]$ be a lower semicontinuous functional on a complete and geodesic space (Y, d_Y). Assume that every $y_0 \in Y$ is the starting point of an* EVI_K*-gradient flow of E. Then E is K-convex along* all *geodesics.*

As we already said, gradient flows in the EVI sense do not necessarily exist, and their existence depends on the properties of the distance d_Y. For instance, it is not hard to see that if we endow $\mathbb{R}^2$ with the L^∞ norm and consider the functional $E(x, y) := x$, then there re is no gradient flow in the EVI_K-sense, regardless of the constant K.

3 Hopf-Lax formula and Hamilton-Jacobi equation

Aim of this subsection is to study the properties of the Hopf-Lax formula in a metric setting and its relations with the Hamilton-Jacobi equation. Here we assume that (X, d) is a compact metric space. Notice that there is no reference measure $\mathfrak{m}$ in the discussion here.

[3] come prima, direi completo e separabile

Let $f : X \to \mathbb{R}$ be a Lipschitz function. For $t > 0$ define

$$F(t, x, y) := f(y) + \frac{\mathsf{d}^2(x, y)}{2t},$$

and the function $Q_t f : X \to \mathbb{R}$ by

$$Q_t f(x) := \inf_{y \in X} F(t, x, y) = \min_{y \in X} F(t, x, y).$$

Also, we introduce the functions D^+, $D^- : X \times (0, \infty) \to \mathbb{R}$ as

$$\begin{aligned} D^+(x, t) &:= \max \mathsf{d}(x, y), \\ D^-(x, t) &:= \min \mathsf{d}(x, y), \end{aligned} \tag{3.1}$$

where, in both cases, the y's vary among all minima of $F(t, x, \cdot)$. We also set $Q_0 f = f$ and $D^{\pm}(x, 0) = 0$. Arguing as in Lemma 3.1.2 of [**?**] it is easy to check that the map $[0, \infty) \times X \ni (t, x) \mapsto Q_t f(x)$ is continuous. Furthermore, the fact that f is Lipschitz easily yields

$$D^-(x, t) \le D^+(x, t) \le 2t \, \mathrm{Lip}(f), \tag{3.2}$$

and from the fact that the functions $\{\mathsf{d}^2(\cdot, y)\}_{y \in Y}$ are uniformly Lipschitz (because (X, d) is bounded) we get that $Q_t f$ is Lipschitz for any $t > 0$.

Proposition 3.1 (Monotonicity of $D^{\pm}$) *For all $x \in X$ it holds*

$$D^+(x, t) \le D^-(x, s), \qquad 0 \le t < s. \tag{3.3}$$

As a consequence, $D^+(x, \cdot)$ and $D^-(x, \cdot)$ are both nondecreasing, and they coincide with at most countably many exceptions in $[0, \infty)$.

Proof Fix $x \in X$. For $t = 0$ there is nothing to prove. Now pick $0 < t < s$ and choose x_t and x_s minimizers of $F(t, x, \cdot)$ and $F(s, x, \cdot)$ respectively, such that $\mathsf{d}(x, x_t) = D^+(x, t)$ and $\mathsf{d}(x, x_s) = D^-(x, s)$. The minimality of x_t, x_s gives

$$\begin{aligned} f(x_t) + \frac{\mathsf{d}^2(x_t, x)}{2t} &\le f(x_s) + \frac{\mathsf{d}^2(x_s, x)}{2t} \\ f(x_s) + \frac{\mathsf{d}^2(x_s, x)}{2s} &\le f(x_t) + \frac{\mathsf{d}^2(x_t, x)}{2s}. \end{aligned}$$

Adding up and using the fact that $\frac{1}{t} \ge \frac{1}{s}$ we deduce

$$D^+(x, t) = \mathsf{d}(x_t, x) \le \mathsf{d}(x_s, x) = D^-(x, s),$$

which is (3.3).

Combining this with the inequality $D^- \le D^+$ we immediately obtain that both functions are nonincreasing. At a point of right continuity of $D^-(x, \cdot)$ we get

$$D^+(x, t) \le \inf_{s > t} D^-(x, s) = D^-(x, t).$$

This implies that the two functions coincide out of a countable set. □

Next, we examine the semicontinuity properties of $D^\pm$. These properties imply that points (x,t) where the equality $D^+(x,t) = D^-(x,t)$ occurs are continuity points for both D^+ and D^-.

Proposition 3.2 (Semicontinuity of $D^\pm$) *The map D^+ is upper semicontinuous and the map D^- is lower semicontinuous in $X \times (0,\infty)$.*

Proof We prove lower semicontinuity of D^-, the proof of upper semicontinuity of D^+ being similar. Let (x_i,t_i) be any sequence converging to (x,t) and, for every i, let (y_i) be a minimum of $F(t_i,x_i,\cdot)$ for which $\mathsf{d}(y_i,x_i) = D^-(x_i,t_i)$. For all i we have

$$f(y_i) + \frac{\mathsf{d}^2(y_i,x_i)}{2t_i} = Q_{t_i}f(x_i),$$

Moreover, the continuity of $(x,t) \mapsto Q_tf(x)$ gives that $\lim_i Q_{t_i}f(x_i) = Q_tf(x)$, thus

$$\lim_{i\to\infty} f(y_i) + \frac{\mathsf{d}^2(y_i,x)}{2t} = Q_tf(x).$$

This means that (y_i) is a minimizing sequence for $F(t,x,\cdot)$. Since (X,d) is compact, possibly passing to a subsequence, not relabeled, we may assume that (y_i) converges to y as $i \to \infty$. Therefore

$$D^-(x,t) \le \mathsf{d}(x,y) = \lim_{i\to\infty} \mathsf{d}(x,y_i) = \lim_{i\to\infty} D^-(x_i,t_i).$$

□

Proposition 3.3 (Time derivative of Q_tf) *The map $t \mapsto Q_tf$ is Lipschitz from $[0,\infty)$ to $C(X)$ is Lipschitz and, for all $x \in X$, it satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_tf(x) = -\frac{[D^\pm(x,t)]^2}{2t^2}, \tag{3.4}$$

for any $t > 0$ with at most countably many exceptions.

Proof Let $t < s$ and x_t, x_s be minima of $F(t,x,\cdot)$ and $F(s,x,\cdot)$. We have

$$Q_sf(x) - Q_tf(x) \le F(s,x,x_t) - F(t,x,x_t) = \frac{\mathsf{d}^2(x,x_t)}{2}\frac{t-s}{ts},$$
$$Q_sf(x) - Q_tf(x) \ge F(s,x,x_s) - F(t,x,x_s) = \frac{\mathsf{d}^2(x,x_s)}{2}\frac{t-s}{ts},$$

which gives that $t \mapsto Q_tf(x)$ is Lipschitz in $(\varepsilon,+\infty)$ for any $\varepsilon > 0$ and $x \in X$. Also, dividing by $(s-t)$ and taking Proposition 3.1 into account, we get (3.4). Now notice that from (3.2) we get that $|\frac{\mathrm{d}}{\mathrm{d}t}Q_tf(x)| \le 2\,\mathrm{Lip}^2(f)$ for any x and a.e. t, which, together with the pointwise convergence of Q_tf to f as $t \downarrow 0$, yields that $t \mapsto Q_tf \in C(X)$ is Lipschitz in $[0,\infty)$. □

Proposition 3.4 (Bound on the local Lipschitz constant of Q_tf) *For $(x,t) \in X \times (0,\infty)$ it holds:*

$$|\nabla Q_tf|(x) \le \frac{D^+(x,t)}{t}. \tag{3.5}$$

Proof Fix $x \in X$ and $t \in (0, \infty)$, pick a sequence (x_i) converging to x and a corresponding sequence (y_i) of minimizers for $F(t, x_i, \cdot)$ and similarly a minimizer y of $F(t, x, \cdot)$. We start proving that

$$\varlimsup_{i \to \infty} \frac{Q_t f(x) - Q_t f(x_i)}{d(x, x_i)} \leq \frac{D^+(x,t)}{t}.$$

Since it holds

$$\begin{aligned} Q_t f(x) - Q_t f(x_i) \leq F(t, x, y_i) - F(t, x_i, y_i) \leq f(y_i) + \frac{\mathsf{d}^2(x, y_i)}{2t} - f(y_i) - \frac{\mathsf{d}^2(x_i, y_i)}{2t} \\ \leq \frac{\mathsf{d}(x, x_i)}{2t}\big(\mathsf{d}(x, y_i) + \mathsf{d}(x_i, y_i)\big) \leq \frac{\mathsf{d}(x, x_i)}{2t}\big(\mathsf{d}(x, x_i) + 2D^+(x_i, t)\big), \end{aligned}$$

dividing by $d(x, x_i)$, letting $i \to \infty$ and using the upper semicontinuity of D^+ we get the claim. To conclude, we need to show that

$$\varlimsup_{i \to \infty} \frac{Q_t f(x_i) - Q_t f(x)}{d(x, x_i)} \leq \frac{D^+(x,t)}{t}.$$

This follows along similar lines starting from the inequality

$$Q_t f(x_i) - Q_t f(x) \leq F(t, x_i, y) - F(t, x, y_i).$$

□

Theorem 3.5 (Subsolution of HJ) *For every $x \in X$ it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_t f(x) + \frac{1}{2} |\nabla Q_t f|^2(x) \leq 0 \tag{3.6}$$

with at most countably many exceptions in $(0, \infty)$.

Proof The claim is a direct consequence of Proposition 3.3 and Proposition 3.4. □

We just proved that in an arbitrary metric space the Hopf-Lax formula produces subsolutions of the Hamilton-Jacobi equation. Our aim now is to prove that if (X, d) is a geodesic space, then the same formula provides also supersolutions.

Theorem 3.6 (Supersolution of HJ) *Assume that (X, d) is a geodesic space. Then equality holds in* (3.5)*. In particular, for all $x \in X$ it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_t f(x) + \frac{1}{2} |\nabla Q_t f|^2(x) = 0,$$

with at most countably many exceptions in $(0, \infty)$.

Proof Let y be a minimum of $F(t, x, \cdot)$ such that $\mathsf{d}(x, y) = D^+(x, t)$. Let $\gamma : [0, 1] \to X$ be a constant speed geodesic connecting x to y. We have

$$\begin{aligned} Q_t f(x) - Q_t f(\gamma_s) \geq f(y) + \frac{\mathsf{d}^2(x, y)}{2t} - f(y) - \frac{\mathsf{d}^2(\gamma_s, y_i)}{2t} \\ = \frac{\mathsf{d}^2(x, y) - \mathsf{d}^2(\gamma_s, y)}{2t} = \frac{\big(D^+(x,t)\big)^2 (2s - s^2)}{2t}. \end{aligned}$$

Therefore we obtain

$$\overline{\lim_{s\downarrow 0}} \frac{Q_t f(x) - Q_t f(\gamma_s)}{\mathsf{d}(x,\gamma_s)} = \overline{\lim_{s\downarrow 0}} \frac{Q_t f(x) - Q_t f(\gamma_s)}{sD^+(x,t)} \geq \frac{D^+(x,t)}{t}$$

Since $s \mapsto \gamma_s$ is a particular family converging to x we deduce

$$|\nabla^- Q_t f|(x) \geq \frac{D^+(x,t)}{t}.$$

Taking into account Proposition 3.3 and Proposition 3.4 we conclude. □

4 Weak definitions of gradient

In this section we introduce two weak notions of gradient, one inspired by Cheeger's seminal paper [?], that we call minimal relaxed slope and one inspired by Shanmugalingam's paper [?], that we call minimal weak upper gradient. We compare our concepts with those of the original papers in Subsection 4.4.

4.1 The vertical approach: minimal relaxed slope

Definition 4.1 (Relaxed slopes) *We say that $G \in L^2(X,\mathfrak{m})$ is a relaxed slope of $f \in L^2(X,\mathfrak{m})$ if there exist $\tilde G$ and Lipschitz functions $f_n \in L^2(X,\mathfrak{m})$ such that:*

(a) $f_n \to f$ in $L^2(X,\mathfrak{m})$ and $|\nabla f_n|$ weakly converges to $\tilde G$ in $L^2(X,\mathfrak{m})$;

(b) $\tilde G \leq G$ $\mathfrak{m}$-a.e. in X.

We say that G is the minimal relaxed slope of f if its $L^2(X,\mathfrak{m})$ norm is minimal among relaxed slopes. We shall denote by $|\nabla f|_*$ the minimal relaxed slope.

Using Mazur's lemma and (2.2a) it is possible to show that an equivalent characterization of relaxed slopes can be given by modifying (a) as follows: $\tilde G$ is the *strong* limit in $L^2(X,\mathfrak{m})$ of $G_n \geq |\nabla f_n|$. The definition of relaxed slope we gave is useful to show existence of relaxed slopes (as soon as an approximating sequence (f_n) with $|\nabla f_n|$ bounded in $L^2(X,\mathfrak{m})$ exists) while the equivalent characterization is useful to perform diagonal arguments and to show that the class of relaxed slopes is a convex closed set. Therefore the definition of $|\nabla f|_*$ is well posed.

Lemma 4.2 (Locality) *Let G_1, G_2 be relaxed slopes of f. Then $\min\{G_1,G_2\}$ is a relaxed slope as well. In particular, for any relaxed slope G it holds*

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

Proof It is sufficient to prove that if $B \in \mathscr{B}(X)$, then $\chi_B G_1 + \chi_{X\setminus B} G_2$ is a relaxed slope of f. By approximation, taking into account the closure of the class of relaxed slopes, we can assume with no loss of generality that B is an open set. We fix $r > 0$ and a Lipschitz function $\phi_r : X \to [0,1]$ equal to 0 on $X \setminus B_r$ and equal to 1 on B_{2r}, where the open sets $B_s \subset B$ are defined by

$$B_s := \{x \in X : \ \mathrm{dist}(x, X \setminus B) > s\} \subset B.$$

Let now $f_{n,i}$, $i = 1, 2$, be Lipschitz and L^2 functions converging to f in $L^2(X, \mathfrak{m})$ as $n \to \infty$, with $|\nabla f_{n,i}|$ weakly convergent to G_i and set $f_n := \phi_r f_{n,1} + (1-\phi_r) f_{n,2}$. Then, $|\nabla f_n| = |\nabla f_{n,1}|$ on B_{2r} and $|\nabla f_n| = |\nabla f_{n,2}|$ on $X \setminus \overline{B_r}$; in $\overline{B_r} \setminus B_{2r}$, by applying (2.2a) and (2.2b), we can estimate

$$|\nabla f_n| \le |\nabla f_{n,2}| + \mathrm{Lip}(\phi_r)|f_{n,1} - f_{n,2}| + \phi_r\big(|\nabla f_{n,1}| + |\nabla f_{n,2}|\big).$$

Since $\overline{B_r} \subset B$, by taking weak limits of a subsequence, it follows that

$$\chi_{B_{2r}} G_1 + \chi_{X\setminus\overline{B_r}} G_2 + \chi_{B\setminus B_{2r}}(G_1 + 2G_2)$$

is a relaxed slope of f. Letting $r \downarrow 0$ gives that $\chi_B G_1 + \chi_{X\setminus B} G_2$ is a relaxed slope as well.

For the second part of the statement argue by contradiction: let G be a relaxed slope of f and assume that $B = \{G < |\nabla f|_*\}$ is such that $\mathfrak{m}(B) > 0$. Consider the relaxed slope $G\chi_B + |\nabla f|_* \chi_{X\setminus B}$: its L^2 norm is strictly less than the L^2 norm of $|\nabla f|_*$, which is a contradiction. □

A trivial consequence of the definition and of the locality principle we just proved is that if $f : X \to \mathbb{R}$ is Lipschitz it holds:

$$|\nabla f|_* \le |\nabla f|, \qquad \mathfrak{m}\text{-a.e. in } X. \tag{4.1}$$

We also remark that it is possible to obtain the minimal relaxed slope as strong limit in L^2 of slopes of Lipschitz functions, and not only weak, as shown in the next proposition.

Proposition 4.3 (Strong approximation) *If $f \in L^2(X, \mathfrak{m})$ has a relaxed slope, there exist Lipschitz functions f_n convergent to f in $L^2(X, \mathfrak{m})$ with $|\nabla f_n|$ convergent to $|\nabla f|_*$ in $L^2(X, \mathfrak{m})$.*

Proof If $g_i \to f$ in L^2 and $|\nabla g_i|$ weakly converges to $|\nabla f|_*$ in L^2, by Mazur's lemma we can find a sequence of convex combinations of $|\nabla g_i|$ strongly convergent to $|\nabla f|_*$ in L^2; the corresponding convex combinations of g_i, that we shall denote by f_n, still converge in L^2 to f and $|\nabla f_n|$ is dominated by the convex combinations of $|\nabla g_i|$. It follows that

$$\varlimsup_{n\to\infty} \int_X |\nabla f_n|^2 \, d\mathfrak{m} \le \varlimsup_{i\to\infty} \int_X |\nabla g_i|^2 \, d\mathfrak{m} = \int_X |\nabla f|_*^2 \, d\mathfrak{m}.$$

This implies at once that $|\nabla f_n|$ weakly converges to $|\nabla f|_*$ (because any limit point in the weak topology is a relaxed with minimal norm) and that the convergence is strong. □

Theorem 4.4 *The Cheeger energy functional*

$$\mathrm{Ch}(f) := \frac{1}{2} \int_X |\nabla f|_*^2 \, d\mathfrak{m}, \tag{4.2}$$

set to $+\infty$ if f has no relaxed slope, is convex and lower semicontinuous in $L^2(X, \mathfrak{m})$.

Proof A simple byproduct of condition (2.2a) is that $\alpha F + \beta G$ is a relaxed slope of $\alpha f + \beta g$ whenever α, β are nonnegative constants and F, G are relaxed slopes of f, g respectively. Taking $F = |\nabla f|_*$ and $G = |\nabla g|_*$ yields the convexity of Ch, while lower semicontinuity follows by a simple diagonal argument based on the strong approximation property stated in Proposition 4.3. □

Proposition 4.5 (Chain rule) *If $f \in L^2(X, \mathfrak{m})$ has a relaxed slope and $\phi : X \to \mathbb{R}$ is Lipschitz and C^1, then $|\nabla\phi(f)|_* \le |\phi'(f)||\nabla f|_*$ $\mathfrak{m}$-a.e. in X. Equality holds if ϕ is nondecreasing.*

Proof We trivially have $|\nabla\phi(f)| \le \phi'(f)|\nabla f|$. If we apply this inequality to the "optimal" approximating sequence of Lipschitz functions given by Proposition 4.3 we get that that $\phi'(f)|\nabla f|_*$ is a relaxed slope of $\phi(f)$, so that $|\nabla\phi(f)|_* \le |\phi'(f)||\nabla f|_*$ $\mathfrak{m}$-a.e. in X. In the case when $\phi' \ge 0$, if we assume (with no loss of generality) $0 \le \phi' \le 1/2$ and apply the inequality to ϕ and $\psi = 1 - \phi$ we get

$$|\nabla f|_* \le |\nabla\phi(f)|_* + |\nabla\psi(f)|_* \le \phi'(f)|\nabla f|_* + \psi'(f)|\nabla f|_* = |\nabla f|_*$$

$\mathfrak{m}$-a.e. in X. Thus, all inequalities are equalities $\mathfrak{m}$-a.e. in X. □

Still by approximation, it is not difficult to show $\phi(f)$ has a relaxed slope if ϕ is nonincreasing and Lipschitz, and that $|\nabla\phi(f)|_* = \phi'(f)|\nabla f|_*$ $\mathfrak{m}$-a.e. in X. In this case $\phi'(f)$ is undefined at points x such that ϕ is not differentiable at $f(x)$, on the other hand the formula still makes sense because $|\nabla f|_* = 0$ $\mathfrak{m}$-a.e. on $f^{-1}(N)$ for any Lebesgue negligible set $N \subset \mathbb{R}$. Particularly useful is the case when ϕ is a truncation function, for instance $\phi(z) = \min\{z, M\}$. In this case

$$|\nabla \min\{f, M\}|_* = \begin{cases} |\nabla f|_* & \text{if } f(x) < M \\ 0 & \text{if } f(x) \ge M. \end{cases}$$

Analogous formulas hold for truncations from below.

4.1.1 Laplacian: definition and basic properties

Since the domain of Ch is dense in $L^2(X, \mathfrak{m})$ (it includes Lipschitz functions), the Hilbertian theory of gradient flows (see for instance [?], [?]) can be applied to Cheeger's functional (4.2) to provide, for all $f_0 \in L^2(X, \mathfrak{m})$, a locally Lipschitz continuous map $t \mapsto f_t$ from $(0, \infty)$ to $L^2(X, \mathfrak{m})$, with $f_t \to f_0$ as $t \downarrow 0$, whose derivative satisfies

$$\frac{d}{dt} f_t \in -\partial \mathrm{Ch}(f_t) \qquad \text{for a.e. } t. \tag{4.3}$$

Here $\partial\mathrm{Ch}(g)$ denotes the subdifferential of Ch at $g \in D(\mathrm{Ch})$ in the sense of convex analysis, i.e.

$$\partial\mathrm{Ch}(g) := \left\{ \xi \in L^2(X, \mathfrak{m}) : \ \mathrm{Ch}(f) \ge \mathrm{Ch}(g) + \int_X \xi (f - g) \, d\mathfrak{m} \ \ \forall f \in L^2(X, \mathfrak{m}) \right\}.$$

Another important regularizing effect of gradient flows of convex l.s.c. functionals lies in the fact that, for a.e. t, $-\frac{d}{dt} f_t$ is actually the element with minimal $L^2(X, \mathfrak{m})$ norm in $\partial^-\mathrm{Ch}(f_t)$. This motivates the next definition:

Definition 4.6 (Laplacian) *The Laplacian Δf of $f \in L^2(X, \mathfrak{m})$ is defined for those f such that $\partial\mathrm{Ch}(f) \ne \emptyset$. For those f, $-\Delta f$ is the element of minimal $L^2(X, \mathfrak{m})$ norm in $\partial\mathrm{Ch}(f)$. The domain of Δ is defined as $D(\Delta)$.*

It is a general property of convex lower semicontinuous functionals that $|\nabla^- \mathrm{Ch}|(f)$ is finite if and only if $\partial\mathrm{Ch}(f)$ is not empty, and coincides with

$$\min \left\{ \|\xi\|_{L^2(X,\mathfrak{m})} : \ \xi \in \partial\mathrm{Ch}(f) \right\}.$$

It follows that the Laplacian is defined at all functions f such that $|\nabla^- \mathrm{Ch}|(f)$ is finite and its domain coincides with the domain of $|\nabla^- \mathrm{Ch}|$.

Remark 4.7 (Potential lack of linearity) It should be observed that in general the Laplacian - as we just defined it - is *not* a linear operator: the potential lack of linearity is strictly related to the fact that potentially the space $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ is not Hilbert, because $f \mapsto \int |\nabla f|_*^2 \, \mathrm{d}\mathfrak{m}$ need not be quadratic. For instance if $X = \mathbb{R}^2$, $\mathfrak{m}$ is the Lebesgue measure and d is the distance induced by the L^∞ norm, then it is easily seen that

$$|\nabla f|_*^2 = \left(\left| \frac{\partial f}{\partial x} \right| + \left| \frac{\partial f}{\partial y} \right| \right)^2.$$

Even though the Laplacian is not linear, the trivial implication

$$v \in \partial^- \mathrm{Ch}(f) \qquad \Rightarrow \qquad \lambda v \in \partial^- \mathrm{Ch}(\lambda f), \quad \forall \lambda \in \mathbb{R},$$

ensures that the Laplacian (and so the gradient flow of Ch) is 1-homogenous. ■

We can now write

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t = \Delta f_t$$

for gradient flows f_t of Ch, the derivative being understood in $L^2(X, \mathfrak{m})$, in accordance with the classical case. The classical Hilbertian theory of gradient flows also ensures that

$$\lim_{t \to \infty} \mathrm{Ch}(f_t) = 0 \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ch}(f_t) = -\|\Delta f_t\|_{L^2(X,\mathfrak{m})}^2, \quad \text{for a.e. } t \in (0, \infty). \tag{4.4}$$

Proposition 4.8 (Integration by parts) *For all $f \in D(\Delta)$, $g \in D(\mathrm{Ch})$ it holds*

$$\left| \int_X g \Delta f \, \mathrm{d}\mathfrak{m} \right| \leq \int_X |\nabla g|_* |\nabla f|_* \, \mathrm{d}\mathfrak{m}. \tag{4.5}$$

Also, let $f \in D(\Delta)$ and $\phi \in C^1(\mathbb{R})$ with bounded derivative on an interval containing the image of f. Then

$$\int_X \phi(f) \Delta f \, \mathrm{d}\mathfrak{m} = -\int_X |\nabla f|_*^2 \phi'(f) \, \mathrm{d}\mathfrak{m}. \tag{4.6}$$

Proof Since $-\Delta f \in \partial^- \mathrm{Ch}(f)$ it holds

$$\mathrm{Ch}(f) - \int_X \varepsilon g \Delta f \, \mathrm{d}\mathfrak{m} \leq \mathrm{Ch}(f + \varepsilon g), \qquad \forall g \in L^2(X, \boldsymbol{\mu}), \ \varepsilon \in \mathbb{R}.$$

For $\varepsilon > 0$, $|\nabla f|_* + \varepsilon |\nabla g|_*$ is a relaxed slope of $f + \varepsilon g$ (possibly not minimal). Thus it holds $2\mathrm{Ch}(f + \varepsilon g) \leq \int_X (|\nabla f|_* + \varepsilon |\nabla g|_*)^2 \, \mathrm{d}\mathfrak{m}$ and therefore

$$-\int_X \varepsilon g \Delta f \, \mathrm{d}\mathfrak{m} \leq \frac{1}{2} \int_X (|\nabla f|_* + \varepsilon |\nabla g|_*)^2 - |\nabla f|_*^2 \, \mathrm{d}\mathfrak{m} = \varepsilon \int_X |\nabla f|_* |\nabla g|_* \, \mathrm{d}\mathfrak{m} + o(\varepsilon).$$

Dividing by ε, letting $\varepsilon \downarrow 0$ and then repeating the argument with $-g$ in place of g we get (4.5).

For the second part we recall that, by the chain rule, $|\nabla (f + \varepsilon \phi(f))|_* = (1 + \varepsilon \phi'(f)) |\nabla f|_*$ for $|\varepsilon|$ small enough. Hence

$$\mathrm{Ch}(f + \varepsilon \phi(f)) - \mathrm{Ch}(f) = \frac{1}{2} \int_X |\nabla f|_*^2 \big((1 + \varepsilon \phi'(f))^2 - 1 \big) \, \mathrm{d}\mathfrak{m} = \varepsilon \int_X |\nabla f|_*^2 \phi'(f) \, \mathrm{d}\mathfrak{m} + o(\varepsilon),$$

which implies that for any $v \in \partial^- \mathrm{Ch}(f)$ it holds $\int_X v \phi(f) \, \mathrm{d}\mathfrak{m} = \int_X |\nabla f|_*^2 \phi'(f) \, \mathrm{d}\mathfrak{m}$, and gives the thesis with $v = -\Delta f$. □

Proposition 4.9 (Some properties of the gradient flow of Ch) *Let $f_0 \in L^2(X, \mathfrak{m})$ and let (f_t) be the gradient flow of* Ch *starting from f_0. Then the following properties hold.*
Mass preservation. $\int f_t \,\mathrm{d}\mathfrak{m} = \int f_0 \,\mathrm{d}\mathfrak{m}$ *for any $t \geq 0$.*
Maximum principle. If $f_0 \leq C$ (resp. $f_0 \geq c$) $\mathfrak{m}$-a.e. in X, then $f_t \leq C$ (resp $f_t \geq c$) $\mathfrak{m}$-a.e. in X for any $t \geq 0$.
Entropy dissipation. Suppose $0 < c \leq f_0 \leq C < \infty$ $\mathfrak{m}$-a.e.. Then $t \mapsto \int f_t \log f_t \,\mathrm{d}\mathfrak{m}$ is locally absolutely continuous and it holds

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

Proof Mass preservation. Just notice that from (4.5) we get

$$\left|\frac{\mathrm{d}}{\mathrm{d}t} \int f_t \,\mathrm{d}\mathfrak{m}\right| = \left|\int \mathbf{1} \cdot \Delta f_t \,\mathrm{d}\mathfrak{m}\right| \leq \int |\nabla \mathbf{1}|_* |\nabla f_t|_* \,\mathrm{d}\mathfrak{m} = 0, \qquad \text{for a.e. } t \in (0, \infty),$$

where $\mathbf{1}$ is the function identically equal to 1, which has minimal relaxed gradient equal to 0.
Maximum principle. Fix $f \in L^2(X, \mathfrak{m})$, $\tau > 0$ and, according to the implicit Euler scheme, let f^τ be the unique minimizer of

$$g \qquad \mapsto \qquad \mathrm{Ch}(g) + \frac{1}{2\tau} \int_X |g - f|^2 \,\mathrm{d}\mathfrak{m}.$$

Assume that $f \leq C$. We claim that in this case $f^\tau \leq C$ as well. Indeed, if this is not the case we can consider the competitor $g := \min\{f^\tau, C\}$ in the above minimization problem. By (a) of Proposition 4.5 we get $\mathrm{Ch}(g) \leq \mathrm{Ch}(f^\tau)$ and the L^2 distance of f and g is strictly smaller than the one of f and f^τ as soon as $\mathfrak{m}(\{f^\tau > C\}) > 0$, which is a contradiction.

Starting from f_0, iterating this procedure, and using the fact that the implicit Euler scheme converges as $\tau \downarrow 0$ (see [**?**], [**?**] for details) to the gradient flow we get the conclusion. The same arguments applies to uniform bounds from below.
Entropy dissipation. The map $z \mapsto z \log z$ is Lipschitz on $[c, C]$ which, together with the maximum principle and the fact that $t \mapsto f_t \in L^2(X, \mathfrak{m})$ is locally absolutely continuous, yields the claimed absolute continuity statement. Now notice that we have $\frac{\mathrm{d}}{\mathrm{d}t} \int f_t \log f_t \,\mathrm{d}\mathfrak{m} = \int (\log f_t + 1) \Delta f_t \,\mathrm{d}\mathfrak{m}$ for a.e. t. Since by the maximum principle $f_t \geq c$ $\mathfrak{m}$-a.e., the function $\log z + 1$ is Lipschitz and C^1 on the image of f_t for any $t \geq 0$, thus from (4.6) we get the conclusion. □

4.2 The horizontal approach: weak upper gradients

In this subsection we introduce a different notion of 'weak norm of gradient' in metric measure space which is Lagrangian in spirit and it does not require a relaxation procedure. This notion of gradient will provide a new estimate of entropy dissipation along the gradient flow of Ch and it will also be useful in the analysis of the derivative of the entropy along geodesics.

While the definition of minimal relaxed slope was taken from Cheeger's work [**?**], the notion we are going to introduce is inspired by the work of Shanmugalingam [**?**], the only difference being that we consider a different notion of null set of curves.

4.2.1 Negligible sets of curves and functions Sobolev along a.e. curve

Recall that the evaluation maps $\mathrm{e}_t : C([0,1],X) \to X$ are defined by $\mathrm{e}_t(\gamma) := \gamma_t$. We also introduce the restriction maps $\mathrm{restr}_t^s : C([0,1],X) \to C([0,1],X)$, $0 \le t \le s \le 1$, given by

$$\mathrm{restr}_t^s(\gamma)_r := \gamma_{((1-r)t+rs)}, \tag{4.7}$$

so that restr_t^s restricts the curve γ to the interval $[t,s]$ and then "stretches" it on the whole of $[0,1]$.

Definition 4.10 (Test plans and negligible sets of curves) *We say that a probability measure $\boldsymbol{\pi} \in \mathscr{P}(C([0,1],X))$ is a* test plan *if $\iint_0^1 |\dot\gamma_t|^2 \mathrm{d}t\, \mathrm{d}\boldsymbol{\pi} < \infty$ and there exists a constant $C(\boldsymbol{\pi})$ such that*

$$(\mathrm{e}_t)_\#\boldsymbol{\pi} \le C(\boldsymbol{\pi})\mathfrak{m}, \qquad \forall t \in [0,1].$$

A Borel set $A \subset AC^2([0,1],X)$ is said negligible *if $\boldsymbol{\pi}(A) = 0$ for any test plan $\boldsymbol{\pi}$. A property which holds for every $\gamma \in AC^2([0,1],X)$, except possibly a negligible set, is said to hold for almost every curve.*

Coupled with the definition of negligible set of curves, there is the definition of function which is Sobolev along a.e. curve.

Definition 4.11 (Sobolev functions along a.e. curve and weak gradients) *A function $f : X \to \mathbb{R} \cup \{\pm\infty\}$ is Sobolev along a.e. curve if for a.e. curve γ the function $t \mapsto f(\gamma_t)$ coincides a.e. in $[0,1]$ and in $\{0,1\}$ with an absolutely continuous map $f_\gamma : [0,1] \to \mathbb{R}$.*

If f is Sobolev along a.e. curve, a function $G : X \to [0,\infty]$ is a weak upper gradient of f if

$$\left|\int_{\partial\gamma} f\right| \le \int_\gamma G, \qquad \text{for a.e. } \gamma, \tag{4.8}$$

where here and in the following we write $\int_{\partial\gamma} f$ for $f(\gamma_1) - f(\gamma_0)$ and $\int_\gamma G$ for $\int_0^1 |\dot\gamma_t| G(\gamma_t)\,\mathrm{d}t$.

Even if we allow $\pm\infty$ as possible values of f, it is easy to see that functions which are Sobolev along a.e. curve are indeed finite $\mathfrak{m}$-a.e. in X. Indeed, suffice to consider the test plan $\boldsymbol{\pi}$ induced by the map assigning to $x \in X$ the curve $\gamma \in C([0,1],X)$ identically equal to x.

Remark 4.12 (Restriction and equivalent formulation) Notice that if $\boldsymbol{\pi}$ is a test plan, so is $(\mathrm{restr}_t^s)_\sharp\boldsymbol{\pi}$. Hence, if G is a weak upper gradient of f, then for every $t < s$ in $[0,1]$

$$|f(\gamma_s) - f(\gamma_t)| \le \int_t^s G(\gamma_r)|\dot\gamma_r|\,\mathrm{d}r \qquad \text{for almost all } \gamma.$$

It follows that for almost all γ the function f_γ satisfies

$$|f_\gamma(s) - f_\gamma(t)| \le \int_t^s G(\gamma_r)|\dot\gamma_r|\,\mathrm{d}r \qquad \text{for all } t < s \in \mathbb{Q} \cap [0,1].$$

Since f_γ is continuous the same holds for all $t < s$ in $[0,1]$, so that we obtain an equivalent pointwise formulation of (4.8):

$$\left|\frac{\mathrm{d}}{\mathrm{d}t} f_\gamma\right| \le G \circ \gamma |\dot\gamma| \quad \text{a.e. in } [0,1], \text{ for almost all } \gamma. \tag{4.9}$$

■

Proposition 4.13 (Locality) *Let $f : X \to \mathbb{R}$ be Sobolev along almost all absolutely continuous curves, and let G_1, G_2 be weak upper gradients of f. Then $\min\{G_1, G_2\}$ is a weak upper gradient of f.*

Proof It is a direct consequence of (4.9). □

The notion of weak upper gradient enjoys natural invariance properties with respect to $\mathfrak{m}$-negligible sets:

Definition 4.14 (Minimal weak upper gradient) *Let $f : X \to \mathbb{R}$ be Sobolev along almost all curves. The minimal weak upper gradient $|\nabla f|_w$ of f is the weak upper gradient characterized, up to $\mathfrak{m}$-negligible sets, by the property*

$$|\nabla f|_w \le G \qquad \mathfrak{m}\text{-a.e. in } X\text{, for every weak upper gradient } G \text{ of } f. \tag{4.10}$$

Uniqueness of the minimal weak upper gradient is obvious. For existence, we take $|\nabla f|_w := \inf_n G_n$, where G_n are weak upper gradients which provide a minimizing sequence in

$$\inf\left\{\int_X \tan^{-1} G \,\mathrm{d}\mathfrak{m} : \ G \text{ is a weak upper gradient of } f\right\}.$$

We immediately see, thanks to Proposition 4.13, that we can assume with no loss of generality that $G_{n+1} \le G_n$. Hence, by monotone convergence, the function $|\nabla f|_w$ is a weak upper gradient of f and $\int_X \tan^{-1} G \,\mathrm{d}\mathfrak{m}$ is minimal at $G = |\nabla f|_w$. This minimality, in conjunction with Proposition 4.13, gives (4.10).

Theorem 4.15 (Stability w.r.t. $\mathfrak{m}$-a.e. convergence) *Assume that f_n are $\mathfrak{m}$-measurable, Sobolev along almost all curves and that G_n are weak upper gradients of f_n. Assume furthermore that $f_n(x) \to f(x) \in \mathbb{R}$ for $\mathfrak{m}$-a.e. $x \in X$ and that (G_n) weakly converges to G in $L^2(X, \mathfrak{m})$. Then G is a weak upper gradient of f.*

Proof Fix a test plan $\boldsymbol{\pi}$. By Mazur's theorem we can find convex combinations

$$H_n := \sum_{i=N_h+1}^{N_{h+1}} \alpha_i G_i \qquad \text{with } \alpha_i \ge 0, \sum_{i=N_h+1}^{N_{h+1}} \alpha_i = 1,\ N_h \to \infty$$

converging strongly to G in $L^2(X, \mathfrak{m})$. Denoting by $\tilde{f}_n$ the corresponding convex combinations of f_n, H_n are weak upper gradients of $\tilde{f}_n$ and still $\tilde{f}_n \to f$ $\mathfrak{m}$-a.e. in X.

Since for every nonnegative Borel function $\varphi : X \to [0, \infty]$ it holds (with $C = C(\boldsymbol{\pi})$)

$$\begin{aligned}
\int \Big(\int_\gamma \varphi\Big)\,\mathrm{d}\boldsymbol{\pi} &= \int \Big(\int_0^1 \varphi(\gamma_t)|\dot\gamma_t|\,\mathrm{d}t\Big)\,\mathrm{d}\boldsymbol{\pi} \le \int \Big(\int_0^1 \varphi^2(\gamma_t)\,\mathrm{d}t\Big)^{1/2}\Big(\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\Big)^{1/2}\mathrm{d}\boldsymbol{\pi} \\
&\le \Big(\int_0^1 \int \varphi^2\,\mathrm{d}(\mathrm{e}_t)_\sharp\boldsymbol{\pi}\,\mathrm{d}t\Big)^{1/2}\Big(\iint_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}\Big)^{1/2} \\
&\le \Big(C\int \varphi^2\,\mathrm{d}\mathfrak{m}\Big)^{1/2}\Big(\iint_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}\Big)^{1/2},
\end{aligned} \tag{4.11}$$

we obtain, for $\bar{C} := \sqrt{C}\Big(\iint_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}\Big)^{1/2}$,

$$\int \Big(\int_\gamma |H_n - G| + \min\{|\tilde{f}_n - f|, 1\}\Big)\,\mathrm{d}\boldsymbol{\pi} \le \bar{C}\Big(\|H_n - G\|_{L^2} + \|\min\{|\tilde{f}_n - f|, 1\}\|_{L^2}\Big) \to 0.$$

By a diagonal argument we can find a subsequence $n(k)$ such that $\int_\gamma |H_{n(k)}-G|+\min\{|\tilde f_{n(k)}-f|,1\}\to 0$ as $k\to\infty$ for $\boldsymbol{\pi}$-a.e. γ. Since $\tilde f_n$ converge $\mathfrak{m}$-a.e. to f and the marginals of $\boldsymbol{\pi}$ are absolutely continuous w.r.t. $\mathfrak{m}$ we have also that for $\boldsymbol{\pi}$-a.e. γ it holds $\tilde f_n(\gamma_0)\to f(\gamma_0)$ and $\tilde f_n(\gamma_1)\to f(\gamma_1)$.

If we fix a curve γ satisfying these convergence properties, since $(\tilde f_{n(k)})_\gamma$ are equi-absolutely continuous (being their derivatives bounded by $H_{n(k)}\circ\gamma|\dot\gamma|$) and a further subsequence of $\tilde f_{n(k)}$ converges a.e. in $[0,1]$ and in $\{0,1\}$ to $f(\gamma_s)$, we can pass to the limit to obtain an absolutely continuous function f_γ equal to $f(\gamma_s)$ a.e. in $[0,1]$ and in $\{0,1\}$ with derivative bounded by $G(\gamma_s)|\dot\gamma_s|$. Since $\boldsymbol{\pi}$ is arbitrary we conclude that f is Sobolev along almost all curves and that G is a weak upper gradient of f. □

Remark 4.16 Notice that in particular, the previous proposition ensures that the notions of being Sobolev along a.e. curve and of weak upper gradient are invariant under modification of the function/weak upper gradient on $\mathfrak{m}$-negligible sets. ■

Remark 4.17 ($|\nabla f|_w\leq|\nabla f|_*$) Another immediate consequence of the previous proposition is that any $f\in D(\mathrm{Ch})$ is Sobolev along a.e. curve and satisfies $|\nabla f|_w\leq|\nabla f|_*$. Indeed, for such f just pick a sequence of Lipschitz functions such that $|\nabla f_n|\to|\nabla f|_*$ in $L^2(X,\mathfrak{m})$ (as in Proposition 4.3) and recall that for Lipschitz functions the local Lipschitz constant is an upper gradient. ■

4.2.2 A bound from below on weak gradients

In this short subsection we show how, using test plans and the very definition of minimal weak gradients, it is possible to use $|\nabla f|_w$ to bound from below the increments of the relative entropy. We start with the following result, proved - in a more general setting - by Lisini in [?]: it shows how to associate to an absolutely continuous curve μ_t w.r.t. W_2 a plan $\boldsymbol{\pi}\in\mathscr{P}(C([0,1],X))$ representing the curve itself (see also Theorem 8.2.1. of [?] for the Euclidean case). We will only sketch the proof.

Proposition 4.18 (Superposition principle) *Let (X,d) be a compact space and let $(\mu_t)\subset\mathscr{P}(X)$ be an absolutely continuous curve w.r.t. W_2. Then there exists $\boldsymbol{\pi}\in\mathscr{P}(C([0,1],X))$ such that $(\mathrm{e}_t)_\sharp\boldsymbol{\pi}=\mu_t$ for any $t\in[0,1]$ and $\int|\dot\gamma_t|^2\,\mathrm{d}\boldsymbol{\pi}(\gamma)=|\dot\mu_t|^2$ for a.e. $t\in[0,1]$.*

Proof Up to a reparametrization, we can assume that $|\dot\mu_t|\in L^2(0,1)$. Now let $\boldsymbol{\pi}\in C([0,1],X)$ be any plan concentrated on $AC^2([0,1],X)$ such that $(\mathrm{e}_t)_\sharp\boldsymbol{\pi}=\mu_t$ for any $t\in[0,1]$ and notice that since $(\mathrm{e}_t,\mathrm{e}_s)_\sharp\boldsymbol{\pi}\in\mathrm{A{\small DM}}(\mu_t,\mu_s)$, for any $t<s$ it holds

$$W_2^2(\mu_t,\mu_s)\leq\int\mathsf{d}^2(\gamma_t,\gamma_s)\,\mathrm{d}\boldsymbol{\pi}(\gamma)\leq\int\left(\int_t^s|\dot\gamma_r|\,\mathrm{d}r\right)^2\mathrm{d}\boldsymbol{\pi}(\gamma)\leq(s-t)\iint_t^s|\dot\gamma_r|^2\,\mathrm{d}r\,\mathrm{d}\boldsymbol{\pi}(\gamma),$$

which shows that $|\dot\mu_t|^2\leq\int|\dot\gamma_t|^2\,\mathrm{d}\boldsymbol{\pi}(\gamma)$ for a.e. t. Hence, to conclude it is sufficient to find a plan $\boldsymbol{\pi}\in\mathscr{P}(C([0,1],X))$, concentrated on $AC^2([0,1],X)$, with $(\mathrm{e}_t)_\sharp\boldsymbol{\pi}=\mu_t$ for any $t\in[0,1]$ such that $\int|\dot\mu_t|^2\,\mathrm{d}t\geq\iint_0^1|\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma)$.

To build such a $\boldsymbol{\pi}$ we make the simplifying assumption that (X,d) is geodesic (the proof for the general case is similar, but rather than interpolating with piecewise geodesic curves one uses piecewise constant ones, this leads to some technical complications that we want to avoid here - see [?] for the complete argument). Fix $n\in\mathbb{N}$ and use a gluing argument to find

$\boldsymbol{\gamma}^n \in \mathscr{P}(X^{n+1})$ such that $(\pi^i,\pi^{i+1})_\sharp \boldsymbol{\gamma}^n \in \mathrm{OPT}(\mu_{\frac{i}{n}},\mu_{\frac{i+1}{n}})$ for $i=0,\ldots,n-1$. By standard measurable selection arguments, there exists a Borel map $T^n : X^{n+1}\to C([0,1],X)$ such that $\gamma := T^n(x_0,\ldots,x_n)$ is a constant speed geodesic on each of the intervals $[i/n,(i+1)/n]$ and $\gamma_{i/n}=x_i$, $i=0,\ldots,n$. Define $\boldsymbol{\pi}^n := T^n_\sharp \boldsymbol{\gamma}^n$. It holds

$$\iint_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}^n(\gamma) = \frac{1}{n}\int\sum_{i=0}^{n-1}\mathsf{d}^2\big(\gamma_{\frac{i}{n}},\gamma_{\frac{i+1}{n}}\big)\,\mathrm{d}\boldsymbol{\pi}(\gamma) = \frac{1}{n}\sum_{i=0}^{n-1}W_2^2\big(\mu_{\frac{i}{n}},\mu_{\frac{i+1}{n}}\big)\le \int_0^1|\dot\mu_t|^2\,\mathrm{d}t. \tag{4.12}$$

Now notice that the map $E : C([0,1],X)\to[0,\infty]$ given by $E(\gamma) := \int_0^1|\dot\gamma_t|^2\,\mathrm{d}t$ if $\gamma\in AC^2([0,1],X)$ and $+\infty$ otherwise, is lower semicontinuous and, via a simple equicontinuity argument, with compact sublevels. Therefore by Prokorov's theorem we get that $(\boldsymbol{\pi}^n)\subset\mathscr{P}(C([0,1],X))$ is a tight sequence, hence for any limit measure $\boldsymbol{\pi}$ the uniform bound (4.12) gives the thesis. □

Proposition 4.19 *Let $[0,1]\ni t\mapsto\mu_t=f_t\mathfrak{m}$ be a curve in $AC^2([0,1],(\mathscr{P}(X),W_2))$. Assume that for some $0<c<C<\infty$ it holds $c\le f_t\le C$ $\mathfrak{m}$-a.e. for any $t\in[0,1]$, and that f_0 is Sobolev along a.e. curve with $|\nabla f_0|_w\in L^2(X,\mathfrak{m})$. Then*

$$\int f_0\log f_0\,\mathrm{d}\mathfrak{m}-\int f_t\log f_t\,\mathrm{d}\mathfrak{m}\le\frac12\int_0^t\int_X\frac{|\nabla f_0|_w^2}{f_0^2}f_s\,\mathrm{d}s\,\mathrm{d}\mathfrak{m}+\frac12\int_0^t|\dot\mu_s|^2\,\mathrm{d}s,\qquad\forall t>0.$$

Proof Let $\boldsymbol{\pi}\in\mathscr{P}(C([0,1],X))$ be a plan associated to the curve (μ_t) as in Proposition 4.18. The assumption $f_t\le C$ $\mathfrak{m}$-a.e. and the fact that $\iint_0^1|\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma)=\int|\dot\mu_t|^2\,\mathrm{d}t<\infty$ guarantee that it is a test plan. Now notice that it holds $|\nabla\log f_t|_w=|\nabla f_t|_w/f_t$ (because $z\mapsto\log z$ is C^1 in $[c,C]$)), thus we get

$$\begin{aligned}\int f_0\log f_0\,\mathrm{d}\mathfrak{m}-\int f_t\log f_t\,\mathrm{d}\mathfrak{m}&\le\int\log f_0(f_0-f_t)\,\mathrm{d}\mathfrak{m}=\int\log f_0\circ\mathrm{e}_0-\log f_0\circ\mathrm{e}_t\,\mathrm{d}\boldsymbol{\pi}\\&\le\iint_0^t\frac{|\nabla f_0|_w(\gamma_s)}{f_0(\gamma_s)}|\dot\gamma_s|\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}(\gamma)\\&\le\frac12\iint_0^t\frac{|\nabla f_0|_w^2(\gamma_s)}{f_0^2(\gamma_s)}\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}(\gamma)+\frac12\iint_0^t|\dot\gamma_s|^2\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}(\gamma)\\&=\frac12\int_0^t\int_X\frac{|\nabla f_0|_w^2}{f_0^2}f_s\,\mathrm{d}s\,\mathrm{d}\mathfrak{m}+\frac12\int_0^t|\dot\mu_s|^2\,\mathrm{d}s.\end{aligned}$$

□

4.3 The two notions of gradient coincide

Here we prove that the two notions of "norm of weak gradient" we introduced coincide. We already noticed in Remark 4.17 that $|\nabla f|_w\le|\nabla f|_*$, so that to conclude we need to show that $|\nabla f|_w\ge|\nabla f|_*$.

The key argument to achieve this, is the following lemma which gives a sharp bound on the W_2-speed of the L^2-gradient flow of Ch. This lemma has been introduced in [?] to study the heat flow on Alexandrov spaces, see also Section 6.

Lemma 4.20 (Kuwada's lemma) *Let $f_0 \in L^2(X,\mathfrak{m})$ and let (f_t) be the L_2-gradient flow of* Ch *starting from f_0. Assume that for some $0 < c \le C < \infty$ it holds $c \le f_0 \le C$ $\mathfrak{m}$-a.e. in X, and that $\int f_0 \,\mathrm{d}\mathfrak{m} = 1$. Then the curve $t \mapsto \mu_t := f_t\mathfrak{m}$ is absolutely continuous w.r.t. W_2 and it holds*

$$|\dot\mu_t|^2 \le \int \frac{|\nabla f_t|_*^2}{f_t}\,\mathrm{d}\mathfrak{m}, \qquad \text{for a.e. } t \in (0,\infty).$$

Proof We start from the duality formula (2.4) with $\varphi = -\psi$: taking into account the factor 2 and using the identity $Q_1(-\psi) = \psi^c$ we get

$$\frac{W_2^2(\mu,\nu)}{2} = \sup_\varphi \int_X Q_1\varphi\,d\nu - \int_X \varphi\,d\mu \tag{4.13}$$

where the supremum runs among all Lipschitz functions φ.

Fix such a φ and recall (Proposition 3.3) that the map $t \mapsto Q_t\varphi$ is Lipschitz with values in $L^\infty(X,\mathfrak{m})$, and a fortiori in $L^2(X,\mathfrak{m})$.

Fix also $0 \le t < s$, set $\ell = (s-t)$ and recall that since (f_t) is the Gradient Flow of Ch in L^2, the map $[0,\ell] \ni \tau \mapsto f_{t+\tau}$ is Lipschitz with values in L^2. Therefore the map $[0,\ell] \ni \tau \mapsto Q_{\frac{\tau}{\ell}}\varphi f_{t+\tau}$ is Lipschitz with values in L^2. The equality

$$\frac{Q_{\frac{\tau+h}{\ell}}\varphi f_{t+\tau+h} - Q_{\frac{\tau}{\ell}}\varphi f_{t+\tau}}{h} = f_{t+\tau}\frac{Q_{\frac{\tau+h}{\ell}} - Q_{\frac{\tau}{\ell}}\varphi}{h} + Q_{\frac{\tau+h}{\ell}}\varphi\frac{f_{t+\tau+h} - f_{t+\tau}}{h},$$

together with the uniform continuity of $(x,\tau) \mapsto Q_{\frac{\tau}{\ell}}\varphi(x)$ shows that the derivative of $\tau \mapsto Q_{\ell\tau}\varphi f_{t+\tau}$ can be computed via the Leibniz rule.

We have:

$$\begin{aligned}\int_X Q_1\varphi\,\mathrm{d}\mu_s - \int_X \varphi\,\mathrm{d}\mu_t &= \int Q_1\varphi f_{t+\ell}\,\mathrm{d}\mathfrak{m} - \int_X \varphi f_t\,\mathrm{d}\mathfrak{m} = \int_X\int_0^\ell \frac{\mathrm{d}}{\mathrm{d}\tau}\big(Q_{\frac{\tau}{\ell}}\varphi f_{t+\tau}\big)d\tau\,\mathrm{d}\mathfrak{m}\\ &= \int_X\int_0^\ell -\frac{|\nabla Q_{\frac{\tau}{\ell}}\varphi|^2}{2\ell}f_{t+\tau} + Q_{\frac{\tau}{\ell}}\varphi\Delta f_{t+\tau}\,\mathrm{d}\tau\,\mathrm{d}\mathfrak{m},\end{aligned}\tag{4.14}$$

having used Theorem 3.6.

Observe that by inequalities (4.5) and (4.1) we have

$$\begin{aligned}\int_X Q_{\frac{\tau}{\ell}}\varphi\Delta f_{t+\tau}\,\mathrm{d}\mathfrak{m} &\le \int_X |\nabla Q_{\frac{\tau}{\ell}}\varphi|_*|\nabla f_{t+\tau}|_*\,\mathrm{d}\mathfrak{m} \le \int_X |\nabla Q_{\frac{\tau}{\ell}}\varphi||\nabla f_{t+\tau}|_*\,\mathrm{d}\mathfrak{m}\\ &\le \frac{1}{2\ell}\int_X |\nabla Q_{\frac{\tau}{\ell}}\varphi|^2 f_{t+\tau}\mathrm{d}\mathfrak{m} + \frac{\ell}{2}\int_X \frac{|\nabla f_{t+\tau}|_*^2}{f_{t+\tau}}\,\mathrm{d}\mathfrak{m}.\end{aligned}\tag{4.15}$$

Plugging this inequality in (4.14), we obtain

$$\int_X Q_1\varphi\,\mathrm{d}\mu_s - \int_X \varphi\,\mathrm{d}\mu_t \le \frac{\ell}{2}\int_0^\ell\int_X \frac{|\nabla f_{t+\tau}|_*^2}{f_{t+\tau}}\,\mathrm{d}\mathfrak{m}.$$

This latter bound does not depend on φ, so from (4.13) we deduce

$$W_2^2(\mu_t,\mu_s) \le \ell\int_0^\ell\int_X \frac{|\nabla f_{t+\tau}|_*^2}{f_{t+\tau}}\,\mathrm{d}\mathfrak{m}.$$

Since $f_\tau \ge c$ for any $\tau \ge 0$ and $\tau \mapsto \mathrm{Ch}(f_\tau)$ is nonincreasing and finite for every $\tau > 0$, we immediately get that $t \mapsto \mu_t$ is locally Lipschitz in $(0,\infty)$. At Lebesgue points of $\tau \mapsto \int_X |\nabla f_\tau|_*^2/f_\tau\,\mathrm{d}\mathfrak{m}$ we obtain the stated pointwise bound on the metric speed. □

Theorem 4.21 *Let $f \in L^2(X, \mathfrak{m})$. Assume that f is Sobolev along a.e. curve and that $|\nabla f|_w \in L^2(X, \mathfrak{m})$. Then $f \in D(\mathrm{Ch})$ and $|\nabla f|_* = |\nabla f|_w$ $\mathfrak{m}$-a.e. in X.*

Proof Up to a truncation argument and addition of a constant, we can assume that $0 < c \le f \le C < \infty$ $\mathfrak{m}$-a.e. in X for some c, C. Let (f_t) be the L_2-gradient flow of Ch and recall that from Proposition 4.9 we have

$$\int f \log f \,\mathrm{d}\mathfrak{m} - \int f_t \log f_t \,\mathrm{d}\mathfrak{m} = \int_0^t \int_X \frac{|\nabla f_s|_*^2}{f_s} \,\mathrm{d}s \,\mathrm{d}\mathfrak{m} < \infty, \qquad \forall t > 0.$$

On the other hand, from Proposition 4.19 and Lemma 4.20 we have

$$\int f \log f \,\mathrm{d}\mathfrak{m} - \int f_t \log f_t \,\mathrm{d}\mathfrak{m} \le \frac{1}{2} \int_0^t \int_X \frac{|\nabla f|_w^2}{f^2} f_s \,\mathrm{d}s \,\mathrm{d}\mathfrak{m} + \frac{1}{2} \int_0^t \int_X \frac{|\nabla f_s|_*^2}{f_s} \,\mathrm{d}s \,\mathrm{d}\mathfrak{m}.$$

Hence we deduce

$$\int_0^t 4\mathrm{Ch}(\sqrt{f_s}) \,\mathrm{d}s = \frac{1}{2} \int_0^t \int_X \frac{|\nabla f_s|_*^2}{f_s} \,\mathrm{d}s \,\mathrm{d}\mathfrak{m} \le \frac{1}{2} \int_0^t \int_X \frac{|\nabla f|_w^2}{f^2} f_s \,\mathrm{d}s \,\mathrm{d}\mathfrak{m}.$$

Letting $t \downarrow 0$, taking into account the L^2-lower semicontinuity of Ch and the fact - easy to check from the maximum principle - that $\sqrt{f_s} \to \sqrt{f}$ as $s \downarrow 0$ in $L^2(X, \mathfrak{m})$, we get $\mathrm{Ch}(\sqrt{f}) \le \varliminf_{t \downarrow 0} \frac{1}{t} \int_0^t \mathrm{Ch}(\sqrt{f_s}) \,\mathrm{d}s$. On the other hand, the bound $f \ge c > 0$ ensures $\frac{|\nabla f|_w^2}{f^2} \in L^1(X, \mathfrak{m})$ and the maximum principle again together with the convergence of f_s to f in $L^2(X, \mathfrak{m})$ when $s \downarrow 0$ grants that the convergence is also weak* in $L^\infty(X, \mathfrak{m})$, therefore $\int \frac{|\nabla f|_w^2}{f} \,\mathrm{d}\mathfrak{m} = \frac{1}{t} \lim_{t \downarrow 0} \int_0^t \int_X \frac{|\nabla f|_w^2}{f^2} f_s \,\mathrm{d}\mathfrak{m} \,\mathrm{d}s$.

In summary, we proved

$$\frac{1}{2} \int \frac{|\nabla f|_*^2}{f} \,\mathrm{d}\mathfrak{m} \le \frac{1}{2} \int \frac{|\nabla f|_w^2}{f} \,\mathrm{d}\mathfrak{m},$$

which, together with the inequality $|\nabla f|_w \le |\nabla f|_*$ $\mathfrak{m}$-a.e. in X, gives the conclusion. □

We are now in the position of defining the Sobolev space $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$. We start with the following simple and general lemma.

Lemma 4.22 *Let $(B, \|\cdot\|)$ be a Banach space and let $E : B \to [0, \infty]$ be a 1-homogeneous, convex and lower semicontinuous map. Then the vector space $\{E < \infty\}$ endowed with the norm*

$$\|v\|_E := \sqrt{\|v\|^2 + E^2(v)},$$

is a Banach space.

Proof It is clear that $(D(E), \|\cdot\|_E)$ is a normed space, so we only need to prove completeness. Pick a sequence $(v_n) \subset D(E)$ which is Cauchy w.r.t. $\|\cdot\|_E$. Then, since $\|\cdot\| \le \|\cdot\|_E$ we also get that (v_n) is Cauchy w.r.t. $\|\cdot\|$, and hence there exists $v \in B$ such that $\|v_n - v\| \to 0$. The lower semicontinuity of E grants that $E(v) \le \varliminf_n E(v_n) < \infty$ and also that it holds

$$\varlimsup_{n \to \infty} \|v_n - v\|_E \le \varlimsup_{n,m \to \infty} \|v_n - v_m\|_E = 0,$$

which is the thesis. □

Therefore, if we want to build the space $W^{1,2}(X, \mathsf{d}, \mathfrak{m}) \subset L^2(X, \mathfrak{m})$, the only thing that we need is an L^2-lower semicontinuous functional playing the role which on $\mathbb{R}^d$ is played by the L^2-norm of the distributional gradient of Sobolev functions. We certainly have this functional, namely the map $f \mapsto \||\nabla f|_*\|_{L^2(X,\mathfrak{m})} = \||\nabla f|_w\|_{L^2(X,\mathfrak{m})}$. Hence the lemma above provides the Banach space $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$. Notice that in general $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ is not Hilbert: this is not surprising, as already the Sobolev space $W^{1,2}$ built over $(\mathbb{R}^d, \|\cdot\|, \mathcal{L}^d)$ is not Hilbert if the underlying norm $\|\cdot\|$ does not come from a scalar product.

4.4 Comparison with previous approaches

It is now time to underline that the one proposed here is certainly not the first definition of Sobolev space over a metric measure space (we refer to [**?**] for a much broader overview on the subject). Here we confine the discussion only to weak notions of (modulus of) gradient, and in particular to [**?**] and [**?**]. Also, we discuss only the quadratic case, referring to [][4] to general power functions p and the independence (in a suitable sense) of p of minimal gradients.

In [**?**] Cheeger proposed a relaxation procedure similar to the one used in Subsection 4.1, but rather than relaxing the local Lipschitz constant of Lipschitz functions, he relaxed upper gradients of arbitrary functions. More precisely, he defined

$$E(f) := \inf \varliminf_{n\to\infty} \|G_n\|_{L^2(X,\mathfrak{m})},$$

where the infimum is taken among all sequences (f_n) converging to f in $L^2(X, \mathfrak{m})$ such that G_n is an upper gradient for f_n. Then, with the same computations done in Subsection 4.1 (which actually inspired our work, which came 10 years later!) he showed that for $f \in D(E)$ there is an underlying notion of weak gradient $|\nabla f|_C$, called minimal generalized upper gradient, such that $E(f) = \||\nabla f|_C\|_{L^2(X,\mathfrak{m})}$ and

$$|\nabla f|_C \le G \qquad \mathfrak{m}\text{-a.e. in } X,$$

for any G weak limit of a sequence (G_n) as in the definition of $E(f)$.

Notice that since the local Lipschitz constant is always an upper gradient for Lipschitz functions, one certainly has

$$|\nabla f|_C \le |\nabla f|_* \qquad \mathfrak{m}\text{-a.e. in } X, \text{ for any } f \in D(\mathrm{Ch}). \tag{4.16}$$

In [**?**] Shanmugalingam used a procedure close to the one used by us to define weak gradients (again: actually Shanmugalingam's work came first and we have been inspired by her) to produce a notion of "norm of weak gradient" which does not require a relaxation procedure. Recall that for $\Gamma \subset AC([0,1], X)$ the 2-Modulus $\mathrm{Mod}_2(\Gamma)$ is defined by

$$\mathrm{Mod}_2(\Gamma) := \inf \Big\{ \|\rho\|^2_{L^2(X,\mathfrak{m})} \ : \ \int_\gamma \rho \ge 1, \ \ \forall \gamma \in \Gamma \Big\}, \qquad \forall \Gamma \subset AC([0,1], X) \tag{4.17}$$

Is it possible to show that the 2-Modulus is an outer measure on $AC([0,1], X)$. Building on this notion, Koskela and MacManus [][5] considered the class of functions f which satisfy the upper gradient inequality not necessarily along all curves, but only out of a Mod_2-negligible set of

[4] Inserire riferimento a Gradientspq

[5] Inserire citazione Koskela-MacManus

curves. In order to compare more properly this concept to Sobolev classes, Shanmugalingam said that $G : X \to [0,\infty]$ is a weak upper gradient for f if there exists $\tilde{f} = f$ $\mathfrak{m}$-a.e. such that

$$\left|\tilde{f}(\gamma_0) - \tilde{f}(\gamma_1)\right| \leq \int_\gamma G \qquad \forall \gamma \in AC([0,1], X) \setminus \mathcal{N},$$

where $\mathrm{Mod}_2(\mathcal{N}) = 0$. Then, she defined the energy $\tilde{E} : L^2 \to [0,\infty]$ by putting

$$\tilde{E}(f) := \inf \|G\|^2_{L^2},$$

where the infimum is taken among all weak upper gradient G of f according to the previous condition. Thanks to the properties of the 2-modulus (a stability property of weak upper gradients analogous to ours), it is possible to show that $\tilde{E}$ is indeed L^2-lower semicontinuous, so that it leads to a good definition of the Sobolev space. Also, using a key lemma due to Fuglede, Shanmugalingam proved that $E = \tilde{E}$ on L^2, so that they produce the same definition of Sobolev space $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ and the underlying gradient $|\nabla f|_S$ which gives a pointwise representation to $\tilde{E}(f)$ is the same $|\nabla f|_C$ behind the energy E.

Observe now that for a Borel set $\Gamma \subset AC^2([0,1], X)$ and a test plan $\boldsymbol{\pi}$, integrating w.r.t. $\boldsymbol{\pi}$ the inequality $\int_\gamma \rho \geq 1$ $\forall \gamma \in \Gamma$ and then minimizing over ρ, we get

$$\left[\boldsymbol{\pi}(\Gamma)\right]^2 \leq C(\boldsymbol{\pi}) \mathrm{Mod}_2(\Gamma) \iint_0^1 |\dot{\gamma}|^2 \, \mathrm{d}s \, \mathrm{d}\boldsymbol{\pi}(\gamma),$$

which shows that any Mod_2-negligible set of curves is also negligible according to Definition 4.10. This fact easily yields that any $f \in D(\tilde{E})$ is Sobolev along a.e. curve and satisfies

$$|\nabla f|_w \leq |\nabla f|_C, \qquad \mathfrak{m}\text{-a.e. in } X. \tag{4.18}$$

Given that we proved in Theorem 4.21 that $|\nabla f|_* = |\nabla f|_w$, inequalities (4.16) and (4.18) also give that $|\nabla f|_* = |\nabla f|_w = |\nabla f|_C = |\nabla f|_S$ (the smallest one among the four notions coincides with the largest one).

What we get by the new approach to Sobolev spaces on metric measure spaces is the following result.

Theorem 4.23 (Density in energy of Lipschitz functions) *Let $(X, \mathsf{d}, \mathfrak{m})$ be a compact normalized metric measure space. Then for any $f \in L^2(X, \mathfrak{m})$ with weak upper gradient in L^2* [6] *there exists a sequence (f_n) of Lipschitz functions converging to f in $L^2(X, \mathfrak{m})$ such that both $|\nabla f_n|$ and $|\nabla f_n|_w$ converge to $|\nabla f|_w$ in $L^2(X, \mathfrak{m})$ as $n \to \infty$.*

Proof Straightforward consequence of the identity of weak and relaxed gradients and of Proposition 4.3. □

As a matter of fact, the construction of f_n could be seen as the combination of two steps: first, the application of the gradient flow of Ch with f as initial datum, to obtain a family of functions g_n convergent to f in $L^2(X, \mathfrak{m})$ and satisfying

$$\limsup_{n\to\infty} \int |\nabla g_n|^2_* \, \mathrm{d}\mathfrak{m} \leq \int |\nabla f|^2_w \, \mathrm{d}\mathfrak{m}.$$

[6] Ho sostituito l'ipotesi $f \in W^{1,2}$ con l'ipotesi che f abbia un weakgrad. Altrimenti non si capisce il punto chiave e uno potrebbe pensare, intendendo il gradiente come quello rilassato, che sia solo coinvolta la Proposizione 4.3

This is the step implicit in the proof of Theorem 4.21. Then, we may approximate g_n by Lipschitz functions $h_{n,k}$ via Proposition 4.3, or consider approximate solutions of the variational problem

$$\inf \left\{ \mathrm{Ch}(h) + k\|h - g_n\|^2_{L^2(X,\mathfrak{m})} :\ h \in \mathrm{Lip}(X) \right\}.$$

A diagonal argument then yields the sequence $f_n = h_{n,k(n)}$.

This density result was previously known (via the use of maximal functions and covering arguments) under the assumption that the space was doubling and supported a local Poincaré inequality for weak upper gradients, see [**?**, Theorem 4.14, Theorem 4.24]. Actually, Cheeger proved more, namely that under these hypotheses Lipschitz functions are dense in the $W^{1,2}$ norm, a result which is still unknown in the general case. Also, notice that another byproduct of our density in energy result is the equivalence of local Poincaré inequality stated for Lipschitz functions on the left hand side and slope on the right hand side, and local Poincaré inequality stated for general functions on the left hand side and upper gradients on the right hand side; this result was previously known [][7] under much more restrictive assumptions on the metric measure structure.

In light of our result, where the construction of the optimal sequence is indirectly provided by the gradient flow of Ch, it is then natural to ask: is there an explicit construction such that, starting from a function f with weak upper gradient in L^2, produces a sequence of Lipschitz functions converging to f as in the statement of Theorem 4.23?

[8] It is unclear to us whether such construction exists; what is known to be possible, and somehow hidden in the proof of Theorem 4.21, is the construction of a sequence of Lipschitz functions (f_n) such that both $|\nabla f_n|_w$ and $|\nabla f_n|$ converge to $|\nabla f|_w$ in $L^2(X,\mathfrak{m})$; however, we are not able to show in general that f_n converge to f in L^2). We remark that although such more explicit construction gives an approximation weaker than the one expressed in Theorem 4.23, this is still an improvement over previous known approximation results.

We now describe how to build such approximating sequence. The key point is the following: the proof of Theorem 4.21 shows that in Kuwada's lemma equality must hold for a.e. t, and this implies that when s is close to t the inequalities in (4.15) must be almost equalities. Therefore, with the notation of Kuwada's lemma, $|\nabla Q_{t+\frac{\tau}{\ell}}\varphi|, |\nabla Q_{t+\frac{\tau}{\ell}}\varphi|_w$ must be close to $\frac{|\nabla f_{t+\tau}|_w}{f_{t+\tau}} = |\nabla \log f_{t+\tau}|_w$ which more or less gives the claim, given that the functions $Q_{t+\frac{\tau}{\ell}}\varphi$ are Lipschitz.

More rigorously, let $h \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$ and, up to a truncation argument, assume that h is bounded. Define $f_0 := ce^h$, c being such that $\int f_0 \,\mathrm{d}\mathfrak{m} = 1$, notice that f_0 is bounded away from 0 and ∞, let (f_t) be the gradient flow of Ch starting from f_0 and $\mu_t := f_t\mathfrak{m}$. The continuity of $t \mapsto f_t \in L^2(X,\mathfrak{m})$, the lower semicontinuity of Ch and the fact that $t \mapsto \mathrm{Ch}(f_t)$ is nonincreasing yield

$$|\nabla f_t|_w \to |\nabla f_0|_w, \qquad \text{in } L^2(X,\mathfrak{m}), \quad \text{as } t \downarrow 0. \tag{4.19}$$

[7] Citare nota di Heinonen-Koskela

[8] Io trovo la discussione da qui in poi un po' speculativa. Contrasta con lo stile del resto dell'articolo, mi chiedo che valenza abbia un risultato di approssimazione nel quale si approssimano solo i gradienti e non le mappe. Inoltre dire che la soluzione di un gradient flow piu' lo studio di un inf (la procedura che ho esplicitato sopra) e' non esplicita, mentre lo e' la soluzione di un gradient flow piu' un problema di minimo di Kantorovich, e' un po' questione di gusti. Ho comunque corretto alcuni misprint e segnalato alcune cose, secondo me o va maggiormente dettagliata o eliminata. Lascio comunque ogni valutazione a voi due, come sempre il parere di Giuseppe sara' determinante

From the proof of Theorem 4.21 we deduce that

$$\int_0^t |\dot{\mu}_t|^2 \,\mathrm{d}t = \int_0^t \int \frac{|\nabla f_s|_w^2}{f_s} \,\mathrm{d}s \,\mathrm{d}\mathfrak{m}.$$

Now for any $t > 0$ let ψ_t be a Kantorovich potential from μ_0 to μ_t (in particular, ψ_t is Lipschitz) and $\varphi_t := -\psi_t$, so that $\frac{1}{2}W_2^2(\mu_0, \mu_t) = \int Q_1\varphi_t \,\mathrm{d}\mu_t - \int \varphi_t \,\mathrm{d}\mu_0$. Notice that from the proof of Lemma 4.20 - in particular inequalities (4.15) - we get that

$$\lim_{t\downarrow 0} \left| \frac{1}{t}\int_0^t \int |\nabla Q_{\frac{\tau}{t}}\varphi_t|_w |\nabla f_\tau|_w \,\mathrm{d}\tau \,\mathrm{d}\mathfrak{m} - \left(\frac{1}{2}\int_0^t \int \frac{|\nabla Q_{\frac{\tau}{t}}\varphi_t|_w^2}{t^2} f_\tau + \frac{|\nabla f_\tau|_w^2}{f_\tau} \,\mathrm{d}\tau \,\mathrm{d}\mathfrak{m} \right) \right| = 0.$$

Therefore, taking into account (4.19), the fact that $f_t(x)$ is uniformly in t, x bounded away from 0 and ∞, we get that as $t \downarrow 0$ and for a.e. $\tau < t$ it holds[9]

$$\frac{|\nabla Q_{\frac{\tau}{t}}\varphi_t|_w}{t} \to \frac{|\nabla f_0|_w}{f_0} = |\nabla h|_w, \qquad \text{in } L^2(X, \mathfrak{m}).$$

The fact that also $\frac{|\nabla Q_{\frac{\tau}{t}}\varphi_t|}{t}$ converges to $|\nabla h|_w$ in $L^2(X, \mathfrak{m})$ follows along the same lines using the fact that the second inequality in (4.15) is infinitesimally an equality.[10]

It might look bizarre that the proof of Theorem 4.21 does not produce any explicit sequence of Lipschitz functions as in the previous construction, however, as far as we can see, this is the case. The problem relies in the equality case for the inequality

$$\int g\Delta f \,\mathrm{d}\mathfrak{m} \le \int |\nabla f|_w |\nabla g|_w \,\mathrm{d}\mathfrak{m}, \tag{4.20}$$

which is used in the proof of Lemma 4.20 (notice that out of the 3 inequalities in (4.15), the equality case for the first one - which is (4.20) - is the only one which we didn't use in the construction depicted above). Indeed, equality certainly holds if $g = -f + c$, $c \in \mathbb{R}$, but this is not the only possibility. For example, if (X, d) is disconnected, the value of the constant can change from one connected component to the other. More dramatically, if $(X, \mathsf{d}, \mathfrak{m})$ is such that $\mathrm{Ch} \equiv 0$ (this is the case, for instance, if $\mathfrak{m}$ is concentrated on a countable set), then $D(\Delta) = L^2(X, \mathfrak{m})$ with $\Delta f = 0$ for any f. This means that in this case equality holds in (4.20) for any couple of functions f, $g \in L^2(X, \mathfrak{m})$, because both sides are zero. In this situation if we pick an arbitrary bounded $h \in W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ and the we start the gradient flow (f_t) of Ch from $f_0 := ce^h$ as before, we obtain $f_t = f_0$ for any $t > 0$, therefore the (opposite of) optimal Kantorovich potentials are just constant functions, and so the same is true for their evolution with the Hopf-Lax formula. Hence there is no chance to get L^2-approximation of h (yet, certainly the weak gradients converge to the weak gradient of h, given that all are zero).

Therefore it looks very difficult to get an explicit approximating sequence of Lipschitz functions by digging down the proofs of Theorem 4.21 and Lemma 4.20.

[9] Qui e' un po' confuso, t tende anche a zero se capisco bene

[10] Anche qui, nel caso, conviene scriverlo per bene

5 The relative entropy and its W_2-gradient flow

In this section we study the W_2-gradient flow of the relative entropy on spaces with Ricci curvature bounded below (in short: $CD(K,\infty)$ spaces). The content is essentially extracted from [?]. As before the space $(X,\mathsf{d},\mathfrak{m})$ is a compact and normalized (i.e. $\mathfrak{m}(X)=1$).

Recall that the relative entropy functional $\mathrm{Ent}_\mathfrak{m}:\mathscr{P}(X)\to[0,\infty]$ is defined by

$$\mathrm{Ent}_\mathfrak{m}(\mu):=\begin{cases}\displaystyle\int_X f\log f\,\mathrm{d}\mathfrak{m} & \text{if } \mu=f\mathfrak{m},\\ +\infty & \text{otherwise.}\end{cases}$$

Definition 5.1 (Weak bound from below on the Ricci curvature) *We say that $(X,\mathsf{d},\mathfrak{m})$ has Ricci curvature bounded from below by K for some $K\in\mathbb{R}$ if the Relative Entropy functional $\mathrm{Ent}_\mathfrak{m}$ is K-convex along geodesics in $(\mathscr{P}(X),W_2)$. More precisely, if for any μ_0, $\mu_1\in D(\mathrm{Ent}_\mathfrak{m})$ there exists a constant speed geodesic $\mu_t:[0,1]\to\mathscr{P}(X)$ between μ_0 and μ_1 satisfying*

$$\mathrm{Ent}_\mathfrak{m}(\mu_t)\le(1-t)\mathrm{Ent}_\mathfrak{m}(\mu_0)+t\mathrm{Ent}_\mathfrak{m}(\mu_1)-\frac{K}{2}t(1-t)W_2^2(\mu_0,\mu_1)\qquad\forall t\in[0,1].$$

This definition was introduced in [?] and [?]. Its two basic features are: **compatiblity** with the Riemannian case (i.e. a compact Riemannian manifold endowed with the normalized volume measure has Ricci curvature bounded below by K in the classical pointwise sense if and only if $\mathrm{Ent}_\mathfrak{m}$ is K-geodesically convex in $(\mathscr{P}(X),W_2)$) and **stability** w.r.t. measured Gromov-Hausdorff convergence.

We also recall that Finsler geometries are included in the class of metric measure spaces with Ricci curvature bounded below. This means that if we have a smooth compact Finsler manifold (that is: a differentiable manifold endowed with a norm - possibly not coming from an inner product - on each tangent space which varies smoothly on the base point) endowed with, say, the Hausdorff measure of correct dimension, then this space has Ricci curvature bounded below by some $K\in\mathbb{R}$ (see the theorem stated at page 926 of [?] for the flat case and [?] for the general one).[11]

The goal now is to study the W_2-gradient flow of $\mathrm{Ent}_\mathfrak{m}$. Notice that the general theory of gradient flows of K-convex functionals ensures the following existence result (see the representation formula for the slope (2.6) and Theorem 2.2).

Theorem 5.2 (Consequences of the general theory of gradient flows) *Let $(X,\mathsf{d},\mathfrak{m})$ be a $CD(K,\infty)$ space. Then the slope $|\nabla^-\mathrm{Ent}_\mathfrak{m}|$ is lower semicontinuous w.r.t. weak convergence and for any $\mu\in D(\mathrm{Ent}_\mathfrak{m})$ there exists a gradient flow of $\mathrm{Ent}_\mathfrak{m}$ starting from μ.*

Thus, existence is granted. The problem is then to show uniqueness of the gradient flow. To this aim, we need to introduce the concept of *push forward via a plan.*

Definition 5.3 (Push forward via a plan) *Let $\mu\in\mathscr{P}(X)$ and let $\boldsymbol{\gamma}\in\mathscr{P}(X^2)$ be such that $\mu\ll\pi^1_\sharp\boldsymbol{\gamma}$. The measure $\boldsymbol{\gamma}_\mu\in\mathscr{P}(X^2)$ is defined as:*

$$\mathrm{d}\boldsymbol{\gamma}_\mu(x,y):=\frac{\mathrm{d}\mu}{\mathrm{d}\pi^1_\#\boldsymbol{\gamma}}(x)\mathrm{d}\boldsymbol{\gamma}(x,y),$$

and the measure $\boldsymbol{\gamma}_\sharp\mu$ as $\boldsymbol{\gamma}_\sharp\mu:=\pi^2_\#\boldsymbol{\gamma}_\mu$.

[11] Qui va controllato se la misura menzionata nel lavoro è o no quella di Hausdorff, o invece una misura di volume Finsleriana, nel caso non piatto potrebbe esserci differenza

Observe that, since $\boldsymbol{\gamma}_\mu \ll \boldsymbol{\gamma}$, we have $\boldsymbol{\gamma}_\sharp\mu \ll \pi^2_\#\boldsymbol{\gamma}$. We will say that $\boldsymbol{\gamma}$ has *bounded compression* if there exist $0 < c \leq C < \infty$ such that $c\mathfrak{m} \leq \pi^i_\sharp\boldsymbol{\gamma} \leq C\mathfrak{m}$, $i = 1, 2$. Writing $\mu = \rho\pi^1_\sharp\boldsymbol{\gamma}$, the definition gives that $\boldsymbol{\gamma}_\sharp\mu = \eta\pi^2_\sharp\boldsymbol{\gamma}$ with η given by

$$\eta(y) = \int \rho(x)\,\mathrm{d}\boldsymbol{\gamma}_y(x), \tag{5.1}$$

where $\{\boldsymbol{\gamma}_y\}_{y\in X}$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. its second marginal.

The operation of push forward via a plan has interesting properties in connection with the relative entropy functional.

Proposition 5.4 *The following properties hold:*

(i) For any $\mu, \nu \in \mathscr{P}(X)$, $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ such that $\mu, \nu \ll \pi^1_\sharp\boldsymbol{\gamma}$ it holds

$$\mathrm{Ent}_{\boldsymbol{\gamma}_\sharp\nu}(\boldsymbol{\gamma}_\sharp\mu) \leq \mathrm{Ent}_\nu(\mu).$$

(ii) For $\mu \in D(\mathrm{Ent}_\mathfrak{m})$ and $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ with bounded compression, it holds $\boldsymbol{\gamma}_\sharp\mu \in D(\mathrm{Ent}_\mathfrak{m})$.

(iii) Given $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ with bounded compression, the map

$$D(\mathrm{Ent}_\mathfrak{m}) \ni \mu \qquad \mapsto \qquad \mathrm{Ent}_\mathfrak{m}(\mu) - \mathrm{Ent}_\mathfrak{m}(\boldsymbol{\gamma}_\sharp\mu),$$

is convex (w.r.t. linear interpolation of measures).

Proof **(i)**. We can assume $\mu \ll \nu$, otherwise there is nothing to prove. Then it is immediate to check from the definition that $\boldsymbol{\gamma}_\sharp\mu \ll \boldsymbol{\gamma}_\sharp\nu$. Let $\mu = \rho\nu$, $\boldsymbol{\gamma}_\sharp\mu = \eta\boldsymbol{\gamma}_\sharp\nu$, and $u(z) := z\log z$. By disintegrating $\boldsymbol{\gamma}$ as in (5.1), we have

$$\begin{aligned}
\mathrm{Ent}_{\boldsymbol{\gamma}_\sharp\nu}(\boldsymbol{\gamma}_\sharp\mu) &= \int u(\eta(y))\,\mathrm{d}\boldsymbol{\gamma}_\sharp\nu(y) = \int u\left(\int \rho(x)\,\mathrm{d}\boldsymbol{\gamma}_y(x)\right)\mathrm{d}\boldsymbol{\gamma}_\sharp\nu(y)\\
&\leq \int\int u(\rho(x))\,\mathrm{d}\boldsymbol{\gamma}_y(x)\,\mathrm{d}\boldsymbol{\gamma}_\sharp\nu(y) = \int u(\rho(x))\frac{\mathrm{d}\nu}{\mathrm{d}\boldsymbol{\gamma}}(x)\,\mathrm{d}\boldsymbol{\gamma}(x,y)\\
&= \int u(\rho(x))\,\mathrm{d}\nu(x) = \mathrm{Ent}_\nu(\mu).
\end{aligned}$$

(ii). Taking into account the identity

$$\mathrm{Ent}_\nu(\mu) = \mathrm{Ent}_\sigma(\mu) + \int \log\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\nu}\right)\mathrm{d}\mu, \tag{5.2}$$

valid for any $\mu, \nu, \sigma \in \mathscr{P}(X)$ with σ having bounded density w.r.t. ν, the fact that $\boldsymbol{\gamma}_\sharp(\pi^1_\sharp\boldsymbol{\gamma}) = \pi^2_\sharp\boldsymbol{\gamma}$ and the fact that $c\mathfrak{m} \leq \pi^1_\#\boldsymbol{\gamma}, \pi^2_\#\boldsymbol{\gamma} \leq C\mathfrak{m}$, the conclusion follows from

$$\mathrm{Ent}_\mathfrak{m}(\boldsymbol{\gamma}_\sharp\mu) \leq \mathrm{Ent}_{\pi^2_\sharp\boldsymbol{\gamma}}(\boldsymbol{\gamma}_\sharp\mu) + \log C \leq \mathrm{Ent}_{\pi^1_\sharp\boldsymbol{\gamma}}(\mu) + \log C \leq \mathrm{Ent}_\mathfrak{m}(\mu) + \log C - \log c.$$

(iii). Let $\mu_0, \mu_1 \in D(\mathrm{Ent}_\mathfrak{m})$ and define $\mu_t := (1-t)\mu_0 + t\mu_1$ and $\nu_t := \boldsymbol{\gamma}_\sharp\mu_t$. A direct computation shows that

$$\begin{aligned}
(1-t)\mathrm{Ent}_\mathfrak{m}(\mu_0) + t\mathrm{Ent}_\mathfrak{m}(\mu_1) - \mathrm{Ent}_\mathfrak{m}(\mu_t) &= (1-t)\mathrm{Ent}_{\mu_t}(\mu_0) + t\mathrm{Ent}_{\mu_t}(\mu_1),\\
(1-t)\mathrm{Ent}_\mathfrak{m}(\nu_0) + t\mathrm{Ent}_\mathfrak{m}(\nu_1) - \mathrm{Ent}_\mathfrak{m}(\nu_t) &= (1-t)\mathrm{Ent}_{\nu_t}(\nu_0) + t\mathrm{Ent}_{\nu_t}(\nu_1),
\end{aligned}$$

and from (i) we have that

$$\mathrm{Ent}_{\mu_t}(\mu_i) \geq \mathrm{Ent}_{\boldsymbol{\gamma}_\sharp\mu_t}(\boldsymbol{\gamma}_\sharp\mu_i) = \mathrm{Ent}_{\nu_t}(\nu_i), \qquad \forall t \in [0,1],\ i = 0, 1,$$

which gives the conclusion. $\square$

In the next lemma and in the sequel we use the short notation

$$C(\boldsymbol{\gamma}) := \int_{X\times X} \mathsf{d}^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}(x,y).$$

Lemma 5.5 (Approximability in Entropy and distance) *Let $\mu,\nu \in D(\mathrm{Ent}_{\mathfrak{m}})$. Then there exists a sequence $(\boldsymbol{\gamma}^n)$ of plans with bounded compression such that $\mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}^n_\sharp\mu) \to \mathrm{Ent}_{\mathfrak{m}}(\nu)$ and $C(\boldsymbol{\gamma}^n_\mu) \to W_2^2(\mu,\nu)$ as $n\to\infty$.*

Proof Pick $\boldsymbol{\gamma} \in \mathrm{O{\scriptsize PT}}(\mu,\mu)$ and, for every $n\in\mathbb{N}$, let $A_n := \{(x,y) : f(x)+g(y)\le n\}$ and

$$\boldsymbol{\gamma}_n := c_n\left(\boldsymbol{\gamma}|_{A_n} + \frac{1}{n}(\mathrm{Id},\mathrm{Id})_\#\mathfrak{m}\right),$$

where $c_n \to 1$ is the normalization constant. It is immediate to check that $\boldsymbol{\gamma}_n$ is of bounded compression and that this sequence satisfies the thesis (see [?] for further details). □

Proposition 5.6 (Convexity of the squared slope) *Let $(X,\mathsf{d},\mathfrak{m})$ be a $CD(K,\infty)$ space. Then the map*

$$D(\mathrm{Ent}_{\mathfrak{m}}) \ni \mu \qquad \mapsto \qquad |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\mu)$$

is convex (w.r.t. linear interpolation of measures).

Notice that the only assumption that we make is the K-convexity of the entropy w.r.t. W_2, and from this we deduce the convexity w.r.t. the classical linear interpolation of measures of the squared slope.

Proof Recall that from (2.6) we know that

$$|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu) = \sup_{\substack{\nu\in\mathscr{P}_2(X)\\ \nu\neq\mu}} \frac{\left[\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\nu) - \frac{K^-}{2}W_2^2(\mu,\nu)\right]^+}{W_2(\mu,\nu)}.$$

We claim that it also holds

$$|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu) = \sup_{\boldsymbol{\gamma}} \frac{\left[\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu) - \frac{\lambda^-}{2}C(\boldsymbol{\gamma}_\mu)\right]^+}{\sqrt{C(\boldsymbol{\gamma}_\mu)}},$$

where the supremum is taken among all plans with bounded compression and the value of the right hand side is taken by definition 0 if $C(\boldsymbol{\gamma}_\mu)=0$.

Indeed, Lemma 5.5 gives that the first expression is not larger than the second. For the converse inequality we can assume $C(\boldsymbol{\gamma}_\mu)>0$ and $\nu_{\boldsymbol{\gamma},\mu}\neq\mu$. Then it is sufficient to apply the simple inequality

$$a,\, b,\, c\in\mathbb{R},\quad 0<b\le c \qquad \Rightarrow \qquad \frac{(a-b)^+}{\sqrt{b}} \ge \frac{(a-c)^+}{\sqrt{c}},$$

with $a := \mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu)$, $b := \frac{K^-}{2}W_2^2(\mu,\boldsymbol{\gamma}_\sharp\mu)$ and $c := \frac{K^-}{2}C(\boldsymbol{\gamma}_\mu)$.

Thus, to prove the thesis it is enough to show that for every $\boldsymbol{\gamma}$ with bounded compression the map

$$D(\mathrm{Ent}_{\mathfrak{m}}) \ni \mu \qquad \mapsto \qquad \frac{\left[\left(\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu) - \frac{K^-}{2}C(\boldsymbol{\gamma}_\mu)\right)^+\right]^2}{C(\boldsymbol{\gamma}_\mu)},$$

is convex w.r.t. linear interpolation of measures.

Clearly the map

$$D(\mathrm{Ent}_{\mathfrak{m}}) \ni \mu \qquad \mapsto \qquad C(\boldsymbol{\gamma}_\mu) = \int \left(\int d^2(x,y) d\boldsymbol{\gamma}_x(y) \right) d\mu(x),$$

where $\{\boldsymbol{\gamma}_x\}$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. its first marginal, is linear. Thus, from (*iii*) of Proposition 5.4 we know that the map

$$\mu \qquad \mapsto \qquad \mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu) - \frac{K^-}{2}C(\boldsymbol{\gamma}_\mu),$$

is convex w.r.t. linear interpolation of measures. Hence the same is true for its positive part. The conclusion follows from the fact that the function $\Psi : [0,\infty)^2 \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\Psi(a,b) := \begin{cases} \dfrac{a^2}{b}, & \text{if } b > 0, \\ +\infty & \text{if } b = 0, a > 0 \\ 0 & \text{if } a = b = 0, \end{cases}$$

is convex, and nondecreasing w.r.t. a. □

The convexity of the squared slope allows to prove uniqueness of the gradient flow of the entropy:

Theorem 5.7 (Uniqueness of the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$**)** *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be a* $CD(K,\infty)$ *space and let* $\mu \in D(\mathrm{Ent}_{\mathfrak{m}})$*. Then there exists a unique gradient flow of* $\mathrm{Ent}_{\mathfrak{m}}$ *starting from* μ *in* $(\mathscr{P}(X), W_2)$*.*

Proof We recall (inequality (2.3)) that the squared Wasserstein distance is convex w.r.t. linear interpolation of measures. Therefore, given two absolutely continuous curves (μ_t^1) and (μ_t^2), the curve $t \mapsto \mu_t := \frac{\mu_t^1 + \mu_t^2}{2}$ is absolutely continuous as well and its metric speed can be bounded from above by

$$|\dot{\mu}_t|^2 \leq \frac{|\dot{\mu}_t^1|^2 + |\dot{\mu}_t^2|^2}{2}, \qquad \text{for a.e. } t \in (0,\infty). \tag{5.3}$$

Let (μ_t^1) and (μ_t^2) be gradient flows of $\mathrm{Ent}_{\mathfrak{m}}$ starting from $\mu \in D(\mathrm{Ent}_{\mathfrak{m}})$. Then we have

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) = \mathrm{Ent}_{\mathfrak{m}}(\mu_T^1) + \frac{1}{2}\int_0^T |\dot{\mu}_t^1|^2 \, \mathrm{d}t + \frac{1}{2}\int_0^T |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t^1) \, \mathrm{d}t, \quad \forall T \geq 0,$$
$$\mathrm{Ent}_{\mathfrak{m}}(\mu) = \mathrm{Ent}_{\mathfrak{m}}(\mu_T^2) + \frac{1}{2}\int_0^T |\dot{\mu}_t^2|^2 \, \mathrm{d}t + \frac{1}{2}\int_0^T |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t^2) \, \mathrm{d}t, \quad \forall T \geq 0.$$

Adding up these two equalities, using the convexity of the squared slope guaranteed by Proposition 5.6, the convexity of the squared metric speed given by (5.3) and the *strict* convexity of the relative entropy, we deduce that for the curve $t \mapsto \mu_t := \frac{\mu_t^1+\mu_t^2}{2}$ it holds

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) > \mathrm{Ent}_{\mathfrak{m}}(\mu_T) + \frac{1}{2}\int_0^T |\dot{\mu}_t|^2 \, \mathrm{d}t + \frac{1}{2}\int_0^T |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t) \, \mathrm{d}t,$$

for every T such that $\mu_T^1 \neq \mu_T^2$. This contradicts inequality (2.8). □

6 The heat flow as gradient flow

It is well known that on $\mathbb{R}^d$ the heat flow can be seen both as gradient flow of the Dirichlet energy in L^2 and as gradient flow of the relative entropy in $(\mathscr{P}_2(\mathbb{R}^d), W_2)$. It is therefore natural to ask whether this identification between the two a priori different gradient flows persists or not. The answer is yes, but the proof follows quite different lines.

The strategy consists in considering a gradient flow (f_t) of Ch with nonnegative initial data and in proving that the curve $t \mapsto \mu_t := f_t \mathfrak{m}$ is a gradient flow of $\mathrm{Ent}_{\mathfrak{m}}(\cdot)$ in $(\mathscr{P}(X), W_2)$: by the uniqueness result of Theorem 5.7 this will be sufficient to conclude.

We already built most of the ingredients needed for the proof to work, the only thing that we should add is the following lemma, where the slope of $\mathrm{Ent}_{\mathfrak{m}}$ is bounded from above in terms of the notions of "norm of weak gradient" that we discussed in Chapter 4. It is worth noticing that we make essential use of the fact that Lipschitz functions are dense in energy, as a direct use of upper gradient seems not to lead to the same estimate.[12]

Lemma 6.1 (Fisher bounds slope) *Let ρ be a probability density such that $0 < c \leq \rho \leq C$ for some c, $C \in \mathbb{R}$. Then*

$$|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\rho\mathfrak{m}) \leq \int \frac{|\nabla \rho|_w^2}{\rho} \, \mathrm{d}\mathfrak{m}.$$

Proof Assume at first that ρ is Lipschitz, and let (ρ_n) be a sequence of probability densities such that $W_2(\rho_n\mathfrak{m}, \rho\mathfrak{m}) \to 0$ and where the slope of $\mathrm{Ent}_{\mathfrak{m}}$ at $\rho\mathfrak{m}$ is attained. Choose $\boldsymbol{\gamma}_n \in \mathrm{O{\scriptsize PT}}(\rho\mathfrak{m}, \rho_n\mathfrak{m})$ and notice that

$$\begin{aligned} \int \rho \log \rho \, \mathrm{d}\mathfrak{m} - \int \rho_n \log \rho_n \, \mathrm{d}\mathfrak{m} &\leq \int (\rho - \rho_n) \log \rho \, \mathrm{d}\mathfrak{m} \\ &= \int \log \rho(x) - \log \rho(y) \, \mathrm{d}\boldsymbol{\gamma}_n(x,y) \\ &\leq \sqrt{\int \frac{\big(\log \rho(x) - \log \rho(y)\big)^2}{\mathsf{d}^2(x,y)} \, \mathrm{d}\boldsymbol{\gamma}_n(x,y)} \sqrt{\int \mathsf{d}^2(x,y) \, \mathrm{d}\boldsymbol{\gamma}_n(x,y)}. \end{aligned} \tag{6.1}$$

Now consider the bounded upper semicontinuous function

$$L(x,y) := \begin{cases} \dfrac{\big(\log \rho(x) - \log \rho(y)\big)^2}{\mathsf{d}^2(x,y)}, & \text{if } x \neq y, \\ |\nabla \log \rho|(x) = \dfrac{|\nabla \rho|(x)}{\rho(x)} & \text{if } x = y, \end{cases}$$

[12] In relazione a questa frase, avevo rilevato che sembra contraddire quanto riusciamo a fare proprio nel primo lavoro, i.e. stimare l'oscillazione dell'entropia usando i weak gradient. Ma questo lo facciamo solo lungo i gradient flow e non puntualmente. Forse possiamo lasciare la frase come l'ha scritta Nicola, o aggiungere qualcosa

and notice that, since $\boldsymbol{\gamma}_n$ weakly converge to $(\mathrm{Id},\mathrm{Id})_\#\rho\mathfrak{m}$, it holds

$$\overline{\lim_{n\to\infty}} \int L(x,y)\,\mathrm{d}\boldsymbol{\gamma}_n(x,y) \le \int L(x,x)\rho(x)\,\mathrm{d}\mathfrak{m}(x) = \int \frac{|\nabla\rho|^2}{\rho}\,\mathrm{d}\mathfrak{m}.$$

Hence (6.1) gives

$$|\nabla^-\mathrm{Ent}_\mathfrak{m}|(\rho\mathfrak{m}) = \overline{\lim_{n\to\infty}} \frac{(\mathrm{Ent}_\mathfrak{m}(\rho\mathfrak{m}) - \mathrm{Ent}_\mathfrak{m}(\rho_n\mathfrak{m}))^+}{W_2(\rho\mathfrak{m},\rho_n\mathfrak{m})} \le \sqrt{\int \frac{|\nabla\rho|^2}{\rho}\,\mathrm{d}\mathfrak{m}}. \tag{6.2}$$

We now turn to the general case. Let $0 < c \le \rho \le C$ be any bounded probability density. If $\rho \notin D(\mathrm{Ch})$ there is nothing to prove. Otherwise, use Theorem 4.23 to find a sequence of Lipschitz functions (ρ_n) converging to ρ in $L^2(X,\mathfrak{m})$ and such that $|\nabla\rho_n| \to |\nabla\rho|_w$ in $L^2(X,\mathfrak{m})$ and $\mathfrak{m}$-a.e.. Up to a truncation and rescaling argument, we can assume that $0 < c' \le \rho_n \le C' < \infty$ $\mathfrak{m}$-a.e. and $\int \rho_n\,\mathrm{d}\mathfrak{m} = 1$, for any $n \in \mathbb{N}$. The conclusion follows passing to the limit in (6.2) by taking into account the weak lower semicontinuity of $|\nabla^-\mathrm{Ent}_\mathfrak{m}|$ (formula (2.6) and discussion thereafter). □

Theorem 6.2 (The heat flow as gradient flow) *Let $f_0 \in L^2(X,\mathfrak{m})$ be such that $\mu_0 = f_0\mathfrak{m} \in \mathscr{P}(X)$ and denote by (f_t) the gradient flow of* Ch *in $L^2(X,\mathfrak{m})$ starting from f_0 and by (μ_t) the gradient flow of* $\mathrm{Ent}_\mathfrak{m}$ *in $(\mathscr{P}(X),W_2)$ starting from μ_0. Then $\mu_t = f_t\mathfrak{m}$ for any $t \ge 0$.*

Proof Thanks to the uniqueness result of Theorem 5.7, it is sufficient to prove that $(f_t\mathfrak{m})$ satisfies the Energy Dissipation Equality for $\mathrm{Ent}_\mathfrak{m}$ in $(\mathscr{P}(X),W_2)$. We assume first that $0 < c \le f_0 \le C < \infty$ $\mathfrak{m}$-a.e. in X, so that the maximum principle (Proposition 4.9) ensures $0 < c \le f_t \le C < \infty$ for any $t > 0$. By Proposition 4.9 we know that $t \mapsto \mathrm{Ent}_\mathfrak{m}(f_t\mathfrak{m})$ is absolutely continuous with derivative equal to $-\int \frac{|\nabla f_t|_w^2}{f_t}\,\mathrm{d}\mathfrak{m}$. Lemma 4.20 ensures that $t \mapsto f_t\mathfrak{m}$ is absolutely continuous w.r.t. W_2 with squared metric speed bounded by $\int \frac{|\nabla f_t|_w^2}{f_t}\,\mathrm{d}\mathfrak{m}$, so that taking into account Lemma 6.1 we get

$$\mathrm{Ent}_\mathfrak{m}(f_0\mathfrak{m}) \ge \mathrm{Ent}_\mathfrak{m}(f_t\mathfrak{m}) + \frac{1}{2}\int_0^t |\dot{f_s\mathfrak{m}}|^2\,\mathrm{d}s + \frac{1}{2}\int_0^t |\nabla^-\mathrm{Ent}_\mathfrak{m}|^2(f_s\mathfrak{m})\,\mathrm{d}s,$$

which, together with (2.8), ensures the thesis.

For the general case we argue by approximation, considering $f_0^n := c_n \min\{n, \max\{f_0, 1/n\}\}$, c_n being the normalizing constant, and the corresponding gradient flow (f_t^n) of Ch. The fact that $f_0^n \to f_0$ in $L^2(X,\mathfrak{m})$ and the convexity of Ch implies that $f_t^n \to f_t$ in $L^2(X,\mathfrak{m})$ for any $t > 0$. In particular, the $W_2(f_t^n\mathfrak{m}, f_t\mathfrak{m}) \to 0$ as $n \to \infty$ for every t (because convergence w.r.t. W_2 is equivalent to weak convergence of measures).

Now notice that we know that

$$\mathrm{Ent}_\mathfrak{m}(f_0^n\mathfrak{m}) = \mathrm{Ent}_\mathfrak{m}(f_t^n) + \frac{1}{2}\int_0^t |\dot{f_s^n\mathfrak{m}}|^2\,\mathrm{d}s + \frac{1}{2}\int_0^t |\nabla^-\mathrm{Ent}_\mathfrak{m}|^2(f_s^n)\,\mathrm{d}s, \qquad \forall t > 0.$$

Furthermore, it is immediate to check that $\mathrm{Ent}_\mathfrak{m}(f_0^n\mathfrak{m}) \to \mathrm{Ent}_\mathfrak{m}(f_0\mathfrak{m})$ as $n \to \infty$. The pointwise convergence of $f_t^n\mathfrak{m}$ to $f_t\mathfrak{m}$ w.r.t. W_2 easily yields that the terms on the right hand side of the last equation are lower semicontinuous when $n \to \infty$ (recall Theorem 5.2 for the slope). Thus it holds

$$\mathrm{Ent}_{\mathfrak{m}}(f_0\mathfrak{m}) \geq \mathrm{Ent}_{\mathfrak{m}}(f_t) + \frac{1}{2}\int_0^t |\dot{f_s\mathfrak{m}}|^2\,\mathrm{d}s + \frac{1}{2}\int_0^t |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(f_s)\,\mathrm{d}s, \qquad \forall t > 0,$$

which, by (2.10), is the thesis.
We know, by Theorem 5.7, that there is at most a gradient flow starting from μ_0. We also know that a gradient flow f'_t of Ch starting from f_0 exists, and part (i) gives that $\mu'_t := f'_t\mathfrak{m}$ is a gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$. The uniqueness of gradient flows gives $\mu_t = \mu'_t$ for all $t \geq 0$. □

7 A metric Brenier theorem

In this section we state and prove the metric Brenier theorem in $CD(L,\infty)$ spaces we announced in the introduction. It was recently proven in [][13] that under an additional non-branching assumption one can really recover an optimal transport map, see also [][14] for related results, obtained under stronger non-branching assumptions and weaker convexity assumptions.

Definition 7.1 (Strong $CD(K,\infty)$ spaces) *We say that a compact normalized metric measure space $(X,\mathsf{d},\mathfrak{m})$ is a strong $CD(K,\infty)$ space if for any μ_0, $\mu_1 \in \mathscr{P}(X)$ there exists $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0,\mu_1)$ with the following property. For any bounded Borel function $F : \mathrm{Geo(X)} \to [0,\infty)$ such that $\int F\,\mathrm{d}\boldsymbol{\pi} = 1$, it holds*

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t^F) \leq (1-t)\mathrm{Ent}_{\mathfrak{m}}(\mu_0^F) + t\mathrm{Ent}_{\mathfrak{m}}(\mu_1^F) - \frac{K}{2}t(1-t)W_2^2(\mu_0^F,\mu_1^F),$$

where $\mu_t^F := (\mathrm{e}_t)_\sharp(F\boldsymbol{\pi})$, for any $t \in [0,1]$.

Thus, the difference between strong $CD(K,\infty)$ spaces and standard $CD(K,\infty)$ ones is the fact that geodesic convexity is required along *all* geodesics induced by the weighted plans $F\boldsymbol{\pi}$, rather than the one induced by $\boldsymbol{\pi}$ only. Notice that the necessary and sufficient optimality conditions ensure that $\boldsymbol{\pi}$ is concentrated on a c-monotone set, hence $F\boldsymbol{\pi}$ has the same property and it is optimal, relative to its marginals.

It is not clear to us whether the notion of being strong $CD(K,\infty)$ is stable or not w.r.t. measured Gromov-Hausdorff convergence and, as such, it should be handled with care. The importance of strong $CD(K,\infty)$ bounds relies on the fact that on these spaces geodesic interpolation between bounded probability densities is made of bounded densities as well, thus granting the existence of many test plans.

Notice that non-branching $CD(K,\infty)$ spaces are always strong $CD(K,\infty)$ spaces, indeed let μ_0, $\mu_1 \in D(\mathrm{Ent}_{\mathfrak{m}})$ and pick $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0,\mu_1)$ such that $\mathrm{Ent}_{\mathfrak{m}}$ is K-convex along $((\mathrm{e}_t)_\sharp\boldsymbol{\pi})$. From the non-branching hypothesis it follows that for F as in Definition 7.1 there exists a unique element in $\mathrm{GeoOpt}(\mu_t^F,\mu_1^F)$ (resp. in $\mathrm{GeoOpt}(\mu_t^F,\mu_0^F)$). Also, since F is bounded, from $\mu_t \in D(\mathrm{Ent}_{\mathfrak{m}})$ we deduce $\mu_t^F \in D(\mathrm{Ent}_{\mathfrak{m}})$. Hence the map $t \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu_t^F)$ is K-convex and bounded on $[\varepsilon,1]$ and on $[0,1-\varepsilon]$ for all $\varepsilon \in (0,1)$, and therefore it is K-convex on $[0,1]$.

[13] Citare Gigli GAFA
[14] Citare Ambrosio-Rajala

Proposition 7.2 (Bound on geodesic interpolant) *Let $(X, \mathsf{d}, \mathfrak{m})$ be a strong $CD(K,\infty)$ space and let $\mu_0, \mu_1 \in \mathscr{P}(X)$ be with bounded densities. Then there exists a geodesic (μ_t) connecting them made of measures with uniformly bounded densities.*

Proof Let M be an upper bound on the densities of μ_0, μ_1, $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0, \mu_1)$ be a plan which satisfies the assumptions of Definition 7.1 and $\mu_t := (\mathrm{e}_t)_\sharp \boldsymbol{\pi}$. We claim that the measures μ_t have uniformly bounded densities. The fact that $\mu_t \ll \mathfrak{m}$ is obvious by geodesic convexity, so let ρ_t be the density of μ_t and assume by contradiction that for some $t_0 \in [0,1]$ it holds

$$\rho_{t_0}(x) > M e^{K^- \mathrm{D}^2/8}, \qquad \forall x \in A, \tag{7.1}$$

where $\mathfrak{m}(A) > 0$ and D is the diameter of X. Define $\tilde{\boldsymbol{\pi}} := c\boldsymbol{\pi}|_{\mathrm{e}_{t_0}^{-1}(A)}$, where c is the normalizing constant (notice that $\tilde{\boldsymbol{\pi}}$ is well defined, because $\boldsymbol{\pi}(\mathrm{e}_{t_0}^{-1}(A)) = \mu_{t_0}(A) > 0$) and observe that the density of $\tilde{\boldsymbol{\pi}}$ w.r.t. $\boldsymbol{\pi}$ is bounded. Let $\tilde{\mu}_t := (\mathrm{e}_t)_\sharp \tilde{\boldsymbol{\pi}}$ and $\tilde{\rho}_t$ its density w.r.t. $\mathfrak{m}$. From (7.1) we get $\tilde{\rho}_{t_0} = c\rho_{t_0}$ on A and $\tilde{\rho}_{t_0} = 0$ on $X \setminus A$, hence

$$\mathrm{Ent}_{\mathfrak{m}}(\tilde{\mu}_{t_0}) = \int \log(\tilde{\rho}_{t_0} \circ \mathrm{e}_{t_0})\, \mathrm{d}\boldsymbol{\pi} > \log c + \log M + \frac{K^-}{8}\mathrm{D}^2. \tag{7.2}$$

On the other hand, we have $\tilde{\rho}_0 \le c\rho_0 \le cM$ and $\tilde{\rho}_1 \le c\rho_1 \le cM$ and thus

$$\mathrm{Ent}_{\mathfrak{m}}(\tilde{\mu}_i) = \int \log(\tilde{\rho}_i \circ \mathrm{e}_i)\, \mathrm{d}\tilde{\boldsymbol{\pi}} \le \log c + \log M, \qquad i = 0, 1. \tag{7.3}$$

Finally, it certainly holds $W_2^2(\tilde{\mu}_0, \tilde{\mu}_1) \le \mathrm{D}^2$, so that (7.2) and (7.3) contradict the K-convexity of $\mathrm{Ent}_{\mathfrak{m}}$ along $(\tilde{\mu}_t)$. Hence (7.1) is false and the ρ_t's are uniformly bounded. □

An important consequence of this uniform bound is the following metric version of Brenier's theorem.
[15]

Theorem 7.3 (A metric Brenier theorem) *Let $(X, \mathsf{d}, \mathfrak{m})$ be a strong $CD(K,\infty)$ space, ρ_0, ρ_1 probability densities and φ any Kantorovich potential for the couple $(\rho_0\mathfrak{m}, \rho_1\mathfrak{m})$. Then for every $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\rho_0\mathfrak{m}, \rho_1\mathfrak{m})$ it holds*

$$\mathsf{d}(\gamma_0, \gamma_1) = |\nabla\varphi|_w(\gamma_0) = |\nabla^+\varphi|(\gamma_0), \qquad \text{for } \boldsymbol{\pi}\text{-a.e. } \gamma.$$

In particular,

$$W_2^2(\rho_0\mathfrak{m}, \rho_1\mathfrak{m}) = \int_X |\nabla\varphi|_*^2 \rho_0 \, \mathrm{d}\mathfrak{m}.$$

Proof φ is Lipschitz, therefore $|\nabla^+\varphi|$ is an upper gradient of φ, and hence $|\nabla\varphi|_w \le |\nabla^+\varphi|$ $\mathfrak{m}$-a.e.. Now fix $x \in X$ and pick any $y \in \partial^c\varphi(x)$. From the c-concavity of φ we get

$$\varphi(x) = \frac{\mathsf{d}^2(x,y)}{2} - \varphi^c(y),$$
$$\varphi(z) \le \frac{\mathsf{d}^2(z,y)}{2} - \varphi^c(y) \qquad \forall z \in X.$$

Therefore

$$\varphi(z) - \varphi(x) \le \frac{\mathsf{d}^2(z,y)}{2} - \frac{\mathsf{d}^2(x,y)}{2} \le \mathsf{d}(z,x)\frac{\mathsf{d}(z,y) + \mathsf{d}(x,y)}{2}.$$

[15] Qui mi sembra che vada aggiunta anche la convergenza $L^2(\boldsymbol{\pi})$ dei rapporti incrementali

Dividing by $\mathsf{d}(x,z)$ and letting $z \to x$, by the arbitrariness of $y \in \partial^c\varphi(x)$ and the fact that $\mathrm{supp}((\mathrm{e}_0,\mathrm{e}_1)_\sharp\boldsymbol{\pi}) \subset \partial^c\varphi$ we get

$$|\nabla^+\varphi|(\gamma_0) \le \min_{y\in\partial^c\varphi(\gamma_0)} \mathsf{d}(\gamma_0,y) \le \mathsf{d}(\gamma_0,\gamma_1) \qquad \text{for } \boldsymbol{\pi}\text{-a.e. } \gamma.$$

Since

$$\int |\nabla\varphi|_w^2\rho_0\,\mathrm{d}\mathfrak{m} \le \int |\nabla^+\varphi|^2(\gamma_0)\,\mathrm{d}\boldsymbol{\pi} \quad \text{and} \quad \int \mathsf{d}^2(\gamma_0,\gamma_1)\,\mathrm{d}\boldsymbol{\pi}(\gamma) = W_2^2(\rho_0\mathfrak{m},\rho_1\mathfrak{m}),$$

to conclude it is sufficient to prove that

$$W_2^2(\rho_0\mathfrak{m},\rho_1\mathfrak{m}) \le \int |\nabla\varphi|_w^2\rho_0\,\mathrm{d}\mathfrak{m}. \tag{7.4}$$

Now assume that ρ_0 and ρ_1 are bounded from above and let $\tilde{\boldsymbol{\pi}} \in \mathrm{GeoOpt}(\rho_0\mathfrak{m},\rho_1\mathfrak{m})$ be a test plan (such $\tilde{\boldsymbol{\pi}}$ exists thanks to Proposition 7.2). Since φ is a Kantorovich potential and $(\mathrm{e}_0,\mathrm{e}_1)_\sharp\tilde{\boldsymbol{\pi}}$ is optimal, it holds $\gamma_1 \in \partial^c\varphi(\gamma_0)$ for any $\gamma \in \mathrm{supp}(\tilde{\boldsymbol{\pi}})$. Hence arguing as before we get

$$\varphi(\gamma_0) - \varphi(\gamma_t) \ge \frac{\mathsf{d}^2(\gamma_0,\gamma_1)}{2} - \frac{\mathsf{d}^2(\gamma_t,\gamma_1)}{2} = \mathsf{d}^2(\gamma_0,\gamma_1)\big(t - t^2/2\big).$$

Dividing by $\mathsf{d}(\gamma_0,\gamma_t) = t\mathsf{d}(\gamma_0,\gamma_1)$, squaring and integrating w.r.t. $\tilde{\boldsymbol{\pi}}$ we obtain

$$\varliminf_{t\downarrow 0} \int \left(\frac{\varphi(\gamma_0)-\varphi(\gamma_t)}{\mathsf{d}(\gamma_0,\gamma_t)}\right)^2 \mathrm{d}\tilde{\boldsymbol{\pi}}(\gamma) \ge \int \mathsf{d}^2(\gamma_0,\gamma_1)\,\mathrm{d}\tilde{\boldsymbol{\pi}}(\gamma) = W_2^2(\rho_0\mathfrak{m},\rho_1\mathfrak{m}). \tag{7.5}$$

Using Remark 4.12 and the fact that $\tilde{\boldsymbol{\pi}}$ is a test plan we have

$$\begin{aligned}\int \left(\frac{\varphi(\gamma_0)-\varphi(\gamma_t)}{\mathsf{d}(\gamma_0,\gamma_t)}\right)^2 \mathrm{d}\tilde{\boldsymbol{\pi}}(\gamma) &\le \int \frac{1}{t^2}\left(\int_0^t |\nabla\varphi|_w(\gamma_s)\,\mathrm{d}s\right)^2 \mathrm{d}\tilde{\boldsymbol{\pi}}(\gamma) \le \frac{1}{t}\iint_0^t |\nabla\varphi|_w^2(\gamma_s)\,\mathrm{d}s\,\mathrm{d}\tilde{\boldsymbol{\pi}}(\gamma)\\ &= \frac{1}{t}\iint_0^t |\nabla\varphi|_w^2\,\mathrm{d}s\,\mathrm{d}(\mathrm{e}_t)_\sharp\tilde{\boldsymbol{\pi}} = \frac{1}{t}\iint_0^t |\nabla\varphi|_w^2\rho_s\,\mathrm{d}s\,\mathrm{d}\mathfrak{m},\end{aligned} \tag{7.6}$$

where ρ_s is the density of $(\mathrm{e}_s)_\sharp\tilde{\boldsymbol{\pi}}$. Since $(\mathrm{e}_t)_\sharp\tilde{\boldsymbol{\pi}}$ weakly converges to $(\mathrm{e}_0)_\sharp\tilde{\boldsymbol{\pi}}$ as $t \downarrow 0$ and $\mathrm{Ent}_\mathfrak{m}((\mathrm{e}_t)_\sharp\tilde{\boldsymbol{\pi}})$ is uniformly bounded (by the K-geodesic convexity), we conclude that $\rho_t \to \rho_0$ weakly in $L^1(X,\mathfrak{m})$ and since $|\nabla\varphi|_w \in L^\infty(X,\mathfrak{m})$ we have

$$\lim_{t\downarrow 0}\frac{1}{t}\iint_0^t |\nabla\varphi|_w^2\rho_s\,\mathrm{d}s\,\mathrm{d}\mathfrak{m} = \int |\nabla\varphi|_w^2\rho_0\,\mathrm{d}\mathfrak{m}. \tag{7.7}$$

Equations (7.5), (7.6) and (7.7) yield (7.4).

Thus the thesis is proved for the case $\rho_0,\rho_1 \in L^\infty(X,\mathfrak{m})$. For the general case, fix a Kantorovich potential φ, $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\rho_0\mathfrak{m},\rho_1\mathfrak{m})$ and for $n \in \mathbb{N}$ define $\boldsymbol{\pi}^n := c_n\boldsymbol{\pi}|_{\{\gamma:\rho_0(\gamma_0)+\rho_1(\gamma_1)\le n\}}$, $c_n \to 1$ being the normalization constant. Then $\boldsymbol{\pi}^n \in \mathrm{GeoOpt}(\rho_0^n\mathfrak{m},\rho_1^n\mathfrak{m})$, where $\rho_i^n := (\mathrm{e}_i)_\sharp\boldsymbol{\pi}^n$, φ is a Kantorovich potential for $(\rho_0^n\mathfrak{m},\rho_1^n\mathfrak{m})$ and $\rho_0^n,\rho_1^n \in L^\infty(X,\mathfrak{m})$. Thus from what we just proved we know that it holds

$$\mathsf{d}(\gamma_0,\gamma_1) = |\nabla\varphi|_w(\gamma_0) = |\nabla^+\varphi|(\gamma_0), \qquad \text{for } \boldsymbol{\pi}^n\text{-a.e. } \gamma.$$

Letting $n \to \infty$ we conclude. □

8 More on calculus on $CD(K,\infty)$ spaces

8.1 On horizontal and vertical derivatives again

Aim of this subsection is to prove another deep relation between "horizontal" and "vertical" derivation, which will allow to compare the derivative of the squared Wasserstein distance along the heat flow with the derivative of the relative entropy along a geodesic (see the next subsection). This will be key in order to understand the properties of spaces with *Riemannian* Ricci curvature bounded from below, illustrated in the last section.

In order to understand the geometric point, consider the following simple example.

Example 8.1 Let $\|\cdot\|$ be a smooth, strictly convex norm on $\mathbb{R}^d$ and let $\|\cdot\|_*$ be the dual norm. Denoting by $\langle\cdot,\cdot\rangle$ the canonical duality from $(\mathbb{R}^d)^*\times\mathbb{R}^d$ into $\mathbb{R}$, let $\mathcal{L}$ be the duality map from $(\mathbb{R}^d,\|\cdot\|)$ to $((\mathbb{R}^d)^*,\|\cdot\|_*)$, characterized by

$$\langle\mathcal{L}(u),u\rangle=\|\mathcal{L}(u)\|_*\|u\| \quad\text{and}\quad \|\mathcal{L}(u)\|_*=\|u\| \qquad \forall u\in\mathbb{R}^d$$

and let $\mathcal{L}^*$ be its inverse, equally characterized by

$$\langle v,\mathcal{L}^*(v)\rangle=\|v\|_*\|\mathcal{L}^*(v)\| \quad\text{and}\quad \|\mathcal{L}^*(v)\|=\|v\|_* \qquad \forall v\in(\mathbb{R}^d)^*.$$

Using the fact that $\epsilon\mapsto\|u\|\|u+\epsilon u'\|-\langle\mathcal{L}u,u+\epsilon u'\rangle$ attains its minimum at $\epsilon=0$ and the analogous relation for $\mathcal{L}^*$ one obtains the useful relations

$$\langle\mathcal{L}(u),u'\rangle=\frac{1}{2}d_u\|\cdot\|^2(u'), \qquad \langle v',\mathcal{L}^*(v)\rangle=\frac{1}{2}d_v\|\cdot\|_*^2(v'). \tag{8.1}$$

For a smooth map $f:\mathbb{R}^d\to\mathbb{R}$ its differential d_xf at any point x is intrinsically defined as cotangent vector, namely as an element of $(\mathbb{R}^d)^*$. To define the gradient $\nabla f(x)\in\mathbb{R}^d$ (which is a tangent vector), the norm comes into play via the formula $\nabla f(x):=\mathcal{L}^*(d_xf)$. Now, given two smooth functions f, g, the real number $d_xf(\nabla g(x))$ is well defined as the application of the cotangent vector d_xf to the tangent vector $\nabla g(x)$. What we want to point out, is that there are two very different ways of obtaining $d_xf(\nabla g(x))$ from a derivation. The first one is the "vertical derivative":

$$Df(\nabla g)(x)=\lim_{\varepsilon\to 0}\frac{\frac{1}{2}\|\nabla(g+\varepsilon f)\|^2(x)-\frac{1}{2}\|\nabla g\|^2(x)}{\varepsilon}. \tag{8.2}$$

The second one, which is usually taken as the definition of $d_xf(\nabla g(x))$, is the "horizontal derivative":

$$\langle d_xf,\nabla g(x)\rangle=\lim_{t\to 0}\frac{f(x+t\nabla g(x))-f(x)}{t}. \tag{8.3}$$

It is not difficult to check that (8.2) is consistent with (8.3): indeed (omitting the x dependence) we have

$$\begin{aligned}\|\nabla(g+\varepsilon f)\|^2 &= \langle dg+\varepsilon df,\mathcal{L}^*(dg+\varepsilon df)\rangle=\langle dg+\varepsilon df,\mathcal{L}^*(dg)+\varepsilon d_{dg}\mathcal{L}^*(df)+o(\varepsilon)\rangle\\ &= \|\nabla g\|^2+\varepsilon\langle df,\nabla g\rangle+\varepsilon\langle dg,d_{dg}\mathcal{L}^*(df)\rangle.\end{aligned}$$

So, to conclude we need only to show that

$$\langle df,\nabla g\rangle=\langle dg,d_{dg}\mathcal{L}^*(df)\rangle.$$

This can be obtained by a variant of the argument leading to (8.1), namely observing that the function

$$\|dg\|_*\|dg+\epsilon df\|_* - \langle dg, \mathcal{L}^*(dg+\epsilon df)\rangle$$

attains its minimum at $\epsilon = 0$ and using the second identity in (8.1) with $v = dg$ and $v' = df$ to see that the derivative of the left hand side is $\langle df, \nabla g\rangle$.[16] ■

The point is that the equality between the right hand sides of formulas (8.2) and (8.3) extends to a genuine metric setting. In the following lemma (where the plan $\boldsymbol{\pi}$ plays the role of $-\nabla g$) we prove one inequality, but we remark that "playing with signs" it is possible to obtain an analogous inequality with $\leq$ in place of $\geq$.

Lemma 8.2 (Horizontal and vertical derivatives) *Let f be Borel function on X such that $|\nabla f|_w \in L^2(X, \mathfrak{m})$, $g : X \to \mathbb{R}$ Lipschitz and $\boldsymbol{\pi}$ a test plan concentrated on* Geo(X)*. Assume that* [17]

$$|\nabla g|_w(\gamma_0) = \lim_{t\downarrow 0} \frac{g(\gamma_0) - g(\gamma_t)}{\mathsf{d}(\gamma_0, \gamma_t)} = \mathsf{d}(\gamma_0, \gamma_1) \qquad \textit{for } \boldsymbol{\pi}\textit{-a.e. } \gamma. \tag{8.4}$$

Then

$$\varliminf_{t\downarrow 0} \int \frac{f(\gamma_t) - f(\gamma_0)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma) \geq \frac{1}{2}\int \frac{|\nabla g|_w^2(\gamma_0) - |\nabla(g+\varepsilon f)|_w^2(\gamma_0)}{\varepsilon}\,\mathrm{d}\boldsymbol{\pi}(\gamma) \qquad \forall \varepsilon > 0. \tag{8.5}$$

Proof Define the functions F_t, G_t : Geo(X) $\to \mathbb{R} \cup \{\pm\infty\}$ by

$$F_t(\gamma) := \frac{f(\gamma_0) - f(\gamma_t)}{\mathsf{d}(\gamma_0, \gamma_t)},$$
$$G_t(\gamma) := \frac{g(\gamma_0) - g(\gamma_t)}{\mathsf{d}(\gamma_0, \gamma_t)}.$$

By (8.4) the functions G_t converge to $|\nabla g|_w \circ \mathrm{e}_0$ for $\boldsymbol{\pi}$-a.e. γ. Since clearly $G_t(\gamma) \leq \mathrm{Lip}(g)$ for any γ, the dominated convergence theorem ensures that $G_t \to |\nabla g|_w \circ \mathrm{e}_0$ as $t \downarrow 0$ in $L^2(\mathrm{Geo}(\mathrm{X}), \boldsymbol{\pi})$. In particular, it holds

$$\int |\nabla g|_w^2 \circ \mathrm{e}_0 \,\mathrm{d}\boldsymbol{\pi}(\gamma) = \lim_{t\downarrow 0} \int G_t^2 \,\mathrm{d}\boldsymbol{\pi}. \tag{8.6}$$

Since $(\mathrm{e}_t)_\sharp\boldsymbol{\pi} \to (\mathrm{e}_0)_\sharp\boldsymbol{\pi}$ weakly in $C_b(X)$ as $t \downarrow 0$ and their densities are uniformly bounded, we obtain that the densities are w^* convergent $L^\infty(X, \mathfrak{m})$. Therefore, using the fact that $|\nabla(g+\varepsilon f)|_w^2 \in L^1(X, \mathfrak{m})$ and taking into account Remark 4.12 we obtain

$$\begin{aligned}
\int |\nabla(g+\varepsilon f)|_w^2 \circ \mathrm{e}_0 \,\mathrm{d}\boldsymbol{\pi}(\gamma) &= \int |\nabla(g+\varepsilon f)|_w^2 \,\mathrm{d}(\mathrm{e}_0)_\sharp\boldsymbol{\pi} = \lim_{t\downarrow 0} \frac{1}{t}\int_0^t \int_X |\nabla(g+\varepsilon f)|_w^2 \,\mathrm{d}(\mathrm{e}_s)_\sharp\boldsymbol{\pi}\,\mathrm{d}s \\
&= \lim_{t\downarrow 0} \frac{1}{t} \iint_0^t |\nabla(g+\varepsilon f)|_w^2(\gamma_s)\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}(\gamma) \geq \varlimsup_{t\downarrow 0} \int \left|\frac{(g+\varepsilon f)(\gamma_0) - (g+\varepsilon f)(\gamma_t)}{t\mathsf{d}(\gamma_0, \gamma_1)}\right|^2 \mathrm{d}\boldsymbol{\pi}(\gamma) \\
&\geq \varlimsup_{t\downarrow 0} \int G_t^2 + 2\varepsilon G_t F_t \,\mathrm{d}\boldsymbol{\pi}.
\end{aligned}$$

[16] Come ormai abbiamo imparato, le dimostrazioni metriche sono meglio anche nei contesti lisci, comunque una differenziale ci voleva....

[17] Cosi' l'ipotesi non va bene, nel senso che e' troppo forte, solo nel lavoro con Tapio riusciamo a ottenerla. Metterei uguaglianza dei gradienti quasi ovunque piu' convergenza L^2 dei rapporti incrementali. Altrimenti sottosuccessioni di t.

Subtracting this inequality from (8.6) and dividing by 2ε we get

$$\frac{1}{2}\int \frac{|\nabla g|_w^2(\gamma_0) - |\nabla(g+\varepsilon f)|_w^2(\gamma_0)}{\varepsilon}\,\mathrm{d}\boldsymbol{\pi}(\gamma) \leq \varliminf_{t\downarrow 0} - \int G_t(\gamma)F_t(\gamma)\,\mathrm{d}\boldsymbol{\pi}(\gamma).$$

We know that $G_t \to |\nabla g|_w \circ \mathrm{e}_0$ in $L^2(\mathrm{Geo(X)}, \boldsymbol{\pi})$ and that $|\nabla g|_w(\gamma_0) = \mathsf{d}(\gamma_0, \gamma_1)$ for $\boldsymbol{\pi}$-a.e. γ. Also, by Remark 4.12 and the fact that $\boldsymbol{\pi}$ is a test plan we easily get $\sup_{t\in[0,1]} \|F_t\|_{L^2(\boldsymbol{\pi})} < \infty$. Thus it holds

$$\varliminf_{t\downarrow 0} - \int G_t(\gamma)F_t(\gamma)\,\mathrm{d}\boldsymbol{\pi}(\gamma) = \varliminf_{t\downarrow 0} - \int \mathsf{d}(\gamma_0,\gamma_1)F_t(\gamma)\,\mathrm{d}\boldsymbol{\pi}(\gamma) = \varliminf_{t\downarrow 0}\int \frac{f(\gamma_t) - f(\gamma_0)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma),$$

which is the thesis. □

8.2 Two important formulas

[18]

Proposition 8.3 (Derivative of $\frac{1}{2}W_2^2$ along the heat flow) *Let $(\rho_t) \subset L^2(X, \mathfrak{m})$ be a heat flow made of probability densities. Then for every $\boldsymbol{\sigma} \in \mathscr{P}(X)$, for a.e. $t \in (0,\infty)$ it holds:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t\mathfrak{m}, \boldsymbol{\sigma}) = \int \varphi_t \Delta\rho_t\,\mathrm{d}\mathfrak{m}, \qquad \textit{for any Kantorovich potential } \varphi \textit{ from } \rho_t \textit{ to } \boldsymbol{\sigma}. \tag{8.7}$$

Proof Since $t \mapsto \rho_t\mathfrak{m}$ is an absolutely continuous curve w.r.t. W_2 (recall Theorem 6.2), the derivative at the left hand side of (8.7) exists for a.e. $t \in (0,\infty)$. Also, for a.e. $t \in (0,\infty)$ it holds $\lim_{h\to 0}\frac{1}{h}(\rho_{t+h} - \rho_t) = \Delta\rho_t$, the limit being understood in $L^2(X, \mathfrak{m})$.

Fix t_0 such that the derivative of the Wasserstein distance exists and the above limit holds and choose any Kantorovich potential φ_{t_0} for $(\rho_{t_0}\mathfrak{m}, \boldsymbol{\sigma})$. We have

$$\begin{aligned}\frac{W_2^2(\rho_{t_0}\mathfrak{m}, \boldsymbol{\sigma})}{2} &= \int_X \varphi_{t_0}\rho_{t_0}\,\mathrm{d}\mathfrak{m} + \int \varphi_{t_0}^c\,\mathrm{d}\boldsymbol{\sigma}\\ \frac{W_2^2(\rho_{t_0+h}\mathfrak{m}, \boldsymbol{\sigma})}{2} &\geq \int_X \varphi_{t_0}\rho_{t_0+h}\,\mathrm{d}\mathfrak{m} + \int \varphi_{t_0}^c\,\mathrm{d}\boldsymbol{\sigma}.\end{aligned}$$

Therefore, since $\varphi_{t_0} \in L^\infty(X, \mathfrak{m})$ we get

$$\frac{W_2^2(\rho_{t_0+h}\mathfrak{m}, \boldsymbol{\sigma})}{2} - \frac{W_2^2(\rho_{t_0}\mathfrak{m}, \boldsymbol{\sigma})}{2} \geq \int \varphi_{t_0}(\rho_{t_0+h} - \rho_{t_0})\,\mathrm{d}\mathfrak{m} = h\int_X \varphi_{t_0}\Delta\rho_{t_0} + o(h).$$

Dividing by $h < 0$ and $h > 0$ and letting $h \to 0$ we get the thesis. □

[19]

Proposition 8.4 (Derivative of the Entropy along a geodesic) *Let $(X, \mathsf{d}, \mathfrak{m})$ be a strong $CD(K,\infty)$ space. Let $\mu_0, \mu_1 \in \mathscr{P}(X)$, $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0,\mu_1)$ and φ a Kantorovich potential for (μ_0,μ_1). Assume that $\boldsymbol{\pi}$ is a test plan and that $\mu_0 \geq c\mathfrak{m}$ from some $c > 0$ and denote by σ_t the density of $\mu_t := (\mathrm{e}_t)_\sharp\boldsymbol{\pi}$. Then*

$$\varliminf_{t\downarrow 0}\frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_t) - \mathrm{Ent}_{\mathfrak{m}}(\mu_0)}{t} \geq \varlimsup_{\varepsilon\downarrow 0}\frac{\mathrm{Ch}(\varphi) - \mathrm{Ch}(\varphi + \varepsilon\sigma_0)}{\varepsilon} \tag{8.8}$$

[18]Dalla proposizione seguente ho tolto l'ipotesi $CD(K,\infty)$

[19]Qui invece strong CD ci vuole, perche' si usa Brenier metrico

Proof The convexity of Ch ensures that the limit at the right hand side exists. From the fact that φ is Lipschitz, it is not hard to see that $\sigma_0 \notin D(\mathrm{Ch})$ implies $\mathrm{Ch}(\varphi + \varepsilon\sigma_0) = +\infty$ for any $\varepsilon > 0$ and in this case there is nothing to prove. Thus, we assume that $\sigma_0 \in D(\mathrm{Ch})$.

The convexity of $z \mapsto z \log z$ gives

$$\frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_t) - \mathrm{Ent}_{\mathfrak{m}}(\mu_0)}{t} \geq \int_X \log \sigma_0 \frac{\sigma_t - \sigma_0}{t}\, \mathrm{d}\mathfrak{m} = \int \frac{\log(\sigma_0 \circ e_t) - \log(\sigma_0 \circ e_0)}{t}\, \mathrm{d}\boldsymbol{\pi}. \tag{8.9}$$

Using the trivial inequality given by Taylor's formula

$$\log b - \log a \geq \frac{b-a}{a} - \frac{|b-a|^2}{2c^2},$$

valid for any a, $b \in [c, \infty)$, we obtain

$$\int \frac{\log(\sigma_0 \circ \mathrm{e}_t) - \log(\sigma_0 \circ \mathrm{e}_0)}{t}\, \mathrm{d}\boldsymbol{\pi} \geq \int \frac{\sigma_0 \circ e_t - \sigma_0 \circ \mathrm{e}_0}{t\sigma_0 \circ \mathrm{e}_0}\, \mathrm{d}\boldsymbol{\pi} - \frac{1}{2tc^2} \int |\sigma_0 \circ \mathrm{e}_t - \sigma_0 \circ \mathrm{e}_0|^2\, \mathrm{d}\boldsymbol{\pi}. \tag{8.10}$$

Taking into account Remark 4.12, the last term in this expression can be bounded from above by

$$\frac{1}{2tc^2} \int \left(\int_0^t |\nabla\sigma_0|_w \circ \mathrm{e}_s \right)^2 \mathrm{d}s\, \mathrm{d}\boldsymbol{\pi} \leq \frac{1}{2c^2} \int \int_0^t |\nabla\sigma_0|_w^2 \circ \mathrm{e}_s\, \mathrm{d}s\, \mathrm{d}\boldsymbol{\pi}, \tag{8.11}$$

which goes to 0 as $t \to 0$.

Now let $S : \mathrm{Geo(X)} \to \mathbb{R}$ be the Borel function defined by $S(\gamma) := \sigma_0 \circ \gamma_0$ and define $\tilde{\boldsymbol{\pi}} := \frac{1}{S}\boldsymbol{\pi}$. It is easy to check that $(\mathrm{e}_0)_\# \tilde{\boldsymbol{\pi}} = \mathfrak{m}$, so that in particular $\tilde{\boldsymbol{\pi}}$ is a probability measure. Also, the bound $\sigma_0 \geq c > 0$ ensures that $\tilde{\boldsymbol{\pi}}$ is a test plan. By definition we have

$$\int \frac{\sigma_0 \circ e_t - \sigma_0 \circ \mathrm{e}_0}{t\sigma_0 \circ \mathrm{e}_0}\, \mathrm{d}\boldsymbol{\pi} = \int \frac{\sigma_0 \circ e_t - \sigma_0 \circ \mathrm{e}_0}{t}\, \mathrm{d}\tilde{\boldsymbol{\pi}}.$$

The latter equality and inequalities (8.9), (8.10) and (8.11) ensure that to conclude it is sufficient to show that

$$\varliminf_{t\downarrow 0} \int \frac{\sigma_0 \circ e_t - \sigma_0 \circ \mathrm{e}_0}{t}\, \mathrm{d}\tilde{\boldsymbol{\pi}} \geq \lim_{\varepsilon\downarrow 0} \frac{\mathrm{Ch}(\varphi) - \mathrm{Ch}(\varphi + \varepsilon\sigma_0)}{\varepsilon}. \tag{8.12}$$

Here we apply the key Lemma 8.2. Observe that Theorem 7.3 ensures that

$$|\nabla\varphi|_w(\gamma_0) = \lim_{t\downarrow 0} \frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{t} = \mathsf{d}(\gamma_0, \gamma_1)$$

where the convergence is understood in $L^2(\boldsymbol{\pi})$. Thus the same holds for $L^2(\tilde{\boldsymbol{\pi}})$ and the hypotheses of Lemma 8.2 are satisfied with $\tilde{\boldsymbol{\pi}}$ as test plan, $g := \varphi$ and $f := \sigma_0$. Equation (8.5) then gives

$$\begin{aligned}\varliminf_{t\downarrow 0} \int \frac{\sigma_0 \circ e_t - \sigma_0 \circ \mathrm{e}_0}{t}\, \mathrm{d}\tilde{\boldsymbol{\pi}} &\geq \varlimsup_{\varepsilon\downarrow 0} \frac{1}{2} \int \frac{|\nabla g|_w^2(\gamma_0) - |\nabla g + \varepsilon f|_w^2(\gamma_0)}{\varepsilon}\, \mathrm{d}\tilde{\boldsymbol{\pi}}(\gamma) \\ &= \varlimsup_{\varepsilon\downarrow 0} \frac{1}{2} \int_X \frac{|\nabla g|_w^2(x) - |\nabla g + \varepsilon f|_w^2(x)}{\varepsilon}\, \mathrm{d}\mathfrak{m}(x),\end{aligned}$$

which concludes the proof. □

9 Riemannian Ricci bounds

We say that $(X, \mathsf{d}, \mathfrak{m})$ has *Riemannian Ricci curvature* bounded below by $K \in \mathbb{R}$ (in short, it is a $RCD(K,\infty)$ space) if any of the 3 equivalent conditions stated in the following theorem is true.

Theorem 9.1 *Let $(X, \mathsf{d}, \mathfrak{m})$ be a compact and normalized metric measure space and $K \in \mathbb{R}$. The following three properties are equivalent.*

(i) $(X, \mathsf{d}, \mathfrak{m})$ *is a strong $CD(K,\infty)$ space (Definition 7.1) and the L^2-gradient flow of* Ch *is linear.*[20]

(ii) $(X, \mathsf{d}, \mathfrak{m})$ *is a strong $CD(K,\infty)$ space (Definition 7.1) and Cheeger's energy is quadratic, i.e.*

$$2\big(\mathrm{Ch}(f)+\mathrm{Ch}(g)\big)=\mathrm{Ch}(f+g)+\mathrm{Ch}(f-g),\qquad \forall f,\, g\in L^2(X,\mathfrak{m}). \tag{9.1}$$

(iii) $(X, \mathsf{d}, \mathfrak{m})$ *is a geodesic space and for any $\mu \in \mathscr{P}(X)$ with* $\mathrm{supp}(\mu) \subset \mathrm{supp}(\mathfrak{m})$ *there exists an* EVI_K*-gradient flow for* $\mathrm{Ent}_{\mathfrak{m}}$ *starting from μ.*

Proof
(i) $\Rightarrow$ **(ii)**. Since the heat semigroup P_t in $L^2(X,\mathfrak{m})$ is linear we obtain that Δ is a linear operator (i.e. its domain $D(\Delta)$ is a subspace of $L^2(X,\mathfrak{m})$ and $\Delta : D(\Delta) \to L^2(X,\mathfrak{m})$ is linear). Since $t \mapsto \mathrm{Ch}(P_t(f))$ is locally Lipschitz, tends to 0 as $t \to \infty$ and $\partial_t \mathrm{Ch}(P_t(f)) = -\|\Delta P_t(f)\|^2_{L^2}$ for a.e. $t > 0$ (see (4.4)), we have

$$\mathrm{Ch}(f)=\int_0^\infty \|\Delta P_t(f)\|^2_{L^2(X,\mathfrak{m})}\,\mathrm{d}t.$$

Therefore Ch, being an integral of quadratic forms, is a quadratic form. Specifically, for any $f,\, g \in L^2(X,\mathfrak{m})$ it holds

$$\begin{aligned}
\mathrm{Ch}(f+g)+\mathrm{Ch}(f-g) &= \int_0^\infty \|\Delta P_t(f+g)\|^2_{L^2(X,\mathfrak{m})}+\|\Delta P_t(f-g)\|^2_{L^2(X,\mathfrak{m})}\,\mathrm{d}t\\
&= \int_0^\infty \|\Delta P_t(f)+\Delta P_t(g)\|^2_{L^2(X,\mathfrak{m})}+\|\Delta P_t(f)-\Delta P_t(g)\|^2_{L^2(X,\mathfrak{m})}\,\mathrm{d}t\\
&= \int_0^\infty 2\|\Delta P_t(f)\|^2_{L^2(X,\mathfrak{m})}+2\|\Delta P_t(g)\|^2_{L^2(X,\mathfrak{m})}\,\mathrm{d}t\\
&= 2\mathrm{Ch}(f)+2\mathrm{Ch}(g).
\end{aligned}$$

(ii) $\Rightarrow$ **(iii)**. [21] Thanks to Remark 2.6 it is sufficient to prove that a gradient flow in the EVI_K sense exists for an initial datum $\mu_0 \ll \mathfrak{m}$ with density bounded away from 0 and infinity. Let ρ_0 be this density, (ρ_t) the heat flow starting from it and recall that from the maximum principle 4.9 we know that the ρ_t's are far from 0 and infinity as well for any $t > 0$. Fix a reference probability measure $\boldsymbol{\sigma}$ with density bounded away from 0 and infinity as well. For any $t \geq 0$ pick a test plan $\boldsymbol{\pi}_t$ optimal for $(\rho_t\mathfrak{m}, \boldsymbol{\sigma})$. Define $\boldsymbol{\sigma}_t^s := (e_s)_\# \pi_t$.

[20] Nel punto (i) ho messo linearita' in L^2, che comunque veniva usata per passare da (i) a (ii). Di consequenza ho modificato l'implicazione da (iii) a (i)

[21] Va inserita anche la prova che (X, d) è geodetico

We claim that for a.e. $t \in (0, \infty)$ it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t\mathfrak{m}, \sigma\mathfrak{m}) \le \varliminf_{s\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\sigma_t^s) - \mathrm{Ent}_{\mathfrak{m}}(\sigma_t^0)}{s}. \tag{9.2}$$

Let φ_t be a Kantorovich potential for $\rho_t\mathfrak{m}, \sigma\mathfrak{m}$. By Proposition 8.3 we know that for a.e. $t \in (0, \infty)$ it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t\mathfrak{m}, \sigma\mathfrak{m}) = \int \varphi\Delta\rho_t\,\mathrm{d}\mathfrak{m} \le \varliminf_{\varepsilon\downarrow 0} \frac{\mathrm{Ch}(\rho_t - \varepsilon\varphi_t) - \mathrm{Ch}(\rho_t)}{\varepsilon},$$

while from Proposition 8.4 we have that for any $t > 0$ it holds

$$\varliminf_{s\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\sigma_t^s) - \mathrm{Ent}_{\mathfrak{m}}(\sigma_t^0)}{s} \ge \varliminf_{\varepsilon\downarrow 0} \frac{\mathrm{Ch}(\varphi_t) - \mathrm{Ch}(\varphi_t + \varepsilon\rho_t)}{\varepsilon}.$$

Here we use the fact that Ch is quadratic. Indeed in this case simple algebraic manipulations show that

$$\frac{\mathrm{Ch}(\rho_t - \varepsilon\varphi_t) - \mathrm{Ch}(\rho_t)}{\varepsilon} = \frac{\mathrm{Ch}(\varphi_t) - \mathrm{Ch}(\varphi_t + \varepsilon\rho_t)}{\varepsilon} + O(\epsilon), \qquad \forall t > 0,$$

and therefore (9.2) is proved.

Now notice that the K-convexity of the entropy yields

$$\varliminf_{s\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\sigma_t^s) - \mathrm{Ent}_{\mathfrak{m}}(\sigma_t^0)}{s} \le \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\sigma}) - \mathrm{Ent}_{\mathfrak{m}}(\rho_t\mathfrak{m}) - \frac{K}{2}W_2^2(\rho_t\mathfrak{m}, \boldsymbol{\sigma}),$$

and therefore we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t\mathfrak{m}, \sigma\mathfrak{m}) + \mathrm{Ent}_{\mathfrak{m}}(\rho_t\mathfrak{m}) + \frac{K}{2}W_2^2(\rho_t\mathfrak{m}, \boldsymbol{\sigma}) \le \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\sigma}), \qquad \text{for a.e. } t \in (0, \infty).$$

By Proposition 2.3 we conclude.

(**iii**) $\Rightarrow$ (**i**). Since (X, d) is geodesic, so is $(\mathscr{P}(X), W_2)$, which together with existence of EVI_K-gradient flows for $\mathrm{Ent}_{\mathfrak{m}}$ yields, via Proposition 2.7, K-geodesic convexity of $\mathrm{Ent}_{\mathfrak{m}}$ along all geodesics in W_2. In particular, $(X, \mathsf{d}, \mathfrak{m})$ is a strong $CD(K, \infty)$ space.

We turn to the linearity. Let (μ_t^0), (μ_t^1) be two EVI_K-gradient flows of the relative entropy and, for $\lambda \in (0, 1)$ fixed, define $\mu_t^\lambda := (1-\lambda)\mu_t^0 + \lambda\mu_t^1$.

We claim that (μ_t) is an EVI_K-gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$. To prove this, fix $\nu \in \mathscr{P}(X)$, $t > 0$ and an optimal plan $\boldsymbol{\gamma} \in \mathrm{O{\scriptsize PT}}(\mu_t^\lambda, \nu)$. Since $\mu_t^i \ll \mu_t^\lambda = \pi^1_\sharp\boldsymbol{\gamma}$ for $i = 0, 1$ we can define, as in Definition 5.3, the plans $\boldsymbol{\gamma}_{\mu_t^i} \in \mathscr{P}(X^2)$ and the measures $\nu^i := \boldsymbol{\gamma}_\sharp\mu_t^i$, $i = 0, 1$. Since $\mathrm{supp}(\gamma_{\mu_t^i}) \subset \mathrm{supp}(\boldsymbol{\gamma})$, we have that $\gamma_{\mu_t^i} \in \mathrm{O{\scriptsize PT}}(\mu_t^i, \nu^i)$, therefore from $\boldsymbol{\gamma} = (1-\lambda)\boldsymbol{\gamma}_{\mu_t^0} + \lambda\boldsymbol{\gamma}_{\mu_t^1}$ we deduce

$$W_2^2(\mu_t^\lambda, \nu) = (1-\lambda)W_2^2(\mu_t^0, \nu^0) + \lambda W_2^2(\mu_t^1, \nu^1). \tag{9.3}$$

On the other hand, from the convexity of the squared Wasserstein distance we immediately get that

$$W_2^2(\mu_{t+h}^\lambda, \nu) \le (1-\lambda)W_2^2(\mu_{t+h}^0, \nu^0) + \lambda W_2^2(\mu_{t+h}^1, \nu^1), \qquad \forall h > 0. \tag{9.4}$$

Furthermore, recalling (*iii*) of Proposition 5.4, we get

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t^\lambda) - \mathrm{Ent}_{\mathfrak{m}}(\nu) \le (1-\lambda)\big(\mathrm{Ent}_{\mathfrak{m}}(\mu_t^0) - \mathrm{Ent}_{\mathfrak{m}}(\nu^0)\big) + \lambda\big(\mathrm{Ent}_{\mathfrak{m}}(\mu_t^1) - \mathrm{Ent}_{\mathfrak{m}}(\nu^1)\big). \tag{9.5}$$

The fact that (μ_t^0) and (μ_t^1) are EVI_K-gradient flows for $\mathrm{Ent}_{\mathfrak{m}}$ (see in particular the characterization (*iii*) given in Proposition 2.3) in conjunction with (9.3), (9.4) and (9.5) yield

$$\overline{\lim_{h\downarrow 0}}\,\frac{W_2^2(\mu_{t+h}^\lambda,\nu) - W_2^2(\mu_t^\lambda,\nu)}{2} + \frac{K}{2}W_2^2(\mu_t^\lambda,\nu) + \mathrm{Ent}_{\mathfrak{m}}(\mu_t^\lambda) \le \mathrm{Ent}_{\mathfrak{m}}(\nu). \tag{9.6}$$

Since $t > 0$ and $\nu \in \mathscr{P}(X)$ were arbitrary, we proved that (μ_t^λ) is a EVI_K-gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ (see again (*iii*) of Proposition 2.3).

Thus, recalling the identification of gradient flows, we proved that the L^2-heat flow is additive in $D(\mathrm{Ent}_{\mathfrak{m}})$. Since the heat flow in $L^2(X,\mathfrak{m})$ commutes with additive and multiplicative constants, it is easy to get from this linearity in the class of bounded functions. By L^2 contractivity, linearity extends to the whole of $L^2(X,\mathfrak{m})$. □

We conclude by discussing some basic properties of the spaces with Riemannian Ricci curvature bounded from below.

We start observing that Riemannian manifolds with Ricci curvature bounded below by K are $RCD(K,\infty)$ spaces, as they are non branching $CD(K,\infty)$ spaces and the heat flow is linear on them. Also, from the studies made in [?], [?], [?] and [?] we also know that finite dimensional Alexandrov spaces with curvature bounded from below are $RCD(K,\infty)$ spaces as well. On the other side, Finsler manifolds are ruled out, as it is known (see for instance [?]) that the heat flow is linear on a Finsler manifold if and only if the manifold is Riemannian.

The stability of the $RCD(K,\infty)$ notion can be deduced by the stability of EVI_K-gradient flows w.r.t. Γ-convergence of functionals, which is an easy consequence of the integral formulation in (*ii*) of Proposition 2.3.

Hence $RCD(K,\infty)$ spaces have the same basic properties of $CD(K,\infty)$ spaces, which gives to this notion the right of being called a synthetic (or weak) notion of Ricci curvature bound.

The point is then to understand the additional analytic/geometric properties of these spaces, which come mainly by the addition of linearity condition. A first consequence is that the heat flow contracts, up to an exponential factor, the distance W_2, i.e.

$$W_2(\mu_t,\nu_t) \le e^{-Kt}W_2(\mu_0,\nu_0), \qquad \forall t \ge 0,$$

whenever (μ_t), $(\nu_t) \subset \mathscr{P}_2(X)$ are gradient flows of the entropy.

By a duality argument (see [?], [?]), this property implies the Bakry-Emery gradient estimate

$$|\nabla \mathsf{h}_t(f)|_w^2(x) \le e^{-2Kt}\mathsf{h}_t(|\nabla f|_w^2)(x), \qquad \text{for } \mathfrak{m}\text{-a.e. } x \in X,$$

for all $t > 0$, where $\mathsf{h}_t : L^2(X,\mathfrak{m}) \to L^2(X,\mathfrak{m})$ is the heat flow seen as gradient flow of Ch. If $(X,\mathsf{d},\mathfrak{m})$ is doubling and supports a local Poincaré inequality, then also the Lipschitz regularity of the heat kernel is deduced (following an argument described in [?]).

Also, since in $RCD(K,\infty)$ spaces Ch is a quadratic form, if we define

$$\mathcal{E}(f,g) := \mathrm{Ch}(f+g) - \mathrm{Ch}(f) - \mathrm{Ch}(g), \qquad \forall f,g \in W^{1,2}(X,\mathsf{d},\mathfrak{m}),$$

we get a closed Dirichlet form on $L^2(X,\mathfrak{m})$ (closure follows from the L^2-lower semicontinuity of Ch). Hence it is natural to compare the calculus on $RCD(K,\infty)$ spaces with the abstract one available for Dirichlet forms (see [**?**]). The picture here is pretty clear and consistent. Recall that to any $f \in D(\mathcal{E})$ one can associate the energy measure $[f]$ defined by

$$[f](\varphi) := -\mathcal{E}(f, f\varphi) + \mathcal{E}(f^2/2, \varphi).$$

Then it is possible to show that the energy measure coincides with $|\nabla f|_*^2 \mathfrak{m}$. Also, the distance d coincides with the intrinsic distance $\mathsf{d}_{\mathcal{E}}$ induced by the form, defined by

$$\mathsf{d}_{\mathcal{E}}(x,y) := \sup\Big\{|g(x) - g(y)| \ : \ g \in D(\mathcal{E}) \cap C(X), \ [g] \leq \mathfrak{m}\Big\}.$$

Taking advantage of these identification and of the locality of $\mathcal{E}$ (which is a consequence of the locality of the notion $|\nabla f|_*$), one can also see that on $RCD(K,\infty)$ spaces a continuous Brownian motion with continuous sample paths associated to h_t exists and is unique.

Finally, for $RCD(K,\infty)$ spaces it is possible to prove tensorization and globalization properties which are in line with those available for $CD(K,\infty)$ spaces.