Equivalent definitions of BV space and of total variation on metric measure spaces

Luigi Ambrosio [*] Simone Di Marino [†]

February 4, 2014

Contents

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

[†]Scuola Normale Superiore, Pisa. email: simone.dimarino@sns.it

1 Introduction

In this paper we study BV functions in metric measure spaces, providing a positive answer to a problem raised in [7], where similar questions are investigated and positively answered in the setting of Sobolev spaces. Let $(X, \mathsf{d}, \mathfrak{m})$ be a complete and separable metric measure space, with $\mathfrak{m}$ locally finite Borel measure. Recall that, according to the notion of BV function given in [27], a function $f \in L^1(X, \mathsf{d}, \mathfrak{m})$ belongs to $BV_*(X, \mathsf{d}, \mathfrak{m})$ if there exist locally Lipschitz functions f_n convergent to f in $L^1(X)$ such that

$$\limsup_{n\to\infty} \int_X |\nabla f_n| \, \mathrm{d}\mathfrak{m} < \infty.$$

Here $|\nabla f_n|$ are the slopes (also called local Lipschitz constants) of f_n, see (2.1) below. By localizing this construction one can define

$$|Df|_*(A) := \inf \left\{ \liminf_{h\to\infty} \int_A |\nabla f_h| \, \mathrm{d}\mathfrak{m} : \ f_h \in \mathrm{Lip}_{\mathrm{loc}}(A), \ f_h \to f \text{ in } L^1(A) \right\}, \tag{1.1}$$

for any open set $A \subseteq X$. In [27], it is proved (with a minor variant, since L^1_{loc} convergence of the functions is considered) that this set function is the restriction to open sets of a finite Borel measure, called total variation measure and, following basically the same strategy, we will extend this result to our more general setup.

A first variant of this definition arises if one considers not only locally Lipschitz approximating functions, but general functions f_n, replacing $|\nabla f_n|$ by upper gradients g_n of f_n. This is the definition considered in [11] for Sobolev functions, with a mention of the possibility of extending it to the construction to BV functions; denoting by $BV_*^c(X, \mathsf{d}, \mathfrak{m})$ the corresponding space, it is clear that BV_*^c is larger than BV_*; in addition, the set function $|Df|_*^c$ obtained by a procedure similar to (1.1) is smaller than $|Df|$, since the class of approximating functions is larger.

We will prove indeed that the two approaches lead to the same BV space and to the same total variation measure. As a matter of fact, this equivalence is part of a more general result, where we consider a new definition of BV function in the spirit of the theory of weak, rather than relaxed, upper gradients [24, 28]. Without entering in this introduction in too many technical details, we say that $f \in w-BV(X, \mathsf{d}, \mathfrak{m})$ if there exists a finite Borel measure μ with this property: for any probability measure $\boldsymbol{\pi}$ on $\mathrm{Lip}([0,1]; X)$ the function $t \mapsto f \circ \gamma_t$ belongs to $BV(0,1)$ for $\boldsymbol{\pi}$-a.e. curve γ_t and

$$\frac{1}{C(\boldsymbol{\pi}) \|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}} \int \gamma_\sharp |D(f\circ\gamma)| \, \mathrm{d}\boldsymbol{\pi} \le \mu.$$

Here $C(\boldsymbol{\pi})$ is the least constant C such that $(e_t)_\sharp \boldsymbol{\pi} \le C\mathfrak{m}$ for all $t \in [0,1]$, where $e_t(\gamma) := \gamma_t$ are the evaluation maps at time t. See also Remark 7.2 for another definition which involves only the oscillation of f at the endpoints of the curve. The smallest measure μ with this property will be denoted by $|Df|_w$, and the proof that $w-BV(X, \mathsf{d}, \mathfrak{m})$ includes the previous two spaces and that $|Df|_* \ge |Df|_*^c \ge |Df|_w$ is not too difficult. It is also worthwhile to mention that the BV property along curves and suitable measures in the space of curves play a role in [14] (see also [12, 13]), for this reason we think that it is interesting to compare the relaxation point of view with the point of view based on measure upper gradients, so to speak.

Notice that proving equivalence of the three definitions amounts to passing from a (quantitative) information on the behavior of the function along random curves to the construction of a Lipschitz approximation. Remarkably, this result does not rely on doubling and Poincaré assumptions on the metric measure structure. As in [7] (based essentially on ideas come from [5], dealing with the case of $W^{1,2}$ Sobolev spaces), the proof is not really constructive: it is obtained with optimal transportation tools and using the theory of gradient flows of convex and lower semicontinuous functionals in Hilbert spaces. Specifically, in our case we shall use the gradient flow in $L^2(X, \mathfrak{m})$ of the functional $f \mapsto |Df|_*(X)$, also called total variation flow in image processing [9].

We can now state the main result of our paper (see also Corollary 7.5, kindly pointed out to us by the reviewer of the paper).

Theorem 1.1 *Let $(X, \mathsf{d}, \mathfrak{m})$ be a complete and separable metric measure space, with $\mathfrak{m}$ nonnegative and locally finite Borel measure (i.e. for all $x \in X$ there exists $r > 0$ such that $\mathfrak{m}(B_r(x)) < \infty$). Then the spaces*

$$BV_*(X, \mathsf{d}, \mathfrak{m}), \qquad BV_*^c(X, \mathsf{d}, \mathfrak{m}), \qquad w - BV(X, \mathsf{d}, \mathfrak{m})$$

and the corresponding total variation measures $|Df|_*$, $|Df|_*^c$, $|Df|_w$ coincide.

The paper is organized as follows. In Section 2 we recall some preliminary facts on absolutely continuous curves, upper gradients and BV functions. In Section 3 we study the properties of the Hopf-Lax semigroup

$$\inf_{y \in X} \phi(y) + \frac{\mathsf{d}^p(x, y)}{pt^{p-1}}.$$

in the limit case when $p = \infty$, where it reduces simply to $Q_t\phi(x) = \inf_{\overline{B}_t(x)} \phi$. In particular the differential inequality

$$\frac{d}{dt} Q_t\phi + |\nabla Q_t\phi| \le 0$$

will play an important role in our analysis. In Section 4 we study some elementary properties of the W_∞ Wasserstein distance, focussing in particular on the dual formulation. In Section 5 we present and compare the three definitions of BV we already mentioned, proving in particular the "easy" inequalities $|Df|_* \ge |Df|_*^c \ge |Df|_w$. In Section 6 we gather a few facts on the gradient flow of $|Df|_*$, that are used in Section 7 to prove our main result.

Finally, Section 8 is devoted to the discussion of 3 potential definitions of the Sobolev space $W^{1,1}$. Since the functional

$$\Phi(f) := \begin{cases} |Df|_*(X) & \text{if } f \in BV_*(X, \mathsf{d}, \mathfrak{m}) \text{ and } |Df|_* \ll \mathfrak{m}; \\ \\ +\infty & \text{otherwise} \end{cases}$$

is not lower semicontinuous, one of our main tools (namely the theory of gradient flows) breaks down and we are not able, at least at this level of generality, to prove equivalence of the three definitions. The Appendix is devoted to the proof of the metric superposition principle in the limiting case $p=\infty$: its proof follows with minor variants [26].

We close this introduction mentioning that some properties of BV functions readily extend to the more general framework considered in this paper. For instance, the coarea formula

$$|Df|_* = \int_0^\infty |D\chi_{\{f>t\}}|_* \,\mathrm{d}t + \int_{-\infty}^0 |D\chi_{\{f<t\}}|_* \,\mathrm{d}t$$

can be achieved following *verbatim* the proof in [27]. On the other hand, more advanced facts, as the decomposition alone curves in absolutely continuous and singular part of the derivative (see [3, Section 3.11]), seem to be open at this level of generality: for instance, Example 7.4 shows that, in contrast to what happens in Euclidean metric measure spaces (here the supremum is understood in the lattice of measures), the measure

$$\sup_{\boldsymbol{\pi}} \frac{1}{C(\boldsymbol{\pi})\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}} \int \gamma_\sharp |D^a(f\circ\gamma)| \,\mathrm{d}\boldsymbol{\pi}, \tag{1.2}$$

which is easily seen to be smaller than the absolutely continuous part of $|Df|_w$, maybe be strictly smaller.

Acknowledgement. The first author acknowledges the support of the ERC ADG GeMeThNES and of the PRIN08-grant from MIUR for the project *Optimal transport theory, geometric and functional inequalities, and applications*. The authors thank an anonymous referee for his/her extremely valuable remarks.

2 Notation and preliminary notions

In this section we introduce some notation and recall a few basic facts on absolutely continuous functions, and BV functions, see also [4], [3] as general references.

2.1 Absolutely continuous curves

Let (X,d) be a metric space, $J\subseteq\mathbb{R}$ a closed interval and consider a curve $\gamma:J\to X$ (sometimes we will denote $\gamma(t)=\gamma_t$). We say that γ is *absolutely continuous* if

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

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

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

See [4, Theorem 1.1.2] for the simple proof.

We will denote by $C([0,1],X)$ the space of continuous curves from $[0,1]$ to (X,d) endowed with the sup norm and by $AC([0,1],X)$ the subset of absolutely continuous curves. For $p>1$, the set $AC^p([0,1],X)\subseteq C([0,1],X)$ consists of all absolutely continuous curves γ such that $|\dot\gamma_t|\in L^p(0,1)$: it is the countable union of the sets $\{\gamma:\|\dot\gamma_t\|_{L^p}\le n\}$, which are easily seen to be closed. Thus the set $AC^p([0,1],X)$ is a Borel subset of $C([0,1],X)$ (this is still true for $AC([0,1],X)$, see [5], but we shall not need this fact in the sequel); in particular we will be interested in $AC^\infty([0,1],X)$, which is easily seen to coincide with the set of Lipschitz curves. The *evaluation maps* $\mathrm{e}_t : C([0,1],X)\to X$ are defined by

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

and are clearly continuous.

2.2 Slopes, locally Lipschitz functions and upper gradients

Let (X,d) be a metric space; given $f:X\to\mathbb{R}$, we define the *slope* of f (also called local Lipschitz constant) by

$$|\nabla f|(x) := \varlimsup_{y\to x}\frac{|f(y)-f(x)|}{\mathsf{d}(y,x)}, \tag{2.1}$$

and, correspondingly, the *ascending* slope $|\nabla^+ f|$ and the *descending* slope $|\nabla^- f|$:

$$|\nabla^\pm f|(x) := \varlimsup_{y\to x}\frac{(f(y)-f(x))^\pm}{\mathsf{d}(y,x)}. \tag{2.2}$$

In the sequel, we say that f is locally Lipschitz in an open set A if for every $x\in A$, the function is Lipschitz in a neighborhood of x. With this definition locally Lipschitz functions in X are Lipschitz if the ambient space X is compact. For f, $g:X\to\mathbb{R}$ locally Lipschitz it clearly holds

$$|\nabla(\alpha f+\beta g)|\le|\alpha||\nabla f|+|\beta||\nabla g| \qquad \forall \alpha,\beta\in\mathbb{R}, \tag{2.3a}$$

$$|\nabla(fg)|\le|f||\nabla g|+|g||\nabla f|. \tag{2.3b}$$

Given a real valued function f on X, we denote by $UG(f)$ the set of upper gradients of f (see also [21, 11]), namely the class of Borel functions $g:X\to[0,\infty]$ such that

$$|f(\gamma_1)-f(\gamma_0)|\le\int_0^1 g(\gamma_t)|\dot\gamma_t|\,\mathrm{d}t \qquad \forall\,\gamma\in AC([0,1];X).$$

With a slight abuse of notation we will write $g\in UG(f)$ with $f\in L^1(X,\mathfrak{m})$, but it should be noticed that a priori the concept of upper gradient is not invariant in the equivalence class of an L^1 function, even though Borel representatives are chosen.

2.3 BV functions and total variation on Euclidean spaces

We refer to Chapter 3 of [3] for a complete review of this topic, with all the proofs; here we will only overview the main properties needed in this paper.

Given an open set $A \subseteq \mathbb{R}^d$, $f \in L^1(A)$ is said to be of *bounded variation* in A if one of the following three equivalent properties hold:

(a) the distributional derivative Df is a $\mathbb{R}^d$-valued measure with finite total variation in A;

(b) The following quantity, called total variation of f in A, is finite:

$$TV_f(A) := \sup\left\{\int_A f\,\mathrm{div}\phi\,\mathrm{d}x \ :\ \phi \in C^1_c(A;\mathbb{R}^d),\ |\phi| \le 1\right\}.$$

(c) There exists a sequence $(f_n) \subseteq C^\infty(A)$ converging fo f in $L^1(A)$ with $\sup_n \int_A |\nabla f_n|\,\mathrm{d}x < \infty$.

The equivalence between (a), (b) and (c) leads to relations between the corresponding quantities involved: in particular we have

$$|Df|(A) = TV_f(A) \le \liminf_{n\to\infty} \int_A |\nabla f_n|\,\mathrm{d}x.$$

Moreover the second definition gives us easily the crucial property that the total variation $|Df|$ of the distributional derivative in open sets is lower semicontinuous with respect to L^1 convergence:

$$\liminf_{n\to\infty} |Df_n|(A) \ge |Df|(A) \qquad \forall A \subseteq \mathbb{R}^d \text{ open set}, \quad f_n \to f \text{ in } L^1(A). \tag{2.4}$$

By means of standard mollifiers and partitions of unity we can get also the following stronger result: there exists a sequence of functions $f_n \in C^\infty(A)$ convergent to f in $L^1(A)$ and such that $|Df_n|(A) \to |Df|(A)$. In our metric context we simply replace $C^\infty(A)$ by the space of locally Lipschitz functions on A.

3 Hopf-Lax formula and Hamilton-Jacobi equation

In this section we study some elementary properties of the Hopf-Lax formula in a metric setting, extending to a limiting case (suitable for the study of the ∞-Wasserstein distances, made in the next section) the analysis made in [5], see also [18]. Here we assume that (X, d) is a metric space: there is no reference measure $\mathfrak{m}$ here and we can drop even the completeness assumption. We are dealing with a very simple convex lower semicontinuous Lagrangian $L : [0,\infty] \to [0,\infty]$:

$$L(s) = \begin{cases} 0 & \text{if } s \le 1; \\ \infty & \text{otherwise.} \end{cases}$$

Let $\phi : X \to \mathbb{R}$ be a Lipschitz function. We set $Q_0\phi(x) = \phi(x)$ and, for $t > 0$,

$$Q_t\phi(x) := \inf_{y\in X} \Big\{\phi(y) + L\Big(\frac{\mathsf{d}(x,y)}{t}\Big)\Big\}. \tag{3.1}$$

Due to the particular form of our Lagrangian, we get

$$Q_t\phi(x) := \inf_{\mathsf{d}(x,y)\le t} \phi(y). \tag{3.2}$$

Obviously, these transformations act almost as a semigroup: in fact, the triangle inequality gives

$$Q_sQ_t\phi(x) = \inf_{\mathsf{d}(y,x)\le s} \Big\{ \inf_{\mathsf{d}(y,z)\le t} \phi(z)\Big\} \ge \inf_{\mathsf{d}(x,z)\le s+t} \phi(z) = Q_{s+t}\phi(x).$$

Moreover, if (X, d) is a length space, we have equality and thus Q_t is a semigroup. In fact, under this assumption, for every z such that $\mathsf{d}(x, z) < s + t$ there exists a constant speed curve $\gamma : [0, 1] \to X$ whose length is less than $s + t$ and such that $\gamma_0 = x$ and $\gamma_1 = z$; in particular there will be a time $\eta := s/(s + t)$ such that $y := \gamma_\eta$ satisfies $\mathsf{d}(x, y) < s$ and $\mathsf{d}(y, z) < t$. It follows that $Q_sQ_t\phi(x) \le \inf_{\mathsf{d}(x,z)<s+t} \phi(z)$. In order to conclude, one has to observe that, if ϕ is continuous, then

$$\inf_{\mathsf{d}(x,z)\le r} \phi(z) = \inf_{\mathsf{d}(x,z)<r} \phi(z) \qquad \forall r > 0,$$

and this is true because in a length space the closure of the open ball is the closed ball.

Also, it is easy to check that the length space property ensures that the Lipschitz constant does not increase:

$$\mathrm{Lip}(Q_t\phi) \le \mathrm{Lip}(\phi). \tag{3.3}$$

Now we look at the time derivative, to get information on the Hamilton-Jacobi equation satisfied by $Q_t\phi(x)$:

Theorem 3.1 (Time derivative of Q_tf) *Let $x \in X$. The map $t \mapsto Q_t\phi(x)$ is nonincreasing in $[0,\infty)$ and satisfies:*

$$\frac{\mathrm{d}}{\mathrm{d}t}Q_t\phi(x) + |\nabla Q_t\phi(x)| \le 0 \qquad \textit{for a.e. } t > 0. \tag{3.4}$$

Moreover, if (X, d) is a length space, the map $t \mapsto Q_t\phi$ is Lipschitz from $[0,\infty)$ to $C(X)$, with Lipschitz constant $\mathrm{Lip}(\phi)$.

Proof. The basic inequality, that we will use in the first part of the proof is:

$$Q_s\phi(y) \le Q_{s'}\phi(y') \qquad \text{whenever } s \ge s' + \mathsf{d}(y, y'). \tag{3.5}$$

It holds because the inequality implies $B(y', s') \subseteq B(y, s)$ and thus it is clear by the very definition of $Q_t\phi$. Now we take x_i and y_i converging to x such that:

$$\lim_{i\to\infty} \frac{Q_t\phi(x_i) - Q_t\phi(x)}{\mathsf{d}(x_i, x)} = -|\nabla^- Q_t\phi|(x), \qquad \lim_{i\to\infty} \frac{Q_t\phi(x) - Q_t\phi(y_i)}{\mathsf{d}(x, y_i)} = -|\nabla^+ Q_t\phi|(x).$$

Now we consider the inequalities, given by (3.5), involving x, x_i, y_i:

$$Q_{t+\mathsf{d}(x_i,x)}\phi(x) \le Q_t\phi(x_i), \qquad Q_t\phi(y_i) \le Q_{t-\mathsf{d}(x,y_i)}\phi(x)$$

and let us define, for brevity, $s_i = \mathsf{d}(x_i, x)$ and $r_i = \mathsf{d}(x, y_i)$. Then we have

$$\begin{aligned}\liminf_{h\to 0^+} \frac{Q_{t+h}\phi(x) - Q_t\phi(x)}{h} &\le \liminf_{i\to\infty} \frac{Q_{t+s_i}\phi(x) - Q_t\phi(x)}{s_i} \\ &\le \lim_{i\to\infty} \frac{Q_t\phi(x_i) - Q_t\phi(x)}{s_i} = -|\nabla^- Q_t\phi|(x)\end{aligned}$$

and, similarly,

$$\begin{aligned}\liminf_{h\to 0^-} \frac{Q_{t+h}\phi(x) - Q_t\phi(x)}{h} &\le \liminf_{i\to\infty} \frac{Q_t\phi(x) - Q_{t-r_i}\phi(x)}{r_i} \\ &\le \lim_{i\to\infty} \frac{Q_t\phi(x) - Q_t\phi(y_i)}{r_i} = -|\nabla^+ Q_t\phi|(x).\end{aligned}$$

Using that $|\nabla f| = \max\{|\nabla^+ f|, |\nabla^- f|\}$, the combination of these inequalities gives

$$\liminf_{h\to 0} \frac{Q_{t+h}\phi(x) - Q_t\phi(x)}{h} \le -|\nabla Q_t\phi|(x) \qquad \forall x \in X \;\; \forall t > 0.$$

Since $Q_t\phi(x)$ is obviously non increasing w.r.t. t, we get that is differentiable almost everywhere and so we get the thesis.

If we suppose that (X, d) is also a length space, using the semigroup property and (3.3) we get that

$$Q_s\phi(x) - Q_t\phi(x) = Q_s\phi(x) - Q_{t-s}(Q_s\phi)(x) \le (t-s)\,\mathrm{Lip}(Q_s\phi) \le (t-s)\,\mathrm{Lip}(\phi) \quad \forall s \in [0,t],$$

and so the thesis. □

Note that, in case (X, d) is not a length space, it might happen that balls are not connected and, as a consequence, that $t \mapsto Q_t\phi(x)$ is discontinuous; as an example we can take X the curve in Figure 1, with the distance induced as subset of $\mathbb{R}^2$.

It is clear that some balls, such as the shaded one centered in x, are disconnected; furthermore if we take a Lipschitz function ϕ equal to 0 in the upper part of the curve and equal to 1 in the lower one, doing an interpolation between two values only in the rightmost and leftmost parts, it is easy to see that $Q_t\phi(p)$ is discontinuous both in time and space.

4 The ∞-Wasserstein distance

Let (X, d) be a complete and separable metric space and let $\mathcal{M}^+(X)$ denote the set of positive and finite Borel measures on X. Given a lower semicontinuous cost $c : X \times X \to$

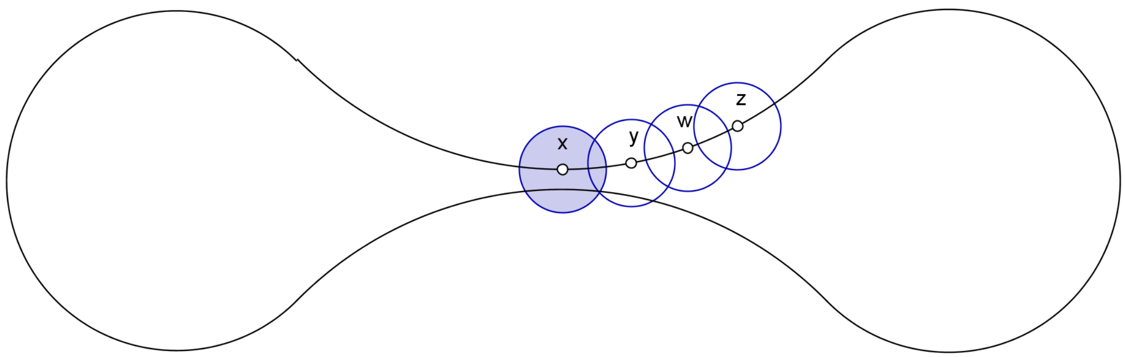

Figure 1: *Example of a compact metric space* (X, d) *that is not a length space, having a time discontinuous Hopf-Lax semigroup* Q_t

$[0, \infty]$, we can consider the classical Kantorovich transport problem on X between measures with same mass, defining

$$\mathscr{C}_c(\mu, \nu) := \min \left\{ \int_{X \times X} c(x, y) \, \mathrm{d}\gamma \mid {\pi^1}_\sharp \gamma = \mu, \ {\pi^2}_\sharp \gamma = \nu \right\},$$

where π^1 and π^2 are respectively the projections on the first and second factors. We shall denote by $\Gamma(\mu, \nu)$ the collection of admissible plans γ in the Kantorovich minimization problem. In the case of $c_p = \mathsf{d}^p$, $1 \leq p < \infty$, we get the classical Wasserstein distances $W_p = (\mathscr{C}_{c_p})^{1/p}$; they can equivalently be written as

$$W_p(\mu, \nu) = \min\{\|\mathsf{d}\|_{L^p(\gamma)} \, | \gamma \in \Gamma(\mu, \nu)\}$$

and so it is somewhat natural to look at the limiting case $p = \infty$:

$$W_\infty(\mu, \nu) := \min \left\{\|\mathsf{d}\|_{L^\infty(\gamma)} \, | \gamma \in \Gamma(\mu, \nu)\right\}.$$

It is known (see for instance [10]) that W_∞ is the monotone limit of W_p as p goes to infinity, at least when we are dealing with probability measures; we want to consider also this limit case as a transport problem, in order to have a dual formulation that will be used later on. The key point is to consider the limit costs: in fact we consider $c(x, y) = L(\mathsf{d}(x, y))$, where L is the function we defined in the previous section. This is indeed the limit cost because d^p converges as $p \to \infty$, in the sense of De Giorgi's theory of Γ-convergence [15], to $L \circ \mathsf{d}$, namely

$$p_n \to \infty, \quad d_n \to d \quad \Longrightarrow \quad \liminf_{n \to \infty} d_n^{p_n} \geq L(d)$$

and

$$\text{for all } d \geq 0 \text{ there exist } d_p \to d \text{ such that } \limsup_{p \to \infty} d_p^p \leq L(d).$$

Notice however that the pointwise limit of d^p is strictly larger than $L(\mathsf{d})$ when the distance equals 1.

We then introduce the "test" distances

$$W_\infty^{(s)}(\mu,\nu) = \min\Big\{ \int_{X\times X} L\Big(\frac{\mathsf{d}(x,y)}{s}\Big) \mathrm{d}\gamma \mid \gamma \in \Gamma(\mu,\nu)\Big\},$$

called this way because $W_\infty^{(s)}(\mu,\nu) = 0$ if and only if $W_\infty(\mu,\nu) \le s$. These "test" distances are given by transport problems with lower semicontinuous costs $c(x,y) = L(\mathsf{d}(x,y)/s)$, so they have a dual formulation [4, Theorem 6.1.1]:

$$W_\infty^{(s)}(\mu,\nu) = \sup_{\psi\in \mathrm{Lip}_b(X)} \int_X \psi^c \, \mathrm{d}\mu + \int_X \psi \, \mathrm{d}\nu,$$

where ψ^c denotes the so called c-transform of f, defined as:

$$\psi^c(x) = \inf_{y\in X} \Big\{ c(x,y) - \psi(y) \Big\}$$

(namely the largest function $g(x)$ satisfying $g(x) + \psi(y) \le c(x,y)$ for all (x,y)). By the definition of $Q_s\phi$ given in the previous section we get:

$$\psi^c(y) = \inf_{y\in X} \left\{ L\left(\frac{\mathsf{d}(x,y)}{s}\right) - \psi(y) \right\} = Q_s(-\psi)(x).$$

Now, setting $\psi = -\phi$ in the dual formulation, and using this characterization of the c-transform, we get

$$W_\infty^{(s)}(\mu,\nu) = \sup_{\phi\in \mathrm{Lip}_b(X)} \int_X Q_s\phi \mathrm{d}\mu - \int_X \phi \mathrm{d}\nu. \tag{4.1}$$

5 Three notions of BV function

Let (X,d) be a complete and separable metric space and let $\mathfrak{m}$ be a nonnegative Borel measure in X. In this section we introduce three notions of BV function and, correspondingly, three notions of total variation. We recall that the aim of this paper is to show that these notions are equivalent.

5.1 BV functions in the relaxed sense

A function f in $L^1(X,\mathfrak{m})$ is said to be BV in the relaxed sense if there exist locally Lipschitz functions f_n converging to f in $L^1(X,\mathfrak{m})$ and with equibounded energies, i.e. such that $\sup_n \int_X |\nabla f_n| \, \mathrm{d}\mathfrak{m} < \infty$. We shall denote this space by $BV_*(X,\mathsf{d},\mathfrak{m})$.
We already noticed that this definition coincides with the classical one in Euclidean spaces.

Associated to this definition is the relaxed total variation $|Df|_*$, already introduced in [27], defined on open sets $A \subseteq X$ as:

$$|Df|_*(A) := \inf\left\{\liminf_{h\to\infty} \int_A |\nabla f_h|\,\mathrm{d}\mathfrak{m} :\ f_h \in \mathrm{Lip}_{\mathrm{loc}}(A),\ f_h \to f \text{ in } L^1(A)\right\}. \tag{5.1}$$

Here "locally Lipschitz in an open set A" means that for all $x \in A$ there exists $r > 0$ such that $B_r(x) \subseteq A$ and the restriction of f to $B_r(x)$ is Lipschitz. In [27] it is proved that for all relaxed BV functions f the set function $A \mapsto |Df|_*(A)$ is the restriction to open sets of a finite Borel measure, for which we keep the same notation. Since the result in [27] is stated and proved in locally compact spaces, we adapt his arguments to our more general framework.

Remark 5.1 If we apply the definition of relaxed total variation to a locally Lipschitz function f, taking $f_h = f$, we get

$$|Df|_*(A) \le \int_A |\nabla f|\,\mathrm{d}\mathfrak{m} \qquad \text{for all } A \subseteq X \text{ open.}$$

Thus, $|Df|_* \ll \mathfrak{m}$ and so f belongs to the Sobolev space $W^{1,1}_*(X, \mathsf{d}, \mathfrak{m})$ consisting of functions $f \in BV_*(X, \mathsf{d}, \mathfrak{m})$ such that $|Df|_* \ll \mathfrak{m}$. See Section 8 for more on the Sobolev space $W^{1,1}$.

5.2 BV functions in Cheeger's relaxed sense

We can imagine a slightly weaker notion than the previous one, not requiring f_n to be locally Lipschitz and replacing $|\nabla f_n|$ with an element of $UG(f_n)$, the set of upper gradients of f_n, see [21, 11]. So the definition becomes: a function $f \in L^1(X, \mathfrak{m})$ belongs to $BV^c_*(X, \mathsf{d}, \mathfrak{m})$ if there exist a sequence $(f_n) \subseteq L^1(X, \mathfrak{m})$ that converges to f in $L^1(X, \mathfrak{m})$ and upper gradients g_n of f_n, such that $\sup_n \int_X g_n \mathrm{d}\mathfrak{m} < \infty$.
For $f \in BV^c_*(X, \mathsf{d}, \mathfrak{m})$ the Cheeger total variation $|Df|^c_*$ is defined on open sets A as

$$|Df|^c_*(A) := \inf\left\{\liminf_{h\to\infty} \int_A g_h\,\mathrm{d}\mathfrak{m} :\ f_h \in L^1(A),\ f_h \to f \text{ in } L^1(A),\ g_h \in UG(f_h)\right\}. \tag{5.2}$$

Of course we have that $BV_* \subseteq BV^c_*$ and $|Df|_* \ge |Df|^c_*$, because the slope is an upper gradient for locally Lipschitz functions (see for instance [11]).

We investigate more closely the properties of the set functions $|Df|_*$ in the following lemma. We will write $A \Subset B$ whenever A, B are open sets and $\mathsf{d}(A, X \setminus B) > 0$ (in particular, $A \Subset B$ implies $A \subseteq B$). We say that A_1 and A_2 are well separated if $\mathrm{dist}(A_1, A_2) > 0$.

Lemma 5.2 *Let $\mathcal{A}(X)$ be the class of open subsets of X, $f \in L^1(X, \mathfrak{m})$ and let $\beta : \mathcal{A}(X) \to [0, \infty]$ be defined as in* (5.1)*, with the convention $\beta(\emptyset) = 0$. Then, β satisfies the following properties:*

(i) $\beta(A_1) \le \beta(A_2)$ whenever $A_1 \subseteq A_2$;

(ii) $\beta(A_1 \cup A_2) \le \beta(A_1) + \beta(A_2)$*, with equality if* A_1 *and* A_2 *are well separated;*

(iii) If A_n *are open and* $A_n \subseteq A_{n+1}$ *it holds*

$$\lim_{n\to\infty} \beta(A_n) = \beta\Big(\bigcup_n A_n\Big). \tag{5.3}$$

In particular the formula

$$\beta(B) := \inf\{\beta(A) \ : \ A \subseteq X \ \text{open}, \ B \subseteq A\}$$

provides a σ*-subadditive extension of* β *whose additive sets, in the sense of Carathéodory, contain* $\mathscr{B}(X)$*. If follows that* $\beta : \mathscr{B}(X) \to [0,\infty]$ *is a* σ*-additive Borel measure.*

Proof. The verifications of monotonicity and the additivity on well separated sets are standard.

Since we will use (iii) in the proof of the first statement of (ii), we prove (iii) first, denoting $A := \cup_n A_n$. It is sufficient to prove that $\sup |Du|_*(A_n) \ge |Du|_*(A)$ because the converse inequality is trivial by monotonicity, so we can assume that $\sup_n |Du|_*(A_n) < \infty$.

First, we reduce ourselves to the case when A_n satisfy the additional condition

$$\operatorname{dist}(\overline{A}_n, X \setminus A_{n+1}) > 0 \qquad \forall n \in \mathbb{N}. \tag{5.4}$$

In order to realize that the restriction to this case is possible, suffices to consider the sets

$$A'_n := \left\{ x \in X \ : \ \operatorname{dist}(x, X \setminus A_n) \ge \frac{1}{n} \right\}$$

which satisfy (5.4), are contained in A_n and whose union is still equal to A.

In particular, if we call

$$\begin{cases} C_1 = A_2 \\ C_k = A_k \setminus \overline{A_{k-2}} & \text{if } k \ge 2, \end{cases}$$

it is clear that the families $\{C_{3k+1}\}$, $\{C_{3k+2}\}$, $\{C_{3k+3}\}$ are well separated, hence $\sum_j |Du|_*(C_{3j+i}) < \infty$ for all $i \in \{1,2,3\}$. It follows that for any $\varepsilon > 0$ we can find an integer $\bar{k}$ such that

$$\sum_{n=\bar{k}}^{\infty} |Du|_*(C_n) \le \varepsilon. \tag{5.5}$$

Now, to prove (5.3) we build a sequence $(u_m) \subseteq \mathrm{Lip}_{\mathrm{loc}}(A)$ such that $u_m \to u$ in $L^1(A, \mathfrak{m})$ and

$$|Du|_*(A_{\bar{k}}) + 2\varepsilon \ge \liminf_{m\to\infty} \int_A |\nabla u_m| \, \mathrm{d}\mathfrak{m}.$$

In order to do so, we fix m and set $D_h = C_{h+\bar{k}}$, $B_h = A_{h+\bar{k}}$ if $h \geq 1$, $D_0 = B_0 = A_{\bar{k}}$. Then we choose $\psi_{k,h} \in \mathrm{Lip_{loc}}(D_h)$ in such a way that

$$\int_{D_h} |\nabla \psi_{k,h}| \, \mathrm{d}\mathfrak{m} \leq |Du|_*(D_h) + \frac{1}{m2^k}. \tag{5.6}$$

We are going to use Lemma 5.4 below with $M = B_h$, $N = D_{h+1}$, so we denote by c_h and $H_h \Subset B_h \cap D_{h+1}$ the constants and the domains given by the lemma. It is then easy to find sufficiently large integers $k(h) \geq h$ satisfying

$$c_h \int_{\overline{H}_h} |\psi_{k(h),h} - u| \, \mathrm{d}\mathfrak{m} \leq \frac{\varepsilon}{2 \cdot 2^h} \qquad \text{and} \qquad c_h \int_{\overline{H}_h} |\psi_{k(h+1),h+1} - u| \, \mathrm{d}\mathfrak{m} \leq \frac{\varepsilon}{2 \cdot 2^h}. \tag{5.7}$$

This is possible because $\overline{H}_h$ is contained in $B_h \cap D_{h+1}$ which, in turn, is contained in D_h. In addition, possibly increasing $k(h)$, we can also have:

$$\int_{D_h} |\psi_{k(h),h} - u| \, \mathrm{d}\mathfrak{m} \leq \frac{1}{m2^h}. \tag{5.8}$$

Now we define by induction on h functions $u_{m,h} \in \mathrm{Lip_{loc}}(B_h)$ for $h \geq 0$: we set $u_{m,0} = \psi_{k(0),0}$ and, given $u_{m,h}$, we build $u_{m,h+1}$ in such a way that:

$$\begin{cases} u_{m,h+1} \equiv u_{m,h} & \text{on } B_{h-1} \\ u_{m,h+1} \equiv \psi_{k(h+1),h+1} & \text{on } B_{h+1} \setminus \overline{B_h}, \end{cases} \tag{5.9}$$

$$\|u_{m,h} - u\|_{L^1(B_h)} \leq \frac{1}{m}\left(1 - \frac{1}{2^h}\right), \tag{5.10}$$

$$\int_{B_{h+1}} |\nabla u_{m,h+1}| \, \mathrm{d}\mathfrak{m} \leq \int_{B_h} |\nabla u_{m,h}| \, \mathrm{d}\mathfrak{m} + \int_{D_{h+1}} |\nabla \psi_{k(h+1),h+1}| \, \mathrm{d}\mathfrak{m} + \frac{\varepsilon}{2^h}. \tag{5.11}$$

Once we have this we are done because we can construct $u_m(x) = u_{m,h}(x)$ if $x \in B_{h-1}$, then it is clear that u_m is well defined thanks to the first equation in (5.9) and locally Lipschitz in A. In addition $\|u_m - u\|_{L^1(A)} \leq 1/m$ thanks to (5.10) and the monotone convergence theorem and, iterating (5.11) and using (5.6) and $k(h) \geq h$, we get

$$\begin{aligned} \int_A |\nabla u_m| \, \mathrm{d}\mathfrak{m} &= \lim_{h\to\infty} \int_{B_h} |\nabla u_{m,h+1}| \, \mathrm{d}\mathfrak{m} \leq \lim_{h\to\infty} \int_{B_{h+1}} |\nabla u_{m,h+1}| \, \mathrm{d}\mathfrak{m} \\ &\leq \sum_{i=0}^{\infty} |Du|_*(D_i) + \frac{2}{m} + \varepsilon \leq |Du|_*(A_{\bar{k}}) + 2\varepsilon + \frac{2}{m}. \end{aligned}$$

In order to prove the induction step in the construction of $u_{m,h}$ we use Lemma 5.4 with $M = B_h$, $N = D_{h+1}$, $u = u_{m,h}$ and $v = \psi_{k(h+1),h+1}$. So, applying (5.12) of the lemma we find a function $w = u_{m,h+1}$ such that

$$\int_{B_{h+1}} |\nabla u_{m,h+1}| \, \mathrm{d}\mathfrak{m} \leq \int_{D_{h+1}} |\nabla \psi_{k(h+1),h+1}| \, \mathrm{d}\mathfrak{m} + \int_{B_h} |\nabla u_{m,h}| \, \mathrm{d}\mathfrak{m} + c_h \int_{\overline{H}_h} |\psi_{k(h+1),h+1} - u_{m,h}| \, \mathrm{d}\mathfrak{m},$$

$$\begin{cases} u_{m,h+1} \equiv u_{m,h} & \text{on } B_h \setminus D_{h+1} \supseteq B_{h-1} \\ u_{m,h+1} \equiv \psi_{k(h+1),h+1} & \text{on } D_{h+1} \setminus B_h \supseteq B_{h+1} \setminus \overline{B}_h. \end{cases}$$

By the induction assumption, $u_{m,h} \equiv \psi_{k(h),h}$ on $B_h \setminus \overline{B}_{h-1}$ which contains $\overline{H}_h$, and so we can use (5.7) to get (5.11). Then (5.13) of Lemma 5.4 with $\sigma = u$ tells us exactly that

$$\int_{B_{h+1}} |u_{m,h+1} - u| \, d\mathfrak{m} \leq \int_{D_{h+1}} |\psi_{k(h+1),h+1} - u| \, d\mathfrak{m} + \int_{B_h} |u_{m,h} - u| \, d\mathfrak{m}$$

and so by (5.7) and the induction assumption we get also (5.10):

$$\int_{B_{h+1}} |u_{m,h+1} - u| \, d\mathfrak{m} \leq \frac{1}{m2^{h+1}} + \frac{1}{m}\left(1 - \frac{1}{2^h}\right) = \frac{1}{m}\left(1 - \frac{1}{2^{h+1}}\right).$$

Now we prove (ii). Having already proved (iii), suffices to show that

$$\beta(A_1' \cup A_2') \leq \beta(A_1) + \beta(A_2) \qquad \text{whenever } A_1' \Subset A_1,\ A_2' \Subset A_2.$$

This inequality can be obtained by applying Lemma 5.4 to join optimal sequences for A_1 and A_2, with $M = (A_1' \cup A_2') \cap A_1$ and $N = (A_1' \cup A_2') \cap A_2$. □

Remark 5.3 We note that the measure $|Df|^c_*$ too has the monotonicity property, thanks to the localizing property of upper gradients. Also, $|Df|^c_*$ is additive on disjoint open sets. In fact, if A_1 and A_2 are disjoint open sets we clearly have that $|Df|^c_*(A_1 \cup A_2) \geq |Df|^c_*(A_1) + |Df|^c_*(A_2)$, by the superadditivity of the lim inf. On the other hand, if we consider pairs (f_i, g_i), $i = 1, 2$, with $f_i \in L^1(A_i)$ and $g_i \in UG(f_i)$, then

$$g(x) = \begin{cases} g_1(x) & \text{if } x \in A_1 \\ g_2(x) & \text{if } x \in A_2 \end{cases} \quad \text{is an upper gradient for} \quad f(x) = \begin{cases} f_1(x) & \text{if } x \in A_1 \\ f_2(x) & \text{if } x \in A_2, \end{cases}$$

since every absolutely continuous curve $\gamma : [0,1] \to A_1 \cup A_2$ lies either entirely in A_1 or entirely in A_2, thanks to the connectedness of $\gamma([0,1])$; this joining property gives the converse inequality $|Df|^c_*(A_1 \cup A_2) \leq |Df|^c_*(A_1) + |Df|^c_*(A_2)$.

We won't need the σ-additivity property of $|Df|^c_*$ in the sequel; however, in the proof of Theorem 1.1 we gain the equality $|Df|_* = |Df|^c_*$ on open sets, and so we recover that $|Df|^c_*$ is the restriction to the open sets of a measure, too.

Lemma 5.4 (Joint lemma) *Let M, N be open sets such that $\mathsf{d}(N \setminus M, M \setminus N) > 0$. There exist an open set $H \Subset M \cap N$ and a constant c depending only on M and N such that for every $u \in \mathrm{Lip}_{\mathrm{loc}}(M)$, $v \in \mathrm{Lip}_{\mathrm{loc}}(N)$ we can find $w \in \mathrm{Lip}_{\mathrm{loc}}(M \cup N)$ such that*

$$\int_{M \cup N} |\nabla w| \, d\mathfrak{m} \leq \int_M |\nabla u| \, d\mathfrak{m} + \int_N |\nabla v| \, d\mathfrak{m} + c(M,N) \int_{\overline{H}} |u - v| \, d\mathfrak{m}; \tag{5.12}$$

$$w \equiv u \text{ on neighborhood of } M \setminus N, \qquad w \equiv v \text{ on neighborhood of } N \setminus M.$$

Furthermore, for every $\sigma \in L^1(M \cup N)$ we have

$$\int_{M\cup N} |w - \sigma| \,\mathrm{d}\mathfrak{m} \le \int_M |u - \sigma| \,\mathrm{d}\mathfrak{m} + \int_N |v - \sigma| \,\mathrm{d}\mathfrak{m}. \tag{5.13}$$

Proof. The assumption on M and N guarantees the existence of a Lipschitz function $\phi : X \to [0, 1]$ such that

$$\phi(x) = \begin{cases} 1 & \text{on a neighborhood of } M \setminus N \\ 0 & \text{on a neighborhood of } N \setminus M, \end{cases}$$

so that $H := \{0 < \phi < 1\} \cap (M \cup N)$ will be an open set contained in $M \cap N$ and well separated from both $M \setminus N$ and $N \setminus M$. Setting $\eta := \mathsf{d}(N \setminus M, M \setminus N)$, it is clear that we can have $\mathrm{Lip}(\phi) \le 3/\eta$; for example we can take

$$\phi(x) := \frac{3}{\eta} \min\left\{ \left(\mathsf{d}(x, N \setminus M) - \frac{\eta}{3}\right)^+, \frac{\eta}{3} \right\}.$$

Now we consider the function $w = \phi u + (1 - \phi)v$ and, using the convexity inequality for the slope $|\nabla w| \le \phi|\nabla u| + (1 - \phi)|\nabla v| + |\nabla\phi||u - v|$ (see [5] for its simple proof) and the fact that $\phi \le \mathbb{1}_M$ and $1 - \phi \le \mathbb{1}_N$ on $M \cup N$, splitting the integration on the interior of $\{\phi = 1\}$, the interior of $\{\phi = 0\}$ and $\overline{H}$ we end up with:

$$\int_{M\cup N} |\nabla w| \,\mathrm{d}\mathfrak{m} \le \int_M |\nabla u| \,\mathrm{d}\mathfrak{m} + \int_N |\nabla v| \,\mathrm{d}\mathfrak{m} + \frac{3}{\eta} \int_{\overline{H}} |u - v| \,\mathrm{d}\mathfrak{m}.$$

To prove (5.13) we simply note that $|w - \sigma| \le \phi|u - \sigma| + (1 - \phi)|v - \sigma|$ on $M \cup N$. □

5.3 Weak-BV functions

Before introducing the third definition we introduce some additional notation and terminology.

Definition 5.5 *A measure $\boldsymbol{\pi} \in \mathscr{P}(C([0, 1]; X))$ is said to be an ∞-*test plan *if the following two properties are satisfied:*

(a) $\boldsymbol{\pi}$ is concentrated on $AC^\infty([0, 1]; X)$ and $\mathrm{Lip}(\gamma)$ belongs to $L^\infty(C([0, 1]; X), \boldsymbol{\pi})$;

(b) there exists $C = C(\boldsymbol{\pi}) \ge 0$ such that $(e_t)_\sharp \boldsymbol{\pi} \le C\mathfrak{m}$ for each $t \in [0, 1]$.

A Borel subset Γ of $C([0, 1]; X)$ is said to be 1-negligible if $\boldsymbol{\pi}(\Gamma) = 0$ for every ∞-plan $\boldsymbol{\pi}$. A property of continuous curves is said to be true 1-almost everywhere if the set for which it is false is contained in a 1-negligible set.

This definition is the limit case of the one that occurs in [7], and also the definition of weak-BV is suggested in there. Given a function f in $L^1(X, \mathfrak{m})$, we say that f is a weak-BV function, and write $f \in w - BV(X, \mathsf{d}, \mathfrak{m})$, if the following two conditions are fulfilled:

(i) for 1-almost every curve we have that $f \circ \gamma \in BV(0,1)$; we require also a mild regularity at the boundary, namely

$$|f(\gamma_1) - f(\gamma_0)| \le |D(f \circ \gamma)|(0,1) \qquad \text{for 1-a.e. } \gamma, \tag{5.14}$$

where $|D(f\circ\gamma)| \in \mathcal{M}^-((0,1))$ is the total variation measure of the map $f\circ\gamma : [0,1] \to \mathbb{R}$;

(ii) there exists $\mu \in \mathcal{M}^+(X)$ such that

$$\int \gamma_\sharp |D(f \circ \gamma)|(B)\, d\boldsymbol{\pi}(\gamma) \le C(\boldsymbol{\pi}) \cdot \|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})} \mu(B) \qquad \forall B \in \mathcal{B}(X). \tag{5.15}$$

Associated to this notion, there is also the concept of weak total variation $|Df|_w$, defined as the least measure μ satisfying (5.15) for every ∞-test plan $\boldsymbol{\pi}$. Equivalently, $|Df|_w$ is the least upper bound, in the complete and separable lattice $\mathcal{M}^+(X)$, of the family of measures

$$\frac{1}{C(\boldsymbol{\pi})\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}} \int \gamma_\sharp |D(f \circ \gamma)|\, d\boldsymbol{\pi}(\gamma) \tag{5.16}$$

as $\boldsymbol{\pi}$ runs in the class of ∞-test plans.

If we fix $t \in (0,1)$ and we consider the rescaling map R_t from $C([0,1], X)$ to $C([0,1], X)$ mapping γ_s to γ_{ts}, we see that the push-forward $\boldsymbol{\pi}_t = (R_t)_\sharp \boldsymbol{\pi}$ is still a ∞-test plan, with $C(\boldsymbol{\pi}_t) \le C(\boldsymbol{\pi})$. In addition

$$\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi}_t)} \le t\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}.$$

By (5.14) we get

$$|f(\gamma_t) - f(\gamma_0)| \le |D(f \circ \gamma)|(0,t) \qquad \text{for } \boldsymbol{\pi}\text{-a.e. } \gamma, \tag{5.17}$$

while (5.15) with $A = X$ gives

$$\int |D(f \circ \gamma)|(0,t)\, \mathrm{d}\boldsymbol{\pi}(\gamma) = \int |D(f \circ \gamma)|(0,1)\, \mathrm{d}\boldsymbol{\pi}_t(\gamma) \le tC(\boldsymbol{\pi})\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}|Df|_w(X). \tag{5.18}$$

Now we prove that the class BV_*^c is contained in the class $w - BV$ and that $|Df|_w \le |Df|_*^c$ on open sets. The proof of this fact is not difficult, and reminiscent of the closure property of weak gradients in a Sobolev context, see [11, 28]. First of all, we state without proof the following elementary lemma:

Lemma 5.6 *Assume that g is an upper gradient of f, that $\gamma:[0,1]\to X$ is Lipschitz and that $\int_\gamma g<\infty$. Then $f\circ\gamma\in W^{1,1}(0,1)$ and $|(f\circ\gamma)'(t)|\le g(\gamma_t)|\dot\gamma_t|$ for a.e. $t\in(0,1)$. In particular*

$$|D(f\circ\gamma)|(B)\le \mathrm{Lip}(\gamma)\int_B g(\gamma_t)\,\mathrm{d}t \qquad \textit{for any Borel set } B\subseteq(0,1).$$

Given an open set $A\subseteq X$, we take a sequence of pairs (f_n,g_n) with $g_n\in UG(f_n)$ such that $f_n\to f$ in $L^1(A,\mathfrak{m})$ and $\int_A g_n\mathrm{d}\mathfrak{m}\to|Df|_*^c(A)$ (whose existence is granted by the definition of Cheeger total variation), and use the lemma to estimate the weak total variation of f_n as follows:

$$\begin{aligned}\int \gamma_\sharp|D(f_n\circ\gamma)|(A)\,\mathrm{d}\boldsymbol{\pi}(\gamma)&=\int|D(f_n\circ\gamma)|(\gamma^{-1}(A))\,\mathrm{d}\boldsymbol{\pi}(\gamma)\\&\le\int\mathrm{Lip}(\gamma)\int_0^1 g_n(\gamma_t)\chi_A(\gamma_t)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}(\gamma)\\&\le\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}\int_0^1\int_A g_n\,\mathrm{d}(\mathrm{e}_t)_\sharp\boldsymbol{\pi}\,\mathrm{d}t\\&\le\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}C(\boldsymbol{\pi})\int_A g_n\,\mathrm{d}\mathfrak{m}.\end{aligned}\tag{5.19}$$

We now introduce a lemma that permits us, up to a subsequence, to localize the L^1 convergence, so that we can estimate the left hand side.

Lemma 5.7 *Let $B\subseteq X$ be a Borel set and let (f_n) be a sequence converging to f in $L^1(B,\mathfrak{m})$. Then, a subsequence of (f_n) converges to f in $L^1(\gamma^{-1}(B),\mathscr{L}^1)$ along 1-almost every curve.*

Proof. We can assume without loss of generality that $B=X$. Possibly extracting a subsequence, we can suppose that

$$\sum_n\|f_n-f\|_{L^1(X,\mathfrak{m})}<\infty.$$

We now fix a ∞-test plan $\boldsymbol{\pi}$ and we show that $\|f_n\circ\gamma-f\circ\gamma\|_{L^1(0,1)}\to0$ for $\boldsymbol{\pi}$-almost every curve γ. Our choice of the subsequence ensures that the function $g:=\sum_n|f_n-f|$ belongs to $L^1(0,1)$. Now, the inequality

$$\int\|g\circ\gamma\|_{L^1(0,1)}\,\mathrm{d}\boldsymbol{\pi}(\gamma)=\iint_0^1(g\circ\gamma)(t)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{\pi}\le C(\boldsymbol{\pi})\int_0^1\int_X g\,\mathrm{d}\mathfrak{m}<\infty$$

guarantees that $g\circ\gamma$ belongs to $L^1(0,1)$ for $\boldsymbol{\pi}$-a.e. curve γ and thus we can say that $f_n\circ\gamma\to f\circ\gamma$ in $L^1(0,1)$ for $\boldsymbol{\pi}$-a.e. γ. By the arbitrariness of $\boldsymbol{\pi}$, we conclude. □

We can now complete the proof of $|Df|_w \le |Df|_*^c$ on open sets, starting from (5.19). Notice that until now we still don't know whether $|Df|_*^c$ is a measure or not, however we can use the additivity properties ensured by Remark 5.3.

Let $A \subseteq X$ be an open set, let (f_n) be a sequence convergent to f in $L^1(A)$ and $g_n \in UG(f_n)$ such that

$$\lim_{n\to\infty} \int_A g_n \, d\mathfrak{m} = |Df|_*^c(A).$$

Thanks to Lemma 5.7 we can find a subsequence $n(s)$ such that $f_{n(s)} \circ \gamma \to f \circ \gamma$ in $L^1(\gamma^{-1}(A))$ along 1-almost every curve γ. By (2.4) in the open set $\gamma^{-1}(A)$ we get

$$\gamma_\sharp |D(f\circ\gamma)|(A) \le \liminf_{s\to\infty} \gamma_\sharp |D(f_{n(s)}\circ\gamma)|(A) \qquad \text{for } \boldsymbol{\pi}\text{-a.e. curve } \gamma.$$

Passing to the limit as $s \to \infty$ in the inequality (5.19) with $n = n(s)$, Fatou's lemma gives $\mu_{\boldsymbol{\pi}}(A) \le |Df|_*^c(A)$ for all ∞-test plan $\boldsymbol{\pi}$, where μ_π is the finite Borel measure in (5.16). If $\pi_1, \ldots, \pi_k$ is a finite collection of ∞-test plans, the formula

$$\bigvee_{i=1}^k \mu_{\pi_i}(A) = \sup \left\{ \sum_{i=1}^k \mu_{\pi_i}(A_i) : \ A_1 \subseteq A, \ldots, A_k \subseteq A \text{ open, pairwise disjoint} \right\}$$

and the additivity of $|Df|_*^c$ yield $|Df|_*^c(A) \ge \vee_1^k \mu_{\pi_i}(A)$ for any open set A. Since this collection is arbitrary, the inequality $|Df|_w(A) \le |Df|_*^c(A)$ is proved.

We're not done yet, because we have to prove also the boundary regularity (5.14) that is part of our axiomatization of $w - BV$ functions. The inequality would clearly follow if we show that $f \circ \gamma_i$, $i = 0, 1$, is the approximate limit of $f \circ \gamma$ as $t \to i$, namely

$$\lim_{t\downarrow 0} \frac{1}{t} \int_0^t |f(\gamma_s) - f(\gamma_0)| \, \mathrm{d}s = 0, \qquad \lim_{t\downarrow 0} \frac{1}{t} \int_{1-t}^1 |f(\gamma_s) - f(\gamma_1)| \, \mathrm{d}s = 0.$$

This is indeed the context of the next lemma, that we state and prove for $t = 0$ only:

Lemma 5.8 (Boundary regularity) *We are given a sequence of pairs (f_n, g_n) where $g_n \in UG(f_n)$, $f_n \to f$ in $L^1(X, \mathfrak{m})$ and $\sup_n \int_X g_n \mathrm{d}\mathfrak{m} < \infty$. Then $t = 0$ is a Lebesgue point for the map $f \circ \gamma : [0,1] \to \mathbb{R}$ for 1-almost every curve γ.*

Proof. Let us fix an ∞-plan $\boldsymbol{\pi}$, set $C_1 := \sup_n \int_X g_n \mathrm{d}\mathfrak{m}$, $C_2 := C(\boldsymbol{\pi})$ and consider the quantities

$$H_t(\gamma) = \frac{1}{t} \int_0^t |f(\gamma_s) - f(\gamma_0)| \mathrm{d}s.$$

By definition, we know that 0 is a Lebesgue point for $f \circ \gamma$ if $H_t(\gamma) \to 0$ as $t \to 0$. Applying Fatou's lemma we get:

$$\int \liminf_{t\to 0} H_t(\gamma) \mathrm{d}\boldsymbol{\pi} \le \liminf_{t\to 0} \int H_t(\gamma) \, \mathrm{d}\boldsymbol{\pi}. \tag{5.20}$$

We can estimate

$$\int H_t(\gamma)\,\mathrm{d}\boldsymbol{\pi} \le \int H_t^n(\gamma)\,\mathrm{d}\boldsymbol{\pi} + \frac{1}{t}\iint_0^t \Big(|f_n(\gamma_s)-f(\gamma_s)| + |f_n(\gamma_0)-f(\gamma_0)|\Big)\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi},$$

where $H_t^n(\gamma) = \frac{1}{t}\int_0^t |f_n(\gamma_s)-f_n(\gamma_0)|\mathrm{d}s$. We now treat separately the two terms on the right: first let's note that

$$\begin{aligned}\int H_t^n(\gamma)\mathrm{d}\boldsymbol{\pi} &= \frac{1}{t}\iint_0^t |f_n(\gamma_s)-f_n(\gamma_0)|\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi} \le \frac{1}{t}\iint_0^t \int_0^s g_n(\gamma_r)\mathrm{d}r\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}\\ &\le \frac{1}{t}\iint_0^t \int_0^t g_n(\gamma_r)\mathrm{d}r\,\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi} = \iint_0^t g_n(\gamma_r)\mathrm{d}r\,\mathrm{d}\boldsymbol{\pi}\\ &\le C_2\int_0^t\int_X g_n(x)\,\mathrm{d}\mathfrak{m}(x)\,\mathrm{d}\mathfrak{m}\,\mathrm{d}t \le tC_1C_2.\end{aligned}$$

For the second term:

$$\begin{aligned}&\frac{1}{t}\iint_0^t \Big(|f_n(\gamma_s)-f(\gamma_s)| + |f_n(\gamma_0)-f(\gamma_0)|\Big)\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi}\\ &= \frac{1}{t}\iint_0^t |f_n(\gamma_s)-f(\gamma_s)|\mathrm{d}s\,\mathrm{d}\boldsymbol{\pi} + \int |f_n(\gamma_0)-f(\gamma_0)|\,\mathrm{d}\boldsymbol{\pi}\\ &\le \frac{1}{t}\int_0^t\int_X |f_n-f|\cdot C_2\,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \int_X |f_n-f|\cdot C_2\,\mathrm{d}\mathfrak{m}\\ &\le 2C_2\cdot\|f_n-f\|_{L^1(X,\mathfrak{m})}.\end{aligned}$$

summing up we get that, choosing n so large that $\|f_n-f\|_{L^1}\le t$,

$$\int H_t(\gamma)\mathrm{d}\boldsymbol{\pi} \le tC_2(C_1+2).$$

Now by (5.20) we conclude that $\int(\liminf_{t\to 0} H_t)\,\mathrm{d}\boldsymbol{\pi} = 0$ and so, thanks to the arbitrariness of $\boldsymbol{\pi}$, we can say that 0 is a Lebesgue point for 1-almost every curve. □

We conclude this section with an auxiliary result regarding weak BV functions.

Lemma 5.9 (Truncations) *Let $f\in w-BV(X,\mathsf{d},\mathfrak{m})$ and $N\in\mathbb{R}$. Then $f\wedge N$ and $f\vee -N$ belong to $w-BV(X,\mathsf{d},\mathfrak{m})$ and*

$$|D(f\wedge N)|_w(X)\le |Df|_w(X), \qquad |D(f\vee -N)|_w(X)\le |Df|_w(X).$$

Proof. It relies on the fact that $|D(\psi\circ g)|(0,1)\le|Dg|(0,1)$ whenever $g\in BV(0,1)$ and $\psi:\mathbb{R}\to\mathbb{R}$ is 1-Lipschitz. □

6 The functional Ch_1 and its gradient flow

Let us consider the convex functional $\mathrm{Ch}_1 : L^2(X, \mathfrak{m}) \to [0, \infty]$ given by

$$\mathrm{Ch}_1(f) := \begin{cases} |Df|_*(X) & \text{if } f \in BV_*(X, \mathsf{d}, \mathfrak{m}); \\ \infty & \text{if } f \notin BV_*(X, \mathsf{d}, \mathfrak{m}), \end{cases} \tag{6.1}$$

where $|Df|_*(X)$ has beed defined in the previous section. Convexity of Ch_1 follows by the more precise inequality between measures

$$|D(\lambda f + \mu g)|_* \le |\lambda||Df|_* + |\mu||Dg|_* \tag{6.2}$$

which simply follows (first on open sets, and then on Borel sets) by homogeneity and convexity of $f \mapsto |\nabla f|(x)$. Also, a simple diagonal argument shows that Ch_1 is lower semicontinuous w.r.t. $L^2(X, \mathfrak{m})$ convergence. In addition its domain

$$D(\mathrm{Ch}_1) = BV_*(X, \mathsf{d}, \mathfrak{m}) \cap L^2(X, \mathfrak{m})$$

is dense in $L^2(X, \mathfrak{m})$, because it contains $\mathrm{Lip}_b(X)$. Thanks to these facts we can apply the standard theory of gradient flows [8] of convex lower semicontinuous functionals in Hilbert spaces to obtain, starting from any $f_0 \in L^2(X, \mathfrak{m})$, a curve f_t such that:

(a) $t \mapsto f_t$ is locally Lipschitz from $(0, \infty)$ to $L^2(X, \mathfrak{m})$ and $f_t \to f_0$ as $t \downarrow 0$;

(b) $t \mapsto \mathrm{Ch}_1(f_t)$ is locally absolutely continuous in $(0, \infty)$;

(c) $\frac{\mathrm{d}}{\mathrm{d}t} f_t = \Delta_1 f_t$ for a.e. $t \in (0, \infty)$.

Here $\Delta_1 f$ denotes the 1-laplacian of f, defined as the opposite of the element of minimal norm of the subdifferential $\partial^- \mathrm{Ch}_1(f)$, when this set is not empty. Namely, $\xi = -\Delta_1 f$ satisfies

$$\mathrm{Ch}_1(g) \ge \mathrm{Ch}_1(f) + \int_X \xi (g - f)\,\mathrm{d}\mathfrak{m} \qquad \forall g \in L^2(X, \mathfrak{m}) \tag{6.3}$$

and is the vector with smallest $L^2(X, \mathfrak{m})$ norm among those with this property. We will denote by $D(\Delta_1)$ the set of functions for which the subdifferential is not empty.

We can think of the gradient flow also as a semigroup S_t that maps f_0 in f_t. When $\mathfrak{m}(X)$ is finite, a property that will be used is that $S_t(f_0 + C) = S_t(f_0) + C$ for all $C \in \mathbb{R}$; this is true because Ch_1 is invariant by addition of a constant and so also $\partial^- \mathrm{Ch}_1$ has the same property.

Proposition 6.1 (Integration by parts) *For all $f \in D(\Delta_1)$ and $g \in D(\mathrm{Ch}_1)$ it holds*

$$-\int_X g \Delta_1 f\,\mathrm{d}\mathfrak{m} \le |Dg|_*(X) = \mathrm{Ch}_1(g), \tag{6.4}$$

with equality if $g = f$.

Proof. Since $-\Delta_1 f \in \partial^- \mathrm{Ch}_1(f)$ it holds

$$\mathrm{Ch}_1(f) - \int_X g\Delta_1 f \,\mathrm{d}\mathfrak{m} \le \mathrm{Ch}_1(f+g), \qquad \forall g \in L^2(X,\mathfrak{m}).$$

Now we can use (6.2) to estimate $\mathrm{Ch}_1(f+g)$ with $\mathrm{Ch}_1(f)+\mathrm{Ch}_1(g)$, and so we get the first statement. For the second statement we need the converse inequality when $f=g$; but this is easy, because it is sufficient to put $g=0$ in (6.3). □

Proposition 6.2 (Some properties of the gradient flow of Ch_1) *Let $f_0 \in L^2(X,\mathfrak{m})$ and let (f_t) be the gradient flow of Ch_1 starting from f_0. Then:*
(Mass preservation) $\int f_t \,\mathrm{d}\mathfrak{m} = \int f_0 \,\mathrm{d}\mathfrak{m}$ for any $t \ge 0$.
(Maximum principle) If $f_0 \le C$ (resp. $f_0 \ge c$) $\mathfrak{m}$-a.e. in X, then $f_t \le C$ (resp $f_t \ge c$) $\mathfrak{m}$-a.e. in X for any $t \ge 0$.
(Energy dissipation) Suppose $0 < c \le f_0 \le C < \infty$ $\mathfrak{m}$-a.e. in X and $\Phi \in C^2([c,C])$. Then $t \mapsto \int \Phi(f_t)\,\mathrm{d}\mathfrak{m}$ is locally absolutely continuous in $(0,\infty)$ and it holds

$$-\frac{\mathrm{d}}{\mathrm{d}t}\int \Phi(f_t)\,\mathrm{d}\mathfrak{m} \le |D\Phi'(f_t)|_* \qquad \textit{for a.e. } t \in (0,\infty),$$

with equality if $\Phi(t) = t^2$.

Proof. (Mass preservation) Just notice that from (6.4) we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \pm\mathbf{1} f_t \,\mathrm{d}\mathfrak{m} = \int \pm\mathbf{1}\cdot\Delta_1 f_t \,\mathrm{d}\mathfrak{m} \le |D(\pm\mathbf{1})|_*(X) = 0 \quad \text{for a.e. } t > 0,$$

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

$$g \qquad \mapsto \qquad \mathrm{Ch}_1(g) + \frac{1}{2\tau}\int_X |g-f|^2 \,\mathrm{d}\mathfrak{m}.$$

Assume that $f \le C$. We claim that in this case $f^\tau \le C$ as well. Indeed, if this is not the case we can consider the competitor $g := \min\{f^\tau, C\}$ in the above minimization problem. By Lemma 5.9 we get $\mathrm{Ch}(g) \le \mathrm{Ch}(f^\tau)$ and the L^2 distance of f and g is strictly smaller than the one of f and f^τ as soon as $\mathfrak{m}(\{f^\tau > C\}) > 0$, which is a contradiction. Starting from f_0, iterating this procedure, and using the fact that the implicit Euler scheme converges as $\tau \downarrow 0$ (see [8], [4] for details) to the gradient flow we get the conclusion.
(Energy dissipation) Since $t \mapsto f_t \in L^2(X,\mathfrak{m})$ is locally absolutely continuous and, by the maximum principle, f_t take their values in $[c,C]$ $\mathfrak{m}$-a.e., from the fact that Φ is Lipschitz in $[c,C]$ we get the claimed absolute continuity statement. Now, we know from the Lagrange mean value theorem that exists a function $\xi_t^h : X \to [c,C]$ such that:

$$\Phi(f_{t+h}) - \Phi(f_t) = \Phi'(f_t)(f_{t+h}-f_t) + \frac{1}{2}\Phi''(\xi_t^h)(f_{t+h}-f_t)^2.$$

Dividing by h and integrating in space, we get that, for times where the L^2 derivative of f_t exists (i.e., for almost every t):

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_X \Phi(f_t)\,\mathrm{d}\mathfrak{m} = \int_X \Phi'(f_t)\Delta_1 f_t\,\mathrm{d}\mathfrak{m}.$$

We can now use Lemma 6.1 with $g = \Phi'(f_t)$ in the right hand side to get the last statement. □

7 Proof of equivalence

In Section 5 we discussed the "easy" inclusions $BV_* \subseteq BV_*^c \subseteq w - BV$, and the corresponding inequalities (localized on open subsets of X)

$$|Df|_w \le |Df|_*^c \le |Df|_*.$$

In this section we prove the main result of the paper, namely the equivalence of the three definitions. So, we have to start from a function $f \in w-BV(X, \mathsf{d}, \mathfrak{m})$, and build a sequence of approximating Lipschitz functions in such a way that

$$\limsup_{n\to\infty}\int_X |\nabla f_n|\,\mathrm{d}\mathfrak{m} \le |Df|_w(X). \tag{7.1}$$

As in [5] for the case $q = 2$ and [7] for the case $1 < q < \infty$, our main tool in the construction will the gradient flow in $L^2(X,\mathfrak{m})$ of the functional Ch_1, starting from f_0. We initially assume that (X, d) is a complete and separable length space (this assumption is used to be able to apply the results of Section 3 and in Lemma 7.3, to apply (4.1)) and that $\mathfrak{m}$ is a finite Borel measure, so that the L^2-gradient flow of Ch_1 can be used. The finiteness and length space assumptions will be eventually removed in the proof of the equivalence result.

We start with the following proposition, which relates energy dissipation to a sharp combination of weak total variation and metric dissipation in W_∞.

Proposition 7.1 *Let $\mu_t = f_t\mathfrak{m}$ be a curve in $AC^\infty([0,1],(\mathcal{M}_+(X),W_\infty))$. Assume that for some $0 < c < C < \infty$ it holds $c \le f_t \le C$ $\mathfrak{m}$-a.e. in X for any $t \in [0,1]$, and that $f_0 \in w - BV(X,\mathsf{d},\mathfrak{m})$. Then for all $\Phi \in C^2([c,C])$ convex it holds*

$$\int \Phi(f_0)\,\mathrm{d}\mathfrak{m} - \int \Phi(f_s)\,\mathrm{d}\mathfrak{m} \le s\mathrm{Lip}(\Phi')|Df_0|_w(X)\cdot C\cdot \mathrm{Lip}(\mu_t) \qquad \forall s > 0.$$

Proof. Let $m = \int_X f_0\,\mathrm{d}\mathfrak{m}$, and let $\boldsymbol{\pi} \in \mathcal{M}_+(C([0,1],X))$ be a plan associated to the curve (μ_t) as in Remark 8.5. The assumption $f_t \le C$ $\mathfrak{m}$-a.e. and the fact that $\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})} = \mathrm{Lip}(\mu_t) < \infty$ guarantee that $\frac{\boldsymbol{\pi}}{m}$ is an ∞-test plan, such that $C(\frac{\boldsymbol{\pi}}{m}) \le C/m$.

Now we get, using our hypothesis that $f_0 \in w-BV$ and (5.17), (5.18):

$$
\begin{aligned}
\int \Phi(f_0) - \int \Phi(f_s)\,\mathrm{d}\mathfrak{m} &\le \int \Phi'(f_0)(f_0 - f_s)\,\mathrm{d}\mathfrak{m} = \int \Phi'(f_0)\circ \mathrm{e}_0 - \Phi'(f_0)\circ \mathrm{e}_s\,\mathrm{d}\boldsymbol{\pi}\\
&\le \int |\Phi'(f_0(\gamma_s)) - \Phi'(f_0(\gamma_0))|\,\mathrm{d}\boldsymbol{\pi}(\gamma)\\
&\le \mathrm{Lip}(\Phi') \int |f_0(\gamma_s) - f_0(\gamma_0)|\,\mathrm{d}\boldsymbol{\pi}(\gamma)\\
&\le m\cdot \mathrm{Lip}(\Phi') \int |D(f_0\circ\gamma)|(0,s)\,\mathrm{d}\left(\frac{\boldsymbol{\pi}}{m}\right)(\gamma)\\
&\le \mathrm{Lip}(\Phi')\cdot s\cdot |Df_0|_w(X)\cdot m\cdot C(\frac{\boldsymbol{\pi}}{m})\cdot \|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}\\
&\le \mathrm{Lip}(\Phi')\cdot s\cdot |Df_0|_w(X)\cdot C\cdot \mathrm{Lip}(\mu_t).
\end{aligned}
$$

□

Remark 7.2 The proof of Proposition 7.1 suggests another possible definition of $w-BV(X,\mathsf{d},\mathfrak{m})$, in the spirit of the classical definitions of BV based on the oscillation, namely, requiring the existence of a constant C_* satisfying

$$
\int |f(\gamma_1) - f(\gamma_0)|\,d\boldsymbol{\pi}(\gamma) \le C_* C(\boldsymbol{\pi})\cdot \|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}
$$

for any ∞-test plan $\boldsymbol{\pi}$. A posteriori, the constant C_* coincides with $|Df|_w(X) = |Df|_*(X)$. Even though this definition is simpler, we have chosen the definition based on (5.15) because it involves explictly a measure, which can be compared with the definition arising from approximation with Lipschitz functions.

The key argument to achieve the identification is the following lemma which gives a sharp bound on the W_∞-speed of the L^2-gradient flow of Ch_1. This lemma, in the W_p case, has been introduced in [25] and then used in [17, 5] to study the heat flow on metric measure spaces.

Lemma 7.3 (Kuwada's lemma for Ch_1) *Let $f_0 \in L^2(X,\mathfrak{m})$ and let (f_t) be the gradient flow of Ch_1 starting from f_0. Assume that for some $0 < c < C < \infty$ it holds $c \le f_0 \le C$ $\mathfrak{m}$-a.e. in X. Then the curve $t \mapsto \mu_t := f_t\mathfrak{m} \in \mathcal{M}_+(X)$ is absolutely continuous w.r.t. W_∞ and it holds*

$$
|\dot\mu_t| \le \frac{1}{c} \qquad \text{for a.e. } t \in (0,\infty).
$$

Proof. We start from the duality formula (4.1)

$$
W_\infty^{(s)}(\mu,\nu) = \sup_{\phi\in \mathrm{Lip}_b(X)} \int_X Q_s\phi\,d\nu - \int_X \phi\,d\mu. \tag{7.2}
$$

where $Q_t\phi$ is defined in (3.1) and (3.2). Fix $\phi \in \mathrm{Lip}_b(X)$ and recall (Theorem 3.1) that the map $t \mapsto Q_t\varphi$ is Lipschitz with values in $C(X)$, in particular also as a $L^2(X, \mathfrak{m})$-valued map.

Fix also $0 \le t < r$, set $\ell = (r - t)$ and recall that since (f_t) is a gradient flow of Ch_1 in $L^2(X, \mathfrak{m})$, the map $[0, \ell] \ni \tau \mapsto f_{t+\tau}$ is absolutely continuous with values in $L^2(X, \mathfrak{m})$. Therefore, since both factors are uniformly bounded, the map $[0, \ell] \ni \tau \mapsto Q_{\frac{s\tau}{\ell}}\varphi f_{t+\tau}$ is absolutely continuous with values in $L^2(X, \mathfrak{m})$. In addition, the equality

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

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

We have:

$$\begin{aligned}\int_X Q_s\varphi \,\mathrm{d}\mu_r - \int_X \varphi \,\mathrm{d}\mu_t = \int Q_s\varphi f_{t+\ell}\,\mathrm{d}\mathfrak{m} - \int_X \varphi f_t \,\mathrm{d}\mathfrak{m} = \int_X\int_0^\ell \frac{\mathrm{d}}{\mathrm{d}\tau}\big(Q_{\frac{s\tau}{\ell}}\varphi f_{t+\tau}\big)d\tau\,\mathrm{d}\mathfrak{m}\\ \le \int_X\int_0^\ell -\frac{s}{\ell}|\nabla Q_{\frac{s\tau}{\ell}}\varphi| f_{t+\tau} + Q_{\frac{s\tau}{\ell}}\varphi\Delta_1 f_{t+\tau}\,\mathrm{d}\tau\,\mathrm{d}\mathfrak{m},\end{aligned} \tag{7.3}$$

having used Theorem 3.1.

Observe that by inequality (6.4) and Remark 5.1 we have

$$\int_X Q_{\frac{s\tau}{\ell}}\varphi\Delta_1 f_{t+\tau}\,\mathrm{d}\mathfrak{m} \le \int_X |\nabla Q_{\frac{s\tau}{\ell}}\varphi|_{*,1}\,\mathrm{d}\mathfrak{m} \le \int_X |\nabla Q_{\frac{s\tau}{\ell}}\varphi|\,\mathrm{d}\mathfrak{m}. \tag{7.4}$$

Plugging this inequality in (7.3), and taking $s = \frac{\ell}{c}$ we obtain

$$\int_X Q_s\varphi\,\mathrm{d}\mu_r - \int_X \varphi\,\mathrm{d}\mu_t \le \int_0^\ell \int_X |\nabla Q_{\frac{s\tau}{\ell}}\varphi|\Big(1 - \frac{sf_{t+\tau}}{\ell}\Big)\,\mathrm{d}\mathfrak{m}\,\mathrm{d}\tau \le 0$$

This latter bound obviously doesn't depend on φ, so from (7.2) we deduce

$$W_\infty(\mu_t, \mu_r) \le \frac{(r - t)}{c}.$$

In particular, we showed that the curve μ_t is $\frac{1}{c}$-Lipschitz. □

We can now prove our main theorem:

Proof. [of Theorem 1.1] Recalling the results of Section 5, to conclude the proof we are only left to show that a function of weak total variation is also a function of relaxed total variation and the two definitions of total variations bring us to the same measure. We first prove that $|Df|_*(X) \le |Df|_w(X)$ and then that the set functions agree on all open sets. This yields the coincidence of the two measures on the Borel σ-algebra.

We split the proof of the inequality $|Df|_*(X) \le |Df|_w(X)$ in three parts: we prove it first for bounded functions and finite measures in length spaces, then we remove the boundedness assumption on f and the length space assumption, and eventually the local finiteness assumption on $\mathfrak{m}$.
Let us consider a bounded function $f_0 \in BV_w$ possibly adding a constant (that doesn't change any of the total variations) we can suppose also that $C \ge f_0 \ge c > 0$. Let us consider as before the gradient flow f_t in $L^2(X, \mathfrak{m})$, with respect to Ch_1, starting from f_0. Now, let $\Phi(x) = x^2$, so that $\Phi'' \equiv 2$, and let's substitute f_0 with $f_0 + H$; our computation is left unchanged, because we know that $S_t(f_0 + H) = f_t + H$ and so we can say, using the energy estimate in Proposition 6.2 and the Lipschitz estimate for the curve $t \mapsto (f_t + H)\mathfrak{m}$ given by Lemma 7.3, combined with Proposition 7.1:

$$\begin{aligned} 2\int_0^s |Df_t|_*(X)\,\mathrm{d}t &= 2\int_0^s |D(f_t+H)|_*(X)\,\mathrm{d}t \\ &= \int_X (f_0+H)^2\,\mathrm{d}\mathfrak{m} - \int_X (f_s+H)^2\,\mathrm{d}\mathfrak{m} \\ &\le 2s \cdot |Df_0|_w(X) \cdot \frac{C+H}{c+H}. \end{aligned}$$

Now, letting $H \to \infty$, we get that

$$\int_0^s |Df_t|_*(X)\,\mathrm{d}t \le s \cdot |Df_0|_w(X).$$

But, knowing that $|Df_t|_*(X) = \mathrm{Ch}_1(f_t)$ is nonincreasing in t we can say

$$s|Df_s|_*(X) \le \int_0^s |Df_t|_*(X)\,\mathrm{d}t \le s \cdot |Df_0|_w(X)$$

and thus $|Df_s|_*(X) \le |Df_0|_w(X)$. Now we have that $|Df|_*$ is lower semicontinuous and so, letting $s \downarrow 0$, we obtain that $f_0 \in BV_*^c(X, \mathsf{d}, \mathfrak{m})$ and that $|Df_0|_*(X) \le |Df_0|_w(X)$.

Now, taking any function $g \in w - BV(X, \mathsf{d}, \mathfrak{m})$, defining $g^N = (g \wedge N) \vee (-N)$, we have $g^N \to g$ in L^1 as N goes to infinity; by the lower semicontinuity of the relaxed total variation we get:

$$|Dg|_w(X) \ge \limsup |Dg^N|_w(X) = \limsup |Dg^N|_*(X) \ge |Dg|_*(X),$$

where the first inequality follows by Lemma 5.9, while the equality is what we proved in the first step, i.e. the thesis for bounded functions.

Now, still assuming $\mathfrak{m}$ finite, we prove that $|Df|_*(A) \le |Df|_w(A)$ for any open set $A \subseteq X$. In fact, the superadditivity of $|Df|_*$ gives

$$\begin{aligned} |Df|_*(A) &\le |Df|_*(X) - |Df|_*(X \setminus \overline{A}) = |Df|_w(X) - |Df|_*(X \setminus \overline{A}) \\ &\le |Df|_w(X) - |Df|_w(X \setminus \overline{A}) = |Df|_w(\overline{A}). \end{aligned}$$

Eventually we use that $|Df|_*$ satisfies (iii) of Lemma 5.2 to conclude that $|Df|_*(A) \leq |Df|_w(A)$ for any open set A.

Now, still assuming $\mathfrak{m}$ to be finite, we see how the length space assumption on X can be easily removed. Indeed, it is not difficult to find an isometric embedding of (X, d) into a complete, separable and length metric space (Y, d_Y): for instance one can use the canonical Kuratowski isometric embedding j of (X, d) into ℓ_∞ and then take as Y the closed convex hull of $j(X)$. For notational simplicity, just assume that $X \subseteq Y$ and that d_Y restricted to $X \times X$ coincides with d. Since X is a closed subset of Y, we may also view $\mathfrak{m}$ as a finite Borel measure in Y supported in X. Then, if $f \in w - BV(X, \mathsf{d}, \mathfrak{m})$, we have also $f \in w - BV(Y, \mathsf{d}_Y, \mathfrak{m})$ and $|Df|_{w,Y}(B) \leq |Df|_{w,X}(B \cap X)$ for any Borel set $B \subseteq Y$, because any ∞-test plan $\boldsymbol{\pi}$ in Y is, by the condition $(\mathrm{e}_t)_\sharp \boldsymbol{\pi} \leq \mathfrak{m}$, supported on Lipschitz curves with values in X. Then, applying the equivalence result in $(Y, \mathsf{d}_Y, \mathfrak{m})$, we find a sequence of locally Lipschitz functions $g_n : Y \to \mathbb{R}$ convergent to f in $L^1(Y, \mathfrak{m})$ satisfying

$$\limsup_{n \to \infty} \int_Y |\nabla g_n| \, \mathrm{d}\mathfrak{m} \leq |Df|_{w,Y}(Y) \leq |Df|_{w,X}(X).$$

Now, if $f_n = g_n|_X$, from the inequality $|\nabla f_n| \leq |\nabla g_n|$ on X we obtain $\limsup_n \int_X |\nabla f_n| \, \mathrm{d}\mathfrak{m} \leq |Df|_{w,X}(X)$. On the other hand, it is immediate to check that f_n are locally Lipschitz in X.

Eventually we show that the theorem is true for all locally finite measures $\mathfrak{m}$. Recall that $\mathfrak{m}$ is said to be locally finite if for any $x \in X$ there exists $r > 0$ such that $\mathfrak{m}(B_r(x)) < \infty$. By the Lindelöf property we can find a sequence of balls $B_{r_i}(x_i)$ with finite $\mathfrak{m}$-measure and $\mathfrak{m}$-negligible boundary whose union is the whole of X. Now, defining

$$A_h := \bigcup_{i=1}^{h} B_{r_i}(x_i),$$

we have a nondecreasing sequence of open sets A_h whose union is X.

Now, notice that the space $w - BV(X, \mathsf{d}, \mathfrak{m})$ satisfies the following global-to-local property: if $f \in w - BV(X, \mathsf{d}, \mathfrak{m})$ and $\mathfrak{m}_C(B) = \mathfrak{m}(C \cap B)$, then $f \in w - BV(C, \mathsf{d}, \mathfrak{m}_C)$ and $|Df|_{\mathfrak{m}_C, w}(X) \leq |Df|_w(X)$ for all closed subsets C of X (this is due to the fact that ∞-test plans in relative to $\mathfrak{m}_C$ can be viewed also as ∞-test plans relative to $\mathfrak{m}$). Then, we can apply first the global-to-local property to all measures $\mathfrak{m}_{C_n}$ relative to the closed sets $C_n := \overline{A}_n$ and then the equivalence theorem for finite measures to obtain that $|Df|_*(A_n) = |Df|_{\mathfrak{m}_{C_n},*}(A_n)$ is uniformly bounded by $|Df|_w(X)$. Eventually we can use Lemma 5.2(iii) to obtain that $|Df|_*(X) = \sup_n |Df|_*(A_n)$ is finite. $\square$

The following example shows that in general the supremum in (1.2) may be strictly smaller than the absolutely continuous parts of $|Df|_w$.

Example 7.4 Let $X = \mathbb{R}^2$, let B be the closed unit ball in $\mathbb{R}^2$, d the Euclidean distance and $\mathfrak{m}(C) = \mathscr{L}^2(C) + \mathscr{H}^1(C \cap \partial B)$, for $C \subseteq X$ Borel. If f is the characteristic function of B, the inequality $\mathfrak{m} \geq \mathscr{L}^2$ gives the inequality between measures $|Df| \leq |Df|_w$. We claim that the two measures coincide. To see this, suffices to show that $|Df|_w(\mathbb{R}^2) \le 2\pi$ and this inequality follows easily by considering the sequence of functions (each one constant in a neighbourhood of ∂B) $f_n(x) = \phi_n(|x|)$ with

$$\phi_n(t) := \begin{cases} 1 & \text{if } t \le 1 + \frac{1}{n}; \\ 1 - n\big(t - 1 - \frac{1}{n}\big) & \text{if } 1 + \frac{1}{n} < t \le 1 + \frac{2}{n}; \\ 0 & \text{if } t > 1 + \frac{2}{n}. \end{cases}$$

Since $|Df|(C) = \mathscr{H}^1(C \cap \partial B)$, it follows that $|Df|_w$ is absolutely continuous w.r.t. $\mathfrak{m}$; on the other hand, since f is a characteristic function the same is true for the maps $f \circ \gamma$, so that $|D^a(f \circ \gamma)| = 0$ whenever $f \circ \gamma$ has bounded variation.

We conclude this section with the following corollary to Theorem 1.1, dealing with the degenerate case $L^1 = BV_*$; similar results could be stated also at the level of the Sobolev spaces $W^{1,q}(X, \mathsf{d}, \mathfrak{m})$ and the corresponding test plans of [7].

Corollary 7.5 *$BV_*(X, \mathsf{d}, \mathfrak{m})$ coincides with $L^1(X, \mathfrak{m})$ if and only if $(X, \mathsf{d}, \mathfrak{m})$ has a ∞-test plan concentrated on nonconstant rectifiable curves. In addition, (X, d) contains one nonconstant rectifiable curve if and only there exists a finite Borel measure $\mathfrak{m}$ in (X, d) satisfying $BV_*(X, \mathsf{d}, \mathfrak{m}) \neq L^1(X, \mathfrak{m})$.*

Proof. In the first statement, the "only if" part is trivial, since absence of ∞-test plans implies that all L^1 functions are BV_w, and therefore BV_*. In order to prove the converse, we notice that for a given countable dense set $D \subset X$, a curve γ is constant iff $t \mapsto \mathsf{d}(\gamma, x)$ is constant for all $x \in D$. Hence, we can find $x \in D$ and a ∞-test plan $\boldsymbol{\pi}$ such that $\mathsf{d}(\gamma, x)$ is nonconstant in a set with $\boldsymbol{\pi}$-positive measure. The composition

$$f(y) := w(\mathsf{d}(y, x)),$$

where $w : [0, \infty) \to [0, 1]$ is a continuous and nowhere differentiable function, provides a function in $L^1 \setminus BV_w = L^1 \setminus BV_*$.

For the second statement, absence of nonconstant rectifiable curves forces the absence of nontrivial ∞-test plans whatever $\mathfrak{m}$ is and, for the reasons explained above, the coincidence $L^1 = BV_*$. On the other hand, existence of a nonconstant rectifiable curve in (X, d) implies existence of a nonconstant injective curve $\gamma : [0, 1] \to X$ with constant speed. If $u \in L^1(0, 1) \setminus BV(0, 1)$, then it is easily seen that $u \circ \gamma^{-1}$ (arbitrarily defined on $X \setminus \gamma([0, 1])$) belongs to $L^1 \setminus BV_w$ provided we choose $\mathfrak{m} := \gamma_\sharp \mathscr{L}^1$, where $\mathscr{L}^1$ is the restriction of Lebesgue measure to $[0, 1]$. □

8 Spaces $W^{1,1}(X, \mathsf{d}, \mathfrak{m})$

In this section we discuss potential definitions of the space $W^{1,1}$. Here the picture is far from being complete, since at least three definitions are available and we are presently not able to prove their equivalence, unlike for BV. For simplicity, here we assume that $(X,\mathsf{d},\mathfrak{m})$ is a compact metric space and that $\mathfrak{m}$ is a probability measure. Recall that $BV_*(X,\mathsf{d},\mathfrak{m})$ denotes the BV space defined by relaxation of the slope of Lipschitz functions, while $w-BV(X,\mathsf{d},\mathfrak{m})$ is the BV space defined with the BV property along curves.

It is immediate to define $w-W^{1,1}(X,\mathsf{d},\mathfrak{m})$ as the subset of $w-BV(X,\mathsf{d},\mathfrak{m})$ consisting of functions $f\in L^1(X,\mathsf{d},\mathfrak{m})$ such that $|Df|_w\ll\mathfrak{m}$. On the other hand, also the construction leading to $BV_*(X,\mathsf{d},\mathfrak{m})$ (or to the relaxed Sobolev spaces) can be adapted to provide a different definition of $W^{1,1}$:

Definition 8.1 (1-relaxed slope) *Let $f\in L^1(X,\mathsf{d},\mathfrak{m})$. We say that a nonnegative function $g\in L^1(X,\mathsf{d},\mathfrak{m})$ is a 1-relaxed slope of f if there exist locally Lipschitz functions f_n converging to f in $L^1(X\mathfrak{m})$ such that $|\nabla f_n|\rightharpoonup h$ weakly in $L^1(X,\mathfrak{m})$, with $g\geq h$ $\mathfrak{m}$-a.e. in X.*

Then, we may define $W^{1,1}_*(X,\mathsf{d},\mathfrak{m})$ as the space of functions in $L^1(X,\mathsf{d},\mathfrak{m})$ having a 1-relaxed slope. It is not difficult to show, using Mazur's lemma, that an equivalent definition of 1-relaxed slope g involves sequences f_n such that $|\nabla f_n|\leq h_n$, with $h_n\to h$ strongly in $L^1(X,\mathfrak{m})$ and $h\leq g$. Then, this gives that $|Df|_w\leq h\mathfrak{m}$ for all $f\in W^{1,1}_*(X,\mathsf{d},\mathfrak{m})$, so that

$$W^{1,1}_*(X,\mathsf{d},\mathfrak{m})\subseteq w-W^{1,1}(X,\mathsf{d},\mathfrak{m}).$$

Finally, also a third intermediate definition of $W^{1,1}(X,\mathsf{d},\mathfrak{m})$ could be considered, in the spirit of [24, 28].

Definition 8.2 (1-upper gradient) *A Borel nonnegative function $g\in L^1(X,\mathsf{d},\mathfrak{m})$ is said to be a 1-upper gradient of $f\in L^1(X,\mathsf{d},\mathfrak{m})$ if there exists a function $\hat f$ that coincides $\mathfrak{m}$-almost everywhere with f such that*

$$|\hat f(\gamma(1))-\hat f(\gamma(0))|\leq\int_0^1 g(\gamma(s))|\dot\gamma_s|\,\mathrm{d}s\qquad\forall\gamma\in AC^\infty([0,1];X)\setminus\Gamma,$$

with $\mathrm{Mod}_1(\Gamma)=0$.

Recall that

$$\mathrm{Mod}_1(\Gamma):=\inf\left\{\int_X\rho\,\mathrm{d}\mathfrak{m}:\ \rho\geq 0,\ \int_\gamma\rho\geq 1\ \ \forall\gamma\in\Gamma\right\}.$$

Since Mod_1-negligible set of curves parametrized on $[0,1]$ are easily seen to be 1-negligible (it suffices to integrate with respect to any ∞-test plan $\boldsymbol{\pi}$ the inequality $\int_0^1\rho(\gamma_t)|\dot\gamma_t|\geq 1$) we see that the space $W^{1,1}_S(X,\mathsf{d},\mathfrak{m})$ of functions having 1-upper gradient is contained in $w-W^{1,1}(X,\mathsf{d},\mathfrak{m})$, while the arguments of [28] provide the inclusion $W^{1,1}_*(X,\mathsf{d},\mathfrak{m})\subseteq W^{1,1}_S(X,\mathsf{d},\mathfrak{m})$. Summing up, we have

$$W^{1,1}_*(X,\mathsf{d},\mathfrak{m})\subseteq W^{1,1}_S(X,\mathsf{d},\mathfrak{m})\subseteq w-W^{1,1}_*(X,\mathsf{d},\mathfrak{m})$$

and we don't know wether the first inclusion may be strict; an example showing that the second inclusion may be strict is provided by Example 7.4. A fourth space could be added to this list, considering general integrable functions f_n and replacing the slopes $|\nabla f_n|$ with upper gradients g_n in Definition 8.1. However, since 1-upper gradients are characterized as strong L^1 limits of upper gradients, this space is easily seen to coincide $W^{1,1}_S(X,\mathsf{d},\mathfrak{m})$.

Appendix: proof of the superposition principle, $p=\infty$

We will need the following result, proved for $1<p<\infty$ in [26]: it shows how to lift, somehow in an optimal way (see (8.5)), a Lipschitz curve μ_t w.r.t. W_∞ to a plan $\boldsymbol{\pi}\in\mathscr{P}(C([0,T];X))$ whose time marginals are μ_t (see also [4, Theorem 8.2.1] for the Euclidean case).

Let us recall some preliminary facts. If Y is a metric space, a function $\varphi:Y\to\mathbb{R}\cup\{+\infty\}$ is said to be *coercive* if for every $c<\infty$ the sublevel set

$$K_c=\{y\in Y\ :\ \varphi(y)\le c\}$$

is compact. In particular, coercive functions are lower semicontininuous. If Y is Polish, Prokhorov's theorem states that a family $\mathcal{A}$ of probability Borel measures in Y is relatively compact for the weak topology if and only if there exists a coercive function φ on Y such that:

$$\sup_{\mu\in\mathcal{A}}\int_Y\varphi\,\mathrm{d}\mu<\infty.$$

We shall use the "if" implication in the sequel, to build our plan $\boldsymbol{\pi}$.

We shall work in the Polish space $\mathcal{M}$ (larger than $C([0,1];X)$) of Borel maps $\gamma:[0,1]\to X$ endowed with the convergence in measure, namely the one induced by the distance

$$d_{\mathcal{M}}(\gamma,\gamma'):=\inf\left\{\varepsilon>0:\ \mathscr{L}^1(\{t\in[0,1]:\ \mathsf{d}(\gamma_t,\gamma'_t)>\varepsilon\})<\varepsilon\right\}.$$

Recall that convergence a.e. implies convergence in measure, and that sequences convergent in measure have subsequences convergent a.e. in $[0,1]$.

We state now the following simple compactness criterion in $\mathcal{M}$: the proof will be obtained by embedding isometrically X into ℓ^∞, and then applying the classical Frechét-Kolmogorov compactness criterion for real-valued maps componentwise.

Proposition 8.3 [Compactness in $\mathcal{M}$] *Let $\mathcal{F}\subseteq\mathcal{M}$ be satisfying the following properties:*

- (equicontinuity) *for every $\varepsilon>0$ there exists $\delta=\delta(\varepsilon)\in(0,1]$ such that*

$$\sup_{0<j\le\delta}\mathscr{L}^1\big(\{t\in[0,1-j]\ :\ \mathsf{d}(\gamma_{t+j},\gamma_t)>\varepsilon\}\big)\le\varepsilon\qquad\forall\gamma\in\mathcal{F};\tag{8.1}$$

- (tightness) *there exists a coercive function Φ on X such that:*

$$\sup_{\gamma\in\mathcal{F}}\int_0^1\Phi(\gamma_t)\,\mathrm{d}t<\infty.$$

Then $\mathcal{F}$ is relatively compact in $(\mathcal{M}, d_{\mathcal{M}})$.

Proof. The statement is well-known if $X = \mathbb{R}^N$: indeed, arguing componentwise one easily reduces to the case $N = 1$. If the functions are equibounded the statement corresponds to the classical Frechét-Kolmogorov relative compactness criterion in $L^p(0,1)$, $1 \le p < \infty$, while in the general case one can combine tightness with a truncation argument to obtain relative compactness w.r.t. convergence in measure.

If (X, d) is complete and separable, possibly embedding (X, d) isometrically into ℓ^∞ endowed with the canonical norm $\|\cdot\|_\infty$, we can assume that $X \subseteq \ell^\infty$ and we denote by $\pi_N : \ell^\infty \to \mathbb{R}^N$ the canonical finite-dimensional projections. By the compactness of the sublevels of Φ, we can find doubly-indexed sequences $\omega_{N,p}$ such that

$$\mathsf{d}(x,y) = \|x-y\|_\infty \le \|\pi_N(x)-\pi_N(y)\|_\infty + \omega_{N,p} \quad \forall x,\, y \in \{\Phi \le p\}, \qquad \lim_{N\to\infty} \omega_{N,p} = 0 \quad \forall p. \tag{8.2}$$

By the relative compactness criterion for $\mathbb{R}^N$-valued maps, applied with

$$\Phi_N(y) := \begin{cases} \min\{\Phi(x) \ : \ \pi_N(x) = y\} & y \in \pi_N(X), \\ +\infty & y \in \mathbb{R}^N \setminus \pi_N(X), \end{cases}$$

the families $\{\pi_N \circ \gamma : \ \gamma \in \mathcal{F}\}$ are relatively compact w.r.t. convergence in measure; by a diagonal argument, given any sequence $(\gamma^k) \subseteq \mathcal{F}$, we can find a subsequence $(\gamma^{k(n)}) \subseteq \mathcal{F}$ such that $\pi_N(\gamma^{k(n)})$ is a Cauchy sequence w.r.t. the distance $d_{\mathcal{M}}$ for all N. In order to conclude the proof, suffices to show that for all $\varepsilon > 0$ and $\delta > 0$ we can find n_0 such that

$$\mathscr{L}^1\big(\{t \in [0,1] \ : \ \|\gamma_t^{k(n)} - \gamma_t^{k(m)}\|_\infty > \varepsilon\}\big) \le \delta \qquad \forall m,\, n \ge n_0. \tag{8.3}$$

Having fixed $\varepsilon > 0$ and $\delta > 0$, choose p so large that $\sup_n \int_0^1 \Phi(\gamma_t^{k(n)})\,\mathrm{d}t < p\delta/3$ and then N so large that $\omega_{N,p} < \varepsilon/2$; then, using (8.2), we easily get that the set in the left hand side of (8.3) is contained in

$$\Big\{\|\pi_N(\gamma_t^{k(n)}) - \pi_N(\gamma_t^{k(m)})\|_\infty > \varepsilon/2\Big\} \cup \Big\{\max\{\Phi(\gamma_t^{k(n)}), \Phi(\gamma_t^{k(m)})\} > p\Big\}.$$

Markov inequality and the fact that $\pi_N(\gamma^{k(n)})$ is a Cauchy sequence, give us an integer n_0 such that

$$\mathscr{L}^1\big(\{t \in [0,1] \ : \ \|\gamma_t^{k(n)} - \gamma_t^{k(m)}\|_\infty > \varepsilon\}\big) \le \frac{\delta}{3} + \frac{\delta}{3} + \frac{\delta}{3} \qquad \forall m,\, n \ge n_0.$$

□

It will be useful in the sequel the following fact. If for every modulus of continuity $\omega : [0,1] \to [0,1]$, we call $\mathcal{P}(\omega)$ the property

$$\sup_{0<j\le\omega(\varepsilon)} \mathscr{L}^1\big(\{t \in [0,1-j] \ : \mathsf{d}(\gamma_{t+j}, \gamma_t) > \varepsilon\}\big) \le \varepsilon \qquad \forall\, \varepsilon \le 1, \tag{8.4}$$

then the set of curves in $\mathcal{M}$ satisfying $\mathcal{P}(\omega)$ is closed w.r.t. convergence in measure; this can be checked verifying the closure w.r.t. convergence almost everywhere, which is a simple matter.

Proposition 8.4 (Superposition principle) *Let* (X, d) *be a complete and separable metric space with and let* $\mu_t \in \mathrm{Lip}\big([0,T];(\mathscr{P}(X), W_\infty)\big)$. *Then there exists* $\boldsymbol{\pi} \in \mathscr{P}(C([0,T];X))$, *concentrated on* $\mathrm{Lip}([0,1];X)$, *such that* $(\mathrm{e}_t)_\sharp\boldsymbol{\pi} = \mu_t$ *for any* $t \in [0,T]$ *and*

$$\|\dot\gamma_t\|_{L^\infty(\boldsymbol{\pi})} = |\dot\mu_t| \qquad \text{for a.e. } t \in [0,T]. \tag{8.5}$$

In particular

$$\|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})} = \mathrm{Lip}(\mu_t). \tag{8.6}$$

Proof. We assume $T = 1$, since the general case follows easily by a rescaling argument. We begin with an inequality: given a plan $\boldsymbol{\pi}$ concentrated on Lipschitz curves with Lipschitz constant less then C, we consider the curve $\mu_t = (\mathrm{e}_t)_\sharp\boldsymbol{\pi}$. Then, using a time rescaled version of $\boldsymbol{\pi}$ as transport plan from μ_t to μ_s it is easy to see that

$$W_p^p(\mu_t, \mu_s) \le \|\mathsf{d}(\gamma_s, \gamma_t)\|^p_{L^p(\boldsymbol{\pi})} \le (t-s)^{p-1}\int_s^t\int |\dot\gamma_r|^p \,\mathrm{d}\boldsymbol{\pi}\,\mathrm{d}r$$

for $1 < p < \infty$ and $0 \le s \le t \le 1$ (notice that, by Fubini's theorem, the metric derivative $|\dot\gamma_t|$ exists $\boldsymbol{\pi}$-a.e. in $C([0,1];X)$ for a.e. t). This yields that the metric derivative of μ_t w.r.t. W_p can be estimated a.e. by $\|\dot\gamma_t\|_{L^p(\boldsymbol{\pi})}$. Since

$$W_\infty(\mu_t, \mu_s) = \lim_{p\to\infty} W_p(\mu_t, \mu_s) \le \int_s^t \|\dot\gamma_r\|_{L^p(\boldsymbol{\pi})}\,\mathrm{d}r \le \int_s^t \|\dot\gamma_r\|_{L^\infty(\boldsymbol{\pi})}\,\mathrm{d}r$$

we obtain the inequality $\ge$ in (8.5), as well as the global inequality $\mathrm{Lip}(\mu_t) \le \|\mathrm{Lip}(\gamma)\|_{L^\infty(\boldsymbol{\pi})}$. Given a curve (μ_t) with Lipschitz constant C, we want to build $\boldsymbol{\pi}$ with the correct marginals that satisfies the opposite inequalities; it is very natural to approximate such a plan. The remaining part of the proof will be split in steps.

Step 1. (Approximating plans Σ_N) We can argue as in [26], with minor changes, to build approximating plans $\boldsymbol{\eta}_N$ in this way:

- we consider Σ_N^i, for $i = 0, \ldots, 2^N - 1$, optimal plans in the ∞-Wasserstein problem between $\mu_{\frac{i}{2^N}}$ and $\mu_{\frac{i+1}{2^N}}$;
- we build as in [26] a probability measure Σ_N on X^{2^N+1} such that

  $$(\pi_i)_\sharp\Sigma_N = \mu_{\frac{i}{2^N}} \quad (0 \le i \le 2^N), \qquad\qquad (\pi_i, \pi_{i+1})_\sharp\Sigma_N = \Sigma_N^i \quad (0 \le i \le 2^N - 1),$$

  where $\pi_i : X^{2^N} \to X$ denotes the canonical projection on the i-th component;
- we consider the map $\rho : X^{2^N+1} \to \mathcal{M}$, taking more precisely values in the class of piecewise constant maps, such that

  $$\rho(x_0, x_1, \ldots, x_{2^N})(t) = x_{\lfloor 2^N t\rfloor}.$$

  Eventually we define $\boldsymbol{\eta}_N = (\rho)_\sharp\Sigma_N$.

Step 2. (Tightness) Now we want to show that the family $\boldsymbol{\eta}_N$ is tight, so that we can extract a converging subsequence: applying Prokhorov compactness theorem it is sufficient to show the existence of a coercive function Ψ on $\mathcal{M}$ such that

$$\sup_{N\in\mathbb{N}} \int_{\mathcal{M}} \Psi(\gamma) \mathrm{d}\boldsymbol{\eta}_N(\gamma) < \infty. \tag{8.7}$$

We claim that the function:

$$\Psi(\gamma) = I_{\mathcal{P}(\omega)}(\gamma) + \int_0^1 \Phi(\gamma_t) \mathrm{d}t$$

satisfies this property, choosing appropriately a coercive function $\Phi : X \to [0,\infty]$ and a modulus of continuity ω; here I sands for the indicator function, namely

$$I_{\mathcal{P}(\omega)}(\gamma) = \begin{cases} 0 & \text{if } \gamma \text{ satisfies } \mathcal{P}(\omega) \\ +\infty & \text{otherwise.} \end{cases}$$

Every function of this kind with Φ coercive is again coercive thanks to Proposition 8.3 and to the fact that $\mathcal{P}(\omega)$ is a closed condition under convergence in measure.
If we want also (8.7) to be satisfied we have to choose carefully Φ and ω. First we note that the family of measure $\{\mu_t\}$ is clearly tight and thus, another application of Prokhorov gives us the existence of a coercive Φ satisfying $\int_X \Phi \, \mathrm{d}\mu_t \le 1$ for all $t \in [0,1]$. In this way we obtain that the second term of Ψ is equibounded; indeed

$$\begin{aligned} \int_{\mathcal{M}} \int_0^1 \Phi(\gamma_t) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{\eta}_N(\gamma) &= \int_{X^{2^N+1}} \int_0^1 \Phi(\rho(x)(t)) \, \mathrm{d}t \, \mathrm{d}\Sigma_N(x) = \int_{X^{2^N+1}} \int_0^1 \Phi(x_{\lfloor t2^N \rfloor}) \, \mathrm{d}t \, \mathrm{d}\Sigma_N(x) \\ &= \int_0^1 \int_{X^{2^N+1}} \Phi(x_{\lfloor t2^N \rfloor}) \, \mathrm{d}\Sigma_N(x) \, \mathrm{d}t = \int_0^1 \int_X \Phi(x) \, \mathrm{d}\mu_{g_N(t)} \, \mathrm{d}t \le 1, \end{aligned}$$

where we used that $(\pi_k)_\sharp \Sigma_N = \mu_{\frac{k}{2^N}}$, then we introduced the functions $g_N(t) := \lfloor 2^N t \rfloor / 2^N$, and finally we used the fact that $\int_X \Phi \, \mathrm{d}\mu_t \le 1$ for all $t \in [0,1]$.

Now let us define the modulus of continuity ω: for all $\varepsilon > 0$ we want to find $\delta > 0$ such that

$$\begin{aligned} &\mathscr{L}^1(A_{\delta,\varepsilon}(\gamma)) \le \varepsilon \qquad \text{for } \boldsymbol{\eta}_N\text{-a.e. } \gamma \in \mathcal{M} \text{ and all } N, \\ &\text{where } A_{\delta,\varepsilon}(\gamma) := \{ t \in [0, 1-\delta] \ : \ \mathsf{d}(\gamma_{t+\delta}, \gamma_t) > \varepsilon \}. \end{aligned}$$

We know that $\boldsymbol{\eta}_N$ is concentrated on equivalence classes of "step"' curves, i.e. curves which remains at the same point in every interval of the form $[\frac{k}{2^N}, \frac{k+1}{2^N})$, with $0 \le k < 2^N$, described by the map ρ; therefore we can estimate $\mathscr{L}^1(A_{\delta,\varepsilon}(\gamma))$ working instead on X^{2^N+1}, with curves $\gamma = \rho(x)$ and with the measure Σ_N, recalling that

$$\|\mathsf{d}(x_k, x_{k+1})\|_{L^\infty(\Sigma_N)} = W_\infty\big(\mu_{\frac{k}{2^N}}, \mu_{\frac{k+1}{2^N}}\big) \le C \cdot 2^{-N}, \qquad 0 \le k \le 2^N - 1. \tag{8.8}$$

We distinguish two cases:

- $\delta < 2^{-N}$. Then it is clear that in $[t, t+\delta]$ there is at most one jump and so $\mathsf{d}(\gamma_t, \gamma_{t+\delta}) \leq C \cdot 2^{-N}$ for $\boldsymbol{\eta}_N$-a.e. γ; in particular if $C \cdot 2^{-N} \leq \varepsilon$ we get that $\mathscr{L}^1(A_{\delta,\varepsilon}(\gamma)) = 0$. Otherwise, in the case $C \cdot 2^{-N} > \varepsilon$, knowing nothing more on the size of the jump, we can say only that $A_{\delta,\varepsilon}(\gamma)$ is contained in the δ-neighourhooud of the set of jumps, which has Lebesgue measure less than $2^N\delta$. So at the end we get that in the case $\delta < 2^N$ we can estimate $\mathscr{L}^1(A_{\delta,\varepsilon}(\gamma))$ on the measure theoretic support of $\boldsymbol{\eta}_N$ as follows:
$$\mathscr{L}^1(A_{\delta,\varepsilon}(\gamma)) \leq 2^N\delta \leq C\frac{\delta}{\varepsilon}.$$
- $k2^{-N} \leq \delta < (k+1)2^{-N}$, for some $k > 0$. This time we know that there exist at most $k+1$ jumps in $[t, t+\delta]$; thus we know that $\mathsf{d}(\gamma_t, \gamma_{t+\delta}) \leq C(k+1)2^{-N}$. Again we get $\mathscr{L}^1(A_{\delta,\epsilon}(\gamma)) = 0$ if $(k+1)2^{-N}C < \varepsilon$; this is always true if $2C\delta < \varepsilon$, in fact, in this case
$$C\frac{k+1}{2^N} = C\frac{k+1}{k}\frac{k}{2^N} \leq C \cdot 2 \cdot \delta < \varepsilon.$$

Summing up, in order to have $\mathscr{L}^1(A_{\delta,\varepsilon}(\gamma)) \leq \varepsilon$ for $\boldsymbol{\eta}_N$-a.e. γ for every N it is sufficient that both the conditions $C\delta/\varepsilon \leq \varepsilon$, $2C\delta < \varepsilon$ hold, and so we can choose the modulus of continuity $\omega(\varepsilon) = \varepsilon^2/(2C)$, $\varepsilon \in [0,1]$.

Step 3. (Construction of $\boldsymbol{\pi}$) We can fix now a limit point $\boldsymbol{\eta}$ of $\boldsymbol{\eta}_N$ and we assume, just for notational simplicity, that the whole family $\boldsymbol{\eta}_N$ weakly converges to $\boldsymbol{\eta}$. Now we show that $\operatorname{supp}\boldsymbol{\eta}$ is contained in the set of equivalence classes of C-Lipschitz curves. We have already seen that, by construction, the support of $\boldsymbol{\eta}_N$ is contained in the closed set:
$$A_N = \left\{ \gamma \in \mathcal{M} \ : \ \sup_{h \geq N2^{-N}} \operatorname*{esssup}_{t \in [0,1-h]} \left\{ \frac{d(\gamma_t, \gamma_{t+h})}{h} \right\} \leq \frac{N+1}{N} C \right\}$$
where the A_N are obviously decreasing. So, we obtain that $\operatorname{supp}\boldsymbol{\eta} \subseteq \bigcap_N A_N$, and it is clear that curves in $\bigcap_N A_N$ have a C-Lipschitz curve in their equivalence class.

Considering the canonical continuous immersion
$$i : C([0,1]; X) \to \mathcal{M}$$
we can define the measure $\boldsymbol{\pi}$ on $C([0,1]; X)$ defined as $\boldsymbol{\pi}(A) = \boldsymbol{\eta}(i(A))$ for every Borel subset A of L. Notice that i is a well-defined Borel probability measure because the continuous image of a Borel set is Suslin, hence $\boldsymbol{\eta}$-measurable. In addition, $\boldsymbol{\pi}$ is concentrated on C-Lipschitz curves.

In order to compute the marginals of $\boldsymbol{\pi}$, for every continuous and bounded function ϕ on X, $h \in [0,1]$ and $t \in [0, 1-h]$ we can consider the bounded continuous map on $\mathcal{M}$ defined by
$$f^{\phi}_{t,h}(\gamma) = \int_t^{t+h} \phi(\gamma_s)\, \mathrm{d}s.$$

Since $\boldsymbol{\eta}_N$ converge weakly to $\boldsymbol{\eta}$, with more or less the same calculation we made before we get:

$$\begin{aligned}\int_{\mathcal{M}} f^{\phi}_{t,h}\,\mathrm{d}\boldsymbol{\eta} &= \lim_{N\to\infty}\int_{\mathcal{M}} f^{\phi}_{t,h}\,\mathrm{d}\boldsymbol{\eta}_N = \lim_{N\to\infty}\int_{X^{2^N+1}} f^{\phi}_{t,h}(\rho(x))\,\mathrm{d}\Sigma_N(x)\\ &= \lim_{N\to\infty}\int_{X^{2^N+1}}\int_t^{t+h}\phi(x_{\lfloor s2^N\rfloor})\,\mathrm{d}s\,\mathrm{d}\Sigma_N(x) = \lim_{N\to\infty}\int_t^{t+h}\int_{X^{2^N+1}}\phi(x_{\lfloor s2^N\rfloor})\,\mathrm{d}\Sigma_N(x)\,\mathrm{d}s\\ &= \lim_{N\to\infty}\int_t^{t+h}\int_X \phi\,\mathrm{d}\mu_{g_N(s)}\,\mathrm{d}s,\end{aligned}$$

with $g_N(t) = \lfloor 2^N t\rfloor/2^N$. It is clear that $g_N(t)\to t$ uniformly in t and, by the continuity of the function $s\to\int\phi\,\mathrm{d}\mu_s$ we get finally that

$$\int_{C([0,1];X)} f^{\phi}_{t,h}\,\mathrm{d}\boldsymbol{\pi} = \int_{\mathcal{M}} f^{\phi}_{t,h}\,\mathrm{d}\boldsymbol{\eta} = \int_t^{t+h}\int_X \phi(x)\,\mathrm{d}\mu_s\,\mathrm{d}s. \tag{8.9}$$

Dividing both sides by h and passing to the limit as $h\downarrow 0$ we can use the fact that $f^{\phi}_{t,h}(\gamma)/h\to\phi(\gamma_t)$ as $h\downarrow 0$ in $C([0,1];X)$ and the arbitrariness of ϕ to obtain that $(e_t)_\sharp\boldsymbol{\pi} = \mu_t$ for all $t\in[0,1]$.

Step 4. (Verification of (8.5)) We need only to show the inequality $\|\dot\gamma_t\|_{L^\infty(\boldsymbol{\pi})}\le|\dot\mu_t|$ for a.e. t. It is clear that the inequality holds by construction if $|\dot\mu_t| = C$ for a.e. t, since in this case we proved that $\boldsymbol{\pi}$ is supported on C-Lipschitz curves. If we drop this assumption, assuming only that $\{|\dot\mu_t|>0\}$ has positive measure in any interval, we can define a strictly increasing map L on $[0,1]$ as follows:

$$L(t) := \int_0^t |\dot\mu_r|\,\mathrm{d}r \qquad t\in[0,1]. \tag{8.10}$$

Set $L = L(1)$. It is immediate to check that $\tilde\mu_s := \mu_{L^{-1}(s)}$, $s\in[0,L]$, is 1-Lipschitz. If we represent $\tilde\mu_s$ as $(e_t)_\sharp\tilde{\boldsymbol{\pi}}$, with $\tilde{\boldsymbol{\pi}}$ concentrated on 1-Lipschitz curves on $[0,L]$, we see immediately that the plan

$$\boldsymbol{\pi} := \Psi_\sharp\tilde{\boldsymbol{\pi}} \qquad \text{with} \qquad \Psi(\gamma)_t := \gamma_{L(t)}$$

represents μ_t, and that $|\dot\gamma_t|\le L'(t) = |\dot\mu_t|$ $\boldsymbol{\pi}$-a.e. in $C([0,1];X)$ for a.e. $t\in[0,1]$.

Finally, we only sketch the argument which allows to remove the assumption that $\{|\dot\mu_t|>0\}$ has positive measure in any interval. One can either use the ε-parameterizations of [4, Lemma 1.1.4] (i.e. adding ε into the integral in (8.10)) and pass to the limit as $\varepsilon\downarrow 0$, or argue as follows: collapsing all open intervals where μ_t is constant, one obtains a new Lipschitz curve $\tilde\mu_t$ defined on an interval $[0,L]$ with $L<1$ which satisfies the nondegeneracy condition. Representing $\tilde\mu_t$ as $(e_t)_\sharp\tilde{\boldsymbol{\pi}}$, with $\tilde{\boldsymbol{\pi}}$ probability measure on $C([0,L];X)$, the intervals can be restored to produce $\boldsymbol{\pi}$, concentrated on curves defined in $[0,1]$ and constant on these intervals. □

Remark 8.5 Let us note that this proposition, stated only for probability measures, holds also for Lipschitz curves $\{\mu_t\} \subseteq (\mathcal{M}_+(X), W_\infty)$. Indeed, first of all we note that $W_\infty(\mu, \nu) = W_\infty(C\mu, C\nu)$ for all $C \geq 0$; then, letting $m = \mu_0(X)$, we can consider the curve of probability measures $\{\frac{\mu_t}{m}\}$ that is still Lipschitz (with the same Lipschitz constant) and so we can apply the proposition, to get a plan $\boldsymbol{\pi}$. Now it is easy to see that $m\boldsymbol{\pi}$ solves the problem for $\{\mu_t\}$.

References

[1] L. Ambrosio, *Fine properties of sets of finite perimeter in Ahlfors regular metric measure spaces*, Advances in Mathematics, **159** (2001), 51–67.

[2] ——, *Fine properties of sets of finite perimeter in doubling metric measure spaces*, Set Valued Analysis, **10** (2002), 111–128.

[3] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, 2000.

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

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

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

[7] ——, *Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces*, Preprint (2011), to appear on Rev. Mat. Iberoamericana.

[8] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[9] G. Bellettini, V. Caselles, M. Novaga, *The total variation flow in* $\mathbb{R}^N$. J. Differential Equations, **184** (2002), 475–525.

[10] T. Champion, L. De Pascale, P. Juutinen, *The* ∞*-Wasserstein distance: local solutions and existence of optimal transport maps* , SIAM Journal of Mathematical Analysis, **40** (2008), 1–20.

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

[12] J. Cheeger, B. Kleiner, *Metric differentiation, monotonicity and maps to L^1.* Invent. Math., **182** (2010), 335–370.

[13] J. Cheeger, B. Kleiner, *Differentiating maps into L^1, and the geometry of BV functions.* Ann. of Math., **171** (2010), 1347–1385.

[14] J. Cheeger, B. Kleiner, A. Naor, *Compression bounds for Lipschitz maps from the Heisenberg group to L^1.* Acta Math., **207** (2011), 291–379.

[15] G. Dal Maso, *An introduction to Γ convergence*, Progress in Nonlinear Differential Equations and Their Applications, Birkhäuser Verlag, Boston, second ed., 1993.

[16] B. Fuglede, *Extremal length and functional completion*, Acta Math., **98** (1957), 171–219.

[17] N. Gigli, K. Kuwada, and S. Ohta, *Heat flow on Alexandrov spaces*, Comm. on Pure and Appl. Math., **66** (2013), 307–331.

[18] N. Gozlan, C. Roberto, and P. Samson, *Hamilton-Jacobi equations on metric spaces and transport entropy inequalities*, Preprint (2012), to appear on Rev. Mat. Iberoamericana.

[19] P. Hajłasz and P. Koskela, *Sobolev met Poincaré*, Mem. Amer. Math. Soc., **145** (2000).

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

[21] J. Heinonen and P. Koskela, *Quasiconformal maps in metric spaces with controlled geometry*, Acta Math., **181** (1998), 1–61.

[22] ——, *A note on Lipschitz functions, upper gradients, and the Poincaré inequality*, New Zealand J. Math., **28** (1999), 37–42.

[23] P. Koskela, *Removable sets for Sobolev spaces*, Ark. Mat., **37** (1999), 291–304.

[24] P. Koskela and P. MacManus, *Quasiconformal mappings and Sobolev spaces*, Studia Math., **131** (1998), 1–17.

[25] K. Kuwada, *Duality on gradient estimates and Wasserstein controls*, Journal of Functional Analysis, **258** (2010), 3758–3774.

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

[27] M. Miranda, Jr., *Functions of bounded variation on "good" metric spaces*, J. Math. Pures Appl. (9), **82** (2003), 975–1004.

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

[29] K.-T. Sturm, *Generalized Orlicz spaces and Wasserstein distances for convex-concave scale functions*, Bulletin des sciences mathématiques, **135** (2011), 795–802.

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