Metric measure spaces with Riemannian Ricci curvature bounded from below

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

July 30, 2011

Abstract

In this paper we introduce a synthetic notion of Riemannian Ricci bounds from below for metric measure spaces $(X, \mathsf{d}, \mathfrak{m})$ which is stable under measured Gromov-Hausdorff convergence and rules out Finsler geometries. It can be given in terms of an enforcement of the Lott, Sturm and Villani geodesic convexity condition for the entropy coupled with the linearity of the heat flow. Besides stability, it enjoys the same tensorization, global-to-local and local-to-global properties. In these spaces, that we call $RCD(K, \infty)$ spaces, we prove that the heat flow (which can be equivalently characterized either as the flow associated to the Dirichlet form, or as the Wasserstein gradient flow of the entropy) satisfies Wasserstein contraction estimates and several regularity properties, in particular Bakry-Emery estimates and the $L^\infty - \mathrm{Lip}$ Feller regularization. We also prove that the distance induced by the Dirichlet form coincides with d, that the local energy measure has density given by the square of Cheeger's relaxed slope and, as a consequence, that the underlying Brownian motion has continuous paths. All these results are obtained independently of Poincaré and doubling assumptions on the metric measure structure and therefore apply also to spaces which are not locally compact, as the infinite-dimensional ones.

Contents

---

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

[†] Nice University. email: nicola.gigli@unice.fr

[‡] Università di Pavia. email: giuseppe.savare@unipv.it

1 Introduction

The problem of finding synthetic notions of Ricci curvature bounds from below has been a central object of investigation in the last few years. What became clear over time (see in particular [15] and [9, Appendix 2]), is that the correct class of spaces where such a synthetic notion can be given, is that of metric measure spaces, i.e. metric spaces equipped with a reference measure which one might think of as volume measure. The goal is then to find a notion consistent with the smooth Riemannian case, which is sufficiently weak to be stable under measured Gromov-Hausdorff limits. The problem of having stability is of course in competition with the necessity to find a condition as restrictive as possible, to describe efficiently the closure of the class of Riemannian manifolds with Ricci curvature uniformly bounded from below.

In their seminal papers Lott-Villani [27] and Sturm [39] independently attacked these questions with tools based on the theory of optimal transportation, devising stable and consistent notions. In these papers, a metric measure space $(X, \mathsf{d}, \mathfrak{m})$ is said to have Ricci curvature bounded from below by $K \in \mathbb{R}$ (in short: it is a $CD(K, \infty)$ space) if the relative entropy functional

$$\mathrm{Ent}_{\mathfrak{m}}(\mu) := \int \rho \log \rho \, \mathrm{d}\mathfrak{m} \qquad \text{with } \mu = \rho\mathfrak{m}$$

is K geodesically convex on the Wasserstein space $(\mathscr{P}_2(X), W_2)$.

Also, in [27], [40] a synthetic notion $CD(K, N)$ of having Ricci curvature bounded from below by K and dimension bounded above by N was given (in [27] only the case $CD(0, N)$ was considered, for $N < \infty$), and a number of geometric consequences of these notions, like Brunn-Minkowski and Bishop-Gromov inequalities, have been derived. In [26] it was also proved that, at least under the nonbranching assumption, the $CD(K, N)$ condition implies also the Poincaré inequaliy, see also [33] for some recent progress in this direction.

An interesting fact, proved by Cordero-Erasquin, Sturm and Villani (see the conclusions of [43]), is that $\mathbb{R}^d$ equipped with any norm and with the Lebesgue measure, is a $CD(0, N)$ space. More generally, Ohta showed in [29] that any smooth compact Finsler manifold is a $CD(K, N)$ space for appropriate finite K, N. However, a consequence of the analysis of tangent spaces done in [9], is that a Finsler manifold arises as limit of Riemannian manifolds with Ricci curvature uniformly bounded below and dimension uniformly bounded from above, if and only if it is Riemannian (the case of possibly unbounded dimension of the approximating sequence is covered by the stability of the heat flow proved in [18] in conjunction with the fact that the heat flow on a Finsler manifold is linear if and only if it is Riemannian [30]).

Therefore it is natural to look for a synthetic and stable notion of Ricci curvature bound which rules Finsler spaces out. This is the scope of this paper. What we do, roughly said, is to add to the $CD(K,\infty)$ condition the linearity of the heat flow, see below for the precise definition.

Before passing to the description of the results of this paper, we recall the main results of the "calculus" in the first paper of ours [2], needed for the development and the understanding of this one. The main goals of [2] have been the identification of two notions of gradient and two gradient flows.

The first notion of gradient, that we call *minimal relaxed gradient* and denote by $|\nabla f|_*$, is inspired by Cheeger's work [8]: it is the local quantity that provides integral representation to the functional $\mathsf{Ch}(f)$ given by

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

(here $|\nabla f_h|$ is the so-called local Lipschitz constant of f_h), so that $\mathsf{Ch}(f)=\frac{1}{2}\int|\nabla f|_*^2\,\mathrm{d}\mathfrak{m}$. The second notion of gradient, that we call *minimal weak upper gradient* and denote by $|\nabla f|_w$ is, instead, inspired by Shanmugalingam's work [36] and based on the validity of the upper gradient property

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

on "almost all" curves γ. We proved that the minimal weak upper gradient and the minimal relaxed gradient coincide. In addition, although our notions of null set of curves differs from [36] and the definition of Ch differs from [8], we prove, a posteriori, that the gradients coincide with those in [8], [36]. Since an approximation by Lipschitz functions is implicit in the formulation (1.1), this provides a density result of Lipschitz function in the weak Sobolev topology without any doubling and Poincaré assumption on $(X,\mathsf{d},\mathfrak{m})$. In the context of the present paper, where Ch will be a quadratic form even when extended to $\mathfrak{m}$-measurable functions using weak upper gradients, this approximation result yields the density of Lipschitz functions in the strong Sobolev topology.

The concept of minimal relaxed gradient can be used in connection with "vertical" variations of the form $\varepsilon\mapsto f+\varepsilon g$, which occur in the study of the $L^2(X,\mathfrak{m})$-gradient flow of Ch, whose semigroup we shall denote by H_t. On the other hand, the concept of minimal weak upper gradient is relevant in connection with "horizontal" variations of the form $t\mapsto f(\gamma_t)$, which play an important role when study the derivative of $\mathrm{Ent}_{\mathfrak{m}}$ along geodesics. For this reason their identification is crucial, as we will see in Section 4. Given this identification for granted, in the present paper most results will be presented and used at the level of minimal weak upper gradients, in order to unify the exposition.

Finally, in $CD(K,\infty)$ spaces we identified the L^2-gradient flow H_t of Ch (in the sense of the Hilbertian theory [6]) with the W_2-gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ in the Wasserstein space of probability measures $\mathscr{P}_2(X)$ (in the sense of De Giorgi's metric theory, see [3] and, at this level of generality, [18]), which we shall denote by $\mathscr{H}_t$. A byproduct of this identification is an equivalent description of the entropy dissipation rate along the flow, equal to $4\int|\nabla\sqrt{\mathsf{H}_t f}|_w^2\,\mathrm{d}\mathfrak{m}$ and to the square of the metric derivative of $\mu_t=\mathscr{H}_t(f\mathfrak{m})$ w.r.t. W_2.

All these results have been obtained in [2] under very mild assumptions on $\mathfrak{m}$, which include all measures such that $\mathrm{e}^{-c\,\mathsf{d}^2(x,x_0)}\mathfrak{m}$ is finite for some $c>0$ and $x_0\in X$. In this paper, in order to minimize the technicalities, we assume that $\mathfrak{m}$ is a probability measure with finite second moment. On the other hand, no local compactness assumption on (X,d) will be needed, so that infinite-dimensional spaces fit well into this theory.

Coming back to this paper, we say that a metric measure space $(X,\mathsf{d},\mathfrak{m})$ has *Riemannian Ricci* curvature bounded from below by $K\in\mathbb{R}$, and write that $(X,\mathsf{d},\mathfrak{m})$ is $RCD(K,\infty)$, if one of the following equivalent conditions hold:

(i) $(X,\mathsf{d},\mathfrak{m})$ is a strong $CD(K,\infty)$ space and the W_2-gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ is additive.

(ii) $(X,\mathsf{d},\mathfrak{m})$ is a strong $CD(K,\infty)$ space and Ch is a quadratic form in $L^2(X,\mathfrak{m})$, so that the L^2-heat flow of Ch is linear.

(iii) The gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ exists for all initial data μ with $\operatorname{supp}\mu\subset\operatorname{supp}\mathfrak{m}$ and satisfies the EVI_K condition.

The equivalence of these conditions is not at all obvious, and its proof is actually one of the main results of this paper.

Observe that in (i) and (ii) the $CD(K,\infty)$ is enforced on the one hand considering a stronger convexity condition (we describe this condition in the end of the introduction, being this aspect less relevant), on the other hand adding linearity of the heat flow. A remarkable fact is that this combination of properties can be encoded in a *single* one, namely the EVI_K property. This latter property can be expressed by saying that for all $\nu\in\mathscr{P}_2(X)$ with finite entropy the gradient flow $\mathscr{H}_t(\mu)$ starting from μ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{W_2^2(\mathscr{H}_t(\mu),\nu)}{2}+\frac{K}{2}W_2^2(\mathscr{H}_t(\mu),\nu)+\mathrm{Ent}_{\mathfrak{m}}(\mathscr{H}_t(\mu))\le\mathrm{Ent}_{\mathfrak{m}}(\nu)\qquad\text{for a.e. } t\in(0,\infty). \tag{1.2}$$

It is immediate to see that the $RCD(K,\infty)$ notion is consistent with the Riemannian case: indeed, uniqueness of geodesics in $(\mathscr{P}_2(M),W_2)$ between absolutely continuous measures and the consistency of the $CD(K,\infty)$ notion, going back to [11, 42], yield that manifolds are strong $CD(K,\infty)$ spaces (see below for the definition), and the fact that Ch is quadratic is directly encoded in the Riemannian metric tensor, yielding the linearity of the heat flow. On the other hand, the stability of $RCD(K,\infty)$ bounds with respect to the measured Gromov-Hausdorff convergence introduced by Sturm [39] is a consequence, not too difficult, of condition (iii) and the general stability properties of EVI_K flows (see also [34, 35] for a similar statement). We remark also that thanks to the results in [20, 29, 32, 44], compact and finite-dimensional spaces with Alexandrov curvature bounded from below are $RCD(K,\infty)$ spaces.

Besides this, we prove many additional properties of $RCD(K,\infty)$ spaces. Having (1.2) at our disposal at the level of measures, it is easy to obtain fundamental solutions, integral representation formulas, regularizing and contractivity properties of the heat flow, which exhibits a strong Feller regularization from $L^\infty(X,\mathfrak{m})$ to Lipschitz. Denoting by $W^{1,2}(X,\mathsf{d},\mathfrak{m})\subset L^2(X,\mathfrak{m})$ the finiteness domain of Ch, the identification of the L^2-gradient flow of Ch and of the W_2-gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ in conjunction with the K-contractivity of W_2 along the heat flow, yields, as in [19], the Bakry-Emery estimate

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

for all $f\in W^{1,2}(X,\mathsf{d},\mathfrak{m})$. As a consequence of this, we prove that functions f whose minimal weak upper gradient $|\nabla f|_w$ belongs to $L^\infty(X,\mathfrak{m})$ have a Lipschitz version $\tilde f$, with $\mathrm{Lip}(\tilde f)\le\||\nabla f|_w\|_\infty$.

In connection with the tensorization property, namely the stability of $RCD(K,\infty)$ metric measure spaces with respect to product (with squared product distance given by the sum of the squares of the distances in the base spaces), we are able to achieve it assuming that the base spaces are nonbranching. This limitation is due to the fact that also for the tensorization of $CD(K,\infty)$ spaces the nonbranching assumption has not been ruled out so far (see [39, Theorem 4.17]). On the other hand, we are able to show that the linearity of the heat flow tensorizes, when coupled just with the strong $CD(K,\infty)$ condition. The nonbranching assumption on the base spaces could be avoided with a proof of the tensorization property directly at the level of EVI_K, but we did not succeed so far in tensorizing the EVI_K condition.

Since Ch is a quadratic form for $RCD(K,\infty)$ spaces, it is tempting to take the point of view of Dirichlet forms and to describe the objects appearing in Fukushima's theory [16] of Dirichlet forms. In this direction, see also the recent work [22] and Remark 6.7. Independently of any curvature bound we show that, whenever Ch is quadratic, a Leibnitz formula holds and there exists a "local" bilinear map $(f,g)\mapsto\nabla f\cdot\nabla g$ from $[W^{1,2}(X,\mathsf{d},\mathfrak{m})]^2$ to $L^1(X,\mathfrak{m})$, that provides an integral representation to the Dirichlet form $\mathcal{E}(u,v)$ associated to Ch. This allows us to show that the local energy measure $[u]$ of Fukushima's theory coincides precisely with $|\nabla u|_w^2\mathfrak{m}$. If the space is $RCD(K,\infty)$ then the intrinsic distance $\mathsf{d}_{\mathcal{E}}$, associated to the Dirichlet form by duality with functions u satisfying $[u]\leq\mathfrak{m}$ is precisely d. The theory of Dirichlet forms can also be applied to obtain the existence of a continuous *Brownian motion* in $RCD(K,\infty)$ spaces, i.e. a Markov process with continuous sample paths and transition probabilities given by $\mathscr{H}_t(\delta_x)$.

Besides the extension to more general classes of reference measures $\mathfrak{m}$, we believe that this paper opens the door to many potential developments: among them we would like to mention the dimensional theory, namely finding appropriate "Riemannian" versions of the $CD(K,N)$ condition, and the study of the tangent space. In connection with the former question, since $CD(K,N)$ spaces are $CD(K,\infty)$, a first step could be analyzing them with the calculus tools we developed and to see the impact of the linearity of the heat flow and of the EVI_K condition stated at the level of $\mathrm{Ent}_{\mathfrak{m}}$. Concerning the latter question, it is pretty natural to expect $RCD(K,\infty)$ spaces to have Hilbertian tangent space for $\mathfrak{m}$-a.e. point. While the proof of this result in the genuine infinite dimensional case seems quite hard to get, if the space is doubling and supports a local Poincaré inequality, one can hope to refine Cheeger's analysis ([8, Section 11]) in order to achieve it.

The paper is organized as follows. In Section 2 we introduce our main notation and the preliminary results needed for the development of the paper. With the exception of Section 2.4, where we quote from [2] the basic results we already alluded to, namely the identification of weak gradients and relaxed gradients and the identification of L^2-gradient flow of Ch and W_2-gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$, the material is basically known. Particularly relevant for us will be the EVI_K formulation of gradient flows, discussed in Section 2.5.

In Section 3 we introduce a convexity condition, that we call *strong* $CD(K,\infty)$, intermediate between the $CD(K,\infty)$ condition, where convexity is required along some geodesic, and convexity along all geodesics. It can be stated by saying that, given any two measures, there is always an optimal geodesic plan $\boldsymbol{\pi}$ joining them such that the K-convexity holds along all the geodesics induced by weighted plans of the form $F\boldsymbol{\pi}$, where F is a bounded, non negative function with $\int F\,\mathrm{d}\boldsymbol{\pi}=1$. We also know from [13] that the EVI_K condition implies convexity along all geodesics supported in $\operatorname{supp}\mathfrak{m}$, and therefore the strong $CD(K,\infty)$ property.
This enforcement of the $CD(K,\infty)$ condition is needed to derive strong L^∞ bounds on the interpolating measures induced by the "good" geodesic plan. These "good" interpolating

measures provide large class of test plans and are used to show, in this framework, that the "metric Brenier" theorem [2, Theorem 10.3] holds. Roughly speaking, this theorem states that, when one transports in an optimal way μ to ν, the transportation distance $\mathsf{d}(\gamma_0,\gamma_1)$ depends $\boldsymbol{\pi}$-almost surely only on the initial point γ_0 (and in particular it is independent on the final point γ_1). Furthermore, the proof of this result provides the equality $\mathsf{d}(\gamma_0,\gamma_1)=|\nabla^+\varphi|(\gamma_0)=|\nabla\varphi|_w(\gamma_1)$ for $\boldsymbol{\pi}$-a.e. γ, where φ is any Kantorovich potential relative to (μ,ν). This equality will be crucial for us when proving optimal bounds for the derivative of $\mathrm{Ent}_{\mathfrak{m}}$ along geodesics.

In Section 4 we enter the core of the paper with two basic formulas, one for the derivative of the Wasserstein distance along the heat flow $(\rho_t\mathfrak{m})$ (Theorem 4.1), obviously important for a deeper understanding of (1.2), the other one for the derivative of the entropy (Theorem 4.9) along a geodesic $\mu_s=\rho_s\mathfrak{m}$. The proof of the first one uses the classical duality method and relates the derivative of $W_2^2(\rho_t\mathfrak{m},\nu)$ to the "vertical" derivative of the density ρ_t in the direction given by Kantorovich potential from $\rho_t\mathfrak{m}$ to ν. The second one involves much more the calculus tools we developed. The key idea is to start from the (classical) convexity inequality for the entropy, written in terms of the optimal geodesic plan $\boldsymbol{\pi}$ from $\rho_0\mathfrak{m}$ to $\rho_1\mathfrak{m}$

$$\mathrm{Ent}_{\mathfrak{m}}(\rho_s\mathfrak{m})-\mathrm{Ent}_{\mathfrak{m}}(\rho_0\mathfrak{m})\geq\int\log\rho_0(\rho_s-\rho_0)\,\mathrm{d}\mathfrak{m}=\int\big(\log\rho_0(\gamma_s)-\log\rho_0(\gamma_0)\big)\,\mathrm{d}\boldsymbol{\pi}(\gamma)\tag{1.3}$$

and then use the crucial Lemma 4.5, relating the "horizontal" derivatives appearing in (1.3) to the "vertical" ones. In section 4.3 the same lemma, applied to suitable plans generated by the heat flow, is the key to deduce a local quadratic structure from a globally quadratic Cheeger energy and to develop useful calculus tools, leading in particular to the identification of $|\nabla u|_w^2\mathfrak{m}$ with the energy measure $[u]$ provided by the general theory of Dirichlet forms.

Section 5 is devoted to the proof of the equivalence of the three conditions defining $RCD(K,\infty)$ spaces, while the final Section 6 treats all properties of $RCD(K,\infty)$ spaces we already discussed: representation, contraction, and regularizing properties of the heat flow, relations with the theory of Dirichlet forms and existence of the Brownian motion, stability, tensorization. We also discuss, in the last section, the so-called global-to-local and local-to-global implications. We prove that the first one always holds if the subset under consideration is convex, with positive $\mathfrak{m}$-measure and $\mathfrak{m}$-negligible boundary. We also prove a partial result in the other direction, from local to global, comparable to those available within the $CD(K,\infty)$ theory.

2 Preliminaries

2.1 Basic notation, metric and measure theoretic concepts

Unless otherwise stated, all metric spaces (Y,d_Y) we will be dealing with are complete and separable. Given a function $E:Y\to\mathbb{R}\cup\{\pm\infty\}$, we shall denote its domain $\{y:\ E(y)\in\mathbb{R}\}$ by $D(E)$. The slope (also called local Lipschitz constant) $|\nabla E|(y)$ of E at $y\in D(E)$ is defined by

$$|\nabla E|(y):=\limsup_{z\to y}\frac{|E(y)-E(z)|}{\mathsf{d}_Y(y,z)}.\tag{2.1}$$

By convention we put $|\nabla E|(y)=+\infty$ if $y\notin D(E)$ and $|\nabla E|(y)=0$ if $y\in D(E)$ is isolated.

We shall also need the one-sided counterparts of this concept, namely the *descending slope* (in the theory of gradient flows) and the *ascending slope* (in the theory of Kantorovich potentials). Their are defined at $y \in D(E)$ by

$$|\nabla^- E|(y) := \limsup_{z \to y} \frac{(E(z) - E(y))^-}{\mathsf{d}_Y(y, z)}, \quad |\nabla^+ E|(y) := \limsup_{z \to y} \frac{(E(z) - E(y))^+}{\mathsf{d}_Y(z, y)},$$

with the usual conventions if either y is isolated or it does not belong to $D(E)$.

We will denote by $\mathrm{C}([0,1]; Y)$ the space of continuous curves on (Y, d_Y); it is a complete and separable metric space when endowed with the sup norm. We also denote with $\mathrm{e}_t : \mathrm{C}([0,1]; Y) \to Y$, $t \in [0,1]$, the evaluation maps

$$\mathrm{e}_t(\gamma) := \gamma_t \qquad \forall \gamma \in \mathrm{C}([0,1]; Y).$$

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

$$\mathsf{d}_Y(\gamma_t, \gamma_s) \leq \int_s^t g(r)\,\mathrm{d}r \qquad \forall s,\, t \in [0,1],\ s \leq t, \tag{2.2}$$

for some $g \in L^1(0,1)$. If γ is absolutely continuous, the *metric speed* $|\dot\gamma| : [0,1] \to [0, \infty]$ is defined by

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

and it is possible to prove that the limit exists for a.e. t, that $|\dot\gamma| \in L^1(0,1)$, and that it is the minimal L^1 function (up to Lebesgue negligible sets) for which the bound (2.2) holds (see [3, Theorem 1.1.2] for the simple proof).

We shall denote by $\mathrm{AC}^2([0,1]; Y)$ the class of absolutely continuous curves with metric derivative in $L^2(0,1)$; it is easily seen to be a countable union of closed sets in $\mathrm{C}([0,1]; Y)$ and in particular a Borel subset.

A curve $\gamma \in \mathrm{C}([0,1]; Y)$ is called constant speed geodesic if $\mathsf{d}_Y(\gamma_t, \gamma_s) = |t - s| \mathsf{d}_Y(\gamma_0, \gamma_1)$ for all s, $t \in [0,1]$. We shall denote by $\mathrm{Geo}(Y)$ the space of constant speed geodesics, which is a closed (thus complete and separable) subset of $\mathrm{C}([0,1]; Y)$.

(Y, d_Y) is called a *length* space if for any y_0, $y_1 \in Y$ and $\varepsilon > 0$ there exists $\gamma \in \mathrm{AC}([0,1]; Y)$ such that

$$\gamma_0 = y_0,\ \gamma_1 = y_1 \quad \text{and} \quad \mathrm{Length}(\gamma) := \int_0^1 |\dot\gamma_t|\,\mathrm{d}t \leq \mathsf{d}_Y(y_0, y_1) + \varepsilon. \tag{2.3}$$

If for any $y_0, y_1 \in Y$ one can find γ satisfying (2.3) with $\varepsilon = 0$ (and thus, up to a reparameterization, $\gamma \in \mathrm{Geo}(Y)$), we say that (Y, d_Y) is a *geodesic* space. We also apply the above definitions to (even non closed) subsets $Z \subset Y$, always endowed with the distance d_Y induced by Y. It is worth noticing that if Z is a length space in Y, then $\overline{Z}$ is a length space in Y as well [7, Ex. 2.4.18].

We use standard measure theoretic notation, as $\mathrm{C}_b(X)$ for bounded continuous maps, $f_\sharp$ for the push forward operator induced by a Borel map f, namely $f_\sharp\mu(A) := \mu(f^{-1}(A))$, $\mu \llcorner A$ for the restriction operator, namely $\mu \llcorner A(B) = \mu(A \cap B)$.

2.2 Reminders on optimal transport

We assume that the reader is familiar with optimal transport, here we just recall the notation we are going to use in this paper and some potentially less known constructions. Standard references are [1, 3, 43] and occasionally we give precise references for the facts stated here.

Given a complete and separable space (X, d), $\mathscr{P}_2(X)$ is the set of Borel probability measures with finite second moment, which we endow with the Wasserstein distance W_2 defined by

$$W_2^2(\mu,\nu) := \min \int \mathsf{d}^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}(x,y), \tag{2.4}$$

the minimum being taken among the collection $\mathrm{Adm}(\mu,\nu)$ of all admissible plans (also called couplings) $\boldsymbol{\gamma}$ from μ to ν, i.e. all measures $\boldsymbol{\gamma} \in \mathscr{P}(X \times X)$ such that $\pi^1_\sharp \boldsymbol{\gamma} = \mu$, $\pi^2_\sharp \boldsymbol{\gamma} = \nu$. All the minimizers of (2.4) are called optimal plans and their collection (always non empty, since μ, $\nu \in \mathscr{P}_2(X)$) is denoted by $\mathrm{Opt}(\mu,\nu)$. The metric space $(\mathscr{P}_2(X), W_2)$ is complete and separable; it is also a length or a geodesic space if and only if X is, see for instance [1, Theorem 2.10, Remark 2.14].

Given a reference measure $\mathfrak{m}$, we shall also use the notation

$$\mathscr{P}_2(X,\mathfrak{m}) := \big\{\mu \in \mathscr{P}_2(X) :\ \operatorname{supp}\mu \subset \operatorname{supp}\mathfrak{m}\big\}.$$

The c-transform of a function $\psi : X \to \mathbb{R} \cup \{-\infty\}$, relative to the cost $c = \frac{1}{2}\mathsf{d}^2$, is defined by

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

Notice that still ψ^c takes its values in $\mathbb{R} \cup \{-\infty\}$, unless $\psi \equiv -\infty$. A function $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is said to be c-concave if $\varphi = \psi^c$ for some $\psi : X \to \mathbb{R} \cup \{-\infty\}$. A set $\Gamma \subset X \times X$ is c-cyclically monotone if

$$\sum_{i=1}^n c(x_i, y_i) \le \sum_{i=1}^n c(x_i, y_{\sigma(i)}) \qquad \forall n \ge 1,\ (x_i, y_i) \in \Gamma,\ \sigma \text{ permutation.}$$

Given μ, $\nu \in \mathscr{P}_2(X)$ there exists a c-cyclically monotone closed set Γ containing the support of all optimal plans $\boldsymbol{\gamma}$. In addition, there exists a (possibly non unique) c-concave function $\varphi \in L^1(X,\mu)$ such that $\varphi^c \in L^1(X,\nu)$ and $\varphi(x) + \varphi^c(y) = c(x,y)$ on Γ. Such functions are called Kantorovich potentials. We remark that the typical construction of φ (see for instance [3, Theorem 6.1.4]) gives that φ is locally Lipschitz in X if the target measure ν has bounded support. Conversely, it can be proved that $\boldsymbol{\gamma} \in \mathrm{Adm}(\mu,\nu)$ and $\operatorname{supp}\boldsymbol{\gamma}$ c-cyclically monotone imply that $\boldsymbol{\gamma}$ is an optimal plan.

It is not hard to check that (see for instance [2, Proposition 3.9])

$$|\nabla^+\varphi|(x) \le \mathsf{d}(x,y) \qquad \text{for } \boldsymbol{\gamma}\text{-a.e. } (x,y), \tag{2.5}$$

for any optimal plan $\boldsymbol{\gamma}$ and Kantorovich potential φ from μ to ν.

If μ and ν are joined by a geodesic in $(\mathscr{P}_2(X), W_2)$, the distance W_2 can be equivalently characterized by

$$W_2^2(\mu,\nu) = \min \int\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma), \tag{2.6}$$

among all measures $\boldsymbol{\pi} \in \mathscr{P}(\mathrm{C}([0,1];X))$ such that $(\mathrm{e}_0)_\sharp \boldsymbol{\pi} = \mu$, $(\mathrm{e}_1)_\sharp \boldsymbol{\pi} = \nu$, where the 2-action $\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t$ is taken by definition $+\infty$ is γ is not absolutely continuous. The set of minimizing plans $\boldsymbol{\pi}$ in (2.6) will be denoted by $\mathrm{GeoOpt}(\mu,\nu)$. It is not difficult to see that $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu,\nu)$ if and only if $\boldsymbol{\gamma} := (\mathrm{e}_0,\mathrm{e}_1)_\sharp \boldsymbol{\pi} \in \mathscr{P}(X\times X)$ is a minimizer in (2.4) and $\boldsymbol{\pi}$ is concentrated on $\mathrm{Geo}(X)$. Furthermore, a curve (μ_t) is a constant speed geodesic from μ_0 to μ_1 if and only if there exists $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0,\mu_1)$ such that

$$\mu_t = (\mathrm{e}_t)_\sharp \boldsymbol{\pi} \qquad \forall t\in[0,1],$$

see for instance [1, Theorem 2.10] and notice that the assumption that (X,d) is geodesic is never used in the proof of $(i) \Leftrightarrow (ii)$.

The linearity of the transport problem immediately yields that the squared Wasserstein distance $W_2^2(\cdot,\cdot)$ is jointly convex. This fact easily implies that if (μ_t^1), $(\mu_t^2) \subset \mathscr{P}_2(X)$ are two absolutely continuous curves, so is $t\mapsto \mu_t := (1-\lambda)\mu_t^1 + \lambda\mu_t^2$ for any $\lambda\in[0,1]$, with an explicit bound on its metric speed:

$$|\dot\mu_t|^2 \le (1-\lambda)|\dot\mu_t^1|^2 + \lambda|\dot\mu_t^2|^2 \qquad \text{for a.e. } t\in[0,1]. \tag{2.7}$$

Finally, we recall the definition of *push forward via a plan*, introduced in [39] (with a different notation) and further studied in [18], [2].

Definition 2.1 (Push forward via a plan) *Let $\boldsymbol{\gamma}\in\mathscr{P}(X\times Y)$. For $\mu\in\mathscr{P}(X)$ such that $\mu = \rho\,(\pi^X_\sharp\boldsymbol{\gamma}) \ll \pi^X_\sharp\boldsymbol{\gamma}$, the push forward $\boldsymbol{\gamma}_\sharp\mu\in\mathscr{P}(Y)$ of μ via $\boldsymbol{\gamma}$ is defined by*

$$\boldsymbol{\gamma}_\sharp\mu := \pi^Y_\sharp\big((\rho\circ\pi^X)\boldsymbol{\gamma}\big).$$

An equivalent representation of $\boldsymbol{\gamma}_\sharp\mu$ is

$$\boldsymbol{\gamma}_\sharp\mu = \eta\,\pi^Y_\sharp\boldsymbol{\gamma} \qquad \text{where} \qquad \eta(y) := \int_X \rho(x)\,\mathrm{d}\boldsymbol{\gamma}_y(x) \tag{2.8}$$

and $\{\boldsymbol{\gamma}_y\}_{y\in Y}\subset\mathscr{P}(X)$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. the projection on Y.

Defining $\boldsymbol{\gamma}^{-1} := (\pi^Y,\pi^X)_\sharp\boldsymbol{\gamma}\in\mathscr{P}(Y\times X)$, we can define in a symmetric way the map $\nu\mapsto\boldsymbol{\gamma}^{-1}_\sharp\nu\in\mathscr{P}(X)$ for any $\nu\ll\pi^Y_\sharp\boldsymbol{\gamma}^{-1} = \pi^Y_\sharp\boldsymbol{\gamma}$.

Notice that if $\boldsymbol{\gamma}$ is concentrated on the graph of a map $T: X\to Y$, it holds $\boldsymbol{\gamma}_\sharp\mu = T_\sharp\mu$ for any $\mu\ll\pi^X_\sharp\boldsymbol{\gamma}$, and that typically $\boldsymbol{\gamma}^{-1}_\sharp(\boldsymbol{\gamma}_\sharp\mu)\neq\mu$. We collect in the following proposition the basic properties of $\boldsymbol{\gamma}_\sharp$ in connection with the Wasserstein distance.

Proposition 2.2 *The following properties hold:*

(i) $\mu\le C\,\pi^1_\sharp\boldsymbol{\gamma}$ *for some* $C>0$ *implies* $\boldsymbol{\gamma}_\sharp\mu\le C\,\pi^2_\sharp\boldsymbol{\gamma}$.

(ii) Let μ, $\nu\in\mathscr{P}_2(X)$ and $\boldsymbol{\gamma}\in\mathrm{Opt}(\mu,\nu)$. Then for every $\tilde\mu\in\mathscr{P}_2(X)$ such that $\tilde\mu\ll\mu$ it holds

$$W_2^2(\tilde\mu,\boldsymbol{\gamma}_\sharp\tilde\mu) = \int \mathsf{d}^2(x,y)\frac{\mathrm{d}\tilde\mu}{\mathrm{d}\mu}(x)\,\mathrm{d}\boldsymbol{\gamma}(x,y) \tag{2.9}$$

and, in particular, $\boldsymbol{\gamma}_\sharp\tilde\mu\in\mathscr{P}_2(Y)$ and $\frac{\mathrm{d}\tilde\mu}{\mathrm{d}\mu}\circ\pi^1\boldsymbol{\gamma}\in\mathrm{Opt}(\tilde\mu,\boldsymbol{\gamma}_\sharp\tilde\mu)$ if any of the two terms is finite.

(iii) Let $\boldsymbol{\gamma} \in \mathscr{P}_2(X \times Y)$*,* $C > 0$ *and* $A_C := \{\mu \in \mathscr{P}_2(X) : \ \mu \leq C\, \pi^1_\sharp \boldsymbol{\gamma}\}$*. Then*

$$\mu \mapsto \boldsymbol{\gamma}_\sharp \mu \quad \textit{is uniformly continuous in } A_C \textit{ w.r.t. the } W_2 \textit{ distances.} \tag{2.10}$$

(iv) Let $\boldsymbol{\gamma} \in \mathscr{P}_2(X \times X)$ *and* $\mu \leq C\, \pi^1_\sharp \boldsymbol{\gamma}$ *for some constant* C*. Then*

$$W_2^2(\mu, \boldsymbol{\gamma}_\sharp \mu) \leq C \int \mathsf{d}^2(x, y)\, \mathrm{d}\boldsymbol{\gamma}(x, y). \tag{2.11}$$

Proof. (i) is obvious.
(ii) Since $\boldsymbol{\gamma}$ is optimal, $\operatorname{supp}\left(\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\mu} \circ \pi^1 \boldsymbol{\gamma}\right) \subset \operatorname{supp} \boldsymbol{\gamma}$ is c-cyclically monotone. Moreover $\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\mu} \circ \pi^1 \boldsymbol{\gamma}$ is an admissible plan from $\tilde{\mu}$ to $\boldsymbol{\gamma}_\sharp \tilde{\mu}$, with cost equal to the right hand side of (2.9). Hence, if the cost is finite, from the finiteness of $W_2(\tilde{\mu}, \boldsymbol{\gamma}_\sharp \tilde{\mu})$ we infer that $\boldsymbol{\gamma}_\sharp \tilde{\mu} \in \mathscr{P}_2(X)$, hence c-cyclical monotonicity implies optimality and equality in (2.9). The same argument works if we assume that $W_2(\tilde{\mu}, \boldsymbol{\gamma}_\sharp \tilde{\mu})$ is finite.
(iii) Since the singleton $\{\pi^1_\sharp \boldsymbol{\gamma}\}$ is both tight and 2-uniformly integrable, the same is true for the set A_C, which, being W_2-closed, is compact (see [3, Section 5.1] for the relevant definitions and simple proofs). Hence it is sufficient to prove the continuity of the map. Let $(\mu_n) \subset A_C$ be W_2-converging to $\mu \in A_C$ and let ρ_n, ρ be the respective densities w.r.t. $\pi^1_\sharp \boldsymbol{\gamma}$. Since (μ_n) converges to μ in duality with $\mathrm{C}_b(X)$ and since the densities are equibounded, we get that ρ_n converge to ρ weakly* in $L^\infty(X, \pi^1_\sharp \boldsymbol{\gamma})$. By (i) and the same argument just used we know that $(\boldsymbol{\gamma}_\sharp \mu_n) \subset \mathscr{P}_2(Y)$ is relatively compact w.r.t. the Wasserstein topology, hence to conclude it is sufficient to show that $(\boldsymbol{\gamma}_\sharp \mu_n)$ converges to $\boldsymbol{\gamma}_\sharp \mu$ in duality with $\mathrm{C}_b(Y)$. To this aim, fix $\varphi \in \mathrm{C}_b(Y)$ and notice that it holds

$$\int_Y \varphi(y)\, \mathrm{d}\boldsymbol{\gamma}_\sharp \mu_n(y) = \int_{X \times Y} \varphi(y) \rho_n(x)\, \mathrm{d}\boldsymbol{\gamma}(x, y) = \int_X \left(\int_Y \varphi(y)\, \mathrm{d}\boldsymbol{\gamma}_x(y) \right) \rho_n(x)\, \mathrm{d}\pi^1_\sharp \boldsymbol{\gamma}(x),$$

where $\{\boldsymbol{\gamma}_x\}$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. the projection on the first component. Since φ is bounded, so is the map $x \mapsto \int \varphi\, \mathrm{d}\boldsymbol{\gamma}_x$, and the claim follows.
(iv) Just notice that $\frac{\mathrm{d}\mu}{\mathrm{d}\pi^1_\sharp \boldsymbol{\gamma}} \circ \pi^1 \boldsymbol{\gamma} \in \mathrm{Adm}(\mu, \boldsymbol{\gamma}_\sharp \mu)$. □

The operation of push forward via a plan has also interesting properties in connection with the relative entropy functional $\mathrm{Ent}_{\mathfrak{m}}$. We recall that, given $\mathfrak{m} \in \mathscr{P}(X)$, the functional $\mathrm{Ent}_{\mathfrak{m}} : \mathscr{P}(X) \to [0, \infty]$ is defined by

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

Proposition 2.3 *For all* $\boldsymbol{\gamma} \in \mathscr{P}(X \times Y)$ *the following properties hold:*

(i) For any $\mathfrak{m}, \mu \ll \pi^X_\sharp \boldsymbol{\gamma}$ *it holds* $\mathrm{Ent}_{\boldsymbol{\gamma}_\sharp \mathfrak{m}}(\boldsymbol{\gamma}_\sharp \mu) \leq \mathrm{Ent}_{\mathfrak{m}}(\mu)$*.*

(ii) For any $\mathfrak{m} \ll \pi^X_\sharp \boldsymbol{\gamma}$*,* $C > 0$*, the map* $\mu \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu) - \mathrm{Ent}_{\boldsymbol{\gamma}_\sharp \mathfrak{m}}(\boldsymbol{\gamma}_\sharp \mu)$ *is convex in* $\{\mu \in \mathscr{P}(X) : \ \mu \leq C\mathfrak{m}\}$*.*

Proof. (i) We follow [1, Lemma 7.4] and [39, Lemma 4.19]. We can assume $\mu \ll \mathfrak{m}$, otherwise there is nothing to prove. Then it is immediate to check from the definition that $\boldsymbol{\gamma}_\sharp\mu \ll \boldsymbol{\gamma}_\sharp\mathfrak{m}$. Let $\mu = \rho\mathfrak{m}$, $\boldsymbol{\gamma}_\sharp\mu = \eta\boldsymbol{\gamma}_\sharp\mathfrak{m}$, and $u(z) := z\log z$. By disintegrating $\boldsymbol{\gamma}$ as in (2.8), we have

$$\begin{aligned}
\mathrm{Ent}_{\boldsymbol{\gamma}_\sharp\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu) &= \int u(\eta(y))\,\mathrm{d}\boldsymbol{\gamma}_\sharp\mathfrak{m}(y) = \int u\left(\int \rho(x)\,\mathrm{d}\boldsymbol{\gamma}_y(x)\right)\,\mathrm{d}\boldsymbol{\gamma}_\sharp\mathfrak{m}(y)\\
&\leq \int\int u(\rho(x))\,\mathrm{d}\boldsymbol{\gamma}_y(x)\,\mathrm{d}\boldsymbol{\gamma}_\sharp\mathfrak{m}(y) = \int u(\rho(x))\frac{\mathrm{d}\mathfrak{m}}{\mathrm{d}\pi^1_\sharp\boldsymbol{\gamma}}(x)\,\mathrm{d}\boldsymbol{\gamma}(x,y)\\
&= \int u(\rho(x))\,\mathrm{d}\mathfrak{m}(x) = \mathrm{Ent}_{\mathfrak{m}}(\mu).
\end{aligned}$$

(ii) This is proved in [2, Lemma 7.7] (see also [18, Proposition 11]). Notice that in [2] we worked under the assumption $X = Y$, but this makes no difference, since the as one can work on the disjoint union $X \sqcup Y$ endowed with a distance which extends those of X, Y. □

Remark 2.4 We remark that the property (*i*) above is true for any internal energy kind functional: as the proof shows, under the same assumptions on $\mathfrak{m}$, μ, $\boldsymbol{\gamma}$ it holds

$$U_{\boldsymbol{\gamma}_\sharp\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu) \leq U_{\mathfrak{m}}(\mu),$$

where $U_{\mathfrak{m}}(\mu)$ is given by $\int u(\rho)\,\mathrm{d}\mathfrak{m} + u'(\infty)\mu^s(X)$ for some convex continuous function $U : [0,\infty) \to \mathbb{R}\cup\{+\infty\}$ and $\mu = \rho\mathfrak{m} + \mu^s$, with $\mu^s \perp \mathfrak{m}$.

On the other hand, part (*ii*) does *not* always hold for these functionals: in [18] it has been shown that for $U(z) := \frac{z^\alpha}{\alpha-1}$ one has that

$$\mu \quad \mapsto \quad U_{\mathfrak{m}}(\mu) - U_{\boldsymbol{\gamma}_\sharp\mathfrak{m}}(\boldsymbol{\gamma}_\sharp\mu)$$

is convex on $\{\mu \in \mathscr{P}(X) :\ \mu \leq C\mathfrak{m}\}$ for any $C > 0$ if and only if $1 < \alpha \leq 2$. In particular, convexity does not hold for the functionals appearing in the definition of $CD(K,N)$ bounds. ■

2.3 Metric measure spaces and Sturm's distance $\mathbb{D}$

Throughout this paper we will always consider normalized metric measure spaces with finite variance, according to [39, §3.1]: in short, we will denote by $\mathbb{X}$ the set of (isomorphism classes of) metric measure spaces that we will consider, namely

$$\mathbb{X} := \Big\{(X,\mathsf{d},\mathfrak{m})\ :\ (X,\mathsf{d}) \text{ is complete, separable, and } \mathfrak{m} \in \mathscr{P}_2(X)\Big\}. \tag{2.12}$$

We say that the two metric measure spaces $(X, \mathsf{d}_X, \mathfrak{m}_X)$ and $(Y, \mathsf{d}_Y, \mathfrak{m}_Y)$ are *isomorphic* if there exists a

$$\text{bijective isometry} \quad f : \operatorname{supp}\mathfrak{m}_X \to \operatorname{supp}\mathfrak{m}_Y \quad \text{such that} \quad f_\sharp\mathfrak{m}_X = \mathfrak{m}_Y. \tag{2.13}$$

We say that $(X,\mathsf{d},\mathfrak{m})$ is *length* or *geodesic* if $(\operatorname{supp}\mathfrak{m},\mathsf{d})$ is so, and these notions are invariant in the isomorphism class.

Notice that $(X,\mathsf{d},\mathfrak{m})$ is always isomorphic to $(\operatorname{supp}\mathfrak{m},\mathsf{d},\mathfrak{m})$, so that it will often be not restrictive to assume the non-degeneracy condition $\operatorname{supp}\mathfrak{m} = X$.

In this section we recall the definition of the distance $\mathbb{D}$ between metric measure spaces, introduced by Sturm in [39], and its basic properties.

Definition 2.5 (Coupling between metric measure spaces) *Given two metric measure spaces* $(X, \mathsf{d}_X, \mathfrak{m}_X)$*,* $(Y, \mathsf{d}_Y, \mathfrak{m}_Y)$*, we consider the product space* $(X \times Y, \mathsf{d}_{XY})$*, where* d_{XY} *is the distance defined by*

$$\mathsf{d}_{XY}\big((x_1, y_1), (x_2, y_2)\big) := \sqrt{\mathsf{d}_X^2(x_1, x_2) + \mathsf{d}_Y^2(y_1, y_2)},$$

We say that a pair $(\mathbf{d}, \boldsymbol{\gamma})$ *is an admissible coupling between* $(X, \mathsf{d}_X, \mathfrak{m}_X)$, $(Y, \mathsf{d}_Y, \mathfrak{m}_Y)$*, and we write* $(\mathbf{d}, \boldsymbol{\gamma}) \in \mathrm{Adm}((\mathsf{d}_X, \mathfrak{m}_X), (\mathsf{d}_Y, \mathfrak{m}_Y))$*, if:*

(a) $\mathbf{d}$ *is a pseudo distance on* $X \sqcup Y$ *(i.e. points at 0* $\mathbf{d}$*-distance are not necessarily equal) which coincides with* d_X *(resp.* d_Y*) when restricted to* $X \times X$ *(resp.* $Y \times Y$*).*

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

It is not hard to see that the set of admissible couplings is always non empty. Notice that the restriction of $\mathbf{d}$ to $X \times Y$ is Lipschitz continuous and therefore Borel (with respect to the product topology), as a simple application of the triangle inequality.

The *cost* $C(\mathbf{d}, \boldsymbol{\gamma})$ of a coupling is given by

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

In analogy with the definition of W_2, the distance $\mathbb{D}\big((X, \mathsf{d}_X, \mathfrak{m}_X), (Y, \mathsf{d}_Y, \mathfrak{m}_Y)\big)$ is then defined as

$$\mathbb{D}^2\big((X, \mathsf{d}_X, \mathfrak{m}_X), (Y, \mathsf{d}_Y, \mathfrak{m}_Y)\big) := \inf C(\mathbf{d}, \boldsymbol{\gamma}), \tag{2.14}$$

the infimum being taken among all couplings $(\mathbf{d}, \boldsymbol{\gamma})$ of $(X, \mathsf{d}_X, \mathfrak{m}_X)$ and $(Y, \mathsf{d}_Y, \mathfrak{m}_Y)$. Since one can use the isometries in (2.13) to transfer couplings between two spaces to couplings between isomorphic spaces, a trivial consequence of the definition is that $\mathbb{D}$ actually depends only on the isomorphism class. In the next proposition we collect the main properties of $\mathbb{D}$, see [39, Section 3.1].

Proposition 2.6 (Properties of $\mathbb{D}$) *The infimum in* (2.14) *is attained and a minimizing coupling will be called optimal. Also,* $\mathbb{D}$ *is a distance on* $\mathbb{X}$*, and in particular* $\mathbb{D}$ *vanishes only on pairs of isomorphic metric measure spaces.*
Finally, $(\mathbb{X}, \mathbb{D})$ *is a complete, separable and length metric space.*

Also, it can be shown [39, Lemma 3.7] that $\mathbb{D}$ metrizes the measured Gromov-Hausdorff convergence, when restricted to compact metric spaces with controlled diameter. We also remark that, in line with what happens with the Gromov-Hausdorff distance, a $\mathbb{D}$-convergent sequence of metric measure spaces can be embedded into a common metric space: in this case the possibility to work in spaces where $\operatorname{supp} \mathfrak{m}$ is not equal to the whole space X turns out to be useful.

Proposition 2.7 *Let* $(X_n, \mathsf{d}_n, \mathfrak{m}_n) \in \mathbb{X}$*,* $n \in \mathbb{N}$*, and* $(X, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$*. Then the following two properties are equivalent.*

(i) $(X_n, \mathsf{d}_n, \mathfrak{m}_n) \overset{\mathbb{D}}{\to} (X, \mathsf{d}, \mathfrak{m})$ *as* $n \to \infty$*.*

(ii) There exist a complete and separable metric space (Y, d_Y) *and isometries* $f_n : \operatorname{supp} \mathfrak{m}_n \to Y$*,* $n \in \mathbb{N}$*,* $f : \operatorname{supp} \mathfrak{m} \to Y$*, such that* $W_2((f_n)_\sharp \mathfrak{m}_n, f_\sharp \mathfrak{m}) \to 0$ *as* $n \to \infty$*.*

Proof. (*i*) $\Rightarrow$ (*ii*). Let $(\mathbf{d}_n, \boldsymbol{\gamma}_n)$ be optimal couplings for $(X, \mathsf{d}, \mathfrak{m})$, $(X_n, \mathsf{d}_n, \mathfrak{m}_n)$, $n \in \mathbb{N}$. Define $Y := \Big(\bigsqcup_i \operatorname{supp} \mathfrak{m}_i\Big) \sqcup \operatorname{supp} \mathfrak{m}$ and the pseudo distance d_Y on Y by

$$\mathsf{d}_Y(x, x') := \begin{cases} \mathbf{d}_n(x, x'), & \text{if } x, x' \in \operatorname{supp} \mathfrak{m} \sqcup \operatorname{supp} \mathfrak{m}_n, \\ \inf\limits_{x'' \in X} \mathbf{d}_n(x, x'') + \mathbf{d}_m(x'', x') & \text{if } x \in \operatorname{supp} \mathfrak{m}_n,\ x' \in \operatorname{supp} \mathfrak{m}_m. \end{cases}$$

By construction the quotient metric space (Y, d_Y) induced by the equivalence relation $x \sim y \Leftrightarrow \mathsf{d}_Y(x, y) = 0$ is separable. Possibly replacing it by its abstract completion we can also assume that it is complete. Denoting by f_n, f the isometric embeddings of X_n, X into Y,

$$\mathbb{D}\big((X, \mathsf{d}, \mathfrak{m}), (X_n, \mathsf{d}_n, \mathfrak{m}_n)\big) = \sqrt{\int_{X \sqcup X_n} \mathbf{d}_n^2(x, y)\, \mathrm{d}\boldsymbol{\gamma}_n(x, y)} \geq W_2((f_n)_\sharp \mathfrak{m}_n, f_\sharp \mathfrak{m}),$$

so the conclusion follows.
(*ii*) $\Rightarrow$ (*i*). Straightforward. $\square$

2.4 Calculus and heat flow in metric measure spaces

2.4.1 Weak upper gradients and gradient flow of Cheeger's energy

Here we recall the definition and basic properties of weak upper gradients of real functions in the metric measure space $(X, \mathsf{d}, \mathfrak{m})$. All the concepts and statements that we consider here have been introduced and proven in [2], see § 5. In particular, here we shall consider measures concentrated in $\mathrm{AC}^2([0,1]; X)$ (see § 2.1).

Definition 2.8 (Test plans and negligible collection of curves) *We say that* $\boldsymbol{\pi} \in \mathscr{P}(\mathrm{AC}^2([0,1]; X))$ *is a test plan with bounded compression if there exists* $C = C(\boldsymbol{\pi}) > 0$ *such that*

$$(\mathrm{e}_t)_\sharp \boldsymbol{\pi} \leq C\, \mathfrak{m} \qquad \text{for every } t \in [0,1].$$

We will denote by $\mathcal{T}$ *the collection of all the test plans with bounded compression. We say that a Borel set* $A \subset \mathrm{AC}^2([0,1]; X)$ *is* $\mathcal{T}$-negligible *if* $\boldsymbol{\pi}(A) = 0$ *for any test plan* $\boldsymbol{\pi} \in \mathcal{T}$.

Since we will always deal with test plans in $\mathcal{T}$, we will often omit to mention exsplicitly $\mathcal{T}$ and the words "bounded compression", and we will refer to them simply as *test plans.*

A property which holds for every curve of $\mathrm{AC}^2([0,1]; X)$, except possibly for a subset of a negligible set, is said to hold for almost every curve.

Definition 2.9 (Functions which are Sobolev along almost all curves) *We say that* $f : X \to \overline{\mathbb{R}}$ *is Sobolev along almost all curves if, for a.e. curve* γ, $f \circ \gamma$ *coincides a.e. in* $[0,1]$ *and in* $\{0,1\}$ *with an absolutely continuous map* $f_\gamma : [0,1] \to \mathbb{R}$.

Notice that the choice of the trivial test plan $\boldsymbol{\pi} := \iota_\sharp \mathfrak{m}$, where $\iota : X \to \mathrm{AC}^2([0,1]; X)$ maps any point $x \in X$ to the constant curve $\gamma \equiv x$, yields that any Sobolev function along almost all curves is finite $\mathfrak{m}$-a.e. in X. In this class of functions we can define the notion of weak upper gradient and of minimal weak upper gradient.

Definition 2.10 (Weak upper gradients) *Given $f : X \to \overline{\mathbb{R}}$ Sobolev along a.e. curve, a $\mathfrak{m}$-measurable function $G : X \to [0,\infty]$ is a weak upper gradient of f if*

$$\left|\int_{\partial\gamma} f\right| \le \int_{\gamma} G \qquad \text{for a.e. curve } \gamma. \tag{2.15}$$

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

It turns out (see [2, Proposition 5.7, Definition 5.9]) that if and G_1, G_2 are weak upper gradients of f, then so is $\min\{G_1, G_2\}$. It follows that there exists a $\mathfrak{m}$-measurable function $|\nabla f|_w : X \to [0,\infty]$ weak upper gradient having the property that

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

for any other weak upper gradient G. Because of this $\mathfrak{m}$-a.e. minimality property, the function $|\nabla f|_w$ will be called the *minimal weak upper gradient* of f. Also, the property of being Sobolev along a.e. curve and the minimal weak upper gradient are invariant under modifications of f in $\mathfrak{m}$-negligible sets ([2, Proposition 5.8]). In addition, the minimal weak gradient is local in the following sense: if both f, g are Sobolev along a.e. curve then it holds

$$|\nabla f|_w = |\nabla g|_w \qquad \mathfrak{m}\text{-a.e. on the set } \{f = g\}. \tag{2.16}$$

Other useful and natural properties are: the restriction inequality [2, Remark 5.6]

$$|f(\gamma_s) - f(\gamma_s)| \le \int_t^s |\nabla f|_w(\gamma_r)|\dot\gamma_r|\,dr \quad \text{for a.e. } \gamma, \text{ for all } [s,t] \subset [0,1] \tag{2.17}$$

the chain rule [2, Proposition 5.14(b)])

$$\begin{aligned} &|\nabla(\phi\circ f)|_w = \phi'\circ f|\nabla f|_w && \mathfrak{m}\text{-a.e. in } X, \text{ if } \phi \text{ is Lipschitz and nondecreasing,}\\ &|\nabla(\phi\circ f)|_w \le |\phi'\circ f||\nabla f|_w && \mathfrak{m}\text{-a.e. in } X, \text{ if } \phi \text{ is Lipschitz,}\end{aligned} \tag{2.18}$$

and the weak Leibnitz rule

$$|\nabla(fg)|_w \le |f||\nabla g|_w + |g||\nabla f|_w \qquad \mathfrak{m}\text{-a.e. in } X. \tag{2.19}$$

The *Cheeger energy* is the functional defined in the class of Borel functions $f : X \to \overline{\mathbb{R}}$ by

$$\mathsf{Ch}(f) := \begin{cases} \dfrac{1}{2}\displaystyle\int |\nabla f|_w^2\,\mathrm{d}\mathfrak{m} & \text{if } f \text{ is Sobolev along a.e. curve,}\\ \\ +\infty & \text{otherwise.}\end{cases}$$

Using the stability properties of weak upper gradients under weak convergence ([2, Theorem 5.12]) it can be proved that Ch is convex and lower semicontinuous w.r.t. convergence in $\mathfrak{m}$-measure (in particular w.r.t. $\mathfrak{m}$-a.e. convergence). For the domain of Ch in $L^2(X,\mathfrak{m})$ we shall also use the traditional notation $W^{1,2}(X,\mathsf{d},\mathfrak{m})$, see [2, Remark 4.7]: it is a Banach space when endowed with the norm $\|f\|_{W^{1,2}}^2 := \|f\|_2^2 + 2\mathsf{Ch}(f)$. A nontrivial approximation theorem (see [2, Theorem 6.2]) shows that

$$\mathsf{Ch}(f) = \frac{1}{2}\inf\left\{\liminf_{h\to\infty}\int |\nabla f_h|^2\,\mathrm{d}\mathfrak{m} :\ f_h \in \mathrm{Lip}(X),\ \|f_h - f\|_2 \to 0\right\} \qquad \forall f \in L^2(X,\mathfrak{m}), \tag{2.20}$$

where $|\nabla f|$ is the local Lipschitz constant of f defined in (2.1).

Given $f \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$, we write $\partial^-\mathsf{Ch}(f) \subset L^2(X,\mathfrak{m})$ for the subdifferential at f of the restriction to $L^2(X,\mathfrak{m})$ of Cheeger's energy, namely $\xi \in \partial^-\mathsf{Ch}(f)$ iff

$$\mathsf{Ch}(g) \geq \mathsf{Ch}(f) + \int_X \xi(g-f)\,\mathrm{d}\mathfrak{m} \qquad \forall g \in L^2(X,\mathfrak{m}).$$

We say that $f \in L^2(X,\mathfrak{m})$ is in the domain of the $(\mathsf{d},\mathfrak{m})$-Laplacian, and write $f \in D(\Delta)$ (in [2] we used the notation $\Delta_{\mathsf{d},\mathfrak{m}}$ to emphasize the dependence on the metric measure structure), if $\partial^-\mathsf{Ch}(f) \neq \emptyset$. In this case we define $\Delta f \in L^2(X,\mathfrak{m})$ by $\Delta f := -v$, where v is the element of minimal $L^2(X,\mathfrak{m})$ norm in $\partial^-\mathsf{Ch}(f)$.

We remark that in this generality Ch is not necessarily a quadratic form, which is the same as to say that its restriction to $L^2(X,\mathfrak{m})$ is not a Dirichlet form. This means that the Laplacian, though 1-homogeneous [2, Remark 4.14], is not necessarily linear.

For the Laplacian we just defined the following rough integration by parts formula holds:

$$\left|\int g\Delta f\,\mathrm{d}\mathfrak{m}\right| \leq \int |\nabla g|_w|\nabla f|_w\,\mathrm{d}\mathfrak{m}, \tag{2.21}$$

for all f, $g \in L^2(X,\mathfrak{m})$ with $f \in D(\Delta)$ and $g \in D(\mathsf{Ch})$, see [2, Proposition 4.15].

The following result is a consequence of the by now classical theory of gradient flows of convex lower semicontinuous functionals on Hilbert spaces.

Theorem 2.11 (Gradient flow of Ch in $L^2(X,\mathfrak{m})$) *For all $f \in L^2(X,\mathfrak{m})$ there exists a unique locally absolutely continuous curve $(0,\infty) \ni t \mapsto f_t \in L^2(X,\mathfrak{m})$ such that $f_t \to f$ in $L^2(X,\mathfrak{m})$ as $t \downarrow 0$ and*

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

the derivative being understood in $L^2(X,\mathfrak{m})$. This curve is also locally Lipschitz, it satisfies $f_t \in D(\Delta)$ for any $t>0$ and

$$\frac{\mathrm{d}^+}{\mathrm{d}t}f_t = \Delta f_t \qquad \forall t>0.$$

Finally, $t \mapsto \mathsf{Ch}(f_t)$ is locally Lipschitz in $(0,\infty)$, infinitesimal at ∞ and, if $f \in D(\mathsf{Ch})$, continuous in 0. Its right derivative is given by $-\|\Delta f_t\|_2^2$ for every $t>0$.

Finally, we recall a property of the minimal weak gradient of Kantorovich potentials [2, Lemma 10.1]:

Proposition 2.12 *Let $(X,\mathsf{d},\mathfrak{m}) \in \mathbb{X}$, μ, $\nu \in \mathscr{P}_2(X)$ with $\mu \geq c\mathfrak{m}$ for some $c>0$ and let φ be a Kantorovich potential relative to (μ,ν). Then φ is finite and absolutely continuous (in particular, Sobolev) along a.e. curve and*

$$|\nabla\varphi|_w \leq |\nabla^+\varphi| \qquad \mathfrak{m}\text{-a.e. in } X.$$

As a consequence of the previous proposition, since (2.5) yields $|\nabla^+\varphi| \in L^2(X,\mu)$, the lower bound on μ yields $|\nabla\varphi|_w \in L^2(X,\mathfrak{m})$.

2.4.2 Convex functionals: gradient flows, entropy, and the $CD(K,\infty)$ condition

Let (Y,d_Y) be a complete and separable metric space, $E : Y \to \mathbb{R} \cup \{+\infty\}$, and $K \in \mathbb{R}$. We say that E is K-geodesically convex if for any y_0, $y_1 \in D(E)$ there exists $\gamma \in \mathrm{Geo}(Y)$ satisfying $\gamma_0 = y_0$, $\gamma_1 = y_1$ and

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

Notice that if E is K-geodesically convex, then $D(E)$ is geodesic in Y and therefore $\overline{D(E)}$ is a length space.

A consequence of K-geodesic convexity is that the descending slope $|\nabla^- E|$ can be calculated at all $y \in D(E)$ as

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

We recall (see [3, Corollary 2.4.10]) that for K-geodesically convex and l.s.c. functionals the descending slope is an upper gradient, in particular the property we shall need is

$$E(y_s) \leq E(y_t) + \int_s^t |\dot{y}_r|\,|\nabla^- E|(y_r)\,\mathrm{d}r \qquad \text{for every } s,\, t \in [0,\infty),\ s<t, \tag{2.23}$$

for *all* locally absolutely continuous curves $y : [0,\infty) \to D(E)$. A metric gradient flow for the K-geodesically convex functional E is a locally absolutely continuous curve $y : [0,\infty) \to D(E)$ along which (2.23) holds as an equality and moreover $|\dot{y}_t| = |\nabla^- E|(y_t)$ for a.e. $t > 0$, so that the energy dissipation rate $\frac{\mathrm{d}}{\mathrm{d}t}E(y_t)$ is equal to $-|\dot{y}_t|^2 = -|\nabla^- E|^2(y_t)$ for a.e. $t > 0$.

An application of Young inequality shows that metric gradient flows for K-geodesically convex and l.s.c. functionals can equivalently be defined as follows.

Definition 2.13 (Metric formulation of gradient flow) *Let $E : Y \to \mathbb{R} \cup \{+\infty\}$ be a K-geodesically convex and l.s.c. functional. We say that a locally absolutely continuous curve $[0,\infty) \ni t \mapsto y_t \in D(E)$ is a gradient flow of E starting from $y_0 \in D(E)$ if*

$$E(y_0) = E(y_t) + \int_0^t \frac{1}{2}|\dot{y}_r|^2 + \frac{1}{2}|\nabla^- E|^2(y_r)\,\mathrm{d}r \qquad \forall t \geq 0. \tag{2.24}$$

We now recall the definition of metric measure space with Ricci curvature bounded from below by $K \in \mathbb{R}$, following [39, §4.2] and [27, §5]. More precisely, we consider here the weaker definition of [39] and we will discuss a stronger version in Section 3: see the bibliographical references of [43, Chapter 17] for a comparison between the two approaches.

Definition 2.14 ($CD(K,\infty)$ spaces) *We say that $(X,\mathsf{d},\mathfrak{m}) \in \mathbb{X}$ has Ricci curvature bounded from below by $K \in \mathbb{R}$ (in short: it is a $CD(K,\infty)$ space) if the relative entropy functional $\mathrm{Ent}_\mathfrak{m}$ is K-geodesically convex on $(\mathscr{P}_2(X),W_2)$, i.e. for any pair of measures μ, $\nu \in D(\mathrm{Ent}_\mathfrak{m}) \cap \mathscr{P}_2(X)$ there exists a constant speed geodesic $(\mu_t) \subset \mathscr{P}_2(X)$ such that $\mu_0 = \mu$, $\mu_1 = \nu$ and*

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

Notice that, in comparison with the definition given in [39] and [27] we are restricting the analysis to the case of a probability reference measure $\mathfrak{m}$ with finite second moment (but we do not assume local compactness). This is actually unneeded from the "Ricci bound" point of view (see also [2, Definition 9.1]), however in this paper we want to focus more on the geometrical aspect, rather than on the - non trivial - analytic tools needed to work in higher generality: the assumption $\mathfrak{m} \in \mathscr{P}_2(X)$ serves to this scope.

Let us also remark that a $CD(K,\infty)$ space $(X,\mathsf{d},\mathfrak{m})$ satisfies the length property, i.e. $\operatorname{supp}\mathfrak{m}$ is a length space if it is endowed with the distance d [39, Remark 4.6(iii)] (the proof therein, based on an approximate midpoint construction, does not use the local compactness).

Now let $(X,\mathsf{d},\mathfrak{m})$ be a $CD(K,\infty)$ space. Then, by assumption, the relative entropy functional $\mathrm{Ent}_{\mathfrak{m}}$ is K-geodesically convex on $(\mathscr{P}_2(X), W_2)$, so that we could ask about the existence and the uniqueness of its gradient flow. The following theorem, proved in [18] for the locally compact case and generalized in [2, Theorem 9.3(ii)] holds:

Theorem 2.15 (Gradient flow of the relative entropy) *Let $(X,\mathsf{d},\mathfrak{m})$ be a $CD(K,\infty)$ space. Then for any $\mu \in D(\mathrm{Ent}_{\mathfrak{m}}) \cap \mathscr{P}_2(X)$ there exists a unique gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ starting from μ.*

Notice that the theorem says nothing about contractivity of the Wasserstein distance along the flow, a property which we address in Section 2.5. Actually, Ohta and Sturm proved in [41] that contractivity *fails* if $(X,\mathsf{d},\mathfrak{m})$ is $\mathbb{R}^d$ endowed with the Lebesgue measure and with a distance coming from a norm not induced by a scalar product.

2.4.3 The heat flow as gradient flow in $L^2(X,\mathfrak{m})$ and in $\mathscr{P}_2(X)$

One of the main result of [2] has been the following identification theorem, see formula (8.5) and Theorem 9.3(iii) therein.

Theorem 2.16 (The heat flow as gradient flow) *Let $(X,\mathsf{d},\mathfrak{m})$ be a $CD(K,\infty)$ space and let $f \in L^2(X,\mathfrak{m})$ be such that $\mu = f\mathfrak{m} \in \mathscr{P}_2(X)$. Let (f_t) be the gradient flow of Ch in $L^2(X,\mathfrak{m})$ starting from f as in Theorem 2.11, and let (μ_t) be the gradient flow of $\mathrm{Ent}_{\mathfrak{m}}$ in $\mathscr{P}_2(X)$ starting from μ, as in Theorem 2.15.*
Then $\mu_t = f_t\mathfrak{m}$ for all $t \geq 0$, $t \mapsto \mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ is locally absolutely continuous in $[0,\infty)$, and

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}_{\mathfrak{m}}(\mu_t) = |\dot{\mu}_t|^2 = \int_{\{f_t>0\}} \frac{|\nabla f_t|_w^2}{f_t}\,\mathrm{d}\mathfrak{m} \qquad \text{for a.e. } t>0. \tag{2.25}$$

In other words, we can unambiguously define the heat flow on a $CD(K,\infty)$ space either as the gradient flow of Cheeger's energy in $L^2(X,\mathfrak{m})$ or as the gradient flow of the relative entropy in $(\mathscr{P}_2(X), W_2)$. A byproduct of this proof is also (see [2, Theorem 9.3(i)]) the equality between slope and the so-called Fisher information functional:

$$|\nabla^-\mathrm{Ent}_{\mathfrak{m}}|^2(\rho\mathfrak{m}) = 4\int |\nabla\sqrt{\rho}|_w^2\,\mathrm{d}\mathfrak{m} \tag{2.26}$$

for all probability densities ρ such that $\sqrt{\rho} \in D(\mathsf{Ch})$. Choosing $f = \sqrt{\rho}$ this identity, in conjunction with the HWI inequality relating entropy, Wasserstein distance and Fisher information (see [26] or [1, Proposition 7.18]) gives the log-Sobolev inequality

$$\int f^2 \log f^2\,\mathrm{d}\mathfrak{m} \leq \frac{2}{K}\int |\nabla f|_w^2\,\mathrm{d}\mathfrak{m} \qquad \text{whenever } f \in D(\mathsf{Ch}) \text{ and } \int f^2\,\mathrm{d}\mathfrak{m} = 1. \tag{2.27}$$

We will denote by $\mathsf{H}_t : L^2(X,\mathfrak{m}) \to L^2(X,\mathfrak{m})$ the heat semigroup in $L^2(X,\mathfrak{m})$ and by $\mathscr{H}_t : \mathscr{P}_2(X) \to \mathscr{P}_2(X)$ the gradient flow of the entropy on $\mathscr{P}_2(X)$. A distinct notation is useful not only for conceptual reasons, but also because the domains of the two gradient flows don't match, even if we identify absolutely continuous measures with their densities.

Some basic properties of the heat flow that we will need later on are collected in the following proposition, see [2, Theorem 4.16] also for further details.

Proposition 2.17 (Some properties of the heat flow) *Let $(X,\mathsf{d},\mathfrak{m}) \in \mathbb{X}$ and $f \in L^2(X,\mathfrak{m})$. Then the following statements hold:*

(i) (Maximum principle) If $f \le C$ (resp. $f \ge C$) $\mathfrak{m}$-a.e. in X for some $C \in \mathbb{R}$, then $\mathsf{H}_t(f) \le C$ (resp. $\mathsf{H}_f(f) \ge C$) $\mathfrak{m}$-a.e. in X for any $t \ge 0$.

(ii) (1-homogeneity) $\mathsf{H}_t(\lambda f) = \lambda \mathsf{H}_t(f)$ for any $\lambda \in \mathbb{R}$, $t \ge 0$.

2.5 EVI formulation of gradient flows

Here we recall a stronger formulation of gradient flows in a complete and separable metric space (Y,d_Y), introduced and extensively studied in [3], [13], [35], which will play a key role in our analysis.

Definition 2.18 (Gradient flows in the EVI sense) *Let $E : Y \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous functional, $K \in \mathbb{R}$ and $(0,\infty) \ni t \mapsto y_t \in D(E)$ be a locally absolutely continuous curve. We say that (y_t) is a K-gradient flow for E in the Evolution Variational Inequalities sense (or, simply, it is an EVI_K gradient flow) if for any $z \in Y$ it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathsf{d}_Y^2(y_t,z)}{2} + \frac{K}{2}\mathsf{d}_Y^2(y_t,z) + E(y_t) \le E(z) \qquad \text{for a.e. } t \in (0,\infty). \tag{2.28}$$

If $\lim_{t\downarrow 0} y_t = y_0 \in \overline{D(E)}$, we say that the gradient flow starts from y_0.

Notice that the derivative in (2.28) exists for a.e. $t > 0$, since $t \mapsto \mathsf{d}_Y(y_t,z)$ is locally absolutely continuous in $(0,\infty)$.

In the next proposition we will consider equivalent formulations of (2.28) involving subsets $D \subset D(E)$ *dense in energy*: it means that for any $y \in D(E)$ there exists a sequence $(y_n) \subset D$ such that $\mathsf{d}_Y(y_n,y) \to 0$ and $E(y_n) \to E(y)$ as $n \to \infty$.

Proposition 2.19 (Equivalent formulations of EVI) *Let E, K be as in Definition 2.18, $D \subset D(E)$ dense in energy, and $y : (0,\infty) \to D(E)$ be a locally absolutely continuous curve with $\lim_{t\downarrow 0} y_t = y_0 \in \overline{D(E)}$. Then, (y_t) is an EVI_K gradient flow if and only if one of the following properties is satisfied:*

(i) (Dense version) The differential inequality (2.28) holds for all $z \in D$.

(ii) (Integral version) For all $z \in D$ it holds

$$\frac{\mathrm{e}^{K(t-s)}}{2}\mathsf{d}_Y^2(y_t,z) - \frac{\mathsf{d}_Y^2(y_s,z)}{2} \le \mathrm{I}_K(t-s)\Big(E(z) - E(y_t)\Big) \qquad \text{for every } 0 \le s \le t, \tag{2.29}$$

where $\mathrm{I}_K(t) := \int_0^t \mathrm{e}^{Kr}\,\mathrm{d}r$.

(iii) (Pointwise version) For all $z \in D$ it holds

$$\limsup_{h \downarrow 0} \frac{\mathsf{d}_Y^2(y_{t+h}, z) - \mathsf{d}_Y^2(y_t, z)}{2} + \frac{K}{2} \mathsf{d}_Y^2(y_t, z) + E(y_t) \le E(z) \qquad \textit{for every } t > 0. \tag{2.30}$$

Proof. To get (2.29) for all $z \in D(E)$ from (2.28), just multiply by e^{Kt} and integrate in time, using the fact that $t \mapsto E(y_t)$ is nonincreasing (see e.g. [10] and the next Proposition); a differentiation provides the equivalence, since y is absolutely continuous. The fact that (2.29) holds for any z if and only if it holds in a set dense in energy is trivial, so that the equivalence of (*ii*) and Definition 2.18 is proved. The equivalences with (*i*) and (*iii*) follow by similar arguments. □

We recall some basic and useful properties of gradient flows in the EVI sense; we give here the essential sketch of the proofs, referring to [3, Chap. 4] and [35] for more details and results. In particular, we emphasize that the maps $\mathsf{S}_t : y_0 \mapsto y_t$ that at every y_0 associate the value at time $t \ge 0$ of the unique K-gradient flow starting from y_0 give raise to a continuous semigroup of K-contractions according to (2.31) in a closed (possibly empty) subset of Y.

Proposition 2.20 (Properties of gradient flows in the EVI sense) *Let Y, E, K, y_t be as in Definition 2.18 and suppose that (y_t) is an EVI_K gradient flow of E starting from y_0. Then:*

(i) If $y_0 \in D(E)$, then y_t is also a metric gradient flow, i.e. (2.24) *holds.*

(ii) If $(\tilde{y}_t)$ is another EVI_K gradient flow for E starting from $\tilde{y}_0$, it holds

$$\mathsf{d}_Y(y_t, \tilde{y}_t) \le e^{-Kt} \mathsf{d}_Y(y_0, \tilde{y}_0). \tag{2.31}$$

In particular, EVI_K gradient flows uniquely depend on the initial condition.

(iii) Existence of EVI_K gradient flows starting from any point in $D \subset Y$ implies existence starting from any point in $\overline{D}$.

(iv) (y_t) is locally Lipschitz in $(0, \infty)$, $y_t \in D(|\nabla^- E|)$ for every $t > 0$, the map $t \mapsto \mathrm{e}^{Kt} |\nabla^- E|(y_t)$ is nonincreasing, and we have the regularization estimate

$$\mathrm{I}_K(t)\, E(y_t) + \frac{\big(\mathrm{I}_K(t)\big)^2}{2} |\nabla^- E|^2(y_t) \le \mathrm{I}_K(t)\, E(z) + \frac{1}{2} \mathsf{d}_Y^2(z, y_0) \quad \forall t > 0,\ z \in D(E). \tag{2.32}$$

Proof. The fact that EVI_K gradient flows satisfy (2.24) has been proved by the third author in [35] (see also [1, Proposition 3.9]). The contractivity property (*ii*) has been proved in [3, Chap. 4]. Statement (*iii*) follows trivially from contractivity and integral formulation (2.29) of the EVI. The fact that $t \mapsto \mathrm{e}^{Kt} |\nabla^- E|(y_t)$ is nonincreasing follows from the energy identity, which shows that $|\nabla^- E|(y_t) = |\dot{y}_t|$, and the K-contraction estimate (2.31), which in particular yields that $t \mapsto \mathrm{e}^{Kt} \mathsf{d}_Y(y_t, y_{t+h})$ is nonincreasing as well as $t \mapsto \mathrm{e}^{Kt} |\dot{y}_t|$.

An easier regularization formula for $t \mapsto E(y_t)$ follows immediately by (2.29) by choosing $s = 0$ and neglecting the term proportional to $\mathsf{d}_Y^2(y_t, z)$. Inequality (2.32) is a consequence of the EVI_K, the identity $\frac{\mathrm{d}}{\mathrm{d}t} E(y_t) = -|\nabla^- E|^2(y_t)$, the previous monotonicity property and the following calculations:

$$\frac{1}{2}\big(\mathrm{I}_K(t)\big)^2|\nabla^- E|^2(y_t) = \frac{1}{2}\big(\mathrm{I}_{-K}(t)\big)^2\mathrm{e}^{2Kt}|\nabla^- E|^2(y_t) \le \int_0^t \mathrm{I}_{-K}(s)\mathrm{e}^{-Ks}\mathrm{e}^{2Ks}|\nabla^- E|^2(y_s)\,\mathrm{d}s$$
$$= -\int_0^t \mathrm{I}_{-K}(s)\mathrm{e}^{Ks}\big(E(y_s)-E(y_t)\big)'\,\mathrm{d}s = \int_0^t \mathrm{e}^{Ks}\big(E(y_s)-E(y_t)\big)\,\mathrm{d}s$$
$$\le \int_0^t -\frac{1}{2}\big(\mathrm{e}^{Ks}\mathsf{d}_Y^2(y_s,z)\big)' + \mathrm{e}^{Ks}\big(E(z)-E(y_t)\big)\,\mathrm{d}s \le \frac{1}{2}\mathsf{d}_Y^2(y_0,z) + \mathrm{I}_K(t)\big(E(z)-E(y_t)\big).$$

□

We point out that in general existence of EVI_K gradient flows is a consequence of the K-geodesic convexity of E and of strong geometric assumptions on the metric space (Y,d_Y): it is well known when Y is a convex set of an Hilbert space, but existence holds even when (Y,d_Y) satisfies suitable lower sectional curvature bounds in the sense of Alexandrov [20, 29, 34, 35], or when suitable compatibility conditions between E and d hold [3, Chapter 4] which include also spaces with nonpositive Alexandrov curvature. In the present paper we will study an important situation where EVI_K gradient flows arise without any assumption on sectional curvature.

In any case, EVI_K gradient flows have the following interesting geometric consequence on the functional E [13, Theorem 3.2]: if EVI_K gradient flows exist for any initial data, then the functional is K-convex along *any* geodesic contained in $\overline{D(E)}$. Recall that the standard definition of geodesic convexity, e.g. the one involved in Definition 2.14 of $CD(K,\infty)$ metric measure spaces, requires convexity along *some* geodesic; this choice is usually motivated by stability properties w.r.t. Γ-convergence [3, Thm. 9.1.4] and Sturm-Gromov-Hausdorff convergence in the case of metric measure spaces (see also the next section). We state this property in a quantitative way, which will turn out to be useful in the following.

Proposition 2.21 *Let E, K, y_t be as in Definition 2.18 and assume that for every $y_0 \in \overline{D(E)}$ there exists the EVI_K gradient flow $y_t := \mathsf{S}_t(y_0)$ for E starting from y_0. If $\varepsilon \ge 0$ and $\gamma : [0,1] \to \overline{D(E)}$ is a Lipschitz curve satisfying*

$$\mathsf{d}_Y(\gamma_{s_1},\gamma_{s_2}) \le L\,|s_1 - s_2|, \quad L^2 \le \mathsf{d}_Y^2(\gamma_0,\gamma_1) + \varepsilon^2 \qquad \textit{for every } s_1,\, s_2 \in [0,1], \tag{2.33}$$

then for every $t > 0$ and $s \in [0,1]$

$$E(\mathsf{S}_t(\gamma_s)) \le (1-s)E(y_0) + sE(y_1) - \frac{K}{2}s(1-s)\mathsf{d}_Y^2(y_0,y_1) + \frac{\varepsilon^2}{2\mathrm{I}_K(t)}s(1-s). \tag{2.34}$$

In particular E is K-convex along all geodesics contained in $\overline{D(E)}$.

The last statement is an immediate consequence of (2.34) by choosing $\varepsilon = 0$ and letting $t \downarrow 0$.

3 Strong $CD(K,\infty)$ spaces

Definition 3.1 (Strong $CD(K,\infty)$ spaces) *We say that $(X,\mathsf{d},\mathfrak{m})$ is a strong $CD(K,\infty)$ space if for every μ_0, $\mu_1 \in D(\mathrm{Ent}_{\mathfrak{m}}) \cap \mathscr{P}_2(X)$ there exists an optimal geodesic plan $\boldsymbol{\pi}$ from μ_0 to μ_1 such that K-convexity of the entropy holds along all weighted plans $\boldsymbol{\pi}_F := F\boldsymbol{\pi}$, where*

$F : \mathrm{Geo}(X) \to \mathbb{R}$ is any Borel, bounded, non negative function such that $\int F \,\mathrm{d}\boldsymbol{\pi} = 1$. More precisely, for any such F, the interpolated measures $\mu_{F,t} := (\mathrm{e}_t)_\sharp \boldsymbol{\pi}_F$ satisfy:

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

It is unclear to us whether this notion is stable w.r.t $\mathbb{D}$-convergence or not. As such, it should be handled with care. We introduced this definition for two reasons. The first one is that applying Proposition 2.21 we will show in Lemma 5.2 that if a length metric measure space $(X, \mathsf{d}, \mathfrak{m})$ admits existence of EVI_K gradient flows of $\mathrm{Ent}_{\mathfrak{m}}$ for any initial measure $\mu \in D(\mathrm{Ent}_{\mathfrak{m}}) \cap \mathscr{P}_2(X)$, then it is a strong $CD(K,\infty)$ space. Given that spaces admitting EVI_K gradient flows for $\mathrm{Ent}_{\mathfrak{m}}$ are the main subject of investigation of this paper, it is interesting to study a priori the properties of strong $CD(K,\infty)$ spaces. The other reason is due to the fact that the proof that linearity of the heat flow implies the existence of EVI_K gradient flows of the entropy requires additional L^∞-estimates for displacement interpolations which looks unavailable in general $CD(K,\infty)$ spaces.

Remark 3.2 (The nonbranching case) If a space $(X, \mathsf{d}, \mathfrak{m})$ is $CD(K,\infty)$ and nonbranching, then it is also strong $CD(K,\infty)$ according to the previous definition.
Indeed, pick μ_0, $\mu_1 \in D(\mathrm{Ent}_{\mathfrak{m}})$, let $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0,\mu_1)$ be such that the relative entropy is K-convex along $((\mathrm{e}_t)_\sharp\boldsymbol{\pi})$, so that $\mathrm{Ent}_{\mathfrak{m}}((\mathrm{e}_t)_\sharp\boldsymbol{\pi})$ is bounded in $[0,1]$. Now, pick F as in Definition 3.1, let $\mu_{F,t} := (\mathrm{e}_t)_\sharp\boldsymbol{\pi}_F$ and notice that the real function $s \mapsto \phi(s) := \mathrm{Ent}_{\mathfrak{m}}(\mu_{F,s})$ is bounded (thus in particular $\mu_{F,s} \in D(\mathrm{Ent}_{\mathfrak{m}})$) in $[0,1]$, since $\mu_{F,t} \leq \sup|F|(\mathrm{e}_t)_\sharp\boldsymbol{\pi}$.
The nonbranching assumption ensures that for any $t \in (0,1)$ there is a unique geodesic connecting $\mu_{F,t}$ to $\mu_{F,0}$ (and similarly to $\mu_{F,1}$). Hence, since $(X, \mathsf{d}, \mathfrak{m})$ is a $CD(K,\infty)$ space, the restriction of ϕ to all the intervals of the form $[0,t]$ and $[t,1]$ for $t \in (0,1)$ is K-convex and finite. It follows that ϕ is K-convex in $[0,1]$. ■

In order to better understand the next basic interpolation estimate, let us consider the simpler case of an otimal geodesic plan $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0,\mu_1)$ in a nonbranching $CD(K,\infty)$ space $(X, \mathsf{d}, \mathfrak{m})$. Assuming that $\mu_i = \rho_i\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}})$ and setting $\mu_t := \rho_t\mathfrak{m}$, along $\boldsymbol{\pi}$-a.e. geodesic γ the real map $t \mapsto \log\rho_t(\gamma_t)$ is K-convex and therefore $\rho_t(\gamma_t)$ can be pointwise estimated by [43, Thm. 30.32, (30.51)]

$$\rho_t(\gamma_t) \leq \mathrm{e}^{-\frac{K}{2}\,t(1-t)\,\mathsf{d}^2(\gamma_0,\gamma_1)}\rho_0(\gamma_0)^{1-t}\,\rho_1(\gamma_1)^t \quad \text{for every } t \in [0,1], \text{ for } \boldsymbol{\pi}\text{-a.e. } \gamma. \tag{3.1}$$

Inequality (3.1) for smooth Riemannian manifolds goes back to [11]. If μ_i have bounded supports, one immediately gets the uniform L^∞-bound:

$$\|\rho_t\|_\infty \leq \mathrm{e}^{\frac{K^-}{2}t(1-t)S^2}\|\rho_0\|_\infty^{1-t}\,\|\rho_1\|_\infty^t, \quad \text{with } S := \sup\big\{\mathsf{d}(x_0,x_1):\ x_i \in \mathrm{supp}(\mu_i)\big\}. \tag{3.2}$$

When $K \geq 0$ (3.2) is also a consequence of the definition (stronger than (2.14)) of spaces with non-negative Ricci curvature given by [27], which in particular yields the geodesic convexity of all the functionals $U_p(\mu) := \int \rho^p \,\mathrm{d}\mathfrak{m}$ whenever $\mu = \rho\mathfrak{m}$ and $p > 1$.

If we know that only ρ_1 is supported in a bounded set, we can still get a weighted L^∞-bound on ρ_t. Let us assume that

$$\mathrm{supp}\,\rho_1 \subset \mathsf{C}, \quad \text{with} \quad \mathsf{Di} := \mathrm{diam}(\mathsf{C}) < \infty, \quad \mathsf{D}(x) := \mathrm{dist}(x,\mathsf{C}) \quad \text{for } x \in X, \tag{3.3}$$

and let us observe that for $\boldsymbol{\pi}$-a.e. γ we have $\gamma_1 \in \operatorname{supp}\mu_1$, so that for every $t \in [0,1)$ it holds

$$\mathsf{d}(\gamma_0,\gamma_1) = \frac{\mathsf{d}(\gamma_t,\gamma_1)}{1-t} \le \frac{\mathsf{D}(\gamma_t)+\mathsf{Di}}{1-t}, \tag{3.4}$$

$$\mathsf{D}(\gamma_0) \ge \mathsf{d}(\gamma_t,\gamma_1) - \mathsf{Di} - \mathsf{d}(\gamma_0,\gamma_t) = (1-2t)\mathsf{d}(\gamma_0,\gamma_1) - \mathsf{Di} \ge \frac{1-2t}{1-t}\mathsf{D}(\gamma_t) - \mathsf{Di}. \tag{3.5}$$

Substituting the above bounds in (3.1) we get

$$\rho_t(x) \le \mathrm{e}^{\frac{K^-}{2}\frac{t}{1-t}(\mathsf{D}(x)+\mathsf{Di})^2}\|\rho_0\|^{1-t}_{L^\infty(R(\mathsf{D}(x),t);\mathfrak{m})}\|\rho_1\|^t_\infty \qquad \mathfrak{m}\text{-a.e. in } X, \tag{3.6}$$

where

$$R(D,t) := \Big\{y \in X : \mathsf{D}(y) \ge \frac{1-2t}{1-t}D - \mathsf{Di}\Big\}, \quad D \ge 0,\ t \in [0,1). \tag{3.7}$$

The next lemma shows that the strong $CD(K,\infty)$ condition is sufficient to obtain the same estimates.

Proposition 3.3 (Interpolation properties) *Let* $(X,\mathsf{d},\mathfrak{m})$ *be a strong* $CD(K,\infty)$ *space and let* ρ_0, ρ_1 *be probability densities such that* $\mu_i = \rho_i\mathfrak{m} \in D(\mathrm{Ent}_{\mathfrak{m}}) \subset \mathscr{P}_2(X)$. *Assume that* ρ_1 *is bounded and with support contained in a bounded set* C *as in* (3.3) *and let* $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0,\mu_1)$ *as in Definition 3.1. Then for all* $t \in [0,1)$ *the density* ρ_t *of* $\mu_t = (\mathrm{e}_t)_\sharp\boldsymbol{\pi}$ *satisfies* (3.6)*. Furthermore, if also* ρ_0 *is bounded with bounded support, then* (3.2) *holds and* $\sup_t\|\rho_t\|_\infty < \infty$.

Proof. Let $\boldsymbol{\pi}$ be given by the strong $CD(K,\infty)$ condition. Fix $t \in (0,1)$ and assume that (3.6) does not hold on a Borel set B of positive $\mathfrak{m}$-measure. Then we can find a Borel set $A \subset B$ with $\mathfrak{m}(A) > 0$ such that

$$\rho_t(x) > \mathrm{e}^{\frac{K^-}{2}\frac{t}{1-t}(\mathsf{D}_1+\mathsf{Di})^2}\mathsf{M}^{1-t}\|\rho_1\|^t_\infty \qquad \forall x \in A,$$

where

$$\mathsf{M} := \|\rho_0\|_{L^\infty(R(\mathsf{D}_2,t);\mathfrak{m})}, \qquad \mathsf{D}_1 := \sup_{x\in A}\mathsf{D}(x), \qquad \mathsf{D}_2 := \inf_{x\in A}\mathsf{D}(x). \tag{3.8}$$

To build A, it suffices to slice B in countably many pieces where the oscillation of D is sufficiently small. We have $\boldsymbol{\pi}((\mathrm{e}_t)^{-1}(A)) = \mu_t(A) > 0$, thus the plan $\tilde{\boldsymbol{\pi}} := c\,\boldsymbol{\pi}\llcorner\mathrm{e}_t^{-1}(A)$, where $c := \big[\mu_t(A)\big]^{-1}$ is the normalizing constant, is well defined. Let $\tilde\rho_s$ be the density of $\tilde\mu_s = (\mathrm{e}_s)_\sharp\tilde{\boldsymbol{\pi}}$. By definition it holds $\tilde\rho_t = c\rho_t$ on A and $\tilde\rho_t = 0$ on $X \setminus A$, thus we have:

$$\mathrm{Ent}_{\mathfrak{m}}(\tilde\mu_t) = \int \tilde\rho_t\log\tilde\rho_t\,\mathrm{d}\mathfrak{m} > \log c + \frac{K^-}{2}\frac{t}{1-t}\big(\mathsf{D}_1+\mathsf{Di}\big)^2 + (1-t)\log\mathsf{M} + t\log\|\rho_1\|_\infty. \tag{3.9}$$

On the other hand, we have $\tilde\rho_0 \le c\rho_0$ and $\tilde\rho_1 \le c\rho_1$ hence

$$\mathrm{Ent}_{\mathfrak{m}}(\tilde\mu_0) = \int \log(\tilde\rho_0\circ\mathrm{e}_0)\,\mathrm{d}\tilde{\boldsymbol{\pi}} \le \log c + \log\Big(\|\rho_0\circ\mathrm{e}_0\|_{L^\infty(\mathrm{Geo}(X),\tilde{\boldsymbol{\pi}})}\Big), \tag{3.10}$$

$$\mathrm{Ent}_{\mathfrak{m}}(\tilde\mu_1) = \int \log(\tilde\rho_1\circ\mathrm{e}_1)\,\mathrm{d}\tilde{\boldsymbol{\pi}} \le \log c + \log\|\rho_1\|_\infty. \tag{3.11}$$

Now observe that $\tilde{\boldsymbol{\pi}}$-a.e. geodesic γ satisfies $\gamma_t \in A$ and $\gamma_1 \in \operatorname{supp}\rho_1 \subset \mathsf{C}$, so that (3.4) and (3.5) yield

$$\mathsf{d}(\gamma_0,\gamma_1) \le \frac{\mathsf{D}_1 + \mathsf{Di}}{1-t}, \qquad \mathsf{D}(\gamma_0) \ge \frac{1-2t}{1-t}\mathsf{D}_2 - \mathsf{Di}, \quad \text{i.e.} \quad \gamma_0 \in R(\mathsf{D}_2, t).$$

Integrating the squared first inequality w.r.t. $\tilde{\boldsymbol{\pi}}$ and combining the second one, (3.10), and (3.8) we get

$$W_2^2(\tilde{\mu}_0,\tilde{\mu}_1) \le \left(\frac{\mathsf{D}_1 + \mathsf{Di}}{1-t}\right)^2, \qquad \mathrm{Ent}_{\mathfrak{m}}(\tilde{\mu}_0) \le \log c + (1-t)\log \mathsf{M}. \tag{3.12}$$

Inequalities (3.9), (3.11), and (3.12) contradict the K-convexity of the entropy along $((\mathrm{e}_s)_\sharp\tilde{\boldsymbol{\pi}})$, so the proof of the first claim is concluded.

The proof of (3.2) when also ρ_0 has bounded support follows the same lines just used. Let $t \in (0,1)$ and assume that (3.2) does not hold. Thus there exists a Borel set A of positive $\mathfrak{m}$-measure such that $\rho_t > e^{K^- t(1-t)S^2/2}\|\rho_0\|_\infty^{1-t}\|\rho_1\|_\infty^t$ in A. As before, we define $\tilde{\boldsymbol{\pi}} := c\boldsymbol{\pi}\llcorner \mathrm{e}_t^{-1}(A)$, where c is the normalizing constant, and $\tilde{\rho}_s$ as the density of $(\mathrm{e}_s)_\sharp\tilde{\boldsymbol{\pi}}$: the inequalities

$$\begin{aligned}
&\mathrm{Ent}_{\mathfrak{m}}((\mathrm{e}_t)_\sharp\tilde{\boldsymbol{\pi}}) > \log c + \frac{K^-}{2}t(1-t)S^2 + (1-t)\log\|\rho_0\|_\infty + t\log\|\rho_1\|_\infty,\\
&\mathrm{Ent}_{\mathfrak{m}}((\mathrm{e}_0)_\sharp\tilde{\boldsymbol{\pi}}) \le \log c + \log\|\rho_0\|_\infty, \quad \mathrm{Ent}_{\mathfrak{m}}((\mathrm{e}_1)_\sharp\tilde{\boldsymbol{\pi}}) \le \log c + \log\|\rho_1\|_\infty,\\
&W_2^2\big((\mathrm{e}_0)_\sharp\tilde{\boldsymbol{\pi}}, (\mathrm{e}_1)_\sharp\tilde{\boldsymbol{\pi}}\big) \le S^2,
\end{aligned}$$

contradict the K-convexity of the entropy along $((\mathrm{e}_s)_\sharp\tilde{\boldsymbol{\pi}})$. □

In the sequel we will occasionally use the stretching/restriction operator restr_0^s in $\mathrm{C}([0,1];X)$, defined for all $s \in [0,1]$ by

$$\mathrm{restr}_0^s(\gamma)_t := \gamma_{ts} \qquad t \in [0,1].$$

Proposition 3.4 (Existence of test plans) *Let $(X,\mathsf{d},\mathfrak{m})$ be a strong $CD(K,\infty)$ space and let ρ_0, ρ_1 be probability densities. Assume that ρ_1 is bounded with bounded support as in* (3.3)*, that ρ_0 is bounded and satisfies*

$$\rho_0(x) \le c\,\mathrm{e}^{-9K^-(\mathsf{D}(x)-C)^2} \qquad \textit{whenever } \mathsf{D}(x) := \operatorname{dist}(x,\operatorname{supp}\rho_1) > R, \tag{3.13}$$

for some nonnegative constants c, C, R. Then, for $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\rho_0\mathfrak{m},\rho_1\mathfrak{m})$ as in Definition 3.1, $(\mathrm{restr}_0^{1/3})_\sharp\boldsymbol{\pi}$ is a test plan (recall Definition 2.8).

Proof. In order to avoid cumbersome formulas, in this proof we switch to the exp notation. We need to prove that $\sup_X \rho_t$ is uniformly bounded in $[0,1/3]$. Let $\mathsf{Di} = \operatorname{diam}(\operatorname{supp}\rho_1)$, M the function defined in (3.7), L a constant to be specified later, $A := \{y : \mathsf{D}(y) \le L\}$ and set $\boldsymbol{\pi}^1 := \boldsymbol{\pi}\llcorner \mathrm{e}_0^{-1}(A)$, $\boldsymbol{\pi}^2 := \boldsymbol{\pi}\llcorner \mathrm{e}_0^{-1}(X \setminus A)$. Choosing L large enough we have $\alpha := \boldsymbol{\pi}(\mathrm{e}_0^{-1}(A)) > 0$ and we can also assume that $\alpha < 1$ (otherwise, ρ_0 has bounded support and the second part of Proposition 3.3 applies). Also, possibly increasing R and taking (3.13) into account, we can assume that $\rho_0(x) \le 1$ wherever $\mathsf{D}(x) \ge R$.

Denoting by $\tilde{\boldsymbol{\pi}}^1$, $\tilde{\boldsymbol{\pi}}^2$ the corresponding renormalized plans, it suffices to show that both have bounded densities in the time interval $[0,1/3]$, because $\boldsymbol{\pi}$ is a convex combination of them. Concerning $\tilde{\boldsymbol{\pi}}^1$, notice that both $(\mathrm{e}_0)_\sharp\tilde{\boldsymbol{\pi}}^1$ and $(\mathrm{e}_1)_\sharp\tilde{\boldsymbol{\pi}}^1$ have bounded support and bounded density, so that the conclusion follows from the second part of Proposition 3.3.

For $\tilde{\boldsymbol{\pi}}^2$ we argue as follows. Pick $\gamma \in \operatorname{supp} \tilde{\boldsymbol{\pi}}^2$ and notice that $\gamma_1 \in \operatorname{supp} \rho_1$ and $t \leq \frac{1}{3}$ give the inequality

$$\mathsf{D}(\gamma_t) \geq \mathsf{d}(\gamma_t, \gamma_1) - \mathsf{Di} = (1-t)\mathsf{d}(\gamma_0, \gamma_1) - \mathsf{Di} \geq (1-t)\mathsf{D}(\gamma_0) - \mathsf{Di} \geq \frac{2}{3} D(\gamma_0) - \mathsf{Di}.$$

So, choosing L sufficiently large (depending only on Di and R), we have

$$\gamma_0 \in X \setminus A \qquad \Rightarrow \qquad \frac{\mathsf{D}(\gamma_t)}{2} - \mathsf{Di} > R.$$

Recalling the definition (3.7) of $R(D,t)$ and using the fact that $t \in [0, 1/3]$, we get that

$$y \in R(\mathsf{D}(\gamma_t), t) \quad \Rightarrow \quad \mathsf{D}(y) \geq \frac{1-2t}{1-t}\mathsf{D}(\gamma_t) - \mathsf{Di} \geq \frac{\mathsf{D}(\gamma_t)}{2} - \mathsf{Di} > R \quad \text{for all } \gamma \in \operatorname{supp} \tilde{\boldsymbol{\pi}}^2,$$

and therefore (3.13) gives

$$\sup_{R(\mathsf{D}(\gamma_t),t)} \rho_0 \leq c \exp\Big(-9K^-\big(\mathsf{D}(\gamma_t) - \mathsf{Di} - C\big)^2\Big) \qquad \text{for all } \gamma \in \operatorname{supp} \tilde{\boldsymbol{\pi}}^2.$$

Now, by applying (3.6) to $\tilde{\boldsymbol{\pi}}^2$, we get that the density η_t of $(\mathrm{e}_t)_\sharp \tilde{\boldsymbol{\pi}}^2$ satisfies

$$\eta_t(\gamma_t) \leq \frac{1}{1-\alpha} \exp\Big(\frac{K^-}{4}(\mathsf{D}(\gamma_t) + \mathsf{Di})^2\Big) \|\rho_0\|^{1-t}_{L^\infty(R(\mathsf{D}(\gamma_t),t),\mathfrak{m})} \|\rho_1\|^t_\infty \qquad \text{for } \tilde{\boldsymbol{\pi}}^2\text{-a.e. } \gamma.$$

Using the fact that t varies in $[0, 1/3]$ and $\rho_0 \leq 1$ in $R(\mathsf{D}(\gamma_t), t)$, we eventually get

$$\eta_t(\gamma_t) \leq c \frac{\|\rho_1\|^{1/3}_\infty}{1-\alpha} \exp\bigg(\frac{K^-}{4}(\mathsf{D}(\gamma_t) + \mathsf{Di})^2 - 6K^-\Big(\frac{\mathsf{D}(\gamma_t)}{2} - \mathsf{Di} - C\Big)^2\bigg) \qquad \text{for } \tilde{\boldsymbol{\pi}}^2\text{-a.e. } \gamma.$$

Since $-\frac{5}{4}K^-\mathsf{D}^2(\gamma_t)$ is the leading term in the exponential, the right-hand side is bounded and we deduce that $\|\eta_t\|_\infty = \|\eta_t \circ \mathrm{e}_t\|_{L^\infty(\mathrm{Geo}(X), \tilde{\boldsymbol{\pi}}^2)}$ is uniformly bounded. □

Proposition 3.5 (Metric Brenier theorem for strong $CD(K,\infty)$ spaces) *Let $(X, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$ be a strong $CD(K,\infty)$ space, $x_0 \in X$, $\mu_0 = \rho_0 \mathfrak{m} \in \mathscr{P}_2(X)$ with*

$$0 < c_R \leq \rho_0 \leq c_R^{-1} \qquad \mathfrak{m}\text{-a.e. in } B_R(x_0) \quad \text{for every } R > 0, \tag{3.14}$$

and $\mu_1 \in \mathscr{P}_2(X)$ with bounded support and bounded density. Then, for $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu_0, \mu_1)$ as in Definition 3.1, there exists $L \in L^2(X, \mu_0)$ such that

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

Furthermore,

$$L(x) = |\nabla \varphi|_w(x) = |\nabla^+ \varphi|(x) \qquad \text{for } \mu_0\text{-a.e. } x \in X,$$

where φ is any Kantorovich potential relative to (μ_0, μ_1).

Proof. We apply the metric Brenier Theorem 10.3 of [2] with $V(x) = \mathsf{d}(x, x_0)$. To this aim, we need only to show that

$$(\mathrm{e}_t)_\sharp \boldsymbol{\pi}(B \cap B_R(x_0)) \le C(R)\mathfrak{m}(B) \qquad \text{for every } t \in [0, 1/2],\ B \in \mathscr{B}(X),\ R > 0. \tag{3.15}$$

Denoting by R_1 the radius of a ball containing the support of μ_1, notice that if a curve γ in the support of $\boldsymbol{\pi}$ hits $B_R(x_0)$ at some time $s \in [0, 1/2]$, then

$$\mathsf{d}(\gamma_0, \gamma_1) \le 2\mathsf{d}(\gamma_s, \gamma_1) \le 2(R + R_1)$$

because $\gamma_1 \in B_{R_1}(x_0)$. Possibly restricting $\boldsymbol{\pi}$ to the set of γ's hitting $B_R(x_0)$ at some $s \in [0, 1/2]$, an operation which does not affect $((\mathrm{e}_t)_\sharp \boldsymbol{\pi}) \llcorner B_R(x_0)$ for $t \in [0, 1/2]$, we get that $(\mathrm{e}_0)_\sharp \boldsymbol{\pi}$, $(\mathrm{e}_1)_\sharp \boldsymbol{\pi}$ have bounded support and bounded densities, thus the conclusion follows from the second part of Proposition 3.3. □

We conclude this section with a technical approximation result of curves in $\mathscr{P}_2(X)$ by "better" curves, having bounded densities w.r.t. $\mathfrak{m}$.

Lemma 3.6 *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be a strong* $CD(K, \infty)$ *space and let* $\mu_t \in \mathrm{AC}^2([0, 1]; \mathscr{P}_2(X))$ *with* $\mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ *bounded in* $[0, 1]$. *Then there exist curves* $\mu_t^n \in \mathrm{AC}^2([0, 1]; \mathscr{P}_2(X))$ *satisfying:*

(i) $\max\limits_{[0,1]} W_2(\mu_t^n, \mu_t) \to 0$ *and* $\big(\max\limits_{[0,1]} \mathrm{Ent}_{\mathfrak{m}}(\mu_t^n) - \max\limits_{[0,1]} \mathrm{Ent}_{\mathfrak{m}}(\mu_t)\big)^+ \to 0$ *as* $n \to \infty$.

(ii) $|\dot\mu_t^n| \to |\dot\mu_t|$ *in* $L^2(0, 1)$ *as* $n \to \infty$.

(iii) For all n *there exists* $C(n)$ *satisfying* $\mu_t^n \le C(n)\mathfrak{m}$ *for all* $t \in [0, 1]$.

Proof. We first notice that suffices to provide approximating sequences μ_t^n satisfying (i), (iii) and a weakened form of (ii), namely $\limsup_n \int_0^1 |\dot\mu_t^n|^2 \, \mathrm{d}t \le \int_0^1 |\dot\mu_t|^2 \, \mathrm{d}t$. Then, the lower semicontinuity of metric derivative yields that any weak limit point g of $|\dot\mu_t^n|$, in the $L^2(0, 1)$ topology, is greater than $|\dot\mu_t|$ a.e. in $(0, 1)$. This provides the lim inf inequality for the norms and (ii).

Now we notice that we can naturally define an approximation σ_t^k of μ_t on a mesh of size $\tau = 1/k$, just choosing geodesics with speed $W_2(\mu_{(i+1)\tau}, \mu_{i\tau})/\tau$ in the interval $[i\tau, (i+1)\tau]$; we can also represent these geodesics by optimal geodesic plans $\boldsymbol{\pi}_i \in \mathscr{P}(\mathrm{C}([i\tau, (i+1)\tau]; X))$ as in Definition 3.1. Then, obviously (i) (thanks to the uniform continuity of $t \mapsto \mu_t$ in $\mathscr{P}_2(X)$ and to the convexity inequality in the intervals $[i\tau, (i+1)\tau]$) and the weakened form of (ii) hold with $\mu_t^k = \sigma_t^k$.

Hence, by a diagonal argument, suffices to prove existence of the approximation when μ_t is a piecewise geodesic curve on a mesh of size $1/k$. We set $\mu_t = \rho_t \mathfrak{m}$ and we represent μ_t as $(\mathrm{e}_t)_\sharp \boldsymbol{\pi}$, where $\boldsymbol{\pi}$, concentrated on piecewise geodesic curves, is obtained by concatenating the k geodesic optimal plans $\boldsymbol{\pi}_i$. Let us fix $\bar x \in X$. Defining $\mu_t^n := c_n (\mathrm{e}_t)_\sharp \boldsymbol{\pi} \llcorner A_n$, where $c_n = \big[\boldsymbol{\pi}(A_n)\big]^{-1} \downarrow 1$ are the normalizing constants and

$$A_n := \left\{ \gamma \in \mathrm{C}([0, 1]; X) : \max_{1 \le i \le k} \rho_{i\tau}(\gamma_{i\tau}) \le n,\ \max_{i \le i \le k} \mathsf{d}(\gamma_{i\tau}, \bar x) \le n \right\},$$

a simple monotonicity argument shows that (i) and (ii) hold. On the other hand, since the marginals at times $i\tau$ are bounded and with bounded support, by applying Proposition 3.3 we obtain that all $c_n \boldsymbol{\pi} \llcorner A_n$ are test plans, so that (iii) holds as well. □

4 Key formulas

4.1 Derivative of the squared Wasserstein distance

In this short section we compute the derivative of the squared Wasserstein distance along a heat flow.

Theorem 4.1 (Derivative of squared Wasserstein distance) *Let* $(X,\mathsf{d},\mathfrak{m})$ *be a* $CD(K,\infty)$ *space,* $\mu=\rho\mathfrak{m}\in\mathscr{P}_2(X)$ *such that* $0<c\le\rho\le C<\infty$ *and define* $\mu_t:=\mathscr{H}_t(\mu)=\rho_t\mathfrak{m}$*. Let* $\sigma\in\mathscr{P}_2(X)$ *and for any* $t>0$ *let* φ_t *be a Kantorovich potential relative to* (μ_t,σ)*. Then for a.e.* $t>0$ *it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\mu_t,\sigma)\le\frac{\mathsf{Ch}(\rho_t-\varepsilon\varphi_t)-\mathsf{Ch}(\rho_t)}{\varepsilon}\qquad\forall\varepsilon>0.\tag{4.1}$$

Proof. Since $t\mapsto\rho_t\mathfrak{m}$ is a locally absolutely continuous curve in $\mathscr{P}_2(X)$, the derivative at the left hand side of (4.1) exists for a.e. $t>0$. Also, the derivative of $t\mapsto\rho_t=\mathsf{H}_t(\rho)\in L^2(X,\mathfrak{m})$ exists for a.e. $t>0$. Fix $t_0>0$ where both derivatives exist and notice that since φ_{t_0} is a Kantorovich potential for (μ_{t_0},σ) it holds

$$\begin{aligned}\frac{1}{2}W_2^2(\mu_{t_0},\sigma)&=\int_X\varphi_{t_0}\,\mathrm{d}\mu_{t_0}+\int\varphi_{t_0}^c\,\mathrm{d}\sigma\\ \frac{1}{2}W_2^2(\mu_{t_0-h},\sigma)&\ge\int_X\varphi_{t_0}\,\mathrm{d}\mu_{t_0-h}+\int\varphi_{t_0}^c\,\mathrm{d}\sigma\qquad\text{for all } h \text{ such that } t_0-h>0.\end{aligned}$$

Taking the difference between the first identity and the second inequality, dividing by $h>0$, and letting $h\to0$ we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\mu_t,\sigma)_{|t=t_0}\le\liminf_{h\downarrow0}\int_X\varphi_{t_0}\frac{\rho_{t_0}-\rho_{t_0-h}}{h}\,\mathrm{d}\mathfrak{m}.$$

Now recall that $\varphi_{t_0}\in L^1(X,\mu_{t_0})$, so that by our assumption on ρ and the maximum principle (Proposition 2.17) we deduce that $\varphi_{t_0}\in L^1(X,\mathfrak{m})$. By Proposition 2.12 we have $|\nabla\varphi_{t_0}|_w\in L^2(X,\mathfrak{m})$.

Now, if $\varphi_{t_0}\in L^2(X,\mathfrak{m})$ the estimate of the lim inf with the difference quotient of Ch is just a consequence of the following three facts: the first one is that, for all $t>0$ we have $h^{-1}(\rho_{t+h}-\rho_t)\to\Delta\rho_t$ as $h\downarrow0$ in $L^2(X,\mathfrak{m})$; the second one is that we have chosen t_0 such that the full limit exists; the third one is the inequality

$$\mathsf{Ch}(\rho_{t_0})+\varepsilon\int\varphi_{t_0}\Delta\rho_{t_0}\,\mathrm{d}\mathfrak{m}\le\mathsf{Ch}(\rho_{t_0}-\varepsilon\varphi_{t_0})\qquad\forall\varepsilon>0$$

provided by the inclusion $-\Delta\rho_{t_0}\in\partial^-\mathsf{Ch}(\rho_{t_0})$.

For the general case, fix $t_0>0$ as before, $\varepsilon>0$ and let $\varphi^N:=\max\{\min\{\varphi_{t_0},N\},-N\}\in L^2(X,\mathfrak{m})$ be the truncated functions. Since the chain rule (2.18) gives $|\nabla\varphi^N|_w\le|\nabla\varphi_{t_0}|_w$, the locality of the minimal weak gradient (2.16) and the dominated convergence theorem ensures that $\mathsf{Ch}(\rho_t-\varepsilon\varphi^N)\to\mathsf{Ch}(\rho_t-\varepsilon\varphi_{t_0})$ as $N\to\infty$. Applying Lemma 4.2 below with $f:=\varphi_{t_0}-\varphi^N$ we get

$$\begin{aligned}&\sup_{h\in(0,t_0/2)}\left|\int(\varphi_{t_0}-\varphi^N)\frac{\rho_{t_0}-\rho_{t_0-h}}{h}\,\mathrm{d}\mathfrak{m}\right|^2\\ &\qquad\le\sup_{h\in(0,t_0/2)}\frac{1}{h}\int_{t_0-h}^{t_0}\left(\int_{\{|\varphi_{t_0}|>N\}}|\nabla\varphi_{t_0}|_w^2\,\rho_s\,\mathrm{d}\mathfrak{m}\int\frac{|\nabla\rho_s|_w^2}{\rho_s}\,\mathrm{d}\mathfrak{m}\right)\mathrm{d}s,\end{aligned}$$

and hence

$$\limsup_{N\to\infty}\sup_{h\in(0,t_0/2)}\left|\int(\varphi_{t_0}-\varphi^N)\frac{\rho_{t_0}-\rho_{t_0-h}}{h}\,d\mathfrak{m}\right|=0,$$

which is sufficient to conclude, applying the liminf estimate to all functions φ^N and then passing to the limit. □

Lemma 4.2 *With the same notation and assumptions of the previous theorem, for every $f\in L^1(X,\mathfrak{m})$ and $[s,t]\subset(0,\infty)$ it holds*

$$\left|\int f\frac{\rho_t-\rho_s}{t-s}\,d\mathfrak{m}\right|^2\le\frac{1}{t-s}\int_s^t\left(\int|\nabla f|_w^2\,\rho_r\,d\mathfrak{m}\int\frac{|\nabla\rho_s|_w^2}{\rho_s}\,d\mathfrak{m}\right)dr.\qquad(4.2)$$

Proof. Assume first that $f\in L^2(X,\mathfrak{m})$. Then from (2.21) we get

$$\left|\int f\Delta\rho_r\,d\mathfrak{m}\right|^2\le\left(\int|\nabla f|_w\,|\nabla\rho_r|_w\,d\mathfrak{m}\right)^2\le\int|\nabla f|_w^2\,\rho_r\,d\mathfrak{m}\int\frac{|\nabla\rho_r|_w^2}{\rho_r}\,d\mathfrak{m},$$

for all $r>0$, and the thesis follows by integration in (s,t).

For the general case, let $f^N:=\max\{\min\{f,N\},-N\}\in L^2(X,\mathfrak{m})$ be the truncated functions. By Proposition 2.17(i) we know that $\rho_t-\rho_s\in L^\infty(X,\mathfrak{m})$, so that

$$\lim_{N\to\infty}\int f^N\,\frac{\rho_t-\rho_s}{t-s}\,d\mathfrak{m}=\int f\,\frac{\rho_t-\rho_s}{t-s}\,d\mathfrak{m},$$

by dominated convergence. Also, by the chain rule (2.18) we have $|\nabla f^N|_w\le|\nabla f|_w$ $\mathfrak{m}$-a.e. in X. The conclusion follows. □

4.2 Derivative of the entropy along a geodesic

We now look for a formula to bound from below the derivative of the entropy along a geodesic, which is going to be a much harder task compared to Theorem 4.1, due to the lack of a change of variable formula. From the technical point of view, we will need to assume that $(X,\mathsf{d},\mathfrak{m})$ is a strong $CD(K,\infty)$ space, it order to apply the metric Brenier theorem 3.5. From the geometric point of view, the key property that we will use is given by Lemma 4.5 where we relate "horizontal" to "vertical" derivatives. In order to better understand the point, we propose the following simple example.

Example 4.3 Let $\|\cdot\|$ be a smooth, strictly convex norm on $\mathbb{R}^d$ and let $\|\cdot\|_*$ be the dual norm. Let $\mathcal{L}$ be the duality map from $(\mathbb{R}^d,\|\cdot\|)$ to $(\mathbb{R}^d,\|\cdot\|_*)$ and let $\mathcal{L}^*$ be its inverse (respectively, the differentials of the maps $\frac12\|\cdot\|^2$ and $\frac12\|\cdot\|_*^2$). For a smooth map $f:\mathbb{R}^d\to\mathbb{R}$ its differential $Df(x)$ at any point x is intrinsically defined as cotangent vector. To define the gradient $\nabla g(x)$ of a function $g:\mathbb{R}^d\to\mathbb{R}$ (which is a tangent vector), the norm comes into play via the formula $\nabla g(x):=\mathcal{L}^*(Dg(x))$. Notice that the gradient can be characterized without invoking the duality map: first of all one evaluates the slope

$$|\nabla g|(x):=\limsup_{y\to x}\frac{|g(x)-g(y)|}{\|x-y\|}=\|Dg(x)\|_*;\qquad(4.3)$$

then one looks for smooth curves $\gamma : (-\delta, \delta) \to \mathbb{R}^d$ such that

$$\begin{gathered}\gamma_0 = x, \quad \frac{\mathrm{d}}{\mathrm{d}t} g(\gamma_t)_{|t=0} = \|\dot\gamma_0\|^2 = |\nabla g|^2(x). \\ \text{In this case} \quad \nabla g(x) = \dot\gamma_0 \quad \text{and} \quad |\nabla g|(x) = \|\nabla g(x)\|.\end{gathered} \tag{4.4}$$

Now, given two smooth functions f, g, the real number $Df(\nabla g)(x)$ is well defined as the application of the cotangent vector $Df(x)$ to the tangent vector $\nabla g(x)$.

What we want to point out, is that there are in principle and in practice two very different ways of obtaining $Df(\nabla g)(x)$ from a derivation. The first one, maybe more conventional, is the "horizontal derivative":

$$Df(\nabla g)(x) = Df(\dot\gamma_0) = \lim_{t\to 0} \frac{f(\gamma(t)) - f(\gamma_0)}{t}, \qquad \text{where } \gamma \text{ is any curve as in (4.4).}$$

The second one is the "vertical derivative", where we consider perturbations of the slope

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

It coincides with the previous quantity thanks to the "dual" representation (4.3). ■

We emphasize that this relation between horizontal and vertical derivation holds in a purely metric setting: compare the statement of the example with that of Lemma 4.5 below (the plan $\boldsymbol{\pi}$ playing the role of a curve γ as in (4.4), moving points in the direction of $-\nabla g$).

For $\gamma \in \mathrm{AC}^2([0,1];X)$ we set

$$E_t(\gamma) := \sqrt{t \int_0^t |\dot\gamma_s|^2 \,\mathrm{d}s}. \tag{4.5}$$

Notice that $E_t(\gamma)$ reduces to $\mathsf{d}(\gamma_0, \gamma_t)$ if $\gamma \in \mathrm{Geo}(X)$. In the sequel it is tacitly understood that the undetermined ratios of the form

$$\frac{f(\gamma_t) - f(\gamma_0)}{E_t(\gamma)}$$

are set equal to 0 whenever $E_t(\gamma) = 0$, i.e. γ is constant in $[0,t]$.

Recall that he notion of negligible collections of curves in $\mathrm{AC}^2([0,1];X)$ has been introduced in Definition 2.8.

Lemma 4.4 *Let $f : X \to \overline{\mathbb{R}}$ be a Borel function, Sobolev on almost every curve, such that $|\nabla f|_w \in L^2(X, \mathfrak{m})$, and let $\boldsymbol{\pi}$ be a test plan. Then*

$$\limsup_{t\downarrow 0} \int \left| \frac{f(\gamma_t) - f(\gamma_0)}{E_t(\gamma)} \right|^2 \mathrm{d}\boldsymbol{\pi}(\gamma) \le \int |\nabla f|_w^2(\gamma_0)\,\mathrm{d}\boldsymbol{\pi}(\gamma). \tag{4.6}$$

In particular, assume that $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\mu, \nu)$ with μ, $\nu \ll \mathfrak{m}$ with bounded densities, ν with bounded support, $\mu \ge c\mathfrak{m}$ for some $c > 0$ and let φ be a Kantorovich potential relative to it. Then it holds

$$\lim_{t\downarrow 0} \frac{\varphi(\gamma_0) - \varphi(\gamma_t)}{E_t(\gamma)} = |\nabla\varphi|_w(\gamma_0) \qquad \text{in } L^2(\mathrm{Geo}(X), \boldsymbol{\pi}). \tag{4.7}$$

Proof. For any $t \in (0,1)$ and $\boldsymbol{\pi}$-a.e. γ it holds

$$\left|\frac{f(\gamma_t)-f(\gamma_0)}{E_t(\gamma)}\right|^2 \le \frac{\left(\int_0^t |\nabla f|_w(\gamma_s)|\dot\gamma_s|\,\mathrm{d}s\right)^2}{E_t^2(\gamma)} \le \frac{1}{t}\int_0^t |\nabla f|_w^2(\gamma_s)\,\mathrm{d}s. \tag{4.8}$$

Hence

$$\int \left|\frac{f(\gamma_t)-f(\gamma_0)}{E_t(\gamma)}\right|^2 \mathrm{d}\boldsymbol{\pi}(\gamma) \le \frac{1}{t}\iint_0^t |\nabla f|_w^2(\gamma_s)\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}(\gamma) = \int \left(\frac{1}{t}\int_0^t \rho_s\,\mathrm{d}s\right)|\nabla f|_w^2\,\mathrm{d}\mathfrak{m},$$

where ρ_s is the density of $(\mathrm{e}_s)_\sharp\boldsymbol{\pi}$. Now notice that $\rho_t\mathfrak{m} \to \rho_0\mathfrak{m}$ as $t \downarrow 0$ in duality with continuous and bounded functions and that $\sup_t \|\rho_t\|_\infty < \infty$. Hence $\rho_t \to \rho_0$ weakly* in $L^\infty(X,\mathfrak{m})$ and the conclusion follows from the fact that $|\nabla f|_w^2 \in L^1(X,\mathfrak{m})$.

The second part of the statement follows by [2, Theorem 10.3], Proposition 3.5, and the identity $E_t(\gamma) = \mathsf{d}(\gamma_0,\gamma_t)$ since $\boldsymbol{\pi}$ in this case is concentrated on $\mathrm{Geo}(X)$. □

Lemma 4.5 (Horizontal and vertical derivatives) *Let f, $g : X \to \overline{\mathbb{R}}$ be Borel functions, Sobolev on almost every curve, such that both $|\nabla f|_w$ and $|\nabla g|_w$ belong to $L^2(X,\mathfrak{m})$, and let $\boldsymbol{\pi}$ be a test plan. Assume that*

$$\lim_{t\downarrow 0}\frac{g(\gamma_0)-g(\gamma_t)}{E_t(\gamma)} = \lim_{t\downarrow 0}\frac{E_t(\gamma)}{t} = |\nabla g|_w(\gamma_0) \qquad \text{in } L^2\big(\mathrm{AC}^2([0,1];X),\boldsymbol{\pi}\big). \tag{4.9}$$

Then

$$\liminf_{t\downarrow 0}\int \frac{f(\gamma_t)-f(\gamma_0)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma) \ge \limsup_{\varepsilon\downarrow 0}\int \frac{|\nabla g|_w^2(\gamma_0)-|\nabla(g+\varepsilon f)|_w^2(\gamma_0)}{2\varepsilon}\,\mathrm{d}\boldsymbol{\pi}(\gamma). \tag{4.10}$$

Proof. Define functions F_t, $G_t : \mathrm{AC}^2([0,1];X) \to \mathbb{R}$ by

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

By (4.9) we get

$$\lim_{t\downarrow 0}\int G_t^2\,\mathrm{d}\boldsymbol{\pi} = \int |\nabla g|_w^2(\gamma_0)\,\mathrm{d}\boldsymbol{\pi}(\gamma). \tag{4.11}$$

Applying Lemma 4.4 to the function $g+\varepsilon f$ we obtain

$$\begin{aligned}\int |\nabla(g+\varepsilon f)|_w^2(\gamma_0)\,\mathrm{d}\boldsymbol{\pi}(\gamma) &\ge \limsup_{t\downarrow 0}\int \left|\frac{(g+\varepsilon f)(\gamma_0)-(g+\varepsilon f)(\gamma_t)}{E_t(\gamma)}\right|^2 \mathrm{d}\boldsymbol{\pi}(\gamma)\\ &\ge \limsup_{t\downarrow 0}\int \big(G_t^2(\gamma)+2\varepsilon G_t F_t\big)\,\mathrm{d}\boldsymbol{\pi}(\gamma).\end{aligned} \tag{4.12}$$

Subtracting this inequality from (4.11) we get

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

By assumption, we know that $\|G_t - E_t/t\|_2 \to 0$ as $t \downarrow 0$. Also, by Lemma 4.4, we have $\sup_t \|F_t\|_2 < \infty$. Thus it holds

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

□

Before turning to the proof of the estimate of the derivative of the entropy along a geodesic, we need a few more lemmas.

Lemma 4.6 (Upper semicontinuity of right derivative) *Let $f_n : [0,1] \to \mathbb{R}$ be K-convex functions pointwise converging to $f : [0,1] \to \mathbb{R}$. Then*

$$\frac{\mathrm{d}^+}{\mathrm{d}t} f(0) \geq \limsup_{n\to\infty} \frac{\mathrm{d}^+}{\mathrm{d}t} f_n(0).$$

Proof. Just notice that, since $t \mapsto f_n(t) - \frac{1}{2}Kt^2$ are convex functions, it holds

$$\frac{f(t) - f(0) - \frac{1}{2}Kt^2}{t} = \lim_{n\to\infty} \frac{f_n(t) - f_n(0) - \frac{1}{2}Kt^2}{t} \geq \limsup_{n\to\infty} \frac{\mathrm{d}^+}{\mathrm{d}t} f_n(0).$$

Then suffices to pass to the limit as $t \downarrow 0$. □

Lemma 4.7 *Let $(X, \mathsf{d}, \mathfrak{m})$ be a metric measure space, $\boldsymbol{\pi}$ a test plan and $\chi_n : X \to [0,1]$ monotonically convergent to 1. Define the plans $\boldsymbol{\pi}^n := c_n\,(\chi_n \circ \mathrm{e}_0)\,\boldsymbol{\pi}$, where c_n is the normalizing constant. Then*

$$\lim_{n\to\infty} \mathrm{Ent}_{\mathfrak{m}}((\mathrm{e}_t)_\sharp \boldsymbol{\pi}^n) = \mathrm{Ent}_{\mathfrak{m}}((\mathrm{e}_t)_\sharp \boldsymbol{\pi}) \qquad \forall t \in [0,1].$$

Proof. If $\rho_{n,t}$ are the densities w.r.t. $\mathfrak{m}$ of $(\mathrm{e}_t)_\sharp(\chi_n \circ \mathrm{e}_0\, \boldsymbol{\pi}_n)$, by monotone convergence we have $\int \rho_{n,t} \log \rho_{n,t}\, \mathrm{d}\mathfrak{m} \to \mathrm{Ent}_{\mathfrak{m}}((\mathrm{e}_t)_\sharp \boldsymbol{\pi})$. Since $c_n \downarrow 1$ the thesis follows. □

Lemma 4.8 *Let f, $g : X \to \mathbb{R}$ be Sobolev functions along a.e. curve, let J be an interval containing $g(X)$ and let $\phi : J \to \mathbb{R}$ be nondecreasing, Lipschitz and C^1. Then*

$$\lim_{\varepsilon\downarrow 0} \frac{|\nabla(f + \varepsilon\phi(g))|_w^2 - |\nabla f|_w^2}{\varepsilon} = \phi'(g) \lim_{\varepsilon\downarrow 0} \frac{|\nabla(f + \varepsilon g)|_w^2 - |\nabla f|_w^2}{\varepsilon} \qquad \mathfrak{m}\text{-a.e. in } X.$$

Similarly, under the same assumptions on ϕ, if Ch is a quadratic form and $|\nabla f|_w \in L^2(X, \mathfrak{m})$, it holds

$$\lim_{\varepsilon\downarrow 0} \int \frac{|\nabla(\phi(g) + \varepsilon f)|_w^2 - |\nabla\phi(g)|_w^2}{\varepsilon}\, \mathrm{d}\mathfrak{m} = \int \phi'(g) \lim_{\varepsilon\downarrow 0} \frac{|\nabla(g + \varepsilon f)|_w^2 - |\nabla g|_w^2}{\varepsilon}\, \mathrm{d}\mathfrak{m}.$$

Proof. Let us consider the first equality. Notice that it is invariant under addition of constants to ϕ and multiplication of ϕ by positive constants, hence if ϕ is affine the thesis is obvious. In addition, since $|\nabla h|_w = 0$ $\mathfrak{m}$-a.e. in all level sets $h^{-1}(c)$, the formula holds $\mathfrak{m}$-a.e. on any level set of g. Then, by locality, the formula holds if ϕ is countably piecewise affine, i.e. if there is a partition of J in countably many intervals where ϕ is affine. In the general case, thanks to the C^1 regularity of ϕ, for any $\delta > 0$ we can find a countably piecewise affine ϕ_δ such that $\|(\phi - \phi_\delta)'\|_\infty < \delta$ and use the estimate

$$\big||\nabla(f + \varepsilon\phi(g))|_w - |\nabla(f + \varepsilon\phi_\delta(g))|_w\big| \leq \varepsilon|\nabla(\phi - \phi_\delta)(g)|_w \leq \varepsilon\delta|\nabla g|_w$$

to conclude the proof of the first equality.

The second integral equality immediately follows by the integrating the first one, pulling the limit out of the integral in the left hand side using the dominated convergence theorem and finally using the identity

$$Q(u + \varepsilon v) - Q(u) - \varepsilon^2 Q(v) = Q(v + \varepsilon u) - Q(v) - \varepsilon^2 Q(u) \tag{4.13}$$

(with $u = \phi(g)$, $v = f$) satisfied by all quadratic forms Q. □

We are finally ready to prove the main result of this section.

Theorem 4.9 (Derivative of the entropy along a geodesic) *Let $(X, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$ be a strong $CD(K,\infty)$ space and let σ_0, σ_1 be bounded probability densities. Assume that σ_1 has bounded support and that $\sigma_0 \geq c$ for some $c > 0$. Then, for $\boldsymbol{\pi} \in \mathrm{GeoOpt}(\sigma_0\mathfrak{m}, \sigma_1\mathfrak{m})$ as in Definition 3.1, it holds*

$$\lim_{t\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_t) - \mathrm{Ent}_{\mathfrak{m}}(\mu_0)}{t} \geq \frac{\mathsf{Ch}(\varphi) - \mathsf{Ch}(\varphi + \varepsilon\sigma_0)}{\varepsilon} \qquad \forall \varepsilon > 0, \tag{4.14}$$

where $\mu_t := (\mathrm{e}_t)_\sharp \boldsymbol{\pi}$ and φ is any Kantorovich potential relative to $(\sigma_0\mathfrak{m}, \sigma_1\mathfrak{m})$.

Proof. Observe that by Proposition 3.5 we know that $\int |\nabla\varphi|_w^2 \,\mathrm{d}\mu_0 < \infty$, so that our assumptions on μ_0 give $|\nabla\varphi|_w \in L^2(X, \mathfrak{m})$ and the statement makes sense. Also, we can assume $\mathsf{Ch}(\sigma_0) < \infty$, indeed if not the inequality $|\nabla(\varphi + \varepsilon\sigma_0)|_w \geq \varepsilon|\nabla\sigma_0|_w - |\nabla\varphi|_w$ implies $\mathsf{Ch}(\varphi + \varepsilon\sigma_0) = \infty$ and there is nothing to prove. Let $\mathsf{D}(x) := \mathsf{d}(x, \operatorname{supp}\sigma_1)$ and $h_n : [0,\infty) \to [0,\infty)$ be given by

$$h_n(r) := \mathrm{e}^{-9K^-((r-n)^+)^2}.$$

Define the cut-off functions $\chi_n(x) := h_n(\mathsf{D}(x))$, and notice that since the h_n's are equi-Lipschitz, so are the χ_n's.

Notice that $\chi_n \uparrow 1$ in X as $n \to \infty$. Define $\boldsymbol{\pi}^n := c_n\,(\chi_n \circ \mathrm{e}_0)\,\boldsymbol{\pi}$, $c_n \downarrow 1$ being the normalizing constant, and $\mu_t^n := (\mathrm{e}_t)_\sharp\boldsymbol{\pi}^n$. By Lemma 4.7 we have that $\mathrm{Ent}_{\mathfrak{m}}(\mu_t^n) \to \mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ for any $t \in [0,1]$, so that from the fact that $(X, \mathsf{d}, \mathfrak{m})$ is a strong $CD(K,\infty)$ space and Lemma 4.6 we deduce

$$\lim_{t\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_t) - \mathrm{Ent}_{\mathfrak{m}}(\mu_0)}{t} \geq \limsup_{n\to\infty} \lim_{t\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_t^n) - \mathrm{Ent}_{\mathfrak{m}}(\mu_0^n)}{t}. \tag{4.15}$$

Let σ_t^n be the density of μ_t^n. We claim that it holds

$$\lim_{t\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_t^n) - \mathrm{Ent}_{\mathfrak{m}}(\mu_0^n)}{t} \geq \frac{\mathsf{Ch}(\varphi) - \mathsf{Ch}(\varphi + \varepsilon\sigma_0^n)}{\varepsilon} \qquad \forall \varepsilon > 0. \tag{4.16}$$

Indeed, by construction it holds

$$\sigma_0^n(x) = c_n\,\chi_n(x)\,\sigma_0(x) \leq 2\|\sigma_0\|_\infty\,\mathrm{e}^{-9K^-(\mathsf{D}(x)-n)^2} \qquad \text{whenever } D(x) \geq n,$$

for n large enough to ensure $c_n \leq 2$. Thus Proposition 3.4 applies (with $c = 2\|\sigma_0\|_\infty$, $C = R = n$) and we get that $(\mathrm{restr}_0^{1/3})_\sharp\boldsymbol{\pi}^n$ is a test plan. A simple scaling argument in (4.16) based on Lemma 4.8 and the fact that $\frac{1}{3}\varphi$ is a Kantorovich potential for $(\mathrm{restr}_0^{1/3})_\sharp\boldsymbol{\pi}^n$, ensures that with no loss of generality we can assume that $\boldsymbol{\pi}^n$ is a test plan.

Now observe that the convexity of $z \mapsto z\log z$ gives

$$\frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_t^n) - \mathrm{Ent}_{\mathfrak{m}}(\mu_0^n)}{t} \geq \int \log\sigma_0^n \frac{\sigma_t^n - \sigma_0^n}{t}\,\mathrm{d}\mathfrak{m} = \int \frac{\log(\sigma_0^n \circ \mathrm{e}_t) - \log(\sigma_0 \circ \mathrm{e}_0)}{t}\,\mathrm{d}\boldsymbol{\pi}^n. \tag{4.17}$$

By definition, we have $\sigma_0^n = c_n\,\chi_n\sigma_0$ and therefore the inequality $|\nabla\log\chi_n|(x) \leq 18K^-\mathsf{D}(x)$ gives

$$|\nabla\log\sigma_0^n|_w(x) \leq |\nabla\log\chi_n|_w(x) + |\nabla\log\sigma_0|_w(x) \leq 18K^-\mathsf{D}(x) + \frac{1}{\tilde{c}}|\nabla\sigma_0|_w(x),$$

so that the sequence $(|\nabla\log\sigma_0^n|_w)$ is dominated in $L^2(X,\mathfrak{m})$. Here we make the fundamental use of Lemma 4.5: take $f:=\log\sigma_0^n$, $g:=\varphi$ and notice that thanks to Proposition 3.5 and the second part of Lemma 4.4 applied with $\boldsymbol{\pi}$, the assumptions of Lemma 4.5 are satisfied by $\boldsymbol{\pi}^n$. Thus we have

$$\begin{aligned}\liminf_{t\downarrow 0}\int \frac{\log(\sigma_0^n\circ e_t)-\log(\sigma_0\circ e_0)}{t}\,\mathrm{d}\boldsymbol{\pi}^n &\geq \int \frac{|\nabla\varphi|_w^2(\gamma_0)-|\nabla(\varphi+\varepsilon\log\sigma_0^n)|_w^2(\gamma_0)}{2\varepsilon}\,\mathrm{d}\boldsymbol{\pi}^n(\gamma)\\ &= \int \frac{|\nabla\varphi|_w^2-|\nabla(\varphi+\varepsilon\log\sigma_0^n)|_w^2}{2\varepsilon}\sigma_0^n\,\mathrm{d}\mathfrak{m}. \end{aligned}\tag{4.18}$$

The inequality

$$\left|\frac{|\nabla\varphi|_w^2-|\nabla(\varphi+\varepsilon\log\sigma_0^n)|_w^2}{\varepsilon}\right|\leq|\nabla\log\sigma_0^n|_w\big(2|\nabla\varphi|_w+|\nabla\log\sigma_0^n|_w\big)\qquad \varepsilon\in(0,1],$$

and the domination of $(|\nabla\log\sigma_0^n|_w)$ in $L^2(X,\mathfrak{m})$ give that the integrand in the r.h.s. of (4.18) is dominated, so that Lemma 4.8 yields

$$\begin{aligned}\lim_{\varepsilon\downarrow 0}\int\frac{|\nabla\varphi|_w^2-|\nabla(\varphi+\varepsilon\log\sigma_0^n)|_w^2}{\varepsilon}\sigma_0^n\,\mathrm{d}\mathfrak{m} &= \int\lim_{\varepsilon\downarrow 0}\frac{|\nabla\varphi|_w^2-|\nabla(\varphi+\varepsilon\log\sigma_0^n))|_w^2}{\varepsilon}\sigma_0^n\,\mathrm{d}\mathfrak{m}\\ &= \int\lim_{\varepsilon\downarrow 0}\frac{|\nabla\varphi|_w^2-|\nabla(\varphi+\varepsilon\sigma_0^n)|_w^2}{\varepsilon}\,\mathrm{d}\mathfrak{m}\\ &= \lim_{\varepsilon\downarrow 0}\int\frac{|\nabla\varphi|_w^2-|\nabla(\varphi+\varepsilon\sigma_0^n)|_w^2}{\varepsilon}\,\mathrm{d}\mathfrak{m}\\ &= 2\lim_{\varepsilon\downarrow 0}\frac{\mathsf{Ch}(\varphi)-\mathsf{Ch}(\varphi+\varepsilon\sigma_0^n)}{\varepsilon}.\end{aligned}$$

The convexity of Ch now proves the claim (4.16).

To conclude, it is enough to show that $\mathsf{Ch}(\varphi+\varepsilon\sigma_0^n)$ converges to $\mathsf{Ch}(\varphi+\varepsilon\sigma_0)$ as $n\to\infty$. To this aim, let L be a uniform bound on the Lipschitz constants of the χ^n and notice that the inequality (2.19) yields $|\nabla(\chi_n\sigma_0)|_w\leq\chi_n|\nabla\sigma_0|_w+\sigma_0|\nabla\chi_n|_w\leq|\nabla\sigma_0|_w+L\|\sigma\|_\infty$, so that the sequence $(|\nabla(\chi_n\sigma_0)|_w)$ is dominated in $L^2(X,\mathfrak{m})$. Now just observe that by the locality principle (2.16) we have $|\nabla\sigma_0|_w=|\nabla(\chi_n\sigma_0)|_w$ $\mathfrak{m}$-a.e. in $\{\chi_n=1\}$. $\square$

4.3 Quadratic Cheeger's energies

Fix a metric measure space $(X,\mathsf{d}\,\mathfrak{m})\in\mathbb{X}$; without assuming any curvature bound, in this section we apply the tools obtained in Lemma 4.5 to derive useful locality and structural properties on the Cheeger energy in the distinguished case when Ch is a quadratic form on $L^2(X,\mathfrak{m})$. Since Ch is 2-homogeneous and convex, this property is easily see to be equivalent to the parallelogram identity (see for instance [12, Proposition 11.9])

$$\mathsf{Ch}(f+g)+\mathsf{Ch}(f-g)=2\mathsf{Ch}(f)+2\mathsf{Ch}(g)\quad\text{for every }\ f,\,g\in L^2(X,\mathfrak{m}).\tag{4.19}$$

If this is the case we will denote by $\mathcal{E}$ the associated *Dirichlet* form with domain $D(\mathcal{E}):=W^{1,2}(X,\mathsf{d},\mathfrak{m})$, i.e. $\mathcal{E}:[D(\mathcal{E})]^2\to\mathbb{R}$ is the unique bilinear symmetric form satisfying (see e.g. [12, Prop. 11.9])

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

Recall that $W^{1,2}(X,\mathsf{d},\mathfrak{m})=D(\mathsf{Ch})\cap L^2(X,\mathfrak{m})$.

We will occasionally use this density criterion in the theory of linear semigroups.

Lemma 4.10 (Density of invariant sets) *Let $\mathcal{E}$ be the bilinear form associated to a nonnegative and lower semicontinuous quadratic form Q in a Hilbert space H, and let S_t be the associated evolution semigroup. If a subspace $V \subset D(Q)$ is dense for the norm of H and S_t-invariant, then V is also dense in $D(Q)$ for the Hilbert norm $\sqrt{\mathcal{E}(u,u)+(u,u)^2}$.*

Proof. If $u \in D(\mathcal{E})$ satisfies $\mathcal{E}(u,w)+(u,w)=0$ for all $w \in V$, we can choose $w = \mathsf{S}_t v$, $v \in V$, and use the fact that S_t is self-adjoint to get

$$\mathcal{E}(\mathsf{S}_t u, v) + (\mathsf{S}_t u, v) = \mathcal{E}(u, \mathsf{S}_t v) + (u, \mathsf{S}_t v) = 0 \qquad \forall v \in V,\ t > 0.$$

Since $\mathsf{S}_t u$ belongs to the domain of the infinitesimal generator of $\mathsf{S}_t u$, $v \mapsto \mathcal{E}(\mathsf{S}_t u, v)$ is continuous in $D(\mathcal{E})$ for the H norm, hence $\mathcal{E}(\mathsf{S}_t u, v) + (\mathsf{S}_t u, v) = 0$ for all $v \in D(\mathcal{E})$ and $t > 0$. Choosing $v = \mathsf{S}_t u$ and letting $t \downarrow 0$ gives $u = 0$. □

Proposition 4.11 (Properties of $W^{1,2}(X,\mathsf{d},\mathfrak{m})$) *If Ch is quadratic in $L^2(X,\mathfrak{m})$ then $W^{1,2}(X,\mathsf{d},\mathfrak{m})$ endowed with the norm $\sqrt{\|f\|_2^2+\mathcal{E}(f,f)}$ is a separable Hilbert space and Lipschitz functions are dense.*

Proof. We already know that $W^{1,2}(X,\mathsf{d},\mathfrak{m})$ is complete [2, Remark 4.7] and therefore it is a Hilbert space since Ch is quadratic. In particular if $f_n, f \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$ satisfy $\|f_n - f\|_2 \to 0$ and $\mathsf{Ch}(f_n) \to \mathsf{Ch}(f)$ then $f_n \to f$ strongly in $W^{1,2}(X,\mathsf{d},\mathfrak{m})$. In fact, by the parallelogram identity and the $L^2(X,\mathfrak{m})$-lower semicontinuity of Ch

$$\begin{aligned}\limsup_{n\to\infty} \mathsf{Ch}(f-f_n) &= \limsup_{n\to\infty} \Big(2\mathsf{Ch}(f) + 2\mathsf{Ch}(f_n) - \mathsf{Ch}(f+f_n)\Big)\\ &= 4\mathsf{Ch}(f) - \liminf_{n\to\infty} \mathsf{Ch}(f+f_n) \le 4\mathsf{Ch}(f) - \mathsf{Ch}(2f) = 0.\end{aligned}$$

The density of Lipschitz function thus follows by (2.20). The separability of $W^{1,2}(X,\mathsf{d},\mathfrak{m})$ follows considering the invariant set $V := \bigcup_{t>0} \mathsf{H}_t L^2(X,\mathfrak{m})$, which is a subspace thanks to the semigroup property, dense in $W^{1,2}(X,\mathsf{d},\mathfrak{m})$ thanks to Lemma 4.10. Using (2.32) and the separability of $L^2(X,\mathfrak{m})$ it is easy to check that V is separable with respect to the $W^{1,2}(X,\mathsf{d},\mathfrak{m})$ norm, whence the separability of $W^{1,2}(X,\mathsf{d},\mathfrak{m})$ follows. □

The terminology Dirichlet form, borrowed from [16], is justified by the fact that $\mathcal{E}$ is closed (because Ch is $L^2(X,\mathfrak{m})$-lower semicontinuous) and Markovian (by the chain rule (2.18)). Good references on the theory of Dirichlet forms are[16, 17] for locally compact spaces and [28]. The second reference (but see also [17, A.4]), where the theory is extended to infinite-dimensional spaces and even to some classes of non-symmetric forms is more appropriate for us, since we are not assuming local compactness of our spaces.

In this section we analyze the basic properties of this form and relate the energy measure $[f]$ appearing in Fukushima's theory, a kind of localization of $\mathcal{E}$, to $|\nabla f|_w$. Recall that for any $f \in D(\mathcal{E}) \cap L^\infty(X,\mathfrak{m})$ the energy measure $[f]$ is defined by

$$[f](\varphi) := -\mathcal{E}(f, f\varphi) - \mathcal{E}(\frac{f^2}{2}, \varphi) \quad \text{for any } \varphi \in D(\mathcal{E}) \cap L^\infty(X,\mathfrak{m}). \tag{4.20}$$

We shall prove in Theorem 4.19 that $[f] = |\nabla f|_w^2 \mathfrak{m}$. The first step concerns locality, which is not difficult to prove in our setting:

Proposition 4.12 *$\mathcal{E}$ is strongly local:*

$$f,\ g \in D(\mathcal{E}),\ g \text{ constant on } \{f \neq 0\} \qquad \Rightarrow \qquad \mathcal{E}(f,g) = 0. \tag{4.21}$$

Proof. By definition we have

$$2\mathcal{E}(f,g) = \int |\nabla(f+g)|_w^2 - |\nabla f|_w^2 - |\nabla g|_w^2 \,\mathrm{d}\mathfrak{m}. \tag{4.22}$$

By the assumption on g, locality (2.16) and chain rule (2.18) we get that $\mathfrak{m}$-a.e. on $\{f \neq 0\}$. $|\nabla(f+g)|_w = |\nabla f|_w$ and $|\nabla g|_w$ vanishes. On the other hand, $\mathfrak{m}$-a.e. on $X \setminus \{f \neq 0\}$ we have $|\nabla(f+g)|_w = |\nabla g|_w$ and $|\nabla f|_w$ vanishes. □

The identification of $[f]$ with $|\nabla f|_w^2\mathfrak{m}$ requires a deeper understanding of the Leibnitz formula in our context. Our goal is to prove the existence of a bilinear symmetric map from $[D(\mathsf{Ch})]^2$ to $L^1(X,\mathfrak{m})$, that we will denote by $\nabla f \cdot \nabla g$, which gives a pointwise representation of the Dirichlet form, in the sense that

$$\mathcal{E}(f,g) = \int \nabla f \cdot \nabla g \,\mathrm{d}\mathfrak{m} \qquad \forall f,\, g \in W^{1,2}(X,\mathsf{d},\mathfrak{m}).$$

In the spirit of Cheeger's theory [8], we may think to differentials of Lipschitz functions as L^∞ sections of the cotangent bundle to X. Then, if we keep f fixed, the map $g \mapsto \nabla f(x) \cdot \nabla g(x)$ may be interpreted as the action of df on the section induced by g, and we may use this map to define the gradient of f.

We remark that $\nabla f \cdot \nabla g$ could be considered as the "carré du champ" operator $\Gamma(f,g)$ of Γ-calculus in this context, widely used in the study of diffusion semigroups. We adopted the appealing notation $\nabla f \cdot \nabla g$ since in our approach this quantity is directly obtained by a pointwise "vertical" differential of the squared weak upper gradient (like when one tries to recover a scalar product from the squared norm). As a byproduct, we will show that the Laplacian satisfies a suitable formulation of the diffusion condition [5, 1.3], so that useful estimates can be derived by the so called Γ-calculus. An example of application will be given in Section 6.2.

Let $f,\, g \in D(\mathsf{Ch})$ and notice that the inequality

$$|\nabla((1-\lambda)f + \lambda g)|_w \le (1-\lambda)|\nabla f|_w + \lambda|\nabla g|_w \qquad \mathfrak{m}\text{-a.e. in } X,$$

valid for any $\lambda \in [0,1]$, immediately yields that $\varepsilon \mapsto |\nabla(f+\varepsilon g)|_w^2$ satisfies the usual convexity inequality $\mathfrak{m}$-a.e. and $\varepsilon \mapsto \varepsilon^{-1}[|\nabla(f+\varepsilon g)|_w^2 - |\nabla f|_w^2]$ is nondecreasing $\mathfrak{m}$-a.e. in $\mathbb{R} \setminus \{0\}$, in the sense that

$$\frac{|\nabla(f+\varepsilon' g)|_w^2 - |\nabla f|_w^2}{\varepsilon'} \le \frac{|\nabla(f+\varepsilon g)|_w^2 - |\nabla f|_w^2}{\varepsilon} \qquad \mathfrak{m}\text{-a.e. in } X \text{ for } \varepsilon', \varepsilon \in \mathbb{R}\setminus\{0\},\ \varepsilon' \le \varepsilon.$$

Definition 4.13 (The function $\nabla f \cdot \nabla g$) *For $f,\, g \in D(\mathsf{Ch})$ we define $\nabla f \cdot \nabla g$ as*

$$\nabla f \cdot \nabla g := \lim_{\varepsilon \downarrow 0} \frac{|\nabla(f+\varepsilon g)|_w^2 - |\nabla f|_w^2}{2\varepsilon} \tag{4.23}$$

where the limit is understood in $L^1(X,\mathfrak{m})$.

Notice that by monotone convergence and the lower bound obtained by taking a negative ε' in the previous monotonicity formula, the limit in (4.23) exists in $L^1(X,\mathfrak{m})$ along any monotonically decreasing sequence $(\epsilon_i) \subset (0,\infty)$. This obviously implies existence of the full limit as $\varepsilon \downarrow 0$; we also have

$$\nabla f \cdot \nabla f = |\nabla f|_w^2 \qquad \mathfrak{m}\text{-a.e. in } X. \tag{4.24}$$

Notice also that we don't know, yet, whether $(f,g)\mapsto \nabla f\cdot\nabla g$ is symmetric, or bilinear, the only trivial consequence of the definition being the positive homogeneity w.r.t. g. Now we examine the continuity properties of $\nabla f\cdot\nabla g$ with respect to g.

Proposition 4.14 *For f, g, $\tilde g\in D(\mathsf{Ch})$ it holds*

$$\big|\nabla f\cdot\nabla g-\nabla f\cdot\nabla\tilde g\big|\leq|\nabla f|_w|\nabla(g-\tilde g)|_w\qquad\mathfrak{m}\text{-a.e. in } X$$

and, in particular, $\nabla f\cdot\nabla g\in L^1(X,\mathfrak{m})$ and

$$|\nabla f\cdot\nabla g|\leq|\nabla f|_w|\nabla g|_w\qquad\mathfrak{m}\text{-a.e. in } X. \tag{4.25}$$

Proof. It follows from

$$\begin{aligned}&\Big|\big(|\nabla(f+\varepsilon g)|_w^2-|\nabla f|_w^2\big)-\big(|\nabla(f+\varepsilon\tilde g)|_w^2-|\nabla f|_w^2\big)\Big|=\Big||\nabla(f+\varepsilon g)|_w^2-|\nabla(f+\varepsilon\tilde g)|_w^2\Big|\\ =\;&\Big|\big(|\nabla(f+\varepsilon g)|_w-|\nabla(f+\varepsilon\tilde g)|_w\big)\big(|\nabla(f+\varepsilon g)|_w+|\nabla(f+\varepsilon\tilde g)|_w\big)\Big|\\ \leq\;&\varepsilon|\nabla(g-\tilde g)|_w\big(|\nabla(f+\varepsilon g)|_w+|\nabla(f+\varepsilon\tilde g)|_w\big),\end{aligned}$$

dividing by ε, letting $\varepsilon\downarrow0$ and using the strong convergence of $|\nabla(f+\varepsilon g)|_w$ and $|\nabla(f+\varepsilon\tilde g)|_w$ to $|\nabla f|_w$. □

Observe that the second chain rule given in Lemma 4.8 grants that

$$\int\nabla(\phi\circ f)\cdot\nabla g\,\mathrm{d}\mathfrak{m}=\int(\phi'\circ f)\,\nabla f\cdot\nabla g\,\mathrm{d}\mathfrak{m} \tag{4.26}$$

for ϕ nondecreasing and C^1 on an interval containing the image of f.

Proposition 4.15 *For any f, $g\in D(\mathsf{Ch})$ it holds*

$$\mathcal{E}(f,g)=\int\nabla f\cdot\nabla g\,\mathrm{d}\mathfrak{m}. \tag{4.27}$$

Also, we have

$$\nabla f\cdot\nabla g=-\nabla(-f)\cdot\nabla g=-\nabla f\cdot\nabla(-g)=\nabla(-f)\cdot\nabla(-g)\qquad\mathfrak{m}\text{-a.e. in } X. \tag{4.28}$$

Proof. The equality (4.27) follows replacing g by εg into (4.22), dividing by ε and letting $\varepsilon\downarrow0$. To get (4.28) notice that the $\mathfrak{m}$-a.e. convexity of $\varepsilon\mapsto|\nabla(f+\varepsilon g)|_w^2(x)$ in $\mathbb{R}$ yields

$$\nabla f\cdot\nabla g+\nabla f\cdot\nabla(-g)\geq0\qquad\mathfrak{m}\text{-a.e. in } X. \tag{4.29}$$

Since $\mathcal{E}(f,g)=-\mathcal{E}(f,-g)$ we can use (4.27) to obtain that the sum in (4.29) is null $\mathfrak{m}$-a.e. in X. We conclude using the identity $\nabla f\cdot\nabla g=\nabla(-f)\cdot\nabla(-g)$, a trivial consequence of the definition. □

Lemma 4.16 *Let $u\in D(\mathsf{Ch})$ be a given bounded function and let $E_t(\gamma)$ be defined as in (4.5). Then there exists a test plan $\boldsymbol{\pi}$ satisfying $(\mathrm{e}_0)_\sharp\boldsymbol{\pi}=\mathfrak{m}$ and*

$$\lim_{t\downarrow0}\frac{E_t}{t}=\lim_{t\downarrow0}\frac{u\circ\mathrm{e}_0-u\circ\mathrm{e}_t}{E_t}=|\nabla u|_w\circ\mathrm{e}_0\qquad\text{in } L^2\big(\mathrm{AC}^2([0,1];X),\boldsymbol{\pi}\big). \tag{4.30}$$

Proof. Let $\rho_0 := c\,\mathrm{e}^u$, where c is the normalization constant, put $\mu_0 := \rho_0\mathfrak{m}$ and $\rho_t := \mathsf{H}_t(\rho_0)$. Notice that ρ_0 is uniformly bounded away from 0 and ∞ and that $\mathsf{Ch}(\rho_t) \to \mathsf{Ch}(\rho_0)$ as $t \downarrow 0$ implies, by the same Hilbertian argument of Proposition 4.11, strong convergence of $|\nabla\rho_t|_w$ to $|\nabla\rho_0|_w$ in $L^2(X,\mathfrak{m})$. Define the functions A_t, B_t, C_t, $D_t : \mathrm{AC}^2([0,1];X) \to \mathbb{R}$ by (with the usual convention if $E_t(\gamma) = 0$)

$$\begin{aligned}
A_t(\gamma) &:= \frac{\log\rho_0(\gamma_0) - \log\rho_0(\gamma_t)}{t} = \frac{u(\gamma_0) - u(\gamma_t)}{t},\\
B_t(\gamma) &:= \frac{\log\rho_0(\gamma_0) - \log\rho_0(\gamma_t)}{E_t(\gamma)} = \frac{u(\gamma_0) - u(\gamma_t)}{E_t(\gamma)},\\
C_t(\gamma) &:= \frac{E_t(\gamma)}{t},\\
D_t(\gamma) &:= \sqrt{\frac{1}{t}\int_0^t \frac{|\nabla\rho_0|_w^2(\gamma_s)}{\rho_0^2(\gamma_s)}\,\mathrm{d}s} = \sqrt{\frac{1}{t}\int_0^t |\nabla u|_w^2(\gamma_s)\,\mathrm{d}s}.
\end{aligned}$$

Now use [25] to get the existence of a plan $\boldsymbol{\pi} \in \mathscr{P}(\mathrm{AC}^2([0,1];X))$ such that $(\mathrm{e}_t)_\sharp\boldsymbol{\pi} = \mu_t := \rho_t\mathfrak{m}$ for all $t \in [0,1]$ and

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

The maximum principle ensures that $\boldsymbol{\pi}$ is a test plan.

By Lemma 4.17 below we get that $D_t \to |\nabla u|_w \circ \mathrm{e}_0$ in $L^2\big(\mathrm{AC}^2([0,1];X),\boldsymbol{\pi}\big)$. From the second equality in (2.25) we have

$$\begin{aligned}
\lim_{t\downarrow 0}\|C_t\|_2^2 &= \lim_{t\downarrow 0}\frac{1}{t}\int\int_0^t |\dot\gamma_s|^2\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi} = \lim_{t\downarrow 0}\frac{1}{t}\int_0^t |\dot\mu_s|^2\,\mathrm{d}s\\
&= \lim_{t\downarrow 0}\frac{1}{t}\int_0^t\int \frac{|\nabla\rho_s|_w^2}{\rho_s}\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s = \int \frac{|\nabla\rho_0|_w^2}{\rho_0}\,\mathrm{d}\mathfrak{m} = \||\nabla u|_w \circ \mathrm{e}_0\|_2^2,
\end{aligned} \tag{4.31}$$

and from Lemma 4.4 and (4.8) we know that

$$|B_t| \le D_t \qquad \text{and} \qquad \limsup_{t\downarrow 0}\|B_t\|_2^2 \le \||\nabla u|_w \circ \mathrm{e}_0\|_2^2. \tag{4.32}$$

Estimates (4.31) and (4.32) imply

$$\begin{aligned}
\limsup_{t\downarrow 0}\int A_t\,\mathrm{d}\boldsymbol{\pi} &= \limsup_{t\downarrow 0}\int B_tC_t\,\mathrm{d}\boldsymbol{\pi} \le \limsup_{t\downarrow 0}\int |B_t|C_t\,\mathrm{d}\boldsymbol{\pi}\\
&\le \limsup_{t\downarrow 0}\|B_t\|_2\|C_t\|_2 \le \||\nabla u|_w \circ \mathrm{e}_0\|_2^2.
\end{aligned} \tag{4.33}$$

Notice that from the convexity of $z \mapsto z\log z$ we have

$$\int A_t\,\mathrm{d}\boldsymbol{\pi} = \int \frac{\log\rho_0(\rho_0 - \rho_t)}{t}\,\mathrm{d}\boldsymbol{\pi} \ge \frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_0) - \mathrm{Ent}_{\mathfrak{m}}(\mu_t)}{t},$$

and from the first equality in (2.25) we deduce

$$\lim_{t\downarrow 0}\frac{\mathrm{Ent}_{\mathfrak{m}}(\mu_0) - \mathrm{Ent}_{\mathfrak{m}}(\mu_t)}{t} = \lim_{t\downarrow 0}\frac{1}{t}\int_0^t\int \frac{|\nabla\rho_s|_w^2}{\rho_s}\,\mathrm{d}s\mathrm{d}\mathfrak{m} = \int \frac{|\nabla\rho_0|_w^2}{\rho_0}\,\mathrm{d}\mathfrak{m} = \||\nabla u|_w \circ \mathrm{e}_0\|_2^2.$$

Thus from (4.33) we deduce that $\int A_t\,\mathrm{d}\boldsymbol{\pi}$ converges to $\||\nabla u|_w \circ \mathrm{e}_0\|_2^2$ as $t \downarrow 0$. Repeating now (4.33) with $\liminf$ we deduce that also $\|B_t\|_2^2$ converges as $t \downarrow 0$ to $\||\nabla u|_w \circ \mathrm{e}_0\|_2^2$. Now, this convergence, the first inequality in (4.32) and the L^2-convergence of D_t to $|\nabla u|_w \circ \mathrm{e}_0$ yield the L^2-convergence of $|B_t|$ to the same limit.

Finally, from the fact that the first inequality in (4.33) is an equality, we get that also B_t converges to $|\nabla u|_w \circ \mathrm{e}_0$ in $L^2(\boldsymbol{\pi})$. Also, since the second inequality in (4.33) is an equality and (4.31) holds one can conclude that C_t converges to $|\nabla u|_w \circ \mathrm{e}_0$ in $L^2(\boldsymbol{\pi})$ as well.

Thus $\boldsymbol{\pi}$ has all the required properties, except the fact that $(\mathrm{e}_0)_\sharp\boldsymbol{\pi}$ is not $\mathfrak{m}$. To conclude, just replace $\boldsymbol{\pi}$ with $\tilde{c}\rho_0^{-1} \circ \mathrm{e}_0\boldsymbol{\pi}$, $\tilde{c}$ being the renormalization constant. □

Lemma 4.17 *Let $f \in L^1(X, \mathfrak{m})$ be nonnegative and define $F_t : \mathrm{AC}^2([0,1]; X) \to [0, \infty]$, $t \in [0,1]$, by*

$$F_t(\gamma) := \sqrt{\frac{1}{t}\int_0^t f(\gamma_s)\,\mathrm{d}s} \quad t \in (0,1], \qquad F_0 := \sqrt{f(\gamma_0)}.$$

Then $F_t \to F_0$ in $L^2\big(\mathrm{AC}^2([0,1]; X), \boldsymbol{\pi}\big)$ as $t \downarrow 0$ for any test plan $\boldsymbol{\pi}$ whose 2-action $\int\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma)$ is finite.

Proof. To prove the thesis it is sufficient to show that $F_t^2 \to F_0^2$ in $L^1(\mathrm{AC}^2([0,1]; X), \boldsymbol{\pi})$.

Now assume first that f is Lipschitz. In this case the conclusion easily follows from the inequality $|F_t^2(\gamma) - F_0^2(\gamma)| \le \mathrm{Lip}(f)\frac{1}{t}\int_0^t \mathsf{d}(\gamma_0, \gamma_s)\,\mathrm{d}s \le \mathrm{Lip}(f)\int_0^t |\dot\gamma_s|\,\mathrm{d}s$ and the fact that $\int\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,d\boldsymbol{\pi}(\gamma) < \infty$. To pass to the general case, notice that Lipschitz functions are dense in $L^1(X, \mathfrak{m})$ and conclude by the continuity estimates

$$\int \frac{1}{t}\left|\int_0^t h(\gamma_s)\,\mathrm{d}s\right|\mathrm{d}\boldsymbol{\pi} \le \frac{1}{t}\int_0^t\int |h(\gamma_s)|\,\mathrm{d}\boldsymbol{\pi}\,\mathrm{d}s \le C\|h\|_1, \qquad \int |h(\gamma_0)|\,\mathrm{d}\boldsymbol{\pi} \le C\|h\|_1,$$

where $C > 0$ satisfies $(\mathrm{e}_t)_\sharp\boldsymbol{\pi} \le C\mathfrak{m}$ for any $t \in [0,1]$. □

Proposition 4.18 (Leibnitz formula for nonnegative functions) *Let f, g, $h \in D(\mathsf{Ch}) \cap L^\infty(X, \mathfrak{m})$ with g, h nonnegative. Then*

$$\mathcal{E}(f, gh) = \int \nabla f \cdot \nabla(gh)\,\mathrm{d}\mathfrak{m} = \int h\nabla f \cdot \nabla g + g\nabla f \cdot \nabla h\,\mathrm{d}\mathfrak{m}. \tag{4.34}$$

Proof. Notice first that if (g_n), (h_n) are equibounded and converge in $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$ to g, h respectively then (2.19) ensures that $g_n h_n$ converge to gh strongly in $W^{1,2}(X, \mathsf{d}, \mathfrak{m})$. Hence, taking Proposition 4.11 and Proposition 4.14 into account, we can assume with no loss of generality g, h to be bounded, nonnegative and Lipschitz.

Now we apply Lemma 4.16 with $u = f$. The definition of $\nabla f \cdot \nabla(gh)$ and inequality (4.10) gives

$$\begin{aligned}\int \nabla f \cdot \nabla(gh)\,\mathrm{d}\mathfrak{m} &= \liminf_{\varepsilon \downarrow 0}\int \frac{|\nabla(f + \varepsilon gh)|_w^2 - |\nabla f|_w^2}{\varepsilon}\,\mathrm{d}\mathfrak{m} \\ &\ge \limsup_{t \downarrow 0}\int \frac{g(\gamma_0)h(\gamma_0) - g(\gamma_t)h(\gamma_t)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma).\end{aligned}$$

Now observe that the convergence

$$\left|\frac{\big(g(\gamma_t) - g(\gamma_0)\big)\big(h(\gamma_t) - h(\gamma_0)\big)}{t}\right| \le \mathrm{Lip}(g)\,\mathrm{Lip}(h)\frac{\mathsf{d}^2(\gamma_0, \gamma_t)}{t} \to 0 \qquad \text{in } L^1(\boldsymbol{\pi}),$$

ensures that

$$\limsup_{t\downarrow 0}\int \frac{g(\gamma_0)h(\gamma_0)-g(\gamma_t)h(\gamma_t)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma)\geq \liminf_{t\downarrow 0}\int \frac{g(\gamma_0)h(\gamma_0)-g(\gamma_t)h(\gamma_t)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma)$$
$$\geq \liminf_{t\downarrow 0}\int g(\gamma_0)\frac{h(\gamma_0)-h(\gamma_t)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma)+\liminf_{t\downarrow 0}\int h(\gamma_0)\frac{g(\gamma_0)-g(\gamma_t)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma).$$

Now applying inequality (4.10) to the plans $(g\circ \mathrm{e}_0)\,\boldsymbol{\pi}$ and $(h\circ \mathrm{e}_0)\,\boldsymbol{\pi}$ we get

$$\liminf_{t\downarrow 0}\int g(\gamma_0)\frac{h(\gamma_0)-h(\gamma_t)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma)\geq \limsup_{\varepsilon\downarrow 0}\int g\frac{|\nabla f|_w^2-|\nabla(f-\varepsilon h)|_w^2}{\varepsilon}\,\mathrm{d}\mathfrak{m},$$
$$\liminf_{t\downarrow 0}\int h(\gamma_0)\frac{g(\gamma_0)-g(\gamma_t)}{t}\,\mathrm{d}\boldsymbol{\pi}(\gamma)\geq \limsup_{\varepsilon\downarrow 0}\int h\frac{|\nabla f|_w^2-|\nabla(f-\varepsilon g)|_w^2}{\varepsilon}\,\mathrm{d}\mathfrak{m}.$$

Recalling the convergence in $L^1(X,\mathfrak{m})$ of the difference quotients in (4.23) we get

$$\limsup_{\varepsilon\downarrow 0}\int g\frac{|\nabla f|_w^2-|\nabla(f-\varepsilon h)|_w^2}{\varepsilon}\,\mathrm{d}\mathfrak{m}=-\int g\nabla f\cdot\nabla(-h)\mathrm{d}\mathfrak{m}=\int g\nabla f\cdot\nabla h\,\mathrm{d}\mathfrak{m},$$
$$\limsup_{\varepsilon\downarrow 0}\int h\frac{|\nabla f|_w^2-|\nabla(f-\varepsilon g)|_w^2}{\varepsilon}\,\mathrm{d}\mathfrak{m}=-\int h\nabla f\cdot\nabla(-g)\mathrm{d}\mathfrak{m}=\int h\nabla f\cdot\nabla g\,\mathrm{d}\mathfrak{m}.$$

Thus we proved that the inequality $\geq$ always holds in (4.34). Replacing f with $-f$ and using (4.28) once more we get the opposite one and the conclusion. $\square$

Theorem 4.19 (Leibnitz formula and identification of $[f]$) *Let $(X,\mathsf{d},\mathfrak{m})\in\mathbb{X}$ and let us assume that Cheeger's energy Ch is quadratic in $L^2(X,\mathfrak{m})$ as in (4.19). Then*

(i) The map $(f,g)\mapsto\nabla f\cdot\nabla g$ from $[D(\mathsf{Ch})]^2$ to $L^1(X,\mathfrak{m})$ is bilinear, symmetric and satisfies (4.25). In particular it is continuous from $[W^{1,2}(X,\mathsf{d},\mathfrak{m})]^2$ to $L^1(X,\mathfrak{m})$.

(ii) For all f, $g\in D(\mathsf{Ch})$ it holds

$$|\nabla(f+g)|_w^2+|\nabla(f-g)|_w^2=2|\nabla f|_w^2+2|\nabla g|_w^2\quad \mathfrak{m}\text{-a.e. in } X. \tag{4.35}$$

In particular Ch is a quadratic form in $L^1(X,\mathfrak{m})$.

(iii) The Leibnitz formula (4.34) holds with equality and with no sign restriction on g, h.

(iv) The energy measure $[f]$ in (4.20) coincides with $|\nabla f|_w^2\mathfrak{m}$.

Proof. The continuity bound follows at once from (4.25). In order to show symmetry and bilinearity it is sufficient to prove (4.35) when f, $g\in D(\mathsf{Ch})$. In turn, this property follows if we are able to prove that

$$f\mapsto\int h|\nabla f|_w^2\,\mathrm{d}\mathfrak{m}\quad\text{is quadratic in } D(\mathsf{Ch}) \tag{4.36}$$

for all $h\in L^\infty(X,\mathfrak{m})$ nonnegative. Since $D(\mathsf{Ch})\cap L^\infty(X,\mathfrak{m})$ is weakly* dense in $L^\infty(X,\mathfrak{m})$, it is sufficient to prove this property for nonnegative $h\in D(\mathsf{Ch})\cap L^\infty(X,\mathfrak{m})$. Pick $f=g\in$

$D(\mathsf{Ch}) \cap L^\infty(X, \mathfrak{m})$ nonnegative in (4.34) to get

$$\begin{aligned}\int h|\nabla f|_w^2 \,\mathrm{d}\mathfrak{m} &= -\int f\nabla f \cdot \nabla h \,\mathrm{d}\mathfrak{m} + \int \nabla f \cdot \nabla(fh) \,\mathrm{d}\mathfrak{m} \\ &= -\int \nabla(\frac{f^2}{2}) \cdot \nabla h \,\mathrm{d}\mathfrak{m} + \int \nabla f \cdot \nabla(fh) \,\mathrm{d}\mathfrak{m} \\ &= -\mathcal{E}(\frac{f^2}{2}, h) + \mathcal{E}(f, fh),\end{aligned}$$

having used equation (4.26) in the second equality. Now, splitting f in positive and negative parts, we can extend the formula to $D(\mathsf{Ch}) \cap L^\infty(X, \mathfrak{m})$, since $\mathcal{E}$ is bilinear and the strong locality ensures $\mathcal{E}(f^+, f^- h) = 0$, $\mathcal{E}(f^-, f^+ h) = 0$. Both maps $f \mapsto \mathcal{E}(f^2/2, h)$ and $f \mapsto \mathcal{E}(f, fh)$ are immediately seen to be quadratic, so the same is true for $\int |\nabla f|_w^2 h \, d\mathfrak{m}$. Thus we proved that the map $\int h|\nabla f|_w^2 \,\mathrm{d}\mathfrak{m}$ is quadratic on $D(\mathsf{Ch}) \cap L^\infty(X, \mathfrak{m})$. The statement for the full domain $D(\mathsf{Ch})$ follows from a simple truncation argument: if $f^N := \max\{\min\{f, N\}, -N\}$, $g^N := \max\{\min\{g, N\}, -N\} \in L^2(X, \mathfrak{m})$ are the truncated functions, the chain rule (2.18) gives

$$\int h|\nabla(f_N + g_N)|_w^2 \,\mathrm{d}\mathfrak{m} + \int h|\nabla(f_N - g_N)|_w^2 \,\mathrm{d}\mathfrak{m} \le 2\int h|\nabla f|_w^2 \,\mathrm{d}\mathfrak{m} + 2\int h|\nabla g|_w^2.$$

which yields in the limit one inequality. A similar argument applied to $f+g$ and $f-g$ provides the converse inequality and proves (4.36). Finally, we can use the fact that h is arbitrary to prove the pointwise formulation (4.35). $\square$

The property (4.36) shows that if $(X, \mathsf{d}, \mathfrak{m})$ gives raise to a quadratic Cheeger's energy, also $(X, \mathsf{d}, h\mathfrak{m})$ enjoys the same property, provided a uniform bound $0 < c \le h \le c^{-1}$ is satisfied (indeed, we can use [2, Lemma 4.11] to prove that $|\nabla f|_w$ is independent of h). The next result consider the case when h is the characteristic function of a closed subset of X.

Theorem 4.20 *Let $(X, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$ and let $Y \subset X$ be a closed set of positive measure. For $f : Y \to \mathbb{R}$, denote by $|\nabla f|_{w,Y}$ the minimal weak upper gradient of f calculated in the metric measure space $(Y, \mathsf{d}, \mathfrak{m}(Y)^{-1}\mathfrak{m}_{|Y})$. Then:*

(i) Let $f : X \to \mathbb{R}$ Borel and Sobolev along a.e. curve of X, and define $g : Y \to \mathbb{R}$ by $g := f_{|Y}$. Then g is Sobolev along a.e. curve of Y and $|\nabla f|_w = |\nabla g|_{w,Y}$ $\mathfrak{m}$-a.e. in Y.

(ii) Let $g : Y \to \mathbb{R}$ be Borel, Sobolev along a.e. curve in Y and such that $\mathrm{dist}(\mathrm{supp}\, g, \partial Y) > 0$. *Define $f : X \to \mathbb{R}$ by $f_{|Y} := g$ and $f_{|X\setminus Y} := 0$. Then f is Sobolev along a.e. curve of X and $|\nabla f|_w = \chi_Y |\nabla g|_{w,Y}$ $\mathfrak{m}$-a.e. in X.*

(iii) If moreover Ch *is a quadratic form in $L^2(X, \mathfrak{m})$ according to (4.19) and $\mathfrak{m}(\partial Y) = 0$, then*

$$\mathsf{Ch}_Y(f) := \begin{cases} \displaystyle\int_Y |\nabla f|_{w,Y}^2 \,\mathrm{d}\mathfrak{m}_Y & \text{if } f \text{ is Sobolev along a.e. curve in } Y, \\ +\infty & \text{otherwise,} \end{cases}$$

is a quadratic form in $L^2(Y, \mathfrak{m}_Y)$.

Proof. (i) The fact that g is Sobolev along a.e. curve in Y is obvious, since this class of curves is smaller. It is also obvious that $|\nabla g|_{w,Y} \leq |\nabla f|_w$ $\mathfrak{m}$-a.e. in Y, so that to conclude it is sufficient to prove the opposite inequality. Let $G : X \to [0,\infty]$ be defined by

$$G(x) := \begin{cases} |\nabla g|_{w,Y}(x) & \text{if } x \in Y, \\ +\infty & \text{otherwise.} \end{cases} \tag{4.37}$$

Then it is trivial from the definition that G is a weak upper gradient for f in X. Thus, the fact that $|\nabla f|_w$ is the minimal weak upper gradient gives that $|\nabla f|_w \leq G$ $\mathfrak{m}$-a.e. in X, which is the thesis.

(ii) From the hypothesis that g is Sobolev along a.e. curve in Y and supported in a set having positive distance from ∂Y it follows that f is Sobolev along a.e. curve in X. To prove this, if C denotes the support of g, notice first that for any absolutely continuous curve γ the set $L_r := \{t \in [0,1] :\ \mathrm{dist}(\gamma_t, C) = r\}$ is finite for a.e. r (if γ_t is Lipschitz we can apply the coarea inequality, see for instance [14, Corollary 2.10.11], in the general case we can reparameterize γ). Now, setting $R := \mathrm{dist}(C, \partial Y) > 0$ and choosing $r \in (0, R)$ such that L_r is finite, we can use this set of times to split γ in finitely many curves contained in Y and finitely many ones not intersecting C. The equality $|\nabla f|_w = \chi_Y |\nabla g|_{w,Y}$ $\mathfrak{m}$-a.e. in X then follows by locality (in $X \setminus Y$) and by (i) (in Y, since $g = f_{|Y}$).

(iii) Fix $r > 0$, define

$$Y_r := \bigl\{x \in Y :\ \ \mathsf{d}(x, \partial Y) > r\bigr\},$$

so that $Y_r \uparrow Y \setminus \partial Y$ as $r \downarrow 0$, and let $\chi_r : Y \to [0,1]$ be a Lipschitz cut-off function with support contained in $Y \setminus Y_{r/2}$ and identically equal to 1 on Y_r. Notice that, since $\mathfrak{m}(\partial Y) = 0$, to prove the quadratic property of Ch_Y it is sufficient to prove, for all $r > 0$, that $f \mapsto \int_{Y_r} |\nabla f|^2_{w,Y}\, \mathrm{d}\mathfrak{m}_Y$ is quadratic in the class of functions which are Sobolev along a.e. curve in Y. By the previous points and the locality principle (2.16) we know that

$$|\nabla f|_{w,Y} = |\nabla (f\chi_r)|_w = |\nabla f|_w \qquad \mathfrak{m}\text{-a.e. in } Y_r,$$

so that the conclusion follows from (4.35) of Theorem 4.19. $\square$

5 Riemannian Ricci bounds: definition

Let $(X, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$. We say that $(X, \mathsf{d}, \mathfrak{m})$ has *Riemannian Ricci curvature* bounded below by K (in short, a $RCD(K,\infty)$ space) if any of the 3 equivalent conditions of Theorem 5.1 below is fulfilled. Basically, one adds to the strong $CD(K,\infty)$ condition a linearity assumption on the heat flow, stated either at the level of $\mathscr{H}_t$ or at the level of H_t. A remarkable fact is that all these conditions are encoded in the EVI_K property of the gradient flow.

Before stating the theorem we observe that linearity at the level of $\mathscr{H}_t$ will be understood as additivity, namely

$$\mathscr{H}_t((1-\lambda)\mu + \lambda\nu) = (1-\lambda)\mathscr{H}_t(\mu) + \lambda\mathscr{H}_t(\nu) \qquad \forall \mu,\, \nu \in \mathscr{P}_2(X, \mathfrak{m}),\ \lambda \in [0,1].$$

Theorem 5.1 (3 general equivalences) *Let $(X, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$. Then the following three properties are equivalent.*

(i) $(X, \mathsf{d}, \mathfrak{m})$ is a strong $CD(K,\infty)$ space and the semigroup $\mathscr{H}_t$ on $\mathscr{P}_2(X, \mathfrak{m})$ is additive.

(ii) $(X, \mathsf{d}, \mathfrak{m})$ *is a strong* $CD(K,\infty)$ *space and* Ch *is a quadratic form on* $L^2(X,\mathfrak{m})$ *according to* (4.19).

(iii) $(X, \mathsf{d}, \mathfrak{m})$ *is a length space and any* $\mu \in \mathscr{P}_2(X,\mathfrak{m})$ *is the starting point of an* EVI_K *gradient flow of* $\mathrm{Ent}_{\mathfrak{m}}$.

If any of these condition holds, the semigroups H_t *and* $\mathscr{H}_t$ *are also related for all* $t \geq 0$ *by*

$$(\mathsf{H}_t f)\mathfrak{m} = \int f(x)\, \mathscr{H}_t(\delta_x)\, \mathrm{d}\mathfrak{m}(x) \qquad \forall f \in L^2(X,\mathfrak{m}), \tag{5.1}$$

meaning that the signed measure $(\mathsf{H}_t f)\mathfrak{m}$ *is the weighted superposition, with weight* $f(x)$, *of the probability measures* $\mathscr{H}_t(\delta_x)$.

Proof. $(i) \Rightarrow (ii)$. The additivity assumption on the heat semigroup, the identification Theorem 2.16 and the 1-homogeneity of the heat semigroup in $L^2(X,\mathfrak{m})$ (Proposition 2.17), easily yield that the heat semigroup H_t is linear in $L^2(X,\mathfrak{m})$ as well (see the proof of (5.1) below). This further implies that its infinitesimal generator $-\Delta$ is a linear operator, so that $D(\Delta)$ is a linear subspace of $L^2(X,\mathfrak{m})$ and $\Delta : D(\Delta) \to L^2$ is linear. Now, given $f \in D(\mathsf{Ch})$, recall that $t \mapsto \mathsf{Ch}(\mathsf{H}_t(f))$ is continuous on $[0,\infty)$ and locally Lipschitz on $(0,\infty)$, goes to 0 as $t \to \infty$ and $\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Ch}(\mathsf{H}_t(f)) = -\|\Delta\mathsf{H}_t(f)\|_2^2$ for a.e. $t > 0$ (Theorem 2.11), thus

$$\mathsf{Ch}(f) = \int_0^\infty \|\Delta\mathsf{H}_t(f)\|_2^2\, \mathrm{d}t \qquad \forall f \in D(\mathsf{Ch}).$$

Now, recall that quadratic forms can be characterized in terms of the parallelogram identity; thus Ch, being on its domain an integral of the quadratic forms $f \mapsto \|\Delta\mathsf{H}_t(f)\|_2^2$, is a quadratic form.

$(ii) \Rightarrow (iii)$. Using (iii) of Proposition 2.20, to conclude it is sufficient to show that $\mathscr{H}_t(\mu)$ is an EVI_K gradient flow for $\mathrm{Ent}_{\mathfrak{m}}$ for any $\mu \ll \mathfrak{m}$ with density uniformly bounded away from 0 and infinity. Thus, choose $\mu = \rho\mathfrak{m} \in \mathscr{P}_2(X)$ such that $0 < c \leq \rho \leq C < \infty$ and define $\mu_t := \mathscr{H}_t(\mu) = \rho_t\mathfrak{m}$. By Proposition 2.19, in order to check that (μ_t) is an EVI_K gradient flow it is sufficient to pick reference measures ν in (2.28) of the form $\nu = \eta\mathfrak{m} \in \mathscr{P}_2(X)$, with η bounded and with bounded support. For any $t > 0$, choose $\boldsymbol{\pi}_t \in \mathrm{GeoOpt}(\mu_t, \nu)$ and define $\nu_t^s := (\mathrm{e}_s)_\sharp\boldsymbol{\pi}_t$. Since $(X, \mathsf{d}, \mathfrak{m})$ is a strong $CD(K,\infty)$ space it holds

$$\lim_{s\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\nu_t^s) - \mathrm{Ent}_{\mathfrak{m}}(\nu_t^0)}{s} \leq \mathrm{Ent}_{\mathfrak{m}}(\nu) - \mathrm{Ent}_{\mathfrak{m}}(\mu_t) - \frac{K}{2}W_2^2(\mu_t,\nu),$$

thus to conclude it is sufficient to show that for a.e. $t > 0$ it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t\mathfrak{m},\nu) \leq \lim_{s\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\nu_t^s) - \mathrm{Ent}_{\mathfrak{m}}(\nu_t^0)}{s}. \tag{5.2}$$

For any $t > 0$, let φ_t be a Kantorovich potential for (μ_t, ν). By Theorem 4.1 we know that for a.e. $t > 0$ it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}W_2^2(\rho_t\mathfrak{m},\sigma\mathfrak{m}) \leq \lim_{\varepsilon\downarrow 0} \frac{\mathsf{Ch}(\rho_t - \varepsilon\varphi_t) - \mathsf{Ch}(\rho_t)}{\varepsilon},$$

while from Theorem 4.9 we have the lower bound

$$\liminf_{s\downarrow 0} \frac{\mathrm{Ent}_{\mathfrak{m}}(\nu_t^s) - \mathrm{Ent}_{\mathfrak{m}}(\nu_t^0)}{s} \geq \lim_{\varepsilon\downarrow 0} \frac{\mathsf{Ch}(\varphi_t) - \mathsf{Ch}(\varphi_t + \varepsilon\rho_t)}{\varepsilon}.$$

Now we use the hypothesis on Cheeger's energy: recalling (4.13), the fact that Ch is a quadratic form ensures the identity

$$\frac{\mathsf{Ch}(\rho_t - \varepsilon\varphi_t) - \mathsf{Ch}(\rho_t)}{\varepsilon} = \frac{\mathsf{Ch}(\varphi_t) - \mathsf{Ch}(\varphi_t + \varepsilon\rho_t)}{\varepsilon} + O(\varepsilon),$$

(see (4.13)) and therefore (5.2) is proved.
$(iii) \Rightarrow (i)$. By Lemma 5.2 below we deduce that $(X, \mathsf{d}, \mathfrak{m})$ is a strong $CD(K,\infty)$ space. We turn to the additivity. Let (μ^0_t), (μ^1_t) be two EVI_K gradient flows of the relative entropy and define $\mu_t := \lambda\mu^0_t + (1-\lambda)\mu^1_t$, $\lambda \in (0,1)$. To conclude it is sufficient to show that (μ_t) is an EVI_K gradient flow of the relative entropy as well. By (2.7) we know that $(\mu_t) \subset \mathscr{P}_2(X)$ is an absolutely continuous curve, so we need only to show that

$$\limsup_{h\downarrow 0} \frac{\mathrm{e}^{Kh} W_2^2(\mu_{t+h},\nu) - W_2^2(\mu,\nu)}{2h} + \mathrm{Ent}_{\mathfrak{m}}(\mu_t) \le \mathrm{Ent}_{\mathfrak{m}}(\nu), \qquad \forall t > 0. \tag{5.3}$$

Thus, given a reference measure ν, for any $t > 0$ let $\boldsymbol{\gamma}_t \in \mathrm{Opt}(\mu_t,\nu)$, define $\nu^0_t := (\boldsymbol{\gamma}_t)_\sharp\mu^0_t$, $\nu^1_t := (\boldsymbol{\gamma}_t)_\sharp\mu^1_t$ (recall Definition 2.1 and notice that $\mu^0_t, \mu^1_t \ll \mu_t = \pi^1_\sharp\boldsymbol{\gamma}_t$). By equation (2.9) we have

$$\begin{aligned}
W_2^2(\mu_t,\nu) &= \int \mathsf{d}^2(x,y)\,\mathrm{d}\boldsymbol{\gamma}_t(x,y) = \int \mathsf{d}^2(x,y)\Big((1-\lambda)\frac{\mathrm{d}\mu^0_t}{\mathrm{d}\mu_t}(x) + \lambda\frac{\mathrm{d}\mu^1_t}{\mathrm{d}\mu_t}(x)\Big)\mathrm{d}\boldsymbol{\gamma}_t(x,y)\\
&= (1-\lambda)\int \mathsf{d}^2(x,y)\frac{\mathrm{d}\mu^0_t}{\mathrm{d}\mu_t}(x)\,\mathrm{d}\boldsymbol{\gamma}_t(x,y) + \lambda\int \mathsf{d}^2(x,y)\frac{\mathrm{d}\mu^1_t}{\mathrm{d}\mu_t}(x)\,\mathrm{d}\boldsymbol{\gamma}_t(x,y)\\
&= (1-\lambda)W_2^2(\mu^0_t,\nu^0_t) + \lambda W_2^2(\mu^1_t,\nu^1_t),
\end{aligned}$$

while the convexity of W_2^2 yields

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

Hence for any $t \ge 0$ we have

$$\begin{aligned}
\limsup_{h\downarrow 0} \frac{\mathrm{e}^{Kh}\, W_2^2(\mu_{t+h},\nu) - W_2^2(\mu_t,\nu)}{2h} &\le (1-\lambda)\limsup_{h\downarrow 0} \frac{\mathrm{e}^{Kh}\, W_2^2(\mu^0_{t+h},\nu^0_t) - W_2^2(\mu^0_t,\nu^0_t)}{2h}\\
&\quad + \lambda\limsup_{h\downarrow 0} \frac{\mathrm{e}^{Kh}\, W_2^2(\mu^1_{t+h},\nu^1_t) - W_2^2(\mu^1_t,\nu^1_t)}{2h}.
\end{aligned} \tag{5.4}$$

Now we use the assumption that (μ^0_t) and (μ^1_t) are gradient flows in the EVI_K sense: fix t and choose respectively ν^0_t and ν^1_t as reference measures in (2.30) to get

$$\begin{aligned}
\limsup_{h\downarrow 0} \frac{\mathrm{e}^{Kh}\, W_2^2(\mu^0_{t+h},\nu^0_t) - W_2^2(\mu^0_t,\nu^0_t)}{2h} &\le \mathrm{Ent}_{\mathfrak{m}}(\nu^0_t) - \mathrm{Ent}_{\mathfrak{m}}(\mu^0_t),\\
\limsup_{h\downarrow 0} \frac{\mathrm{e}^{Kh}\, W_2^2(\mu^1_{t+h},\nu^1_t) - W_2^2(\mu^1_t,\nu^1_t)}{2h} &\le \mathrm{Ent}_{\mathfrak{m}}(\nu^1_t) - \mathrm{Ent}_{\mathfrak{m}}(\mu^1_t).
\end{aligned} \tag{5.5}$$

Finally, from Proposition 2.3 we have

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t) - \mathrm{Ent}_{\mathfrak{m}}(\nu) \le (1-\lambda)\Big(\mathrm{Ent}_{\mathfrak{m}}(\mu^0_t) - \mathrm{Ent}_{\mathfrak{m}}(\nu^0_t)\Big) + \lambda\Big(\mathrm{Ent}_{\mathfrak{m}}(\mu^1_t) - \mathrm{Ent}_{\mathfrak{m}}(\nu^1_t)\Big). \tag{5.6}$$

Inequalities (5.4), (5.5) and (5.6) yield (5.3).

Finally, we prove (5.1). By linearity we can assume that f is a probability density. Notice that the additivity of the semigroup $\mathscr{H}_t$ gives $\mathscr{H}_t(\sum_i a_i\delta_{x_i}) = \sum_i a_i\mathscr{H}_t(\delta_{x_i})$ whenever $a_i \geq 0$, $\sum_i a_i = 1$ and $x_i \in \operatorname{supp}\mathfrak{m}$. Hence, if f is a continuous probability density in $\operatorname{supp}\mathfrak{m}$ with bounded support, by a Riemann sum approximation we can use the continuity of $\mathscr{H}_t$ to get

$$\mathscr{H}_t(f\mathfrak{m}) = \int f(x)\mathscr{H}_t(\delta_x)\,\mathrm{d}\mathfrak{m}(x)$$

and the identification of gradient flows provides (5.1). By a monotone class argument we extend the validity of the formula from continuous to Borel functions $f \in L^2(X,\mathfrak{m})$. $\square$

Lemma 5.2 *Any $(X,\mathsf{d},\mathfrak{m}) \in \mathbb{X}$ satisfying condition (iii) of Theorem 5.1 is a strong $CD(K,\infty)$ space.*

Proof. Let us first prove that $(X,\mathsf{d},\mathfrak{m})$ is a $CD(K,\infty)$ space. We notice that, since by assumption $(\operatorname{supp}\mathfrak{m},\mathsf{d})$ is a length space, then $\mathscr{P}_2(X,\mathfrak{m})$ is a length metric space. Therefore, up to a suitable reparameterization, for every $\mu_0,\,\mu_1 \in D(\mathrm{Ent}_\mathfrak{m}) \subset \mathscr{P}_2(X,\mathfrak{m})$ and $\varepsilon > 0$ there exists a L_ε-Lipschitz curve $(\mu^\varepsilon) \in \mathrm{Lip}([0,1];\mathscr{P}_2(X,\mathfrak{m}))$ connecting μ_0 to μ_1 with $L_\varepsilon^2 \leq W_2^2(\mu_0,\mu_1) + \varepsilon^2$.

We can thus set $\tilde\mu_s^\varepsilon := \mathscr{H}_\varepsilon(\mu_\varepsilon)$, where $\mathscr{H}_t(\mu)$ denotes the K-gradient flow starting from μ, so that by (2.34) of Proposition 2.21 we get

$$\mathrm{Ent}_\mathfrak{m}(\tilde\mu_s^\varepsilon) \leq (1-s)\mathrm{Ent}_\mathfrak{m}(\mu_0) + s\mathrm{Ent}_\mathfrak{m}(\mu_1) - \frac{K}{2}s(1-s)W_2^2(\mu_0,\mu_1) + \frac{\varepsilon^2}{\mathrm{I}_K(\varepsilon)}$$

Since $\varepsilon^2/\mathrm{I}_K(\varepsilon) \to 0$ as $\varepsilon \downarrow 0$, the family $\{\tilde\mu_s^\varepsilon\}$ has uniformly bounded entropy and therefore it is tight. By (ii) of Proposition 2.20 we know that

$$W_2(\tilde\mu_r^\varepsilon,\tilde\mu_s^\varepsilon) \leq \mathrm{e}^{-K\varepsilon}L_\varepsilon|r-s| \quad \text{for every } r,s \in [0,1],\ \varepsilon > 0.$$

Since $\limsup_{\varepsilon\downarrow 0} L_\varepsilon \leq W_2(\mu_0,\mu_1)$, we can applying the refined Ascoli-Arzelà compactness theorem of [3, Prop. 3.3.1] to find a vanishing sequence $\varepsilon_n \downarrow 0$ and a limit curve $(\mu_s) \subset \mathscr{P}_2(X,\mathfrak{m})$ connecting μ_0 to μ_1 such that

$$\mu_s^{\varepsilon_n} \to \mu_s \quad \text{in } \mathscr{P}(X), \quad W_2(\mu_r,\mu_s) \leq W_2(\mu_0,\mu_1)|r-s|, \quad \mathrm{Ent}_\mathfrak{m}(\mu_s) < \infty \quad \text{for every } r,s \in [0,1].$$

It turns out that (μ_s) is a geodesic in $D(\mathrm{Ent}_\mathfrak{m})$ connecting μ_0 to μ_1 and therefore Proposition 2.21 shows that $\mathrm{Ent}_\mathfrak{m}$ is K-convex along μ_s.

The same Proposition shows that $\mathrm{Ent}_\mathfrak{m}$ is K-convex along any geodesic contained in $\mathscr{P}_2(X,\mathfrak{m})$: in particular, taking any optimal geodesic plan $\boldsymbol{\pi}$ as the one induced by the geodesic obtained by the previous argument, $\mathrm{Ent}_\mathfrak{m}$ satisfies the K-convexity inequality associated to $\boldsymbol{\pi}_F$ as in Definition 3.1, since all the measures $\mu_{F,t}$ belong to $D(\mathrm{Ent}_\mathfrak{m})$. $\square$

6 Riemannian Ricci bounds: properties

6.1 Heat Flow

In this section we study more in detail the properties of the L^2-semigroup H_t in a $RCD(K,\infty)$ space $(X,\mathsf{d},\mathfrak{m}) \in \mathbb{X}$ and the additional informations that one can obtain from the relation (5.1) with the W_2-semigroup $\mathscr{H}_t$. First of all, let us remark that since Ch is quadratic the operator Δ is in fact the infinitesimal generator of H_t, and therefore is linear. Furthermore, denoting by $\mathcal{E}(u,v) : [D(\mathsf{Ch})]^2 \to \mathbb{R}$ the Dirichlet form induced by Ch, the relation $\mathcal{E}(u,v) = -\int v\Delta u \, d\mathfrak{m}$ for $u \in D(\Delta)$, $v \in D(\mathsf{Ch})$ implies that Δ is self-adjoint in $L^2(X,\mathfrak{m})$ and the same is true for H_t.

Also, again by Proposition 2.20, and the definition of $RCD(K,\infty)$ spaces, we know that for any $x \in \operatorname{supp}\mathfrak{m}$ there exists a unique EVI_K gradient flow $\mathscr{H}_t(\delta_x)$ of $\mathrm{Ent}_\mathfrak{m}$ starting from δ_x, related to H_t by (5.1).

Since $\mathrm{Ent}_\mathfrak{m}(\mathscr{H}_t(\delta_x)) < \infty$ for any $t > 0$, it holds $\mathscr{H}_t(\delta_x) \ll \mathfrak{m}$, so that $\mathscr{H}_t(\delta_x)$ has a density, that we shall denote by $\rho_t[x]$. The functions $\rho_t[x](y)$ are the so-called transition probabilities of the semigroup. By standard measurable selection arguments we can choose versions of these densities in such a way that the map $(x,y) \mapsto \rho_t[x](y)$ is $\mathfrak{m}\times\mathfrak{m}$-measurable for all $t > 0$.

In the next theorem we prove additional properties of the flows. The information on both benefits of the identification theorem: for instance the symmetry property of transition probabilities is not at all obvious when looking at $\mathscr{H}_t$ only from the optimal transport point of view, and heavily relies on (5.1), whose proof in turn relies on the identification Theorem 2.16 proved in [2]. On the other hand, the regularizing properties of H_t are deduced by duality by those of $\mathscr{H}_t$, using in particular the contractivity estimate (see (2.31))

$$W_2(\mathscr{H}_t(\mu), \mathscr{H}_t(\nu)) \le e^{-Kt} W_2(\mu,\nu) \qquad t \ge 0,\ \mu,\, \nu \in \mathscr{P}_2(X,\mathfrak{m}). \tag{6.1}$$

and the regularization estimates for the Entropy and its slope (apply (2.32) with $z := \mathfrak{m}$)

$$\mathrm{I}_K(t)\mathrm{Ent}_\mathfrak{m}(\mathscr{H}_t(\mu)) + \frac{(\mathrm{I}_K(t))^2}{2} |\nabla^- \mathrm{Ent}_\mathfrak{m}|^2(\mathscr{H}_t(\mu)) \le \frac{1}{2} W_2^2(\mu,\mathfrak{m}). \tag{6.2}$$

Notice also that (6.1) yields $W_1(\mathscr{H}_t(\delta_x), \mathscr{H}_t(\delta_y)) \le \mathrm{e}^{-Kt}\mathsf{d}(x,y)$ for all x, $y \in \operatorname{supp}\mathfrak{m}$ and $t \ge 0$. This implies that $RCD(K,\infty)$ spaces have Ricci curvature bounded from below by K according to [31], [21]. Notice also that using (5.1), the identification of gradient flows and a simple convexity argument, we can recover the inequality

$$W_1(\mathscr{H}_t(\mu), \mathscr{H}_t(\nu)) \le \mathrm{e}^{-Kt} W_1(\mu,\nu)$$

first with when μ, $\nu \in \mathscr{P}_2(X,\mathfrak{m})$ have densities in $L^2(X,\mathfrak{m})$ and then, by approximation, in the general case when μ, $\nu \in \mathscr{P}_1(X,\mathfrak{m})$.

Notice also that, as a consequence of [33, Theorem 1.3] and the density of Lipschitz functions one has the weak local $(1,1)$-Poincaré inequality

$$\int_{B_r(x)} |u - \bar{u}| \, d\mathfrak{m} \le 4r \int_{B_{2r}(x)} |\nabla u|_w \, d\mathfrak{m} \quad \text{with} \quad \bar{u} := \frac{1}{\mathfrak{m}(B_r(x))} \int_{B_r(x)} u \, d\mathfrak{m}$$

for all $x \in \operatorname{supp}\mathfrak{m}$, $r > 0$, $u \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$, which implies the standard weak local $(1,1)$-Poincaré inequality under doubling assumptions on $\mathfrak{m}$.

Further relevant properties will be obtained in the next section.

Theorem 6.1 (Regularizing properties of the heat flow) *Let $(X,\mathsf{d},\mathfrak{m}) \in \mathbb{X}$ be a $RCD(K,\infty)$ space. Then:*

(i) The transition probability densities are symmetric

$$\rho_t[x](y) = \rho_t[y](x) \qquad \mathfrak{m}\times\mathfrak{m}\text{-a.e. in } X \times X, \text{ for all } t > 0, \tag{6.3}$$

and satisfy for all $x \in X$ the Chapman-Kolmogorov formula:

$$\rho_{t+s}[x](y) = \int \rho_t[x](z)\rho_s[z](y)\,\mathrm{d}\mathfrak{m}(z) \qquad \text{for } \mathfrak{m}\text{-a.e. } y \in X, \text{ for all } t,\ s \geq 0. \tag{6.4}$$

(ii) The formula

$$\tilde{\mathsf{H}}_t f(x) := \int f(y)\,\mathrm{d}\mathscr{H}_t(\delta_x)(y) \qquad x \in \operatorname{supp}\mathfrak{m} \tag{6.5}$$

provides a version of $\mathsf{H}_t f$ for every $f \in L^2(X,\mathfrak{m})$, an extension of H_t to a continuous contraction semigroup in $L^1(X,\mathfrak{m})$ which is pointwise everywhere defined if $f \in L^\infty(X,\mathfrak{m})$.

(iii) The semigroup $\tilde{\mathsf{H}}_t$ maps contractively $L^\infty(X,\mathfrak{m})$ in $\mathrm{C}_b(\operatorname{supp}\mathfrak{m})$ and, in addition, $\tilde{\mathsf{H}}_t f(x)$ belongs to $\mathrm{C}_b\big((0,\infty)\times\operatorname{supp}\mathfrak{m}\big)$.

(iv) If $f : \operatorname{supp}\mathfrak{m} \to \mathbb{R}$ is Lipschitz, then $\tilde{\mathsf{H}}_t f$ is Lipschitz on $\operatorname{supp}\mathfrak{m}$ as well and $\operatorname{Lip}(\tilde{\mathsf{H}}_t f) \leq e^{-Kt}\operatorname{Lip}(f)$.

Proof. (*i*). Fix f, $g \in \mathrm{C}_b(X)$ and notice that (5.1) gives

$$\int g\,\mathsf{H}_t f\,\mathrm{d}\mathfrak{m} = \int f(x)\int g\,\mathrm{d}\mathscr{H}_t(\delta_x)\,\mathrm{d}\mathfrak{m}(x) = \int\int f(x)g(y)\rho_t[x](y)\,\mathrm{d}\mathfrak{m}(y)\,\mathrm{d}\mathfrak{m}(x).$$

Reversing the roles of f and g and using the fact that H_t is self-adjoint it follows that $\int\int(\rho_t[x](y) - \rho_t[y](x))f(x)g(y)\,\mathrm{d}\mathfrak{m}\,\mathrm{d}\mathfrak{m}$ vanishes, and since f and g are arbitrary we obtain (6.3). Formula (6.4) is a direct consequence of the semigroup property $\mathscr{H}_{t+s}(\delta_x) = \mathscr{H}_t(\mathscr{H}_s(\delta_x))$.
(*ii*) Using the symmetry of transition probabilities, for $f \in L^1(X,\mathfrak{m})$ nonnegative we get $\|\tilde{\mathsf{H}}_t f\|_1 = \int\int f(y)\rho_t[y](x)\,\mathrm{d}\mathfrak{m}(x)\,\mathrm{d}\mathfrak{m}(y) = \|f\|_1$. By linearity this shows that $\tilde{\mathsf{H}}_t$ is well defined $\mathfrak{m}$-a.e. and defines a contraction semigroup in $L^1(X,\mathfrak{m})$. The fact that the right hand side in (6.5) provides a version of $\mathsf{H}_t f$ follows once more from (5.1) and the symmetry of transition probabilities.
(iii) Contractivity of $\tilde{\mathsf{H}}_t$ in $L^\infty(X,\mathfrak{m})$ is straightforward. By (6.1) we get that $\mathscr{H}_s(\delta_y) \to \mathscr{H}_t(\delta_x)$ in duality with $\mathrm{C}_b(X)$ when $y \to x$ in X and $s \to t$. Also, the a priori estimate (2.32) shows that $(t,y) \mapsto \mathrm{Ent}_\mathfrak{m}(\mathscr{H}_t(\delta_y))$ is bounded on sets of the form $(\epsilon,\infty)\times B$, with B bounded and $\epsilon > 0$. Thus the family $\{\rho_t[y]\}_{y\in B, t\geq\epsilon}$ is equi-integrable. This shows that $\rho_s[y] \to \rho_t[x]$ weakly in $L^1(X,\mathfrak{m})$ when $(y,s) \to (x,t) \in X\times(0,\infty)$ and proves the continuity of $\tilde{\mathsf{H}}_t f(x)$.
(*iv*) By (6.5) we get $|\tilde{\mathsf{H}}_t f(x) - \tilde{\mathsf{H}}_t f(y)| \leq \operatorname{Lip}(f)W_1(\mathscr{H}_t(\delta_x),\mathscr{H}_t(\delta_y)) \leq \operatorname{Lip}(f)W_2(\mathscr{H}_t(\delta_x),\mathscr{H}_t(\delta_y))$. We can now use (6.1) to conclude (see [23] for a generalization of this duality argument). □

Using Lemma 6.5 below we can refine (*iv*) of Theorem 6.1 and prove by a kind of duality argument [23] a Bakry-Emery estimate in $RCD(K,\infty)$ spaces.

Theorem 6.2 (Bakry-Emery in $RCD(K,\infty)$ spaces) *If $|\nabla f|_w \in L^\infty(X,\mathfrak{m})$ and $t > 0$, then $e^{-Kt}(\tilde{\mathsf{H}}_t|\nabla f|_w^2)^{1/2}$ is an upper gradient of $\tilde{\mathsf{H}}_t f$ on $\operatorname{supp}\mathfrak{m}$, $|\nabla\tilde{\mathsf{H}}_t f| \leq \tilde{\mathsf{H}}_t|\nabla f|_w$ pointwise in $\operatorname{supp}\mathfrak{m}$ and f has a Lipschitz version $\tilde{f} : X \to \mathbb{R}$, with $\operatorname{Lip}(\tilde{f}) \leq \||\nabla f|_w\|_\infty$. Moreover, for any $f \in D(\mathsf{Ch})$ and $t > 0$ we have*

$$|\nabla(\mathsf{H}_t f)|_w^2 \leq \mathrm{e}^{-2Kt}\mathsf{H}_t(|\nabla f|_w^2) \qquad \mathfrak{m}\text{-a.e. in } X. \tag{6.6}$$

Proof. With no loss of generality we can assume, by a truncation argument, that $f \in L^\infty(X,\mathfrak{m})$. Given x, $y \in X$, let $\gamma_s \in \mathrm{AC}^2([0,1];\operatorname{supp}\mathfrak{m})$ be connecting x to y. Given $t > 0$, we can then apply Lemma 6.5 below with $\mu_s := \mathscr{H}_t(\delta_{\gamma_s})$ to get

$$\begin{aligned}|\tilde{\mathsf{H}}_t f(x) - \tilde{\mathsf{H}}_t f(y)| &= \left|\int f \,\mathrm{d}\mathscr{H}_t(\delta_x) - \int f\,\mathrm{d}\mathscr{H}_t(\delta_y)\right| \le \int_0^1 \Bigl(\int |\nabla f|_w^2 \,\mathrm{d}\mu_s\Bigr)^{1/2} |\dot\mu_s|\,\mathrm{d}s \qquad (6.7)\\ &\le \mathrm{e}^{-Kt}\int_0^1 \Bigl(\int |\nabla f|_w^2\,\mathrm{d}\mu_s\Bigr)^{1/2}|\dot\gamma_s|\,\mathrm{d}s = \mathrm{e}^{-Kt}\int_0^1 \bigl(\mathsf{H}_t(|\nabla f|_w^2)(\gamma_s)\bigr)^{1/2}|\dot\gamma_s|\,\mathrm{d}s\end{aligned}$$

(in the last inequality we used the contractivity property, which provides the upper bound on $|\dot\mu_s|$). Notice that we can use the length property of $\operatorname{supp}\mathfrak{m}$ to get, by a limiting argument,

$$|\tilde{\mathsf{H}}_t f(x) - \tilde{\mathsf{H}}_t f(y)| \le \mathsf{d}(x,y)\sup\Bigl\{\mathrm{e}^{-Kt}\bigl(\tilde{\mathsf{H}}_t(|\nabla f|_w^2)\bigr)^{1/2}(z) :\ \mathsf{d}(z,x) \le 2\mathsf{d}(x,y)\Bigr\} \qquad (6.8)$$

for all x, $y \in \operatorname{supp}\mathfrak{m}$. In particular, if $L := \||\nabla f|_w\|_\infty$ is finite, this estimate implies $\mathrm{Lip}(\tilde{\mathsf{H}}_t f) \le L\mathrm{e}^{-Kt}$. Choosing a subsequence $(t_i) \downarrow 0$ such that $\tilde{\mathsf{H}}_{t_i} f \to f$ $\mathfrak{m}$-a.e. we obtain a set $Y \subset \operatorname{supp}\mathfrak{m}$ of full $\mathfrak{m}$-measure such that $f|_Y$ is L-Lipschitz and $\tilde f$ is a L-Lipschitz extension of this restriction to X.

In order to prove (6.6) notice that, if $|\nabla f|_w \in L^\infty(X,\mathfrak{m})$, the continuity of $\tilde{\mathsf{H}}_t|\nabla f|_w^2$ together with (6.8), yield $|\nabla\tilde{\mathsf{H}}_t f|^2 \le \mathrm{e}^{-2Kt}(\tilde{\mathsf{H}}_t(|\nabla f|_w^2))^{1/2}$, stronger than (6.6). To prove (6.6) for general functions $f \in D(\mathsf{Ch})$ we approximate f in the strong $W^{1,2}$ topology by Lipschitz functions and use the stability properties of weak upper gradients. □

Remark 6.3 (Weak Bochner inequality) Following *verbatim* the proof in [19, Theorem 4.6], relative to the Alexandrov case, one can use the Leibnitz rule of Theorem 4.19 and the Bakry-Emery estimate to prove Bochner's inequality (in the case $N = \infty$)

$$\frac{1}{2}\Delta(\nabla f\cdot\nabla f) - \nabla(\Delta f)\cdot\nabla f \ge K\nabla f\cdot\nabla f.$$

in a weak form. Precisely, for all $f \in D(\Delta)$ with $\Delta f \in W^{1,2}(X,\mathsf{d},\mathfrak{m})$ and all $g \in D(\Delta)$ bounded and nonnegative, with $\Delta g \in L^\infty(X,\mathfrak{m})$, it holds:

$$\frac{1}{2}\int_X \Delta g|\nabla f|_w^2\,\mathrm{d}\mathfrak{m} - \int_X g\nabla(\Delta f)\cdot\nabla f\,\mathrm{d}\mathfrak{m} \ge K\int_X g|\nabla f|_w^2\,\mathrm{d}\mathfrak{m}.$$

■

Remark 6.4 (Lipschitz continuity of $\mathsf{H}_t f$ and $\mathscr{H}_t(\mu)$) If we assume the stronger $L^1 \mapsto L^p$ regularization property

$$\|\mathsf{H}_t f\|_p \le C(t)\|f\|_1 \qquad \text{for every } f \in L^2(X,\mathfrak{m}),\ t > 0 \qquad (6.9)$$

for some $p > 1$, then we can improve (6.6) to a pointwise inequality as follows:

$$|\nabla(\tilde{\mathsf{H}}_t f)|^2 \le \mathrm{e}^{-2Kt}\tilde{\mathsf{H}}_t(|\nabla f|_w^2) \qquad \text{in } \operatorname{supp}\mathfrak{m}, \text{ for all } f \in L^2(X,\mathfrak{m}).$$

Indeed, we can first use (6.9) to get, by approximation, $\|\rho_t[x]\|_p \le C(t)$ for all $x \in \operatorname{supp}\mathfrak{m}$. Using the Young inequality for linear semigroups, this gives the implication

$$\|\mathsf{H}_t f\|_{q^*} \le C(t)\|f\|_q \qquad \text{whenever } \frac{1}{q^*} + 1 \le \frac{1}{q} + \frac{1}{p}.$$

Then, choosing $N \geq 1$ so large that $p \geq N/(N-1)$, by iterating this estimate N times the semigroup property yields the $L^1 \mapsto L^\infty$ regularization

$$\sup_{\operatorname{supp}\mathfrak{m}} |\tilde{\mathsf{H}}_t f| \leq (C(t/N))^N \|f\|_1 \qquad \forall f \in L^1(X,\mathfrak{m}). \tag{6.10}$$

Now we can apply the first part of Theorem 6.2 to $u_s := \mathsf{H}_{t-s}f$, whose minimal weak upper gradient is in $L^\infty(X,\mathfrak{m})$, to obtain that $\mathrm{e}^{-Ks}(\tilde{\mathsf{H}}_s(|\nabla u_s|_w^2))^{1/2}$ is an upper gradient of $\tilde{\mathsf{H}}_t f$ and then pass to the limit as $s \downarrow 0$ to obtain that $(\tilde{\mathsf{H}}_t(|\nabla f|_w^2))^{1/2}$ is an upper gradient of $\tilde{\mathsf{H}}_t f$ on $\operatorname{supp}\mathfrak{m}$. Using the length property as in the proof of Theorem 6.2, from this estimate the bound on the slope of $\tilde{\mathsf{H}}_t f$ follows.

In particular we obtain that $\tilde{\mathsf{H}}_t f$ is Lipschitz on $\operatorname{supp}\mathfrak{m}$ for all $t > 0$. Using again the inequality $\|\rho_t[x]\|_\infty \leq C(t)$ for all $x \in \operatorname{supp}\mathfrak{m}$ and the semigroup property we obtain that also $\mathscr{H}_t(\mu)$ has a Lipschitz density for all $\mu \in \mathscr{P}_2(X,\mathfrak{m})$.

The stronger regularizing property (6.9) is known to be true, for instance, if doubling and Poincaré hold in $(X,\mathsf{d},\mathfrak{m})$, see [37, Corollary 4.2]. Notice also that Theorem 6.9 below ensures in any case the Lipschitz regularization, starting from bounded functions. ■

Lemma 6.5 *Let $(X,\mathsf{d},\mathfrak{m})$ be a strong $CD(K,\infty)$ space. Let $\psi \in D(\mathsf{Ch}) \cap L^\infty(X,\mathfrak{m})$ with $|\nabla\psi|_w \in L^\infty(X,\mathfrak{m})$ and let $\mu_t \in \mathrm{AC}^2([0,1];\mathscr{P}_2(X))$ with $\mathrm{Ent}_{\mathfrak{m}}(\mu_t)$ bounded in $[0,1]$. Then*

$$\left|\int_X \psi\,\mathrm{d}\mu_1 - \int_X \psi\,\mathrm{d}\mu_0\right| \leq \int_0^1 \Big(\int_X |\nabla\psi|_w^2\,\mathrm{d}\mu_t\Big)^{1/2} |\dot{\mu}_t|\,\mathrm{d}t. \tag{6.11}$$

Proof. Taking Lemma 3.6 into account we can assume with no loss of generality that $\mu_t \leq C\mathfrak{m}$ for all $t \in [0,1]$, for some constant C. Applying [25], we can find a probability measure $\boldsymbol{\pi}$ in $\mathrm{C}([0,1];X)$ concentrated on $\mathrm{AC}^2([0,1];X) = 0$ and satisfying

$$\mu_t = (\mathrm{e}_t)_\sharp\boldsymbol{\pi} \quad \text{for all } t \in [0,1], \qquad |\dot{\mu}_t|^2 = \int |\dot{\gamma}_t|^2\,\mathrm{d}\boldsymbol{\pi}(\gamma) \quad \text{for a.e. } t \in (0,1). \tag{6.12}$$

Since $\mu_t \leq C\mathfrak{m}$ we obtain that $\boldsymbol{\pi}$ is a test plan, hence ψ is Sobolev along $\boldsymbol{\pi}$-a.e. curve. By the weak upper gradient property we get

$$\begin{aligned}
\left|\int_X \psi\,\mathrm{d}\mu_1 - \int_X \psi\,\mathrm{d}\mu_0\right| &= \left|\int (\psi\circ\mathrm{e}_1 - \psi\circ\mathrm{e}_0))\,\mathrm{d}\boldsymbol{\pi}\right| \leq \int \Big(\int_0^1 |\nabla\psi|_w(\gamma_t)|\dot{\gamma}_t|\,\mathrm{d}t\Big)\,\mathrm{d}\boldsymbol{\pi}(\gamma) \\
&= \int_0^1 \Big(\int |\nabla\psi|_w(\gamma_t)|\dot{\gamma}_t|\,\mathrm{d}\boldsymbol{\pi}(\gamma)\Big)\,\mathrm{d}t \\
&\leq \int_0^1 \Big(\int |\nabla\psi|_w^2(\gamma_t)\,\mathrm{d}\boldsymbol{\pi}(\gamma)\Big)^{1/2}\Big(\int |\dot{\gamma}_t|^2\,\mathrm{d}\boldsymbol{\pi}(\gamma)\Big)^{1/2}\,\mathrm{d}t \\
&= \int_0^1 \Big(\int |\nabla\psi|_w^2\,\mathrm{d}\mu_t\Big)^{1/2} |\dot{\mu}_t|\,\mathrm{d}t.
\end{aligned}$$

□

6.2 Dirichlet form and Brownian motion

In this section we fix a $RCD(K,\infty)$ space $(X,\mathsf{d},\mathfrak{m})$. Recalling that the associated Cheeger's energy is a quadratic form, we will denote by $\mathcal{E}$ the associated *Dirichlet* form as in Section 4.3. In particular Ch satisfies all the properties stated in Theorem 4.19.

Notice also that (see for instance [16, Theorem 5.2.3]) it is not difficult to compute $[f]$ in terms of H_t or in terms of $\mathscr{H}_t(\delta_x)$ by

$$[f]=\lim_{t\downarrow 0}\frac{1}{2t}(f^2+\mathsf{H}_tf^2-2f\mathsf{H}_tf),\qquad [f](\varphi)=\lim_{t\downarrow 0}\frac{1}{2t}\int\int(f(x)-f(y))^2\varphi(y)\,\mathrm{d}\mathscr{H}_t(\delta_x)(y)\,\mathrm{d}\mathfrak{m}(y). \tag{6.13}$$

A direct application of the theory of Dirichlet forms yields the existence of a Brownian motion in $(X,\mathsf{d},\mathfrak{m})$ with continuous sample paths. Continuity of sample paths depends on a locality property, which in our context holds in a particularly strong form, see (4.21).

Theorem 6.6 (Brownian motion) *Let $(X,\mathsf{d},\mathfrak{m})$ be a RCD(K,∞) space. There exists a unique (in law) Markov process $\{\mathbf{X}_t\}_{\{t\geq 0\}}$ in $(\operatorname{supp}\mathfrak{m},\mathsf{d})$ with continuous sample paths in $[0,\infty)$ and transition probabilities $\mathscr{H}_t(\delta_x)$, i.e.*

$$\mathbf{P}\big(\mathbf{X}_{s+t}\in A\big|\mathbf{X}_s=x\big)=\mathscr{H}_t(\delta_x)(A)\qquad\forall s,\,t\geq 0,\ A\ \textit{Borel} \tag{6.14}$$

for $\mathfrak{m}$-a.e. $x\in\operatorname{supp}\mathfrak{m}$.

Proof. Uniqueness in law is obvious, since all finite-dimensional distributions are uniquely determined by (6.14), (6.4) and the Markov property.

First, in the case when (X,d) is not locally compact, we prove a tightness property arguing exactly as in [4, Theorem 1.2], [28, Proposition IV.4.2] (the construction therein uses only distance functions and the inequality $[\mathsf{d}(\cdot,x)]\leq\mathfrak{m}$) to prove a tightness property, namely the existence of a nondecreasing sequence of compact sets $F_n\subset\operatorname{supp}\mathfrak{m}$ satisfying $\mathrm{cap}_{\mathcal{E}}(\operatorname{supp}\mathfrak{m}\setminus F_n)\to 0$ (here $\mathrm{cap}_{\mathcal{E}}$ is the capacity associated to $\mathcal{E}$).

Since $\mathcal{E}$ is a strongly local Dirichlet form, and Lipschitz functions are dense in $D(\mathcal{E})$ for the $W^{1,2}$ norm (Proposition 4.11) we may apply [16, Theorem 4.5.3] in the locally compact case or [28, Theorem IV.3.5, Theorem V.1.5] in the general case to obtain a Markov family $\{\mathbf{P}_x\}_{x\in\operatorname{supp}\mathfrak{m}}$ of probability measures in $\mathrm{C}\big([0,\infty);X\big)$ satisfying

$$\tilde{\mathsf{H}}_tf(x)=\int f(\gamma_t)\,\mathrm{d}\mathbf{P}_x(\gamma)\qquad\text{for all } t\geq 0,\ f\in\mathrm{C}_b(X),\ x\in X\setminus N$$

with $\mathfrak{m}(N)=0$. Then we can take the law $\mathbb{P}:=\int\mathbb{P}_x\,\mathrm{d}\mathfrak{m}(x)$ in $\mathrm{C}\big([0,\infty);X\big)$ and consider the canonical process $\mathbf{X}_t(\gamma)=\gamma(t)$ to obtain the result. □

As a further step we consider the distance induced by the bilinear form $\mathcal{E}$

$$\mathsf{d}_{\mathcal{E}}(x,y):=\sup\{|\tilde g(x)-\tilde g(y)|:\ g\in D(\mathcal{E}),\ [g]\leq\mathfrak{m}\}\qquad\forall(x,y)\in\operatorname{supp}\mathfrak{m}\times\operatorname{supp}\mathfrak{m}, \tag{6.15}$$

which we identify in Theorem 6.8 with d (the function $\tilde g$ is the continuous representative in the Lebesgue class of g, see (v) of Theorem 6.1).

Remark 6.7 In [22] the techniques of [19, 2] are applied to a case slightly different than the one considered here. The starting point of [22] is a Dirichlet form $\mathcal{E}$ on a measure space $(X,\mathfrak{m})$ and X is endowed with the distance $\mathsf{d}_{\mathcal{E}}$. Assuming compactness of $(X,\mathsf{d}_{\mathcal{E}})$, K-geodesic convexity of $\mathrm{Ent}_{\mathfrak{m}}$ in $\mathscr{P}_2(X)$ with cost function $c=d^2_{\mathcal{E}}$, doubling, weak $(1,2)$-Poincaré inequality and the validity of the so-called Newtonian property, the authors prove that the $L^2(X,\mathfrak{m})$ heat flow induced by $\mathcal{E}$ coincides with $\mathscr{H}_t$. The authors also analyze some consequences of this identification, as Bakry-Emery estimates and the short time asymptotic of the heat kernel (a theme discussed neither here nor in [19]). As a consequence of [22, Theorem 5.1] and [2, Theorem 9.3] the Dirichlet form coincides with the Cheeger energy of $(X, \mathsf{d}_{\mathcal{E}}, \mathfrak{m})$ (because their flows coincide). This is a non trivial property, because as shown in [38], a Dirichlet form is not uniquely determined by its intrinsic distance, see also the next result. ■

Theorem 6.8 (Identification of $\mathsf{d}_{\mathcal{E}}$ and d) *The function* $\mathsf{d}_{\mathcal{E}}$ *in* (6.15) *coincides with* d *on* $\operatorname{supp}\mathfrak{m} \times \operatorname{supp}\mathfrak{m}$.

Proof. Choosing $g(z) = \mathsf{d}(z, x)$, since $[g] = |\nabla g|_w^2 \mathfrak{m} \le \mathfrak{m}$ we obtain immediately that $\mathsf{d}_{\mathcal{E}}(x, y) \ge \mathsf{d}(x, y)$ on $\operatorname{supp}\mathfrak{m} \times \operatorname{supp}\mathfrak{m}$. In order to prove the converse inequality we notice that $[g] \le \mathfrak{m}$ implies, by Theorem 6.2, that the continuous representative $\tilde{g}$ has Lipschitz constant less than 1 in X, hence $|\tilde{g}(x) - \tilde{g}(y)| \le \mathsf{d}(x, y)$. □

We conclude this section with an example of application, following the ideas of [5], of the calculus tools developed in Section 4.3 combined with lower Ricci curvature bounds, in particular with the Bakry-Emery estimate (6.2).

Theorem 6.9 (Lipschitz regularization) *If* $f \in L^\infty(X, \mathfrak{m})$ *then* $\tilde{\mathsf{H}}_t f \in \operatorname{Lip}(\operatorname{supp}\mathfrak{m})$ *for every* $t > 0$ *with*

$$\sqrt{\mathrm{I}_{2K}(t)}\,\operatorname{Lip}(\tilde{\mathsf{H}}_t f) \le \|f\|_\infty \quad \textit{for every } t > 0. \tag{6.16}$$

Proof. Let us consider two bounded Lipschitz functions f, φ with φ nonnegative, and let us set

$$G(s) := \int \big(\mathsf{H}_{t-s} f\big)^2\, \mathsf{H}_s \varphi \,\mathrm{d}\mathfrak{m}, \quad G(0) = \int \big(\mathsf{H}_t f\big)^2 \varphi \,\mathrm{d}\mathfrak{m}, \quad G(t) = \int f^2 \mathsf{H}_t \varphi \,\mathrm{d}\mathfrak{m}. \tag{6.17}$$

It is easy to check that G is of class C^1 and, evaluating the derivative of G, we obtain thanks to (4.20)

$$\begin{aligned} G'(s) &= -\mathcal{E}\big((\mathsf{H}_{t-s} f)^2, \mathsf{H}_s\varphi\big) - 2\int \mathsf{H}_{t-s} f\, \Delta \mathsf{H}_{t-s} f\, \mathsf{H}_s \varphi \,\mathrm{d}\mathfrak{m} \\ &= -\mathcal{E}\big((\mathsf{H}_{t-s} f)^2, \mathsf{H}_s\varphi\big) + 2\mathcal{E}\big(\mathsf{H}_{t-s} f, \mathsf{H}_{t-s} f\, \mathsf{H}_s\varphi\big)\,\mathrm{d}\mathfrak{m} = \int |\nabla(\mathsf{H}_{t-s} f)|_w^2 \mathsf{H}_s \varphi \,\mathrm{d}\mathfrak{m}. \end{aligned}$$

Using the fact that H_t is selfadjoint and applying the Bakry-Emery estimate (6.6) we get

$$G'(s) = \int \mathsf{H}_s\Big(|\nabla(\mathsf{H}_{t-s} f)|_w^2\Big)\varphi \,\mathrm{d}\mathfrak{m} \ge \mathrm{e}^{2Ks} \int |\nabla(\mathsf{H}_t f)|_w^2\, \varphi \,\mathrm{d}\mathfrak{m}$$

and an integration in time yields

$$\int \Big(\mathsf{H}_t f^2 - \big(\mathsf{H}_t f\big)^2 - \mathrm{I}_{2K}(t) |\nabla(\mathsf{H}_t f)|_w^2\Big)\varphi \,\mathrm{d}\mathfrak{m} \ge 0. \tag{6.18}$$

Since φ is arbitrary nonnegative, neglecting the term $(\mathsf{H}_t f)^2$ we get an L^∞ bound on $|\nabla(\mathsf{H}_t f)|_w$. We can now use Theorem 6.2 to obtain (6.16). □

By duality one immediately gets:

Corollary 6.10 (W_1-L^1 regularization) *For every* x, $y \in \operatorname{supp}\mathfrak{m}$ *and* $t > 0$ *we have*

$$\sqrt{\mathrm{I}_{2K}(t)} \int \Big|\rho_t[x](z) - \rho_t[y](z)\Big| \,\mathrm{d}\mathfrak{m}(z) \le \mathsf{d}(x, y). \tag{6.19}$$

More generally, the map $\mathsf{h}_t : \mu \mapsto \mathrm{d}\mathscr{H}_t(\mu)/\mathrm{d}\mathfrak{m}$ *satisfies*

$$\sqrt{\mathrm{I}_{2K}(t)}\, \|\mathsf{h}_t \mu - \mathsf{h}_t \nu\|_{L^1(X,\mathfrak{m})} \le W_1(\mu, \nu). \tag{6.20}$$

6.3 Stability

Here we prove that the Riemannian Ricci curvature bounds are stable w.r.t. $\mathbb{D}$-convergence. Notice that we will prove this by showing that condition (*iii*) of Theorem 5.1, namely the EVI property, is stable w.r.t. $\mathbb{D}$-convergence.

Theorem 6.11 (Stability) *Let $(X_n, \mathsf{d}_n, \mathfrak{m}_n) \in \mathbb{X}$, $n \in \mathbb{N}$, be RCD(K, ∞) spaces. If*

$$\lim_{n\to\infty} \mathbb{D}\big((X_n, \mathsf{d}_n, \mathfrak{m}_n), (X, \mathsf{d}, \mathfrak{m})\big) = 0,$$

then $(X, \mathsf{d}, \mathfrak{m})$ is a RCD(K, ∞) space as well.

Proof. We pass to the limit in (*iii*) of Theorem 5.1. By Proposition 2.20 it is sufficient to prove that for any measure $\mu = \rho\mathfrak{m}$ with $\rho \in L^\infty(X, \mathfrak{m})$, there exists a continuous curve (μ_t) on $[0, \infty)$ starting from μ which is locally absolutely continuous on $(0, \infty)$ and satisfies

$$\frac{\mathrm{e}^{K(s-t)}}{2} W_2^2(\mu_s, \nu) - \frac{1}{2} W_2^2(\mu_t, \nu) + \mathrm{I}_K(s-t)\mathrm{Ent}_{\mathfrak{m}}(\mu_s) \le \mathrm{I}_K(s-t)\mathrm{Ent}_{\mathfrak{m}}(\nu) \qquad \forall t \le s, \tag{6.21}$$

for any $\nu \in \mathscr{P}_2(X)$ with bounded density. Let $C := \|\rho\|_\infty$, choose optimal couplings $(\mathbf{d}_n, \boldsymbol{\gamma}_n) \in \mathrm{Opt}((\mathsf{d}, \mathfrak{m}), (\mathsf{d}_n, \mathfrak{m}_n))$ and define $\mu^n := (\boldsymbol{\gamma}_n)_\sharp \mu \in \mathscr{P}_2(X_n)$. Since $(X_n, \mathsf{d}_n, \mathfrak{m}_n)$ is a RCD(K, ∞) space, we know that there exists a curve $t \mapsto \mu^n_t \in \mathscr{P}_2(X_n)$ starting from μ^n such that

$$\frac{\mathrm{e}^{K(s-t)}}{2} W_2^2(\mu^n_s, \nu^n) - \frac{1}{2} W_2^2(\mu^n_t, \nu^n) + \mathrm{I}_K(s-t)\,\mathrm{Ent}_{\mathfrak{m}}(\mu^n_s) \le \mathrm{I}_K(s-t)\,\mathrm{Ent}_{\mathfrak{m}}(\nu^n) \qquad \forall t \le s, \tag{6.22}$$

where $\nu^n := (\boldsymbol{\gamma}_n)_\sharp \nu$. By the maximum principle (Proposition 2.17) we get $\mu^n_t \le C\mathfrak{m}_n$ for any n, t. Also, the energy dissipation equality (2.24) yields that

$$\frac{1}{2} \int_t^s |\dot{\mu}^n_r|^2 dr \le \mathrm{Ent}_{\mathfrak{m}^n}(\mu^n) \le C \log C, \tag{6.23}$$

so that the curves (μ^n_t) are equi-absolutely continuous.

Now, define $\tilde{\mu}^n_t := (\boldsymbol{\gamma}_n)^{-1}_\sharp \mu^n_t \in \mathscr{P}_2(X)$ for any n, t and notice that by (*i*) of Proposition 2.2 we have $\tilde{\mu}^n_t \le C\mathfrak{m}$ for any n, t.

We claim that the set of measures in $\mathscr{P}_2(X)$ which are absolutely continuous w.r.t. $\mathfrak{m}$ and with density bounded bounded above by C is compact w.r.t. W_2. Indeed the measure $\mathfrak{m}$ is tight and, since it has finite second moment, also 2-uniformly integrable. Thus the same is true for the set of measures less than $C\mathfrak{m}$, which is therefore compact (see [3] Section 5.1 for the relevant definitions and properties).

By a diagonal argument we obtain a subsequence $n_k \uparrow \infty$ such that $\tilde{\mu}^{n_k}_t \to \mu_t$ in $(\mathscr{P}_2(X), W_2)$ as $k \to \infty$ for any $t \in \mathbb{Q} \cap [0, \infty)$ and some $\mu_t \in \mathscr{P}_2(X)$. The equi-absolute continuity of the (μ^n_t)'s granted by (6.23), the uniform bound on the densities and the equicontinuity of $(\boldsymbol{\gamma}_n)^{-1}_\sharp$ ((*ii*) of Proposition 2.2) grant that there is convergence for all times to a limit curve $(\mu_t) \subset \mathscr{P}_2(X)$ which is absolutely continuous as well.

To conclude, notice that by (2.11) we have $W_2(\mu^{n_k}_t, \nu^{n_k}) \to W_2(\mu_t, \nu)$ for any $t \in [0, \infty)$, that lower semicontinuity and marginal monotonicity of the entropy yield

$$\mathrm{Ent}_{\mathfrak{m}}(\mu_t) \le \liminf_{k\to\infty} \mathrm{Ent}_{\mathfrak{m}}(\tilde{\mu}^{n_k}_t) \le \mathrm{Ent}_{\mathfrak{m}_{n_k}}(\mu^{n_k}_t)$$

and $\mathrm{Ent}_{\mathfrak{m}_n}(\nu^n) \le \mathrm{Ent}_{\mathfrak{m}}(\nu)$. Thus, we can pass to the limit in (6.22) to get (6.21). □

We remark that it looks much harder to pass to the limit in (ii) of Theorem 5.1, because in general we gain no information about convergence of Cheeger's energies by the $\mathbb{D}$-convergence of the spaces. To see why, just observe that in [39] it has been proved that any space $(X, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$ can be $\mathbb{D}$-approximated by a sequence of finite spaces and that in these spaces Cheeger's energy is trivially null.

6.4 Tensorization

In this section we shall prove the following tensorization property of $RCD(K, \infty)$ spaces:

Theorem 6.12 (Tensorization) *Let* $(X, \mathsf{d}_X, \mathfrak{m}_X)$, $(Y, \mathsf{d}_Y, \mathfrak{m}_Y) \in \mathbb{X}$ *and define the product space* $(Z, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$ *as* $Z := X \times Y$, $\mathfrak{m} := \mathfrak{m}_X \times \mathfrak{m}_Y$ *and*

$$\mathsf{d}\big((x, y), (x', y')\big) := \sqrt{\mathsf{d}_X^2(x, x') + \mathsf{d}_Y^2(y, y')}.$$

Assume that both $(X, \mathsf{d}_X, \mathfrak{m}_X)$ *and* $(Y, \mathsf{d}_Y, \mathfrak{m}_Y)$ *are* $RCD(K, \infty)$ *and nonbranching. Then* $(Z, \mathsf{d}, \mathfrak{m})$ *is* $RCD(K, \infty)$ *and non branching as well.*

The proof of this result is not elementary. Before turning to the details, we comment on the statement of the theorem: the non branching assumption is needed in particular because, up to now, it is not known whether the $CD(K, \infty)$ tensorizes or not: what is known is that the product of two *nonbranching* $CD(K, \infty)$ spaces is $CD(K, \infty)$ [39, Proposition 4.16]. Thus, the result follows combining this tensorization property with another tensorization property at the level of Cheeger's energies, proved in Theorem 6.18, that ensures that Cheeger's energy in Z is a quadratic form. Finally we use the nonbranching assumption once more to show that (Z, d) is nonbranching as well and therefore strong $CD(K, \infty)$ holds.

Throughout this section we assume that the base spaces $(X, \mathsf{d}_X, \mathfrak{m}_X)$, $(Y, \mathsf{d}_Y, \mathfrak{m}_Y)$ are $RCD(K, \infty)$, even though for the proof some intermediate results suffice weaker assumptions.

Keeping the notation of Theorem 6.12 in mind, given $f : Z \to \mathbb{R}$ we shall denote f^x the function $f(x, \cdot)$ and by f^y the function $f(\cdot, y)$. Having in mind Beppo-Levi's pioneering paper [24], we denote by $BL^{1,2}(Z, \mathsf{d}, \mathfrak{m})$ the space of functions $f \in L^2(Z, \mathfrak{m})$ satisfying:

(a) $f^x \in D(\mathsf{Ch}^Y)$ for $\mathfrak{m}_X$-a.e. $x \in X$ and $f^y \in D(\mathsf{Ch}^X)$ for $\mathfrak{m}_Y$-a.e. $y \in X$.

(b) $|\nabla f^y|_w^2(x) \in L^1(Z, \mathfrak{m})$ and $|\nabla f^x|_w^2(y) \in L^1(Z, \mathfrak{m})$.

For any $f \in BL^{1,2}(Z, \mathsf{d}, \mathfrak{m})$ the *cartesian* gradient

$$|\nabla f|_c(x, y) := \sqrt{|\nabla f^y|_w^2(x) + |\nabla f^x|_w^2(y)}$$

is well defined and belongs to $L^2(Z, \mathfrak{m})$.

Accordingly, we shall denote by $\mathsf{Ch}^c : L^2(Z, \mathfrak{m}) \to [0, \infty)$ the quadratic form associated to $|\nabla f|_c$, namely

$$\mathsf{Ch}^c(f) := \int \mathsf{Ch}^X(f^y)\, \mathrm{d}\mathfrak{m}(y) + \int \mathsf{Ch}^Y(f^x)\, \mathrm{d}\mathfrak{m}(x) = \frac{1}{2} \int |\nabla f|_c^2(x, y)\, \mathrm{d}\mathfrak{m}(x, y),$$

if $f \in BL^{1,2}(Z, \mathsf{d}, \mathfrak{m})$, $+\infty$ otherwise. It is not hard to show that the two terms which define Ch^c are $L^2(Z, \mathfrak{m})$-lower semicontinuous, which implies in particular that Ch^c is lower semicontinuous: indeed, considering for instance $\int \mathsf{Ch}^Y(f^x)\, \mathrm{d}\mathfrak{m}(x)$, suffices to check the lower semicontinuity on (fastly) converging sequences satisfying $\sum_n \|f_n - f\|_2^2 < \infty$. By Fubini's theorem these sequences satisfy $\sum_n \|f^x_n - f^x\|_2^2 \,\mathrm{d}\mathfrak{m}(x) < \infty$, so that $f^n_x \to f^x$ in $L^2(Y, \mathfrak{m}_Y)$ for $\mathfrak{m}$-a.e. $x \in X$; then, the lower semicontinuity of Cheeger's functional Ch^Y in the base space Y and Fatou's lemma provide the lower semicontinuity (the same argument applies to $\int \mathsf{Ch}^X(f^y)\,\mathrm{d}\mathfrak{m}(y)$).

Lemma 6.13 *If $f \in \mathrm{Lip}(Z)$ then $|\nabla f|_w \le \sqrt{|\nabla f^x|^2 + |\nabla f^y|^2}$ $\mathfrak{m}$-a.e. in Z. In particular*

$$|\nabla f|_w \le \sqrt{g_1^2 + g_2^2} \qquad \mathfrak{m}\text{-a.e. in } Z \tag{6.24}$$

whenever g_1, $g_2 : Z \to \mathbb{R}$ are bounded Borel functions such that $g_1(x,\cdot)$ is a upper semicontinuous upper gradient of f^x and $g_2(\cdot,y)$ is a upper semicontinuous upper gradient of f^y.

Proof. We will prove that the cartesian slope $\sqrt{|\nabla f^x|^2 + |\nabla f^y|^2}$ is a weak upper gradient for Lipschitz functions f. If $\gamma = (\gamma^X, \gamma^Y) \in \mathrm{AC}^2([0,1]; Z)$ we need to prove that for a.e. t the inequality

$$|\frac{\mathrm{d}}{\mathrm{d}t}(f\circ\gamma)|(t) \le \sqrt{|\nabla f^{\gamma^X_t}|^2(\gamma^Y_t) + |\nabla f^{\gamma^Y_t}|^2(\gamma^X_t)}\sqrt{|\dot\gamma^X_t|^2 + |\dot\gamma^Y_t|^2} \tag{6.25}$$

holds. A pointwise proof of this inequality seems not to be easy, on the other hand, working at the level of distributional derivatives, in [3, Lemma 4.3.4] it is proved that a.e. in $[0,1]$ it holds

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

so that

$$|\frac{\mathrm{d}}{\mathrm{d}t}(f\circ\gamma)|(t) \le |\nabla f^{\gamma^Y_t}|(\gamma^X_t)|\dot\gamma^X_t| + |\nabla f^{\gamma^X_t}|(\gamma^Y_t)|\dot\gamma^Y_t| \qquad \text{a.e. in } [0,1],$$

from which (6.25) readily follows. The estimate (6.24) follows noticing that any upper semicontinuous upper gradient bounds the slope from below. □

In the next lemma we will improve the inequality $|\nabla f|_w \le \sqrt{|\nabla f^x|^2 + |\nabla f^y|^2}$ obtaining $|\nabla f|_c$ in the right hand side. To this aim we consider, as a regularizing operator, the product semigroup $\tilde{\mathsf{H}}^c_t$ in $L^2(Z, \mathfrak{m})$, pointwise defined by

$$\tilde{\mathsf{H}}^c_t f(x,y) := \int\int f(x', y')\rho^X_t[x](x')\rho^Y_t[y](y')\,\mathrm{d}\mathfrak{m}_X(x')\,\mathrm{d}\mathfrak{m}_Y(y') \tag{6.26}$$

where $\rho^X_t[x](x')$ and $\rho^Y_t[y](y')$ are the transition probability densities in the base spaces (see also (6.28) below for an equivalent description in terms of iterated operators). It is easy to show that $\tilde{\mathsf{H}}_t$ retains the same properties of its "factors" $\tilde{\mathsf{H}}^X_t$, $\tilde{\mathsf{H}}^Y_t$ in the base spaces, in particular it is mass preserving, self-adjoint, satisfies the maximum principle, regularizes from $L^\infty(Z, \mathfrak{m})$ to $\mathrm{C}_b(Z)$ and leaves $\mathrm{Lip}(Z)$ invariant. In addition, $\tilde{\mathsf{H}}_t$ can also be viewed as the $L^2(Z, \mathfrak{m})$-gradient flow of Ch^c, namely the solution to

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t = \Delta^c f_t \tag{6.27}$$

where the linear operator Δ^c is defined in terms of the Laplacians in the base spaces Δ_X, Δ_Y by $\Delta^c f(x,y) := \Delta_X f^y(x) + \Delta_Y f^x(y)$.

Lemma 6.14 *For all $f \in \mathrm{Lip}(Z)$ it holds $|\nabla f|_w \le |\nabla f|_c$ $\mathfrak{m}$-a.e. in Z.*

Proof. Set $F_1(x,y) := |\nabla f^y|_w(x)$ and $F_2(x,y) := |\nabla f^x|_w(y)$. We consider the regularization $f_t := \tilde{\mathsf{H}}_t^c f$ of f. Writing

$$f_t(x,y) = \tilde{\mathsf{H}}_t^X G(\cdot,y)(x) \tag{6.28}$$

with $G(x',y) := \tilde{\mathsf{H}}_t^Y f(x',\cdot)(y)$, we can use first Theorem 6.2 and then the convexity of $g \mapsto |\nabla g|_w$ to get

$$\begin{aligned} |\nabla f_t^y|(x) &\le \mathrm{e}^{-Kt}\big(\tilde{\mathsf{H}}_t^X |\nabla G(\cdot,y)|_w^2\big)^{1/2}(x) \\ &\le \mathrm{e}^{-Kt}\big(\tilde{\mathsf{H}}_t^X \tilde{\mathsf{H}}_t^Y F_1^2\big)^{1/2}(x,y) \\ &= \mathrm{e}^{-Kt}\big(\tilde{\mathsf{H}}_t^c F_1^2\big)^{1/2}(x,y). \end{aligned}$$

Analogously, reversing the role of the variables we get

$$|\nabla f_t^x|(y) \le \mathrm{e}^{-Kt}\big(\tilde{\mathsf{H}}_t^c F_2^2\big)^{1/2}.$$

So, we may take $g_1(x,y) := \mathrm{e}^{-Kt}(\tilde{\mathsf{H}}_t^c F_1^2)^{1/2}$ and $g_2 := \mathrm{e}^{-Kt}(\tilde{\mathsf{H}}_t^c F_2^2)^{1/2}$ in Lemma 6.13 to get

$$|\nabla f_t|_w^2 \le \mathrm{e}^{-2Kt}\tilde{\mathsf{H}}_t^c(F_1^2 + F_2^2) = \tilde{\mathsf{H}}_t^c |\nabla f|_c^2 \qquad \mathfrak{m}\text{-a.e. in } Z.$$

Letting $t \downarrow 0$ the stability property of weak upper gradients and the strong continuity of the semigroup provide the result. □

The proof of the converse inequality is more involved. It rests mainly in an improvement in product spaces of the Hamilton-Jacobi inequality satisfied by the Hopf-Lax semigroup (see Lemma 6.15 below) and on its consequence, an improved metric derivative that we obtain in Lemma 6.16 along solutions to the $L^2(Z,\mathfrak{m})$-gradient flow of Ch^c defined in (6.26) or, equivalently, in (6.27).

In [2, Section 3], a very detailed analysis of the differentiability properties of the Hopf-Lax semigroup

$$Q_t g(w) := \inf_{w' \in W} g(w') + \frac{1}{2t}\mathsf{d}_W^2(w',w) \tag{6.29}$$

in a metric space (W, d_W) has been made. The analysis is based on the quantities

$$D_g^+(w,t) := \sup \limsup_{n\to\infty} \mathsf{d}_W(w,w_n'), \qquad D_g^-(w,t) := \inf \liminf_{n\to\infty} \mathsf{d}_W(w,w_n),$$

where the supremum and the infimum run among all minimizing sequences (w_n) in (6.29). These quantities reduce respectively to the maximum and minimum distance from w of minimizers in the locally compact case. Confining for simplicity our discussion to the case of bounded functions, which suffices for our purposes, it has been shown that D_g^+ and D_g^- are respectively upper and lower semicontinuous in $W \times (0,\infty)$, that $D_g^-(\cdot,t)/t$ is an upper gradient of $Q_t g$ and that the following pointwise equality holds:

$$\frac{\mathrm{d}^+}{\mathrm{d}t} Q_t g(w) + \frac{(D_g^+(w,t))^2}{2t^2} = 0, \tag{6.30}$$

where we recall that $\mathrm{d}^+/\mathrm{d}t$ stands for right derivative (part of the statement is its existence at every point). Notice that, since $D_g^+(\cdot,t)/t \ge D_g^-(\cdot,t)/t$ is an upper semicontinuous upper gradient of $Q_t g$, (6.30) implies the Hamilton-Jacobi subsolution property $\frac{\mathrm{d}^+}{\mathrm{d}t} Q_t g + |\nabla Q_t g|^2/2 \le 0$, but in the sequel we shall need the sharper form (6.30).

Lemma 6.15 *Let $g : Z \to \mathbb{R}$ be a bounded function. Then, for all $t > 0$ the function $Q_t g$ satisfies*

$$\frac{\mathrm{d}^+}{\mathrm{d}t}Q_t g + \frac{1}{2}|\nabla Q_t g|_c^2 \le 0 \qquad \mathfrak{m}\text{-}a.e.\ in\ Z. \tag{6.31}$$

Proof. Taking (6.30) into account and the definition of $|\nabla Q_t g|_c$ (recall the notation $f^x(y) = f(x,y) = f^y(x)$), suffices to show that for all $t > 0$ it holds

$$\frac{[D_g^+((x,y),t)]^2}{t^2} \ge |\nabla (Q_t g)^y|_w^2(x) + |\nabla (Q_t g)^y|_w^2(x) \qquad \mathfrak{m}\text{-a.e. in } Z. \tag{6.32}$$

In order to prove (6.32), notice that we can minimize first in one variable and then in the other one to get

$$(Q_t g)^y(x) = Q_t^X(L_{t,y})(x), \qquad (Q_t g)^x(y) = Q_t^Y(R_{t,x})(y), \tag{6.33}$$

where $L_{t,y}(x') := Q_t^Y g(x',\cdot)(y)$ and $R_{t,x}(y') := Q_t^X g(\cdot,y')(x)$. Since $D^-(\cdot,t)/t$ is is an upper gradient, we see that (6.32) is a consequence of the pointwise inequality

$$[D_g^+((x,y),t)]^2 \ge [D_{L_{t,y}}^-(x,t)]^2 + [D_{R_{t,x}}^-(y,t)]^2. \tag{6.34}$$

In order to prove (6.34), let us consider a minimizing sequence (x_n, y_n) for $Q_t g(x,y)$; since

$$\begin{aligned} Q_t g(x,y) &= \lim_{n\to\infty} g(x_n,y_n) + \frac{1}{2t}\mathsf{d}_Y^2(y_n,y) + \frac{1}{2t}\mathsf{d}_X^2(x_n,x) \\ &\ge \liminf_{n\to\infty} L_t^y(x_n) + \frac{1}{2t}\mathsf{d}_X^2(x_n,x) \ge Q_t^X(L_{t,y})(x) \end{aligned}$$

we can use (6.33) to obtain that all inequalities are equalities: this implies that the liminf is a limit and that and that (x_n) is a minimizing sequence for $Q_t^X\varphi(x)$, with $\varphi(x) = L_{t,y}(x)$. Analogously, (y_n) is a minimizing sequence for $Q_t^Y\psi(y)$, where $\psi(y) = R_{t,x}(y)$. Taking into account the definitions of $D^\pm$, this yields (6.34). □

Lemma 6.16 (Kuwada's lemma in product spaces) *Let $f \in L^\infty(Z,\mathfrak{m})$ be a probability density and let f_t be the solution of the L^2-gradient flow of Ch^c starting from f. Then $\mu_t = f_t\mathfrak{m} \in \mathscr{P}_2(X)$ for all $t \ge 0$ and*

$$|\dot\mu_t|^2 \le \int_{\{f_t>0\}} \frac{|\nabla f_t|_c^2}{f_t}\,\mathrm{d}\mathfrak{m} \qquad \text{for a.e. } t > 0. \tag{6.35}$$

Proof. The proof can be achieved following *verbatim* the proof of the analogous result [2, Lemma 6.1], this time working with $|\nabla f_t|_c$ in place of $|\nabla f_t|_w$: this replacement is possible in view of the improved Hamilton-Jacobi inequality (6.31) and of the calculus rules

$$-\int g\Delta^c f\,\mathrm{d}\mathfrak{m} \le \int |\nabla f|_c|\nabla g|_c\,\mathrm{d}\mathfrak{m}, \qquad -\int \phi(f)\Delta^c f\,\mathrm{d}\mathfrak{m} = \int \phi'(f)|\nabla f|_c^2\,\mathrm{d}\mathfrak{m}, \tag{6.36}$$

which follow immediately by the analogous properties of the partial Laplacians. □

Proposition 6.17 *We have $D(\mathsf{Ch}) \subset BL^{1,2}(Z,\mathsf{d},\mathfrak{m})$. In addition, for all $f \in D(\mathsf{Ch})$ there exist $f_n \in D(\mathsf{Ch}^c)$ converging to f in $L^2(Z,\mathfrak{m})$ and satisfying*

$$\limsup_{n\to\infty} \mathsf{Ch}^c(f_n) \le \mathsf{Ch}(f). \tag{6.37}$$

Proof. We argue exactly as in [2, Theorem 6.2], where we identify weak upper gradients and relaxed gradients, the only difference being the use of the gradient flow of Ch^c and the improved estimate (6.35).

Pick $f \in D(\mathsf{Ch})$. With a truncation argument, we can assume that $c^{-1} \geq f \geq c > 0$ $\mathfrak{m}$-almost everywhere in Z with $\int f^2 \,\mathrm{d}\mathfrak{m} = 1$. We consider the gradient flow (h_t) of Ch^c with initial datum $h := f^2$, setting $\mu_t = h_t\mathfrak{m}$, and we apply Lemma 6.16. The maximum principle yields $c^{-1} \geq f_t \geq c$ and a standard argument based on (6.27) and (6.36) yields the energy dissipation identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\int f_t \log f_t \,\mathrm{d}\mathfrak{m} = -\int \frac{|\nabla f_t|_c^2}{f_t}\,\mathrm{d}\mathfrak{m}. \tag{6.38}$$

Let $g = h^{-1}|\nabla h|_w$, notice that by the chain rule we know that $\log h$ is Sobolev along almost every curve and use the same argument of [2, Theorem 6.2] to get

$$\int \big(h\log h - h_t \log h_t\big)\,\mathrm{d}\mathfrak{m} \leq \int \log h(h-h_t)\,\mathrm{d}\mathfrak{m} \leq \Big(\int_0^t\int g^2 h_s \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s\Big)^{1/2}\Big(\int_0^t |\dot\mu_s|^2\,\mathrm{d}s\Big)^{1/2}.$$

Now, inequality (6.35) gives

$$\begin{aligned}\int \big(h\log h - h_t \log h_t\big)\,\mathrm{d}\mathfrak{m} &\leq \frac{1}{2}\int_0^t\int g^2 h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \frac{1}{2}\int_0^t |\dot\mu_s|^2\,\mathrm{d}s\\ &\leq \frac{1}{2}\int_0^t\int g^2 h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \frac{1}{2}\int_0^t \int \frac{|\nabla h_s|_c^2}{h_s}\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s.\end{aligned}$$

Recalling the entropy dissipation formula (6.38) we obtain

$$\int_0^t\int \frac{|\nabla h_s|_c^2}{h_s}\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s \leq \int_0^t\int g^2 h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s.$$

Now, the chain rule and the identity $g = 2f^{-1}|\nabla f|_w$ give $\int_0^t \mathsf{Ch}^c(\sqrt{h_s})\,\mathrm{d}s \leq \int_0^t\int |\nabla f|_w^2 f^{-2} h_s\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s$, so that dividing by t and passing to the limit as $t \downarrow 0$ we get (6.37), since $\sqrt{h_s}$ are equibounded and converge strongly to f in $L^2(Z,\mathfrak{m})$ as $s \downarrow 0$. ☐

Theorem 6.18 *Let $f \in L^2(Z,\mathfrak{m})$. Then $f \in D(\mathsf{Ch})$ if and only if $f \in D(\mathsf{Ch}^c)$ and $|\nabla f|_w = |\nabla f|_c$ $\mathfrak{m}$-a.e. in Z. In particular $\mathsf{Ch} = \mathsf{Ch}^c$ is a quadratic form.*

Proof. By Proposition 6.17 we obtain that $f \in D(\mathsf{Ch})$ implies $f \in D(\mathsf{Ch}^c)$ and $\mathsf{Ch}^c(f) \leq \mathsf{Ch}(f)$. If $f \in \mathrm{Lip}(Z)$, Lemma 6.14 yields $|\nabla f|_w \leq |\nabla f|_c$ $\mathfrak{m}$-a.e. in Z and the converse inequality $\mathsf{Ch}(f) \leq \mathsf{Ch}^c(f)$. It follows that the functionals and the gradients coincide in $\mathrm{Lip}(Z)$. Since $\mathrm{Lip}(Z)$ is a $L^2(Z,\mathfrak{m})$-dense and invariant subset for H_t^c, for all $f \in D(\mathsf{Ch}^c)$ we can apply Lemma 4.10 to obtain $(f_n) \subset \mathrm{Lip}(Z)$ satisfying $\mathsf{Ch}^c(f - f_n) \to 0$ and we can pass to the limit as $n \to \infty$ in the inequality $|\nabla f_n|_w \leq |\nabla f_n|_c$ to get $|\nabla f|_w \leq |\nabla f|_c$. Hence, $\mathsf{Ch}(f) = \mathsf{Ch}^c(f)$ and the respective gradients coincide. ☐

Proof. (*of Theorem 6.12*) By [39, Proposition 4.16] we know that $(Z,\mathsf{d},\mathfrak{m})$ is $CD(K,\infty)$, while Theorem 6.18 ensures that Cheeger's energy in this space is a quadratic form.

The proof that $(Z,\mathsf{d},\mathfrak{m})$ is nonbranching is simple, and we just sketch the argument. It is immediately seen that the non branching property is implied by the stability of constant speed geodesics under projections, namely if $\gamma = (\gamma^X,\gamma^Y) \in \mathrm{Geo}(Z)$, then $\gamma^X \in \mathrm{Geo}(X)$ and

$\gamma^Y \in \mathrm{Geo}(Y)$. This stability property can be shown as follows: in any metric space, constant speed geodesics are characterized by

$$\int_0^1 |\dot\gamma_t|^2 \, \mathrm{d}t = \mathsf{d}^2(\gamma_0, \gamma_1),$$

while for all other curves the inequality $\geq$ holds. Since $|\dot\gamma_t|^2 = |\dot\gamma_t^X|^2 + |\dot\gamma_t^Y|^2$ wherever the metric derivatives of the components exist, we obtain $\int_0^1 |\dot\gamma_t^X|^2 \, \mathrm{d}t = \mathsf{d}_X^2(\gamma_0^X, \gamma_1^X)$ and $\int_0^1 |\dot\gamma_t^Y|^2 \, \mathrm{d}t = \mathsf{d}_Y^2(\gamma_0^Y, \gamma_1^Y)$, so that both γ^X and γ^Y are constant speed geodesics.

Finally, we prove the strong $CD(K, \infty)$ property. Since the space is nonbranching, by Remark 3.2 it is sufficient to prove that it is a $CD(K, \infty)$ space. To prove this, we argue exactly as in [39, Lemma 4.7 and Proposition 4.16], taking into account the tightness of the sublevels of $\mathrm{Ent}_{\mathfrak{m}}$ to remove the compactness assumption. We omit the details. □

6.5 Locality

Here we study the locality properties of $RCD(K, \infty)$ spaces. As for the tensorization, we will adopt the point of view of the definition coming from the Dirichlet form, rather than the ones coming from the properties of the heat flow. The reason is simple. On one side, the heat flow does not localize at all: even on $\mathbb{R}^d$ to know how the heat flow behaves on the whole space gives little information about the behavior of the flow on a bounded region (we recall that, with our definitions, the heat flow that we consider reduces to the classical one with homogeneous Neumann boundary condition). On the other hand, Cheeger's energy comes out as a local object, and we will see that the analysis carried out in Section 4.3 and Section 6.2 will allow us to quickly derive the locality properties we are looking for.

There are two questions we want to answer. The first one is: say that we have a $RCD(K, \infty)$ space and a convex subregion, can we say that this subregion - endowed with the restricted distance and measure - is a $RCD(K, \infty)$ space as well? The second one is: suppose that a space is covered by subregions, each one being a $RCD(K, \infty)$ space, can we say that the whole space is $RCD(K, \infty)$?

The first question has a simple answer: yes. The second one is more delicate, the problem coming from proving the convexity of the entropy. The analogous question for $CD(K, \infty)$ spaces has, as of today, two different answers. On one side there is Sturm's result [39, Theorem 4.17] saying that this local-to-global property holds if the space is nonbranching and the domain of the entropy is geodesically convex. On the other side there is Villani's result [43, Theorem 30.42]) which still requires the space to be nonbranching, but replaces the global convexity of the domain on the entropy, with a local one one, roughly speaking "$(X, \mathsf{d}, \mathfrak{m})$ is finite-dimensional near to every point" (in a sense which we won't specify).

Our answer to the local-to-global question in the $RCD(K, \infty)$ setting will be based on the following assumptions, besides the obvious one that the covering subregions are $RCD(K, \infty)$: the space is nonbranching and $CD(K, \infty)$, so that independently from the approach one has at disposal to prove the local to global for $CD(K, \infty)$, as soon as the space is nonbranching, $RCD(K, \infty)$ globalizes as well.

We say that a subset Y of a metric space (X, d) is *convex* if, for any x, $y \in Y$, there exists a geodesic γ connecting x to y is contained in Y.

Theorem 6.19 (Global to Local) *Let $(X, \mathsf{d}, \mathfrak{m}) \in \mathbb{X}$ be a $RCD(K, \infty)$ space and let $Y \subset$*

X be a closed convex set such that $\mathfrak{m}(\partial Y)=0$ and $\mathfrak{m}(Y)>0$. Then $(Y,\mathsf{d},\mathfrak{m}_Y)$ is a $RCD(K,\infty)$ space as well, where $\mathfrak{m}_Y:=(\mathfrak{m}(Y))^{-1}\mathfrak{m}\llcorner Y$.

Proof. Since Y is closed, $(Y,\mathsf{d},\mathfrak{m}_Y)\in\mathbb{X}$. Let us first remark that for every $\mu\in\mathscr{P}_2(X)$

$$\mathrm{Ent}_{\mathfrak{m}_Y}(\mu)<\infty \qquad \Leftrightarrow \qquad \operatorname{supp}\mu\subset Y, \quad \mathrm{Ent}_{\mathfrak{m}}(\mu)<\infty, \tag{6.39}$$

and in this case $\mathrm{Ent}_{\mathfrak{m}_Y}(\mu)=c_Y+\mathrm{Ent}_{\mathfrak{m}}(\mu)$, where $c_Y=\log(\mathfrak{m}(Y))$. Therefore, thanks to the $RCD(K,\infty)$ property of $(X,\mathsf{d},\mathfrak{m})$, the functional $\mathrm{Ent}_{\mathfrak{m}_Y}$ is K-geodesically convex on any Wasserstein geodesic (μ_s) with $\operatorname{supp}\mu_s\subset Y$ for all $s\in[0,1]$. Such a geodesic exists since (Y,d) and thus $(\mathscr{P}_2(Y),W_2)$ are geodesic spaces: in particular, $(Y,\mathsf{d}_Y,\mathfrak{m}_Y)$ is a strong $CD(K,\infty)$ space. Thus, to conclude we simply apply (iii) of Theorem 4.20. □

The previous result is similar to the following lower Ricci curvature bound for weighted spaces:

Proposition 6.20 (Weighted spaces) *Let $(X,\mathsf{d},\mathfrak{m})\in\mathbb{X}$ be a $RCD(K,\infty)$ space and let $V:X\mapsto\mathbb{R}$ be a continuous H-geodesically convex function bounded from below with $\int \mathrm{e}^{-V}\,\mathrm{d}\mathfrak{m}=1$. Then $(X,\mathsf{d},\mathrm{e}^{-V}\mathfrak{m})$ is a $RCD(K+H,\infty)$ space.*

The proof follows by the same arguments, applying [39, Proposition 4.14] (showing that $(X,\mathsf{d},\mathrm{e}^{-V}\mathfrak{m})$ is a strong $CD(K+H,\infty)$ space), [2, Lemma 4.11] for the invariance of weak gradients with respect to the multiplicative perturbation, and (iii) of Theorem 4.19.

We conclude this section with the globalization result.

Theorem 6.21 (Local to Global) *Let $(X,\mathsf{d},\mathfrak{m})\in\mathbb{X}$ and let $\{Y_i\}_{i\in I}$ be a cover of X made of finitely or countably many closed sets of positive $\mathfrak{m}$-measure, with $\mathfrak{m}_i:=[\mathfrak{m}(Y_i)]^{-1}\mathfrak{m}\llcorner Y_i$. Assume that the Cheeger functional associated to $(Y_i,\mathsf{d},\mathfrak{m}_i)$ is quadratic (in particular when $(Y_i,\mathsf{d},\mathfrak{m}_i)$ is a $RCD(K,\infty)$ space) for every $i\in I$. Assume also that $(X,\mathsf{d},\mathfrak{m})$ is nonbranching and $CD(K,\infty)$. Then $(X,\mathsf{d},\mathfrak{m})$ is a $RCD(K,\infty)$ space.*

Proof. We start proving that Ch is a quadratic form. Notice that it holds

$$2\mathsf{Ch}(f)=\int_X|\nabla f|_w^2\,\mathrm{d}\mathfrak{m}=\sum_i\int_{X_i}|\nabla f|_w^2\,\mathrm{d}\mathfrak{m}$$

where $X_i:=Y_i\setminus\cup_{j<i}Y_j$. Let $f_i:=f_{|Y_i}$ and recall that by Theorem 4.20 we know that $|\nabla f_i|_{w,Y_i}=|\nabla f|_w$ $\mathfrak{m}$-a.e. on Y_i. Also, by Theorem 4.19 we have that for i and any Borel subset A of Y_i, the map $f\mapsto\int_A|\nabla f|_{w,Y_i}^2\,\mathrm{d}\mathfrak{m}$ is quadratic. Choosing $A=X_i$ the conclusion follows.

The fact that $(X,\mathsf{d},\mathfrak{m})$ is a strong $CD(K,\infty)$ space follows from the fact that it is nonbranching and $CD(K,\infty)$, as in Remark 3.2. □

References

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

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Calculus and heat flows in metric measure spaces with Ricci curvature bounded from below*, Submitted paper, arXiv:1106.2090.

[3] ——, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[4] 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.

[5] D. Bakry, *Functional inequalities for Markov semigroups*, in Probability measures on groups: recent directions and trends, Tata Inst. Fund. Res., Mumbai, 2006, pp. 91–147.

[6] 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).

[7] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, vol. 33 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2001.

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

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

[10] P. Clément and W. Desch, *Some remarks on the equivalence between metric formulations of gradient flows*, Boll. Unione Mat. Ital. (9), 3 (2010), pp. 583–588.

[11] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math., 146 (2001), pp. 219–257.

[12] G. Dal Maso, *An Introduction to* Γ*-Convergence*, vol. 8 of Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser, Boston, 1993.

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

[14] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften, Band 153, Springer-Verlag New York Inc., New York, 1969.

[15] K. Fukaya, *Collapsing of riemannian manifolds and eigenvalues of the laplace operator*, Invent. Math., 87 (1987), pp. 517–547.

[16] M. Fukushima, *Dirichlet forms and Markov processes*, vol. 23 of North-Holland Mathematical Library, North-Holland Publishing Co., Amsterdam, 1980.

[17] M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet forms and symmetric Markov processes*, vol. 19 of de Gruyter Studies in Mathematics, Walter de Gruyter & Co., Berlin, extended ed., 2011.

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

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

[20] N. Gigli and S. Ohta, *First variation formula in Wasserstein spaces over compact Alexandrov spaces*, Canad. Math. Bull., to appear, (2010).

[21] A. Joulin, *A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature*, Bernoulli, 15 (2009), pp. 532–549.

[22] P. Koskela and Y. Zhou, *Geometry and Analysis of Dirichlet forms*, Preprint, 2011.

[23] K. Kuwada, *Duality on gradient estimates and wasserstein controls*, Journal of Functional Analysis, 258 (2010), pp. 3758–3774.

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

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

[26] J. Lott and C. Villani, *Weak curvature bounds and functional inequalities*, J. Funct. Anal., 245 (2007), pp. 311–333.

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

[28] Z.-M. Ma and M. Röckner, *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*, Springer, Heidelberg, 1992.

[29] S.-i. Ohta, *Gradient flows on Wasserstein spaces over compact Alexandrov spaces*, Amer. J. Math., 131 (2009), pp. 475–516.

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

[31] Y. Ollivier, *Ricci curvature of Markov chains on metric measure spaces*, J. Functional Analysis, 256 (2009), pp. 810–864.

[32] A. Petrunin, *Alexandrov meets lott-villani-sturm*, Münster J. of Mathematics, to appear, (2009).

[33] T. Rajala, *Local Poincaré inequalities from stable curvature conditions in metric spaces*, 2011.

[34] G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris, 345 (2007), pp. 151–154.

[35] G. Savaré, *Gradient flows and evolution variational inequalities in metric spaces*, In preparation, (2011).

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

[37] K. T. Sturm, *Analysis on local Dirichlet spaces. III. The parabolic Harnack inequality*, J. Math. Pures Appl. (9), 75 (1996), pp. 273–297.

[38] K.-T. Sturm, *Is a diffusion process determined by its intrinsic metric?*, Chaos, Solitons & Fractals, 8 (1997), pp. 1855 – 1860. Relativity, Locality and Random Fractals in Quantum Theory.

[39] ——, *On the geometry of metric measure spaces. I*, Acta Math., 196 (2006), pp. 65–131.

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

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

[42] K.-T. Sturm and M.-K. von Renesse, *Transport inequalities, gradient estimates, entropy, and Ricci curvature*, Comm. Pure Appl. Math., 58 (2005), pp. 923–940.

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

[44] H.-C. Zhang and X.-P. Zhu, *Ricci curvature on Alexandrov spaces and rigidity theorems*, Comm. Anal. Geom., 18 (2010), pp. 503–553.