# Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below

Luigi Ambrosio [*] Nicola Gigli [†] Giuseppe Savaré [‡]

September 21, 2012

**Abstract**

This paper is devoted to a deeper understanding of the heat flow and to the refinement of calculus tools on metric measure spaces $(X, \mathsf{d}, \mathfrak{m})$. Our main results are:

- A general study of the relations between the Hopf-Lax semigroup and Hamilton-Jacobi equation in metric spaces $(X, \mathsf{d})$.
- The equivalence of the heat flow in $L^2(X, \mathfrak{m})$ generated by a suitable Dirichlet energy and the Wasserstein gradient flow of the relative entropy functional $\mathrm{Ent}_{\mathfrak{m}}$ in the space of probability measures $\mathscr{P}(X)$.
- The proof of density in energy of Lipschitz functions in the Sobolev space $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ under the only assumption that $\mathfrak{m}$ is locally finite.
- A fine and very general analysis of the differentiability properties of a large class of Kantorovich potentials, in connection with the optimal transport problem.

Our results apply in particular to spaces satisfying Ricci curvature bounds in the sense of Lott & Villani [28] and Sturm [35, 36] and require neither the doubling property nor the validity of the local Poincaré inequality.

**MSC-classification:** 52C23, 49J52, 49Q20, 58J35, 35K90, 31C25

## Contents



[*] l.ambrosio@sns.it
[†] nicola.gigli@unice.fr
[‡] giuseppe.savare@unipv.it

# 1 Introduction

Aim of this paper is to provide a deeper understanding of analysis in metric measure spaces, with a particular focus on the properties of the heat flow. Our main results, whose validity does not depend on doubling and Poincaré assumptions, are:

(i) The proof that the Hopf-Lax formula produces sub-solutions of the Hamilton-Jacobi equation on general metric spaces $(X, \mathsf{d})$, and solutions if $(X, \mathsf{d})$ is a length space.

(ii) The proof of equivalence of the heat flow in $L^2(X, \mathfrak{m})$ generated by a suitable Dirichlet energy and the Wasserstein gradient flow in $\mathscr{P}(X)$ of the relative entropy functional $\mathrm{Ent}_{\mathfrak{m}}$ w.r.t. $\mathfrak{m}$.

(iii) The proof that Lipschitz functions are always dense in energy in the Sobolev space $W^{1,2}$. This is achieved by showing the equivalence of two weak notions of modulus of the gradient: the first one (inspired by Cheeger [10], see also [21], [19], and the recent review [20]), that we call *relaxed gradient*, is defined by $L^2(X, \mathfrak{m})$-relaxation of the pointwise Lipschitz constant in the class of Lipschitz functions; the second one (inspired by Shanmugalingam [34]), that we call *weak upper gradient*, is based on the validity of the fundamental theorem of calculus along almost all curves. These two notions of gradient will be compared and identified, assuming only $\mathfrak{m}$ to be locally finite. We might consider the former gradient as a "vertical" derivative, related to variations in the dependent variable, while the latter is an "horizontal" derivative, related to variations in the independent variable.

(iv) A fine and very general analysis of the differentiability properties of a large class of Kantorovich potentials, in connection with the optimal transport problem.

Our results apply in particular to spaces satisfying Ricci curvature bounds in the sense of Lott & Villani [28] and Sturm [35, 36], that we call in this introduction *LSV* spaces. Indeed, the development of a "calculus" in this class of spaces has been one of our motivations. In particular we are able to prove the following result (see Theorem 9.3 for a more precise and general statement): if $(X,\mathsf{d},\mathfrak{m})$ is a $CD(K,\infty)$ space and $\mathfrak{m}\in\mathscr{P}(X)$, then

(a) For every $\mu=f\mathfrak{m}\in\mathscr{P}(X)$ the Wasserstein slope $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\mu)$ of the relative entropy $\mathrm{Ent}_{\mathfrak{m}}$ coincides with the Fisher information functional $\int_{\{\rho>0\}}|\nabla\rho|_*^2/\rho\,\mathrm{d}\mathfrak{m}$, where $|\nabla\rho|_*$ is the relaxed gradient of $\rho$ (see the brief discussion before (1.3)).

(b) For every $\mu_0=f_0\mathfrak{m}\in D(\mathrm{Ent}_{\mathfrak{m}})\cap\mathscr{P}_2(X)$ there exists a unique gradient flow $\mu_t=f_t\mathfrak{m}$ of $\mathrm{Ent}_{\mathfrak{m}}$ starting from $\mu_0$ in $(\mathscr{P}_2(X),W_2)$, and if $f_0\in L^2(X,\mathfrak{m})$ the functions $f_t$ coincide with the $L^2(X,\mathfrak{m})$ gradient flow of Cheeger's energy $\mathsf{Ch}_*$, defined by (see also (1.3) for an equivalent definition)

$$\mathsf{Ch}_*(f):=\frac{1}{2}\inf\left\{\liminf_{h\to\infty}\int|\nabla f_h|^2\,\mathrm{d}\mathfrak{m}:f_h\in\mathrm{Lip}(X),\ \int_X|f_h-f|^2\,\mathrm{d}\mathfrak{m}\to 0\right\}. \tag{1.1}$$

On the other hand, we believe that the"calculus" results described in (iii) are of a wider interest for analysis in metric measure spaces, beyond the application to *LSV* spaces. Particularly important is not only the identification of heat flows, but also the identification of weak gradients that was previously known only under doubling and Poincaré asssumptions. The key new idea is to use the heat flow and the rate of energy dissipation, instead of the usual covering arguments, to prove the optimal approximation by Lipschitz functions, see also Remark 4.6 and Remark 5.10 for a detailed comparison with the previous approaches.

In connection with (ii), notice that the equivalence so far has been proved in Euclidean spaces by Jordan-Kinderleher-Otto, in the seminal paper [22], in Riemannian manifolds by Erbar and Villani [13, 38], in Hilbert spaces by [5], in Finsler spaces by Ohta-Sturm [29] and eventually in Alexandrov spaces by Gigli-Kuwada-Ohta [17]. In fact, the strategy pursued in [17], that we shall describe later on, had a great influence on our work. The distinguished case when the gradient flows are linear will be the object, in connection with *LSV* spaces, of a detailed investigation in [3].

We exploit as much as possible the variational formulation of gradient flows on one hand (based on sharp energy dissipation rate and the notion of descending slope) and the variational structure of the optimal transportation problem to develop a theory that does not rely on finite dimensionality and doubling properties; we are even able to cover situations where the distance $\mathsf{d}$ is allowed to take the value $+\infty$, as it happens for instance in optimal transportation problems in Wiener spaces (see for instance [15, 14]). We are also able to deal with $\sigma$-finite measures $\mathfrak{m}$, provided they are representable in the form $\mathrm{e}^{V^2}\,\tilde{\mathfrak{m}}$ with $\tilde{\mathfrak{m}}(X)\le 1$ and $V:X\to[0,\infty)$ $\mathsf{d}$-Lipschitz weight function bounded from above on compact sets.

In order to reach this level of generality, it is useful to separate the roles of the topology $\tau$ of $X$ (used for the measure-theoretic structure) and of the possibly extended distance $\mathsf{d}$

involved in the optimal transport problem, introducing the concept of Polish extended metric measure space $(X, \mathsf{d}, \tau, \mathfrak{m})$. Of course, the case when $\mathsf{d}$ is a distance inducing the Polish topology $\tau$ is included. Since we assume neither doubling properties nor the validity of the Poincaré inequalities, we can't rely on Cheeger's theory [10], developed precisely under these assumptions. The only known connection between synthetic curvature bounds and this set of assumptions is given in [27], where the authors prove that in non-branching $LSV$ spaces the Poincaré inequality holds under the so-called $CD(K, N)$ assumption ($N < \infty$), a stronger curvature assumption which involves also the dimension.

Now we pass to a more detailed description of the results of the paper, the problems and the related literature. In Section 2 we introduce all the basic concepts used in the paper: first we define extended metric spaces $(X, \mathsf{d})$, Polish extended spaces $(X, \mathsf{d}, \tau)$ (in our axiomatization $\mathsf{d}$ and $\tau$ are not completely decoupled, see (iii) and (iv) in Definition 2.3), absolutely continuous curves, metric derivative $|\dot{x}_t|$, local Lipschitz constant $|\nabla f|$, one-sided slopes $|\nabla^{\pm} f|$. Then, we see how in Polish extended spaces one can naturally state the optimal transport problem with cost $c = \mathsf{d}^2$ either in terms of transport plans (i.e. probability measures in $X \times X$) or, when the space is geodesic, in terms of geodesic transport plans, namely probability measures, with prescribed marginals at $t = 0$, $t = 1$, in the space $\mathrm{Geo}(X)$ of constant speed geodesics in $X$. In Subsection 2.5 we recall the basic definition of gradient flow $(y_t)$ of an energy functional $E$: it is based on the integral formulation of the sharp energy dissipation rate

$$-\frac{\mathrm{d}}{\mathrm{d}t} E(y_t) \geq \frac{1}{2}|\dot{y}_t|^2 + \frac{1}{2}|\nabla^- E(y_t)|^2$$

which, under suitable additional assumptions (for instance the fact that $|\nabla^- E|$ is an upper gradient of $E$, as it happens for $K$-geodesically convex functionals), turns into an equality for almost every time. These facts will play a fundamental role in our analysis.

In Section 3 we study the fine properties of the Hopf-Lax semigroup

$$Q_t f(x) := \inf_{y \in X} f(y) + \frac{\mathsf{d}^2(x, y)}{2t}, \qquad (x, t) \in X \times (0, \infty) \tag{1.2}$$

in a extended metric space $(X, \mathsf{d})$. Here the main technical novelty, with respect to [26], is the fact that we do not rely on Cheeger's theory (in fact, no reference measure $\mathfrak{m}$ appears here) to show in Proposition 3.6 that in length spaces $(x, t) \mapsto Q_t f(x)$ is a pointwise solution to the Hamilton-Jacobi equation $\partial_t Q_t f + |\nabla Q_t|^2/2 = 0$: precisely, for given $x$, the equation does not hold for at most countably many times $t$. This is achieved refining the estimates in [4, Lemma 3.1.2] and looking at the monotonicity properties w.r.t. $t$ of the quantities

$$D^+(x, t) := \sup \limsup_{n \to \infty} \mathsf{d}(x, y_n), \qquad D^-(x, t) := \inf \liminf_{n \to \infty} \mathsf{d}(x, y_n)$$

where the supremum and the infimum run among all minimizing sequences $(y_n)$ in (1.2). Although only the easier subsolution property $\partial_t Q_t f + |\nabla Q_t|^2/2 \leq 0$ (which does not involve the length condition) will play a crucial role in the results of Sections 6 and 8, another byproduct of this refined analysis is a characterization of the slope of $Q_t f$ (see Proposition 3.6) which applies, to some extent, also to Kantorovich potentials (see 10).

In Section 4 we follow quite closely [10], defining the collection of relaxed gradients of $f$ as the weak $L^2$ limits of $|\nabla f_n|$, where $f_n$ are $\mathsf{d}$-Lipschitz and $f_n \to f$ in $L^2(X, \mathfrak{m})$ (the differences with respect to [10] are detailed in Remark 4.6). The collection of all these weak limits is a convex closed set in $L^2(X,\mathfrak{m})$, whose minimal element is called *relaxed gradient*, and denoted by $|\nabla f|_*$. One can then see that Cheeger's convex and lower semicontinuous functional (1.1) can be equivalently represented as

$$\mathsf{Ch}_*(f) = \frac{1}{2}\int_X |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m} \tag{1.3}$$

(set to $+\infty$ if $f$ has no relaxed gradient) and get a canonical gradient flow in $L^2(X,\mathfrak{m})$ of $\mathsf{Ch}_*$ and a notion of Laplacian $\Delta_{\mathsf{d},\mathfrak{m}}$ associated to $\mathsf{Ch}_*$. As explained in Remark 4.12 and Remark 4.14, this construction can be trivial if no other assumption on $(X,\mathsf{d},\tau,\mathfrak{m})$ is imposed, and in any case $\mathsf{Ch}_*$ is not necessarily a quadratic form and the Laplacian, though 1-homogeneous, is not necessarily linear. Precisely because of this potential nonlinearity we avoided the terminology "Dirichlet form", usually associated to quadratic forms, in connection with $\mathsf{Ch}_*$.
It is also possible to consider the one-sided slopes $|\nabla^\pm f|$, getting one-sided relaxed gradients $|\nabla^\pm f|_*$ and Cheeger's corresponding functionals $\mathsf{Ch}_*^\pm$; eventually, but this fact is not trivial, we prove that the one-sided relaxed functionals coincide with $\mathsf{Ch}_*$, see Remark 6.4.

Section 5 is devoted to the "horizontal" notion of modulus of gradient, that we call *weak upper gradient*, along the lines of [34]: roughly speaking, we say that $G$ is a weak upper gradient of $f$ if the inequality $|f(\gamma_0)-f(\gamma_1)|\le\int_\gamma G$ holds along "almost all" curves with respect to a suitable collection $\mathcal{T}$ of probability measures concentrated on absolutely continuous curves, see Definition 5.4 for the precise statement. The class of weak upper gradients has good stability properties that allow to define a minimal weak upper gradient, that we shall denote by $|\nabla f|_{w,\mathcal{T}}$, and to prove that $|\nabla f|_{w,\mathcal{T}}\le|\nabla f|_*$ $\mathfrak{m}$-a.e. in $X$ for all $f\in D(\mathsf{Ch}_*)$ if $\mathcal{T}$ is concentrated on the class of all the absolutely continuous curves with finite 2-energy.

Section 6 is devoted to prove the converse inequality and therefore to show that in fact the two notions of gradient coincide. The proof relies on the fine analysis of the rate of dissipation of the entropy $\int_X h_t\log h_t\,\mathrm{d}\mathfrak{m}$ along the gradient flow of $\mathsf{Ch}_*$, and on the representation of $h_t\mathfrak{m}$ as the time marginal of a random curve. The fact that $h_t\mathfrak{m}$ (having a priori only $L^2(X,\mathfrak{m})$ regularity in time and Sobolev regularity in space) can be viewed as an absolutely continuous curve with values in $(\mathscr{P}(X),W_2)$ is a consequence of Lemma 6.1, inspired by [17, Proposition 3.7]. More precisely, the metric derivative of $t\mapsto h_t\mathfrak{m}$ w.r.t. the Wasserstein distance can be estimated as follows:

$$|\dot{h_t\mathfrak{m}}|^2 \le 4\int_X |\nabla\sqrt{h_t}|_*^2\,\mathrm{d}\mathfrak{m} \qquad \text{for a.e. } t\in(0,\infty). \tag{1.4}$$

The latter estimate, written in an integral form, follows by a delicate approximation procedure, the Kantorovich duality formula and the fine properties of the Hopf-Lax semigroup we proved.

In Section 7 we introduce the relative entropy functional and the Fisher information

$$\mathrm{Ent}_\mathfrak{m}(\rho\mathfrak{m}) = \int_X \rho\log\rho\,\mathrm{d}\mathfrak{m}, \qquad \mathsf{F}(\rho) = 4\int_X |\nabla\sqrt{\rho}|_*^2\,\mathrm{d}\mathfrak{m},$$

and prove two crucial inequalities for the descending slope of $\mathrm{Ent}_\mathfrak{m}$: the first one, still based on Lemma 6.1, provides the lower bound via the Fisher information

$$\mathsf{F}(\rho) = 4\int_X |\nabla\sqrt{\rho}|_*^2\ \mathrm{d}\mathfrak{m} \le |\nabla^-\mathrm{Ent}_\mathfrak{m}|^2(\mu) \quad \text{if } \mu=\rho\mathfrak{m}, \tag{1.5}$$

and the second one, combining [38, Theorem 20.1] with an approximation argument, the upper bound when $\rho$ is $\mathsf{d}$-Lipschitz (and satisfies further technical assumptions if $\mathfrak{m}(X)=\infty$)

$$|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu) \le 4\int_X |\nabla^- \sqrt{\rho}|^2 \ \mathrm{d}\mathfrak{m} \quad \text{if } \mu=\rho\mathfrak{m}. \tag{1.6}$$

The identification of the squared descending slope of $\mathrm{Ent}_{\mathfrak{m}}$ (which is always a convex functional, as we show in §7.3) with the Fisher information thus follows, whenever $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ satisfies a lower semicontinuity property, as in the case of LSV spaces.

In Section 8 we show how the uniqueness proof written by the second author in [16] for the case of finite reference measures can be adapted, thanks to the tightness properties of the relative entropy, to our more general framework: we prove uniqueness of the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ first for flows with uniformly bounded densities and then, assuming that $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ is an upper gradient, without any restriction on the densities. In this way we obtain the key property that the Wasserstein gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$, understood in the metric sense of Subsection 2.5, has a unique solution for a given initial condition with finite entropy. This uniqueness phenomenon should be compared with the recent work [30], where it it shown that in $LSV$ spaces (precisely in Finsler spaces) contractivity of the Wasserstein distance along the semigroup may fail.

In Section 8.3 we prove the equivalence of the two gradient flows, in the natural class where a comparison is possible, namely nonnegative initial conditions $f_0 \in L^1 \cap L^2(X,\mathfrak{m})$ (if $\mathfrak{m}(X)=\infty$ we impose also that $\int_X f_0 V^2 \,\mathrm{d}\mathfrak{m} < \infty$). In the proof of this result, that requires suitable assumptions on $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$, we follow the new strategy introduced in [17]: while the traditional approach aims at showing that the Wasserstein gradient flow $\mu_t = f_t\mathfrak{m}$ solves a "conventional" PDE, here we show the converse, namely that the gradient flow of Cheeger's energy provides solutions to the Wasserstein gradient flow. Then, uniqueness (and existence) at the more general level of Wasserstein gradient flow provides equivalence of the two gradient flows. The key properties to prove the validity of the sharp dissipation rate

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

where $f_t$ is the gradient flow of $\mathsf{Ch}_*$, are the slope estimate (1.6) and the metric derivative estimate (1.4).

We also emphasize that some results of ours, as the uniqueness provided in Theorem 8.1 for flows with bounded densities, or the full convergence as the time step tends to 0 of the Jordan-Kinderleher-Otto scheme in Corollary 8.2, require *no* assumption on the space (except for an exponential volume growth condition) and $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$, so that they are applicable even to spaces which are known to be not $LSV$ or for which the lower semicontinuity of $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ fails or it is unknown, as Carnot groups endowed with the Carnot-Carathéodory distance and the Haar measure.

In Section 9 we show, still following to a large extent [16], the crucial lower semicontinuity of $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ in $LSV$ spaces; this shows that all existence and uniqueness results of Section 8.3 are applicable to $LSV$ spaces and that the correspondence between the heat flows is complete.

The paper ends, in the last section, with results that are important for the development of a "calculus" with Kantorovich potentials. They will play a key role in some proofs of [3]. We included these results here because their validity does not really depend on curvature properties, but rather on their implications, namely the existence of geodesic interpolations satisfying suitable $L^\infty$ bounds.

Under these assumptions, in Theorem 10.3 we prove that the ascending slope $|\nabla^+\varphi|$ is the minimal weak upper gradient for Kantorovich potentials $\varphi$, A nice byproduct of this proof is a "metric" Brenier theorem, namely the fact that the transport distance $\mathsf{d}(x,y)$ coincides for $\boldsymbol{\gamma}$-a.e. $(x,y)$ with $|\nabla^+\varphi|(x)$ even when the transport plan $\boldsymbol{\gamma}$ is multi-valued. In addition, $|\nabla^+\varphi|$ coincides $\mathfrak{m}$-a.e. with the relaxed and weak upper gradients. To some extent, the situation here is "dual" to the one appearing in the transport problem with cost=Euclidean distance: in that situation, one knows the direction of transport, without knowing the distance. In addition, we obtain in Theorem 10.4 a kind of differentiability property of $\varphi$ along transport geodesics.

Eventually, we want to highlight an important application to the present paper to the theory of Ricci bounds from below for metric measure spaces. It is well known [29] that *LSV* spaces, while stable under Gromov-Hausdorff convergence and consistent with the smooth Riemannian case, include also Finsler geometries. It is therefore natural to look for additional axioms, still stable and consistent, that rule out these geometries, thus getting a finer description of Gromov-Hausdorff limits of Riemannian manifolds. In [3] we prove, relying in particular on the results obtained in Section 6, Section 9 and Section 10 of this paper, that *LSV* spaces whose associated heat flow is linear have this stability property. In addition, we show that *LSV* bounds and linearity of the heat flow are equivalent to a single condition, namely the existence of solutions to the Wasserstein gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ in the *EVI* sense, implying nice contraction and regularization properties of the flow; we call these *Riemannian* lower bounds on Ricci curvature. Finally, for this stronger notion we provide good tensorization and localization properties.

**Acknowledgement.** The authors acknowledge the support of the ERC ADG GeMeThNES.

## 2 Preliminary notions

In this section we introduce the basic metric, topological and measure-theoretic concepts used in the paper.

### 2.1 Extended metric and Polish spaces

In this paper we consider metric spaces whose distance function may attain the value $\infty$, we call them *extended metric spaces.*

**Definition 2.1 (Extended distance and extended metric spaces)** *An extended distance on $X$ is a map $\mathsf{d}: X^2 \to [0,\infty]$ satisfying*

$$\begin{aligned}
&\mathsf{d}(x,y) = 0 \qquad \text{if and only if } x = y,\\
&\mathsf{d}(x,y) = \mathsf{d}(y,x) \qquad \forall x,\, y \in X,\\
&\mathsf{d}(x,y) \le \mathsf{d}(x,z) + \mathsf{d}(z,y) \qquad \forall x,\, y,\, z \in X.
\end{aligned}$$

*If $\mathsf{d}$ is an extended distance on $X$, we call $(X,\mathsf{d})$ an extended metric space.*

Most of the definitions concerning metric spaces generalize verbatim to extended metric spaces, since extended metric spaces can be written as a disjoint union of metric spaces, which are simply defined as

$$X_{[x]} := \big\{y \in X : \mathsf{d}(y,x) < \infty\big\}, \qquad x \in X. \tag{2.1}$$

For instance it makes perfectly sense to speak about a complete or length extended metric space.

**Definition 2.2 ($\mathsf{d}$-Lipschitz functions and Lipschitz constant)** *We say that $f : X \to \mathbb{R}$ is $\mathsf{d}$-Lipschitz if there exists $C \geq 0$ satisfying*

$$|f(x) - f(y)| \leq C\mathsf{d}(x, y) \qquad \forall\, x,\, y \in X.$$

*The least constant $C$ with this property will be denoted by* $\mathrm{Lip}(f)$*.*

In our framework the roles of the distance $\mathsf{d}$ (used to define optimal transport) and of the topology are distinct. This justifies the following definition. Recall that a topological space $(X, \tau)$ is said to be *Polish* if $\tau$ is induced by a complete and separable distance.

**Definition 2.3 (Polish extended spaces)** *We say that $(X, \tau, \mathsf{d})$ is a Polish extended space if:*

*(i) $\tau$ is a topology on $X$ and $(X, \tau)$ is Polish;*

*(ii) $\mathsf{d}$ is an extended distance on $X$ and $(X, \mathsf{d})$ is a complete extended metric space;*

*(iii) For $(x_h) \subset X$, $x \in X$, $\mathsf{d}(x_h, x) \to 0$ implies $x_h \to x$ w.r.t. to the topology $\tau$;*

*(iv) $\mathsf{d}$ is lower semicontinuous in $X \times X$, with respect to the $\tau \times \tau$ topology.*

In the sequel, when $\mathsf{d}$ is not explicitly mentioned, all the topological notions (in particular the class of compact sets, the class of Borel sets $\mathscr{B}(X)$, the class $C_b(X)$ of bounded continuous functions and the class $\mathscr{P}(X)$ of Borel probability measures) are always referred to the topology $\tau$, even when $\mathsf{d}$ is a distance. When $(X, \mathsf{d})$ is separable (thus any $\mathsf{d}$-open set is a countable union of $\mathsf{d}$-closed balls, which are also $\tau$-closed by (iv)), then a subset of $X$ is $\mathsf{d}$-Borel if and only if it is $\tau$-Borel, but when $(X, \mathsf{d})$ is not separable $\mathscr{B}(X)$ can be a strictly smaller class than the Borel sets generated by $\mathsf{d}$.

The Polish condition on $\tau$ guarantees that all Borel probability measures $\mu \in \mathscr{P}(X)$ are tight, a property (shared with the more general class of Radon spaces, see e.g. [4, Def. 5.1.4]) which justifies the introduction of the weaker topology $\tau$. In fact most of the results of the present paper could be extended to Radon spaces, thus including Lusin and Suslin topologies [33].

Notice that the only compatibility conditions between the possibly extended distance $\mathsf{d}$ and $\tau$ are (iii) and (iv). Condition (iii) guarantees that convergence in $(\mathscr{P}(X), W_2)$, as defined in Section 2.4, implies weak convergence, namely convergence in the duality with $C_b(X)$. Condition (iv) enables us, when the cost function $c$ equals $\mathsf{d}^2$, to use the standard results of the Kantorovich theory (existence of optimal plans, duality, etc.) and other useful properties, as the lower semicontinuity of the length and the $p$-energy of a curve w.r.t. pointwise $\tau$-convergence, or the representation results of [23].

An example where the roles of the distance and the topology are different is provided by bounded closed subsets of the dual of a separable Banach space: in this case $\mathsf{d}$ is the distance induced by the dual norm and $\tau$ is the weak$^*$ topology. In this case $\tau$ enjoys better compactness properties than $\mathsf{d}$.

The typical example of Polish extended space is a separable Banach space $(X, \|\cdot\|)$ endowed with a Gaussian probability measure $\gamma$. In this case $\tau$ is the topology induced by the norm and $\mathsf{d}$ is the Cameron-Martin extended distance induced by $\gamma$ (see [6]): thus, differently from $(X,\tau)$, $(X,\mathsf{d})$ is not separable if $\dim X=\infty$.

It will be technically convenient to use also the class $\mathscr{B}^*(X)$ of *universally* measurable sets (and the associated universally measurable functions): it is the $\sigma$-algebra of sets which are $\mu$-measurable for any $\mu\in\mathscr{P}(X)$.

## 2.2 Absolutely continuous curves and slopes

If $(X,\mathsf{d})$ is an extended metric space, $J\subset\mathbb{R}$ is an open interval, $p\in[1,\infty]$ and $\gamma:J\to X$, we say that $\gamma$ belongs to $\mathrm{AC}^p(J;(X,\mathsf{d}))$ if

$$\mathsf{d}(\gamma_s,\gamma_t)\le\int_s^t g(r)\,\mathrm{d}r\qquad\forall s,\,t\in J,\ s<t$$

for some $g\in L^p(J)$. The case $p=1$ corresponds to *absolutely continuous* curves, whose space is simply denoted by $\mathrm{AC}(J;(X,\mathsf{d}))$. It turns out that, if $\gamma$ belongs to $\mathrm{AC}^p(J;(X,\mathsf{d}))$, there is a minimal function $g$ with this property, called *metric derivative* and given for a.e. $t\in J$ by

$$|\dot\gamma_t|:=\lim_{s\to t}\frac{\mathsf{d}(\gamma_s,\gamma_t)}{|s-t|}.$$

See [4, Theorem 1.1.2] for the simple proof. We say that an absolutely continuous curve $\gamma_t$ has *constant speed* if $|\dot\gamma_t|$ is (equivalent to) a constant.

Notice that, by the completeness of $(X,\mathsf{d})$, $\mathrm{AC}^p(J;(X,\mathsf{d}))\subset C(\bar J;X)$, the set of $\tau$-continuous curves $\gamma:\bar J\to X$. For $t\in\bar J$ we define the evaluation map $\mathrm{e}_t:C(\bar J;X)\to X$ by

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

We endow $C(\bar J;X)$ with the sup extended distance

$$\mathsf{d}^*(\gamma,\tilde\gamma):=\sup_{t\in\bar J}\mathsf{d}(\gamma_t,\tilde\gamma_t)$$

and with the compact-open topology $\tau^*$, whose fundamental system of neighborhoods is

$$\left\{\gamma\in C(\bar J;X):\ \gamma(K_i)\subset U_i,\quad i=1,2,\ldots,n\right\},\qquad K_i\subset\bar J\text{ compact, }U_i\in\tau,\ n\ge1.$$

With these choices, it can be shown that $(C(\bar J;X),\tau^*,\mathsf{d}^*)$ inherits a Polish extended structure from $(X,\tau,\mathsf{d})$. Also, with this topology it is clear that the evaluation maps are continuous from $(C(\bar J;X),\tau^*)$ to $(X,\tau)$. Since for $p>1$ the $p$-energy

$$\mathcal{E}_p[\gamma]:=\int_J|\dot\gamma|^p\,\mathrm{d}t\quad\text{if }\gamma\in\mathrm{AC}^p(J;(X,\mathsf{d})),\quad\mathcal{E}_p[\gamma]:=\infty\quad\text{otherwise,}\tag{2.2}$$

is $\tau^*$-lower-semicontinuous thanks to (iv) of Definition 2.3, $\mathrm{AC}^p(J;(X,\mathsf{d}))$ is a Borel subset of $C(\bar J;X)$. It is not difficult to check that $\mathrm{AC}(J;(X,\mathsf{d}))$ is a Borel set as well; indeed, denoting $J=(a,b)$ and defining

$$\mathsf{TV}\big(\gamma,(a,s)\big):=\sup\left\{\sum_{i=0}^{n-1}\mathsf{d}(\gamma_{t_{i+1}},\gamma_{t_i}):\ n\in\mathbb{N},\ a<t_0<\cdots<t_n<s\right\}\qquad s\in(a,b],$$

it can be immediately seen that $\mathsf{TV}\big(\gamma,(a,s)\big)$ is lower semicontinuous in $\gamma$ and nonincreasing in $s$. Also, a continuous $\gamma$ is absolutely continuous iff the Stieltjes measure associated to $\mathsf{TV}\big(\gamma,(a,\cdot)\big)$ is absolutely continuous w.r.t. $\mathscr{L}^1$; by an integration by parts, this can be characterized in terms of $m_\varepsilon(\gamma)\downarrow 0$ as $\varepsilon\downarrow 0$, where

$$m_\varepsilon(\gamma) := \sup\left\{\int_a^b \mathsf{TV}\big(\gamma,(a,s)\big)\psi'(s)\,\mathrm{d}s : \psi\in C^1_c(a,b),\ \max|\psi|\le 1,\ \int_a^b |\psi(s)|\,\mathrm{d}s\le\varepsilon\right\} \tag{2.3}$$

if $\mathsf{TV}\big(\gamma,(a,b)\big)$ is finite, $m_\varepsilon(\gamma)=+\infty$ otherwise. Since $m_\epsilon$ are Borel in $C(\bar J;X)$, thanks to the separability of $C^1_c(a,b)$ w.r.t. the $C^1$ norm, the Borel regularity of $\mathrm{AC}(J;(X,\mathsf{d}))$ follows.

We call $(X,\mathsf{d})$ *a geodesic space* if for any $x_0,\,x_1\in X$ with $\mathsf{d}(x_0,x_1)<\infty$ there exists a curve $\gamma:[0,1]\to X$ satisfying $\gamma_0=x_0$, $\gamma_1=x_1$ and

$$\mathsf{d}(\gamma_s,\gamma_t)=|t-s|\mathsf{d}(\gamma_0,\gamma_1)\qquad\forall s,\,t\in[0,1]. \tag{2.4}$$

We will denote by $\mathrm{Geo}(X)$ the space of all constant speed geodesics $\gamma:[0,1]\to X$, namely $\gamma\in\mathrm{Geo}(X)$ if (2.4) holds. Given $f:X\to\overline{\mathbb{R}}$ we define its effective domain $D(f)$ by

$$D(f):=\{x\in X:\ f(x)\in\mathbb{R}\}\,. \tag{2.5}$$

Given $f:X\to\overline{\mathbb{R}}$ and $x\in D(f)$, we define *the local Lipschitz constant at $x$* by

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

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

$$|\nabla^- f|(x):=\limsup_{y\to x}\frac{[f(y)-f(x)]^-}{\mathsf{d}(y,x)},\qquad |\nabla^+ f|(x):=\limsup_{y\to x}\frac{[f(y)-f(x)]^+}{\mathsf{d}(y,x)}. \tag{2.6}$$

When $x\in D(f)$ is an isolated point of $X$, we set $|\nabla f|(x)=|\nabla^- f|(x)=|\nabla^+ f|(x):=0$, while all slopes are conventionally set to $+\infty$ on $X\setminus D(f)$.

Notice that for all $x\in D(f)$ it holds

$$|\nabla f|(x)=\max\big\{|\nabla^- f|(x),|\nabla^+ f|(x)|\big\},\qquad |\nabla^- f|(x)=|\nabla^+(-f)|(x). \tag{2.7}$$

Also, for $f,\,g:X\to\overline{\mathbb{R}}$ it is not difficult to check that

$$|\nabla(\alpha f+\beta g)|\le|\alpha||\nabla f|+|\beta||\nabla g|,\qquad\forall\alpha,\beta\in\mathbb{R} \tag{2.8a}$$
$$|\nabla(fg)|\le|f||\nabla g|+|g||\nabla f| \tag{2.8b}$$

on $D(f)\cap D(g)$. Also, if $\chi:X\to[0,1]$, it holds

$$|\nabla^\pm(\chi f+(1-\chi)g)|\le\chi|\nabla^\pm f|+(1-\chi)|\nabla^\pm g|+|\nabla\chi|\,|f-g|. \tag{2.9}$$

Indeed, adding the identities

$$\chi(y)f(y)-\chi(x)f(x)=\chi(y)(f(y)-f(x))+f(x)(\chi(y)-\chi(x)),$$
$$\tilde\chi(y)g(y)-\tilde\chi(x)g(x)=\tilde\chi(y)(g(y)-g(x))+g(x)(\tilde\chi(y)-\tilde\chi(x))$$

with $\tilde\chi = 1-\chi$ one obtains

$$\frac{(\chi f+\tilde\chi g)(y)-(\chi f+\tilde\chi g)(x)}{\mathsf{d}(y,x)} = \chi(y)\frac{f(y)-f(x)}{\mathsf{d}(y,x)} + \tilde\chi(y)\frac{g(y)-g(x)}{\mathsf{d}(y,x)} + \frac{\chi(y)-\chi(x)}{\mathsf{d}(x,y)}\left(f(x)-g(x)\right)$$

from which the inequality readily follows by taking the positive or negative parts and letting $y\to x$. We shall also need the measurability of slopes, ensured by the following lemma.

**Lemma 2.4** *If $f : X \to \overline{\mathbb{R}}$ is Borel, then its slopes $|\nabla^{\pm} f|$ (and therefore $|\nabla f|$) are $\mathscr{B}^*(X)$-measurable in $D(f)$. In particular, if $\gamma : [0,1] \to X$ is a continuous curve with $\gamma_t \in D(f)$ for a.e. $t \in [0,1]$, then the ($\mathscr{L}^1$-almost everywhere defined) functions $|\nabla^{\pm} f|\circ\gamma$ are Lebesgue measurable.*

*Proof.* By (2.7) it is sufficient to consider the case of the ascending slope and, since the functions

$$G_r(x) := \sup_{\{y:\ 0<\mathsf{d}(x,y)<r\}} \frac{(f(y)-f(x))^+}{\mathsf{d}(y,x)}$$

(with the convention $\sup\emptyset = 0$, so that $G_r(x) = 0$ for $r$ small enough if $x$ is an isolated point) monotonically converge to $|\nabla^+ f|$ on $D(f)$, it is sufficient to prove that $G_r$ is universally measurable for any $r > 0$. For any $r > 0$ and $\alpha \geq 0$ we see that the set

$$\{x \in D(f) :\ G_r(x) > \alpha\}$$

is the projection on the first factor of the Borel set

$$\{(x,y) \in D(f)\times X :\ f(y)-f(x) > \alpha\mathsf{d}(x,y),\quad 0 < \mathsf{d}(x,y) < r\},$$

so it is a Suslin set (see [7, Proposition 1.10.8]) and therefore it is universally measurable (see [7, Theorem 1.10.5]).

To check the last statement of the lemma it is sufficient to recall [12, Remark 32 (c2)] that a continuous curve $\gamma$ is $(\mathscr{B}^*([0,1]),\mathscr{B}^*(X))$ measurable, since any set in $\mathscr{B}^*(X)$ is measurable for all images of measures $\mu \in \mathscr{P}([0,1])$ under $\gamma$. □

Finally, for completeness we include the simple proof of the fact that $|\nabla^- f| = |\nabla^+ f|$ $\mathfrak{m}$-a.e. if $\mathsf{d}$ is finite, $f$ is $\mathsf{d}$-Lipschitz and $(X,\mathsf{d},\mathfrak{m})$ is a doubling metric measure space. We will be able to prove a weaker version of this result even in non-doubling situations, see Remark 6.4.

**Proposition 2.5** *If $\mathsf{d}$ is finite, and $(X,\mathsf{d},\mathfrak{m})$ is doubling, for all $\mathsf{d}$-Lipschitz $f : X \to \mathbb{R}$, $|\nabla^- f| = |\nabla^+ f|$ $\mathfrak{m}$-a.e. in $X$.*

*Proof.* Let $\alpha' > \alpha > 0$ and consider the set $H := \{|\nabla^- f| \leq \alpha\}$. Let $H_m$ be the subset of points $x$ such that $f(x)-f(y) \leq \alpha'\mathsf{d}(x,y)$ for all $y$ satisfying $\mathsf{d}(x,y) < 1/m$. By the doubling property, the equality $H = \cup_m H_m$ ensures that $\mathfrak{m}$-a.e. $x \in H$ is a point of density 1 for some set $H_m$. If we fix $\bar x$ with this property and $\mathsf{d}(x_n,\bar x) \to 0$, we can estimate

$$f(x_n)-f(\bar x) = f(x_n)-f(y_n)+f(y_n)-f(\bar x) \leq \mathrm{Lip}(f)\mathsf{d}(x_n,y_n)+\alpha'\mathsf{d}(y_n,\bar x)$$

choosing $y_n \in H_m \cap B_{1/m}(\bar x)$. But, thanks to the doubling property, since the density of $H_m$ at $\bar x$ is 1 we can choose $y_n$ in such a way that $\mathsf{d}(x_n,y_n) = o(\mathsf{d}(x_n,\bar x))$. Indeed, if for some

$\delta > 0$ the ball $B_{\delta \mathsf{d}(x_n,\bar x)}(x_n)$ does not intersect $H_m$ for infinitely many $n$, the upper density of $X \setminus H_m$ in the balls $B_{\mathsf{d}(x_n,\bar x)}(\bar x)$ is strictly positive. Dividing both sides by $\mathsf{d}(x_n,\bar x)$ the arbitrariness of the sequence $(x_n)$ yields $|\nabla^+ f|(\bar x) \le \alpha'$.

Since $\alpha$ and $\alpha'$ are arbitrary we conclude that $|\nabla^+ f| \le |\nabla^- f|$ $\mathfrak{m}$-a.e. in $X$. The proof of the converse inequality is similar. □

## 2.3 Upper gradients

According to [10], we say that a function $g : X \to [0,\infty]$ is an *upper gradient* of $f : X \to \overline{\mathbb{R}}$ if, for any curve $\gamma \in \mathrm{AC}((0,1);(D(f),\mathsf{d}))$, $s \mapsto g(\gamma_s)|\dot\gamma_s|$ is measurable in $[0,1]$ (with the convention $0\cdot\infty = 0$) and

$$\left|\int_{\partial\gamma} f\right| \le \int_\gamma g, \tag{2.11}$$

Here and in the following we write $\int_{\partial\gamma} f$ for $f(\gamma_1) - f(\gamma_0)$ and $\int_\gamma g = \int_0^1 g(\gamma_s)|\dot\gamma_s|\,\mathrm{d}s$.

It is not difficult to see that if $f$ is a Borel and $\mathsf{d}$-Lipschitz function then the two slopes and the local Lipschitz constant are upper gradients. More generally, the following remark will be useful.

**Remark 2.6 (When slopes are upper gradients along a curve)** Notice that if one a priori knows that $t \mapsto f(\gamma_t)$ is absolutely continuous along a given absolutely continuous curve $\gamma : [0,1] \to D(f)$, then $|\nabla^\pm f|$ are upper gradients of $f$ along $\gamma$. Indeed, $|\nabla^\pm (f\circ\gamma)|$ are bounded from above by $|\nabla^\pm f|\circ\gamma|\dot\gamma|$ wherever the metric derivative $|\dot\gamma|$ exists; then, one uses the fact that at any differentiability point both slopes of $f\circ\gamma$ coincide with $|(f\circ\gamma)'|$. ■

The next lemma is a refinement of [4, Lemma 1.2.6]; as usual, we adopt the convention $0\cdot\infty = 0$.

**Lemma 2.7 (Absolute continuity criterion)** *Let $L \in L^1(0,1)$ be nonnegative and let $g : [0,1] \to [0,\infty]$ be a measurable map with $\int_0^1 L\,\mathrm{d}t > 0$ and $\int_0^1 g(t)L(t)\,\mathrm{d}t < \infty$. Let $w : [0,1] \to \mathbb{R}\cup\{-\infty\}$ be an upper semicontinuous map, with $w > -\infty$ a.e. on $\{L \ne 0\}$, satisfying*

$$w(s) - w(t) \le g(t)\left|\int_s^t L(r)\,\mathrm{d}r\right| \qquad \textit{for all } t \in \{w > -\infty\} \tag{2.12}$$

*and, for arbitrary $0 \le a < b \le 1$,*

$$\int_a^b L\,\mathrm{d}t = 0 \qquad \Longrightarrow \qquad w \textit{ is constant in } [a,b]. \tag{2.13}$$

*Then $\{w = -\infty\}$ is empty and $w$ is absolutely continuous in $[0,1]$.*

*Proof.* It is not restrictive to assume $\int_0^1 L(t)\,\mathrm{d}t = 1$ and set $\lambda := L\mathscr{L}^1|_{[0,1]}$. We introduce the monotone, right continuous map $\mathsf{t} : [0,1] \to [0,1]$ pushing $\mathscr{L}^1$ onto $\lambda$: setting

$$\mathsf{x}(t) := \int_0^t L(r)\,\mathrm{d}r = \lambda([0,t]) \quad \text{it holds} \quad \mathsf{t}(x) := \sup\{t \in [0,1] :\ \mathsf{x}(t) \le x\},$$

and considering the function $\tilde g := g\circ\mathsf{t}$ we easily get

$$\int_{\mathsf{t}(x)}^{\mathsf{t}(y)} L(r)\,\mathrm{d}r = |x-y| \quad 0 \le x \le y \le 1, \quad \int_a^b g(t)L(t)\,\mathrm{d}t = \int_{\mathsf{x}(a)}^{\mathsf{x}(b)} \tilde g(z)\,\mathrm{d}z \quad a,\, b \in [0,1], \tag{2.14}$$

so that, defining also $\tilde w := w \circ \mathsf{t}$, (2.12) becomes

$$\tilde w(y) - \tilde w(x) \le \tilde g(x)|x-y| \quad \text{for all } x \in \{\tilde w > -\infty\}. \tag{2.15}$$

Notice that $\tilde w$ is still upper semicontinuous: since it is the composition of an upper semicontinuous function with the increasing right continuous map $\mathsf{t}$, we have just to check this property at the jump set of $\mathsf{t}$. If $x \in (0,1]$ satisfies $\mathsf{t}_-(x) = \lim_{y\uparrow x} \mathsf{t}(y) < \mathsf{t}(x)$, since $w$ is constant in $[\mathsf{t}_-(x), \mathsf{t}(x)]$ we have

$$\limsup_{y\uparrow x} \tilde w(y) = \limsup_{s\uparrow \mathsf{t}_-(x)} w(s) \le w(\mathsf{t}_-(x)) = w(\mathsf{t}(x)) = \tilde w(x).$$

In particular $\tilde w$ is bounded from above and choosing $y_0$ such that $w(y_0) > -\infty$ we get $\tilde w(x) \ge \tilde w(y_0) - \tilde g(x)$ for every $x \in \{\tilde w > -\infty\}$, so that $\tilde w$ is integrable. Since

$$|\tilde w(y) - \tilde w(x)| \le (\tilde g(x) + \tilde g(y))|x-y| \quad \text{for every } x, y \in (0,1) \setminus \{\tilde w > -\infty\}$$

applying [4, Lemma 2.1.6] we obtain that $\tilde w \in W^{1,1}(0,1)$ and $|\tilde w'| \le 2\tilde g$ a.e. in $(0,1)$.

Since $\tilde w \in W^{1,1}(0,1)$ there exists a continuous representative $\bar w$ of $\tilde w$ in the Lebesgue equivalence class of $\tilde w$. Since any point in $[0,1]$ can be approximated by points in the coincidence set we obtain that $\tilde w \ge \bar w > -\infty$ in $[0,1]$. We can apply (2.15) to obtain (in the case $y < 1$)

$$\tilde w(y) - \frac{1}{h}\int_y^{y+h} \tilde w(x)\,\mathrm{d}x \le \int_y^{y+h} \tilde g(x)\,\mathrm{d}x \to 0 \quad \text{as } h \downarrow 0.$$

Since $\int_y^{y+h} \tilde w(x)\,\mathrm{d}x \sim h\bar w(y)$ as $h \downarrow 0$, we obtain the opposite inequality $\tilde w(y) \le \bar w(y)$ for every $y \in [0,1)$. In the case $y = 1$ the argument is similar.

We thus obtain $|\tilde w(x) - \tilde w(y)| \le 2\int_x^y \tilde g(z)\,\mathrm{d}r$ for every $0 \le x \le y \le 1$. Now, the fact that $w$ is constant in any closed interval where $\mathsf{x}$ is constant ensures the validity of the identity $w(s) = w(\mathsf{t}(\mathsf{x}(s)))$, so that $w(s) = \tilde w(\mathsf{x}(s))$ and the second equality in (2.14) yields

$$|w(s) - w(t)| = |\tilde w(\mathsf{x}(s)) - \tilde w(\mathsf{x}(t))| \le 2\int_{\mathsf{x}(s)}^{\mathsf{x}(t)} \tilde g(z)\,\mathrm{d}z = 2\int_s^t g(r)L(r)\,\mathrm{d}r \quad 0 \le s < t \le 1.$$

□

**Corollary 2.8** *Let $\gamma \in \mathrm{AC}([0,1];(X,\mathsf{d}))$ and let $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ be a $\mathsf{d}$-upper semicontinuous map such that $\varphi(\gamma_s) > -\infty$ a.e. in $[0,1]$. Let $g : X \to [0,\infty]$ be such that $g \circ \gamma|\dot\gamma| \in L^1(0,1)$ and*

$$\varphi(\gamma_s) - \varphi(\gamma_t) \le g(\gamma_t)\left|\int_s^t |\dot\gamma|(r)\,\mathrm{d}r\right| \quad \textit{for all } t \textit{ such that } \varphi(\gamma_t) > -\infty. \tag{2.16}$$

*Then the map $s \mapsto \varphi(\gamma_s)$ is real valued and absolutely continuous.*

The *proof* is an immediate application of Lemma 2.7 with $L := |\dot\gamma|$ and $w := \varphi \circ \gamma$; (2.13) is true since $\gamma$ (and thus $\varphi \circ \gamma$) is constant on every interval where $|\dot\gamma|$ vanishes a.e.

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

Here we assume that $(X, \tau, \mathsf{d})$ is a Polish extended space. Given $\mu$, $\nu \in \mathscr{P}(X)$, we define the Wasserstein distance $W_2$ between them as

$$W_2^2(\mu, \nu) := \inf \int_{X \times X} \mathsf{d}^2(x, y) \, \mathrm{d}\boldsymbol{\gamma}(x, y), \tag{2.17}$$

where the infimum is taken among all $\boldsymbol{\gamma} \in \mathscr{P}(X \times X)$ such that

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

Such measures are called admissible plans (or couplings) for the pair $(\mu, \nu)$. As usual, if $\mu \in \mathscr{P}(X)$ and $T : X \to Y$ is a $\mu$-measurable map with values in the topological space $Y$, the push-forward measure $T_\sharp \mu \in \mathscr{P}(Y)$ is defined by $T_\sharp \mu(B) := \mu(T^{-1}(B))$ for every set $B \in \mathscr{B}(Y)$.

We are not restricting ourselves to the space of measures with finite second moments, so that it can possibly happen that $W_2(\mu, \nu) = \infty$. Still, via standard arguments one can prove that $W_2$ is an extended distance in $\mathscr{P}(X)$. Also, we point out that if we define

$$\mathscr{P}_{[\mu]}(X) := \Big\{ \nu \in \mathscr{P}(X) \ : \ W_2(\mu, \nu) < \infty \Big\}$$

for some $\mu \in \mathscr{P}(X)$, then the space $(\mathscr{P}_{[\mu]}(X), W_2)$ is actually a complete metric space (which reduces to the standard one $(\mathscr{P}_2(X), W_2)$ if $\mu$ is a Dirac mass and $\mathsf{d}$ is finite).

Concerning the relation between $W_2$ convergence and weak convergence, the implication

$$W_2(\mu_n, \mu) \to 0 \qquad \Longrightarrow \qquad \int_X \varphi \, \mathrm{d}\mu_n \to \int_X \varphi \, \mathrm{d}\mu \quad \forall \varphi \in C_b(X) \tag{2.18}$$

is well known if $(X, \mathsf{d})$ is a metric space and $\tau$ is induced by the distance $\mathsf{d}$, see for instance [4, Proposition 7.1.5]; the implication remains true in our setting, with the same proof, thanks to the compatibility condition (iii) of Definition 2.3.

Since $\mathsf{d}^2$ is $\tau$-lower semicontinuous, when $W_2(\mu, \nu) < \infty$ the infimum in the definition (2.17) of $W_2^2$ is attained and we call optimal all the plans $\boldsymbol{\gamma}$ realizing the minimum; Kantorovich's duality formula holds:

$$\frac{1}{2} W_2^2(\mu, \nu) = \sup \left\{ \int_X \varphi \, \mathrm{d}\mu + \int_X \psi \, \mathrm{d}\nu : \ \varphi(x) + \psi(y) \leq \frac{1}{2} \mathsf{d}^2(x, y) \right\}, \tag{2.19}$$

where the functions $\varphi$ and $\psi$ in the supremum are respectively $\mu$-measurable and $\nu$-measurable, and in $L^1$. One can also restrict, without affecting the value of the supremum, to bounded and continuous functions $\varphi$, $\psi$ (see [4, Theorem 6.1.1]).

Recall that the *c-transform* $\varphi^c$ of $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is defined by

$$\varphi^c(y) := \inf \left\{ \frac{\mathsf{d}^2(x, y)}{2} - \varphi(x) : \ x \in X \right\}$$

and that $\psi$ is said to be *c-concave* if $\psi = \varphi^c$ for some $\varphi$.

$c$-concave functions are always $\mathsf{d}$-upper semicontinuous, hence Borel in the case when $\mathsf{d}$ is finite and induces $\tau$. More generally, it is not difficult to check that

$$\varphi \ \text{ Borel} \qquad \Longrightarrow \qquad \varphi^c \ \ \mathscr{B}^*(X)\text{-measurable}. \tag{2.20}$$

The proof follows, as in Lemma 2.4, from Suslin's theory: indeed, the set $\{\varphi^c < \alpha\}$ is the projection on the second coordinate of the Borel set of points $(x, y)$ such that $\mathsf{d}^2(x, y)/2 - \varphi(x) < \alpha$, so it is a Suslin set and therefore universally measurable.

If $\varphi(x)$, $\psi(y)$ satisfy $\varphi(x) + \psi(y) \le \mathsf{d}^2(x, y)/2$, since $\varphi^c \ge \psi$ still satisfies $\varphi + \varphi^c \le \mathsf{d}^2/2$ and since we may restrict ourselves to bounded continuous functions, we obtain

$$\frac{1}{2}W_2^2(\mu, \nu) = \sup\left\{\int_X \varphi \,\mathrm{d}\mu + \int_X \varphi^c \,\mathrm{d}\nu : \ \varphi \in C_b(X)\right\}. \tag{2.21}$$

**Definition 2.9 (Kantorovich potential)** *Assume that* $\mathsf{d}$ *is a finite distance. We say that a map* $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ *is a Kantorovich potential relative to an optimal plan* $\boldsymbol{\gamma}$ *if:*

*(i)* $\varphi$ *is c-concave, not identically equal to* $-\infty$ *and Borel;*

*(ii)* $\varphi(x) + \varphi^c(y) = \frac{1}{2}\mathsf{d}^2(x, y)$ *for* $\boldsymbol{\gamma}$*-a.e.* $(x, y) \in X \times X$.

Since $\varphi$ is not identically equal to $-\infty$ the function $\varphi^c$ still takes values in $\mathbb{R} \cup \{-\infty\}$ and the $c$-concavity of $\varphi$ ensures that $\varphi = (\varphi^c)^c$. Notice that we are not requiring integrability of $\varphi$ and $\varphi^c$, although condition (ii) forces $\varphi$ (resp. $\varphi^c$) to be finite $\mu$-a.e. (resp. $\nu$-a.e.).

The existence of maximizing pairs in the duality formula can be a difficult task if $\mathsf{d}$ is unbounded, and no general result is known when $\mathsf{d}$ may attain the value $\infty$. For this reason we restrict ourselves to finite distances $\mathsf{d}$ in the previous definition and in the next proposition, concerning the main existence and integrability result for Kantorovich potentials.

**Proposition 2.10 (Existence of Kantorovich potentials)** *If* $\mathsf{d}$ *is finite and* $\boldsymbol{\gamma}$ *is an optimal plan with finite cost, then Kantorovich potentials* $\varphi$ *relative to* $\boldsymbol{\gamma}$ *exist. In addition, if* $\mathsf{d}(x, y) \le a(x) + b(y)$ *with* $a \in L^2(X, \mu)$ *and* $b \in L^2(Y, \nu)$*, the functions* $\varphi$, $\varphi^c$ *are respectively* $\mu$*-integrable and* $\nu$*-integrable and provide maximizers in the duality formula* (2.19). *In this case* $\varphi$ *is a Kantorovich potential relative to any optimal plan* $\boldsymbol{\gamma}$.

*Proof.* Existence of $\varphi$ follows by a well-known argument, see for instance [4, Theorem 6.1.4]: one makes the Rüschendorf-Rockafellar construction of a $c$-concave function $\varphi$ starting from a $\sigma$-compact and $\mathsf{d}^2$-monotone set $\Gamma$ on which $\boldsymbol{\gamma}$ is concentrated. The last statement follows by

$$\frac{1}{2}W_2^2(\mu, \nu) = \int_X \varphi \,\mathrm{d}\mu + \int_X \varphi^c \,\mathrm{d}\nu \le \int_{X \times X} \frac{1}{2}\mathsf{d}^2 \,\mathrm{d}\boldsymbol{\gamma}$$

for any admissible plan $\boldsymbol{\gamma}$. □

## 2.5 Geodesically convex functionals and gradient flows

Given an extended metric space $(Y, \mathsf{d}_Y)$ (in the sequel it will mostly be a Wasserstein space) and $K \in \mathbb{R}$, a functional $E : Y \to \mathbb{R} \cup \{+\infty\}$ is said to be $K$-geodesically convex if for any $y_0$, $y_1 \in D(E)$ with $\mathsf{d}_Y(y_0, y_1) < \infty$ there exists $\gamma \in \mathrm{Geo}(Y)$ such that $\gamma_0 = y_0$, $\gamma_1 = y_1$ and

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

A consequence of $K$-geodesic convexity is that the descending slope defined in (2.6) can be calculated as

$$|\nabla^- E|(y) = \sup_{z \in Y \setminus \{y\}} \left(\frac{E(y) - E(z)}{\mathsf{d}_Y(y, z)} + \frac{K}{2}\mathsf{d}_Y(y, z)\right)^+, \tag{2.22}$$

so that $|\nabla^- E|(y)$ is the smallest constant $S \geq 0$ such that

$$E(z) \geq E(y) - S\mathsf{d}_Y(z,y) + \frac{K}{2}\mathsf{d}_Y^2(z,y) \quad \text{for every } z \in Y_{[y]}. \tag{2.23}$$

We recall (see [4, Corollary 2.4.10]) that for $K$-geodesically convex functionals the descending slope is a upper gradient, as defined in Section 2.3: in particular

$$E(y_t) \geq E(y_s) - \int_s^t |\dot{y}_r|\,|\nabla^- E|(y_r)\,\mathrm{d}r \qquad \text{for every } s,t \in [0,\infty),\ s<t \tag{2.24}$$

for all locally absolutely continuous curves $y : [0,\infty) \to D(E)$. A metric gradient flow for $E$ is a locally absolutely continuous curve $y : [0,\infty) \to D(E)$ along which (2.24) holds as an equality and moreover $|\dot{y}_t| = |\nabla^- E|(y_t)$ for a.e. $t \in (0,\infty)$.

An application of Young inequality shows that gradient flows for functionals can be characterized by the following definition.

**Definition 2.11 ($E$-dissipation inequality and metric gradient flow)** *Let $E : Y \to \mathbb{R} \cup \{+\infty\}$ be a functional. We say that a locally absolutely continuous curve $[0,\infty) \ni t \mapsto y_t \in D(E)$*

*satisfies the* $E$-dissipation inequality *if*

$$E(y_0) \geq E(y_t) + \frac{1}{2}\int_0^t |\dot{y}_r|^2\,\mathrm{d}r + \frac{1}{2}\int_0^t |\nabla^- E|^2(y_r)\,\mathrm{d}r \qquad \forall t \geq 0. \tag{2.25}$$

*$y$ is a gradient flow of $E$ starting from $y_0 \in D(E)$ if* (2.25) *holds as an equality, i.e.*

$$E(y_0) = E(y_t) + \frac{1}{2}\int_0^t |\dot{y}_r|^2\,\mathrm{d}r + \frac{1}{2}\int_0^t |\nabla^- E|^2(y_r)\,\mathrm{d}r \qquad \forall t \geq 0. \tag{2.26}$$

By the remarks above, it is not hard to check that (2.26) is equivalent to the $E$-dissipation inequality (2.25) whenever $t \mapsto E(y_t)$ is absolutely continuous, in particular if $|\nabla^- E|$ is an upper gradient of $E$ (as for $K$-geodesically convex functionals). In this case (2.26) is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}t}E(y_t) = -|\dot{y}_t|^2 = -|\nabla^- E|^2(y_t) \quad \text{for a.e. } t \in (0,\infty). \tag{2.27}$$

If $E : \mathbb{R}^d \to \mathbb{R}$ is a smooth functional, then a $C^1$ curve $(y_t)$ is a gradient flow according to the previous definition if and only if it satisfies $y_t' = -\nabla E(y_t)$ for all $t \in (0,\infty)$, so that the metric definition reduces to the classical one when specialized to Euclidean spaces and to regular curves and functionals.

# 3 Hopf-Lax semigroup in metric spaces

In this section we study the properties of the functions given by Hopf-Lax formula in a metric setting and the relations with the Hamilton-Jacobi equation. Here we only assume that $(X,\mathsf{d})$ is an extended metric space until Theorem 3.5 (in particular, $(X,\mathsf{d})$ is not necessarily $\mathsf{d}$-complete or $\mathsf{d}$-separable) and the measure structure $(X,\tau,\mathfrak{m})$ does not play a role, except in Proposition 3.8 and Proposition 3.9. After Theorem 3.5 we will also assume that our space is a length space.

Let $(X, \mathsf{d})$ be an extended metric space and $f : X \to \mathbb{R} \cup \{+\infty\}$. We define

$$F(t, x, y) := f(y) + \frac{\mathsf{d}^2(x, y)}{2t}, \tag{3.1}$$

and

$$Q_t f(x) := \inf_{y \in X} F(t, x, y) \qquad (x, t) \in X \times (0, \infty). \tag{3.2}$$

The map $(x, t) \mapsto Q_t f(x)$, $X \times (0, \infty) \to \overline{\mathbb{R}}$, is obviously $\mathsf{d}$-upper semicontinuous, The behavior of $Q_t f$ is not trivial only in the set

$$\mathcal{D}(f) := \{x \in X : \ \mathsf{d}(x, y) < \infty \text{ for some } y \text{ with } f(y) < \infty\} \tag{3.3}$$

and we shall restrict our analysis to $\mathcal{D}(f)$, so that $Q_t f(x) \in \mathbb{R} \cup \{-\infty\}$ for $(x, t) \in \mathcal{D}(f) \times (0, \infty)$. For $x \in \mathcal{D}(f)$ we set also

$$t_*(x) := \sup\{t > 0 : \ Q_t f(x) > -\infty\}$$

with the convention $t_*(x) = 0$ if $Q_t f(x) = -\infty$ for all $t > 0$. Since $Q_t f(x) > -\infty$ implies $Q_s f(y) > -\infty$ for all $s \in (0, t)$ and all $y$ at a finite distance from $x$, it follows that $t_*(x)$ depends only on the equivalence class $X_{[x]}$ of $x$, see (2.1).

Finally, we introduce the functions $D^+(x, t)$, $D^-(x, t)$ as

$$D^+(x, t) := \sup_{(y_n)} \limsup_n \mathsf{d}(x, y_n), \qquad D^-(x, t) := \inf_{(y_n)} \liminf_n \mathsf{d}(x, y_n), \tag{3.4}$$

where, in both cases, the $(y_n)$'s vary among all minimizing sequences of $F(t, x, \cdot)$. It is easy to check (arguing as in [4, Lemma 2.2.1, Lemma 3.1.2]) that $D^+(x, t)$ is finite for $0 < t < t_*(x)$ and that

$$\lim_{i \to \infty} \mathsf{d}(x_i, x) = 0, \quad \lim_{i \to \infty} t_i = t \in (0, t_*(x)) \qquad \Longrightarrow \qquad \lim_{i \to \infty} Q_{t_i} f(x_i) = Q_t f(x), \tag{3.5}$$

$$\sup \left\{D^+(y, t) : \ \mathsf{d}(x, y) \le R, \ 0 < t < t_*(x) - \varepsilon\right\} < \infty \qquad \forall R > 0, \ \varepsilon > 0. \tag{3.6}$$

Simple diagonal arguments show that the supremum and the infimum in (3.4) are attained.

Obviously $D^-(x, \cdot) \le D^+(x, \cdot)$; the next proposition shows that both functions are nonincreasing, and that they coincide out of a countable set.

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

$$D^+(x, t) \le D^-(x, s) < \infty, \qquad 0 < t < s < t_*(x). \tag{3.7}$$

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

*Proof.* Fix $x \in \mathcal{D}(f)$, $0 < t < s < t_*(x)$ and choose minimizing sequences $(x_t^n)$ and $(x_s^n)$ for $F(t, x, \cdot)$ and $F(s, x, \cdot)$ respectively, such that $\lim_n \mathsf{d}(x, x_t^n) = D^+(x, t)$ and $\lim_n \mathsf{d}(x, x_s^n) = D^-(x, s)$. As a consequence, there exist the limits of $f(x_t^n)$ and $f(x_s^n)$ as $n \to \infty$. The minimality of the sequences gives

$$\lim_n f(x_t^n) + \frac{\mathsf{d}^2(x_t^n, x)}{2t} \le \lim_n f(x_s^n) + \frac{\mathsf{d}^2(x_s^n, x)}{2t}$$
$$\lim_n f(x_s^n) + \frac{\mathsf{d}^2(x_s^n, x)}{2s} \le \lim_n f(x_t^n) + \frac{\mathsf{d}^2(x_t^n, x)}{2s}.$$

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

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

which is (3.7). 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 $D^\pm$: they 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 $D^\pm$)** *Let $x_n \xrightarrow{\mathsf{d}} x$ and $t_n \to t \in (0, t_*(x))$. Then*

$$D^-(x,t) \le \liminf_{n\to\infty} D^-(x_n,t_n), \qquad D^+(x,t) \ge \limsup_{n\to\infty} D^+(x_n,t_n).$$

*In particular, for every $x \in X$ the map $t \mapsto D^-(x,t)$ is left continuous in $(0,t_*(x))$ and the map $t \mapsto D^+(x,t)$ is right continuous in $(0,t_*(x))$.*

*Proof.* For every $n \in \mathbb{N}$, let $(y_n^i)_{i\in\mathbb{N}}$ be a minimizing sequence for $F(t_n, x_n, \cdot)$ for which the limit of $\mathsf{d}(y_n^i, x_n)$ as $i \to \infty$ equals $D^-(x_n,t_n)$. From (3.6) we see that we can assume that $\sup_{i,n} \mathsf{d}(y_n^i, x^i)$ is finite. For all $n$ we have

$$\lim_{i\to\infty} f(y_n^i) + \frac{\mathsf{d}^2(y_n^i, x_n)}{2t_n} = Q_{t_n} f(x_n).$$

Moreover, the $\mathsf{d}$-upper semicontinuity of $(x,t) \mapsto Q_t f(x)$ gives that $\limsup_n Q_{t_n} f(x_n) \le Q_t f(x)$. Since $\mathsf{d}(y_n^i, x_n)$ is bounded we have $\sup_i |\mathsf{d}^2(y_n^i, x_n) - \mathsf{d}^2(y_n^i, x)|$ is infinitesimal, hence by a diagonal argument we can find a sequence $n \mapsto i(n)$ such that

$$\limsup_{n\to\infty} f(y_n^{i(n)}) + \frac{\mathsf{d}^2(y_n^{i(n)}, x)}{2t} \le Q_t f(x), \quad \left|\mathsf{d}(x_n, y_n^{i(n)}) - D^-(x_n,t_n)\right| \le \frac{1}{n}.$$

This implies that $n \mapsto y_n^{i(n)}$ is a minimizing sequence for $F(t,x,\cdot)$, therefore

$$D^-(x,t) \le \liminf_{n\to\infty} \mathsf{d}(x, y_n^{i(n)}) = \liminf_{n\to\infty} \mathsf{d}(x_n, y_n^{i(n)}) = \liminf_{i\to\infty} D^-(x_n,t_n).$$

If we choose, instead, sequences $(y_n^i)_{i\in\mathbb{N}}$ on which the supremum in the definition of $D^+(x_n,t_n)$ is attained, we obtain the upper semicontinuity property. □

Before stating the next proposition we recall that semiconcave functions $g$ on an open interval are local quadratic perturbations of concave functions; they inherit from concave functions all pointwise differentiability properties, as existence of right and left derivatives $\frac{\mathrm{d}^-}{\mathrm{d}t} g \ge \frac{\mathrm{d}^+}{\mathrm{d}t} g$, and similar.

**Proposition 3.3 (Time derivative of $Q_t f$)** *The map $(0,t_*(x)) \ni t \mapsto Q_t f(x)$ is locally Lipschitz and locally semiconcave. For all $t \in (0,t_*(x))$ it satisfies*

$$\frac{\mathrm{d}^-}{\mathrm{d}t} Q_t f(x) = -\frac{(D^-(x,t))^2}{2t^2}, \qquad \frac{\mathrm{d}^+}{\mathrm{d}t} Q_t f(x) = -\frac{(D^+(x,t))^2}{2t^2}. \tag{3.8}$$

*In particular, $s \mapsto Q_s f(x)$ is differentiable at $t \in (0,t_*(x))$ if and only if $D^+(x,t) = D^-(x,t)$.*

*Proof.* Let $(x_t^n)$, $(x_s^n)$ be minimizing sequences for $F(t,x,\cdot)$ and $F(s,x,\cdot)$. We have

$$Q_s f(x) - Q_t f(x) \le \liminf_{n\to\infty} F(s,x,x_t^n) - F(t,x,x_t^n) = \liminf_{n\to\infty} \frac{\mathsf{d}^2(x,x_t^n)}{2}\left(\frac{1}{s}-\frac{1}{t}\right), \tag{3.9}$$

$$Q_s f(x) - Q_t f(x) \ge \limsup_{n\to\infty} F(s,x,x_s^n) - F(t,x,x_s^n) = \limsup_{n\to\infty} \frac{\mathsf{d}^2(x,x_s^n)}{2}\left(\frac{1}{s}-\frac{1}{t}\right). \tag{3.10}$$

If $s > t$ we obtain

$$\frac{(D^-(x,s))^2}{2}\left(\frac{1}{s}-\frac{1}{t}\right) \le Q_s f(x) - Q_t f(x) \le \frac{(D^+(x,t))^2}{2}\left(\frac{1}{s}-\frac{1}{t}\right); \tag{3.11}$$

recalling that $\lim_{s\downarrow t} D^-(x,s) = D^+(x,t)$, a division by $s-t$ and a limit as $s \downarrow t$ gives the identity for the right derivative in (3.8). A similar argument, dividing by $t-s<0$ and passing to the limit as $t \uparrow s$ yields the left derivative in (3.8).

The local Lipschitz continuity follows by (3.11) recalling that $D^\pm(x,\cdot)$ are locally bounded functions; we easily get the quantitative bound

$$\left\|\frac{\mathrm{d}}{\mathrm{d}t} Q_t f(x)\right\|_{L^\infty(\tau,\tau')} \le \frac{1}{2\tau^2}\|D^+(x,\cdot)\|_{L^\infty(\tau,\tau')} \qquad \text{for every } 0<\tau<\tau'<t_*(x). \tag{3.12}$$

Since the distributional derivative of the function $t \mapsto [D^+(x,t)]^2/(2t^2)$ is locally bounded from below, we also deduce that $t \mapsto Q_t f$ is locally semiconcave. $\square$

**Proposition 3.4 (Slopes and upper gradients of $Q_t f$)** *For $x \in \mathcal{D}(f)$ it holds:*

$$t \in (0, t_*(x)) \quad\Longrightarrow\quad |\nabla Q_t f|(x) \le \frac{D^+(x,t)}{t}, \tag{3.13a}$$

$$Q_t f(x) > -\infty \quad\Longrightarrow\quad |\nabla^+ Q_t f|(x) \le \frac{D^-(x,t)}{t}. \tag{3.13b}$$

*In addition, for all $t \in (0,t_*(x))$, $D^-(\cdot,t)/t$ is an upper gradient of $Q_t f$ restricted to $X_{[x]} = \{y : \mathsf{d}(x,y) < \infty\}$.*

*Proof.* Let us first prove that for arbitrary $x$, $y$ be at finite distance with $Q_t f(y) > -\infty$ we have the estimate

$$Q_t f(x) - Q_t f(y) \le \mathsf{d}(x,y)\Big(\frac{D^-(y,t)}{t} + \frac{\mathsf{d}(x,y)}{2t}\Big). \tag{3.14}$$

It is sufficient to take a minimizing sequence $(y_n)$ for $F(t,y,\cdot)$ on which the infimum in the definition of $D^-(y,t)$ is attained, obtaining

$$\begin{aligned} Q_t f(x) - Q_t f(y) &\le \liminf_{n\to\infty} F(t,x,y_n) - F(t,y,y_n) = \liminf_{n\to\infty} \frac{\mathsf{d}^2(x,y_n)}{2t} - \frac{\mathsf{d}^2(y,y_n)}{2t} \\ &\le \liminf_{n\to\infty} \frac{\mathsf{d}(x,y)}{2t}\big(\mathsf{d}(x,y_n) + \mathsf{d}(y,y_n)\big) \le \frac{\mathsf{d}(x,y)}{2t}\big(\mathsf{d}(x,y) + 2D^-(y,t)\big). \end{aligned}$$

Dividing both sides of (3.14) by $\mathsf{d}(x,y)$ and taking the $\limsup$ as $y \to x$ we get (3.13a) for the descending slope, since Proposition 3.2 yields the upper-semicontinuity of $D^+$. The implication (3.13b) follows by the same argument, by inverting the role of $x$ and $y$ in (3.14) and still taking the lim sup as $y \to x$ after a division by $\mathsf{d}(x,y)$. The complete inequality in (3.13a) follows by (2.7).

We conclude with the proof of the upper gradient property. Let $t \in (0, t_*(x))$, let $\gamma : [0,1] \to X_{[x]}$ be an absolutely continuous curve with constant speed (this is not restrictive, up to a reparameterizazion), and notice that (3.5) gives that $Q_t f(\gamma_s)$ is continuous in $[0,1]$ whereas Proposition 3.2 shows the upper-semicontinuity (and thus the measurability) of $D^-(\gamma_s, t)$. By applying (3.14) with $x = \gamma_{s'}$, $y = \gamma_s$, if $s \mapsto D^-(\gamma_s, t) \in L^1(0,1)$ we can use Corollary 2.8 to obtain that $s \mapsto Q_t f(\gamma_s)$ is absolutely continuous. Coming back to (3.14) we obtain that $|\frac{\mathrm{d}}{ds} Q_t f(\gamma_s)| \le D^-(\gamma_s, t)/t$ for a.e. $s \in [0,1]$. □

**Theorem 3.5 (Subsolution of HJ)** *For $x \in \mathcal{D}(f)$ and $t \in (0, t_*(x))$ the right and left derivatives $\frac{\mathrm{d}^\pm}{\mathrm{d}t} Q_t f(x)$ satisfy*

$$\frac{\mathrm{d}^+}{\mathrm{d}t} Q_t f(x) + \frac{|\nabla Q_t f|^2(x)}{2} \le 0, \qquad \frac{\mathrm{d}^-}{\mathrm{d}t} Q_t f(x) + \frac{|\nabla^+ Q_t f|^2(x)}{2} \le 0.$$

*In particular*

$$\frac{\mathrm{d}}{\mathrm{d}t} Q_t f(x) + \frac{|\nabla Q_t f|^2(x)}{2} \le 0 \tag{3.15}$$

*with at most countably many exceptions in $(0, t_*(x))$.*

*Proof.* The first claim is a direct consequence of Propositions 3.3 and 3.4. The second one (3.15) follows by the fact that the larger derivative, namely the left one, coincides with $-[D^-(x,t)]^2/(2t^2)$, and then with $-[D^+(x,t)]^2/(2t^2)$ with at most countably many exceptions. The latter is smaller than $-|\nabla Q_t f|^2(x)/2$ by (3.13a). □

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

We say that $(X, \mathsf{d})$ is a *length* space if for all $x$, $y \in X$ the infimum of the length of continuous curves joining $x$ to $y$ is equal to $\mathsf{d}(x,y)$. We remark that under this assumption it can be proved that the Hopf-Lax formula produces a semigroup (see for instance the proof in [26]), while in general only the inequality $Q_{s+t} f \le Q_s(Q_t f)$ holds.

**Proposition 3.6 (Solution of HJ and agreement of slopes)** *Assume that $(X, \mathsf{d})$ is a length space. Then for all $x \in \mathcal{D}(f)$ and $t \in (0, t_*(x))$ it holds*

$$|\nabla^- Q_t f|(x) = |\nabla Q_t f|(x) = \frac{D^+(x,t)}{t}, \tag{3.16}$$

*so that equality holds in (3.13a). In particular, the right time derivative of $Q_t f$ satisfies*

$$\frac{\mathrm{d}^+}{\mathrm{d}t} Q_t f(x) + \frac{|\nabla Q_t f|^2(x)}{2} = 0 \quad \text{for every } t \in (0, t_*(x)), \tag{3.17}$$

*and equality holds in (3.15), with at most countably many exceptions.*

*Proof.* Let $(y_i)$ be a minimizing sequence for $F(t, x, \cdot)$ on which the supremum in the definition of $D^+(x,t)$ is attained. Let $\gamma^i : [0,1] \to X$ be continuous curves connecting $x$ to $y_i$ whose lengths $\mathcal{L}(\gamma^i)$ converge to $D^+(x,t)$. For every $s \in (0,1)$ we have

$$\begin{aligned}\limsup_{i\to\infty} Q_t f(x) - Q_t f(\gamma^i_s) &\geq \limsup_{i\to\infty} F(t,x,y_i) - F(t,\gamma^i_s,y_i)\\ &= \limsup_{i\to\infty} \frac{\mathsf{d}^2(x,y_i) - \mathsf{d}^2(\gamma^i_s,y_i)}{2t},\end{aligned}$$

and our assumption on the $\gamma^i$'s ensures that

$$\lim_{i\to\infty} \frac{\mathsf{d}(x,\gamma^i_s)}{s\mathsf{d}(x,y_i)} = 1, \qquad \lim_{i\to\infty} \frac{\mathsf{d}(\gamma^i_s,y_i)}{(1-s)\mathsf{d}(x,y_i)} = 1 \quad \text{for every } s \in (0,1).$$

Therefore we obtain

$$\begin{aligned}\limsup_{i\to\infty} \frac{Q_t f(x) - Q_t f(\gamma^i_s)}{\mathsf{d}(x,\gamma^i_s)} &\geq \limsup_{i\to\infty} \frac{\big(\mathsf{d}(x,y_i) - \mathsf{d}(\gamma^i_s,y_i)\big)\big(\mathsf{d}(x,y_i) + \mathsf{d}(\gamma^i_s,y_i)\big)}{2t\mathsf{d}(x,\gamma^i_s)}\\ &= \frac{(2-s)D^+(x,t)}{2t} \qquad \text{for all } s \in (0,1).\end{aligned}$$

With a diagonal argument we find $s(i) \downarrow 0$ such that

$$\limsup_{i\to\infty} \frac{Q_t f(x) - Q_t f(\gamma^i_{s(i)})}{\mathsf{d}(x,\gamma^i_{s(i)})} \geq \frac{D^+(x,t)}{t}.$$

Since $i \mapsto \gamma^i_{s(i)}$ is a particular sequence converging to $x$ we deduce

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

Thanks to (3.13a) and to the inequality $|\nabla^- Q_t| \leq |\nabla Q_t|$, this proves that $|\nabla^- Q_t f|(x) = |\nabla Q_t f|(x) = D^+(x,t)/t$.

Recalling (3.13b) we have $|\nabla^+ Q_t f| \leq |\nabla^- Q_t f|$ and therefore $|\nabla Q_t f| = |\nabla^- Q_t f|$ by (2.7). Taking Proposition 3.3 into account we obtain (3.17) and that the Hamilton-Jacobi equation is satisfied at all points $x$ such that $D^+(x,t) = D^-(x,t)$. $\square$

When $f$ is bounded the maps $Q_t f$ are easily seen to be bounded and $\mathsf{d}$-Lipschitz. It is immediate to see that

$$\inf_X f \leq \inf_X Q_t f \leq \sup_X Q_t f \leq \sup_X f. \tag{3.18}$$

A quantitative global estimate we shall need later on is:

$$\mathrm{Lip}(Q_t f) \leq 2\sqrt{\frac{\mathrm{osc}(f)}{t}}, \qquad \text{where} \quad \mathrm{osc}(f) := \sup_X f - \inf_X f. \tag{3.19}$$

It can be derived noticing that choosing a minimizing sequence $(y_n)_{n\in\mathbb{N}}$ for $F(t,x,\cdot)$ attaining the supremum in (3.4), the energy comparison

$$\frac{(D^+(x,t))^2}{2t} - \mathrm{osc}(f) \leq \lim_{n\to\infty} f(y_n) + \frac{\mathsf{d}^2(x,y_n)}{2t} - f(x) = Q_t f(x) - f(x) \leq 0$$

yields

$$D^+(x,t) \leq \sqrt{2t\,\mathrm{osc}(f)}. \tag{3.20}$$

Since $D^-(x,t) \le D^+(x,t)$, setting $R := (\sqrt{2}-1)\sqrt{2t\,\mathrm{osc}(f)}$, (3.14) and simple calculations yield

$$\frac{Q_t f(x) - Q_t f(y)}{\mathsf{d}(x,y)} \le 2\Big(\frac{\mathrm{osc}(f)}{t}\Big)^{1/2} \quad \text{if } 0 < \mathsf{d}(x,y) \le R,$$

and, since $\mathrm{osc}(Q_t f) \le \mathrm{osc}(f)$ by (3.18),

$$\frac{Q_t f(x) - Q_t f(y)}{\mathsf{d}(x,y)} \le \frac{\mathrm{osc}(Q_t f)}{R} \le 2\Big(\frac{\mathrm{osc}(f)}{t}\Big)^{1/2} \quad \text{if } \mathsf{d}(x,y) \ge R.$$

The constant 2 in (3.19) can be reduced to $\sqrt{2}$ if $X$ is a length space: it is sufficient to combine (3.20) with (3.13a).

We conclude this section with a simple observation, a technical lemma, where also a Polish structure is involved, and with some relations between slope of Kantorovich potentials and Wasserstein distance.

**Remark 3.7 (Continuity of $Q_t$ at $t = 0$)** If $(X, \tau, \mathsf{d})$ is an extended Polish space and $\varphi$ is bounded and $\tau$-lower semicontinuous, then $Q_t\varphi \uparrow \varphi$ as $t \downarrow 0$. This is a simple consequence of assumption (iii) in Definition 2.3. ■

**Proposition 3.8** *Let $(X, \tau, \mathsf{d})$ be an extended Polish space.*

*(i) if $K \subset X$ is compact, $\psi \in C(K)$, $M \ge \max\psi$ and*

$$\varphi(x) = \begin{cases} \psi(x) & \text{if } x \in K, \\ M & \text{if } x \in X \setminus K, \end{cases} \tag{3.21}$$

*then $Q_t\varphi$ is $\tau$-lower semicontinuous in $X$ for all $t > 0$;*

*(ii) if $\mathcal{D}(f) = X$, $t_*(x) \ge T > 0$ for all $x \in X$ and $Q_t\varphi$ is Borel measurable for all $t > 0$ then $\frac{\mathrm{d}^+}{\mathrm{d}t}Q_t\varphi(x)$ is Borel measurable in $X \times (0,T)$ and the slopes*

$$(x,t) \mapsto |\nabla^+ Q_t\varphi|(x), \qquad (x,t) \mapsto |\nabla^- Q_t\varphi|(x)$$

*are $\mathscr{B}^*(X \times (0,T))$-measurable in $X \times (0,T)$.*

*Proof.* (i) The proof is straightforward, using the identity

$$Q_t\varphi(x) = \min_{y \in K} \psi(y) + \frac{1}{2t}\mathsf{d}^2(x,y).$$

(ii) The Borel measurability of $\frac{\mathrm{d}^+}{\mathrm{d}t}Q_t\varphi(x)$ is a simple consequence of the continuity of $t \mapsto Q_t\varphi(x)$, together with the Borel measurability of $Q_t\varphi$. A simple time discretization argument also shows that $(x,t) \mapsto Q_t\varphi(x)$ is Borel measurable. Then, the proof of the measurability of slopes follows as in Lemma 2.4. □

In the next proposition we consider the ascending slope of Kantorovich potentials, for finite distances $\mathsf{d}$.

**Proposition 3.9 (Slope and approximation of Kantorovich potentials)** *Let $\mu$, $\nu \in \mathscr{P}(X)$ with $W_2(\mu,\nu) < \infty$ and let $\boldsymbol{\gamma} \in \mathscr{P}(X \times X)$ be an optimal plan with marginals $\mu$, $\nu$. If $\varphi$ is a Kantorovich potential relative to $\boldsymbol{\gamma}$, we have*

$$|\nabla^+\varphi|(x) \le \mathsf{d}(x,y) \quad \text{for } \boldsymbol{\gamma}\text{-a.e. } (x,y). \tag{3.22}$$

*In particular $|\nabla^+\varphi| \in L^2(X,\mu)$ and $\int_X |\nabla^+\varphi|^2\,\mathrm{d}\mu \le W_2^2(\mu,\nu)$.*

*Proof.* We set $f := -\varphi^c$, so that from $\varphi = (\varphi^c)^c$ we have $\varphi = Q_1 f$. In addition, the definition of Kantorovich potential tells us that $\varphi(x) = f(y) + \mathsf{d}^2(x, y)/2$ for $\boldsymbol{\gamma}$-a.e. $(x, y)$, so that

$$D^-(x, 1) \le \mathsf{d}(x, y) \quad \text{for } \boldsymbol{\gamma}\text{-a.e. } (x, y). \tag{3.23}$$

Taking (3.13b) into account we obtain (3.22). □

In general the inequality $\int_X |\nabla^+\varphi|^2 \,\mathrm{d}\mu \le W_2^2(\mu, \nu)$ can be strict, as the following simple example shows:

**Example 3.10** Let $X = [0, 1]$ endowed with the Euclidean distance, $\mu_0 = \delta_0$ and $\mu_t = t^{-1}\chi_{[0,t]}\mathscr{L}^1$ for $t \in (0, 1]$. Then clearly $(\mu_t)$ is a constant speed geodesic connecting $\mu_0$ to $\mu_1$ and the corresponding Kantorovich potential is $\varphi(x) = x^2/2 - x$, so that $\int |\nabla^+\varphi|^2 \,\mathrm{d}\mu_0 = 0$, while $W_2^2(\mu_0, \mu_1) = 1/3$.

# 4 Relaxed gradient, Cheeger's energy, and its $L^2$-gradient flow

In this section we assume that $(X, \tau, \mathsf{d})$ is a Polish extended space. Furthermore, $\mathfrak{m}$ is a nonnegative, Borel and $\sigma$-finite measure on $X$. Recall that

$$\text{there exists a bounded Borel function } \vartheta : X \to (0, \infty) \text{ such that } \int_X \vartheta \,\mathrm{d}\mathfrak{m} \le 1. \tag{4.1}$$

Notice that $\mathfrak{m}$ and the finite measure $\tilde{\mathfrak{m}} := \vartheta\mathfrak{m}$ share the same class of negligible sets. In the following we will often assume that $\mathfrak{m}$ and $\vartheta$ satisfy some further structural conditions, which will be described as they occur. For future references, let us just state here our strongest assumption in advance: we will often assume that $\vartheta$ has the form $\mathrm{e}^{-V^2}$, where

$$\begin{gathered} V : X \to [0, \infty) \text{ is a Borel } \mathsf{d}\text{-Lipschitz map,} \\ \text{it is bounded on each compact set } K \subset X, \text{ and } \quad \int_X \mathrm{e}^{-V^2} \,\mathrm{d}\mathfrak{m} \le 1. \end{gathered} \tag{4.2}$$

When $\tau$ is the topology induced by the finite distance $\mathsf{d}$, then the facts that $V$ is Borel and bounded on compact sets are obvious consequences of the $\mathsf{d}$-Lipschitz property. In this case a simple choice is $V(x) = \sqrt{\kappa/2}\,\mathsf{d}(x, x_0)$ for some $x_0 \in X$ and $\kappa > 0$. It is not difficult to check that (4.2) is then equivalent to

$$\exists\, \kappa > 0 : \quad m(r) \le \mathrm{e}^{\frac{\kappa}{2}r^2} \quad \text{where} \quad m(r) := \mathfrak{m}\big(\{x \in X : \mathsf{d}(x, x_0) < r\}\big). \tag{4.3}$$

In fact, for every $h > 0$

$$\int_X \mathrm{e}^{-\frac{h}{2}\mathsf{d}^2(x,x_0)} \,\mathrm{d}\mathfrak{m} = \int_X \int_{r > \mathsf{d}(x,x_0)} h\, r\, \mathrm{e}^{-\frac{h}{2}r^2} \,\mathrm{d}r \,\mathrm{d}\mathfrak{m}(x) = \int_0^\infty h\, r\, m(r)\, \mathrm{e}^{-\frac{h}{2}r^2} \,\mathrm{d}r. \tag{4.4}$$

Since $r \mapsto m(r)$ is nondecreasing, if the last integral in (4.4) is less than 1 for $h := \kappa$, then Chebichev inequality yields $m(r)\mathrm{e}^{-\frac{1}{2}\kappa r^2} \le 1$; on the other hand, if (4.4) holds, then there exists $h > \kappa$ sufficiently big such that the integral in (4.4) is less than 1, so that (4.2) holds.

### 4.1 Minimal relaxed gradient

The content of this subsection is inspired by Cheeger's work [10]. We are going to relax the integral of the squared local Lipschitz constant of Lipschitz functions with respect to the $L^2(X,\mathfrak{m})$ topology. By Lemma 2.4, $|\nabla f|$ is $\mathscr{B}^*(X)$-measurable whenever $f$ is $\mathsf{d}$-Lipschitz and Borel.

**Proposition 4.1** *Let $(X,\tau,\mathsf{d})$ be an extended Polish space and let $\mathfrak{m}$ be a nonnegative, Borel measure in $(X,\tau)$ satisfying the following condition (weaker than (4.2)):*

$$\forall\, K\subset X \text{ compact} \quad \exists r>0: \quad \mathfrak{m}\big(\{x\in X:\mathsf{d}(x,K)\le r\}\big)<\infty. \tag{4.5}$$

*Then the class of bounded, Borel and $\mathsf{d}$-Lipschitz functions $f\in L^2(X,\mathfrak{m})$ with $|\nabla f|\in L^2(X,\mathfrak{m})$ is dense in $L^2(X,\mathfrak{m})$.*

*Proof.* It suffices to approximate functions $\varphi: X\to\mathbb{R}$ such that for some compact set $K\subset X$

$$\varphi|_K\in C^0(K),\quad \varphi\equiv 0 \quad \text{in } X\setminus K.$$

By taking the positive and negative part, we can always assume that $\varphi$ is, e.g., nonnegative. We can thus define

$$\varphi_n(x):=\sup_{y\in K}\Big[\varphi(y)-n\mathsf{d}(x,y)\Big]^+.$$

It is not difficult to check that $\varphi_n$ is upper semicontinuous, nonnegative, $n$-Lipschitz and bounded above by $S:=\max_K\varphi\ge 0$; moreover

$$\varphi_n(x)=|\nabla\varphi_n|(x)=0 \quad \text{if } \mathsf{d}(x,K)>S/n.$$

If $r>0$ is given by (4.5), choosing $n>S/r$ we get that $\varphi_n,|\nabla\varphi_n|$ are supported in the set $\{x\in X:\mathsf{d}(x,K)\le r\}$ of finite measure, so that they belong to $L^2(X,\mathfrak{m})$; since $S\ge\varphi_n(x)\ge\varphi(x)$ and $\varphi_n(x)\downarrow\varphi(x)$ for every $x\in X$, we conclude. □

**Definition 4.2 (Relaxed gradients)** *We say that $G\in L^2(X,\mathfrak{m})$ is a relaxed gradient of $f\in L^2(X,\mathfrak{m})$ if there exist Borel $\mathsf{d}$-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 converge to $\tilde G$ in $L^2(X,\mathfrak{m})$;*

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

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

The definition of minimal relaxed gradient is well posed; indeed, thanks to (2.8a) and to the reflexivity of $L^2(X,\mathfrak{m})$, the collection of relaxed gradients of $f$ is a convex set, possibly empty. Its closure follows by the following lemma, which also shows that it is possible to obtain the minimal relaxed gradient as *strong* limit in $L^2$.

**Lemma 4.3 (Closure and strong approximation of the minimal relaxed gradient)**

*(a) If $G \in L^2(X,\mathfrak{m})$ is a relaxed gradient of $f \in L^2(X,\mathfrak{m})$ then there exist Borel* $\mathsf{d}$*-Lipschitz functions $f_n$ converging to $f$ in $L^2(X,\mathfrak{m})$ and $G_n \in L^2(X,\mathfrak{m})$ strongly convergent to $\tilde G$ in $L^2(X,\mathfrak{m})$ with $|\nabla f_n| \le G_n$ and $\tilde G \le G$.*

*(b) If $G_n \in L^2(X,\mathfrak{m})$ is a relaxed gradient of $f_n \in L^2(X,\mathfrak{m})$ and $f_n \rightharpoonup f$, $G_n \rightharpoonup G$ weakly in $L^2(X,\mathfrak{m})$, then $G$ is a relaxed gradient of $f$.*

*(c) In particular, the collection of all the relaxed gradients of $f$ is closed in $L^2(X,\mathfrak{m})$ and there exist Borel* $\mathsf{d}$*-Lipschitz functions $f_n \in L^2(X,\mathfrak{m})$ such that*

$$f_n \to f, \quad |\nabla f_n| \to |\nabla f|_* \quad \textit{strongly in } L^2(X,\mathfrak{m}). \tag{4.6}$$

*Proof.* (a) Since $G$ is a relaxed gradient, we can find Borel $d$-Lipschitz functions $g_i \in L^2(X,\mathfrak{m})$ such that $g_i \to f$ in $L^2(X,\mathfrak{m})$ and $|\nabla g_i|$ weakly converges to $\tilde G \le G$ in $L^2(X,\mathfrak{m})$; by Mazur's lemma we can find a sequence of convex combinations $G_n$ of $|\nabla g_i|$, starting from an index $i(n) \to \infty$, strongly convergent to $\tilde G$ in $L^2(X,\mathfrak{m})$; the corresponding convex combinations of $g_i$, that we shall denote by $f_n$, still converge in $L^2(X,\mathfrak{m})$ to $f$ and $|\nabla f_n|$ is bounded from above by $G_n$.

(b) Let us prove now the weak closure in $L^2(X,\mathfrak{m}) \times L^2(X,\mathfrak{m})$ of the set

$$S := \big\{(f,G) \in L^2(X,\mathfrak{m}) \times L^2(X,\mathfrak{m}) : G \text{ is a relaxed gradient for } f\big\}.$$

Since $S$ is convex, it is sufficient to prove that $S$ is strongly closed. If $S \ni (f^i, G^i) \to (f,G)$ strongly in $L^2(X,\mathfrak{m}) \times L^2(X,\mathfrak{m})$, we can find sequences of Borel $d$-Lipschitz functions $(f^i_n)_n \in L^2(X,\mathfrak{m})$ and of nonnegative functions $(G^i_n)_n \in L^2(X,\mathfrak{m})$ such that

$$f^i_n \stackrel{n\to\infty}{\longrightarrow} f^i, \quad G^i_n \stackrel{n\to\infty}{\longrightarrow} \tilde G^i \text{ strongly in } L^2(X,\mathfrak{m}), \quad |\nabla f^i_n| \le G^i_n, \quad \tilde G^i \le G^i.$$

Possibly extracting a suitable subsequence, we can assume that $\tilde G^i \rightharpoonup \tilde G$ weakly in $L^2(X,\mathfrak{m})$ with $\tilde G \le G$; by a standard diagonal argument and the reflexivity of $L^2(X,\mathfrak{m})$ we can also find an increasing sequence $i \mapsto n(i)$ such that $f^i_{n(i)} \to f$, $|\nabla f^i_{n(i)}| \rightharpoonup H$, and $G^i_{n(i)} \rightharpoonup \tilde G$ in $L^2(X,\mathfrak{m})$. It follows that $H \le \tilde G \le G$ so that $G$ is a relaxed gradient for $f$.

(c) Let us consider now the minimal relaxed gradient $G := |\nabla f|_*$ and let $f_n$, $G_n$ be sequences in $L^2(X,\mathfrak{m})$ as in the first part of the present Lemma. Since $|\nabla f_n|$ is uniformly bounded in $L^2(X,\mathfrak{m})$ it is not restrictive to assume that it is weakly convergent to some limit $H \in L^2(X,\mathfrak{m})$ with $0 \le H \le \tilde G \le G$. This implies at once that $H = \tilde G = G$ and $|\nabla f_n|$ weakly converges to $|\nabla f|_*$ (because any limit point in the weak topology of $|\nabla f_n|$ is a relaxed gradient with minimal norm) and that the convergence is strong, since

$$\limsup_{n\to\infty} \int_X |\nabla f_n|^2 \,\mathrm{d}\mathfrak{m} \le \limsup_{n\to\infty} \int_X G_n^2 \,\mathrm{d}\mathfrak{m} = \int_X G^2 \,\mathrm{d}\mathfrak{m} = \int_X H^2 \,\mathrm{d}\mathfrak{m}.$$

□

The distinguished role of the minimal relaxed gradient is also illustrated by the following lemma.

**Lemma 4.4 (Locality)** *Let $G_1$, $G_2$ be relaxed gradients of $f$. Then $\min\{G_1,G_2\}$ and $\chi_B G_1 + \chi_{X\setminus B} G_2$, $B \in \mathscr{B}(X)$, are relaxed gradients of $f$ as well. In particular, for any relaxed gradient $G$ of $f$ it holds*

$$|\nabla f|_* \le G \qquad \mathfrak{m}\text{-}a.e. \text{ in } X. \tag{4.7}$$

*Proof.* It is sufficient to prove that if $B \in \mathscr{B}(X)$, then $\chi_{X\setminus B} G_1 + \chi_B G_2$ is a relaxed gradient of $f$. By approximation, taking into account the closure of the class of relaxed gradients, we can assume with no loss of generality that $X \setminus B$ is a compact set, so that the $\mathsf{d}$-Lipschitz function

$$\rho(y) := \inf \{\mathsf{d}(y,x) : \ x \in X \setminus B\}$$

is $\tau$-lower semicontinuous and therefore $\mathscr{B}(X)$-measurable. Notice that, because of condition (iii) in Definition 2.3, $\rho$ is strictly positive in $B$ and null on $X \setminus B$. Therefore it will be sufficient to show that, setting $\chi_r := \min\{1, \rho/r\}$, $\chi_r G_1 + (1-\chi_r) G_2$ is a relaxed gradient for all $r > 0$.

Let now $f_{n,i}$, $i = 1, 2$, be Borel, $\mathsf{d}$-Lipschitz and $L^2(X, \mathfrak{m})$ functions converging to $f$ in $L^2$ as $n \to \infty$ with $|\nabla f_{n,i}|$ weakly convergent to $\tilde G_i \le G_i$, and set $f_n := \chi_r f_{n,1} + (1-\chi_r) f_{n,2}$. Then (2.9) immediately gives that $\chi_r G_1 + (1-\chi_r) G_2 \ge \chi_r \tilde G_1 + (1-\chi_r)\tilde G_2$ is a relaxed gradient.

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

By (4.7), for $f$ Borel and $\mathsf{d}$-Lipschitz we get

$$|\nabla f|_* \le |\nabla f| \qquad \mathfrak{m}\text{-a.e. in } X. \tag{4.8}$$

A direct byproduct of this characterization of $|\nabla f|_*$ is its invariance under multiplicative perturbations of $\mathfrak{m}$ of the form $\theta\,\mathfrak{m}$, with

$$0 < c \le \theta \le C < \infty \quad \mathfrak{m}\text{-a.e. on } X. \tag{4.9}$$

Indeed, the class of relaxed gradients is invariant under these perturbations.

**Theorem 4.5** *Cheeger's functional*

$$\mathsf{Ch}_*(f) := \frac{1}{2}\int_X |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m}, \tag{4.10}$$

*set equal to $+\infty$ if $f$ has no relaxed slope, is convex and lower semicontinuous in $L^2(X, \mathfrak{m})$. If* (4.5) *holds, then its domain is dense in $L^2(X, \mathfrak{m})$.*

*Proof.* A simple byproduct of condition (2.8a) is that $\alpha F + \beta G$ is a relaxed gradient of $\alpha f + \beta g$ whenever $\alpha$, $\beta$ are nonnegative constants and $F$, $G$ are relaxed gradients of $f$, $g$ respectively. Taking $F = |\nabla f|_*$ and $G = |\nabla g|_*$ yields

$$|\nabla(\alpha f + \beta g)|_* \le \alpha |\nabla f|_* + \beta |\nabla g|_* \quad \text{for every } f, g \in D(\mathsf{Ch}_*),\ \alpha, \beta \ge 0. \tag{4.11}$$

This proves the convexity of $\mathsf{Ch}_*$, while lower semicontinuity follows by (b) of Lemma 4.3. □

**Remark 4.6 (Cheeger's original functional)** Our definition of $\mathsf{Ch}_*$ can be compared with the original one in [10]: the relaxation procedure is similar, but the approximating functions $f_n$ are not required to be Lipschitz and $|\nabla f_n|$ are replaced by upper gradients $G_n$ of $f_n$. Obviously, this leads to a *smaller* functional, that we shall denote by $\underline{\mathsf{Ch}}_*$; this functional can still be represented by the integration of a local object, smaller $\mathfrak{m}$-a.e. than $|\nabla f|_*$, that we shall denote by $|\nabla f|_C$. Relating $\mathsf{Ch}_*$ to $\underline{\mathsf{Ch}}_*$ amounts to find, for any $f \in L^2(X, \mathfrak{m})$ and any upper gradient $G$ of $f$, a sequence of Lipschitz functions $f_n$ such that $f_n \to f$ in $L^2(X, \mathfrak{m})$ and

$$\limsup_{n\to\infty} \int_X |\nabla f_n|^2 \, d\mathfrak{m} \le \int_X G^2 \, d\mathfrak{m}. \tag{4.12}$$

It is well known, see [10], that this approximation is possible (even in strong $W^{1,2}$ norm) if Poincaré and doubling hold.

A byproduct of our identification result, see Remark 5.10 in the next section, is the fact that $\underline{\mathsf{Ch}}_* = \mathsf{Ch}_*$, i.e. that the approximation (4.12) with Lipschitz functions and their corresponding slopes instead of upper gradients is possible, without *any* regularity assumption on $(X, \mathsf{d}, \mathfrak{m})$, besides (4.2). Also, in the case when $\mathsf{d}$ is a distance, taking into account the locality properties of the weak gradients, the result can be extended to locally finite measures. ■

**Remark 4.7 (The Sobolev space $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$)** As a simple consequence of the lower semicontinuity of the Cheeger's functional, it can be proved that the domain of $\mathsf{Ch}_*$ endowed with the norm

$$\|f\|_{W^{1,2}} := \sqrt{\|f\|_2^2 + \||\nabla f|_*\|_2^2},$$

is a Banach space (for a detailed proof see [10, Theorem 2.7]). Call $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ this space.

This notation may be misleading because, in general, $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ is *not* an Hilbert space. This is the case, for example, of the metric measure space $(\mathbb{R}^d, \|\cdot\|, \mathscr{L}^d)$ where $\|\cdot\|$ is any norm not coming from an inner product. The fact that $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ may fail to be Hilbert is strictly related to the potential lack of linearity of the heat flow, see also Remark 4.14 below (for computations in smooth spaces with non linear heat flows see [29]). Also, the reflexivity of $W^{1,2}$ and the density of Lipschitz functions in $W^{1,2}$ norm seem to be difficult problems at this level of generality, while it is known that both these facts are true in doubling spaces satisfying a local Poincaré inequality, see [10]. ■

**Proposition 4.8 (Chain rule)** *If $f \in L^2(X, \mathfrak{m})$ has a relaxed gradient, the following properties hold:*

*(a) for any $\mathscr{L}^1$-negligible Borel set $N \subset \mathbb{R}$ it holds $|\nabla f|_* = 0$ $\mathfrak{m}$-a.e. on $f^{-1}(N)$;*

*(b) $|\nabla f|_* = |\nabla g|_*$ $\mathfrak{m}$-a.e. on $\{f - g = c\}$ for all constants $c \in \mathbb{R}$ and $g \in L^2(X, \mathfrak{m})$ with $\mathsf{Ch}_*(g) < \infty$;*

*(c) $\phi(f) \in D(\mathsf{Ch}_*)$ and $|\nabla \phi(f)|_* \le |\phi'(f)||\nabla f|_*$ for any Lipschitz function $\phi$ on an interval $J$ containing the image of $f$ (with $0 \in J$ and $\phi(0) = 0$ if $\mathfrak{m}$ is not finite);*

*(d) $\phi(f) \in D(\mathsf{Ch}_*)$ and $|\nabla \phi(f)|_* = \phi'(f)|\nabla f|_*$ for any nondecreasing and Lipschitz function $\phi$ on an interval $J$ containing the image of $f$ (with $0 \in J$ and $\phi(0) = 0$ if $\mathfrak{m}$ is not finite).*

*Proof.* **(a)** We claim that for $\phi : \mathbb{R} \to \mathbb{R}$ continuously differentiable and Lipschitz on the image of $f$ it holds

$$|\nabla \phi(f)|_* \le |\phi' \circ f||\nabla f|_*, \qquad \mathfrak{m}\text{-a.e. in } X, \tag{4.13}$$

for any $f \in D(\mathsf{Ch}_*)$. To prove this, observe that the pointwise inequality $|\nabla \phi(f)| \le |\phi' \circ f||\nabla f|$ trivially holds for $f \in L^2(X, \mathfrak{m})$ Borel and $\mathsf{d}$-Lipschitz. The claim follows by an easy approximation argument, thanks to (4.6) of Lemma 4.3; when $\mathfrak{m}$ is not finite, we also require $\phi(0) = 0$ in order to be sure that $\phi \circ f \in L^2(X, \mathfrak{m})$.

Now, assume that $N$ is compact. In this case, let $A_n \subset \mathbb{R}$ be open sets such that $A_n \downarrow N$. Also, let $\psi_n : \mathbb{R} \to [0, 1]$ be a continuous function satisfying $\chi_N \leq \psi_n \leq \chi_{A_n}$, and define $\phi_n : \mathbb{R} \to \mathbb{R}$ by

$$\begin{cases} \phi_n(0) &= 0, \\ \phi_n'(z) &= 1 - \psi_n(z). \end{cases}$$

The sequence $(\phi_n)$ uniformly converges to the identity map, and each $\phi_n$ is 1-Lipschitz and $C^1$. Therefore $\phi_n \circ f$ converge to $f$ in $L^2$. Taking into account that $\phi_n' = 0$ on $N$ and (4.13) we deduce

$$\begin{aligned} \int_X |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m} &\leq \liminf_{n\to\infty} \int_X |\nabla \phi_n(f)|_*^2 \,\mathrm{d}\mathfrak{m} \leq \liminf_{n\to\infty} \int_X |\phi_n' \circ f|^2 |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m} \\ &= \liminf_{n\to\infty} \int_{X \setminus f^{-1}(N)} |\phi_n' \circ f|^2 |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m} \leq \int_{X \setminus f^{-1}(N)} |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m}. \end{aligned}$$

It remains to deal with the case when $N$ is not compact. In this case we consider the measure $\mu := f_\sharp \tilde{\mathfrak{m}} \in \mathscr{P}(\mathbb{R})$, where $\tilde{\mathfrak{m}} = \vartheta \mathfrak{m}$ is the finite measure defined as in (4.1). Then there exists an increasing sequence $(K_n)$ of compact subsets of $N$ such that $\mu(K_n) \uparrow \mu(N)$. By the result for the compact case we know that $|\nabla f|_* = 0$ $\mathfrak{m}$-a.e. on $\cup_n f^{-1}(K_n)$, and by definition of push forward and the fact that $\tilde{\mathfrak{m}}$ and $\mathfrak{m}$ have the same negligible subsets, we know that $\mathfrak{m}(f^{-1}(N \setminus \cup_n K_n)) = 0$.
**(b)** By (a) the claimed property is true if $g$ is identically 0. In the general case we notice that $|\nabla(f - g)|_* + |\nabla g|_*$ is a relaxed gradient of $f$, hence on $\{f - g = c\}$ we conclude that $\mathfrak{m}$-a.e. it holds $|\nabla f|_* \leq |\nabla g|_*$. Reversing the roles of $f$ and $g$ we conclude.
**(c)** By (a) and Rademacher Theorem we know that the right hand side is well defined, so that the statement makes sense (with the convention to define $|\phi' \circ f|$ arbitrarily at points $x$ such that $\phi'$ does not exist at $f(x)$). Also, by (4.13) we know that the thesis is true if $\phi$ is $C^1$. For the general case, just approximate $\phi$ with a sequence $(\phi_n)$ of equi-Lipschitz and $C^1$ functions, such that $\phi_n' \to \phi'$ a.e. on the image of $f$.
**(d)** Arguing as in (c) we see that it is sufficient to prove the claim under the further assumption that $\phi$ is $C^1$, thus we assume this regularity. Also, with no loss of generality we can assume that $0 \leq \phi' \leq 1$. We know that $(1 - \phi'(f))|\nabla f|_*$ and $\phi'(f)|\nabla f|_*$ are relaxed gradients of $f - \phi(f)$ and $f$ respectively. Since

$$|\nabla f|_* \leq |\nabla(f - \phi(f))|_* + |\nabla \phi(f)|_* \leq \Big((1 - \phi'(f)) + \phi'(f)\Big)|\nabla f|_* = |\nabla f|_*$$

it follows that all inequalities are equalities $\mathfrak{m}$-a.e. in $X$. □

Taking the locality into account, we can extend the relaxed gradient from $L^2(X, \mathfrak{m})$ to the class of $\mathfrak{m}$-measurable maps $f$ whose truncates $f_N := \min\{N, \max\{f, -N\}\}$ belong to $D(\mathsf{Ch}_*) \subset L^2(X, \mathfrak{m})$ for any integer $N$ in the following way:

$$|\nabla f|_* := |\nabla f_N|_* \qquad \mathfrak{m}\text{-a.e. on } \{|f| < N\}. \tag{4.14}$$

Accordingly, we can extend Cheeger's functional (4.10) as follows:

$$\tilde{\mathsf{Ch}}_*(f) := \begin{cases} \frac{1}{2} \int_X |\nabla f|_*^2 \,\mathrm{d}\mathfrak{m} & \text{if } f_N \in D(\mathsf{Ch}_*) \text{ for all } N \geq 1 \\ +\infty & \text{otherwise.} \end{cases} \tag{4.15}$$

It is obvious that $\tilde{\mathsf{Ch}}_*$ is convex and, when $\mathfrak{m}(X) < \infty$, it is sequentially lower semicontinuous with respect to convergence $\mathfrak{m}$-a.e. in $X$: we shall see that this property holds even when $\mathfrak{m}$ is not finite but satisfies (4.2). We shall use this extension when we will compare relaxed and weak upper gradient, see Theorem 6.2.

Here it is useful to introduce the *Fisher information functional*:

**Definition 4.9 (Fisher information)** *We define the Fisher information $\mathsf{F}(f)$ of a Borel function $f : X \to [0, \infty)$ as*

$$\mathsf{F}(f) := 4 \int_X |\nabla \sqrt{f}|_*^2 \, d\mathfrak{m} = 8 \, \mathsf{Ch}_*(\sqrt{f}), \tag{4.16}$$

*if $\sqrt{f} \in D(\mathsf{Ch}_*)$ and we define $\mathsf{F}(f) = +\infty$ otherwise.*

**Lemma 4.10 (Properties of $\mathsf{F}$)** *For every Borel function $f : X \to [0, \infty)$ we have the equivalence*

$$f \in D(\mathsf{F}) \quad \Longleftrightarrow \quad f, \, |\nabla f|_* \in L^1(X, \mathfrak{m}), \quad \int_{\{f>0\}} \frac{|\nabla f|_*^2}{f} \, d\mathfrak{m} < \infty, \tag{4.17}$$

*and in this case it holds*

$$\mathsf{F}(f) = \int_{\{f>0\}} \frac{|\nabla f|_*^2}{f} \, d\mathfrak{m}. \tag{4.18}$$

*In addition, the functional $\mathsf{F}$ is convex and sequentially lower semicontinuous with respect to the weak topology of $L^1(X, \mathfrak{m})$.*

*Proof.* By the definition of extended relaxed gradient it is sufficient to consider the case when $f$ is bounded. The right implication in (4.17) is an immediate consequence of Proposition 4.8 with $\phi(r) = r^2$. The reverse one still follows by applying the same property to $\phi_\varepsilon(r) = \sqrt{r + \varepsilon} - \sqrt{\varepsilon}$, $\varepsilon > 0$, and then passing to the limit as $\varepsilon \downarrow 0$.

The strong lower semicontinuity in $L^1(X, \mathfrak{m})$ is an immediate consequence of the lower semicontinuity of the Cheeger's energy in $L^2(X, \mathfrak{m})$. The convexity of $\mathsf{F}$ follows by the representation of $\mathsf{F}$ given in (4.18), the convexity of $g \mapsto |\nabla g|_*$ stated in (4.11), and the convexity of the function $(x, y) \mapsto y^2/x$ in $(0, \infty) \times \mathbb{R}$. Since $\mathsf{F}$ is convex, its weak lower semicontinuity in $L^1(X, \mathfrak{m})$ is a consequence of the strong one. □

We conclude this section with a result concerning general multiplicative perturbations of the measure $\mathfrak{m}$. Notice that the choice $\theta = \mathrm{e}^{-V^2}$ with $V$ as in (4.2) implies (4.19) for arbitrary $r > 0$.

**Lemma 4.11 (Invariance with respect to multiplicative perturbations of $\mathfrak{m}$)** *Let $\mathfrak{m}' = \theta \, \mathfrak{m}$ be another $\sigma$-finite Borel measure whose density $\theta$ satisfies the following condition: for every $K$ compact in $X$ there exist $r > 0$ and positive constants $c(K)$, $C(K)$ such that*

$$0 < c(K) \le \theta \le C(K) < \infty \quad \mathfrak{m}\text{-}a.e. \text{ on } K(r) := \{x \in X : \mathsf{d}(x, K) \le r\}. \tag{4.19}$$

*Then the relaxed gradient $|\nabla f|_*'$ induced by $\mathfrak{m}'$ coincides $\mathfrak{m}$-a.e. with $|\nabla f|_*$ for every $f \in W^{1,2}(X, \mathsf{d}, \mathfrak{m}) \cap W^{1,2}(X, \mathsf{d}, \mathfrak{m}')$. If moreover there exists $r > 0$ such that (4.19) holds for every compact set $K \subset X$ then*

$$f \in W^{1,2}(X, \mathsf{d}, \mathfrak{m}), \quad f, |\nabla f|_* \in L^2(X, \mathfrak{m}') \quad \Longrightarrow \quad f \in W^{1,2}(X, \mathsf{d}, \mathfrak{m}'). \tag{4.20}$$

*Proof.* Let us first notice that the role of $\mathfrak{m}$ and $\mathfrak{m}'$ in (4.19) can be inverted, since also $\mathfrak{m}$ is absolutely continuous w.r.t. $\mathfrak{m}'$ ((4.19) yields $\mathfrak{m}(K) = 0$ for every compact set $K$ with $\mathfrak{m}'(K) = 0$) and therefore its density $\mathrm{d}\mathfrak{m}/\mathrm{d}\mathfrak{m}' = \theta^{-1}$ w.r.t. $\mathfrak{m}'$ still satisfies (4.19).

Let us prove that $|\nabla f|_* \le |\nabla f|'_*$: we argue by contradiction and we suppose that for some $f \in W^{1,2}(X, \mathsf{d}, \mathfrak{m}) \cap W^{1,2}(X, \mathsf{d}, \mathfrak{m}')$ the strict inequality $|\nabla f|_* > |\nabla f|'_*$ holds in a Borel set $B$ with $\mathfrak{m}'(B) > 0$.

By the regularity of $\mathfrak{m}'$ we can find a compact set $K \subset B$ with $\mathfrak{m}'(K) > 0$ (and therefore $\mathfrak{m}(K) > 0$ by (4.19)) and $r > 0$ such that (4.19) holds. Introducing a Lipschitz real function $\phi_r : \mathbb{R} \to [0, 1]$ such that $\phi_r(v) \equiv 1$ in $[0, r/3]$ and $\phi_r(v) \equiv 0$ in $[2r/3, \infty)$, we consider the corresponding functions $\chi_r(x) := \phi_r(\mathsf{d}(x, K))$, which are lower semicontinuous, $\mathsf{d}$-Lipschitz, and satisfy $\chi_r(x) = |\nabla \chi_r(x)| = 0$ for every $x$ with $\mathsf{d}(x, K) > 2r/3$.

Applying Lemma 4.3 we find a sequence of Borel and $\mathsf{d}$-Lipschitz function $f_n \in L^2(X, \mathfrak{m})$ satisfying (4.6). It is easy to check that $f'_n := \chi_r\, f_n$ is a sequence of Borel $\mathsf{d}$-Lipschitz functions which converges strongly to $f' := \chi_r\, f$ in $L^2(X, \mathfrak{m}')$ by (4.19). Moreover, since

$$|\nabla f'_n| \le \chi_r |\nabla f_n| + |f_n| \,\mathrm{Lip}(\chi_r) \qquad \text{and} \qquad |\nabla f'_n| \equiv 0 \quad \text{on the open set } X \setminus K(r), \tag{4.21}$$

$|\nabla f'_n|$ is clearly uniformly bounded in $L^2(X, \mathfrak{m}')$ by (4.19), so that up to subsequence, it weakly converges to some function $G' \ge |\nabla f|'_*$. Since $|\nabla f'_n| = |\nabla f_n|$ in a $\mathsf{d}$-open set containing $K$, (4.6) yields $G' = |\nabla f|_*$ $\mathfrak{m}'$-a.e. in $K$ so that $|\nabla f|'_* \le |\nabla f|_*$ $\mathfrak{m}'$-a.e. in $K$. Inverting the role of $\mathfrak{m}$ and $\mathfrak{m}'$, we can also prove the converse inequality $|\nabla f|'_* \le |\nabla f|_*$.

In order to prove (4.20), let $K_n$ be an sequence of compact sets such that $\chi_{K_n} \uparrow 1$ as $n \to \infty$ $\mathfrak{m}$-a.e. in $X$ (recall that the finite measure $\tilde{\mathfrak{m}} = \vartheta \mathfrak{m}$ defined by (4.1) is tight); by the previous argument and (4.19) (which now, by assumption, holds uniformly with respect to $K_n$) we find a sequence $\chi_n(x) := \phi_r(\mathsf{d}(x, K_n))$ uniformly $\mathsf{d}$-Lipschitz such that $\chi_n f \in W^{1,2}(X, \mathsf{d}, \mathfrak{m}')$. Since $\chi_n f$ converges strongly to $f$ in $L^2(X, \mathfrak{m}')$ and (4.21) yields $|\nabla(\chi_n f)|_* \le |\nabla f|_* + \frac{3}{r}|f|$, we deduce that $|\nabla(\chi_n f)|_* = |\nabla(\chi_n f)|'_*$ is uniformly bounded in $L^2(X, \mathfrak{m}')$; applying (b) of Lemma 4.3 we conclude. $\square$

**Remark 4.12** Although the content of this section makes sense in a general metric measure space, it should be remarked that if no additional assumption is made it may happen that the constructions presented here are trivial.
Consider for instance the case of the interval $[0, 1] \subset \mathbb{R}$ endowed with the Euclidean distance and a probability measure $\mathfrak{m}$ concentrated on the set $\{q_n\}_{n \in \mathbb{N}}$ of rational points in $(0, 1)$. For every $n \ge 1$ we consider an open set $A_n \supset \mathbb{Q} \cap (0, 1)$ with Lebesgue measure less than $1/n$ and the 1-Lipschitz function $j_n(x) = \mathscr{L}^1(([0, x] \setminus A_n)$, locally constant in $A_n$. If $f$ is any $L$-Lipschitz function in $[0, 1]$, then $f_n(x) := f(j_n(x))$ is still $L$-Lipschitz and satisfies

$$\int_{[0,1]} |\nabla f_n|^2(x) \,\mathrm{d}\mathfrak{m}(x) = 0.$$

Since $j_n(x) \to x$, $f_n \to f$ strongly in $L^2([0, 1]; \mathfrak{m})$ as $n \to \infty$ and we obtain that $\mathsf{Ch}_*(f) = 0$. Hence Cheeger's functional is identically 0 and the corresponding gradient flows that we shall study in the sequel are simply the constant curves.

Another simple example is $X = [0, 1]$ endowed with the Lebesgue measure $\mathfrak{m}$ and the distance $\mathsf{d}(x, y) := |y - x|^{1/2}$. It is easy to check that $|\nabla f|(x) \equiv 0$ for every $f \in C^1([0, 1])$ (which is in particular $\mathsf{d}$-Lipschitz), so that a standard approximation argument yields $\mathsf{Ch}_*(f) = 0$ for every $f \in L^2([0, 1]; \mathfrak{m})$. $\blacksquare$

## 4.2 Laplacian and $L^2$ gradient flow of Cheeger's energy

In this subsection we assume, besides $\sigma$-finiteness that the measure $\mathfrak{m}$ satisfies the condition in (4.5) (weaker than (4.2)), so that the domain of $\mathsf{Ch}_*$ is dense in $L^2(X,\mathfrak{m})$ by Proposition 4.1.

The Hilbertian theory of gradient flows (see for instance [8], [4]) can be applied to Cheeger's functional (4.10) to provide, for all $f_0 \in L^2(X,\mathfrak{m})$, a locally Lipschitz map $t \mapsto f_t = \mathsf{H}_t(f_0)$ from $(0,\infty)$ to $L^2(X,\mathfrak{m})$, with $f_t \to f_0$ as $t \downarrow 0$, whose derivative satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t \in -\partial^-\mathsf{Ch}_*(f_t) \qquad \text{for a.e. } t \in (0,\infty). \tag{4.22}$$

Recall that the subdifferential $\partial^-\mathsf{Ch}_*$ of convex analysis is the multivalued operator in $L^2(X,\mathfrak{m})$ defined at all $f \in D(\mathsf{Ch}_*)$ by the family of inequalities

$$\ell \in \partial^-\mathsf{Ch}_*(f) \quad \Longleftrightarrow \quad \int_X \ell(g-f)\,\mathrm{d}\mathfrak{m} \le \mathsf{Ch}_*(g) - \mathsf{Ch}_*(f) \quad \text{for every } g \in L^2(X,\mathfrak{m}). \tag{4.23}$$

The map $\mathsf{H}_t : f_0 \mapsto f_t$ is uniquely determined by (4.22) and defines a semigroup of contractions in $L^2(X,\mathfrak{m})$. Furthermore, we have the regularization estimate

$$\mathsf{Ch}_*(f_t) \le \inf\left\{\mathsf{Ch}_*(g) + \frac{1}{2t}\int_X |g-f_0|^2\,\mathrm{d}\mathfrak{m} :\ g \in W^{1,2}(X,\mathsf{d},\mathfrak{m})\right\}. \tag{4.24}$$

Another important regularizing effect of gradient flows lies in the fact that, for every $t>0$, the right derivative $\frac{\mathrm{d}^+}{\mathrm{d}t} f_t$ exists and it is actually the element with minimal $L^2(X,\mathfrak{m})$ norm in $\partial^-\mathsf{Ch}_*(f_t)$. This motivates the next definition:

**Definition 4.13 ($(\mathsf{d},\mathfrak{m})$-Laplacian)** *The Laplacian $\Delta_{\mathsf{d},\mathfrak{m}} f$ of $f \in L^2(X,\mathfrak{m})$ is defined for those $f$ such that $\partial^-\mathsf{Ch}_*(f) \neq \emptyset$. For those $f$, $-\Delta_{\mathsf{d},\mathfrak{m}} f$ is the element of minimal $L^2(X,\mathfrak{m})$ norm in $\partial^-\mathsf{Ch}_*(f)$.*

The domain of $\Delta_{\mathsf{d},\mathfrak{m}}$ will be denoted by $D(\Delta_{\mathsf{d},\mathfrak{m}})$, since there is no risk of confusion with the notation (2.5) introduced for extended real valued maps; in this connection, notice that convexity and lower semicontinuity of $\mathsf{Ch}_*$ ensure the identity $D(\Delta_{\mathsf{d},\mathfrak{m}}) = D(|\nabla^-\mathsf{Ch}_*|)$, see [4, Proposition 1.4.4]. We can now write

$$\frac{\mathrm{d}^+}{\mathrm{d}t} f_t = \Delta_{\mathsf{d},\mathfrak{m}} f_t \qquad \text{for every } t \in (0,\infty)$$

for gradient flows $f_t$ of $\mathsf{Ch}_*$, in agreement with the classical case. However, not all classical properties remain valid, as illustrated in the next remark.

**Remark 4.14 (Potential lack of linearity)** It should be observed that, in general, the Laplacian we just defined is *not* a linear operator: the potential lack of linearity is strictly related to the fact that the space $W^{1,2}(X,\mathsf{d},\mathfrak{m})$ needs not be Hilbert, see also Remark 4.7. However, the Laplacian (and the corresponding gradient flow $\mathsf{H}_t$) is always 1-homogeneous, namely

$$\Delta_{\mathsf{d},\mathfrak{m}}(\lambda f) = \lambda \Delta_{\mathsf{d},\mathfrak{m}} f, \quad \mathsf{H}_t(\lambda f) = \lambda \mathsf{H}_t(f) \quad \text{for all } f \in D(\Delta_{\mathsf{d},\mathfrak{m}})) \text{ and } \lambda \in \mathbb{R}.$$

This is indeed a property true for the subdifferential of any 2-homogeneous functional $\Phi$; to prove it, if $\lambda \neq 0$ (the case $\lambda = 0$ being trivial) and $\xi \in \partial\Phi(x)$ it suffices to multiply the subdifferential inequality $\Phi(\lambda^{-1}y) \ge \Phi(x) + \langle \xi, \lambda^{-1}y - x\rangle$ by $\lambda^2$ to get $\lambda\xi \in \partial\Phi(\lambda x)$. ■

**Proposition 4.15 (Some properties of the Laplacian)** *For all $f \in D(\Delta_{\mathsf{d},\mathfrak{m}})$, $g \in D(\mathsf{Ch}_*)$ it holds*

$$-\int_X g\Delta_{\mathsf{d},\mathfrak{m}} f \,\mathrm{d}\mathfrak{m} \le \int_X |\nabla g|_* |\nabla f|_* \,\mathrm{d}\mathfrak{m}. \tag{4.25}$$

*Also, let $f \in D(\Delta_{\mathsf{d},\mathfrak{m}})$ and $\phi : J \to \mathbb{R}$ Lipschitz, with $J$ closed interval containing the image of $f$ ($\phi(0) = 0$ if $\mathfrak{m}(X) = \infty$). Then*

$$-\int_X \phi(f)\Delta_{\mathsf{d},\mathfrak{m}} f \,\mathrm{d}\mathfrak{m} = \int_X |\nabla f|_*^2 \phi'(f) \,\mathrm{d}\mathfrak{m}. \tag{4.26}$$

*Proof.* Since $-\Delta_{\mathsf{d},\mathfrak{m}} f \in \partial^- \mathsf{Ch}_*(f)$ it holds

$$\mathsf{Ch}_*(f) - \int_X \varepsilon g \Delta_{\mathsf{d},\mathfrak{m}} f \,\mathrm{d}\mathfrak{m} \le \mathsf{Ch}_*(f + \varepsilon g) \qquad \forall \varepsilon \in \mathbb{R}.$$

For $\varepsilon > 0$, $|\nabla f|_* + \varepsilon |\nabla g|_*$ is a relaxed gradient of $f + \varepsilon g$. Thus it holds $2\mathsf{Ch}_*(f + \varepsilon g) \le \int_X (|\nabla f|_* + \varepsilon |\nabla g|_*)^2 \,\mathrm{d}\mathfrak{m}$ and therefore

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

Dividing by $\varepsilon$, letting $\varepsilon \downarrow 0$ we get (4.25).

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

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

which implies that for any $v \in \partial^- \mathsf{Ch}_*(f)$ 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 $v = -\Delta_{\mathsf{d},\mathfrak{m}} f$. □

**Theorem 4.16 (Comparison principle, convex entropies and contraction)** *Let $f_t = \mathsf{H}_t(f_0)$ be the gradient flow of $\mathsf{Ch}_*$ starting from $f_0 \in L^2(X, \mathfrak{m})$.*

*(a) Assume that $f_0 \le C$ (resp. $f_0 \ge c$). Then $f_t \le C$ (resp. $f_t \ge c$) for every $t \ge 0$.*

*(b) If $g_0 \in L^2(X, \mathfrak{m})$ and $f_0 \le g_0 + c$ $\mathfrak{m}$-a.e. in $X$ for some constant $c \ge 0$, then $f_t \le g_t + c$ $\mathfrak{m}$-a.e. in $X$, where $g_t = \mathsf{H}_t(g_0)$ is the gradient flow starting from $g_0$. In particular the semigroup $\mathsf{H}_t : L^2(X, \mathfrak{m}) \to L^2(X, \mathfrak{m})$ satisfies the contraction property*

$$\|\mathsf{H}_t(f_0) - \mathsf{H}_t(g_0)\|_{L^p(X,\mathfrak{m})} \le \|f_0 - g_0\|_{L^p(X,\mathfrak{m})} \quad \forall\, f_0,\, g_0 \in L^2(X, \mathfrak{m}) \cap L^p(X, \mathfrak{m}) \tag{4.27}$$

*for every $p \in [2, \infty]$.*

*(c) If $e : \mathbb{R} \to [0, \infty]$ is a convex lower semicontinuous function and $E(f) := \int_X e(f) \,\mathrm{d}\mathfrak{m}$ is the associated convex and lower semicontinuous functional in $L^2(X, \mathfrak{m})$ it holds*

$$E(f_t) \le E(f_0) \quad \text{for every } t \ge 0. \tag{4.28}$$

*In particular, if $p \in [1, \infty]$ and $f_0 \in L^p(X, \mathfrak{m})$, then also $f_t \in L^p(X, \mathfrak{m})$. Moreover, if $e'$ is locally Lipschitz in $\mathbb{R}$ and $E(f_0) < \infty$, then we have*

$$E(f_t) + \int_0^t \int_X e''(f_s) |\nabla f_s|_*^2 \,\mathrm{d}\mathfrak{m} \,\mathrm{d}s = E(f_0) \qquad \forall t \ge 0. \tag{4.29}$$

*(d) When* $\mathfrak{m}(X) < \infty$ *we have*

$$\int_X f_t \, d\mathfrak{m} = \int_X f_0 \, d\mathfrak{m} \quad \text{for every } t \geq 0, \tag{4.30}$$

*and the evolution semigroup* $\mathsf{H}_t$ *satisfies* (4.27) *also for* $p \in [1, 2]$ *and, more generally, for* $E$ *as in (c) it holds*

$$E(f_t - g_t) \leq E(f_0 - g_0). \tag{4.31}$$

*Proof.* By [8, Theorem 4.4], a convex and lower semicontinuous functional $E$ on $L^2(X, \mathfrak{m})$ is nonincreasing along the gradient flow generated by $\mathsf{Ch}_*$ if and only if for every $f \in L^2(X, \mathfrak{m})$ with $E(f) < \infty$ and $\lambda > 0$ we have $E(f^\lambda) \leq E(f)$, where $f^\lambda$ is the unique minimizer of

$$u \qquad \mapsto \qquad \mathsf{Ch}_*(u) + \frac{\lambda}{2} \int_X |u - f|^2 \, d\mathfrak{m}. \tag{4.32}$$

Notice that $f^\lambda$ is also the unique solution of

$$f^\lambda - \lambda \Delta_{\mathsf{d},\mathfrak{m}} f^\lambda = f. \tag{4.33}$$

In order to prove (a), we can apply the classical Stampacchia's truncation argument and prove that $f \leq C$ entails $f^\lambda \leq C$. Indeed, if this is not the case we can consider the competitor $g := \min\{f^\lambda, C\}$ in the above minimization problem. Its Cheeger energy is less or equal than the one of $f^\lambda$ (by applying Proposition 4.8 (d)) and the $L^2$ distance between $f$ and $g$ is strictly smaller than the one between $f$ and $f^\lambda$, if $\mathfrak{m}(\{f^\lambda > C\}) > 0$. The same arguments applies to uniform bounds from below.

In order to prove the first part of (b) it suffices to show a discrete maximum principle analogous to (a) (but, notice that constants need not be in $L^2(X, \mathfrak{m})$), i.e. if $f^\lambda$ is as above and $g^\lambda$ minimizes $u \mapsto 2\mathsf{Ch}_*(u) + \lambda \|u - g_0\|_2^2$, then $f^\lambda \leq g^\lambda + c$ $\mathfrak{m}$-a.e. in $X$. Indeed, iterating this estimate the convergence of the Euler scheme to gradient flows provides the result. We can assume with no loss of generality that $f_0 < g_0 + c$ $\mathfrak{m}$-a.e. in $X$.

Let $\tilde{f}^\lambda := \min\{f^\lambda, g^\lambda + c\}$, $\tilde{g}^\lambda := \max\{f^\lambda - c, g^\lambda\}$, $A = \{f^\lambda > g^\lambda + c\}$, $B = X \setminus A$. Notice that $\tilde{f}^\lambda = f^\lambda - (f^\lambda - g^\lambda - c)^+$ belongs to $W^{1,2}(X, \mathsf{d}\mathfrak{m})$ by (c) of Proposition 4.8 (here $\phi(r) = (r - c)^+$ and $c \geq 0$), and the same property holds for $\tilde{g}^\lambda$. Our goal is to show that $\mathfrak{m}(A) = 0$; to this aim, adding the inequalities

$$\mathsf{Ch}_*(f^\lambda) + \frac{\lambda}{2} \int_X |f^\lambda - f_0|^2 \, d\mathfrak{m} \leq \mathsf{Ch}_*(\tilde{f}^\lambda) + \frac{\lambda}{2} \int_X |\tilde{f}^\lambda - f_0|^2 \, d\mathfrak{m},$$

$$\mathsf{Ch}_*(g^\lambda) + \frac{\lambda}{2} \int_X |g^\lambda - g_0|^2 \, d\mathfrak{m} \leq \mathsf{Ch}_*(\tilde{g}^\lambda) + \frac{\lambda}{2} \int_X |\tilde{g}^\lambda - g_0|^2 \, d\mathfrak{m}$$

and taking the identity $\mathsf{Ch}_*(f^\lambda) + \mathsf{Ch}_*(g^\lambda) = \mathsf{Ch}_*(\tilde{f}^\lambda) + \mathsf{Ch}_*(\tilde{g}^\lambda)$ into account by Proposition 4.8(b), we get

$$\begin{aligned} & \int_A |f^\lambda - f_0|^2 + |g^\lambda - g_0|^2 \, d\mathfrak{m} + \int_B |f^\lambda - f_0|^2 + |g^\lambda - g_0|^2 \, d\mathfrak{m} \\ \leq \; & \int_A |g^\lambda + c - f_0|^2 + |f^\lambda - c - g_0|^2 \, d\mathfrak{m} + \int_B |f^\lambda - f_0|^2 + |g^\lambda - g_0|^2 \, d\mathfrak{m}, \end{aligned}$$

so that, simplifying and rearranging terms we get

$$2\int_A (f_0 - g_0 - c)(g^\lambda + c - f^\lambda)\,\mathrm{d}\mathfrak{m} \le 0.$$

Since $f_0 < g_0 + c$ and $g^\lambda + c < f^\lambda$ in $A$, this can happen only if $\mathfrak{m}(A) = 0$.

The estimate (4.27) for $p = \infty$ follows immediately by applying the previous comparison principle with $c := \|f_0 - g_0\|_{L^\infty(X,\mathfrak{m})}$ and then reverting the role of $f_t$ and $g_t$ in the inequality. When $p = 2$ (4.27) is the well known contraction property for gradient flows of convex functionals in Hilbert spaces; the general case $p \in (2,\infty)$ follows then by interpolation, see [37, 32].

In order to prove the first claim of (c) we can assume $E(f_0) < \infty$. Assuming temporarily $e'$ to be Lipschitz, we can multiply (4.33) by $e'(f^\lambda) \in L^2(X,\mathfrak{m})$ (notice that if $\mathfrak{m}(X) = \infty$ then $e(0) = 0$, since $E(f_0) < \infty$, and $e'(0) = 0$, since 0 is a minimum point for $e$), obtaining after an integration and the application of (4.26)

$$\int_X f^\lambda e'(f^\lambda)\,\mathrm{d}\mathfrak{m} + \lambda\int_X e''(f^\lambda)|\nabla f^\lambda|_*^2\,\mathrm{d}\mathfrak{m} = \int_X f e'(f^\lambda)\,\mathrm{d}\mathfrak{m},$$

so that, by the convexity of $e$,

$$E(f^\lambda) - E(f) \le \int_X (f^\lambda - f)e'(f^\lambda)\,\mathrm{d}\mathfrak{m} \le 0.$$

A standard approximation, replacing $e(r)$ by its Yosida approximation yields the same result for general $e$.

The second claim of (c) follows by a similar argument: notice that by (a) we know that the image of $f_t$ is contained in the same interval containing the image of $f_0$. We can also assume, by truncation, that this interval is closed, bounded, and that $e'$ is Lipschitz in it (the interval also contains 0 if $\mathfrak{m}(X) = \infty$ and $e'(0) = 0$). Also, we know that the map $t \mapsto f_t$ is locally Lipschitz in $(0,\infty)$ with values in $L^2(X,\mathfrak{m})$, the same is true for the map $t \mapsto e(f_t)$ and $\frac{\mathrm{d}}{dt}e(f_t) = e'(f_t)\frac{\mathrm{d}}{\mathrm{d}t}f_t = e'(f_t)\Delta_{\mathsf{d},\mathfrak{m}} f_t$. Thus the map $t \mapsto E(f_t)$ is locally Lipschitz in $L^2(X,\mathfrak{m})$ and using equation (4.26) with $\phi = e'$ we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_X e(f_t)\,\mathrm{d}\mathfrak{m} = \int_X e'(f_t)\Delta_{\mathsf{d},\mathfrak{m}} f_t \mathrm{d}\,\mathfrak{m} = -\int_X e''(f_t)|\nabla f_t|_*^2\,\mathrm{d}\mathfrak{m}. \tag{4.34}$$

In order to prove (d) we notice that $\mathfrak{m}(X) < \infty$ allows to the choice of $g = \pm 1$ in (4.25), to obtain $\int_X \Delta_{\mathsf{d},\mathfrak{m}} h\,\mathrm{d}\mathfrak{m} = 0$ for all $h \in D(\Delta_{\mathsf{d},\mathfrak{m}})$. Hence (4.30) follows by

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_X f_t\,\mathrm{d}\mathfrak{m} = \int_X \Delta_{\mathsf{d},\mathfrak{m}} f_t\,\mathrm{d}\mathfrak{m} = 0.$$

The contraction property (4.27) when $p = 1$ follows now by the classical argument of Crandall and Tartar [11, Prop. 1]; (4.31) can be obtained by applying [9, Lemma 3] to the map $T(h) := \mathsf{H}_t(h + g_0)) - \mathsf{H}_t(g_0)$ and choosing $h := f_0 - g_0$. $\square$

## 4.3 Increasing family of measures and variational approximation of Cheeger's energy

In this section we study a monotone approximation scheme for the Cheeger's energy and its gradient flow, which turns to be quite useful when $\mathfrak{m}(X) = \infty$ and one is interested to extend the validity of suitable estimates, which can be more easily obtained in the case of measures with finite total mass.

Let us consider an increasing sequence of $\sigma$-finite, Borel measures $\mathfrak{m}^0 \le \mathfrak{m}^1 \le \cdots \le \mathfrak{m}^k \le \mathfrak{m}^{k+1} \le \cdots$ converging to the limit measure $\mathfrak{m}$ in the sense that

$$\lim_{k\to\infty} \mathfrak{m}^k(B) = \mathfrak{m}(B) \quad \text{for every } B \in \mathscr{B}(X). \tag{4.35}$$

Let us assume that, as in (4.19), $\mathfrak{m} \ll \mathfrak{m}^0$ with density $\theta = \dfrac{\mathrm{d}\mathfrak{m}}{\mathrm{d}\mathfrak{m}^0}$ satisfying

$$0 < c(K) \le \theta \le C(K) < \infty \quad \mathfrak{m}^0\text{-a.e. on } K(r) := \{x \in X : \mathsf{d}(x, K) \le r\} \tag{4.36}$$

for any compact set $K \subset X$, with $r = r(K) > 0$. Notice that the measures $\mathfrak{m}^k$ share the same collection of negligible sets and of measurable functions. We denote by $\mathcal{H}^k := L^2(X, \mathfrak{m}^k)$ and by $\mathsf{Ch}^k_*$ the Cheeger's energy associated to $\mathfrak{m}^k$ in $W^{1,2}(X, \mathsf{d}, \mathfrak{m}^k) \subset \mathcal{H}^k$, extended to $+\infty$ in $\mathcal{H}^0 \setminus W^{1,2}(X, \mathsf{d}, \mathfrak{m}^k)$. We have $\mathcal{H}^{k+1} \subset \mathcal{H}^k \subset \mathcal{H}^0$, with continuous inclusion and, by Lemma 4.11, $\mathsf{Ch}^k_* \le \mathsf{Ch}^{k+1}_*$.

**Proposition 4.17 ($\Gamma$-convergence)** *Let $(\mathfrak{m}^k)$ be an increasing sequence of $\sigma$-finite measures satisfying (4.35) and (4.36). If $f^k \in \mathcal{H}^k$ weakly converge in $\mathcal{H}^0$ to $f$ with $S := \limsup_k \int_X |f^k|^2 \,\mathrm{d}\mathfrak{m}^k < \infty$ then $f \in L^2(X, \mathfrak{m})$,*

$$\liminf_{k\to\infty} \int_X |f^k|^2 \,\mathrm{d}\mathfrak{m}^k \ge \int_X |f|^2 \,\mathrm{d}\mathfrak{m}, \qquad \liminf_{k\to\infty} \mathsf{Ch}^k_*(f^k) \ge \mathsf{Ch}_*(f), \tag{4.37}$$

*and*

$$\lim_{k\to\infty} \int_X f^k\, g \,\mathrm{d}\mathfrak{m}^k = \int_X f\, g \,\mathrm{d}\mathfrak{m} \quad \textit{for every } g \in L^2(X, \mathfrak{m}). \tag{4.38}$$

*Finally, if $S \le \int_X |f|^2 \,\mathrm{d}\mathfrak{m}$ then*

$$f^k \to f \quad \textit{strongly in } \mathcal{H}^0 \quad \textit{and} \quad \lim_{k\to\infty} \int_X |f^k|^2 \,\mathrm{d}\mathfrak{m}^k = \int_X |f|^2 \,\mathrm{d}\mathfrak{m}. \tag{4.39}$$

*Proof.* (4.37) is an easy consequence of the monotonicity of $\mathfrak{m}^k$, the lower semicontinuity of the $L^2$-norm with respect to weak convergence, and (4.20) of Lemma 4.11.

In order to check (4.38) notice that for every $g \in L^2(X, \mathfrak{m})$ and every $\upsilon > 0$

$$\int_X f^k\, g \,\mathrm{d}\mathfrak{m}^k = \frac{1}{2} \int_X (\upsilon f^k + \upsilon^{-1} g)^2 \,\mathrm{d}\mathfrak{m}^k - \frac{\upsilon^2}{2} \int_X |f^k|^2 \,\mathrm{d}\mathfrak{m}^k - \frac{1}{2\upsilon^2} \int_X |g|^2 \,\mathrm{d}\mathfrak{m}^k, \tag{4.40}$$

so that, taking the limit as $k \to \infty$,

$$\begin{aligned} \liminf_{k\to\infty} \int_X f^k\, g \,\mathrm{d}\mathfrak{m}^k &\ge \frac{1}{2} \int_X (\upsilon f + \upsilon^{-1} g)^2 \,\mathrm{d}\mathfrak{m} - \frac{\upsilon^2}{2} S - \frac{1}{2\upsilon^2} \int_X |g|^2 \,\mathrm{d}\mathfrak{m} \\ &= \int_X f g \,\mathrm{d}\mathfrak{m} + \frac{\upsilon^2}{2} \Big( \int_X |f|^2 \,\mathrm{d}\mathfrak{m} - S \Big). \end{aligned}$$

Passing to the limit as $\upsilon \downarrow 0$ and applying the same inequality with $-g$ in place of $g$ we get (4.38). Finally, (4.39) follows easily by (4.38) and the inequality $S \le \int_X |f|^2 \,\mathrm{d}\mathfrak{m}$, passing to the limit in

$$\int_X |f^k - f|^2 \,\mathrm{d}\mathfrak{m}^k = -2\int_X f^k f \,\mathrm{d}\mathfrak{m}^k + \int_X |f^k|^2 \,\mathrm{d}\mathfrak{m}^k + \int_X |f|^2 \,\mathrm{d}\mathfrak{m}^k.$$

□

Let us now consider the gradient flow $\mathsf{H}^k_t$ generated by $\mathsf{Ch}^k_*$ in $\mathcal{H}^k$ and the "limit" semigroup $\mathsf{H}_t$ generated by $\mathsf{Ch}_*$ in $\mathcal{H} = L^2(X,\mathfrak{m}) \subset \mathcal{H}^0$. Since any element $f_0$ of $\mathcal{H}$ belongs also to $\mathcal{H}^k$, the evolution $f^k_t := \mathsf{H}^k_t(f_0)$ is well defined for every $k$ and it is interesting to prove the convergence of $f^k_t$ to $f_t = \mathsf{H}_t(f_0)$ as $k \to \infty$ in the larger space $\mathcal{H}^0$.

**Theorem 4.18** *Let $f_0 \in L^2(X,\mathfrak{m}) \subset \mathcal{H}^0$ and let $f^k_t = \mathsf{H}^k_t(f_0) \in \mathcal{H}^0$ be the heat flow in $L^2(X,\mathfrak{m}^k)$, $f_t := \mathsf{H}_t(f_0) \in L^2(X,\mathfrak{m})$. Then for every $t \ge 0$ we have*

$$\lim_{k\to\infty} f^k_t = f_t \quad \textit{strongly in } \mathcal{H}^0, \quad \lim_{k\to\infty} \int_X |f^k_t|^2 \,\mathrm{d}\mathfrak{m}^k = \int_X |f_t|^2 \,\mathrm{d}\mathfrak{m}. \tag{4.41}$$

*Proof.* The following classical argument combines the Γ-convergence result of the previous proposition with resolvent estimates; the only technical issue here is that the gradient flows are settled in Hilbert spaces $\mathcal{H}^k$ also depending on $k$.

Let us fix $\lambda > 0$ and let us consider the family of resolvent operators $J^k_\lambda : \mathcal{H}^k \to \mathcal{H}^k$ which to every $f^k \in \mathcal{H}^k$ associate the unique minimizer $f^k_\lambda$ of

$$\mathcal{C}^k_\lambda(g; f^k) := \mathsf{Ch}^k_*(g) + \frac{\lambda}{2}\int_X |g - f^k|^2 \,\mathrm{d}\mathfrak{m}^k. \tag{4.42}$$

We first prove that if $\limsup_k \int_X |f^k|^2 \,\mathrm{d}\mathfrak{m} \le \int_X |f|^2 \,\mathrm{d}\mathfrak{m}$ then $f^k_\lambda := J^k_\lambda(f^k)$ converge to $f_\lambda := J_\lambda(f)$ as $k \to \infty$ according to (4.39). In fact we know that for every $g \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$

$$\mathsf{Ch}^k_*(f^k_\lambda) + \frac{\lambda}{2}\int_X |f^k_\lambda - f^k|^2 \,\mathrm{d}\mathfrak{m}^k \le \mathsf{Ch}^k_*(g) + \frac{\lambda}{2}\int_X |g - f^k|^2 \,\mathrm{d}\mathfrak{m}^k.$$

By the assumption on $f^k$ the right hand side of the previous inequality converges to $\mathsf{Ch}_*(g) + \frac{\lambda}{2}\int_X |g - f|^2 \,\mathrm{d}\mathfrak{m}$. Since the sequence $(f^k_\lambda)$ is uniformly bounded in $\mathcal{H}^0 = L^2(X,\mathfrak{m}_0)$, up to extracting a suitable subsequence we can assume that $f^k_\lambda$ weakly converge to some limit $\tilde f$ in $\mathcal{H}^0$; (4.37) yields

$$\begin{aligned}\mathsf{Ch}_*(\tilde f) + \frac{\lambda}{2}\int_X |\tilde f - f|^2 \,\mathrm{d}\mathfrak{m} &\le \liminf_{k\to\infty} \mathsf{Ch}^k_*(f^k_\lambda) + \frac{\lambda}{2}\int_X |f^k_\lambda - f^k|^2 \,\mathrm{d}\mathfrak{m}_k \\ &\le \mathsf{Ch}_*(g) + \frac{\lambda}{2}\int_X |g - f|^2 \,\mathrm{d}\mathfrak{m} = \mathcal{C}_\lambda(g; f),\end{aligned}$$

for every $g \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$. We deduce that $\tilde f = J_\lambda f$ is the unique minimizer of $g \mapsto \mathcal{C}_\lambda(g; f)$. In particular the whole sequence converge weakly to $J_\lambda f$ in $\mathcal{H}^0$ and moreover

$$\limsup_{k\to\infty} \int_X |f^k_\lambda - f^k|^2 \,\mathrm{d}\mathfrak{m}^k \le \int_X |f_\lambda - f|^2 \,\mathrm{d}\mathfrak{m}, \tag{4.43}$$

so that we can apply Proposition 4.17 and obtain (4.39) for the sequence $(f^k_\lambda)$.

Iterating this resolvent convergence property, we get the same result for the operator $(J^k_\lambda)^n$ obtained by $n$ iterated compositions of $J^k_\lambda$, for every $n \in \mathbb{N}$. By the general estimates for gradient flows, choosing $\lambda := n/t$, we know that

$$\int_X |H_t f_0 - (J_{n/t})^n f_0|^2 \,\mathrm{d}\mathfrak{m} \le \frac{t}{n}\mathsf{Ch}_*(f_0), \quad \int_X |H^k_t f_t - (J^k_{n/t})^n f_0|^2 \,\mathrm{d}\mathfrak{m}^k \le \frac{t}{n}\mathsf{Ch}^k_*(f_0) \le \frac{t}{n}\mathsf{Ch}_*(f_0).$$

Since for every $n$ and every $t > 0$ we have $\lim_{k\to\infty}(J^k_{n/t})^n f = (J_{n/t})^n f$ strongly in $\mathcal{H}^0$, combining the previous estimates we get the first convergence property of (4.41) when $\mathsf{Ch}_*(f_0) < \infty$. Since the domain of $\mathsf{Ch}_*$ is dense in $L^2(X,\mathfrak{m})$ and $\mathsf{H}_t$ is a contraction semigroup, a simple approximation argument yields the general case when $f_0 \in L^2(X,\mathfrak{m})$. Passing to the limit as $k \to \infty$ in the identities

$$\frac{1}{2}\int_X |f^k_t|^2 \,\mathrm{d}\mathfrak{m}^k + \int_0^t \mathsf{Ch}^k_*(f^k_s)\,\mathrm{d}s = \frac{1}{2}\int_X |f_0|^2\,\mathrm{d}\mathfrak{m}^k, \tag{4.44}$$

and taking into account the corresponding identity for $\mathfrak{m}$ and $\mathsf{Ch}_*$ and the lower semicontinuity property (4.37) for $\mathsf{Ch}^k_*$, we obtain the second limit of (4.41). □

### 4.4 Mass preservation and entropy dissipation when $\mathfrak{m}(X) = \infty$

Let us start by deriving useful "moment-entropy estimates", in the case of a measure $\mathfrak{m}$ with finite mass.

**Lemma 4.19 (Moment-entropy estimate)** *Let $\mathfrak{m}$ be a finite measure, let $V : X \to [0,\infty)$ be a Borel and $\mathsf{d}$-Lipschitz function, let $f_0 \in L^2(X,\mathfrak{m})$ be nonnegative with*

$$\int_X \mathrm{e}^{-V^2}\,\mathrm{d}\mathfrak{m} \le \int_X f_0\,\mathrm{d}\mathfrak{m}, \qquad \int_X V^2 f_0\,\mathrm{d}\mathfrak{m} < \infty, \tag{4.45}$$

*and let $f_t = \mathsf{H}_t(f_0)$ be the solution of (4.22). Then the map $t \mapsto \int_X V^2 f_t\,\mathrm{d}\mathfrak{m}$ is locally absolutely continuous in $[0,\infty)$ and for every $t \ge 0$*

$$\int_X V^2 f_t\,\mathrm{d}\mathfrak{m} \le \mathrm{e}^{4\mathrm{Lip}^2(V)t}\int_X f_0\Big(\log f_0 + 2V^2\Big)\,\mathrm{d}\mathfrak{m}, \tag{4.46}$$

$$\int_0^t \int_{\{f_s>0\}} \frac{|\nabla f_s|^2_*}{f_s}\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s \le 2\mathrm{e}^{4\mathrm{Lip}^2(V)t}\int_X f_0\Big(\log f_0 + V^2\Big)\,\mathrm{d}\mathfrak{m}. \tag{4.47}$$

*Proof.* We set $L = \mathrm{Lip}(V)$ and

$$M^2(t) := \int_X V^2 f_t\,\mathrm{d}\mathfrak{m}, \quad E(t) := \int_X f_t \log f_t\,\mathrm{d}\mathfrak{m}, \quad F^2(t) := \int_{\{f_t>0\}} \frac{|\nabla f_t|^2_*}{f_t}\,\mathrm{d}\mathfrak{m}. \tag{4.48}$$

Applying (4.29) to $(f_t + \varepsilon)$ and letting $\varepsilon \downarrow 0$ we get $F \in L^2(0,T)$ for every $T > 0$ with

$$\frac{\mathrm{d}}{\mathrm{d}t}E(t) = -F^2(t) \quad \text{a.e. in } (0,T). \tag{4.49}$$

The convexity inequality $r\log r \ge r - r_0 + r\log r_0$ with $r = f_t$, $r_0 = \mathrm{e}^{-V^2}$, and the conservation of the total mass (4.30) and (4.45) yield for every $t \ge 0$

$$E(t) \ge \int_X (f_t - \mathrm{e}^{-V^2})\,\mathrm{d}\mathfrak{m} - M^2(t) = \int_X (f_0 - \mathrm{e}^{-V^2})\,\mathrm{d}\mathfrak{m} - M^2(t) \ge -M^2(t).$$

We introduce now the truncated weight $V_k(x) = \min(V(x), k)$ and the corresponding functional $M_k^2(t)$ defined as in (4.48). Since the map $t \mapsto M_k^2(t)$ is Lipschitz continuous we get for a.e. $t > 0$

$$\Big|\frac{\mathrm{d}}{\mathrm{d}t} M_k^2(t)\Big| = \Big|\int_X V_k^2\, \Delta_{\mathsf{d},\mathfrak{m}} f_t \,\mathrm{d}\mathfrak{m}\Big| \le 2\int_X |\nabla f_t|_* |\nabla V_k|_* V_k \,\mathrm{d}\mathfrak{m} \le 2L\, F(t)\, M_k(t). \tag{4.50}$$

We deduce that

$$M_k(t) \le M_k(0) + L\int_0^t F(s)\,\mathrm{d}s \le M(0) + L\int_0^t F(s)\,\mathrm{d}s,$$

so that $M_k(t)$ is uniformly bounded. Passing to the limit in (an integral form of) (4.50) as $k \to \infty$ by monotone convergence, we obtain the same differential inequality for $M$

$$\Big|\frac{\mathrm{d}}{\mathrm{d}t} M^2(t)\Big| \le 2L\, F(t)\, M(t).$$

Combining with (4.49) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(E + 2M^2\big) + F^2 \le 4L\, F\, M \le F^2 + 4L^2\, M^2.$$

Since $E + 2M^2 \ge M^2$, Gronwall lemma yields

$$M^2(t) \le E(t) + 2M^2(t) \le \big(E(0) + 2M^2(0)\big)\mathrm{e}^{4L^2 t},$$

i.e. (4.46). Integrating now (4.49) we get $\int_0^t F^2(s)\,\mathrm{d}s \le E(0) - E(t) \le E(0) + M^2(t)$, which yields (4.47). □

We want now to extend the validity of (4.30), (4.46) and (4.47) to the case when $\mathfrak{m}(X) = \infty$, at least when (4.2) holds. Notice that this assumption also includes the cases when $\mathfrak{m}(X) \in (0,\infty)$.

**Theorem 4.20** *If $\mathfrak{m}$ is a $\sigma$-finite measure satisfying* (4.2)*, then the gradient flow $\mathsf{H}_t$ of the Cheeger's energy is mass preserving (i.e.* (4.30) *holds) and satisfies the contraction estimate* (4.31)*, in particular* (4.27) *for every $p \in [1,\infty]$. Moreover, for every nonnegative $f_0 \in L^2(X,\mathfrak{m})$ with*

$$f_0 \log f_0 \in L^1(X,\mathfrak{m}), \qquad \int_X V^2\, f_0 \,\mathrm{d}\mathfrak{m} < \infty, \qquad \int_X f_0 \,\mathrm{d}\mathfrak{m} = 1, \tag{4.51}$$

*the solution $f_t = \mathsf{H}_t(f_0)$ of* (4.22) *satisfies* (4.46) *and* (4.47) *for every $t \ge 0$.*

*Proof.* Let us first prove mass preservation and (4.46), (4.47) for a nonnegative initial datum satisfying (4.51). The proof is based on a simple approximation result. We set $\mathfrak{m}^0 := \mathrm{e}^{-V^2}\mathfrak{m}$, $V_k := \min(V,k)$ and $\mathfrak{m}^k := \mathrm{e}^{V_k^2}\mathfrak{m}^0 = \mathrm{e}^{V_k^2 - V^2}\mathfrak{m}$, so that that $\mathfrak{m}^k$ is an increasing family of finite measures satisfying conditions (4.35), by monotone convergence. In addition, since by (4.2) $V$ is $\mathsf{d}$-Lipschitz and bounded from above on compact sets, (4.36) holds.

We define $f_t^k = \mathsf{H}_t^k(f_0)$ as in the Theorem 4.18 and $z_k := \int_X f_0\,\mathrm{d}\mathfrak{m}^k$. We apply (4.30) to obtain that $\int_X f_t^k\,\mathrm{d}\mathfrak{m}^k = z_k$ for all $t \ge 0$; then, since $z_k \uparrow 1$ we can find $\eta_k \downarrow 1$ such that

$\int \mathrm{e}^{-\eta_k^2 V^2}\,\mathrm{d}\mathfrak{m}^k \le z_k$ and apply the estimates of Lemma 4.19 with $\mathfrak{m} := \mathfrak{m}^k$ and weight $\eta_k V$ to obtain

$$\eta_k^2 \int_X V^2\, f_t^k\,\mathrm{d}\mathfrak{m}^k \le \mathrm{e}^{4\eta_k^2 \mathrm{Lip}^2(V)t} \int_X f_0(\log f_0 + 2V^2)\,\mathrm{d}\mathfrak{m}^k \le \mathrm{e}^{4\eta_k^2 \mathrm{Lip}^2(V)t} \int_X f_0(\log f_0 + 2V^2)\,\mathrm{d}\mathfrak{m}. \tag{4.52}$$

Since, thanks to (4.39), $f_t^k \to f_t$ strongly in $L^2(X,\mathfrak{m}^0)$ as $k \to \infty$, we get up to subsequences $f_t^k \to f_t$ $\mathfrak{m}$-a.e., so that Fatou's lemma and the monotonicity of $\mathfrak{m}^k$ yield

$$\int_X V^2\, f_t\,\mathrm{d}\mathfrak{m} \le \liminf_{k\to\infty} \int_X V^2\, f_t^k\,\mathrm{d}\mathfrak{m}^k,$$

and (4.46) follows by (4.52).

Let us consider now $A_h := \{x \in X : V(x) \le h\}$ and observe that (4.2) and (4.38) yield

$$\mathfrak{m}(A_h) \le \int_X \mathrm{e}^{h^2 - V^2}\,\mathrm{d}\mathfrak{m} \le \mathrm{e}^{h^2} < \infty, \qquad \int_{A_h} f_t\,\mathrm{d}\mathfrak{m} = \lim_{k\to\infty} \int_{A_h} f_t^k\,\mathrm{d}\mathfrak{m}^k.$$

From (4.46) we obtain for every $t > 0$ a constant $C$ satisfying $h^2 \int_{X\setminus A_h} f_t^k\,\mathrm{d}\mathfrak{m}^k \le C$ for every $h > 0$, so that

$$\int_X f_t\,\mathrm{d}\mathfrak{m} \ge \int_{A_h} f_t\,\mathrm{d}\mathfrak{m} = \lim_{k\to\infty} \int_{A_h} f^k\,\mathrm{d}\mathfrak{m}^k \ge 1 - \limsup_{k\to\infty} \int_{X\setminus A_h} f_t^k\,\mathrm{d}\mathfrak{m}^k \ge 1 - C/h^2.$$

Since $h$ is arbitrary and the integral of $f_t$ does not exceed 1 by (4.28), we showed that $\int_X f_t\,\mathrm{d}\mathfrak{m} = 1$. Finally, (4.47) follows now by the lower semicontinuity (4.37) of the Cheeger's energy from the corresponding estimate for $f_t^k$, recalling (4.18).

Let us now consider an initial datum $f_0 \in L^2(X,\mathfrak{m})$ with arbitrary sign and vanishing outside some $A_n$, so that $|f_0|$ satisfies (4.51) (up to a multiplication for a suitable constant). The comparison principle yields $\big|\mathsf{H}_t^k(f_0)\big| \le \mathsf{H}_t^k(|f_0|)$, so that for every $t > 0$ there exists a constant $C$ such that $h^2 \int_{X\setminus A_h} |f_t^k|\,\mathrm{d}\mathfrak{m}^k \le C$. Since $\int_X f_t^k\,\mathrm{d}\mathfrak{m}^k = \int_X f_0\,\mathrm{d}\mathfrak{m}^k$ by (4.30), we thus have

$$\begin{aligned}\Big|\int_X (f_t - f_0)\,\mathrm{d}\mathfrak{m}\Big| \le \int_{X\setminus A_h} |f_t^k|\,\mathrm{d}\mathfrak{m}^k + \int_{X\setminus A_h} |f_t|\,\mathrm{d}\mathfrak{m} + \Big|\int_{A_h} f_t\,\mathrm{d}\mathfrak{m} - \int_{A_h} f_t^k\,\mathrm{d}\mathfrak{m}^k\Big| \\ + \int_X |f_0|\,\mathrm{d}(\mathfrak{m} - \mathfrak{m}^k).\end{aligned}$$

Passing to the limit in the previous inequality first as $k \to \infty$, taking (4.38) into account, and then as $h \to \infty$ we obtain that the integral of $f_t$ is constant in time. As in the proof of Theorem 4.16(d) we can show that $\mathsf{H}_t$ satisfies the contraction estimate (4.27) for $p = 1$ and arbitrary couples of initial data vanishing outside $A_n$. Approximating any $f_0 \in L^2(X,\mathfrak{m}) \cap L^1(X,\mathfrak{m})$ by the sequence $\chi_{A_n} f_0$ we can easily extend the contraction property and the mass conservation to arbitrary initial data.

The contraction property (4.31) (and (4.27) for $p \in (1,2)$) then follows as in the proof of Theorem 4.16, (d). □

**Remark 4.21** It is interesting to compare the mass preservation property of Theorem 4.20 relying on (4.2) with the well known results for the Heat flow on a smooth, complete, finite dimensional, Riemannian manifold $(X, \mathsf{d}, \mathfrak{m})$, where $\mathsf{d}$ (resp. $\mathfrak{m}$) is the induced Riemannian distance (resp. volume measure). In this case, a sufficient condition [18, Theorem 9.1] is

$$\int_{r_0}^{\infty} \frac{r}{\log\big(m(r)\big)}\,\mathrm{d}r = \infty, \quad \text{for some } \ r_0 > 0, \quad m(r) := \mathfrak{m}\big(\{x : \mathsf{d}(x, x_0) < r\}\big), \tag{4.53}$$

which is obviously a consequence of (4.3). On the other hand, (4.3) is always satisfied if the Ricci curvature of $X$ is bounded from below: more generally (4.3) holds in metric spaces satisfying the $CD(K, \infty)$ condition, see Section 9 and [35, Theorem 4.24].

**Proposition 4.22 (Entropy dissipation)** *Let $\mathfrak{m}$ be a $\sigma$-finite measure satisfying* (4.2)*, let $f_0 \in L^2(X, \mathfrak{m})$ be a nonnegative initial datum with $\int_X f_0\,\mathrm{d}\mathfrak{m} = 1$ and $\int_X V^2 f_0\,\mathrm{d}\mathfrak{m} < \infty$ and let $(f_t)$ be the corresponding gradient flow of Cheeger's energy. Then the map $t \mapsto \int_X f_t \log f_t\,\mathrm{d}\mathfrak{m}$ is locally absolutely continuous in $(0, \infty)$ and it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_X f_t \log f_t\,\mathrm{d}\mathfrak{m} = -\int_{\{f_t>0\}} \frac{|\nabla f_t|_*^2}{f_t}\,\mathrm{d}\mathfrak{m} \qquad \textit{for a.e. } t \in (0, \infty). \tag{4.54}$$

*Proof.* The case when $\mathfrak{m}(X) < \infty$ can be easily deduced by Proposition 4.16(c). If $\mathfrak{m}(X) = \infty$ we first consider regularized $C^{1,1}(0, \infty)$ and convex entropies $e_\varepsilon$, $0 < \varepsilon < \mathrm{e}^{-1}$, of $e(r) := r \log r$:

$$\begin{cases} e_\varepsilon(r) = (1 + \log\varepsilon)r = e'(\varepsilon)r & \text{in } [0, \varepsilon],\\ e_\varepsilon(r) = r\log r + \varepsilon = e(r) - e(\varepsilon) + \varepsilon e'(\varepsilon) & \text{in } [\varepsilon, \infty). \end{cases}$$

Notice that $e_\varepsilon'(r) = \max\{e'(r), e'(\varepsilon)\} \le (1 + \log r)^+$ because of our choice of $\varepsilon$; since $(1 + \log r)^+ \le r$, we deduce that $e(r) \le e_\varepsilon(r) \le \frac{1}{2}r^2$ and $e_\varepsilon(r) \downarrow e(r)$ as $\varepsilon \downarrow 0$.

We can now define a convex and $C^{1,1}(\mathbb{R})$ function by setting $\tilde e_\varepsilon(r) := e_\varepsilon(r) - (1+\log\varepsilon)r$ for $r \ge 0$ and $\tilde e_\varepsilon(r) \equiv 0$ for $r < 0$; applying (4.29) (by the previous estimates $\int_X \tilde e_\varepsilon(f_0)\,\mathrm{d}\mathfrak{m} < \infty$) and recalling that the integral of $f_t$ is constant for every $t \ge 0$ we obtain

$$\int_X e_\varepsilon(f_t)\,\mathrm{d}\mathfrak{m} + \int_0^t \int_{\{f_t > \varepsilon\}} \frac{|\nabla f_t|_*^2}{f_t}\,\mathrm{d}\mathfrak{m}\,\mathrm{d}t = \int_X e_\varepsilon(f_0)\,\mathrm{d}\mathfrak{m}.$$

Passing to the limit as $\varepsilon \downarrow 0$ and recalling the uniform bounds (4.46) and (4.47), guaranteed by Theorem 4.20, we conclude. □

**Remark 4.23** Although these facts will not play a role in the paper, we emphasize that it is also possible to define one-sided Cheeger energies $\mathsf{Ch}_*^+(f)$, $\mathsf{Ch}_*^-(f)$, by relaxing respectively the ascending and descending slopes of Borel and $\mathsf{d}$-Lipschitz functions w.r.t. $L^2(X, \mathfrak{m})$ convergence. We still have the representation

$$\mathsf{Ch}_*^+(f) = \frac{1}{2}\int_X |\nabla^+ f|_*^2\,\mathrm{d}\mathfrak{m}, \qquad \mathsf{Ch}_*^-(f) = \frac{1}{2}\int_X |\nabla^- f|_*^2\,\mathrm{d}\mathfrak{m}$$

for suitable one-sided relaxed gradients $|\nabla^\pm f|_*$ with minimal norm and it is easily seen that the functionals $\mathsf{Ch}_*^\pm$ are convex and lower semicontinuous in $L^2(X, \mathfrak{m})$.

Obviously $\mathsf{Ch}_* \ge \max\{\mathsf{Ch}_*^+, \mathsf{Ch}_*^-\}$ and $\mathsf{Ch}_*^+(f) = \mathsf{Ch}_*^-(-f)$. Lemma 4.3 still holds with the same proof and, using (2.9), locality can be proved for the one-sided relaxed gradients as well, so that $|\nabla^\pm f|_* \le |\nabla^\pm f|$ $\mathfrak{m}$-a.e. for $f$ Borel and $\mathsf{d}$-Lipschitz. We shall see in the next section that if $\mathfrak{m}$ satisfies (4.2) then for every Borel function $|\nabla^\pm f|_* = |\nabla f|_*$ and $\mathsf{Ch}_* = \mathsf{Ch}_*^+ = \mathsf{Ch}_*^-$.
■

# 5 Weak upper gradients

In this subsection we define a new notion for the "weak norm of the gradient" (which we will call "minimal weak upper gradient") of a real valued functions $f$ on an extended metric space $(X, \mathsf{d})$ and we will show that this new notion essentially coincides with the relaxed gradient. The approach that we use here is inspired by the work [34], i.e. rather than proceeding by relaxation, as we did for $|\nabla f|_*$, we ask the fundamental theorem of calculus to hold along "most" absolutely continuous curves, in a sense that we will specify soon. Our definition of null set of curves is different from [34], natural in the context of optimal transportation, and leads to an a priori larger class of null sets, see Remark 5.3; also, another difference is that we consider Sobolev regularity (and not absolute continuity) along every curve, so that our theory does not depend on the choice of precise representatives in the Lebesgue equivalence class. In Remark 5.10 we compare more closely the two approaches and show, as a nontrivial consequence of our identification results, that they lead to the same Sobolev space.

The advantages of working with a direct definition, rather than proceeding by relaxation, can be appreciated by looking at Lemma 5.15, where we prove absolute continuity of functionals $t \mapsto \int_X \phi(f_t)\,\mathrm{d}\mathfrak{m}$ even along curves $t \mapsto f_t\mathfrak{m}$ that are absolutely continuous in the Wasserstein sense, compare with Proposition 4.22 for $L^2$-gradient flows; we can also compute the minimal weak upper gradient for Kantorovich potentials, as we will see in Section 10.

We assume in this section that $(X, \tau, \mathsf{d})$ is an extended Polish space and that $\mathfrak{m}$ is a $\sigma$-finite Borel measure in $X$ representable in the form $\mathrm{e}^{V^2}\tilde{\mathfrak{m}}$ with $\tilde{\mathfrak{m}}(X) \le 1$ and $V : X \to [0, \infty)$ Borel and $\mathsf{d}$-Lipschitz. Recall that the $p$-energy of an absolutely continuous curve has been defined in (2.2), as well as the collection of curves of finite $p$-energy $\mathrm{AC}^p((0,1); (X, \mathsf{d}))$, which we will consider as a Borel subset of $C([0,1]; X)$ (and in particular a Borel subset of a Polish space).

## 5.1 Test plans, Sobolev functions along a.c. curves, and weak upper gradients

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

$$\mathrm{restr}_t^s(\gamma)_r := \gamma_{((1-r)t+rs)}, \tag{5.1}$$

so that $\mathrm{restr}_t^s$ restricts the curve $\gamma$ to the interval $[t, s]$ and then "stretches" it on the whole of $[0, 1]$.

**Definition 5.1 (Test plans)** *We say that a probability measure $\boldsymbol{\pi} \in \mathscr{P}(C([0,1]; X))$ is a test plan if it is concentrated on $\mathrm{AC}((0,1); (X, \mathsf{d}))$, i.e. $\boldsymbol{\pi}\big(C([0,1]; X) \setminus \mathrm{AC}((0,1); (X, \mathsf{d}))\big) = 0$, and*

$$(\mathrm{e}_t)_\sharp \boldsymbol{\pi} \ll \mathfrak{m} \quad \text{for all} \quad t \in [0,1]. \tag{5.2}$$

*A collection $\mathfrak{T}$ of test plans is stretchable if*

$$\boldsymbol{\pi} \in \mathfrak{T} \quad \Longrightarrow \quad (\mathrm{restr}_t^s)_\sharp \boldsymbol{\pi} \in \mathfrak{T} \quad \text{for every } 0 \le t \le s \le 1. \tag{5.3}$$

We will often impose additional quantitative assumptions on test plans, besides (5.2). The most important one, which we call *bounded compression*, provides a locally uniform upper bound on the densities of $(\mathrm{e}_t)_\sharp\boldsymbol{\pi}$. More precisely, a test plan $\boldsymbol{\pi}$ has bounded compression on the sublevels of $V$ if for all $M \geq 0$ there exists $C = C(\boldsymbol{\pi}, M) \in [0, \infty)$ satisfying

$$(\mathrm{e}_t)_\sharp\boldsymbol{\pi}(B \cap \{V \leq M\}) \leq C(\boldsymbol{\pi}, M)\,\mathfrak{m}(B) \qquad \forall B \in \mathscr{B}(X),\ t \in [0,1]. \tag{5.4}$$

The above condition (5.4) depends not only on $\mathfrak{m}$, but also on $V$. For finite measures $\mathfrak{m}$ it will be understood that we take $V$ equal to a constant, so that (5.4) does not depend on the value of the constant.

Taking (5.4) into account, typical examples of stretchable collections $\mathcal{T}$ are the families of all the test plans with bounded compression which are concentrated on absolutely continuous curves, or on the curves of finite 2-energy, or on the geodesics in $X$.

**Definition 5.2 (Negligible sets of curves)** *Let $\mathcal{T}$ be a stretchable collection of test plans. We say that a set $N \subset \mathrm{AC}((0,1);(X,\mathsf{d}))$ is $\mathcal{T}$-negligible provided $\boldsymbol{\pi}(N) = 0$ for any test plan $\boldsymbol{\pi} \in \mathcal{T}$. A property which holds for every curve of some Borel set $A \subset \mathrm{AC}((0,1);(X,\mathsf{d}))$ except possibly for a negligible subset of $A$ is said to hold for $\mathcal{T}$-almost all curves in $A$.*

When $\mathcal{T}$ is a stretchable collection of test plans, we say that $f : X \to \mathbb{R}$ is *Sobolev* along $\mathcal{T}$-almost all curves if, for $\mathcal{T}$-almost all absolutely continuous curves $\gamma$, $f \circ \gamma$ coincides a.e. in $[0,1]$ and in $\{0,1\}$ with an absolutely continuous map $f_\gamma : [0,1] \to \mathbb{R}$.

In the next remark we compare our definition with the more classical notion of $\mathrm{Mod}_2$-null set of absolutely continuous curve used in [34].

**Remark 5.3** Recall that, for a collection $\Gamma$ of absolutely continuous curves in $(X,\mathsf{d})$, the 2-modulus $\mathrm{Mod}_2(\Gamma)$ is defined by

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

If $\mathcal{T}$ denotes the class of plans with bounded compression defined by (5.4), it is not difficult to show that Borel and $\mathrm{Mod}_2$-null sets of curves are $\mathcal{T}$-negligible. Indeed, if $\boldsymbol{\pi} \in \mathcal{T}$ has (with no loss of generality) finite 2-action and is concentrated on curves contained in $\{V \leq M\}$ and $\int_\gamma g \geq 1$ for all $\gamma \in \Gamma$, we can integrate w.r.t. $\boldsymbol{\pi}$ and then minimize w.r.t. $g$ to get

$$\left[\boldsymbol{\pi}(\Gamma)\right]^2 \leq C(\boldsymbol{\pi}, M)\mathrm{Mod}_2(\Gamma)\int\int_0^1 |\dot\gamma|^2\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}(\gamma).$$

Proving equivalence of the two concepts seems to be difficult, also because one notion is independent of parameterization, while the other one (because of the bounded compression condition) takes into account also the way curves are parameterized. ■

**Definition 5.4 (Weak upper gradients)** *Let $\mathcal{T}$ be a stretchable collection of test plans. Given $f : X \to \mathbb{R}$ Sobolev along $\mathcal{T}$-almost all curves, a $\mathfrak{m}$-measurable function $G : X \to [0,\infty]$ is a $\mathcal{T}$-weak upper gradient of $f$ (or a weak upper gradient w.r.t. $\mathcal{T}$) if*

$$\left|\int_{\partial\gamma} f\right| \leq \int_\gamma G \qquad \text{for } \mathcal{T}\text{-almost all } \gamma \in \mathrm{AC}((0,1);(X,\mathsf{d})). \tag{5.5}$$

Notice that the measurability of $s \mapsto G(\gamma_s)$ in $[0,1]$ for $\mathcal{T}$-almost every $\gamma$ is a direct consequence of the $\mathfrak{m}$-measurability of $G$: indeed, if $\tilde{G}$ is a Borel modification of $G$, $A \supset \{G \neq \tilde{G}\}$ is a $\mathfrak{m}$-negligible Borel set and $\boldsymbol{\pi}$ is a test plan we have by (5.2) that $\boldsymbol{\pi}(\{\gamma_t \in A\}) = (\mathrm{e}_t)_\sharp \boldsymbol{\pi}(A) = 0$ for every $t \in [0,1]$, so that

$$0 = \int_0^1 \boldsymbol{\pi}(\{\gamma_t \in A\})\,\mathrm{d}t = \int_0^1 \int \chi_{\{\gamma_t \in A\}}\,\mathrm{d}\boldsymbol{\pi}(\gamma)\,\mathrm{d}t = \int \Big(\int_0^1 \chi_{\{\gamma_t \in A\}}\,\mathrm{d}t\Big)\mathrm{d}\boldsymbol{\pi}(\gamma).$$

$\int_0^1 \chi_{\{\gamma_t \in A\}}\,\mathrm{d}t$ is therefore null for $\boldsymbol{\pi}$-a.e. $\gamma$. For any curve $\gamma$ for which the integral is null $G(\gamma_t)$ coincides a.e. in $[0,1]$ with the Borel map $\tilde{G}(\gamma_t)$.

**Remark 5.5 (Slopes of $\mathsf{d}$-Lipschitz functions are weak upper gradients)** As we explained in Remark 2.6, if $f : X \to \mathbb{R}$ is Borel and $\mathsf{d}$-Lipschitz, then the local Lipschitz constant $|\nabla f|$ and the one-sided slopes are upper gradients. Therefore they are also weak upper gradients w.r.t. any stretchable collection of test plans sense we just defined. Notice that the $\mathfrak{m}$-measurability of the slopes is ensured by Lemma 2.4. ■

**Remark 5.6 (Restriction and equivalent formulation)** If $G$ is a $\mathcal{T}$-weak upper gradient of $f$, then the strechable condition (5.3) yields for every $t < s$ in $[0,1]$

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

It follows that for $\mathcal{T}$-almost all $\gamma$ the function $f_\gamma$ satisfies

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

Since $f_\gamma$ is continuous the same holds for all $t < s$ in $[0,1]$, so that we obtain an equivalent pointwise formulation of (5.5):

$$\left|\frac{\mathrm{d}}{\mathrm{d}t} f_\gamma\right| \leq G \circ \gamma |\dot{\gamma}| \quad \text{a.e. in } [0,1], \text{ for } \mathcal{T}\text{-almost all } \gamma \in \mathrm{AC}((0,1);(X,\mathsf{d})). \tag{5.6}$$

■

## 5.2 Calculus with weak upper gradients

**Proposition 5.7 (Locality)** *Let $\mathcal{T}$ be a strechable collection of test plans, let $f : X \to \mathbb{R}$ be Sobolev along $\mathcal{T}$-almost all absolutely continuous curves, and let $G_1$, $G_2$ be weak upper gradients of $f$ w.r.t. $\mathcal{T}$. Then $\min\{G_1, G_2\}$ is a $\mathcal{T}$-weak upper gradient of $f$.*

*Proof.* It is a direct consequence of (5.6). □

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

**Proposition 5.8 (Invariance under modifications in $\mathfrak{m}$-negligible sets)** *Let $\mathcal{T}$ be a stretchable collection of test plans, let $f$, $\tilde{f} : X \to \mathbb{R}$ and $G$, $\tilde{G} : X \to [0,\infty]$ be such that both $\{f \neq \tilde{f}\}$ and $\{G \neq \tilde{G}\}$ are $\mathfrak{m}$-negligible. Assume that $f$ is Sobolev along $\mathcal{T}$-almost all curves and that $G$ is a $\mathcal{T}$-weak upper gradient of $f$. Then $\tilde{f}$ is Sobolev along $\mathcal{T}$-almost all curves and $\tilde{G}$ is a $\mathcal{T}$-weak upper gradient of $\tilde{f}$*

*Proof.* Fix a test plan $\boldsymbol{\pi}$: it is sufficient to prove that the sets $\{\gamma:\ f(\gamma_t)\neq\tilde f(\gamma_t)\}$, $t=0,1$, the set $\{\gamma:\ \int_\gamma G\neq\int_\gamma\tilde G\}$ and the set $\{\gamma:\ \mathscr{L}^1(\{s:f(\gamma_s)\neq\tilde f(\gamma_s)\})>0\}$ are contained in $\boldsymbol{\pi}$-negligible Borel sets.

For the first two sets the proof is obvious, because $(\mathrm{e}_t)_\sharp\boldsymbol{\pi}\ll\mathfrak{m}$, which implies that if $A$ is a $\mathfrak{m}$-negligible Borel set containing $\{f\neq\tilde f\}$ we have $\boldsymbol{\pi}(\{\gamma:\ \gamma_t\in A\})=(\mathrm{e}_t)_\sharp\boldsymbol{\pi}(A)=0$. For the third one we choose as $A$ a $\mathfrak{m}$-negligible Borel set containing $\{G\neq\tilde G\}$ and we use the argument described immediately after Definition 5.4. For the fourth one we choose a $\mathfrak{m}$-negligible Borel set $A$ containing $\{f\neq\tilde f\}$ and argue as for the third. □

Thanks to the previous proposition we can also consider extended real valued $f$ (as Kantorovich potentials), provided the set $N=\{|f|=\infty\}$ is $\mathfrak{m}$-negligible: as a matter of fact the curves $\gamma$ which intersect $N$ at $t=0$ or $t=1$ are negligible, hence $\int_{\partial\gamma}f$ is defined for almost every $\gamma$.

**Definition 5.9 (Minimal weak upper gradient)** *Let $\mathcal{T}$ be a stretchable collection of test plans and let $f:X\to\mathbb{R}$ be Sobolev along $\mathcal{T}$-almost all absolutely continuous curves. The $\mathcal{T}$-minimal weak upper gradient $|\nabla f|_{w,\mathcal{T}}$ of $f$ is the weak upper gradient characterized, up to $\mathfrak{m}$-negligible sets, by the property*

$$|\nabla f|_{w,\mathcal{T}}\leq G\qquad \mathfrak{m}\text{-a.e. in } X,\text{ for every }\mathcal{T}\text{-weak upper gradient } G \text{ of } f. \tag{5.7}$$

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

$$\inf\left\{\int_X\tan^{-1}G\,\mathrm{d}\mathfrak{m}^0:\ G\text{ is a }\mathcal{T}\text{-weak upper gradient of }f\right\}.$$

We immediately see, thanks to Proposition 5.7, that we can assume with no loss of generality that $G_{n+1}\leq G_n$. Hence, by monotone convergence, the function $|\nabla f|_{w,\mathcal{T}}$ is a weak upper gradient of $f$ on $\mathcal{A}$ and $\int_X\tan^{-1}G\,\mathrm{d}\mathfrak{m}^0$ is minimal at $G=|\nabla f|_{w,\mathcal{T}}$. This minimality, in conjunction with Proposition 5.7, gives (5.7).

**Remark 5.10 (Comparison with Newtonian spaces)** Shanmugalingam introduced in [34] the Newtonian space $N^{1,2}(X,\mathsf{d},\mathfrak{m})$ of all functions $f:X\to\mathbb{R}$ such that $\int f^2\,\mathrm{d}\mathfrak{m}<\infty$ and the inequality

$$|f(\gamma_1)-f(\gamma_0)|\leq\int_\gamma G \tag{5.8}$$

holds out of a $\mathrm{Mod}_2$-null set of curves, for some $G\in L^2(X,\mathfrak{m})$. Then, she defined $|\nabla f|_S$ as the function $G$ in (5.8) with smallest $L^2$ norm and proved [34, Proposition 3.1] that functions in $N^{1,2}(X,\mathsf{d},\mathfrak{m})$ are absolutely continuous along $\mathrm{Mod}_2$-almost every curve.

Remarkably, Shanmugalingam proved (the proofs in [34] work, with no change, even in the case of extended metric measure spaces) this connection between Newtonian spaces and Cheeger's functional $\underline{\mathsf{Ch}}_*$ described in Remark 4.6: $f\in D(\underline{\mathsf{Ch}}_*)$ if and only if there is $\tilde f\in N^{1,2}(X,\mathsf{d},\mathfrak{m})$ in the Lebesgue equivalence class of $f$, and the two notions of gradient $|\nabla f|_S$ and $|\nabla f|_C$ coincide $\mathfrak{m}$-a.e. in $X$.

Now, the inclusion between null sets provided by Remark 5.3 shows that the situation described in Remark 4.6 is reversed. Indeed, while $|\nabla f|_C\leq|\nabla f|_*$, the gradient $|\nabla f|_S$ is larger $\mathfrak{m}$-a.e. than $|\nabla f|_{w,\mathcal{T}}$, so that

$$|\nabla f|_{w,\mathcal{T}}\leq|\nabla f|_S=|\nabla f|_C\leq|\nabla f|_*\qquad\mathfrak{m}\text{-a.e. in } X.$$

Although we are not presently able to reverse the inclusion between null sets, a nontrivial consequence of our identification of $|\nabla f|_{w,\mathcal{T}}$ and $|\nabla f|_*$, proved in the next section, is that all these gradients coincide $\mathfrak{m}$-a.e. in $X$.

Since $D(\mathsf{Ch}_*) \subset D(\underline{\mathsf{Ch}}_*)$, a byproduct of the absolute continuity of functions in Newtonian spaces, that however will not play a role in our paper, is that functions in $D(\mathsf{Ch}_*)$ have a version which is absolutely continuous along $\mathrm{Mod}_2$-a.e. curve. ■

**Remark 5.11** Notice that the notion of weak gradient do depend on the class $\mathcal{T}$ of test plans (which, in turn, might depend on $V$).

If $\mathcal{T}_1 \subset \mathcal{T}_2$ are stretchable collections of test plans and a function $f : X \to \mathbb{R}$ is Sobolev along $\mathcal{T}_2$-almost all absolutely continuous curves, then $f$ is Sobolev along $\mathcal{T}_1$-almost all absolutely continuous curves and

$$|\nabla f|_{w,\mathcal{T}_1} \le |\nabla f|_{w,\mathcal{T}_2}. \tag{5.9}$$

Thus larger classes of test plans induce smaller classes of weak upper gradients, hence larger minimal weak upper gradients. ■

Another important property of weak upper gradients is their stability w.r.t. $L^p$ convergence: we state it for all the stretchable classes of test plans satisfying a condition weaker than bounded compression, inspired to the "democratic" condition introduced by [27].

**Theorem 5.12 (Stability w.r.t. $\mathfrak{m}$-a.e. convergence)** *Let us suppose that $\mathcal{T}$ is a stretchable collection of test plans concentrated on $\mathrm{AC}^p((0,1);(X,\mathsf{d}))$ for some $p \in (1,\infty]$ such that for all $\boldsymbol{\pi} \in \mathcal{T}$ and all $M \ge 0$ there exists $C = C(\boldsymbol{\pi}, M) \in [0,\infty)$ satisfying*

$$\int_0^1 (\mathrm{e}_t)_\sharp \boldsymbol{\pi}(B \cap \{V \le M\})\,\mathrm{d}t \le C(\boldsymbol{\pi}, M)\,\mathfrak{m}(B) \qquad \forall B \in \mathscr{B}(X). \tag{5.10}$$

*Assume that $f_n$ are $\mathfrak{m}$-measurable, Sobolev along $\mathcal{T}$-almost all curves and that $G_n$ are $\mathcal{T}$-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 $L^q(\{V \le M\}, \mathfrak{m})$ for all $M \ge 0$, where $q \in [1,\infty)$ is the conjugate exponent of $p$. Then $G$ is a $\mathcal{T}$-weak upper gradient of $f$.*

*Proof.* Fix a test plan $\boldsymbol{\pi}$ and assume with no loss of generality that $\mathcal{E}_p[\gamma] \le L < \infty$ $\boldsymbol{\pi}$-a.e. (recall (2.2)). 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 $L^q(X, \{V \le M\})$. 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 $\{V \le M\}$.

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

$$
\begin{aligned}
\int \Big( \int_{\gamma \cap \{V \le M\}} \varphi \Big) \, \mathrm{d}\boldsymbol{\pi} &= \int \Big( \int_0^1 \chi_{\{V \le M\}}(\gamma_t) \varphi(\gamma_t) |\dot\gamma_t| \, \mathrm{d}t \Big) \, \mathrm{d}\boldsymbol{\pi} \\
&\le \int \Big( \int_0^1 \chi_{\{V \le M\}}(\gamma_t) \varphi^q(\gamma_t) \, \mathrm{d}t \Big)^{1/q} \Big( \int_0^1 |\dot\gamma_t|^p \, \mathrm{d}t \Big)^{1/p} \mathrm{d}\boldsymbol{\pi} \\
&\le \Big( \int_0^1 \int_{\{V \le M\}} \varphi^q \, \mathrm{d}(\mathrm{e}_t)_\sharp \boldsymbol{\pi} \, \mathrm{d}t \Big)^{1/q} \Big( \int \mathcal{E}_p[\gamma] \, \mathrm{d}\boldsymbol{\pi} \Big)^{1/p} \\
&\le \Big( C \int_{\{V \le M\}} \varphi^q \, \mathrm{d}\mathfrak{m} \Big)^{1/q} \Big( \int \mathcal{E}_p[\gamma] \, \mathrm{d}\boldsymbol{\pi} \Big)^{1/p}, \qquad (5.11)
\end{aligned}
$$

we obtain, for $\bar C := C^{1/q} L^{1/p}$

$$
\begin{aligned}
&\int \Big( \int_{\gamma \cap \{V \le M\}} |H_n - G| + \min\{|\tilde f_n - f|, 1\} \Big) \, \mathrm{d}\boldsymbol{\pi} \\
&\quad \le \bar C \Big( \|H_n - G\|_{L^q(\{V \le M\}, \mathfrak{m})} + \|\min\{|\tilde f_n - f|, 1\}\|_{L^q(\{V \le M\}, \mathfrak{m})} \Big) \to 0.
\end{aligned}
$$

By a diagonal argument we can find a subsequence $n(k)$ independent of $M \in \mathbb{N}$ 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. $\gamma$ contained in $\{V \le M\}$, and thus for $\boldsymbol{\pi}$-a.e. $\gamma$. 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. $\gamma$ 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 $\gamma$ 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 we obtain an absolutely continuous function $f_\gamma$ 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 in $\mathcal{T}$ we conclude that $f$ is Sobolev along $\mathcal{T}$-almost all curves and that $G$ is a $\mathcal{T}$-weak upper gradient of $f$. □

**Corollary 5.13** *Let $\mathcal{T}$ be a stretchable collection of test plans satisfying* (5.10) *and concentrated on* $\mathrm{AC}^2((0, 1); (X, \mathsf{d}))$*, and let $\tilde{\mathsf{Ch}}$ be defined as in* (4.15)*. If $f \in D(\tilde{\mathsf{Ch}})$ then $f$ is Sobolev along $\mathcal{T}$-almost all curves and $|\nabla f|_{w,\mathcal{T}} \le |\nabla f|_*$ $\mathfrak{m}$-a.e. in $X$.*

*Proof.* By the very definition of $\tilde{\mathsf{Ch}}$ and the chain rule for relaxed gradients it is sufficient to consider the case when $f$ is bounded. We already observed in Remark 5.5 that, for a Borel $\mathsf{d}$-Lipschitz function $f$, the local Lipschitz constant is a $\mathcal{T}$-weak upper gradient. Now, pick a sequence $(f_n)$ of Borel $\mathsf{d}$-Lipschitz functions converging to $f$ in $L^2(X, \mathfrak{m})$ such that $|\nabla f_n|$ converge weakly in $L^2(X, \mathfrak{m})$ to $|\nabla f|_*$, thus in particular weakly in $L^2(\{V \le M\}, \mathfrak{m})$ to $|\nabla f|_*$ for all $M \ge 0$. Then, Theorem 5.12 ensures that $|\nabla f|_*$ is a $\mathcal{T}$-weak upper gradient for $f$. □

We shall also need chain rules for minimal weak upper gradients; the proofs are very analogous to those of relaxed gradients, so we omit a few details.

**Proposition 5.14 (Chain rule for minimal weak upper gradients)** *Let $\mathcal{T}$ be as in Theorem 5.12. If $f : X \to \mathbb{R}$ is Sobolev along $\mathcal{T}$-almost all curves, the following properties hold:*

*(a) for any $\mathscr{L}^1$-negligible Borel set $N \subset \mathbb{R}$ it holds $|\nabla f|_{w,\mathcal{T}} = 0$ $\mathfrak{m}$-a.e. on $f^{-1}(N)$;*

*(b) $|\nabla \phi(f)|_{w,\mathcal{T}} = \phi'(f)|\nabla f|_{w,\mathcal{T}}$, with the convention $0 \cdot \infty = 0$, for any nondecreasing function $\phi$, locally Lipschitz on an interval containing the image of $f$.*

*Proof.* First we prove (b) in the case when $\phi$ is everywhere differentiable. By the same minimality argument in Proposition 4.8(d), it suffices to show that $|\nabla \phi(f)|_{w,\mathcal{T}} \le \phi'(f)|\nabla f|_{w,\mathcal{T}}$. This inequality is a direct consequence of (5.6): indeed, if $f_\gamma$ is the absolutely continuous function equal to $f \circ \gamma$ a.e. in $[0,1]$ and in $\{0,1\}$, we have $\phi \circ f_\gamma = \phi \circ f$ on $\{0,1\}$ and $|(\phi \circ f_\gamma)'| = \phi'(f_\gamma)|f_\gamma'| \le \phi'(f_\gamma)|\nabla f|_{w,\mathcal{T}} \circ \gamma |\dot\gamma|$. Since $f_\gamma = f \circ \gamma$ a.e. in $[0,1]$, by integration we get that $\phi'(f)|\nabla f|_{w,\mathcal{T}}$ is a weak upper gradient.

Having estabilished the chain rule when $\phi$ is differentiable, the proof of (a) follows by the stability of weak gradients, as in the proof of Proposition 4.8(a). Eventually we extend (b), which now makes sense defining arbitrarily $\phi'(f)$ at points where $x$ such that $\phi$ is not differentiable at $f(x)$, by a further approximation, as in Proposition 4.8. □

**Lemma 5.15** *Let $\mathcal{T}$ be the collection of all the test plans concentrated on $\mathrm{AC}^2((0,1);(X,\mathsf{d}))$ with bounded compression on the sublevels of $V$ (i.e. satisfying (5.4)).*
*Let $\mu \in \mathrm{AC}^2((0,T);(\mathscr{P}(X),W_2))$ be an absolutely continuous curve with uniformly bounded densities $f_t = \mathrm{d}\mu_t/\mathrm{d}\mathfrak{m}$. Let $\phi : [0,\infty) \to \mathbb{R}$ be a convex function with $\phi(0) = 0$ and $\phi'$ locally Lipschitz in $(0,\infty)$. We suppose that for a.e. $t \in (0,T)$ $f_t$ is Sobolev along $\mathcal{T}$-almost all curves*

$$H_t^2 := \int_X |\nabla f_t|^2_{w,\mathcal{T}}\,\mathrm{d}\mathfrak{m} < \infty, \quad G_t^2 := \int_{\{f_t>0\}} \Big(\phi''(f_t)|\nabla f_t|_{w,\mathcal{T}}\Big)^2 f_t\,\mathrm{d}\mathfrak{m} < \infty, \tag{5.12}$$

*for a.e. $t \in (0,T)$. Assume in addition that $G$, $H \in L^2(0,T)$ and that $\int_X |\phi(f_0)|\,\mathrm{d}\mathfrak{m} < \infty$. Then $t \mapsto \int_X |\phi(f_t)|\,\mathrm{d}\mathfrak{m}$ is bounded in $[0,T]$,*

$$\Phi_t := \int_X \phi(f_t)\,\mathrm{d}\mathfrak{m} \quad \textit{is absolutely continuous in } [0,T] \textit{ and } \quad \Big|\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t\Big| \le G_t\,|\dot\mu_t| \quad \textit{a.e. in } (0,T). \tag{5.13}$$

*If moreover $\phi'$ is Lipschitz on an interval containing the image of $f_t$, $t \in [0,T]$, then the pointwise estimates hold*

$$\limsup_{s\downarrow t} \frac{\Phi_t - \Phi_s}{s-t} \le G_t \limsup_{s\downarrow t} ⨍_t^s |\dot\mu_r|\,\mathrm{d}r, \quad \liminf_{s\downarrow t} \frac{\Phi_t - \Phi_s}{s-t} \le G_t \liminf_{s\downarrow t} ⨍_t^s |\dot\mu_r|\,\mathrm{d}r. \tag{5.14}$$

*Proof.* It is not restrictive to assume $T = 1$. Let $C$ be a constant satisfying $\mu_t \le C\mathfrak{m}$ for all $t \in [0,1]$ and notice that, by interpolation, $f_t$ are uniformly bounded in all spaces $L^p(X,\mathfrak{m})$. In addition, since $f_s$ weakly converge to $f_t$ as $s \to t$ in the duality with $C_b(X)$, and $C_b(X) \cap L^p(X,\mathfrak{m})$, $1 \le p \le \infty$, is dense in $L^p(X,\mathfrak{m})$ (thanks to the existence of the $\mathsf{d}$-Lipschitz weight function $V$ whose sublevels have finite $\mathfrak{m}$ measure), we obtain that $t \mapsto f_t$ is continuous in the weak topology of $L^q(X,\mathfrak{m})$ (weak* if $q = \infty$), with $q$ dual exponent of $p$. It follows that $\Phi_t$ is lower semicontinuous. Arguing as in [23, 2], see also the work in progress [24] for the case of extended metric spaces, we can find $\boldsymbol{\pi} \in \mathscr{P}(C([0,1];X))$ concentrated in $\mathrm{AC}^2((0,1);(X,\mathsf{d}))$ and satisfying

$$\mu_t = (\mathrm{e}_t)_\sharp \boldsymbol{\pi} \quad \text{for every } t \in [0,1], \qquad |\dot\mu_t|^2 = \int |\dot\gamma_t|^2\,\mathrm{d}\boldsymbol{\pi}(\gamma) \quad \text{for a.e. } t \in (0,1), \tag{5.15}$$

so that $\boldsymbol{\pi} \in \mathcal{T}$. Let us first suppose that $\phi'$ is locally Lipschitz continuous in $[0,\infty)$, so that $\Phi_t$ is everywhere finite. Possibly replacing $\phi(z)$ by $\phi(z)-\phi'(0)z$ we can assume that $\phi$ is nonnegative and nondecreasing.

We pick a point $t$ such that $f_t$ is Sobolev along $\boldsymbol{\pi}$-almost all curves and $H_t<\infty$ and we set $h_t := \phi'(f_t)$, $g_t := |\nabla h_t|_{w,\mathcal{T}} = \phi''(f_t)|\nabla f_t|_{w,\mathcal{T}}$. Then for every $s\in(0,t)$ we have

$$\begin{aligned}\Phi_t - \Phi_s &\le \int_X \phi'(f_t)(f_t-f_s)\,\mathrm{d}\mathfrak{m} = \int \big(h_t(\gamma_t)) - h_t(\gamma_s)\big)\,\mathrm{d}\boldsymbol{\pi}(\gamma) \le \int\int_s^t g_t(\gamma_r)|\dot\gamma_r|\,\mathrm{d}r\mathrm{d}\boldsymbol{\pi}(\gamma)\\ &\le \int_s^t \Big(\int_X |g_t|^2\,f_r\,\mathrm{d}\mathfrak{m}\Big)^{1/2}\Big(\int |\dot\gamma_r|^2\,\mathrm{d}\boldsymbol{\pi}(\gamma)\Big)^{1/2}\mathrm{d}r = \int_s^t \Big(\int_X |g_t|^2\,f_r\,\mathrm{d}\mathfrak{m}\Big)^{1/2}|\dot\mu_r|\,\mathrm{d}r. \quad (5.16)\end{aligned}$$

Since $H\in L^2(0,1)$ we deduce from Lemma 2.7 (with $w=-\Phi$, $L=|\dot\mu|$, $g=C^{1/2}H$ at all points $t$ such that $f_t$ is Sobolev along $\boldsymbol{\pi}$-almost all curves, $g=+\infty$ elsewhere) that $\Phi$ is absolutely continuous.

Writing the inequalities analogous to (5.16) for $s>t$, dividing by $s-t$, and passing to the limit as $s\downarrow t$, thanks to the $w^*$ continuity of $r\mapsto f_r$ in $L^\infty(X,\mathfrak{m})$ we get the bound (5.14) (and thus (5.13) when $t$ is also a differentiability point for $\Phi$ and a Lebesgue point for $|\dot\mu|$).

When $\phi$ is an arbitrary convex function, for $\varepsilon\in(0,1]$ we set

$$\phi_\varepsilon(r) := \begin{cases} r\phi'(\varepsilon) & \text{if } 0\le r\le\varepsilon,\\ \phi(z)-\phi(\varepsilon)+\varepsilon\phi'(\varepsilon) & \text{if } r\ge\varepsilon;\end{cases}$$

it is easy to check that $\phi_\varepsilon$ is convex, with locally Lipschitz derivative in $[0,\infty)$ and that $\phi_\varepsilon\downarrow\phi$ as $\varepsilon\downarrow 0$, since $\varepsilon\mapsto\varepsilon\phi'(\varepsilon)-\phi(\varepsilon)$ is increasing and converges to 0 as $\varepsilon\downarrow 0$. Notice moreover that $(\phi_\varepsilon)''\le\phi''$. Applying the integral form of (5.13) to $\Phi^\varepsilon_t := \int_X \phi_\varepsilon(f_t)\,\mathrm{d}\mathfrak{m}$ we get

$$\big|\Phi^\varepsilon_t - \Phi^\varepsilon_s\big| \le \int_s^t G_r\,|\dot\mu_r|\,\mathrm{d}r \quad \text{for every } 0\le s<t\le 1. \quad (5.17)$$

Since $\Phi^\varepsilon_0\to\Phi_0$, it follows that all the functions $\Phi^\varepsilon_t$ are uniformly bounded. In addition, (5.16) with $\phi=\phi_\varepsilon$ gives that

$$\int_X (\phi_\varepsilon)^-(f_s)\,\mathrm{d}\mathfrak{m} \le \int_X (\phi_\varepsilon)^+(f_s)\,\mathrm{d}\mathfrak{m} + R \le \int_X (\phi_1)^+(f_s)\,\mathrm{d}\mathfrak{m} + R$$

with $R$ uniformly bounded in $s$ and $\varepsilon$ (notice that $t$ can be chosen independently of $s$ and $\varepsilon$). Hence, applying the monotone convergence theorem we obtain the uniform bound on $\|\phi(f_t)\|_{L^1(X,\mathfrak{m})}$ and pass to the limit in (5.17) as $\varepsilon\downarrow 0$, obtaining (5.13). □

**Remark 5.16 (Invariance properties)** If $\mathcal{T}$ is the collection of all test plans concentrated on $\mathrm{AC}^2((0,1);(X,\mathsf{d}))$ with bounded compression on the sublevels of $V$ (according to (5.4)) all the concepts introduced so far (test plans, negligible sets of curves, weak upper gradient and minimal weak upper gradient) are immediately seen to be invariant if one replaces $\mathfrak{m}$ with the finite measure $\tilde{\mathfrak{m}} := \mathrm{e}^{-V^2}\mathfrak{m}$ (recall (4.2)): indeed, any test plan with bounded compression relative to $\tilde{\mathfrak{m}}$ is a test plan with bounded compression relative to $\mathfrak{m}$ and any test plan bounded compression relative to $\mathfrak{m}$ can be monotonically approximated by analogous test plans relative to $\tilde{\mathfrak{m}}$. A similar argument holds for plans satisfying (5.10). ■

**Remark 5.17** As for Cheeger's energy and the relaxed gradient, if no additional assumption on $(X,\tau,\mathsf{d},\mathfrak{m})$ is made, it is well possible that the weak upper gradient is trivial.

This is the case of the second example considered in Remark 4.12, where it is easy to check that the class of absolutely continuous curves contains just the constants, so that $|\nabla f|_{w,\mathcal{T}}\equiv 0$ for every $f\in L^2([0,1];\mathfrak{m})$ independently from the choice of $\mathcal{T}$. In order to exclude such situations, we are going to make additional assumptions on $(X,\tau,\mathsf{d},\mathfrak{m})$ in the next sections, as the lower semicontinuity of $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(f\mathfrak{m})$: this ensures, as we will see in Theorem 7.6, its agreement with $8\mathsf{Ch}_*(\sqrt{f})$. Since $\mathrm{Ent}_{\mathfrak{m}}$ is not trivial, the same is true for $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ and for $\mathsf{Ch}_*$. In turn, we will see that lower semicontinuity of $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2$ is implied by $CD(K,\infty)$. ■

# 6 Identification between relaxed gradient and weak upper gradient

The key statement that will enable us to prove the main identification result of this section is provided in the following lemma. It corresponds precisely to [17, Proposition 3.7]: the main improvement here is the use of the refined analysis of the Hamilton-Jacobi equation semigroup we did in Section 3, together with the use of relaxed gradients, in place of the standard Sobolev spaces in Alexandrov spaces. In this way we can also avoid any lower curvature bound on $(X,\mathsf{d})$ and we do not even require that $(X,\mathsf{d})$ is a length space.

**Lemma 6.1 (A key estimate for the Wasserstein velocity)** *Let $(X,\tau,\mathsf{d},\mathfrak{m})$ be a Polish extended measure space satisfying*

$$\mathfrak{m}\big(\{x\in X:\mathsf{d}(x,K)\le r\}\big)<\infty\quad \textit{for every compact } K\subset X \textit{ and } r>0. \tag{6.1}$$

*Let $(f_t)$ be the gradient flow of $\mathsf{Ch}_*$ in $L^2(X,\mathfrak{m})$ starting from a nonnegative $f_0\in L^2(X,\mathfrak{m})$ and let us assume that*

$$\int_X f_t\,\mathrm{d}\mathfrak{m}=1,\qquad \int_0^t\int_{\{f_s>0\}}\frac{|\nabla f_s|_*^2}{f_s}\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s<\infty\quad \textit{for every } t\ge 0. \tag{6.2}$$

*Then, setting $\mu_t:=f_t\mathfrak{m}\in\mathscr{P}(X)$, the curve $t\mapsto\mu_t:=f_t\mathfrak{m}$ is locally absolutely continuous from $(0,\infty)$ to $(\mathscr{P}_{[\mu_0]}(X),W_2)$ and its metric speed $|\dot\mu_t|$ satisfies*

$$|\dot\mu_t|^2\le\int_{\{f_t>0\}}\frac{|\nabla f_t|_*^2}{f_t}\,\mathrm{d}\mathfrak{m}\qquad \textit{for a.e. } t\in(0,\infty). \tag{6.3}$$

*Proof.* We start from the duality formula (2.21): it is easy to check that it can be written as

$$\frac{W_2^2(\mu,\nu)}{2}=\sup_\varphi\int_X Q_1\varphi\,\mathrm{d}\nu-\int_X\varphi\,\mathrm{d}\mu \tag{6.4}$$

where the supremum runs in $C_b(X)$. Now, if $\mu\ll\mathfrak{m}$, we may equivalently consider $\tau$-lower semicontinuous functions $\varphi$ of the form (3.21); indeed, given $\varphi\in C_b(X)$, considering a sequence of compact sets $K_n\subset X$ whose union is of full $\mathfrak{m}$-measure and setting

$$\varphi_n(x):=\begin{cases}\varphi(x) & \text{if } x\in K_n;\\ \sup\varphi & \text{if } x\in X\setminus K_n\end{cases}$$

we obtain $\varphi_n \downarrow \varphi$ $\mathfrak{m}$-a.e. and $Q_1\varphi_n \geq Q_1\varphi$.

Moreover, since (6.4) is invariant by adding constants to $\varphi$, we can always assume that $M=0$ in (3.21) and that $\varphi$ vanishes outside a compact set $K$.

Now, if $\varphi$ is of the form (3.21) with $M=0$, we notice that for all $\varepsilon>0$ the map $Q_\varepsilon\varphi$ is $\mathsf{d}$-Lipschitz, bounded and lower semicontinuous (the latter property follows by Proposition 3.8), and

$$Q_\varepsilon\varphi(x)=0 \quad \text{if} \quad \mathsf{d}(x,K)\geq 2\sqrt{\max_K(-\phi)} \text{ and } \varepsilon\leq 2, \tag{6.5}$$

so that, by (6.1), $(Q_\varepsilon(\varphi))_{\varepsilon\in[0,2]}$ is uniformly bounded in each $L^p(X,\mathfrak{m})$.

In addition, since $\varphi$ is $\tau$-lower semicontinuous, by Remark 3.7 we have $Q_\varepsilon\varphi\uparrow\varphi$ as $\varepsilon\downarrow 0$. Hence, by approximating $\varphi$ with $Q_\varepsilon\varphi$ and taking also the convergence of $Q_{1+\varepsilon}\varphi\leq Q_1(Q_\varepsilon\varphi)$ to $Q_1\varphi$ as $\varepsilon\downarrow 0$ into account (recall the continuity property (3.5) and that $t_*\equiv\infty$ in this case), we see that the supremum in (6.4) can be taken over the set of bounded $\varphi$'s with $Q_t\varphi$ $\tau$-lower semicontinuous for all $t>0$ and uniformly $\mathsf{d}$-Lipschitz.

Fix such a function $\varphi$ and observe that, thanks to the pointwise estimates (3.12) and (3.20), the map $t\mapsto Q_t\varphi$ is Lipschitz with values in $L^\infty(X,\mathfrak{m})$, and a fortiori in $L^2(X,\mathfrak{m})$ by (6.5). In addition, the "functional" derivative (i.e. the strong limit in $L^2$ of the difference quotients) $\partial_t Q_t\varphi$ of this $L^2(X,\mathfrak{m})$-valued map is easily seen to coincide, for a.e. $t$, with the map $\frac{d^+}{dt}Q_t\varphi(x)$. Recall also that, still thanks to Proposition 3.8, the latter map is Borel and $|\nabla Q_t\varphi|$ is $\mathscr{B}^*(X\times(0,\infty))$-measurable.

Fix also $0<t<s$, set $\ell=(s-t)$ and recall that since $(f_t)$ is the gradient flow of $\mathsf{Ch}_*$ in $L^2(X,\mathfrak{m})$, the map $[0,\ell]\ni r\mapsto f_{t+r}$ is Lipschitz with values in $L^2(X,\mathfrak{m})$.

Now, for $a$, $b:[0,\ell]\to L^2(X,\mathfrak{m})$ Lipschitz, it is well known that $t\mapsto\int_X a_tb_t\,\mathrm{d}\mathfrak{m}$ is Lipschitz in $[0,\ell]$ and that $\big(\int_X a_tb_t\,\mathrm{d}\mathfrak{m}\big)'=\int_X b_t\partial_t a_t\,\mathrm{d}\mathfrak{m}+\int_X a_t\partial_t b_t\,\mathrm{d}\mathfrak{m}$ for a.e. $t\in[0,\ell]$. Therefore we get

$$\frac{\mathrm{d}}{\mathrm{d}r}\int_X Q_{r/\ell}\varphi\, f_{t+r}\,\mathrm{d}\mathfrak{m}=\int_X\frac{1}{\ell}\xi_{t/\ell}\, f_{t+r}+Q_{r/\ell}\varphi\,\Delta_{\mathsf{d},\mathfrak{m}}f_{t+r}\,\mathrm{d}\mathfrak{m} \qquad \text{for a.e. } r>0,$$

where $\xi_s(x):=\frac{\mathrm{d}^+}{\mathrm{d}t}Q_t\varphi\big|_{t=s}(x)$; we have then:

$$\begin{aligned}
\int_X Q_1\varphi\,\mathrm{d}\mu_s-\int_X\varphi\,\mathrm{d}\mu_t&=\int_X Q_1\varphi f_{t+\ell}\,\mathrm{d}\mathfrak{m}-\int_X\varphi f_t\,\mathrm{d}\mathfrak{m}\\
&=\int_0^\ell\int_X\frac{1}{\ell}\xi_{r/\ell}\, f_{t+r}+Q_{r/\ell}\varphi\,\Delta_{\mathsf{d},\mathfrak{m}}f_{t+r}\,\mathrm{d}\mathfrak{m}\mathrm{d}r\\
&=\int_X\int_0^\ell\frac{1}{\ell}\xi_{r/\ell}\, f_{t+r}+Q_{r/\ell}\varphi\,\Delta_{\mathsf{d},\mathfrak{m}}f_{t+r}\,\mathrm{d}r\mathrm{d}\mathfrak{m}\\
&\leq\int_X\int_0^\ell -\frac{|\nabla Q_{r/\ell}\varphi|^2}{2\ell}+Q_{r/\ell}\varphi\,\Delta_{\mathsf{d},\mathfrak{m}}f_{t+r}\,\mathrm{d}r\,\mathrm{d}\mathfrak{m}.
\end{aligned}\tag{6.6}$$

In the last two steps we used first Fubini's theorem and then Theorem 3.5. Observe that by inequalities (4.25) and (4.8) we have (using also that $|\nabla f_s|_*=0$ $\mathfrak{m}$-a.e. on $\{f_s=0\}$)

$$\begin{aligned}
\int_X Q_{r/\ell}\varphi\,\Delta_{\mathsf{d},\mathfrak{m}}f_{t+r}\,\mathrm{d}\mathfrak{m}&\leq\int_X|\nabla Q_{r/\ell}\varphi|_*\,|\nabla f_{t+r}|_*\,\mathrm{d}\mathfrak{m}\\
&\leq\frac{1}{2\ell}\int_X|\nabla Q_{r/\ell}\varphi|^2 f_{t+r}\,\mathrm{d}\mathfrak{m}+\frac{\ell}{2}\int_{\{f_{t+r}>0\}}\frac{|\nabla f_{t+r}|_*^2}{f_{t+r}}\,\mathrm{d}\mathfrak{m}.
\end{aligned}$$

Plugging this inequality in (6.6) and using once more Fubini's theorem we obtain

$$\int_X Q_1\varphi\,\mathrm{d}\mu_s - \int_X \varphi\,\mathrm{d}\mu_t \le \frac{\ell}{2}\int_0^\ell \int_{\{f_{t+r}>0\}} \frac{|\nabla f_{t+r}|_*^2}{f_{t+r}}\,\mathrm{d}r\mathrm{d}\mathfrak{m}.$$

This latter bound does not depend on $\varphi$, so from (6.4) we deduce

$$W_2^2(\mu_t,\mu_s) \le \ell \int_0^\ell \int_{\{f_{t+r}>0\}} \frac{|\nabla f_{t+r}|_*^2}{f_{t+r}}\mathrm{d}r\mathrm{d}\mathfrak{m}, \qquad \ell = s-t.$$

By (6.2) we immediately get that $\mu_t \in \mathscr{P}_{[\mu_0]}(X)$ and (6.3) holds. □

In the next two results we will consider the class of (measurable) functions

$$f : X \to \mathbb{R} \quad \text{such that } f_N := \min\{N, \max\{f,-N\}\} \in L^2(X,\mathfrak{m}) \quad \text{for every } N > 0. \tag{6.7}$$

**Theorem 6.2 (Relaxed and weak upper gradients coincide)** *Let $(X,\tau,\mathsf{d},\mathfrak{m})$ be a Polish extended measure space with $\mathfrak{m}$ satisfying (4.2) and let $\mathcal{T}$ be the collection of all test plans concentrated on $\mathrm{AC}^2((0,1);(X,\mathsf{d}))$ with bounded compression on the sublevels of $V$ (5.4). A measurable function $f : X \to \mathbb{R}$ satisfying (6.7) has relaxed gradient $|\nabla f|_*$ (according to (4.14)) in $L^2(X,\mathfrak{m})$ iff $f$ is Sobolev on $\mathcal{T}$-almost all curves and $|\nabla f|_{w,\mathcal{T}} \in L^2(X,\mathfrak{m})$. In this case*

$$|\nabla f|_* = |\nabla f|_{w,\mathcal{T}} \quad \mathfrak{m}\text{-a.e. in } X. \tag{6.8}$$

*Proof.* Taking into account Remark 5.16 and Lemma 4.11 It is not restrictive to assume that $\mathfrak{m} \in \mathscr{P}(X)$, so that we can choose $V \equiv 1$. Moreover, we can assume that $0 < M^{-1} \le f \le M < \infty$ $\mathfrak{m}$-almost everywhere in $X$ with $\int_X f^2\,\mathrm{d}\mathfrak{m} = 1$. By Corollary 5.13 we have to prove that if $f$ is Sobolev on $\mathcal{T}$-almost all curves with $|\nabla f|_{w,\mathcal{T}} \in L^2(X,\mathfrak{m})$ then

$$\mathsf{Ch}_*(f) \le \frac{1}{2}\int_X |\nabla f|_{w,\mathcal{T}}^2\,\mathrm{d}\mathfrak{m}. \tag{6.9}$$

We consider the gradient flow $(h_t)$ of the Cheeger's energy with initial datum $h := f^2$, setting $\mu_t = h_t\mathfrak{m}$, and we apply Lemma 6.1. If $g = h^{-1}|\nabla h|_{w,\mathcal{T}}$, we easily get arguing as in (5.16) and using inequality (6.3)

$$\begin{aligned}\int_X \big(h\log h - h_t\log h_t\big)\,\mathrm{d}\mathfrak{m} &\le \Big(\int_0^t\int_X g^2 h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s\Big)^{1/2}\Big(\int_0^t |\dot\mu_s|^2\,\mathrm{d}s\Big)^{1/2}\\ &\le \frac{1}{2}\int_0^t\int_X g^2 h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \frac{1}{2}\int_0^t |\dot\mu_s|^2\,\mathrm{d}s \le \frac{1}{2}\int_0^t\int_X g^2 h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \frac{1}{2}\int_0^t\int_{\{h_s>0\}} \frac{|\nabla h_s|_*^2}{h_s}\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s.\end{aligned}$$

Recalling the entropy dissipation formula (4.54) we obtain

$$\int_0^t\int_{\{h_s>0\}} \frac{|\nabla h_s|_*^2}{h_s}\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s \le \int_0^t\int_X g^2 h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s.$$

Now, (4.16) and the identity $g = 2f^{-1}|\nabla f|_{w,\mathcal{T}}$ give $\int_0^t \mathsf{Ch}_*(\sqrt{h_s})\,\mathrm{d}s \le \int_0^t\int_X |\nabla f|_{w,\mathcal{T}}^2 f^{-2}h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s$, so that dividing by $t$ and passing to the limit as $t \downarrow 0$ we get (6.9), since $\sqrt{h_s}$ are equibounded and converge strongly to $f$ in $L^2(X,\mathfrak{m})$ as $s \downarrow 0$. □

**Corollary 6.3** *Let $\mathcal{T}_1$ be the collection of all test plans concentrated on* $\mathrm{AC}^2((0,1);(X,\mathsf{d}))$ *with bounded compression on the sublevels of $V$ as in (5.4), and let $\mathcal{T}_2$ be the collection of all test plans concentrated on* $\mathrm{AC}^2((0,1);(X,\mathsf{d}))$ *satisfying (5.10). Let us suppose that a measurable function $f : X \to \mathbb{R}$ satisfying (6.7) is Sobolev on $\mathcal{T}_1$-almost all curves with $|\nabla f|_{w,\mathcal{T}_1} \in L^2(X,\mathfrak{m})$. Then $f$ is Sobolev on $\mathcal{T}_2$-almost all curves and*

$$|\nabla f|_{w,\mathcal{T}_1} = |\nabla f|_{w,\mathcal{T}_2} = |\nabla f|_* \quad \mathfrak{m}\text{-a.e. in } X. \tag{6.10}$$

*Proof.* Applying Theorem 6.2 and Corollary 5.13 we prove that $f$ is Sobolev on $\mathcal{T}_2$-almost all curves with $|\nabla f|_{w,\mathcal{T}_1} \geq |\nabla f|_{w,\mathcal{T}_2}$. The converse inequality follows by Remark 5.11. □

**Remark 6.4 (One-sided relaxed gradients)** Theorem 6.2 shows that the one-sided Cheeger's functionals $\mathsf{Ch}_*^{\pm}(f)$ (and the corresponding relaxed gradients $|\nabla^{\pm} f|$) introduced in Remark 4.23 coincide with $\mathsf{Ch}_*(f)$ (resp. with $|\nabla f|_*$). In fact, we already observed that $\mathsf{Ch}_*(f) \geq \mathsf{Ch}_*^{\pm}(f)$. On the other hand, since $|\nabla^{\pm} f|$ are weak upper gradients for Borel $\mathsf{d}$-Lipschitz functions, arguing as in Corollary 5.13 we get $|\nabla f|_{w,\mathcal{T}} \leq |\nabla^{\pm} f|_*$, if $f \in D(\mathsf{Ch}_*^{\pm})$ and $\mathcal{T}$ is the class of all test plans concentrated on $\mathrm{AC}^2((0,1);(X,\mathsf{d}))$ with boundedcompression on the sublevels of $V$. Corollary 6.3 yields

$$\mathsf{Ch}_*(f) = \mathsf{Ch}_*^{\pm}(f), \qquad |\nabla f|_* = |\nabla^{\pm} f|_* \quad \mathfrak{m}\text{-a.e. in } X.$$

# 7 Relative entropy, Wasserstein slope, and Fisher information

In this section we assume that $(X,\tau,\mathsf{d})$ is a Polish extended space equipped with a $\sigma$-finite Borel reference measure $\mathfrak{m}$ such that $\tilde{\mathfrak{m}} := e^{-V^2}\mathfrak{m}$ has total mass less than 1 for some Borel and $\mathsf{d}$-Lipschitz $V : X \to [0,\infty)$. We shall work in the subspace

$$\mathscr{P}_V(X) := \left\{ \mu \in \mathscr{P}(X) : \int_X V^2 \,\mathrm{d}\mu < \infty \right\}. \tag{7.1}$$

We say that $(\mu_n) \subset \mathscr{P}_V(X)$ *weakly converges with moments* to $\mu \in \mathscr{P}_V(X)$ if $\mu_n \to \mu$ weakly in $\mathscr{P}(X)$ and $\int_X V^2 \,\mathrm{d}\mu_n \to \int_X V^2 \,\mathrm{d}\mu$. Analogously we define strong convergence with moments in $\mathscr{P}(X)$, by requiring that $|\mu_n - \mu|(X) \to 0$, instead of the weak convergence.

Since for every $\mu \in \mathscr{P}_V(X)$, $\nu \in \mathscr{P}(X)$ with $W_2(\mu,\nu) < \infty$ we have $\nu \in \mathscr{P}_V(X)$ and

$$\Big(\int_X V^2 \,\mathrm{d}\nu\Big)^{1/2} \leq \mathrm{Lip}(V) W_2(\nu,\mu) + \Big(\int_X V^2 \,\mathrm{d}\mu\Big)^{1/2}, \tag{7.2}$$

we obtain that weak convergence with moments is implied by $W_2$ convergence. When the topology $\tau$ is induced by the distance $\mathsf{d}$ and $V(x) := A\,\mathsf{d}(x,x_0)$ for some $A > 0$ and $x_0 \in X$, then weak convergence with moments is in fact equivalent to $W_2$ convergence. When $\mathfrak{m}(X) < \infty$ we may take $V$ equal to a constant, so that $\mathscr{P}(X) = \mathscr{P}_V(X)$ and weak convergence with moments reduces to weak convergence.

## 7.1 Relative entropy

**Definition 7.1 (Relative entropy)** *The relative entropy functional* $\mathrm{Ent}_{\mathfrak{m}} : \mathscr{P}(X) \to (-\infty,+\infty]$ *is defined as*

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

Notice that, according to our definition, $\mu \in D(\mathrm{Ent}_{\mathfrak{m}})$ implies $\int_X V^2 \,\mathrm{d}\mu < \infty$, and that $D(\mathrm{Ent}_{\mathfrak{m}})$ is convex. Strictly speaking the notation $\mathrm{Ent}_{\mathfrak{m}}$ is a slight abuse, since the functional depends also on the choice of $V$, which is not canonically induced by $\mathfrak{m}$ (not even in Euclidean spaces endowed with the Lebesgue measure). It is tacitly understood that we take $V$ equal to a constant whenever $\mathfrak{m}(X) < \infty$ and in this case $\mathrm{Ent}_{\mathfrak{m}}$ is independent of the chosen constant.

When $\mathfrak{m} \in \mathscr{P}(X)$ the functional $\mathrm{Ent}_{\mathfrak{m}}$ is sequentially lower semicontinuous w.r.t. weak convergence in $\mathscr{P}(X)$. In addition, it is nonnegative, thanks to Jensen's inequality. More generally, if $\mathfrak{m}$ is a finite measure and $\bar{\mathfrak{m}} := \mathfrak{m}(X)^{-1}\mathfrak{m}$,

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) = \mathrm{Ent}_{\bar{\mathfrak{m}}}(\mu) - \log(\mathfrak{m}(X)) \geq -\log(\mathfrak{m}(X)) \quad \text{for every } \mu \in \mathscr{P}(X), \tag{7.3}$$

and we have the general inequality (see for instance [4, Lemma 9.4.5])

$$\mathrm{Ent}_{\pi_\sharp \mathfrak{m}}(\pi_\sharp \mu) \leq \mathrm{Ent}_{\mathfrak{m}}(\mu) \quad \text{for every } \mu \in \mathscr{P}(X) \quad \text{and } \pi : X \to Y \text{ Borel map}, \tag{7.4}$$

which turns out to be an equality if $\pi$ is injective. When $\mathfrak{m}(X) = \infty$, since the density $\tilde{\rho}$ of $\mu$ w.r.t. $\tilde{\mathfrak{m}}$ equals $\rho\, \mathrm{e}^{V^2}$ we obtain that the negative part of $\rho \log \rho$ is $L^1(\mathfrak{m})$-integrable for $\mu \in \mathscr{P}_V(X)$, so that Definition 7.1 is well posed and $\mathrm{Ent}_{\mathfrak{m}}$ does not attain the value $-\infty$. We also obtain the useful formula

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) = \mathrm{Ent}_{\tilde{\mathfrak{m}}}(\mu) - \int_X V^2 \,\mathrm{d}\mu \qquad \forall \mu \in \mathscr{P}_V(X). \tag{7.5}$$

The same formula shows $\mathrm{Ent}_{\mathfrak{m}}$ is sequentially lower semicontinuous in $\mathscr{P}_V(X)$ w.r.t. convergence with moments, i.e.

$$\mu_n \rightharpoonup \mu \text{ in } \mathscr{P}(X), \ \int_X V^2 \,\mathrm{d}\mu_n \to \int_X V^2 \,\mathrm{d}\mu < \infty \quad \Longrightarrow \quad \liminf_{n\to\infty} \mathrm{Ent}_{\mathfrak{m}}(\mu_n) \geq \mathrm{Ent}_{\mathfrak{m}}(\mu). \tag{7.6}$$

From (7.2) we also get

$$\mu \in \mathscr{P}_V(X), \quad W_2(\mu_n, \mu) \to 0 \quad \Longrightarrow \quad \liminf_{n\to\infty} \mathrm{Ent}_{\mathfrak{m}}(\mu_n) \geq \mathrm{Ent}_{\mathfrak{m}}(\mu). \tag{7.7}$$

The following lemma for the change of reference measure in the entropy, related to (7.5), will be useful.

**Lemma 7.2 (Change of reference measure in the entropy)** *Let $\nu \in \mathscr{P}(X)$ and the positive finite measure $\mathfrak{n}$ be satisfying $\mathrm{Ent}_{\mathfrak{n}}(\nu) < \infty$. If $\nu = g\mathfrak{m}$ for some $\sigma$-finite Borel measure $\mathfrak{m}$, then $g \log g \in L^1(X, \mathfrak{m})$ if and only if $\log(\mathrm{d}\mathfrak{n}/\mathrm{d}\mathfrak{m}) \in L^1(X, \nu)$ and*

$$\mathrm{Ent}_{\mathfrak{m}}(\nu) = \mathrm{Ent}_{\mathfrak{n}}(\nu) + \int_X \log\left(\frac{\mathrm{d}\mathfrak{n}}{\mathrm{d}\mathfrak{m}}\right) \mathrm{d}\nu. \tag{7.8}$$

*Proof.* Write $\nu = f\mathfrak{n}$ and let $\mathfrak{n} = h\mathfrak{m} + \mathfrak{n}^s$ be the Radon-Nikodým decomposition of $\mathfrak{n}$ w.r.t. $\mathfrak{m}$. Since $g\mathfrak{m} = \nu = fh\mathfrak{m} + f\mathfrak{n}^s$ we obtain that $g = fh$ $\mathfrak{m}$-a.e. in $X$ and $f = 0$ $\mathfrak{n}^s$-a.e. in $X$. Since $f \log f \in L^1(X, \mathfrak{n})$ we obtain that $\chi_{\{h>0\}} g \log(g/h)$ belongs to $L^1(X, \mathfrak{m})$, so that (taking into account that $\{g > 0\} \subset \{h > 0\}$ up to $\mathfrak{m}$-negligible sets) $g \log g \in L^1(X, \mathfrak{m})$ if and only if $g \log h \in L^1(X, \mathfrak{m})$. The latter property is equivalent to $\log h \in L^1(X, \nu)$. $\square$

**Remark 7.3 (Tightness of sublevels of** $\mathrm{Ent}_{\mathfrak{m}}(\mu)$ **and setwise convergence)** We remark that the sublevels of the relative entropy functional are tight if $\mathfrak{m}(X) < \infty$. Indeed, by Ulam's theorem $\mathfrak{m}$ is tight. Then, using first the inequality $z\log(z) \geq -e^{-1}$ and then Jensen's inequality, for $\mu = \rho\mathfrak{m}$ we get

$$\frac{\mathfrak{m}(X)}{e} + C \geq \frac{\mathfrak{m}(X \setminus E)}{e} + \mathrm{Ent}_{\mathfrak{m}}(\mu) \geq \int_E \rho\log\rho\,\mathrm{d}\mathfrak{m} \geq \mu(E)\log\left(\frac{\mu(E)}{\mathfrak{m}(E)}\right) \tag{7.9}$$

whenever $E \in \mathscr{B}(X)$ and $\mathrm{Ent}_{\mathfrak{m}}(\mu) \leq C$. This shows that $\mu(E) \to 0$ as $\mathfrak{m}(E) \to 0$ uniformly in the set $\{\mathrm{Ent}_{\mathfrak{m}} \leq C\}$.

In general, when $\int_X \mathrm{e}^{-V^2}\,\mathrm{d}\mathfrak{m} \leq 1$, we see that (7.5) yields

$$\Big\{\mu \in \mathscr{P}(X) : \int_X V^2\,\mathrm{d}\mu + \mathrm{Ent}_{\mathfrak{m}}(\mu) \leq C\Big\} \quad \text{is tight in } \mathscr{P}(X) \text{ for every } C \in \mathbb{R}. \tag{7.10}$$

Moreover, if a sequence $(\mu_n)$ belongs to a sublevel (7.10) and weakly converges to $\mu$, then the sequence of the corresponding densities $\rho_n = \frac{\mathrm{d}\mu_n}{\mathrm{d}\mathfrak{m}}$ converges to $\rho = \frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}$ weakly in $L^1(X,\mathfrak{m})$: since $L^\infty(X,\tilde{\mathfrak{m}}) = L^\infty(X,\mathfrak{m})$, it is sufficient to recall (7.3) and to apply de la Vallée Puissen's criterion for uniform integrability [7, §4.5.10] to the densities of $\mu_n$ w.r.t. the finite measure $\tilde{\mathfrak{m}} = \mathrm{e}^{-V^2}\mathfrak{m}$. In particular the sequence $(\mu_n)$ setwise converges to $\mu$, i.e. $\mu_n(B) \to \mu(B)$ for every $B \in \mathscr{B}(X)$. ■

## 7.2 Entropy dissipation, slope and Fisher information

In this subsection we collect some general properties of the relative entropy, its Wasserstein slope and the Fisher information functional defined via the relaxed gradient that we introduced in the previous section.

We will always assume that $\mathfrak{m}$ satisfies condition (4.2), so that Theorem 4.20 will be applicable.

**Theorem 7.4** *Let* $\mu = \rho\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}})$ *with* $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu) < \infty$*. Then* $\sqrt{\rho} \in D(\mathsf{Ch}_*)$ *and*

$$4\int_X |\nabla\sqrt{\rho}|_*^2\,\mathrm{d}\mathfrak{m} \leq |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\mu). \tag{7.11}$$

*Proof.* Let us first assume that $\rho \in L^2(X,\mathfrak{m})$ and let $(\rho_t)$ be the gradient flow of the Cheeger's functional starting from $\rho$; we set $\mu_t := \rho_t\mathfrak{m}$ and recall the definition 4.9 of Fisher information functional $\mathsf{F}$. Applying Proposition 4.22 and Lemma 6.1 we get

$$\begin{aligned}\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\mu_t) &\geq \frac{1}{2}\int_0^t \mathsf{F}(\rho_s)\,\mathrm{d}s + \frac{1}{2}\int_0^t |\dot\mu_s|^2\,\mathrm{d}s \qquad (7.12)\\ &\geq \frac{1}{2}\Big(\frac{1}{\sqrt{t}}\int_0^t \sqrt{\mathsf{F}(\rho_s)}\,\mathrm{d}s\Big)^2 + \frac{1}{2}\Big(\frac{1}{\sqrt{t}}\int_0^t |\dot\mu_s|\,\mathrm{d}s\Big)^2 \geq \frac{1}{t}\Big(\int_0^t \sqrt{\mathsf{F}(\rho_s)}\,\mathrm{d}s\Big)W_2(\mu,\mu_t).\end{aligned}$$

Dividing by $W_2(\mu,\mu_t)$ and passing to the limit as $t \downarrow 0$ we get (7.11), since the lower semicontinuity of Cheeger's functional yields

$$\sqrt{\mathsf{F}(\rho)} \leq \liminf_{t\downarrow 0}\frac{1}{t}\int_0^t \sqrt{\mathsf{F}(\rho_s)}\,\mathrm{d}s.$$

In the general case when only the integrability conditions $\int_X \rho\log\rho\,\mathrm{d}\mathfrak{m}<\infty$ and $\int_X V^2\rho\,\mathrm{d}\mathfrak{m}<\infty$ are available, we can still prove (7.12) by approximation. We set $\rho^n:=z_n\min\{\rho,n\}$, with $z_n\uparrow 1$ normalizing constants, and denote by $\rho^n_t$ the gradient flows of Cheeger's energy starting from $\rho^n$. Since Proposition 4.16(b) provides the monotonicity property $\rho^n_t\le\rho^m_t$ $\mathfrak{m}$-a.e. in $X$ for $n\le m$, we can define $\rho_t:=\sup_n\rho^n_t$. Since $\int_X\rho^n_t\,\mathrm{d}\mathfrak{m}=z_n$ it is immediate to check that $\rho_t\mathfrak{m}\in\mathscr{P}(X)$ and a simple monotonicity argument based on the apriori estimate (4.46) guaranteed by Theorem 4.20 also gives that $\mu_t:=\rho_t\mathfrak{m}\in\mathscr{P}_V(X)$ and that $z_n^{-1}\rho^n_t\mathfrak{m}$ converge with moments to $\mu_t$. It is then easy to pass to the limit in (7.12), using the sequential lower semicontinuity of entropy with respect to convergence with moments, to get

$$\mathrm{Ent}_{\mathfrak{m}}(\mu)-\mathrm{Ent}_{\mathfrak{m}}(\mu_t)\ge\frac{1}{t}\Big(\int_0^t\sqrt{\mathsf{F}(\rho_s)}\,\mathrm{d}s\Big)W_2(\mu,\mu_t)$$

and then conclude as before. $\square$

**Theorem 7.5** *Let $\mu=\rho\mathfrak{m}\in D(\mathrm{Ent}_{\mathfrak{m}})$. Assume that $\rho=\max\{\rho_0,ce^{-2V^2}\}$, where $c>0$ and $\rho_0$ is a $\mathsf{d}$-Lipschitz and bounded map identically 0 for $V$ sufficiently large.*
*Then*

$$|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\mu)\le\int_X\frac{|\nabla^-\rho|^2}{\rho}\,\mathrm{d}\mathfrak{m}=4\int_X|\nabla^-\sqrt{\rho}|^2\,\mathrm{d}\mathfrak{m}. \tag{7.13}$$

*Proof.* We set $L=\mathrm{Lip}(V)$, $M:=\sup\rho_0$ and choose $C\ge 0$ in such a way that $\rho_0=0$ on $\{2V^2>C\}$. Possibly multiplying $\rho$ and $\mathfrak{m}$ by constants we assume $\mathrm{Lip}(\rho_0)=1$. Let us introduce the non negative $\mathscr{B}^*(X\times X)$-measurable function

$$L(x,y):=\begin{cases}\dfrac{\big(\log\rho(x)-\log\rho(y)\big)^+}{\mathsf{d}(x,y)} & \text{if } x\ne y,\\[2ex] \dfrac{|\nabla^-\rho(x)|}{\rho(x)} & \text{if } x=y,\end{cases} \tag{7.14}$$

and notice that for every $x\in X$, the map $y\mapsto L(x,y)$ is $\mathsf{d}$-upper semicontinuous.

We claim that for some constants $C'$, $C''$ depending only on $M$, $c$ and $C$ it holds

$$L(x,y)\le C'+C''\big(\mathsf{d}(x,y)+V(x)\big),\qquad \forall x,\,y\in X. \tag{7.15}$$

To prove this, let $A:=\{\rho_0>ce^{-2V^2}\}$ and notice that $\log\rho\ge\log c-2V^2$ gives

$$\big(\log\rho(x)-\log\rho(y)\big)^+\le\begin{cases}|\log(\rho_0(x))-\log(\rho_0(y))|, & \text{if } x,\,y\in A,\\ 2|V^2(x)-V^2(y)|, & \text{if } x\notin A,\\ \big(\log(\rho_0(x))+2V^2(y)-\log c\big)^+, & \text{if } x\in A,\ y\notin A.\end{cases} \tag{7.16}$$

Since $2V^2\le C$ on $A$, the function $\rho_0$ is $\mathsf{d}$-Lipschitz and bounded from below by $c\,\mathrm{e}^{-C}$ on $A$, so that

$$|\log(\rho_0(x))-\log(\rho_0(y))|\le\frac{\mathrm{e}^C}{c}|\rho_0(x)-\rho_0(y)|\le\frac{\mathrm{e}^C}{c}\mathsf{d}(x,y)\qquad\forall x,\,y\in A. \tag{7.17}$$

Also, for all $x$, $y\in X$ it holds

$$|V^2(x)-V^2(y)|=|V(x)-V(y)||V(x)+V(y)|\le L\mathsf{d}(x,y)\big(L\mathsf{d}(x,y)+2V(x)\big). \tag{7.18}$$

Finally, let us consider the case $x \in A$, $y \notin A$; since $2V^2(y) - \log c \leq -\log \rho_0(y)$ and $\rho_0(y) \geq c\,\mathrm{e}^{-C}/2$ if $\mathsf{d}(x,y) \leq \bar a := c\,\mathrm{e}^{-C}/2$, we get

$$\log(\rho_0(x)) + 2V^2(y) - \log c \leq \log(\rho_0(x)) - \log(\rho_0(y)) \leq \frac{2\mathrm{e}^C}{c}\mathsf{d}(x,y) \tag{7.19}$$

for $\mathsf{d}(x,y) \leq \bar a$. If, instead, $\mathsf{d}(x,y) > \bar a$ we use the fact that $\rho$ is bounded from above, the bound $2V^2(x) \leq C$ for $x \in A$, and (7.18) to get

$$\begin{aligned} \log(\rho_0(x)) + 2V^2(y) - \log c &= \log(\rho_0(x)) + 2V^2(x) - \log c + 2\big(V^2(y) - V^2(x)\big) \\ &\leq \frac{\mathsf{d}(x,y)}{\bar a}\big(\log(M/c) + C\big) + 2L\mathsf{d}(x,y)\big(L\mathsf{d}(x,y) + 2V(x)\big). \end{aligned} \tag{7.20}$$

Inequalities (7.16), (7.17), (7.18), (7.19), (7.20) give the claim (7.15).

Let us now consider a sequence $(\rho_n\mathfrak{m})$, converging to $\mu$ in $\mathscr{P}_2(X)$ and such that

$$|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu) = \lim_{n\to\infty} \frac{\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\mu_n)}{W_2(\mu,\mu_n)}.$$

From the convexity of the function $r \mapsto r\log r$ we have

$$\begin{aligned} \mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\mu_n) &= \int_X \Big(\rho\log\rho - \rho_n\log\rho_n\Big)\,\mathrm{d}\mathfrak{m} \leq \int_X \log\rho\Big(\rho - \rho_n\Big)\,\mathrm{d}\mathfrak{m} \\ &= \int_X \log\rho\,\mathrm{d}\mu - \int_X \log\rho\,\mathrm{d}\mu_n = \int_{X\times X}\Big(\log\rho(x) - \log\rho(y)\Big)\,\mathrm{d}\boldsymbol{\gamma}_n \\ &\leq \int_{X\times X} L(x,y)\mathsf{d}(x,y)\,\mathrm{d}\boldsymbol{\gamma}_n \leq W_2(\mu,\mu_n)\Big(\int_{X\times X} L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_n\Big)^{1/2} \\ &= W_2(\mu,\mu_n)\Big(\int_X\Big(\int_X L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x}\Big)\,\mathrm{d}\mu(x)\Big)^{1/2}, \end{aligned}$$

where $\boldsymbol{\gamma}_n$ is any optimal plan between $\mu$ and $\mu_n$ and $\boldsymbol{\gamma}_{n,x}$ is its disintegration w.r.t. its first marginal $\mu$. Since $\int_X\big(\int_X \mathsf{d}^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x}(y)\big)\,\mathrm{d}\mu(x) \to 0$ as $n\to\infty$ we can assume with no loss of generality that

$$\lim_{n\to\infty}\int_X \mathsf{d}^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x}(y) = 0 \quad \text{for } \mu\text{-a.e. } x \in X,$$

thus taking into account (7.15) we get

$$\int_{X\setminus B_r(x)} L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x}(y) \to 0 \quad \text{for } \mu\text{-a.e. } x \in X,$$

for all $r > 0$. Taking an arbitrary radius $r > 0$ we get

$$\begin{aligned} \limsup_{n\to\infty}\int_X L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x} &\leq \limsup_{n\to\infty}\int_{B_r(x)} L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x} + \limsup_{n\to\infty}\int_{X\setminus B_r(x)} L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x} \\ &\leq \limsup_{n\to\infty}\int_{B_r(x)} L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x} \leq \sup_{y\in B_r(x)} L^2(x,y). \end{aligned}$$

Since $L(x,\cdot)$ is $\mathsf{d}$-upper semicontinuous, taking the limit as $r \downarrow 0$ in the previous estimate we get $\limsup_n \int_X L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x} \leq L^2(x,x)$ for $\mu$-a.e. $x \in X$. Using again (7.15), which provides a domination from above with a strongly convergent sequence, we are entitled to use Fatou's lemma to obtain

$$\begin{aligned}|\nabla \mathrm{Ent}_{\mathfrak{m}}|(\mu) &= \lim_{n\to\infty} \frac{\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\mu_n)}{W_2(\mu,\mu_n)} \le \int_X \limsup_{n\to\infty} \Big(\int_X L^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_{n,x}(y)\Big)^{1/2}\,\mathrm{d}\mu(x)\\ &\le \Big(\int_X L^2(x,x)\,\mathrm{d}\mu(x)\Big)^{1/2}.\end{aligned}$$

□

**Theorem 7.6** *Let* $(X,\tau,d,\mathfrak{m})$ *be a Polish extended space with* $\mathfrak{m}$ *satisfying* (4.2)*. Then* $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ *is sequentially lower semicontinuous w.r.t. strong convergence with moments in* $\mathscr{P}(X)$ *on sublevels of* $\mathrm{Ent}_{\mathfrak{m}}$ *if and only if*

$$|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\mu) = 4\int_X |\nabla\sqrt{\rho}|_*^2\,\mathrm{d}\mathfrak{m} \qquad \forall \mu = \rho\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}}). \tag{7.21}$$

*In this case* $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ *satisfies the following stronger lower semicontinuity property:*

$$\mu_n(B) \to \mu(B) \text{ for every } B \in \mathscr{B}(X) \quad \Longrightarrow \quad \liminf_{n\to\infty} |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu_n) \ge |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu). \tag{7.22}$$

*Proof.* If (7.21) holds then $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ coincides on its domain with a convex functional (by Lemma 4.10) which is lower semicontinuous with respect to strong convergence in $L^1(X,\mathfrak{m})$: therefore it is also weakly lower semicontinuous and (7.22) holds [7, §4.7(v)]. In particular $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ is sequentially lower semicontinuous with respect to strong convergence with moments.

To prove the converse implication, by Theorem 7.4 it is sufficient to prove the inequality $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\mu) \le 4\int_X |\nabla\sqrt{\rho}|_*^2\,\mathrm{d}\mathfrak{m}$. Assume first that $\rho \le M^2$ $\mathfrak{m}$-a.e. in $X$ for some $M \in [0,\infty)$. Taking Theorem 7.5 into account, it suffices to find a sequence of functions $\rho_n = \max\{f_m^2, c_m\mathrm{e}^{-2V^2}\}$ convergent to $\rho$ in $L^1(X,\mathfrak{m})$ and satisfying:

(a) $f_m$ is $\mathsf{d}$-Lipschitz, nonnegative, bounded from above by $M$ and null for $V$ sufficiently large;

(b) $\limsup_{n\to\infty} \frac{1}{2}\int_X |\nabla\sqrt{\rho_n}|^2\,\mathrm{d}\mathfrak{m} \to \mathsf{Ch}_*(\sqrt{\rho})$;

(c) $\int_X V^2\rho_n\,\mathrm{d}\mathfrak{m} \to \int_X V^2\rho\,\mathrm{d}\mathfrak{m}$.

Since the $\mathsf{d}$-Lipschitz property of the weight implies $\mathrm{e}^{-V^2} \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$ and $\int V^2\mathrm{e}^{-2V^2}\,\mathrm{d}\mathfrak{m} < \infty$, if we choose $c_m > 0$ infinitesimal it suffices to find $f_m$ satisfying (a), $\frac{1}{2}\int_X |\nabla f_m|^2\,\mathrm{d}\mathfrak{m} \to \mathsf{Ch}_*(\sqrt{\rho})$ and $\int_X V^2 f_m^2\,\mathrm{d}\mathfrak{m} \to \int_X V^2\rho\,\mathrm{d}\mathfrak{m}$.

To this aim, given $m > 0$, we fix a compact set $K \subset X$ such that $\int_{X\setminus K}\rho\,\mathrm{d}\mathfrak{m} < (1+m)^{-2}$ and a 1-Lipschitz function $\phi : X \to [0,1]$ equal to 1 on $K$ and equal to 0 out of the 1-neighbourhood of $K$, denoted by $\tilde{K}$. Notice that $\mathfrak{m}(\tilde{K}) < \infty$, since $V$ is bounded from above in $\tilde{K}$.

By approximation, for all $f \in D(\mathsf{Ch}_*)$ we have

$$|\nabla(f\phi)|_* \le \phi|\nabla f|_* + |f||\nabla\phi|_* \qquad \mathfrak{m}\text{-a.e. in } X.$$

Let now $(g_n)$ be a sequence of $\mathsf{d}$-Lipschitz functions convergent to $\sqrt{\rho}$ in $L^2(X,\mathfrak{m})$ and satisfying $\frac{1}{2}\int_X |\nabla g_n|^2\,\mathrm{d}\mathfrak{m} \to \mathsf{Ch}_*(\sqrt{\rho})$; by a simple truncation argument we can assume that all

$g_n$ satisfy $0 \le g_n \le M$. The bounded, nonnegative, $\mathsf{d}$-Lipschitz functions $g_n\phi$ converge in $L^2(X,\mathfrak{m})$ to $\sqrt{\rho}\phi$ and, thanks to the inequality $|\nabla\phi|_* \le \chi_{X\setminus K}$ $\mathfrak{m}$-a.e. in $X$ satisfy,

$$\limsup_{n\to\infty} \mathsf{Ch}_*(g_n\phi) \le (1+\frac{1}{m})\mathsf{Ch}_*(\sqrt{\rho}) + (1+m)\int_{X\setminus K}\rho\,\mathrm{d}\mathfrak{m} \le (1+\frac{1}{m})\mathsf{Ch}_*(\sqrt{\rho}) + \frac{1}{1+m}.$$

In addition, $g_n^2\phi^2 \to \rho\phi^2$ in $L^1(X,\mathfrak{m})$ because the functions vanish out of $\tilde{K}$. We conclude that we have also

$$\lim_{n\to\infty}\int_X V^2 g_n^2\phi^2\,\mathrm{d}\mathfrak{m} = \int_X V^2\rho\phi^2\,\mathrm{d}\mathfrak{m} \le \int_X V^2\rho\,\mathrm{d}\mathfrak{m}.$$

By a diagonal argument, choosing $f_m = g_n$ with $n = n(m)$ sufficiently large, the existence of a sequence $f_m$ with the stated properties is proved.

In the case when $\rho$ is not bounded we truncate $\rho$, without increasing its Cheeger's energy, and use once more the lower semicontinuity of $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$. □

## 7.3 Convexity of the squared slope

This subsection adapts and extends some ideas extracted from [16] to the more general framework considered in this paper. The main result of the section shows that the squared Wasserstein slope of the entropy is always convex (with respect to the linear structure in the space of measures), independently from the identification with the Fisher information considered in Theorem 7.6 (the identification therein relies on the assumption that $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ is sequentially lower semicontinuous with respect to strong convergence with moments).

Let us first introduce the notion of *push forward of a measure through a transport plan*: given $\boldsymbol{\gamma} \in \mathscr{P}(X\times X)$ with marginals $\gamma^i = \pi^i_\sharp\boldsymbol{\gamma}$ and given $\mu \in \mathscr{P}(X)$ we set

$$\boldsymbol{\gamma}_\mu := (\rho\circ\pi^1)\boldsymbol{\gamma} \quad \text{with} \quad \mu = \rho\boldsymbol{\gamma}^1 + \mu^s,\ \mu^s \perp \boldsymbol{\gamma}^1, \qquad \boldsymbol{\gamma}_\sharp\mu := \pi^2_\sharp\boldsymbol{\gamma}_\mu. \tag{7.23}$$

We recall that this construction first appeared, with a different notation, in Sturm's paper [35]. Notice that $\boldsymbol{\gamma}_\mu$ is a probability measure and $\pi^1_\sharp\boldsymbol{\gamma}_\mu = \mu$ if $\mu \ll \gamma^1$; in this case, if $(\boldsymbol{\gamma}_x)_{x\in X}$ is the disintegration of $\boldsymbol{\gamma}$ with respect to its first marginal $\gamma^1$, we have

$$\boldsymbol{\gamma}_\sharp\mu(B) = \int_X \boldsymbol{\gamma}_x(B)\,\mathrm{d}\mu(x) \quad \text{for every } B \in \mathscr{B}(X). \tag{7.24}$$

Since moreover $\boldsymbol{\gamma}_\mu \ll \boldsymbol{\gamma}$ we also have that $\boldsymbol{\gamma}_\sharp\mu \ll \gamma^2$.

Notice that

$$\boldsymbol{\gamma} = \int_X \delta_{\boldsymbol{r}(x)}\,\mathrm{d}\nu(x), \quad \mu \ll \nu \qquad \Longrightarrow \qquad \boldsymbol{\gamma}_\sharp\mu = \boldsymbol{r}_\sharp\mu. \tag{7.25}$$

In the next lemma we consider the real-valued map

$$\mu \mapsto G_{\boldsymbol{\gamma}}(\mu) := \mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu), \tag{7.26}$$

defined in the convex set

$$R_{\boldsymbol{\gamma}} := \big\{\mu \in \mathscr{P}(X) : \mu \ll \pi^1_\sharp\boldsymbol{\gamma},\ \mu, \boldsymbol{\gamma}_\sharp\mu \in D(\mathrm{Ent}_{\mathfrak{m}})\big\}. \tag{7.27}$$

In the simple case when $\boldsymbol{\gamma} = \int_X \delta_{\boldsymbol{r}(x)}\,\mathrm{d}\mathfrak{m}(x)$ with $\boldsymbol{r} : X \to X$ Borel bijection we may use first the representation formula (7.25) for $\boldsymbol{\gamma}_\sharp\mu$ and then (7.4) to obtain

$$\mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu) = \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{r}_\sharp\mu) = \mathrm{Ent}_{\mathfrak{m}'}(\mu)$$

with $\mathfrak{m}' := (\boldsymbol{r}^{-1})_\sharp\mathfrak{m}$. Since $\mu \in R_{\boldsymbol{\gamma}}$ we have $\mathrm{Ent}_{\mathfrak{m}'}(\mu) < \infty$ and we can use (7.8) for the change of reference measure in the relative entropy to get

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu) = \int_X \log\bigl(\frac{\mathrm{d}\mathfrak{m}'}{\mathrm{d}\mathfrak{m}}\bigr)\,\mathrm{d}\mu,$$

so that $G_{\boldsymbol{\gamma}}$ is linear w.r.t. $\mu$. In general, when $\boldsymbol{r}$ is not injective or $\boldsymbol{r}$ is multivalued, convexity persists:

**Lemma 7.7** *For every $\boldsymbol{\gamma} \in \mathscr{P}(X \times X)$ the map $G_{\boldsymbol{\gamma}}$ in (7.26) is convex in $R_{\boldsymbol{\gamma}}$.*

*Proof.* Let $\mu_1 = \rho_1\mathfrak{m}$, $\mu_2 = \rho_2\mathfrak{m} \in R_{\boldsymbol{\gamma}}$, set $\mu = \alpha_1\mu_1 + \alpha_2\mu_2$ with $\alpha_1 + \alpha_2 = 1$, $\alpha_1$, $\alpha_2 \in (0,1)$ and denote by $\theta_i \le 1/\alpha_i$ the densities of $\mu_i$ w.r.t. $\mu$.

We apply (7.8) of Lemma 7.2 with $\nu := \mu_i$ and $\mathfrak{n} := \mu$ to get

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_i) = \mathrm{Ent}_{\mu}(\mu_i) + \int_X \log\rho\,\mathrm{d}\mu_i,$$

where $\rho = \alpha_1\rho_1 + \alpha_2\rho_2$ is the density of $\mu$ w.r.t. $\mathfrak{m}$. Taking a convex combination of the previous equalities for $i = 1$, 2, we obtain

$$\alpha_1\mathrm{Ent}_{\mathfrak{m}}(\mu_1) + \alpha_2\mathrm{Ent}_{\mathfrak{m}}(\mu_2) = \alpha_1\mathrm{Ent}_{\mu}(\mu_1) + \alpha_2\mathrm{Ent}_{\mu}(\mu_2) + \mathrm{Ent}_{\mathfrak{m}}(\mu). \tag{7.28}$$

Analogously, setting $\nu_i := \boldsymbol{\gamma}_\sharp\mu_i$ and $\nu := \boldsymbol{\gamma}_\sharp\mu = \alpha_1\nu_1 + \alpha_2\nu_2$, we have

$$\alpha_1\mathrm{Ent}_{\mathfrak{m}}(\nu_1) + \alpha_2\mathrm{Ent}_{\mathfrak{m}}(\nu_2) = \alpha_1\mathrm{Ent}_{\nu}(\nu_1) + \alpha_2\mathrm{Ent}_{\nu}(\nu_2) + \mathrm{Ent}_{\mathfrak{m}}(\nu). \tag{7.29}$$

Combining (7.28) and (7.29) we obtain

$$\alpha_1 G_{\boldsymbol{\gamma}}(\mu_1) + \alpha_2 G_{\boldsymbol{\gamma}}(\mu_2) = G_{\boldsymbol{\gamma}}(\mu) + \sum_{i=1,2}\alpha_i\Bigl(\mathrm{Ent}_{\mu}(\mu_i) - \mathrm{Ent}_{\nu}(\nu_i)\Bigr). \tag{7.30}$$

Since $\nu_i = \pi^2_\sharp(\boldsymbol{\gamma}_{\mu_i})$ and $\nu = \pi^2_\sharp(\boldsymbol{\gamma}_\mu)$, (7.4) yields

$$\mathrm{Ent}_{\nu}(\nu_i) \le \mathrm{Ent}_{\boldsymbol{\gamma}_\mu}(\boldsymbol{\gamma}_{\mu_i}) = \mathrm{Ent}_{\boldsymbol{\gamma}_\mu}(\theta_i\boldsymbol{\gamma}_\mu) = \mathrm{Ent}_{\mu}(\mu_i),$$

where in the last equality we used that the first marginal of $\boldsymbol{\gamma}_\mu$ is $\mu$. Therefore (7.30) yields $\alpha_1 G_{\boldsymbol{\gamma}}(\mu_1) + \alpha_2 G_{\boldsymbol{\gamma}}(\mu_2) \ge G_{\boldsymbol{\gamma}}(\mu)$. □

**Theorem 7.8** *The squared descending slope $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2$ of the relative entropy is convex.*

*Proof.* Let $\mu_1$, $\mu_2 \in D(\mathrm{Ent}_{\mathfrak{m}})$ be measures with finite descending slope and let $\mu = \alpha_1\mu_1 + \alpha_2\mu_2$ with $\alpha_1, \alpha_2 \in (0,1)$, $\alpha_1 + \alpha_2 = 1$. Obviously $\mu \in D(\mathrm{Ent}_{\mathfrak{m}})$ and since it is not restrictive to assume $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu) > 0$, by definition of descending slope we can find a sequence $(\nu^n) \subset D(\mathrm{Ent}_{\mathfrak{m}})$ with $\mathrm{Ent}_{\mathfrak{m}}(\nu^n) \le \mathrm{Ent}_{\mathfrak{m}}(\mu)$ such that

$$W_2(\nu^n, \mu) \to 0, \qquad \frac{\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\nu^n)}{W_2(\mu, \nu^n)} \to |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu).$$

Let $\boldsymbol{\gamma}^n$ be optimal plans with marginals $\mu$ and $\nu^n$ respectively and let $\nu_i^n := \boldsymbol{\gamma}^n_\sharp \mu_i$. Since $\boldsymbol{\gamma}^n_{\mu_i}$ are optimal plans with marginals $\mu_i$ and $\nu_i^n$, we have

$$W_2^2(\mu_i, \nu_i^n) = \int_{X\times X} \mathsf{d}^2(x,y)\theta_i(x)\,\mathrm{d}\boldsymbol{\gamma}^n(x,y) \to 0 \quad \text{as } n \to \infty,$$

where $\theta_i \le \alpha_i^{-1}$ are the densities of $\mu_i$ w.r.t. $\mu$; in particular

$$W_2^2(\mu, \nu^n) = \alpha_1 W_2^2(\mu_1, \nu_1^n) + \alpha_2 W_2^2(\mu_2, \nu_2^n). \tag{7.31}$$

Since $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu_i) < \infty$, by the very definition of descending slope for every $S_i > |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu_i)$ there exists $\bar n \in \mathbb{N}$ satisfying

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_i) - \mathrm{Ent}_{\mathfrak{m}}(\nu_i^n) \le S_i W_2(\mu_i, \nu_i^n) \quad \text{for every } n \ge \bar n.$$

By Lemma 7.7 and (7.31) we get, for $n \ge \bar n$

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\nu^n) \le \alpha_1 S_1 W_2(\mu_1, \nu_1^n) + \alpha_2 S_2 W_2(\mu_2, \nu_2^n) \le \Big(\alpha_1 S_1^2 + \alpha_2 S_2^2\Big)^{1/2} W_2(\mu, \nu^n),$$

so that, passing to the limit as $n \to \infty$, our choice of $(\nu^n)$ yields that $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|(\mu)$ does not exceed $\big(\alpha_1 S_1^2 + \alpha_2 S_2^2\big)^{1/2}$. Taking the infimum with respect to $S_i$ we conclude. $\square$

# 8 The Wasserstein gradient flow of the entropy and its identification with the $L^2$ gradient flow of Cheeger's energy

## 8.1 Gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$: the case of bounded densities.

In the next result we show that any Wasserstein gradient flow (recall Definition 2.11) of the entropy functional with uniformly bounded densities coincides with the $L^2$-gradient flow of the Cheeger's functional. We prove in fact a slightly stronger result, starting from the energy dissipation inequality (2.25) instead of the identity (2.26), where we use the Fisher information functional $\mathsf{F}$ defined by Definition 4.9 instead of the squared slope of $\mathrm{Ent}_{\mathfrak{m}}$. Recall that $\mathsf{F}(f) \le |\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(f\mathfrak{m})$ by Theorem 7.4.

**Theorem 8.1** *Let $(X, \tau, \mathsf{d}, \mathfrak{m})$ be an extended Polish space satisfying (4.2) and let $\mu_t = f_t\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}})$, $t \in [0,T]$, be a curve in $\mathrm{AC}^2((0,T); (\mathscr{P}(X), W_2))$ satisfying the Entropy-Fisher dissipation inequality*

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_0) \ge \mathrm{Ent}_{\mathfrak{m}}(\mu_T) + \frac{1}{2}\int_0^T |\dot\mu_t|^2\,\mathrm{d}t + \frac{1}{2}\int_0^T \mathsf{F}(f_t)\,\mathrm{d}t. \tag{8.1}$$

*If $\sup_{t\in[0,T]} \|f_t\|_{L^\infty(X,\mathfrak{m})} < \infty$ then $f_t$ coincides in $[0,T]$ with the gradient flow $\mathsf{H}_t(f_0)$ of Cheeger's energy starting from $f_0$.*
*In particular, for all $f_0 \in L^\infty(X,\mathfrak{m})$ there exists at most one Wasserstein gradient flow $\mu_t = f_t\mathfrak{m}$ of $\mathrm{Ent}_{\mathfrak{m}}$ in $(\mathscr{P}_{[\mu]}(X), W_2)$ starting from $\mu_0 = f_0\mathfrak{m}$ with uniformly bounded densities $f_t$.*

*Proof.* Let us set $\mu_t^1 = \mu_t$, $f_t^1 = f_t$ and let us first observe that by Lemma 5.15 and (6.8) the curve $\mu_t^1$ satisfies

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_0^1) = \mathrm{Ent}_{\mathfrak{m}}(\mu_t^1) + \frac{1}{2}\int_0^t |\dot\mu_s^1|^2 \,\mathrm{d}s + \frac{1}{2}\int_0^t \mathsf{F}(f_s^1)\,\mathrm{d}s \quad \text{for every } t \in [0,T]. \tag{8.2}$$

Indeed, (5.13) and (6.8) show that the function defined by the right-hand side of (8.2) is nondecreasing with respect to $t$ and coincides with $\mathrm{Ent}_{\mathfrak{m}}(\mu_0^1)$ at $t = 0$ and at $t = T$ by (8.1).

Let $\mu_t^2 = f_t^2\mathfrak{m}$, with $f_t^2 := \mathsf{H}_t(f_0)$, be the solution of the $L^2$-gradient flow of the Cheeger's energy. Theorem 4.16 shows that $\|f_t^2\|_{L^\infty(X,\mathfrak{m})} \le \|f_0\|_{L^\infty(X,\mathfrak{m})}$; by Lemma 6.1 and Proposition 4.22 we get

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_0^2) \ge \mathrm{Ent}_{\mathfrak{m}}(\mu_t^2) + \frac{1}{2}\int_0^t |\dot\mu_s^2|^2 \,\mathrm{d}s + \frac{1}{2}\int_0^t \mathsf{F}(f_s^2)\,\mathrm{d}s \quad \text{for every } t \in [0,T]. \tag{8.3}$$

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

$$|\dot\mu_t|^2 \le \frac{|\dot\mu_t^1|^2 + |\dot\mu_t^2|^2}{2} \qquad \text{for a.e. } t \in (0,T). \tag{8.4}$$

Adding up (8.2) and (8.3) and using the convexity of the Fisher information functional (see Lemma 4.10), the convexity of the squared metric speed guaranteed by (8.4) and taking into account the *strict* convexity of $\mathrm{Ent}_{\mathfrak{m}}$ we deduce that for the curve $t \mapsto \mu_t := (\mu_t^1 + \mu_t^2)/2$ it holds

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) > \mathrm{Ent}_{\mathfrak{m}}(\mu_t) + \frac{1}{2}\int_0^t |\dot\mu_s|^2 \,\mathrm{d}s + \frac{1}{2}\int_0^t \mathsf{F}(f_s)\,\mathrm{d}s$$

for every $t$ such that $\mu_t^1 \ne \mu_t^2$, where $f_t := \frac{1}{2}(f_t^1 + f_t^2)$ is the density of $\mu_t$. This contradicts Lemma 5.15, which yields the opposite inequality. □

Although the result will not play a role in the paper, let's see that we can apply the previous theorem to characterize all limits of the JKO [22] – Minimizing Movement Scheme (see [4, Definition 2.0.6]) generated by the entropy functional in $\mathscr{P}_V(X)$. The result shows that starting from an initial datum with bounded density, the JKO scheme *always converges* to the $L^2$-gradient flow of Cheeger's energy, without any extra assumption on the space, except for the integrability condition (4.2).

For a given initial datum $\mu_0 = f_0\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}})$ and a time step $h > 0$ we consider the sequence $\mu_n^h = f_n^h\mathfrak{m}$ defined by the recursive variational problem

$$\mu_n^h \in \operatorname*{argmin}_{\mu \in \mathscr{P}_V(X)} \Big\{ \frac{1}{2h} W_2^2(\mu, \mu_{n-1}^h) + \mathrm{Ent}_{\mathfrak{m}}(\mu) \Big\},$$

and we set $\mu^h(t) = f^h(t)\mathfrak{m} := \mu_n^h$ if $t \in ((n-1)h, nh]$.

**Corollary 8.2 (Convergence of the minimizing movement scheme)** *Let $(X,\tau,\mathsf{d},\mathfrak{m})$ be an extended Polish space satisfying (4.2) and let $\mu_0 = f_0\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}})$ with $f_0 \in L^\infty(X,\mathfrak{m})$. Then for every $t \ge 0$ the family $\mu^h(t)$ weakly converges to $\mu_t = f_t\mathfrak{m}$ as $h \downarrow 0$, where $f_t = \mathsf{H}_t(f_0)$ is the $L^2$-gradient flow of Cheeger's energy.*

*Proof.* Arguing exactly as in [1, §2.1], [31, Proposition 2] it is not hard to show that $\|f_n^h\|_\infty \le \|f_0\|_\infty$.

We want to apply the theory developed in [4, Chap. 2-3]: according to the notation therein notation, $\mathscr{S}$ is the metric space $\mathscr{P}_{[\mu_0]}(X)$ endowed with the Wasserstein distance $W_2$, $\sigma$ is the weak topology in $\mathscr{P}(X)$, and $\phi$ is the Entropy functional $\mathrm{Ent}_\mathfrak{m}$. Since by (7.2) and (7.5) the negative part of $\mathrm{Ent}_\mathfrak{m}$ has at most quadratic growth in $\mathscr{P}_{\mu_0}(X)$, the basic assumptions [4, 2.1(a,b,c)] are satisfied and we can apply the compactness result [4, Corollary 3.3.4]: from any vanishing sequence of time steps $h_m \downarrow 0$ we can extract a subsequence (still denoted by $h_m$) such that $\mu^{h_m}(t) \to \mu_t = f_t\mathfrak{m}$ weakly in $\mathscr{P}(X)$, with $f^{h_m}(t) \rightharpoonup f_t$ weakly in any $L^p(X,\mathfrak{m})$, $p \in [1,\infty)$, and $\|f_t\|_\infty \le \|f_0\|_\infty$. Since the relaxed slope of the entropy functional, defined as

$$|\partial^-\mathrm{Ent}_\mathfrak{m}|(\mu) := \inf\Big\{\liminf_{n\to\infty} |\nabla^-\mathrm{Ent}_\mathfrak{m}|(\mu_n) : \mu_n \rightharpoonup \mu, \quad \sup_n W_2(\mu_n,\mu), \mathrm{Ent}_\mathfrak{m}(\mu_n) < \infty\Big\}$$

still satisfies the lower bound (7.11) $|\partial^-\mathrm{Ent}_\mathfrak{m}|(\rho\mathfrak{m}) \ge \mathsf{F}(\rho)$ thanks to the lower semicontinuity of the Fisher information with respect to the weak $L^1(X,\mathfrak{m})$-topology, the energy inequality [4, (3.4.1)] based on De Giorgi's variational interpolation yields

$$\mathrm{Ent}_\mathfrak{m}(\mu_0) \ge \mathrm{Ent}_\mathfrak{m}(\mu_T) + \frac{1}{2}\int_0^T |\dot\mu_t|^2\,\mathrm{d}t + \frac{1}{2}\int_0^T \mathsf{F}(f_t)\,\mathrm{d}t.$$

Applying the previous theorem we conclude that $f_t = \mathsf{H}_t(f_0)$. Since the limit is uniquely characterized, all the family $\mu^h(t)$ converges to $\mu_t$ as $h \downarrow 0$. □

## 8.2 Uniqueness of the Wasserstein gradient flow if $|\nabla^-\mathrm{Ent}_\mathfrak{m}|$ is an upper gradient

In the next theorem we prove uniqueness of the gradient flow of $\mathrm{Ent}_\mathfrak{m}$, a result that will play a key role in the equivalence results of the next section. Here we can avoid the uniform $L^\infty$ bound assumed in Theorem 8.1, but we need to suppose that $|\nabla^-\mathrm{Ent}_\mathfrak{m}|$ is an upper gradient for the entropy functional (a condition which is ensured by its geodesically $K$-convexity, see the next section).

**Theorem 8.3 (Uniqueness of the gradient flow of $\mathrm{Ent}_\mathfrak{m}$)** *Let $(X,\tau,\mathsf{d},\mathfrak{m})$ be a Polish extended space be such that $|\nabla^-\mathrm{Ent}_\mathfrak{m}|$ is an upper gradient of $\mathrm{Ent}_\mathfrak{m}$ and let $\mu \in D(\mathrm{Ent}_\mathfrak{m})$. Then there exists at most one gradient flow of $\mathrm{Ent}_\mathfrak{m}$ starting from $\mu$ in $(\mathscr{P}_{[\mu]}(X), W_2)$.*

*Proof.* As in [16] and in the proof of Theorem 8.1, assume that starting from some $\mu \in D(\mathrm{Ent}_\mathfrak{m})$ we can find two different gradient flows $(\mu_t^1)$ and $(\mu_t^2)$. Then we have

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

Adding up these two equalities and using the convexity of the squared slope guaranteed by Theorem 7.8, the convexity of the squared metric speed guaranteed by (8.4) and taking into account the *strict* convexity of $\mathrm{Ent}_\mathfrak{m}$ we deduce that for the curve $t \mapsto \mu_t := (\mu_t^1 + \mu_t^2)/2$ 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 $\mu_T^1 \neq \mu_T^2$. Taking the upper gradient property into account, this contradicts (2.24). $\square$

**Remark 8.4** The proofs of Theorem 8.3 and Theorem 8.1 do not rely on contractivity of the Wasserstein distance. Actually, as proved by Ohta and Sturm in [30], the property

$$W_2(\mu_t, \nu_t) \le e^{Kt} W_2(\mu_0, \nu_0)$$

for gradient flows of $\mathrm{Ent}_{\mathfrak{m}}$ in Minkowski spaces $(\mathbb{R}^n, \|\cdot\|, \mathscr{L}^n)$ whose norm is not induced by an inner product fails for any $K \in \mathbb{R}$. ■

### 8.3 Identification of the two gradient flows

Here we prove one of the main results of this paper, namely the identification of the gradient flow of $\mathsf{Ch}_*$ in $L^2(X, \mathfrak{m})$ and the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ in $(\mathscr{P}(X), W_2)$. The strategy consists in considering a gradient flow $(f_t)$ of $\mathsf{Ch}_*$ with nonnegative initial data and in proving that the curve $t \mapsto \mu_t := f_t \mathfrak{m}$ is a gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ in $(\mathscr{P}(X), W_2)$. All these results will be applied to the case of metric spaces satisfying a $CD(K, \infty)$ condition in the next section.

**Theorem 8.5 (Identification of the two gradient flows)** *Let $(X, \tau, \mathsf{d}, \mathfrak{m})$ be a Polish extended space such that (4.2) holds and let us assume that $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ is lower semicontinuous with respect to strong convergence with moments in $\mathscr{P}(X)$ on sublevels of $\mathrm{Ent}_{\mathfrak{m}}$. For all $f_0 \in L^2(X, \mathfrak{m})$ such that $\mu_0 = f_0 \mathfrak{m} \in \mathscr{P}_V(X)$ the following equivalence holds:*

*(i) If $f_t$ is the gradient flow of $\mathsf{Ch}_*$ in $L^2(X, \mathfrak{m})$ starting from $f_0$, then $\mu_t := f_t \mathfrak{m}$ is the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ in $(\mathscr{P}_{[\mu_0]}(X), W_2)$ starting from $\mu_0$, $t \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ is locally absolutely continuous in $(0, \infty)$ and*

$$-\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ent}_{\mathfrak{m}}(\mu_t) = |\dot{\mu}_t|^2 = |\nabla^- \mathrm{Ent}_{\mathfrak{m}}(\mu_t)|^2 \qquad \textit{for a.e. } t \in (0, \infty). \tag{8.5}$$

*(ii) Conversely, if $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ is an upper gradient of $\mathrm{Ent}_{\mathfrak{m}}$, and $\mu_t$ is the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ in $(\mathscr{P}_{[\mu_0]}(X), W_2)$ starting from $f_0 \mathfrak{m}$, then $\mu_t = f_t \mathfrak{m}$ and $f_t$ is the gradient flow of $\mathsf{Ch}_*$ in $L^2(X, \mathfrak{m})$ starting from $f_0$.*

*Proof.* (i) First of all, we remark that assumption (6.2) of Lemma 6.1 is satisfied, thanks to Theorem 4.20; in addition, the same theorem ensures that $\int_X V^2 f_t^2 \,\mathrm{d}\mathfrak{m} < \infty$ for all $t \ge 0$. Defining $\mu_t := f_t \mathfrak{m}$, we know by Proposition 4.22 that the map $t \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ is locally absolutely continuous in $(0, \infty)$ and that (4.54) holds.

On the other hand, since we assumed the lower semicontinuity of $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$, we can prove that $\mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ satisfies the energy dissipation inequality (2.25). Indeed, by Lemma 6.1 and Theorem 7.6 it holds:

$$\int_{\{f_t > 0\}} \frac{|\nabla f_t|_*^2}{f_t} \,\mathrm{d}\mathfrak{m} \ge \frac{1}{2} |\dot{\mu}_t|^2 + \frac{1}{2} |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_t) \qquad \text{for a.e. } t \in (0, \infty).$$

This proves that $\mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ satisfies the energy dissipation inequality. But, since we know that $t \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ is locally absolutely continuous we can apply Remark 2.6 to obtain that $|\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Ent}_{\mathfrak{m}}(\mu_t)| \le |\nabla^- \mathrm{Ent}_{\mathfrak{m}}|(\mu_t) |\dot{\mu}_t|$ for a.e. $t \in (0, \infty)$. Hence, as explained in Section 2.5,

(2.25) in combination with Young inequality and the previous inequality yield that all the inequalities turn a.e. into equalities, so that (8.5) holds.
(ii) We know that a gradient flow $\tilde f_t$ of $\mathsf{Ch}_*$ starting from $f_0$ exists, and part (i) gives that $\tilde\mu_t := \tilde f_t \mathfrak{m}$ is a gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$. By Theorem 8.3, there is at most one gradient flow starting from $\mu_0$, hence $\mu_t = \tilde\mu_t$ for all $t \geq 0$. □

As a consequence of the identification result, we present a general existence result of the Wasserstein gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ which includes also the case of $\sigma$-finite measures.

**Theorem 8.6 (Existence of the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$)** *Let $(X, \tau, d, \mathfrak{m})$ be a Polish extended space satisfying assumption (4.2) and such that $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ is lower semicontinuous with respect to strong convergence with moments in $\mathscr{P}(X)$ on sublevels of $\mathrm{Ent}_{\mathfrak{m}}$.*
*Then for all $\mu = \rho\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}})$ there exists a gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ starting from $\mu$ in $(\mathscr{P}_{[\mu]}(X), W_2)$.*

*Proof.* For completeness we provide a proof that does not use the identification of gradient flows in the case when $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ is also an upper gradient for the entropy functional: indeed, we can apply the existence result [4, Prop. 2.2.3, Thm. 2.3.3], achieved via the so-called minimizing movements technique, with the topology of weak convergence in duality with $C_b(X)$. Remark 7.3, (7.2), and the lower semicontinuity part of Theorem 7.6 give that the assumptions are satisfied, and we get measures $\mu_t$ satisfying

$$\mathrm{Ent}_{\mathfrak{m}}(\rho\mathfrak{m}) = \mathrm{Ent}_{\mathfrak{m}}(\mu_t) + \int_0^t \frac{1}{2}|\dot\mu_s|^2 + \frac{1}{2}|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|^2(\mu_s)\,\mathrm{d}s \qquad \forall t \geq 0. \tag{8.6}$$

In the general case, we can take advantage of the identification of gradient flows and immediately obtain existence when $\rho \in L^2(X, \mathfrak{m})$. If only the integrability conditions $\int_X \rho\log\rho\,\mathrm{d}\mathfrak{m} < \infty$ and $\int_X V^2\rho\,\mathrm{d}\mathfrak{m} < \infty$ are available, we can use the same monotone approximation argument as in the proof of Theorem 7.4: keeping that notation, we set $\mu_t^n := \rho_t^n\mathfrak{m}$, and we decompose the function $\rho\log\rho$ into the sum $h_-(\rho) + h_+(\rho)$ where $h_-(\rho) = \min(\rho, \mathrm{e}^{-1})\log(\min(\rho, \mathrm{e}^{-1}))$ and $h_+(\rho) = \max(\rho, \mathrm{e}^{-1})\log(\max(\rho, \mathrm{e}^{-1})) + \mathrm{e}^{-1}$ are decreasing and increasing functions respectively. Applying the monotone convergence theorem to $h_\pm(\rho_t^n)$ we easily get $\mathrm{Ent}_{\mathfrak{m}}(\mu_t^n) \to \mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ as $n \to \infty$. By the lower semicontinuity of the Fisher information we can pass to the limit in the integral form of (4.54), written for $\rho_t^n$, obtaining

$$\big|\mathrm{Ent}_{\mathfrak{m}}(\mu_{t_0}) - \mathrm{Ent}_{\mathfrak{m}}(\mu_{t_1})\big| \leq \int_{t_0}^{t_1} \mathsf{F}(\rho_s)\,\mathrm{d}s \quad \text{for every } 0 \leq t_0 < t_1.$$

It follows that $t \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ is absolutely continuous. We can now pass to the limit in (8.6) written for $\mu_t^n$ by using the lower semicontinuity of $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ and of the 2-energy to obtain a curve $\mu_t$ satisfying the entropy-dissipation inequality (2.25). Since $t \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ is absolutely continuous, (2.25) yields (2.26) and we conclude that $\mu_t$ is a gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$. □

# 9 Metric measure spaces satisfying $CD(K, \infty)$

In this section we present the applications of the previous theory in the case when the Polish extended space $(X, \tau, d, \mathfrak{m})$ has Ricci curvature bounded from below, according to [25] and [35]. Under this condition the Wasserstein slope $|\nabla^- \mathrm{Ent}_{\mathfrak{m}}|$ turns out to be a lower semicontinuous upper gradient of the entropy, so that all the assumptions of Theorems 8.3, 8.5, and 8.6 are satisfied.

**Definition 9.1** ($CD(K,\infty)$) *We say that $(X,\tau,d,\mathfrak{m})$ has Ricci curvature bounded from below by $K \in \mathbb{R}$ if $\mathrm{Ent}_{\mathfrak{m}}$ is $K$-convex along geodesics in $(\mathscr{P}(X),W_2)$. More precisely, this means that for any $\mu_0$, $\mu_1 \in D(\mathrm{Ent}_{\mathfrak{m}}) \subset \mathscr{P}(X)$ with $W_2(\mu_0,\mu_1) < \infty$ there exists a constant speed geodesic $\mu_t : [0,1] \to \mathscr{P}(X)$ between $\mu_0$ and $\mu_1$ satisfying*

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t) \le (1-t)\mathrm{Ent}_{\mathfrak{m}}(\mu_0) + t\mathrm{Ent}_{\mathfrak{m}}(\mu_1) - \frac{K}{2}t(1-t)W_2^2(\mu_0,\mu_1) \qquad \forall t \in [0,1]. \tag{9.1}$$

Notice that unlike the definitions given in [25] and [35], here we are allowing the distance $\mathsf{d}$ to attain the value $+\infty$. Also, even if $\mathsf{d}$ were finite, this definition slightly differs from the standard one, as typically geodesic convexity is required only in the space $(\mathscr{P}_2(X),W_2)$, while here we are assuming it to hold for any couple of probability measures with finite entropy and distance. Actually, the two are equivalent, as a simple approximation argument based on the tightness given by Remark 7.3 shows.

**Remark 9.2 (The integrability condition** (4.2)**)** If $(X,\tau,\mathsf{d}\,\mathfrak{m})$ satisfies a $CD(K,\infty)$ condition and $\tau$ is the topology induced by the finite distance $\mathsf{d}$, then (4.2) is equivalent (see [35, Theorem 4.24]) to assume that for every $x \in X$ there exists $r > 0$ such that $\mathfrak{m}(B_r(x)) < \infty$. In this case one can choose $V(x) := A\mathsf{d}(x,x_0)$ for a suitable constant $A \ge 0$ and $x_0 \in X$. ■

**Theorem 9.3 (Slope, Fisher, and gradient flows)** *Let $(X,\tau,\mathsf{d},\mathfrak{m})$ be a Polish extended space satisfying $CD(K,\infty)$ and* (4.2)*.*

*(i) For every $\mu = f\mathfrak{m} \in \mathscr{P}_V(X)$ the Wasserstein slope $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\mu)$ coincides with the Fisher information of $f$, it is lower semicontinuous under setwise convergence, according to* (7.22)*, and it is an upper gradient for* $\mathrm{Ent}_{\mathfrak{m}}$*.*

*(ii) For every $\mu_0 = f_0\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}})$ there exists a unique gradient flow $\mu_t = f_t\mathfrak{m}$ of $\mathrm{Ent}_{\mathfrak{m}}$ starting from $\mu_0$ in $(\mathscr{P}_{[\mu]}(X),W_2)$.*

*(iii) If moreover $f_0 \in L^2(X,\mathfrak{m})$, the gradient flow $f_t = \mathsf{H}(f_0)$ of $\mathsf{Ch}_*$ in $L^2(X,\mathfrak{m})$ starting from $f_0$ and the gradient flow $\mu_t$ of $\mathrm{Ent}_{\mathfrak{m}}$ in $(\mathscr{P}_{[\mu_0]}(X),W_2)$ starting from $\mu_0$ coincide, i.e. $\mu_t = f_t\mathfrak{m}$ for every $t > 0$.*

Thanks to this theorem, under the $CD(K,\infty)$ assumption we can unambiguously say that a *Heat Flow* on $(X,\tau,\mathsf{d},\mathfrak{m})$ is either a gradient flow of Cheeger's energy in $L^2(X,\mathfrak{m})$ or a gradient flow of the relative entropy in $(\mathscr{P}(X),W_2)$, at least for square integrable initial conditions with finite moment.

Concerning the *proof* of Theorem 9.3, we observe that applying the results of the previous section it is sufficient to show that the Wasserstein slope $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ is lower semicontinuous w.r.t. strong convergence with moments in $\mathscr{P}(X)$ on the sublevel of $\mathrm{Ent}_{\mathfrak{m}}$. In fact, if this property holds, (7.21) of Theorem 7.6 shows that $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ coincides with the Fisher functional and thus satisfies the lower semicontinuity property (7.22): in particular it is lower semicontinuous w.r.t. weak convergence with moments in $\mathscr{P}(X)$. Applying Theorem 8.6 we prove the existence of the Wasserstein gradient flow starting from $\mu_0$; its uniqueness follows from Theorem 8.3, since the slope is always an upper gradient of $\mathrm{Ent}_{\mathfrak{m}}$ under $CD(K,\infty)$. Applying Theorem 8.5 we can thus obtain the identification of the two gradient flows.

In order to prove the lower semicontinuity of the slope $|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|$ w.r.t. strong convergence with moments in $\mathscr{P}(X)$ (Proposition 9.7) we proceed in various steps, adapting the arguments of [16].

**Definition 9.4 (Plans with bounded deformation)** *Let $\tilde{\mathfrak{m}} := \mathrm{e}^{-V^2}\mathfrak{m}$, where $V$ satisfies (4.2). We say that $\boldsymbol{\gamma} \in \mathscr{P}(X^2)$ has bounded deformation if*

$$\mathsf{d} \in L^\infty(X \times X, \boldsymbol{\gamma}) \quad \textit{and} \quad c\tilde{\mathfrak{m}} \le \pi^i_\sharp \boldsymbol{\gamma} \le \frac{1}{c}\tilde{\mathfrak{m}},\ i = 1,\, 2, \textit{ for some } c > 0. \tag{9.2}$$

**Proposition 9.5 (Sequential lower semicontinuity of $G_{\boldsymbol{\gamma}}$)** *For any plan $\boldsymbol{\gamma}$ with bounded deformation the map $\mu \mapsto G_{\boldsymbol{\gamma}}(\mu) = \mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp \mu)$ (recall Section 7.3) is sequentially lower semicontinuous with respect to weak convergence with moments, on sequences with* $\mathrm{Ent}_{\mathfrak{m}}$ *uniformly bounded from above.*

*Proof.* Let $\mu_n = \eta_n \tilde{\mathfrak{m}} \in \mathscr{P}_V(X)$ be weakly convergent with moments to $\mu = \eta\tilde{\mathfrak{m}}$, with $\mathrm{Ent}_{\mathfrak{m}}(\mu_n)$ uniformly bounded. If $\rho$ denotes the density of $\pi^1_\sharp\boldsymbol{\gamma}$ w.r.t. $\tilde{\mathfrak{m}}$, we have that $(\eta_n/\rho) \circ \pi^1 \boldsymbol{\gamma}$ is an admissible plan between $\mu_n$ and $\boldsymbol{\gamma}_\sharp \mu_n$, hence $\rho^{-1} \in L^\infty(X, \tilde{\mathfrak{m}})$ and $\mathsf{d} \in L^\infty(X \times X, \boldsymbol{\gamma})$ ensure that $\boldsymbol{\gamma}_\sharp\mu_n$ belong to $\mathscr{P}_V(X)$ as well. A similar argument also shows that $\boldsymbol{\gamma}_\sharp\mu_n$ converge with moments to $\boldsymbol{\gamma}_\sharp\mu$. From (7.5) we obtain that

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_n) - \mathrm{Ent}_{\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu_n) = \mathrm{Ent}_{\tilde{\mathfrak{m}}}(\mu_n) - \mathrm{Ent}_{\tilde{\mathfrak{m}}}(\boldsymbol{\gamma}_\sharp\mu_n) - \int_X V^2 \,\mathrm{d}\mu_n + \int_X V^2 \,\mathrm{d}\boldsymbol{\gamma}_\sharp\mu_n$$

and that $\mathrm{Ent}_{\tilde{\mathfrak{m}}}(\mu_n)$ are uniformly bounded. So, we are basically led, after a normalization, to the case of a probability reference measure $\tilde{\mathfrak{m}}$. In this case the proof uses the equiintegrability in $L^1(\tilde{\mathfrak{m}})$ of $\eta_n$, ensured by the upper bound on entropy, see [16, Proposition 11] for details. □

**Lemma 9.6 (Approximation)** *If $\mu,\, \nu \in D(\mathrm{Ent}_{\mathfrak{m}})$ satisfy $W_2(\mu,\nu) < \infty$ then there exist plans $\boldsymbol{\gamma}_n$ with bounded deformation satisfying*

$$\int_{X\times X} \mathsf{d}^2 \,\mathrm{d}(\boldsymbol{\gamma}_n)_\mu \to W_2^2(\mu,\nu) \quad \textit{and} \quad \mathrm{Ent}_{\mathfrak{m}}((\boldsymbol{\gamma}_n)_\sharp\mu) \to \mathrm{Ent}_{\mathfrak{m}}(\nu).$$

*Proof.* Let $\rho$, $\eta$ be respectively the densities of $\mu$ and $\nu$ w.r.t. $\tilde{\mathfrak{m}}$, let $\boldsymbol{\gamma}$ be an optimal plan relative to $\mu$ and $\nu$ and set $\boldsymbol{\gamma}_n := z_n^{-1}\chi_{E_n}\boldsymbol{\gamma}$, where

$$E_n := \{(x,y) :\ \rho(x) + \mathsf{d}(x,y) + \eta(y) \le n\}, \qquad z_n := \boldsymbol{\gamma}(E_n) \uparrow 1.$$

By monotone convergence it is immediate to check that $\boldsymbol{\gamma}_n$ satisfy the first convergence property, because $z_n\boldsymbol{\gamma}_n \uparrow \boldsymbol{\gamma}$. In connection with the second one, since $z_n(\boldsymbol{\gamma}_n)_\mu \uparrow (\boldsymbol{\gamma})_\mu = \boldsymbol{\gamma}$, considering the second marginals $z_n(\boldsymbol{\gamma}_n)_\sharp\mu := \eta_n\mathfrak{m}$ we see that $\eta_n \uparrow \eta$ $\mathfrak{m}$-a.e. in $X$.

It is clear that $\mathsf{d} \in L^\infty(X \times X, \boldsymbol{\gamma}_n)$ and that the marginals of $\boldsymbol{\gamma}_n$ have bounded densities with respect to $\tilde{\mathfrak{m}}$. Considering convex combinations $\tilde{\boldsymbol{\gamma}}_n := (1 - 1/n)\boldsymbol{\gamma}_n + \boldsymbol{\gamma}^0/n$, with $\boldsymbol{\gamma}^0 := (Id, Id)_\sharp\tilde{\mathfrak{m}}$, we obtain that the densities are bounded also from below, so that $\tilde{\boldsymbol{\gamma}}_n$ have bounded deformation; in addition, since $\boldsymbol{\gamma}^0_\sharp\mu = \mu$, we obtain

$$(\tilde{\boldsymbol{\gamma}}_n)_\sharp\mu = \big(1 - \frac{1}{n}\big) z_n^{-1}\eta_n + \frac{1}{n}\mu,$$

so that monotone convergence (see the argument in the proof of Theorem 8.6), convexity and lower semicontinuity of the entropy ensure that still $\mathrm{Ent}_{\mathfrak{m}}((\tilde{\boldsymbol{\gamma}}_n)_\sharp\mu) \to \mathrm{Ent}_{\mathfrak{m}}(\nu)$. □

**Proposition 9.7** **($|\nabla^-\mathrm{Ent}_{\mathfrak m}|$ is a l.s.c. slope in $CD(K,\infty)$ spaces)** *Assume that $(X,\tau,d,\mathfrak m)$ is a Polish extended space satisfying $CD(K,\infty)$ and (4.2) holds. Then $D(\mathrm{Ent}_{\mathfrak m})\ni\mu\mapsto|\nabla^-\mathrm{Ent}_{\mathfrak m}|^2(\mu)$ is sequentially lower semicontinuous w.r.t. weak convergence with moments on the sublevels of $\mathrm{Ent}_{\mathfrak m}$. In particular (7.21) holds.*

*Proof.* In this proof we denote by $C(\boldsymbol\gamma)$ the cost of $\boldsymbol\gamma$, i.e. $C(\boldsymbol\gamma):=\int \mathsf d^2\,\mathrm d\boldsymbol\gamma$. We closely follow [16, Theorem 12 and Corollary 13].

Let $\mu=\rho\mathfrak m$ in the domain of the entropy. Taking (2.22) and the $K$-geodesic convexity of $\mathrm{Ent}_{\mathfrak m}$ into account, we first prove that it holds

$$|\nabla^-\mathrm{Ent}_{\mathfrak m}|(\mu)=\sup_{\gamma}\frac{\big(G_{\boldsymbol\gamma}(\mu)-\frac{K^-}{2}C(\boldsymbol\gamma_\mu)\big)^+}{\sqrt{C(\boldsymbol\gamma_\mu)}},\tag{9.3}$$

where the supremum runs in the class of plans $\boldsymbol\gamma$ with bounded deformation. Indeed, inequality $\geq$ follows choosing $\nu=\boldsymbol\gamma_\sharp\mu$ in (2.22) and using the trivial inequality

$$a\in\mathbb R,\ c\geq b>0\qquad\Longrightarrow\qquad\frac{(a-b)^+}{\sqrt b}\geq\frac{(a-c)^+}{\sqrt c},$$

with $a=G_{\boldsymbol\gamma}(\mu)$, $b=W_2^2(\mu,\nu)$ and $c=C(\boldsymbol\gamma_\mu)$, together with the fact that $C(\boldsymbol\gamma_\mu)\geq W_2^2(\mu,\nu)$. The other inequality is a consequence of the approximation Lemma 9.6.

To conclude, it is sufficient to prove that for all $\boldsymbol\gamma$ with bounded deformation the map $\mu\mapsto\big(G_{\boldsymbol\gamma}(\mu)-\frac{K^-}{2}C(\boldsymbol\gamma_\mu)\big)^+/C(\boldsymbol\gamma_\mu)^{1/2}$ is sequentially lower semicontinuous with respect to weak convergence with moments on the sublevels of the entropy. This follow by Proposition 9.5 and the fact that $\mu\mapsto C(\boldsymbol\gamma_\mu)$ is continuous along these sequences. In turn, the continuity property along these sequences follows by the representation

$$C(\boldsymbol\gamma_\mu)=\int_{X\times X}\frac{\mathrm d\mu}{\mathrm d\tilde{\mathfrak m}}(x)\Big(\frac{\mathrm d\pi^1_\sharp\boldsymbol\gamma}{\mathrm d\tilde{\mathfrak m}}(x)\Big)^{-1}\mathsf d^2(x,y)\,\mathrm d\boldsymbol\gamma(x,y).$$

Indeed, both $\big(\mathrm d\pi^1_\sharp\boldsymbol\gamma/\mathrm d\tilde{\mathfrak m}\big)^{-1}$ and $\mathsf d$ are essentially bounded, while the densities $\mathrm d\mu/\mathrm d\tilde{\mathfrak m}$ are equiintegrable, as we saw in the proof of Proposition 9.5. □

# 10 A metric Brenier theorem and gradients of Kantorovich potentials

In this section we provide a "metric" version of Brenier's theorem and we identify ascending slope and minimal weak upper gradient of Kantorovich potentials. These results depend on $L^\infty$ upper bound on interpolations, a property that holds in spaces with Riemannian lower bounds on Ricci curvature, see [3], or in nonbranching $CD(K,\infty)$ metric spaces (because the nonbranching property is inherited by $(\mathscr P(X),W_2)$, see [38, Corollary 7.32], [2, Proposition 2.16], and all $p$-entropies are convex). If $\mathsf d$ is bounded, the $L^\infty$ bound can be relaxed to an easier bound on entropy, but modifying the class of test plans, see Remark 10.7. We assume throughout this section that

$$\mathsf d\text{ is a finite distance and }\mathfrak m\text{ satisfies (4.2).}$$

However, we keep the possibility of considering the case when $\tau$ is not induced by $\mathsf d$.

In this section we denote by $\mathcal{T}$ the class of test plans concentrated on $\mathrm{AC}^2([0,1];(X,\mathsf{d}))$ with bounded compression on the sublevels of $V$ and by $\mathcal{G} \subset \mathcal{T}$ the subclass of test plans concentrated on $\mathrm{Geo}(X)$. By Remark 5.11 we have the obvious relation

$$|\nabla f|_{w,\mathcal{G}} \le |\nabla f|_{w,\mathcal{T}}. \tag{10.1}$$

In the next lemma we prove that, for Kantorovich potentials $\varphi$, $t \mapsto \varphi(\gamma_t)$ is not only Sobolev but also absolutely continuous along $\mathcal{T}$-almost all curves in $\mathrm{AC}^2([0,1];(X,\mathsf{d}))$. This holds even though in our general framework no Lipschitz continuity property (not even a local one) of $\varphi$ can be hoped for; in particular, by Remark 2.6 we obtain that $|\nabla^+\varphi|$ is a $\mathcal{T}$-weak upper gradient of $\varphi$.

**Lemma 10.1 (Slope is a weak upper gradient for Kantorovich potentials)** *Let $\mu = \rho\mathfrak{m} \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(X)$ with $W_2(\mu,\nu) < \infty$ and let $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ be a Kantorovich potential relative to some optimal plan $\boldsymbol{\gamma}$ between $\mu$ and $\nu$. If $\rho$ satisfies*

$$\rho \ge c_M > 0 \quad \mathfrak{m}\text{-a.e. in } \{V \le M\}, \text{ for all } M \ge 0 \tag{10.2}$$

*then $\varphi$ is absolutely continuous along $\mathcal{T}$-almost every curve of $\mathrm{AC}^2([0,1];(X,\mathsf{d}))$ and the slope $|\nabla^+\varphi|$ is a $\mathcal{T}$-weak upper gradient of $\varphi$.*

*Proof.* Set $f = -\varphi^c$, so that $\varphi = Q_1 f$ (here we adopt the notation of §3) and the set $\mathcal{D}(f)$ in (3.3) coincides with $X$. By Proposition 3.9 we know that the function

$$D^*(x) := \int_X \mathsf{d}(x,y)\,\mathrm{d}\boldsymbol{\gamma}_x(y)$$

(where $\{\boldsymbol{\gamma}_x\}_{x\in X}$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. $\mu$) belongs to $L^2(X,\mu)$ and bounds $\mu$-a.e. from above $D^-(x,1)$ by (3.23), and then $\mathfrak{m}$-a.e.; we know also from (3.13b) that $|\nabla^+\varphi| \in L^2(X,\mu)$ and that $D^-(x,1) \ge |\nabla^+\varphi|(x)$ wherever $\varphi(x) > -\infty$. We modify $D^*$ in a $\mathfrak{m}$-negligible set, getting a function $\tilde D \in L^2(X,\mu)$ larger than $D^-(x,1)$ everywhere and equal to $+\infty$ on the $\mathfrak{m}$-negligible set $\{\varphi = -\infty\}$.

We claim now that the condition $\int_\gamma \tilde D < \infty$ is fulfilled for $\mathcal{T}$-almost every $\gamma$ in $\mathrm{AC}^2([0,1];(X,\mathsf{d}))$. Indeed, arguing as in (5.11), for any test plan $\boldsymbol{\pi} \in \mathcal{T}$ with $\mathcal{E}_2[\gamma] \le N^2 < \infty$ $\boldsymbol{\pi}$-a.e. we have

$$\int\int_{\gamma\cap\{V\le M\}} \tilde D\,\mathrm{d}\boldsymbol{\pi} \le N\Big(C(\boldsymbol{\pi},M)\int_{\{V\le M\}} \tilde D^2\,\mathrm{d}\mathfrak{m}\Big)^{1/2} \le N\Big(c_M^{-1}C(\boldsymbol{\pi},M)\int_X \tilde D^2\,\mathrm{d}\mu\Big)^{1/2} < \infty,$$

thanks to the fact that $\rho \ge c_M$ on $\{V \le M\}$. Since $M$ is arbitrary and since $\boldsymbol{\pi}$-a.e. curve $\gamma$ is contained in $\{V \le M\}$ for sufficiently large $M$, the claim follows.

Now, let $\gamma \in \mathrm{AC}^2([0,1];(X,\mathsf{d}))$ with $\int_\gamma \tilde D < \infty$, and $\lambda = |\dot\gamma|\mathscr{L}^1|_{[0,1]}$. $\varphi$ is $\mathsf{d}$-upper semicontinuous and $\varphi\circ\gamma$ is finite $\lambda$-a.e. (since $\tilde D \circ \gamma$ is finite $\lambda$-a.e.). By (3.14) with $x = \gamma_s$ and $y = \gamma_t$, taking also the inequality $D^-(x,1) \le \tilde D(x)$ into account, we get

$$\varphi(\gamma_s) - \varphi(\gamma_t) \le \mathsf{d}(\gamma_s,\gamma_t)\Big(\tilde D(\gamma_t) + \frac{\mathsf{d}(\gamma_s,\gamma_t)}{2}\Big) \le \Big|\int_s^t |\dot\gamma_r|\,\mathrm{d}r\Big|\Big(\tilde D(\gamma_t) + \mathrm{diam}(\gamma)\Big)$$

for all $t$ such that $\varphi(\gamma_t) > -\infty$. Hence we can apply Corollary 2.8 to conclude that $\varphi\circ\gamma$ is absolutely continuous in $[0,1]$. Recalling Remark 2.6 we get

$$\Big|\int_{\partial\gamma}\varphi\Big| \le \int_\gamma |\nabla^+\varphi|.$$

□

Using (10.1), the previous lemma and Proposition 3.9, we have the chain of inequalities

$$|\nabla\varphi|_{w,\mathcal{G}}(x) \le |\nabla\varphi|_{w,\mathcal{T}}(x) \le |\nabla^+\varphi|(x) \le \mathsf{d}(x,y) \qquad \boldsymbol{\gamma}\text{-a.e. in } X\times X \tag{10.3}$$

for any optimal plan $\boldsymbol{\gamma}$. In the next theorem we show that an $L^\infty$ bound on geodesic interpolation ensures that the inequalities are actually equalities.

To perform geodesic interpolation we will assume that $(X,\mathsf{d})$ is a geodesic space. In such spaces, the optimal transport problem can be "lifted" to $\mathrm{Geo}(X)$ considering all $\boldsymbol{\pi}\in\mathscr{P}(\mathrm{Geo}(X))$ whose marginals at time 0 and at time 1 are respectively $\mu$ and $\nu$ and minimizing

$$\int\int_0^1 |\dot\gamma_s|^2\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}(\gamma) = \int \mathsf{d}^2(\gamma_0,\gamma_1)\,\mathrm{d}\boldsymbol{\pi}(\gamma)$$

in this class. Since $(e_0,e_1)_\sharp\boldsymbol{\pi}$ is an admissible plan between $\mu$ and $\nu$, it turns out that the infimum is larger than $W_2^2(\mu,\nu)$. But a simple measurable geodesic selection argument provides equivalence of the problems and existence of optimal $\boldsymbol{\pi}$. This motivates the next definition.

**Definition 10.2 (Optimal geodesic plans)** *Let $\mu$, $\nu\in\mathscr{P}(X)$ be such that $W_2(\mu,\nu)<\infty$. A plan $\boldsymbol{\pi}\in\mathscr{P}(\mathrm{Geo}(X))$ is an optimal geodesic plan between $\mu$ and $\nu$ if*

$$(\mathrm{e}_0)_\sharp\boldsymbol{\pi}=\mu,\quad (\mathrm{e}_1)_\sharp\boldsymbol{\pi}=\nu,\quad \int \mathsf{d}^2(\gamma_0,\gamma_1)\,\mathrm{d}\boldsymbol{\pi}(\gamma) = \int\int_0^1 |\dot\gamma_s|^2\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}(\gamma) = W_2^2(\mu,\nu).$$

It is easy to check that

$$t \quad\mapsto\quad (\mathrm{e}_t)_\sharp\boldsymbol{\pi}, \tag{10.4}$$

is a constant speed geodesic in $\mathscr{P}(X)$ from $\mu$ to $\nu$ for all optimal geodesic plans between $\mu$ and $\nu$. In particular, $(\mathscr{P}(X),W_2)$ is geodesic as well. Also, $(e_0,e_1)_\sharp\boldsymbol{\pi}$ is an optimal coupling whenever $\boldsymbol{\pi}$ is an optimal geodesic plan.

Adapting the arguments in [38, Theorem 7.21, Corollary 7.22] for the locally compact case and [23, 2] for the complete case, it can be shown that in any geodesic extended Polish space $(X,\tau,\mathsf{d})$ (10.4) provides a description of *all* constant speed geodesics, see [24].

**Theorem 10.3 (A metric Brenier's theorem)** *Let $\mu=\rho\mathfrak{m}\in\mathscr{P}(X)$ be satisfying (10.2), let $\nu\in\mathscr{P}(X)$ with $W_2(\mu,\nu)<\infty$, let $\boldsymbol{\pi}$ be an optimal geodesic plan between $\mu$ and $\nu$ and let $\varphi: X\to\mathbb{R}\cup\{-\infty\}$ be a Kantorovich potential relative to $(\mathrm{e}_0,\mathrm{e}_1)_\sharp\boldsymbol{\pi}$. Assume that $(\mathrm{e}_s)_\sharp\boldsymbol{\pi}=\mu_s=\rho_s\mathfrak{m}$ for all $s>0$ sufficiently small and that*

$$\limsup_{s\downarrow 0}\|\rho_s\|_{L^\infty(\{V\le M\},\mathfrak{m})}<\infty \qquad \forall M>0. \tag{10.5}$$

*Then*

$$\mathsf{d}(\gamma_1,\gamma_0) = |\nabla^+\varphi|(\gamma_0) = |\nabla\varphi|_{w,\mathcal{T}}(\gamma_0) = |\nabla\varphi|_{w,\mathcal{G}}(\gamma_0) \qquad \text{for } \boldsymbol{\pi}\text{-a.e. } \gamma\in\mathrm{Geo}(X). \tag{10.6}$$

*As a consequence, $W_2^2(\mu,\nu)=\int_X|\nabla^+\varphi|^2\,\mathrm{d}\mu$ and $|\nabla^+\varphi|=|\nabla\varphi|_{w,\mathcal{T}}=|\nabla\varphi|_{w,\mathcal{G}}$ $\mathfrak{m}$-a.e. in $X$.*

*Proof.* Set $g := |\nabla\varphi|_{w,\mathcal{G}} \in L^2(X,\mathfrak{m})$. Taking (10.3) into account, (10.6) can be achieved if we show that $\int \mathsf{d}^2(\gamma_1,\gamma_0)\,\mathrm{d}\boldsymbol{\pi} \le \int g^2(\gamma_0)\,\mathrm{d}\boldsymbol{\pi}$. Setting $f = -\varphi^c$ so that $\varphi = Q_1 f$, for $\boldsymbol{\pi}$-a.e. $\gamma \in \mathrm{Geo}(X)$ we have

$$\begin{aligned}\varphi(\gamma_0)-\varphi(\gamma_t) &\ge \Big(f(\gamma_1)+\frac{\mathsf{d}^2(\gamma_0,\gamma_1)}{2}\Big)-\Big(f(\gamma_1)+\frac{\mathsf{d}^2(\gamma_t,\gamma_1)}{2}\Big)\\ &= \frac{1-(1-t)^2}{2}\mathsf{d}^2(\gamma_0,\gamma_1) = \Big(\frac{2t-t^2}{2}\Big)\,\mathsf{d}^2(\gamma_0,\gamma_1).\end{aligned} \tag{10.7}$$

Since the speed of $\gamma$ is $\mathsf{d}(\gamma_0,\gamma_1)$ we have

$$\Big(\varphi(\gamma_0)-\varphi(\gamma_t)\Big)^2 \le \Big(\int_0^t |\nabla\varphi|_{w,\mathcal{G}}(\gamma_s)\mathsf{d}(\gamma_1,\gamma_0)\,\mathrm{d}s\Big)^2 \le t\mathsf{d}^2(\gamma_1,\gamma_0)\int_0^t g^2(\gamma_s)\,\mathrm{d}s.$$

Set now $Z_M := \{\gamma \in \mathrm{Geo}(X) : V(\gamma_0) \le M,\ \mathsf{d}(\gamma_0,\gamma_1) \le M\}$. Dividing by $t^2\mathsf{d}^2(\gamma_1,\gamma_0) = \mathsf{d}^2(\gamma_t,\gamma_0)$ and integrating on $Z_M$ with respect to $\boldsymbol{\pi}$, we obtain

$$\frac{1}{t}\int_0^t\int_{Z_M} g^2(\gamma_s)\,\mathrm{d}\boldsymbol{\pi}(\gamma)\mathrm{d}s \ge \int_{Z_M}\Big(\frac{\varphi(\gamma_0)-\varphi(\gamma_t)}{\mathsf{d}(\gamma_0,\gamma_t)}\Big)^2\mathrm{d}\boldsymbol{\pi} \ge \frac{(2-t)^2}{4}\int_{Z_M}\mathsf{d}^2\,\mathrm{d}\boldsymbol{\pi}.$$

Setting $L = \mathrm{Lip}(V)$, for all $\delta > Lt$ our choice of $Z_M$ gives for $\mu_s = (\mathrm{e}_s)_\sharp\boldsymbol{\pi}$

$$\frac{1}{t}\int_0^t\int_{\{V\le M+\delta\}} g^2\,\mathrm{d}\mu_s\mathrm{d}s \ge \int_{Z_M}\Big(\frac{\varphi(\gamma_0)-\varphi(\gamma_t)}{\mathsf{d}(\gamma_0,\gamma_t)}\Big)^2\mathrm{d}\boldsymbol{\pi} \ge \frac{(2-t)^2}{4}\int_{Z_M}\mathsf{d}^2\,\mathrm{d}\boldsymbol{\pi}. \tag{10.8}$$

In order to pass to the limit as $t \downarrow 0$, we observe that (10.5) gives

$$\forall\, N > 0: \quad \int_{\{V\le N\}} f\,\mathrm{d}\mu_s \to \int_{\{V\le N\}} f\,\mathrm{d}\mu \quad \text{as } s\downarrow 0 \text{ for all } f\chi_{\{V\le N\}} \in L^1(X,\mathfrak{m}). \tag{10.9}$$

Indeed for every bounded, Borel, and $\mathsf{d}$-Lipschitz function $h : X \to \mathbb{R}$ we have

$$\Big|\int_X h\,\mathrm{d}\mu_s - \int_X h\,\mathrm{d}\mu\Big| \le \int |h(\gamma_s)-h(\gamma_0)|\,\mathrm{d}\boldsymbol{\pi}(\gamma) \le s\,\mathrm{Lip}(h)\,W_2(\mu,\nu). \tag{10.10}$$

On the other hand, arguing exactly as in the proof of Proposition 4.1, if $f\chi_{\{V\le N\}} \in L^1(X,\mathfrak{m})$ we can find a sequence $(h_n) \subset L^1(X,\mathfrak{m})$ of bounded, Borel, $\mathsf{d}$-Lipschitz functions strongly converging to $f\chi_{\{V\le N\}}$ in $L^1(X,\mathfrak{m})$. Upon multiplying $h_n$ by the $\mathsf{d}$-Lipschitz function $k_N(x) := \min\{1,(N+1-V(x))^+\}$, it is not restrictive to assume that $h_n$ identically vanishes on $\{V > N+1\}$. If $\|\rho_s\|_{L^\infty(\{V\le N+1\},\mathfrak{m})} \le C$ for sufficiently small $s$ according to (10.5), we thus have

$$\Big|\int_{\{V\le N\}} f\,\mathrm{d}\mu_s - \int_{\{V\le N\}} f\,\mathrm{d}\mu\Big| \le 2C\|f\chi_{\{V\le N\}}-h_n\|_{L^1(X,\mathfrak{m})} + \Big|\int_X h_n\,\mathrm{d}\mu_s - \int_X h_n\,\mathrm{d}\mu\Big|.$$

Taking first the lim sup as $s \downarrow 0$ thanks to (10.10) and then the limit as $n \to \infty$ we obtain (10.9).

By (10.2) the functions $g^2\chi_{\{V\le M+\delta\}}$ belong to $L^1(X,\mathfrak{m})$. Therefore, using (10.9) with $f := g$ and $N := M+\delta$, passing to the limit in (10.8) first as $t \downarrow 0$ and then as $\delta \downarrow 0$ gives

$$\int_{\{V\le M\}} g^2\,\mathrm{d}\mu \ge \limsup_{t\downarrow 0}\int_{Z_M}\Big(\frac{\varphi(\gamma_0)-\varphi(\gamma_t)}{\mathsf{d}(\gamma_0,\gamma_t)}\Big)^2\mathrm{d}\boldsymbol{\pi}(\gamma) \ge \int_{Z_M}\mathsf{d}^2(\gamma_1,\gamma_0)\,\mathrm{d}\boldsymbol{\pi}(\gamma). \tag{10.11}$$

Letting $M \to \infty$ this completes the proof of (10.6). $\square$

The identification (10.6) could be compared to Theorem 6.1 of [10], where Cheeger identified the relaxed gradient of Lipschitz functions with the local Lipschitz constant, assuming that the metric measure space $(X, \mathsf{d}, \mathfrak{m})$ is doubling and satisfies the Poincaré inequality. Without doubling conditions, but assuming the validity of good interpolation properties, we are able to obtain an analogous identification at least in a suitable class of $c$-concave functions.

For finite reference measures $\mathfrak{m}$ and densities $\rho$ uniformly bounded from below, we can also prove a more precise convergence result for the difference quotients of $\varphi$.

**Theorem 10.4** *Let $\mu = \rho\mathfrak{m} \in \mathscr{P}(X)$ be satisfying $\rho \geq c > 0$ $\mathfrak{m}$-a.e. in $X$ and let $\varphi$, $\boldsymbol{\pi}$ as in Theorem 10.3. Then*

$$\lim_{t\downarrow 0} \frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{\mathsf{d}(\gamma_0, \gamma_t)} = |\nabla^+\varphi|(\gamma_0) \qquad \text{in } L^2(\mathrm{Geo}(X), \boldsymbol{\pi}). \tag{10.12}$$

*Proof.* The lower bound on $\rho$ yields in this case $|\nabla^+\varphi| \in L^2(X, \mathfrak{m})$, hence one can argue as in the proof of Theorem 10.3, this time integrating on the whole of $\mathrm{Geo}(X)$, to get

$$\int_X |\nabla\varphi|^2_{w,\mathcal{G}}\, \mathrm{d}\mu \geq \limsup_{t\downarrow 0} \int_{\mathrm{Geo}(X)} \Big(\frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{\mathsf{d}(\gamma_0, \gamma_t)}\Big)^2 \mathrm{d}\boldsymbol{\pi}(\gamma) \geq \int_{\mathrm{Geo}(X)} \mathsf{d}^2(\gamma_1, \gamma_0)\, \mathrm{d}\boldsymbol{\pi}(\gamma)$$

in place of (10.11). Since (10.6) yields that all inequalities are equalities, and (10.7) yields

$$\liminf_{t\downarrow 0} \frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{\mathsf{d}(\gamma_0, \gamma_t)} \geq |\nabla^+\varphi|(\gamma_0) \qquad \text{for } \boldsymbol{\pi}\text{-a.e. } \gamma \in \mathrm{Geo}(X)$$

we can use Lemma 10.5 below to obtain (10.12). $\square$

**Lemma 10.5** *Let $\sigma$ be a positive, finite measure in a measurable space $(Z, \mathcal{F})$ and let $f_n$, $f \in L^2(Z, \mathcal{F}, \sigma)$ be satisfying*

$$\limsup_{n\to\infty} \int_Z f_n^2\, \mathrm{d}\sigma \leq \int_Z f^2\, \mathrm{d}\sigma < \infty \tag{10.13}$$

*and $\liminf_n f_n \geq f \geq 0$ $\sigma$-a.e. in $Z$. Then $f_n \to f$ in $L^2(Z, \mathcal{F}, \sigma)$.*

*Proof.* If $f_n \geq 0$, it suffices to expand the square $(f_n - f)^2$ and to apply Fatou's lemma. In the general case we obtain first the convergence of $f_n^+$ to $f$ in $L^2$, and then use (10.13) once more to obtain that $f_n^- \to 0$ in $L^2$. $\square$

Example 3.10 shows that the localization technique provided by the potential $V$ and (4.2) plays an important role: indeed, in the same situation of that example, let $\mathfrak{m} = \delta_0 + x^{-1}\mathscr{L}^1$ be a $\sigma$-finite measure in $X = [0,1]$, so that $\mathrm{d}\mu_t/\mathrm{d}\mathfrak{m}(x) \leq 1$ for any $t$, $x$. In this case the conclusions of the metric Brenier theorem are not valid, since $\mu_0$ is concentrated at 0 and $d(0, y)$ takes all values in $[0,1]$. Notice that $\mathfrak{m}$ is not locally finite and the class of continuous $\mathfrak{m}$-integrable functions is not dense in $L^1([0,1]; \mathfrak{m})$ (any continuous and integrable function must vanish at $x = 0$).

**Remark 10.6** We remark that in the generality we are working with, it is not possible to prove uniqueness of the optimal plan, and the fact that it is induced by a map, not even if we add a $CD(K, \infty)$ assumption. To see why, consider the following example. Let $X = \mathbb{R}^2$ with the $L^\infty$ distance and the Lebesgue measure. Let $\mu_0 := \chi_{[0,1]^2}\mathscr{L}^2$ and $\mu_1 := \chi_{[3,4]\times[0,1]}\mathscr{L}^2$. Then, using standard tools of optimal transport theory, one can see that the only information that one can get by analyzing the $c$-superdifferential of an optimal Kantorovich potential is, shortly said, that any vertical line $\{t\}\times[0,1]$ must be sent onto the vertical line $\{t+3\}\times[0,1]$. The constraint on the marginals gives that this transport of $\{t\}\times[0,1]$ on $\{t+3\}\times[0,1]$ must send the 1-dimensional Hausdorff measure on $\{t\}\times[0,1]$ in the 1-dimensional Hausdorff measure on $\{t+3\}\times[0,1]$ for a.e. $t$. Apart from this, there is no other constraint, so we see that there are quite many optimal plans and that most of them are not induced by a map. Yet, the metric Brenier theorem is true, as the distance each point travels is independent of the optimal plan chosen (and equal to 3 for $\mu_0$-a.e. $x$). ■

**Remark 10.7** Theorem 10.3 and Theorem 10.4, with the same proof, hold if we replace condition (10.5) with the weaker one (at least in finite measure spaces)

$$\limsup_{s\downarrow 0}\int_X \rho_s\log\rho_s\,\mathrm{d}\mathfrak{m}<\infty,$$

but adding the condition $|\nabla^+\varphi|\in L^\infty(\{V\le M\},\mathfrak{m})$ for all $M\ge 0$. This, however, requires a slight modification of the class of test plans, and consequently of the concept of minimal weak upper gradient, requiring that the marginals have only bounded entropy instead of bounded density. This approach, that we do not pursue here, might be particularly appropriate when $\mathsf{d}$ is a bounded distance (e.g. in compact metric spaces), because in this situation Kantorovich potentials are Lipschitz. ■

# References


[1] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Adv. Differential Equations, 10 (2005), pp. 309–360.

[2] L. Ambrosio and N. Gigli, *User's guide to optimal transport theory*, to appear.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, arXiv:1109.0222v1.

[4] ——, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[5] L. Ambrosio, G. Savaré, and L. Zambotti, *Existence and stability for Fokker-Planck equations with log-concave reference measure.*, Probab. Theory Relat. Fields, 145 (2009), pp. 517–564.

[6] V. I. Bogachev, *Gaussian measures*, vol. 62 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 1998.

[7] V. I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007.

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

[9] H. Brézis and W. A. Strauss, *Semi-linear second-order elliptic equations in $l^1$*, J. Math. Soc. Japan, 25 (1973), pp. 565–590.

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

[11] M. G. Crandall and L. Tartar, *Some relations between nonexpansive and order preserving mappings*, Proc. Amer. Math. Soc., 78 (1980), pp. 385–390.

[12] C. Dellacherie and P.-A. Meyer, *Probabilities and potential*, vol. 29 of North-Holland Mathematics Studies, North-Holland Publishing Co., Amsterdam, 1978.

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

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

[15] D. Feyel and A. S. Üstünel, *Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space*, Probab. Theory Related Fields, 128 (2004), pp. 347–385.

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

[17] N. Gigli, K. Kuwada, and S. Ohta, *Heat flow on alexandrov spaces*, submitted, (2010).

[18] A. Grigor′yan, *Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds*, Bull. Amer. Math. Soc. (N.S.), 36 (1999), pp. 135–249.

[19] P. Hajłasz and P. Koskela, *Sobolev met Poincaré*, Mem. Amer. Math. Soc., 145 (2000), pp. x+101.

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

[21] J. Heinonen and P. Koskela, *Quasiconformal maps in metric spaces with controlled geometry*, Acta Math., 181 (1998), pp. 1–61.

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

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

[24] S. Lisini, paper in preparation.

[25] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, ArXiv Mathematics e-prints, (2004).

[26] J. Lott and C. Villani, *Hamilton-Jacobi semigroup on length spaces and applications*, J. Math. Pures Appl. (9), 88 (2007), pp. 219–229.

[27] ——, *Weak curvature conditions and functional inequalities*, J. Funct. Anal., 245 (2007), pp. 311–333.

[28] ——, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2), 169 (2009), pp. 903–991.

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

[30] ——, *Non-contraction of heat flow on Minkowski spaces.* Preprint, (2010).

[31] F. Otto, *Doubly degenerate diffusion equations as steepest descent*, Manuscript, (1996).

[32] J. Peetre, *Interpolation of Lipschitz operators and metric spaces*, Mathematica (Cluj), 12(35) (1970), pp. 325–334.

[33] L. Schwartz, *Radon measures on arbitrary topological spaces and cylindrical measures*, Published for the Tata Institute of Fundamental Research, Bombay by Oxford University Press, London, 1973. Tata Institute of Fundamental Research Studies in Mathematics, No. 6.

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

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

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

[37] L. Tartar, *Interpolation non linéaire et régularité*, J. Funct. Anal., (1972).

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