Sobolev and Bounded Variation Functions on Metric Measure Spaces

Luigi Ambrosio, Roberta Ghezzi

Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
luigi.ambrosio@sns.it.
Institut de Mathématiques de Bourgogne, UBFC, 9 Avenue Alain Savary, 21078 Dijon, France roberta.ghezzi@u-bourgogne.fr.

Contents

CHAPTER 1

Introduction

These notes reflect, with minor modification and updates, the lectures given by the first author in occasion of the Trimester in *Geometry, Analysis and Dynamics on Sub-Riemannian Manifolds* held at Institut Henri Poincaré in Paris, September 2014. Later on this series of lectures has been in part repeated by the first author within the Research Term in *Analysis and Geometry in Metric Spaces* at Instituto de Ciencias Matemáticas in Madrid, May 2015. The style of the presentation is similar to the one of the lectures, in that we aim at the illustration of the key ideas and of the fundamental technical concepts, not looking for the most general statements, whose proofs are occasionally cumbersome. The references on the quite broad topic of Analysis in Metric Spaces do not pretend to be exhaustive, even looking only at the Sobolev theory; nevertheless we think that these lectures provide a good entry point to the literature and to the state of the art on this topic.

1. History

Consider the classical Euclidean space $(\mathbb{R}^n, |\cdot|, \mathscr{L}^n)$, where $\mathscr{L}^n$ denotes the n-dimensional Lebesgue measure. In this context one can give three definitions of Sobolev spaces, that we now know to be equivalent.

Let $1 < p < \infty$ and denote by $\mathcal{C}_c^\infty(\mathbb{R}^n)$ the vector space of $\mathcal{C}^\infty$ functions on $\mathbb{R}^n$ with compact support.

Definition 1.1 (W-definition via integration by parts).

$$W^{1,p}(\mathbb{R}^n) := \Big\{ f \in L^p(\mathbb{R}^n) \mid \exists\, g \in L^p(\mathbb{R}^n, \mathbb{R}^n) \text{ such that } \forall\, \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n), \int_{\mathbb{R}^n} f \frac{\partial \varphi}{\partial x_i} dx = -\int_{\mathbb{R}^n} \varphi g_i dx,\ i = 1, \ldots, n \Big\}.$$

The vector g in the definition above, called weak gradient, is unique and denoted by ∇f. The $W^{1,p}$ norm is then defined by

$$\|f\|_{W^{1,p}} := \big(\|f\|_p^p + \|\nabla f\|_p^p \big)^{1/p}.$$

Definition 1.2 (Sobolev definition or H-definition).

$$H^{1,p}(\mathbb{R}^n) := \overline{\mathcal{C}_c^\infty(\mathbb{R}^n)}^{W^{1,p}(\mathbb{R}^n)},$$

i.e., the Sobolev space $H^{1,p}(\mathbb{R}^n)$ is the closure of $\mathcal{C}_c^\infty(\mathbb{R}^n)$ with respect to the $W^{1,p}(\mathbb{R}^n)$-norm.

Let us recall how the equivalence between Definitions 1.1 and 1.2 can be proved.

Theorem 1.1. *$H^{1,p}(\mathbb{R}^n) = W^{1,p}(\mathbb{R}^n)$.*

Sketch of the proof. The inclusion $H^{1,p}(\mathbb{R}^n) \subset W^{1,p}(\mathbb{R}^n)$ is rather obvious, once the notion of convergence in $W^{1,p}(\mathbb{R}^n)$ is well understood. Indeed, let $f \in H^{1,p}(\mathbb{R}^n)$ and let $(f_k) \subset \mathcal{C}_c^\infty(\mathbb{R}^n)$ be an approximating sequence of f, i.e.,

$$\lim_{k\to\infty} \|f_k - f\|_{W^{1,p}(\mathbb{R}^n)} = 0. \tag{1}$$

Fix any $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$. Since f_k is smooth and φ has compact support, using the integration by parts formula we obtain

$$\int_{\mathbb{R}^n} f_k \frac{\partial \varphi}{\partial x_i} dx = -\int_{\mathbb{R}^n} \varphi \frac{\partial f_k}{\partial x_i} dx. \tag{2}$$

Thanks to (1), the sequence (f_k) converges to f strongly in $L^p(\mathbb{R}^n)$ and there exists $g \in L^p(\mathbb{R}^n, \mathbb{R}^n)$ such that $\frac{\partial f_k}{\partial x_i}$ converges to g_i strongly in $L^p(\mathbb{R}^n)$. Thus formula (2) goes to the limit and we get that g_i is indeed the weak derivative of f with respect to x_i in the sense of distributions. (In other words, integration by parts formula is preserved when approximating by smooth functions).

The opposite inclusion is due to Meyers and Serrin [**48**]: even in general domains Ω, their result is that $C^\infty(\Omega) \cap W^{1,p}(\Omega)$ is dense in $W^{1,p}(\Omega))$ and it is based on an approximation technique using mollifiers. Pick $f \in W^{1,p}(\mathbb{R}^n)$ and a sequence $\{\rho_\epsilon\}_{\epsilon>0}$ of mollifiers[1]. Then $f * \rho_\epsilon \in \mathcal{C}^\infty(\mathbb{R}^n) \cap W^{1,p}(\mathbb{R}^n)$. Multiplying $f * \rho_\epsilon$ by a cut-off function and using a diagonal argument, one can build up a sequence of smooth functions with compact support converging to f strongly in $W^{1,p}(\mathbb{R}^n)$. □

Remark 1.1. In connection with sub-Riemannian geometry, a Meyers–Serrin's type result was proved by Franchi, Serapioni and Serra Cassano in [**29**] for metric measure spaces associated with systems of vector fields in $\mathbb{R}^n$ satisfying mild hypotheses. For the definition of Sobolev space in this context, see Example 3.4 and references therein.

Definition 1.3 (Beppo Levi definition). Given a point $x \in \mathbb{R}^n$ and an index $i = 1, \ldots, n$ we denote by $\hat{x}_i$ the point $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in \mathbb{R}^{n-1}$. Then, we define

$$\begin{aligned} BL^{1,p}(\mathbb{R}^n) \;:=\; \Big\{ & f : \mathbb{R}^n \to \mathbb{R} \mid \ \forall\, i = 1, \ldots, n \text{ and at } \mathscr{L}^{n-1}\text{-a.e. } \hat{x}_i \in \mathbb{R}^{n-1} \text{ the map} \\ & t \mapsto f(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n) \text{ is absolutely continuous and} \\ & \int_{\mathbb{R}^{n-1}} \left(\int_{\mathbb{R}} |f|^p + \left| \frac{\partial f}{\partial x_i} \right|^p dt \right) d\hat{x}_i < \infty \Big\}. \end{aligned}$$

In the case $n = p = 2$, Definition 1.3 appeared much earlier than Definitions 1.1 and 1.2 in the work [**45**] by Beppo Levi, where the author constructs the space above in order to solve a specific Dirichlet problem in two-dimensional domains. Definition 1.3 is based on absolute continuity along almost every line parallel to every coordinate axes; at that time, the theory of absolutely continuous functions on the line was well understood.

The goal of these lectures is to define W, H, and BL-spaces in the more general context of metric measure spaces and to show their equivalence. As we already wrote, this topic is also extensively covered in many research papers and monographs (see for instance [**36, 37, 38**] and the recent book [**40**]), covering in details also many more aspects of the theory

[1]Given an even, smooth, nonnegative convolution kernel ρ in $\mathbb{R}^n$ with compact support we denote by $\rho_\varepsilon(x) = \varepsilon^{-n}\rho(x/\varepsilon)$ the sequence of rescaled kernels.

(potential estimates, regularity of solutions, trace theorems, etc.). We want to convey a new point of view in this subject and show that the equivalence between these points of view and of the corresponding notions of gradient is a general phenomenon, true under almost no assumption on the metric measure structure. In addition, some concepts presented in these notes (even though with a few details, since we mostly focus on the Sobolev theory) are new even from the point of view of the classical theory, as the notion of "measure upper gradient" developed in connection with the theory of functions of bounded variation.

Let us remark that, already in the Euclidean case, the three definitions are conceptually quite different. Indeed the three constructions rely on different objects: Definition 1.1 is based on coordinate vector fields, Definition 1.2 exploits approximations by smooth functions, whereas Definition 1.3 takes account of the behavior of a function along special curves. Moreover, the BL-space is characterized by a pointwise definition: there is no evidence there of Lebesgue equivalence classes of functions, as in the definition of the W-space. This is essentially due to the fact that at the time of Beppo Levi's work, the mathematical tools needed to deal with this kind of space were not completely developed yet. As a drawback, it was not clear whether the BL-space depended upon the choice of coordinate systems. It was much later that this criticism was overcome in Fuglede's paper [**31**], with a frame invariant definition that we will illustrate.

2. Motivations

Most of the material in these lecture notes has been developed in a series of papers by Ambrosio, Gigli and Savaré [**11, 12, 14, 13, 15**], as well as Ambrosio, Colombo, Di Marino [**5**].

In this series of papers, one of the main motivations for revisiting the theory of Sobolev spaces in metric measure spaces is that this theory provides basic mathematical tools for the theory of synthetic Ricci lower bounds in metric measure space. Let us describe roughly the two fundamental approaches in this context. One side of the theory was developed by Bakry and Émery who introduced Γ-calculus and used the language of Dirichlet forms to give a meaning to the inequality

$$\Delta\left(\frac{1}{2}|\nabla f|^2\right) - \langle \nabla f, \Delta \nabla f\rangle \geq \frac{1}{N}(\Delta f)^2 + K|\nabla f|^2,$$

which is known as $CD(K,N)$. This inequality is inspired by the classical Bochner's identity on Riemannian manifolds

$$\Delta\left(\frac{1}{2}|\nabla f|^2\right) - \langle \nabla f, \Delta \nabla f\rangle = \|\mathrm{Hess} f\|^2 + \mathrm{Ric}(\nabla f, \nabla f).$$

We will refer to this theory as "Eulerian theory", in the sense that it involves gradient, Laplacian etc., and all these concepts make sense also in the abstract setup provided by Dirichlet forms and the associated heat semigroup (whose fundamental generator is precisely the Laplacian). One of the main advantages of Bakry–Émery approach is that it is well-suited to get useful functional inequalities even in sharp form (e.g., Li–Yau inequality).

The other side of the theory is concerned with optimal transport and was developed independently by Lott–Villani and Sturm. We will refer to their approach as "Lagrangian theory", as it involves the study of geodesics and paths in suitable metric spaces. The underlying idea is to study the K-convexity properties of two functionals

$$\int \rho \log \rho dm \qquad \text{without upper bounds on space dimension,}$$
$$-N \int \rho^{1-\frac{1}{N}} dm \qquad \text{when the dimension is } \leq N < \infty.$$

In the paper [**15**] it is established a basic equivalence between the Lott–Villani–Sturm theory for (X, d, m) asymptotically Hilbertian (in a sense that we will specify below, see Definition 2.1) and the Bakry–Émery theory in the case $N = \infty$ and $k \in \mathbb{R}$. This result has been subsequently extended to the case of upper bounds on dimensions in [**27**], see also the paper in preparation [**17**].

Remark 2.1. In connection with sub-Riemannian geometry, let us mention the works [**1**] and [**22**] related to the Eulerian theory. Concerning the Lagrangian aspect, we refer the reader to [**41**] for a negative result and [**51**] for a positive result on the measure contraction property.

In order to understand why equivalence between Definitions 1.1, 1.2, 1.3 in the context of metric measure spaces can lead to nontrivial information, let us consider an extreme situation.

Let (X, d) be a metric space containing no non constant rectifiable curve, for instance, one can take snowflake-type spaces, such as $\big(\mathbb{R}, \sqrt{|\cdot|}\big)$. Then $BL^{1,p}(X, d, m)$ is essentially $L^p(X, m)$ (although in that definition equivalence classes are not considered), since Definition 1.3 is concerned with oscillations of functions along curves. Hence, the equality $BL^{1,p} = H^{1,p}$ implies that any $f \in L^p(X, m)$ can be approximated by $f_n \in \mathrm{Lip}(X)$ (in metric spaces Lipschitz functions play the role of smooth functions) in the sense that

$$\lim_{n\to\infty} \int (|f_n - f|^p + |\nabla f_n|^p)\, dm = 0,$$

where the slope $|\nabla g|$ of a Lipschitz map g is defined by

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

(see also Definiton 1.1 below). In other words, triviality of (rectifiable) curves provides a global approximation of L^p functions by Lipschitz functions with small gradient, in the L^p sense. Note that even in the case of $\big(\mathbb{R}, \sqrt{|\cdot|}\big)$ the implication is non-trivial; the proof comes in an indirect way, via the equivalence of Sobolev spaces (i.e., the approximating sequence is not built explicitely).

3. Examples of metric measure spaces

Example 3.1 (Euclidean space). $(\mathbb{R}^n, |\cdot|, \mathscr{L}^n)$

Example 3.2 (Weighted Euclidean spaces). $(\mathbb{R}^n, |\cdot|, m)$, where $m = w\mathscr{L}^n$ and $w(x) = e^{-V(x)}$, for some potential $V : \mathbb{R}^n \to \mathbb{R}$. The case $V(x) = \frac{|x|^2}{2}$ is also known (up to a normalization of m) as *Gaussian case*. Note that when w is sufficiently regular, say of class $\mathcal{C}^1$, the integration by parts formula for $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ holds

$$\int_{\mathbb{R}^n} \frac{\partial \varphi}{\partial x_i}\, dm = \int_{\mathbb{R}^n} \varphi \frac{\partial V}{\partial x_i}\, dm.$$

As a consequence, when w is sufficiently smooth, one easily defines the derivative of a function f along x_i in the sense of distributions as a function g_i satisfying

$$\int_{\mathbb{R}^n} g_i \varphi\, dm = -\int_{\mathbb{R}^n} f \frac{\partial \varphi}{\partial x_i}\, dm + \int_{\mathbb{R}^n} f\varphi \frac{\partial V}{\partial x_i}\, dm,\ \forall\, \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n).$$

Then, the question is how to deal with non-smooth weights w. For instance, weights of this kind occur in the regularity theory for degenerate elliptic PDEs. More precisely, is it possible to get an integration by parts formula in this more general context? We shall see that the answer is yes, provided that we give up the idea of using (only) constant vector fields.

In this connection, concerning theory of elliptic PDEs, let us mention that there is another notion of weighted Sobolev spaces which reads as follows

$$W_w^{1,p}(\mathbb{R}^n) = \left\{ f \in W_{loc}^{1,1}(\mathbb{R}^n) \mid \left(\int |f|^p + |\nabla f|^p \right) w dx < \infty \right\}.$$

One can say that this notion is in some sense an extrinsic one, because one appeals to the standard Euclidean structure to define the weak derivative and then exploits the weight w only for the the integrability of the function and its gradient. In the recent work [**18**] a precise comparison is made between this point of view and the totally intrinsic point of view of Sobolev spaces in metric measure spaces.

Example 3.3 (Weighted Riemannian manifolds). (M, g, m), where (M, g) is a Riemannian manifold, $m = w\mathrm{vol}$, where vol denotes the Riemannian volume on M and $w(x) = e^{-V(x)}$, for some potential V. In this context, the gradient of a $\mathcal{C}^1$ function f is defined through the Riemannian structure as the only vector ∇f such that

$$df(v) = g(\nabla f, v) = v(f),$$

where $v(f)$ denotes the action of a vector v as a derivation on the germ of f at a point. Moreover, one can define the divergence of a vector field b through m as the only function, denoted by $\mathrm{div}_m b$, such that

$$\int_M \varphi \mathrm{div}_m b\, dm = -\int_M g(b, \nabla\varphi)\, dm,\ \forall\, \varphi \in \mathcal{C}_c^\infty(M). \tag{3}$$

In coordinates, it is easy to check that

$$\mathrm{div}_m b = \mathrm{div}\, b - g(b, \nabla V),$$

where $\mathrm{div}\, b$ is the standard divergence associated to the volume form. The Riemannian metric allows to define an integration by parts formula through (3). Moreover, together with the given measure m, it provides a notion of Laplacian of a $\mathcal{C}^2$ function f as

$$\Delta f := \mathrm{div}_m(\nabla f).$$

The operator $f \mapsto \Delta f$ encodes an interaction between the Riemannian tensor, which appears in the gradient, and the measure, which appears in the divergence. Even in the general context of metric measure spaces this dependence should be emphasized, writing $\Delta = \Delta_{d,m}$.

EXAMPLE 3.4 (Carnot–Carathéodory spaces). Consider a family of vector fields $\chi = \{X_1, \ldots, X_m\} \subset L^\infty(\mathbb{R}^n, \mathbb{R}^n)$, $i = 1, \ldots, m$, such that $\mathrm{div} X_i \in L^\infty(\mathbb{R}^n)$. Requiring divergences to be in L^∞ is sufficient to get a notion of weak derivative and thus to define

$$W^{1,p}_\chi(\mathbb{R}^n) = \{f \in L^p(\mathbb{R}^n) \mid X_i f \in L^p(\mathbb{R}^n), i = 1, \ldots m\}.$$

(here we consider the Lebesgue measure on $\mathbb{R}^n$), where $X_i f$ denotes the derivative of f along X_i in the sense of distributions. We associate to the family of vector fields the Carnot–Carathéodory distance, denoted by d_{cc}, defined by the following control problem

$$d_{cc}(x, y) = \inf \left\{ \int_0^1 \sqrt{\sum_{i=1}^m u_i^2(s)} ds \mid \dot{\gamma} = \sum_{i=1}^m u_i X_i(\gamma),\ \gamma(0) = x, \gamma(1) = y \right\}.$$

We will define the set $W^{1,p}(\mathbb{R}^n, d_{cc}, \mathscr{L}^n)$, see Definition 2.1, and we will prove the following fact: if $X_1, \ldots, X_m$ are smooth, then $W^{1,p}_\chi(\mathbb{R}^n) = W^{1,p}(\mathbb{R}^n, d_{cc}, \mathscr{L}^n)$. Note that to prove the equivalence we need neither the Lie bracket generating condition nor the connectivity of the metric space $(\mathbb{R}^n, d_{cc})$. In this connection, we mention the works [**29**], [**32**] and [**9**].

EXAMPLE 3.5 (Infinite dimensional Gaussian space). Denote by $\gamma_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ the Gaussian map with null mean and variance 1 on $\mathbb{R}$ and consider the measure $\gamma_1 \mathscr{L}^1$. Take the product measure $m = \prod_{n\in\mathbb{N}} \gamma_1 \mathscr{L}^1$ on a countable product $\mathbb{R}^\infty = \prod_{n\in\mathbb{N}} \mathbb{R}$. Then $m \in \mathcal{P}(\mathbb{R}^\infty)$, i.e., it is a probability measure on $\mathbb{R}^\infty$. An alternative presentation, which provides an artificial Banach structure, is the following. Fix $(c_n)_{n\in\mathbb{N}} \in \ell_1$ with $c_n \geq 0$ and consider the Hilbert weighted space H_c given by

$$H_c = \left\{ (x_n)_{n\in\mathbb{N}} \in \mathbb{R}^\infty \mid \sum c_n x_n^2 < \infty \right\},$$

endowed with the scalar product $\langle (x), (y) \rangle_{H_c} = \sum c_n x_n y_n$. Then,

$$\int \sum c_n x_n^2 \, dm = \sum c_n < \infty,$$

which allows to interpret m as a probability measure on H_c. Define the extended (i.e., possibly infinite) distance

$$d((x), (y)) = \begin{cases} \|(x) - (y)\|_{\ell_2} & \text{if } (x) - (y) \in \ell_2, \\ +\infty & \text{otherwise,} \end{cases}$$

called Cameron–Martin distance. Clearly, $\mathbb{R}^\infty$ is not connected, in the sense that two points have finite distance (i.e., they are connected) if and only if they belong to the same ℓ_2-leaf. In other words, $\mathbb{R}^\infty$ is foliated by ℓ_2. Nevertheless, if one considers random points instead of single points, the situation changes drastically and connectivity is recovered, thanks to the following fact. Let ρ_1, ρ_2 be densities functions such that $\int \rho_i \log \rho_i \, dm < \infty$, $i = 1, 2$ and set $\mu_i = \rho_i m$. Then μ_1, μ_2 have finite distance as stated in the result below.[2]

[2] We introduce the following notation. Given two measurable spaces (X, ξ_X), (Y, ξ_Y) and a measurable map $T : X \to Y$, we define the push-forward operator $T_\# : \mathcal{P}(X) \to \mathcal{P}(Y)$ which maps a probability measure

THEOREM 3.1 (Talagrand, Feyel–Ustunel). *There exists a displacement map $D : H_c \to \ell_2$ such that setting $T(x) = x + D(x)$ there hold $T_\# \mu_1 = \mu_2$ and the quantitative estimate*

$$\int_{H_c} d^2(x, T(x)) d\mu_1(x) \le 2 \left(\int_{H_c} \rho_1 \log \rho_1 \, dm + \int_{H_c} \rho_2 \log \rho_2 \, dm \right).$$

In the statement of Theorem 42 we combine an infinite-dimensional extension of Brenier's fundamental theorem of optimal transportation with the so-called Transport inequality.

EXAMPLE 3.6 (Measured Gromov–Hausdorff limits). Let $(M_n, g_n, m_n)_{n\in\mathbb{N}}$ be a sequence of weighted Riemannian manifolds of fixed dimension (see Example 3.3) such that the measured Gromov–Hausdorff limit exists and is denoted by (X, d, m). In the program developed in [**25**] several properties (rectifiability, tangent cones, etc) of the limit space are recovered by approximation. One of the scopes of the synthetic theory is the development of an "intrinsic" differential calculus, without reference to an approximating sequence. Ideally, one would like to fill the gap between these two viewpoints, i.e. identify the intrinsic structural properties which ensure the existence of an approximation; for the moment this goal seems to be very far.

$\mu \in \mathcal{P}(X)$ to the measure

$$T_\# \mu(B) = \mu(T^{-1}(B)).$$

More generally, the operator $T_\#$ is also a mass-preserving operator from $\mathcal{M}_+(X) \to \mathcal{M}_+(Y)$. This definition provides the change of variable formula

$$\int_Y \varphi \, dT_\# \mu = \int_X \varphi \circ T \, d\mu,$$

for every bounded or nonnegative ξ_Y-measurable map $\varphi : Y \to \mathbb{R}$.

CHAPTER 2

H-Sobolev space and first tools of differential calculus

Throughout the notes, (X, d, m) denotes a metric measure space, satisfying the following assumptions:

- (X, d) is a complete and separable metric space;
- m is a finite Borel measure on X.

The finiteness assumption is made just for simplicity, and it can be relaxed (at the price of many technical complications and less readable proofs) assuming that $m(B_r(x)) \leq Ce^{cr^2}$ for some positive constants c, C (here and in the sequel $B_r(x)$ denotes the open ball centered at x of radius r). Also, many results illustrated in these notes have a local nature and they hold simply under the assumption that m is finite on bounded sets. Anyhow, the goal of the notes is to illustrate the key ideas, not to try to reach the highest generality of the original papers.

1. Relaxed slope and Cheeger energy

Definition 1.1 (Slope and asymptotic Lipschitz constant). Given $f \in \mathrm{Lip}(X)$ we define the *slope of f at x* as

$$|\nabla f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|}{d(y, x)}, \tag{4}$$

and the *asymptotically Lipschitz constant of f at x* as

$$\mathrm{lip}_a(f, x) := \lim_{r \downarrow 0} \mathrm{Lip}(f, B_r(x)), \tag{5}$$

where $\mathrm{Lip}(f, B_r(x))$ is the Lipschitz constant of f on the ball $B_r(x)$.

One easily checks, by monotonicity, that the limit in (5) exists and that the function $x \mapsto \mathrm{lip}_a(f, x)$ is upper semicontinuous.

Given $f \in \mathrm{Lip}(X)$, $|\nabla f|$ and $\mathrm{lip}_a(f, \cdot)$ satisfy

$$|\nabla f|(x) \leq |\nabla f|^*(x) \leq \mathrm{lip}_a(f, x). \tag{6}$$

where $|\nabla f|^*(\cdot)$ is the upper semicontinuous relaxation of $|\nabla f|$. If (X, d) is a length space,[1] then $|\nabla f|^*(x) = \mathrm{lip}_a(f, x)$ for every Lipschitz function f.

[1] (X, d) is a *length space* if

$$d(x, y) = \inf\{\ell(\gamma) \mid \gamma(0) = x, \gamma(1) = y\} \qquad \forall x,\, y \in X,$$

the infimum being taken in the class of rectifiable curves. A length space is called *geodesic* if the infimum is attained.

DEFINITION 1.2 (H Sobolev space). Let $1 < p < \infty$.

$$H^{1,p}(X,d,m) := \Big\{ f \in L^p(X,m) \mid \exists\, f_n \in \mathrm{Lip}_b(X), \|f_n - f\|_{L^p} \to 0, \\ \limsup_{n\to\infty} \int_X \mathrm{lip}_a^p(f_n,x)\,dm(x) < \infty \Big\}, \tag{7}$$

where $\mathrm{Lip}_b(X) = \{f \in \mathrm{Lip}(X) \mid f \text{ is bounded}\}$.

For simplicity, we may use the short notation $H^{1,p}$ or H-space for $H^{1,p}(X,d,m)$.

Definition 1.2 is slightly different and in particular stronger than the original one given in [**24**], since we use here a stronger notion of pseudogradient (the asymptotic Lipschitz constant, instead of the slope) and a smaller class of approximating functions (bounded Lipschitz functions, instead of all functions with an upper gradient in L^p). An equivalent definition is based on the notion of relaxed slope.

DEFINITION 1.3. The *relaxed slope* of a function $f \in L^p(X,m)$ is the set

$$RS(f) := \{g \mid g \geq \tilde{g}\ m\text{-a.e., where } \tilde{g} \text{ is a } w\text{-}L^p \text{ limit of } \mathrm{lip}_a(f_n,\cdot), \|f_n - f\|_{L^p} \to 0\}. \tag{8}$$

By reflexivity, we immediately get

$$\{f \in L^p(X,m) \mid RS(f) \neq \emptyset\} = H^{1,p}(X,d,m).$$

Let us see how to identify the smallest relaxed slope of f, which amounts to find the best possible approximating sequence in (7). To do this, let us remark that, thanks to Mazur's Lemma, weak-L^p limits in (8) can be turned into strong-L^p limits. More precisely

$$RS(f) = \{g \mid g \geq \tilde{g}\ m\text{-a.e., where } \tilde{g} \text{ is a } s\text{-}L^p \text{ limit of } g_n \geq \mathrm{lip}_a(f_n,\cdot), \|f_n - f\|_{L^p} \to 0\}.$$

By diagonal arguments based on the characterization above, one gets the following property which, in the classical framework, corresponds to the closure of the gradient operator.

PROPOSITION 1.1. *Assume f_n converge strongly in L^p to f and $g_n \in RS(f_n)$ converge weakly in L^p to g. Then $g \in RS(f)$.*

Let us list some important consequences of the proposition above.

(A) $RS(f)$ is a weakly closed set. Moreover, since the asymptotically Lipschitz constant of f is convex as a function of f, $RS(f)$ is also convex, and therefore there exists a unique $g \in RS(f)$ with minimal L^p norm. We denote by $|\nabla f|_*$ such g. This notation is motivated by the fact that $|\nabla f|_*$ consists of a weak notion of modulus of the gradient of f obtained through an approximation procedure.

(B) Taking an approximating sequence for the distinguished element $|\nabla f|_*$, we immediately find a representation for the Cheeger energy. We define the *Cheeger functional* (or *Cheeger energy*) as

$$\mathscr{C}(f) := \inf\left\{ \liminf_{n\to\infty} \frac{1}{p} \int_X \mathrm{lip}_a^p(f_n,x)\,dm(x) \mid \|f_n - f\|_{L^p} \to 0 \right\} \tag{9}$$

i.e., the normalized lower semicontinuous relaxation of the energy defined in (7). Thanks to the previous property, the Cheeger energy can be represented through the minimal relaxed slope

$$\mathscr{C}(f) = \frac{1}{p} \int_X |\nabla f|_*^p\,dm.$$

(C) The Cheeger energy $\mathscr{C}$ is a convex lower semicontinuous functional on $L^p(X,m)$, and it has a dense domain in $L^p(X,m)$ (since it includes $\mathrm{Lip}_b(X)$). These ingredients are crucial to apply the theory of gradient flows.

2. Elements of differential calculus

PROPOSITION 2.1. *The following properties hold.*

(i) *Strong approximation: for every $f \in H^{1,p}$ there exists $f_n \in \mathrm{Lip}_b(X)$ such that $\mathrm{lip}_a(f_n)$ converges to $|\nabla f|_*$ strongly in $L^p(X,m)$.*
(ii) *Pointwise minimality: for every $g \in RS(f)$, $|\nabla f|_* \le g$ m-almost everywhere.*
(iii) *Locality: given f, $g \in H^{1,p}$, $|\nabla f|_* = |\nabla g|_*$ m-almost everywhere on $\{f = g\}$.*
(iv) *Chain rule: for every $\varphi \in \mathrm{Lip}(\mathbb{R})$ one has*

$$|\nabla(\varphi \circ f)|_* = |\varphi' \circ f||\nabla f|_* \quad m\text{-a.e. in } X. \tag{10}$$

Property (iv) is useful in connection with the Laplacian, see Section 4.

PROOF. Proof of (i). This follows by the fact weak convergence in L^p together with convergence of norms implies strong convergence in L^p.

Proof of (ii). Take a function $\chi \in \mathrm{Lip}(X,[0,1])$ and $f, \tilde{f} \in \mathrm{Lip}(X)$. Then one checks easily the inequality for asymptotically Lipschitz constants

$$\mathrm{lip}_a((1-\chi)f + \chi\tilde{f}) \le (1-\chi)\mathrm{lip}_a(f) + \chi\mathrm{lip}_a(\tilde{f}) + \mathrm{lip}(\chi)|f - \tilde{f}|. \tag{11}$$

Let $f \in H^{1,p}$ and $g, \tilde{g} \in RS(f)$. Take two sequences f_n and $\tilde{f}_n$ approximating f as in (7). Consider two sequences $g_n, \tilde{g}_n$ converging strongly in L^p to g and $\tilde{g}$, respectively, and such that $g_n \ge \mathrm{lip}_a(f_n)$, $\tilde{g}_n \ge \mathrm{lip}_a(\tilde{f}_n)$. Then, applying (11) to $f_n, \tilde{f}_n$ we deduce that

$$(1-\chi)g + \chi\tilde{g} \in RS(f), \text{ for every } g, \tilde{g} \in RS(f), \chi \in \mathrm{Lip}(X,[0,1]). \tag{12}$$

Since $RS(f)$ is closed, we get by approximation that property (12) holds for every g, $\tilde{g} \in RS(f)$ and every $\chi \in L^\infty(X,m)$, with $0 \le \chi \le 1$.

By contradiction, assume the set $B = \{g < |\nabla f|_*\}$ has positive measure $m(B) > 0$, for some $g \in RS(f)$. Take $\tilde{g} = |\nabla f|_*$ and $\chi = \chi_{X \setminus B}$, then the function $\hat{g} := (1-\chi)g + \chi\tilde{g}$ belongs to $RS(f)$. Then $B \subset \{x \mid \hat{g}(x) < |\nabla f|_*(x)\}$, which violates the minimality property that defines $|\nabla f|_*$.

Proof of (iii). Let $f, g \in H^{1,p}(X,d,m)$. By subadditivity of $h \mapsto |\nabla h|_*$, one reduces to the case $g = 0$. Therefore, it suffices to prove that $|\nabla f|_*(x) = 0$ m-almost everywhere on $\{f = 0\}$. We follow a classical proof's idea for Sobolev spaces, which relies on lower semicontinuity and not on the integration by parts formula. Take a small bump perturbation of the identity on the line, i.e., a sequence $\varphi_n \in \mathcal{C}^1(\mathbb{R})$, such that $0 \le \varphi_n' \le 1$, $\varphi_n(t) \to t$, $\varphi_n'(0) = 0$. To deduce (iii) we need a weaker version of property (iv), namely, the inequality

$$|\nabla(\varphi_n \circ f)|_* \le (\varphi_n' \circ f)|\nabla f|_*. \tag{13}$$

To obtain (13), since $\varphi_n \in \mathcal{C}^1(\mathbb{R})$, it suffices to check the same inequality for the asymptotically Lipschitz constants

$$\mathrm{lip}_a(\varphi_n \circ u, x) \le (\varphi_n' \circ u)\mathrm{lip}_a(u,x),$$

for every Lipschitz function u. Then one approximates $f \in H^{1,p}$ by Lipschitz functions to get (13). Let us show how (13) implies property (iii). By (13),

$$\int_X |\nabla(\varphi_n \circ f)|_*^p \, dm \le \int_X ((\varphi_n' \circ f)|\nabla f|_*)^p \, dm \le \int_{\{f \neq 0\}} |\nabla f|_*^p \, dm,$$

where the last inequality follows by our choice of φ_n. Then, letting n go to ∞, by lower semicontinuity, we infer

$$\int_X |\nabla f|_*^p \, dm \le \int_{\{f \neq 0\}} |\nabla f|_*^p \, dm,$$

which gives precisely that $|\nabla f|_*$ must vanish on the zero level set of f.

Proof of (iv). We give only the proof of the equality when $\varphi \in \mathrm{Lip}(\mathbb{R})$, $\varphi(0) = 0$, $\varphi' \ge 0$ and skip the general case.

Recall that, thanks to (13), we have

$$|\nabla(\varphi \circ f)|_* \le |\varphi' \circ f||\nabla f|_*, \quad \text{for every } \varphi \in \mathcal{C}^1(\mathbb{R}) \cap \mathrm{Lip}(\mathbb{R}). \tag{14}$$

Let us show how to get this inequality, in the case where $\varphi \in \mathrm{Lip}(X)$, $\varphi(0) = 0$ and $\varphi' \ge 0$.

First of all, remark that, when φ is only Lipschitz, φ' exists only almost everywhere and therefore formula (10) is undefined on the set $f^{-1}(N)$, where $N = \{t \mid \nexists\, \varphi'(t)\}$. Nevertheless, there holds $|\nabla f|_* = 0$ m-almost everywhere on $f^{-1}(E)$ whenever $\mathscr{L}^1(E) = 0$. Indeed, this has been shown above when E is a singleton. For a general negligible set E, take a compact set $K \subset E$ and a sequence $\varphi_n \in \mathcal{C}^1(\mathbb{R})$ such that $0 \le \varphi_n' \le 1$, $\varphi_n(t) \to t$ and $\varphi_n' = 0$ on K. Then the same proof works in the general case. Hence, since in (10) we multiply $\varphi' \circ f$ by $|\nabla f|_*$, whatever definition of φ' we may take on N, formula (10) holds true since both $|\nabla f|_*$ and $|\nabla(\varphi \circ f)|_*$ vanish.

Take any $\varphi \in \mathrm{Lip}(\mathbb{R})$, $\varphi' \ge 0$, and apply inequality (14) to $\varphi * \rho_\epsilon$, where ρ_ϵ is a family of mollifiers. Then, taking the limit, we deduce

$$|\nabla(\varphi \circ f)|_* \le (\varphi' \circ f)|\nabla f|_*. \tag{15}$$

Finally let us show the converse inequality, when $\varphi \in \mathrm{Lip}(\mathbb{R})$, $\varphi(0) = 0$, $\varphi' \ge 0$. Set $M = \|\varphi'\|_{L^\infty}$ and define $\psi(t) = Mt - \varphi(t)$. Then $\psi \in \mathrm{Lip}(\mathbb{R})$, $\psi' \ge 0$ and $\psi(0) = 0$. Moreover, $Mf = \psi \circ f + \varphi \circ f$. Hence

$$|\nabla(Mf)|_* \le |\nabla(\varphi \circ f)|_* + |\nabla(\psi \circ f)|_* \le (\varphi' \circ f + \psi' \circ f)|\nabla f|_* = M|\nabla f|_*,$$

where the first inequality follows by subadditivity and the second one by (15) applied to ψ. Finally, since $|\nabla(Mf)|_* = M|\nabla f|_*$ all inequalities above become equalities. □

Before going further let us make two important observations.

First observation. So far, the only assumptions needed on (X, d, m) are that (X, d) is complete and m is finite. It may happen, without further conditions, that the H-space is trivial, i.e., it may coincide with L^p (which, by a simple exercise, is *equivalent* to say that the Cheeger energy vanishes identically). To see this, let us show a simple example.

EXAMPLE 2.1. Consider $(\mathbb{R}, |\cdot|, m)$, where the measure is concentrated on rational numbers, that is $m = \sum_k 2^{-k}\delta_{q_k}$, with $k \to q_k$ any enumeration of $\mathbb{Q}$. In some sense, m is too concentrated to get a reasonable Sobolev space. To see why, the argument is somehow dual to the one used to prove locality property of $|\nabla f|_*$. Take open sets A_n with $\mathbb{Q} \subset A_n$ and $\mathscr{L}^1(A_n) \downarrow 0$. Build up a sequence $\varphi_n \in \mathrm{Lip}(\mathbb{R})$ such that $\varphi_n(t) \to t$, $0 \le \varphi_n' \le 1$ and

$\varphi_n' = 0$ on A_n. Let $f \in \mathrm{Lip}_b(\mathbb{R})$ and consider $f \circ \varphi_n$. Then $\mathrm{lip}_a(f \circ \varphi_n) \equiv 0$ on A_n and, since m is concentrated on $\mathbb{Q}$, $\mathscr{C}(f \circ \varphi_n) = 0$. Now, since f is Lipschitz, $f \circ \varphi_n$ converges to f and by lower semicontinuity of the Cheeger energy we deduce $\mathscr{C}(f) = 0$. Finally, by density of Lipschitz functions and by lower semicontinuity of $\mathscr{C}$, we conclude that $\mathscr{C}(f) = 0$ for every $f \in L^p(X, m)$.

This shows that the theory outlined above is very general, at the expense of a risk of triviality of the H-space.

Second observation. Recall that the Cheeger energy is computed as

$$\mathscr{C}(f) = \frac{1}{p} \int_X |\nabla f|_*^p dm.$$

One should be aware that $\mathscr{C}(\cdot)$ is not a quadratic functional, even if $p = 2$ (see Example 2.2 below). This motivates the following definition.

DEFINITION 2.1. [Asymptotically Hilbertian spaces] The space (X, d, m) is said to be *asymptotically Hilbertian* if $\mathscr{C}_2$ is a quadratic form[2].

Obviously $\mathscr{C}_2$ is 2-homogeneous but it may not be quadratic and, in general, the parallelogram identity may fail.

As we mention in Section 2, being Hilbertian is the condition needed in the Lott–Sturm–Villani theory to get (a basic) equivalence with the Bakry–Émery theory. In this connection, an interesting open question is to find other properties or characterizations of asymptotically Hilbertian spaces, maybe independent of $\mathscr{C}_2$ (in the Lagrangian side of the theory, a sufficient condition is provided by the so-called EVI property of the entropy, see [**14**]).

REMARK 2.1. Sub-Riemannian spaces are asymptotically Hilbertian.

In general, we cannot expect $\mathscr{C}_2$ to be quadratic.

EXAMPLE 2.2. Consider the space $(\mathbb{R}^2, \|\cdot\|_\infty, \mathscr{L}^2)$. Let $f \in \mathcal{C}^1(\mathbb{R}^2)$. Then, it is easy to check that the slope of f coincides with the ℓ_1-norm of f, i.e., $|\nabla f| = \|\nabla f\|_1$. Using standard tools (namely convolutions to pass from $\mathcal{C}^1$ functions to Sobolev functions), one deduces that

$$\mathscr{C}_2(f) = \frac{1}{2} \int_{\mathbb{R}^2} \left(\left| \frac{\partial f}{\partial x_1} \right| + \left| \frac{\partial f}{\partial x_2} \right| \right)^2 dx_1 dx_2$$

which is 2-homogeneous, but not a quadratic form.

This phenomenon is well-known in Finsler geometry where the metric is not induced by (the square of) a quadratic form, but by a norm. Let us explain where the lack of quadraticity of $\mathscr{C}_2$ comes from. Consider the space $(E, \|\cdot\|, \mathscr{L}^n)$, where E is $\mathbb{R}^n$ and $\|\cdot\|$ is a generic norm. Then the operator $f \mapsto df$ associating to a function its differential is linear. The gradient is defined as $\nabla f = J(df)$, where $J : E^* \to E$ is the (possibly multivalued) duality map built as follows. Given $v^* \in E^*$, $J(v^*)$ is any vector $v \in E$ such that $\|v^*\|_{E^*} = \|v\|_E$ and

$$\langle v^*, v \rangle = \|v^*\|_{E^*} \|v\|_E,$$

[2]To emphasize the dependence of $\mathscr{C}$ on p we will sometimes use the notation $\mathscr{C}_p(f) = \frac{1}{p}\int_X |\nabla f|_*^p dm$.

where $\langle \cdot, \cdot \rangle$ is the canonical duality between E and E^*. Note that, in general, one has

$$\langle v^*, w \rangle \leq \|v^*\|_{E^*} \|w\|_E,$$

so to define J we are essentially requiring the equality. Of course, v may not be unique, unless we ask some strict convexity assumption on E, hence J may be multi-valued. Nevertheless, even in the single-valued case, J is not linear if the norm is not Hilbertian. More precisely, J is linear if and only if the original norm comes from a quadratic form. Hence, in general, the operator $f \mapsto \nabla f$ is not linear and $f \mapsto \mathscr{C}_2(f)$ is not quadratic.

3. Reminders of convex analysis

To introduce the notion of Laplacian, let us start with some reminders of convex analysis.

Definition 3.1 (Subdifferential). Let E be a Banach space, let $\Phi : E \to \mathbb{R} \cup \{+\infty\}$ be convex, lower semicontinuous and let $x \in \operatorname{dom} \Phi = \{\Phi < \infty\}$. The *subdifferential of* Φ *at* x is the set

$$\partial\Phi(x) = \{p \in E^* \mid \Phi(y) \geq \Phi(x) + \langle p, y - x \rangle, \ \forall\, y \in E\}.$$

Namely, $\partial\Phi(x)$ is the collection of slopes of hyperplanes touching from below the graph of Φ at x.

If E is reflexive and E^* is strictly convex, the *gradient of* Φ at x, denoted $\nabla\Phi(x)$, is the unique element of $\partial\Phi(x)$ having minimal norm.

The subdifferential is always a closed, possibly empty, set.

Let us see the relation between $\nabla\Phi$ and the metric slope defined in Definition 1.1. At this point, one should be aware that the notation $|\nabla\Phi|$ indicates both the modulus of $\nabla\Phi$ and the slope of Φ. In the sequel, one can deduce from the context which of the two objects is actually involved.

Dealing with (downward) gradient flows of convex functions, what matters is how much one can decrease a convex function, rather than how much it increases. This motivates the following definition of a one-sided slope.

Definition 3.2 (Descending slope). Given a function f on a metric space, we define the *descending slope of* f *at* x as

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

To see how the descending slope quantifies how much f decreases, note that $|\nabla^- f|(x) = 0$ if x is a minimum point of f.

The connection between the descending slope and the gradient of a convex lower semicontinuous functional is given by the inequality

$$|\nabla^- \Phi|(x) \leq \|\nabla\Phi(x)\|_E^*.$$

Indeed, let $p \in \partial\Phi(x)$. Then, for every $y \in E$ one has

$$\Phi(x) - \Phi(y) \leq \langle p, y - x \rangle \leq \|p\|_{E^*} \|x - y\|_E = \|p\|_{E^*} d(x, y).$$

Thus

$$(\Phi(x) - \Phi(y))^+ \leq \|p\|_{E^*} d(x, y).$$

and we conclude $|\nabla^-\Phi|(x) \le \|p\|_{E^*}$. Finally, minimizing the norm of $p \in \partial\Phi(x)$ we get

$$|\nabla^-\Phi|(x) \le \inf\{\|p\|_{E^*} \mid p \in \partial\Phi(x)\} = \|\nabla\Phi(x)\|_E.$$

Notice that this proof used neither the convexity nor the lower semicontinuity of Φ. As an exercise, the reader may use Hahn-Banach Theorem to prove, under these assumptions, that

$$|\nabla^-\Phi|(x) = \|\nabla\Phi(x)\|_E. \tag{16}$$

4. Laplacian and integration by parts formula

To simplify the situation, in this section and in Section 5 we consider only the case $p = 2$, so that $L^2(X, m)$ is a Hilbert space and $\mathscr{C} = \mathscr{C}_2$.

Definition 4.1 (Laplacian). Let $f \in \operatorname{dom}\mathscr{C}$ and $\partial\mathscr{C}(f) \neq \emptyset$. We define the *Laplacian of f* by

$$-\Delta f = \nabla\mathscr{C}(f),$$

that is, $-\Delta f$ is the unique element in $\partial\mathscr{C}(f)$ having minimal L^2-norm.

To motivate the construction above, we prove the following simple fact.

Proposition 4.1. *Consider $(\mathbb{R}^n, |\cdot|, \mathscr{L}^n)$ and let $f \in D(\mathscr{C})$, $p \in L^2(\mathbb{R}^n)$. Then*

$$\frac{1}{2}\int_{\mathbb{R}^n} |\nabla(f+\epsilon g)|^2 dx - \frac{1}{2}\int_{\mathbb{R}^n} |\nabla f|^2 dx \ge \int_{\mathbb{R}^n} \langle p, \epsilon g\rangle dx,\ \forall\, \epsilon > 0, \forall\, g \in \mathcal{C}^\infty_c(\mathbb{R}^n), \tag{17}$$

if and only if $p = -\Delta f$ in the sense of distributions.

Proof. Assume (17) holds. By differentiation, we get

$$\int_{\mathbb{R}^n} \nabla f \cdot \nabla g\, dx \ge \int_{\mathbb{R}^n} \langle p, \epsilon g\rangle dx,$$

and replacing g with $-g$ we conclude the equality

$$\int_{\mathbb{R}^n} \nabla f \cdot \nabla g\, dx = \int_{\mathbb{R}^n} \langle p, \epsilon g\rangle dx.$$

This means that

$$p = -\Delta f,\ \text{in } \mathscr{D}',$$

i.e., p coincides with $-\Delta f$ in the sense of distributions. Conversely, if $p \in \partial\mathscr{C}(f)$ is such that $p = -\Delta f$, in $\mathscr{D}'$, then p satisfies (17) because of convexity. □

Therefore, in $\mathbb{R}^n$ we can say that $\mathscr{C}(\cdot)$ is subdifferentiable at f if and only if the Laplacian of f is in L^2 in the sense of distributions and in this case $\partial\mathscr{C}(f)$ contains a unique element.

Remark 4.1. The Laplacian operator $f \mapsto \Delta f$ is 1-homogeneous but not linear, unless X is asymptotically Hilbertian, see the second observation in Section 2.

Despite of the non-linearity of Δ, one can recover some tools of differential calculus, in particular an integration by parts formula.

Proposition 4.2. *Let $f \in \operatorname{dom}\mathscr{C}$ and $\partial\mathscr{C}(f) \neq \emptyset$. Then, for every $g \in \operatorname{dom}\mathscr{C}$,*

$$-\int_X g\Delta f\, dm \leq \int_X |\nabla g|_* |\nabla f|_*\, dm,$$

and equality holds whenever $g = \varphi \circ f$, with $\varphi \in \mathrm{Lip}(\mathbb{R})$, φ nondecreasing.

Proof. Take $\epsilon > 0$. Since $-\Delta f \in \partial\mathscr{C}(f)$ and by subadditivity

$$-\int_X \epsilon g\Delta f\, dm \leq \frac{1}{2}\int_X \left(|\nabla(f+\epsilon g)|_*^2 - |\nabla f|_*^2\right) dm \leq \frac{1}{2}\int_X \left((|\nabla f|_* + \epsilon|\nabla g|_*)^2 - |\nabla f|_*^2\right) dm.$$

Dividing by ϵ and letting ϵ to 0 gives the required inequality.

Finally, a refinement of the above argument gives the equality in the special case where $g = \varphi \circ f$ and φ is nondecreasing. □

5. Heat flow in (X, d, m)

In order to introduce the heat flow in a metric measure framework, we will apply the classical theory of evolution equations for maximal monotone operators developed in the '70s by Komura and Brezis.

Reminders about Komura–Brezis theory. Let H be a Hilbert space. Note that in this theory the Hilbert structure is essential, as a generalization to the general Banach case is not straightforward and provides much weaker results. This is why in the sequel we deal only with the case $p = 2$ (see [**12**] and [**5**] for the general treatment of the cases $1 < p < \infty$).

Let $\Phi : H \to \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous functional and pick an element $\bar{u} \in \overline{\{\Phi < \infty\}}$. Consider the Cauchy problem[3]

$$\begin{cases} u'(t) \in -\partial\Phi(u(t)), \text{ a.e. } t, \\ u \in AC^2_{loc}((0,\infty), H), \\ \lim_{t\to 0^+} u(t) = \bar{u}. \end{cases} \tag{18}$$

Then we have the following facts.

- Problem (18) has a unique solution and it generates a contraction semigroup in $\overline{\{\Phi < \infty\}}$, that is, if $v(\cdot)$ is the solution of the Cauchy problem

$$\begin{cases} v'(t) \in -\partial\Phi(v(t)), \text{ a.e. } t, \\ v \in AC^2_{loc}((0,+\infty), H), \\ \lim_{t\to 0^+} v(t) = \bar{v}, \end{cases} \tag{19}$$

  then

$$|u(t) - v(t)| \leq |\bar{u} - \bar{v}| \qquad \forall t \geq 0. \tag{20}$$

- The solution satisfies the following regularizing effects.

[3]Here $AC^2_{loc}((0,\infty), H) = \{u \in AC_{loc}((0,\infty), H) \mid u' \in L^2_{loc}((0,\infty), H)\}$. Note that this is the good space to look for a solution as the derivative of u might blow-up at 0, so one needs to consider functions which are locally absolutely continuous and have $L^2_{\rm loc}$ derivative.

1. For any $\bar{u} \in \overline{\{\Phi < \infty\}}$, in particular even for initial data having infinite energy, $\Phi(u(t)) < \infty$ for any positive t. Moreover, the quantitative inequality
$$\Phi(u(t)) \leq \inf_v \left\{ \Phi(v) + \frac{1}{2t}\|v - \bar{u}\|^2 \right\}$$
provides an estimate of the blow-up rate of Φ as t goes to 0.
2. The first condition in (18) is a differential inclusion. Then the derivative $u'(t)$ of the unique solution $u(\cdot)$ precisely selects for a.e. t the opposite of the element with minimal norm in the sub-differential, i.e. $u'(t) = -\nabla\Phi(u(t))$ for a.e. $t > 0$.
3. Even if Φ is only lower semicontinuous, the map $t \mapsto \Phi(u(t))$ is $AC_{loc}(0, \infty)$ and
$$\frac{d}{dt}(\Phi \circ u)(t) = -|u'(t)||\nabla\Phi|(u(t)) = -|u'(t)||\nabla^-\Phi|(u(t)),$$
where the last equality follows by the previous item.

The most nontrivial part of this theory is existence of solutions, as we do not have compactness, so good a priori Cauchy estimates in a possibly noncompact (but complete) setting are to be found. Also deducing the regularizing effects 1 and 2 is not trivial. On the other hand, the contraction property (20) and, as a consequence, uniqueness can be easily proved.

Proof of (20). Let $p \in \partial\Phi(x)$. Then, by definition,
$$\Phi(y) - \Phi(x) \geq \langle p, y - x\rangle.$$
Exchanging x and y, for every $q \in \partial\Phi(y)$,
$$\Phi(x) - \Phi(y) \geq \langle q, x - y\rangle.$$
Summing the above inequalities we obtain the monotonicity inequality[4]
$$\langle p - q, y - x\rangle \leq 0, \ \forall\, p \in \partial\Phi(x), \ \forall\, q \in \partial\Phi(y). \tag{21}$$
Let now $u(\cdot)$ be a solution of (18) and $v(\cdot)$ be a solution of (19). Then
$$\frac{1}{2}\frac{d}{dt}\|u(t) - v(t)\|^2 = \langle u'(t) - v'(t), u(t) - v(t)\rangle \leq 0,$$
where the last inequality is a consequence of $u'(t) \in -\partial\Phi(u(t))$, $v'(t) \in -\partial\Phi(v(t))$ and of (21). This proves the contractivity property (20). Now, if $\bar{u} = \bar{v}$, (20) gives directly $u = v$. □

Heat flow. Let us apply the theory above in our framework. Recall that, in the case $p = 2$, the Cheeger energy $\mathscr{C}$ is convex, lower semi-continuous on $L^2(X, m)$, and has dense domain, i.e., $\overline{\{\mathscr{C} < \infty\}} = L^2(X, m)$. Applying the Komura–Brezis theory we get that, when $\Phi = \mathscr{C}$, the Cauchy problem (18) has a unique solution for every initial condition in L^2.

[4] This is why this theory is concerned with the so-called monotone operators.

According to our definition of Laplacian, see Definition 4.1, since by definition $-\Delta u$ is the element in $\partial\mathscr{C}(u)$ of minimal norm, Cauchy problem (18) becomes precisely

$$\begin{cases} \partial_t u = \Delta u, \\ u(0) := \lim_{t\to 0^+} u(t) \in L^2(X,m). \end{cases}$$

The Laplacian and the heat flow defined above will play a fundamental role in proving equivalence of Sobolev spaces.

Remark 5.1. In these notes we develop a theory of first order Sobolev spaces $H^{1,p}(X,d,m)$. Let us mention that getting a general higher order theory of Sobolev spaces is an open problem, object of current investigations. This requires a good notion of vector field on metric spaces (see [**21**], [**52**], [**20**] for recent papers on this subject), as well as a way of differentiating vector fields. So far, this has been done only in the asymptotically Hilbertian case. The problem is still open even for the low dimensional case of tangent vector fields to curves. We refer the reader to the recent work by Gigli [**33**], where the authour deals with the case $H^{2,2}(X,d,m)$, under some curvature assumptions on the metric measure structure.

CHAPTER 3

The Lagrangian (Beppo Levi) approach

In this chapter we deal with definitions and properties of Sobolev space based on a Lagrangian point of view. More precisely, we introduce a second notion of Sobolev space, denoted by $BL^{1,p}$, prove the rather easy inclusion $H^{1,p} \subset BL^{1,p}$, and finally prove the converse inclusion which is much less trivial and requires, among others, the tools provided in the previous chapter.

1. Basic tools

Definition 1.1 (Absolute continuity). Let $1 \leq q \leq \infty$. A curve $\gamma : [0,1] \to X$ is *absolutely continuous* if there exists a nonnegative $g \in L^q(0,1)$ such that

$$d(\gamma(s), \gamma(t)) \leq \int_s^t g(\tau) d\tau, \ \forall \, s, \, t \in [0,1], \ s \leq t. \tag{22}$$

We denote by $AC^q([0,1], X)$ (occasionally shortened to AC^q) the set of absolutely continuous curves satisfying (22). Clearly $AC^\infty([0,1], X) = \mathrm{Lip}([0,1], X)$.

Assume that an absolutely continuous curve is given. It is derirable to identify the best (minimal) function g satisfying (22). This motivates the notion of metric derivative.

Theorem 1.1 (Existence of metric derivative). *Let $\gamma \in AC^q([0,1], X)$. Then the limit*

$$|\gamma'|(t) := \lim_{h \to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|}$$

exists for almost every $t \in (0,1)$ and it is minimal a.e. with respect to condition (22), i.e., for every g satisfying (22) one has

$$|\gamma'| \leq g, \ \mathscr{L}^1\text{-almost everywhere in } (0,1).$$

Definition 1.2 (q-action). Given $\gamma \in AC^q$, we define the *q-action* of γ as

$$\mathscr{A}_q(\gamma) := \int_0^1 |\gamma'|^q(t) dt.$$

Obviously, a curve has finite q-action if and only if it is absolutely continuous. We use the notation q, since typically to study $W^{1,p}$ we need AC^q, for $q = p/(p-1)$ (and, in particular, in the limiting case of BV or $W^{1,1}$ functions, we need Lipschitz curves).

Definition 1.3 (Curvilinear integral). Let $g : X \to [0, \infty]$ be a Borel function, we define the *curvilinear integral of g* on $\gamma \in AC^q$ by

$$\int_\gamma g := \int_0^1 g(\gamma(t)) |\gamma'|(t) dt.$$

The curvilinear integral defined in this way retains the usual properties valid in the context of Riemannian manifolds (invariance by reparameterization, concatenations, etc.).

Given $\gamma \in AC^q$, the operator $\gamma_\# : \mathcal{M}_+([0,1]) \to \mathcal{M}_+(X)$ pushes positive measures on $[0,1]$ to positive measures on X. Setting $\mathcal{J}\gamma = \gamma_\#(|\gamma'|\mathscr{L}^1)$, by the change of variable formula, the curvilinear integral has the equivalent representation

$$\int_\gamma g = \int_X g \, d\mathcal{J}\gamma.$$

The following notion comes from complex analysis [**2**] and was later fully developed by Fuglede [**31**].

Definition 1.4 (p-Modulus). Let $\Gamma \subset AC^q([0,1], X)$ be a family of curves. The p-*Modulus* of Γ is

$$\mathrm{Mod}_p(\Gamma) := \inf \left\{ \int_X \rho^p dm \mid \rho : X \to [0, +\infty], \text{ Borel s.t. } \int_\gamma \rho \geq 1, \ \forall \gamma \in \Gamma \right\}$$

By construction, Mod_p is an outer measure in the space of absolutely continuous curves. Moreover, the notion of p-Modulus is invariant under reparameterization.

The following definition of Sobolev space has been introduced in [**31**], more or less at the same time when the points of view of Schwarz (distributions, integration by parts) and Sobolev (completion of smooth functions) where being developed. We can now state the result as a theorem, namely all these points of view are equivalent.

Theorem 1.2 (Fuglede).

$$H^{1,p}(\mathbb{R}^n) = \Big\{ f : X \to \mathbb{R} \mid \exists F \in L^p(\mathbb{R}^n, \mathbb{R}^n), \exists \tilde{f} \text{ representative of } f \text{ s.t. } \int |f|^p \, dx < \infty \text{ and}$$
$$\tilde{f}(\gamma_1) - \tilde{f}(\gamma_0) = \int_\gamma F, \text{ for } \mathrm{Mod}_p\text{-almost every curve } \gamma \Big\}.$$

Requiring the property to hold for Mod_p*-almost every curve means that it may fail on a family* Γ *such that* $\mathrm{Mod}_p(\Gamma) = 0$*. For every* $f \in H^{1,p}$ *the vector field* F *is unique and it coincides with the weak derivative of* f *in the sense of distributions.*

Remark 1.1. This approach is consistent with the original one by Beppo Levi (see Definition 1.3). Indeed, the collection of vertical (or horizontal) lines in $\mathbb{R}^2$ has positive Mod_p modulus. Thus, as a consequence of the invariance of p-Modulus under the action of the rotation group, Fuglede obtained that the Beppo Levi definition has a frame invariant counterpart. As a matter of fact, using the fine theory of Sobolev and BV functions which characterizes in a frame-indifferent way the good representative, see [**58, 8**], one can prove that even the original definition of Levi is frame invariant.

2. The metric case

Following the ideas of Koskela–MacManus [**43**] and Shanmugalingam [**53**], let us adapt Fuglede's point of view to the metric framework. The following definition is due to Heinonen–Koskela [**39**].

Definition 2.1 (Upper gradient). Let $f : X \to \mathbb{R}$. We say that a Borel $g : X \to [0, \infty]$ is an *upper gradient* of f if

$$|f(\gamma_1) - f(\gamma_0)| \leq \int_\gamma g, \tag{23}$$

for every $\gamma \in AC^q$.

For instance, if $f \in \mathrm{Lip}(X)$ then the slope $|\nabla f|$ is an upper gradient and the same holds for the asymptotically Lipschitz constant $\mathrm{lip}_a(f)$, which is larger than $|\nabla f|$. To see this, note that the real curve $t \mapsto f(\gamma(t))$ is absolutely continuous, since $\gamma \in AC^q$ and $f \in \mathrm{Lip}(X)$. Therefore

$$|f(\gamma_1) - f(\gamma_0)| \leq \int_0^1 \left| \frac{d}{dt}(f \circ \gamma) \right| dt \leq \int_0^1 |\nabla f| \circ \gamma |\gamma'| \, dt = \int_\gamma |\nabla f|,$$

where the last inequality follows by the pointwise inequality $|(f \circ \gamma)'| \leq |\nabla f||\gamma'|$, which is easy to check.

In [**43**] a *weak* upper gradient is defined by the validity of (23) along Mod_p-a.e. curve γ (we will actually borrow their terminology and use the locution *weak upper gradient* for a more probabilistic notion of negligibility of curves). This was used in [**53**] to define a corresponding Sobolev class, in the same spirit of Fuglede's Euclidean definition of Theorem 1.2 (with an inequality replacing the equality).

Definition 2.2 (Newtonian space). We define the *Newtonian space*

$$N^{1,p}(X, d, m) := \Big\{ f \in L^p(X, m) \mid \exists \, g \in L^p(X, m), \exists \, \tilde{f} \text{ representative of } f \text{ s.t.} \\ |\tilde{f}(\gamma_1) - \tilde{f}(\gamma_0)| \leq \int_\gamma g, \ \ \mathrm{Mod}_p\text{-a.e. } \gamma \Big\} \tag{24}$$

Remark 2.1. It is possible, modifying g in $L^p(X, m)$ by an arbitrarily small amount, to obtain that the inequality (24) holds *for every* curve.[1] However, when passing to limits, we will see that the "a.e. curve" formulation is much more flexible.

Let us see how condition (24) is meaningful. Of course, when $\int_\gamma g = \infty$ the condition is trivial. The question is, given $g \in L^p(X, m)$, how many curves γ satisfy $\int_\gamma g < \infty$.

Proposition 2.1 (Markov–Chebishev inequality for Mod_p). *For every $g \in L^p(X, m)$,*

$$\int_\gamma g < \infty, \quad \textit{for } \mathrm{Mod}_p\textit{-almost every } \gamma.$$

Proof. For every $M > 0$, set $\Gamma_M = \{\gamma \mid \int_\gamma g \geq M\}$. Then, by definition of Mod_p,

$$\mathrm{Mod}_p(\Gamma_M) \leq \int_X \left(\frac{g}{M}\right)^p dm = M^{-p} \|g\|_p^p.$$

Hence, letting M go to $+\infty$ we get the conclusion. □

[1] The proof of this fact uses the very definition of Mod_p and is left as an exercise.

3. p-test plans and their relation with p-Modulus

Our final goal is to provide a connection between the Newtonian space and the H-space in Definition 1.2, and show their equivalence with a third notion of Sobolev space that stems from an integration by parts formula. In particular, concerning the relation with the integration by parts formula, it will turn out that Mod_p is not the optimal tool to provide a connection with vector fields. This is one of the motivations of the new approach, proposed in [**13**], which yields a more probabilistic way to measure exceptional curves.

In the sequel we endow the complete and separable metric space $\mathcal{C}([0,1],X)$ (with the sup norm) with the Borel σ-algebra and consider the class $\mathcal{P}(\mathcal{C}([0,1],X))$ of Borel probability measures in $\mathcal{C}([0,1],X)$.

Definition 3.1 (p-test plan). A probability measure $\pi \in \mathcal{P}(\mathcal{C}([0,1],X))$ is a *p-test plan* if

(1) π is concentrated on $AC^q([0,1],X)$, with $q = \frac{p}{p-1}$;
(2) there exists $C = C(\pi) \geq 0$ such that

$$(e_t)_{\#}\pi \leq Cm, \quad \forall\, t \in [0,1], \tag{25}$$

where $e_t : \mathcal{C}([0,1],X) \to X$ is $e_t(\gamma) = \gamma_t$, the evaluation map at t.

In more explicit terms, condition (25) means that for every test function $\psi \geq 0$ one has

$$\int \psi(\gamma_t)\, d\pi(\gamma) \leq C \int_X \psi\, dm \qquad \forall t \in [0,1].$$

Heuristically, property 2 is a "non-concentration" condition, in the sense that at any given time t, this family of curves does not concentrate too much relatively to the reference measure m. With this in mind, we call the smallest $C(\pi)$ in the previous definition the *compression constant* of π.

Using p-test plans, we introduce the dual notion of p-negligibility as follows.

Definition 3.2 (p-negligibility). A family of curves $\Gamma \subset AC([0,1],X)$ is said to be *p-negligible* if $\pi(\Gamma) = 0$ for every p-test plan π.

Remark 3.1. To check whether a family of curves is p-negligible it is sufficient to consider p-test plans with bounded $\mathscr{A}_q$ action. More precisely, given a p-test plan π and a positive constant M we set

$$\pi_M = \frac{\pi \llcorner_{\{\gamma | \mathscr{A}_q(\gamma) \leq M\}}}{\pi(\{\gamma \mid \mathscr{A}_q(\gamma) \leq M\})}.$$

Then, by monotone approximation, Γ is p-negligible if and only if, for every test plan π and for every $M > 0$, $\pi_M(\Gamma) = 0$.

Definition 3.3 (Beppo-Levi space). The *Beppo-Levi space* is

$$BL^{1,p}(X,d,m) := \left\{ f \in L^p(X,m) \mid \exists\, g \in L^p(X,m) \text{ s.t. } |f(\gamma_1) - f(\gamma_0)| \leq \int_\gamma g, \text{ for } p\text{-a.e.}\gamma \right\}.$$

A function g such that the inequality above holds for p-almost every γ is called a *weak upper gradient* of f, and we write $g \in WUG(f)$.

Let us compare the notions of Mod_p-negligibility and p-negligibility. An important difference is that, due to condition 2 in Definition 3.1, which is a non concentration condition at every time t, p-negligibility is not parametric-free. However, one always has the following implication.

Proposition 3.1. *Let* $\Gamma \subset AC^q([0,1],X)$*. Then* Γ *is p-negligible whenever* $\mathrm{Mod}_p(\Gamma) = 0$*. In particular the Newtonian space is contained in the Beppo-Levi space.*

Proof. Let π be a p-test plan such that $\mathscr{A}_q \in L^\infty(\pi)$ and let $\rho \geq 0$ be such that $\int_\gamma \rho \geq 1$ for every $\gamma \in \Gamma$. Integrating this inequality with respect to π over Γ,

$$\int_\Gamma \int_\gamma \rho \, d\pi(\gamma) \geq \int_\Gamma d\pi(\gamma) = \pi(\Gamma).$$

By Hölder inequality,

$$\int_\Gamma \int_\gamma \rho \, d\pi(\gamma) = \int_\Gamma \int_0^1 \rho(\gamma_t)|\gamma'|(t) dt \, d\pi(\gamma) \leq \left(\int_\Gamma \int_0^1 \rho^p(\gamma_t) dt \, d\pi(\gamma) \right)^{1/p} \left(\int_\Gamma \mathscr{A}_q(\gamma) d\pi(\gamma) \right)^{1/q}.$$

Moreover, using the bounded compression property, one has

$$\left(\int_\Gamma \int_0^1 \rho^p(\gamma_t) dt \, d\pi(\gamma) \right)^{1/p} \leq C(\pi)^{1/p} \|\rho\|_{L^p(X,m)}.$$

Collecting all the inequalities, since π has bounded action, we proved that for a certain constant $\tilde{c}(\pi)$,

$$\pi(\Gamma) \leq C(\pi)^{1/p} \tilde{c}(\pi) \|\rho\|_{L^p(X,m)}.$$

Then the statement is proved, once we recall that $\mathrm{Mod}_p(\Gamma) = 0$ implies there exists $\rho_n \geq 0$ such that $\|\rho_n\|_{L^p} \to 0$ and for which the inequality above holds with $\rho = \rho_n$. □

The obvious consequence of Proposition 3.1 is the easy inclusion $N^{1,p} \subset BL^{1,p}$.

Even though this will not play a role in the rest of the notes, let us go further in the analysis of the connection between probability measures and p-Modulus. Let $\eta \in \mathcal{P}(\mathcal{C}([0,1],X))$ be a probability measure concentrated on $AC^q([0,1],X)$ and consider the map $\gamma \mapsto \mathscr{J}\gamma$ which associates with $\gamma \in AC^q([0,1],X)$ the measure $\mathscr{J}\gamma = \gamma_\#(|\gamma'|\mathscr{L}^1) \in \mathcal{M}_+(X)$. We call *barycenter of* η the positive measure on X given by

$$\mathrm{Bar}(\eta)(E) := \int_{\mathcal{C}([0,1],X)} \mathscr{J}\gamma(E) \, d\eta(\gamma) \qquad E \subset X \text{ Borel.}$$

Notive that $\mathscr{J}$ is invariant under reparameterizations. The analogue of the non concentration property of test plans in this parametric-free context is to consider probabilities measures η for which the barycenter is not too concentrated in the sense below. We set

$$\|\mathrm{Bar}(\eta)\|_q = \begin{cases} +\infty, & \mathrm{Bar}(\eta) \not\ll m, \\ \left(\int_X g^q\right)^{1/q}, & \mathrm{Bar}(\eta) = gm. \end{cases}$$

Theorem 3.2 (Duality formula [**7**]). *For all Borel (or even Suslin) sets* Γ *relative to a suitable topology on the set of non-parametric curves, one has*

$$\mathrm{Mod}_p(\Gamma) = \sup \left\{ \frac{1}{\|\mathrm{Bar}(\eta)\|_q^q} \mid \eta \in \mathcal{P}(\Gamma) \right\}.$$

The proof of the inequality "$\geq$" is rather easy as a direct consequence of the definitions, and it is similar to the one of Proposition 3.1. The proof of the opposite inequality is more involved and rests essentially on an application of Hahn–Banach Theorem.

Let us mention that the notion of p-Modulus as well as the result above apply to families of hypersurfaces (instead of curves) and, more generally, to families of measures. Indeed, all the concepts introduced so far do not really depend on γ, but rather on the measure $\mathscr{J}(\gamma)$.

4. The inclusion $H^{1,p}(X,d,m) \subset BL^{1,p}(X,d,m)$

The comparison between p- and Mod_p-negligibility reveals how the notion of Beppo Levi space in Definition 3.3 is more flexible (although somehow weaker) than the one of Newtonian spaces of Definition 2.2, in the sense that one does not need to care about the good representative in (24).

PROPOSITION 4.1. *Let $f \in BL^{1,p}(X,d,m)$ and $\tilde{f} \in L^p(X,d,m)$ such that $m(\{f \neq \tilde{f}\}) = 0$. Then $\tilde{f} \in BL^{1,p}(X,d,m)$. Morever, if $g \in WUG(f)$ and $\tilde{g} \in L^p_+(X,m)$ is such that $m(\{g \neq \tilde{g}\}) = 0$ then $\tilde{g} \in WUG(f)$.*

PROOF. Let f and g satisfy

$$|f(\gamma_1) - f(\gamma_0)| \leq \int_\gamma g, \quad \text{for } p\text{-almost every } \gamma,$$

and let $\tilde{f}$ be such that $m(\{f \neq \tilde{f}\}) = 0$. It is sufficient to prove that, given $t \in [0,1]$,

$$\tilde{f}(\gamma_t) = f(\gamma_t), \quad \text{for } p\text{-almost every } \gamma,$$

that is, for every p-test plan π,

$$\tilde{f}(\gamma_t) = f(\gamma_t), \quad \text{for } \pi\text{-almost every } \gamma.$$

This is a direct consequence of the fact that $(e_t)_\#\pi \ll m$ and $f = \tilde{f}$ m-almost everywhere. The proof of the second statement is analogous. $\square$

As we did for the H-space, let us try to find a "minimal" upper gradient for a given function $f \in BL^{1,p}(X,d,m)$. This will come from the following closure property of $WUG(f)$.

PROPOSITION 4.2. *Let $f_n \in BL^{1,p}$, $f \in L^p$ be such that f_n converge to f strongly in L^p and let $g_n \in WUG(f_n)$ be such that g_n converge to g weakly in L^p. Then $f \in BL^{1,p}$ and $g \in WUG(f)$.*

PROOF. By Mazur's Lemma and convexity of $BL^{1,p}$, we reduce ourselves to the case where $\|g_n - g\|_{L^p} \to 0$. Moreover, up to subsequences, we assume that

$$\sum_n (\|f_n - f\|_{L^p} + \|g_n - g\|_{L^p}) < \infty.$$

Fix a p-test plan π with $\mathscr{A}_q(\gamma) \in L^\infty(\pi)$. Then

$$|f_n(\gamma_1) - f_n(\gamma_0)| \leq \int_\gamma g_n, \quad \text{for } \pi\text{-almost every } \gamma.$$

The idea is to pass to the limit in the inequality above. To do this, let us show that for π-almost every γ we have $\int_\gamma g_n \to \int_\gamma g$. It is sufficient to prove that

$$\int \left(\int_\gamma \sum |g_n - g| \right) d\pi(\gamma) < \infty.$$

By Hölder inequality,

$$\int \int_0^1 \sum |g_n - g|(\gamma_t)|\gamma'|(t) dt \, d\pi(\gamma) \le \left(\int \int_0^1 \sum |g_n - g|^p(\gamma_t) dt \, d\pi(\gamma) \right)^{1/p} \left(\int \mathscr{A}_q(\gamma) d\pi(\gamma) \right)^{1/q}.$$

Thanks to the compression constant and the subadditivity of L^p-norm, we can bound from above the right-hand side by

$$C(\pi)^{1/p} \sum \|g_n - g\|_{L^p} \left(\int \mathscr{A}_q(\gamma) d\pi(\gamma) \right)^{1/q},$$

which gives the required property.

A similar and simpler argument gives, for all $t \in [0,1]$, $f_n(\gamma_t) \to f(\gamma_t)$ for π-a.e. γ. □

Remark 4.1. An analogous and stronger statement holds in Fuglede's theory: if $g_n \to g$ in L^p then $\int_g g_n \to \int_\gamma g$, for Mod_p-almost every γ. Using this fact, arguing as in Proposition 4.3 below, one can prove that $H^{1,p} \subset N^{1,p}$.

By the proposition above, since $WUG(f)$ is convex, it is also weakly closed and hence admits a unique element of minimal norm, denoted $|\nabla f|_{BL}$. In the end, we will prove not only that H-space and BL-space coincide, but also that their respective gradients coincide.

Proposition 4.3. *The inclusion $H^{1,p}(X,d,m) \subset BL^{1,p}(X,d,m)$ and the inequality*

$$|\nabla f|_{BL} \le |\nabla f|_* \quad m\text{-almost everywhere in } X \tag{26}$$

hold.

Proof. Let $f \in H^{1,p}$. By property (i) in Proposition 2.1, there exist $f_n \in \mathrm{Lip}_b(X)$ such that $\mathrm{lip}_a(f_n) \to |\nabla f|_*$ in L^p. Since $|\nabla f_n|_*$ is an upper gradient for f_n, $\mathrm{lip}_a(f_n) \in WUG(f_n)$ and thus by the closure property $|\nabla f|_* \in WUG(f)$. Finally, to get (26) we need the pointwise minimality property

$$|\nabla f|_{BL} \le g, \quad \forall g \in WUG(f),$$

which can be proved as the analogous one for the relaxed gradient (see the proof of property (ii) in Proposition 2.1). □

5. Equivalence between H-space and BL-space

Our final program is

- to reverse the inclusion and the inequality above,
- to introduce a $W^{1,p}$ Sobolev space and prove the chain of inclusions $H^{1,p} \subset W^{1,p} \subset BL^{1,p}$.

A posteriori, these two facts will give equivalence between the three different notions. Moreover, since also the Newtonian space $N^{1,2}$ is contained in the Beppo-Levi space (Proposition 3.1) and contains the $H^{1,2}$ space (by Remark 4.1), all 4 spaces coincide.

Since the proofs are not elementary, from now on we will only deal with the case $p = 2$.

The inclusion $BL^{1,p} \subset H^{1,p}$ and the corresponding inequality $|\nabla f|_* \le |\nabla f|_{BL}$ between weak gradients are more difficult and require some tools, such as the Hopf–Lax formula and the superposition principle. Before going into the details, let us outline the strategy. The basic idea is to look at the energy dissipation rate of the entropy along the heat flow. Indeed, in our general metric context the heat flow plays the role convolution does in the Euclidean case.

THEOREM 5.1. *$BL^{1,2}(X,d,m) \subset H^{1,2}(X,d,m)$ and, for every $g \in BL^{1,2}(X,d,m)$,*

$$\int_X |\nabla g|_*^2 \, dm \le \int_X |\nabla g|_{BL}^2 \, dm.$$

SKETCH OF THE PROOF. Let $g \in BL^{1,2}(X,d,m)$. To show $g \in H^{1,2}(X,d,m)$ we must show that g can be approximated by Lipschitz functions in an optimal way.

Step 1. By approximation and truncation we can reduce to the case where $0 < c \le g \le C < \infty$ and, by linearity, to the case where $\int_X g^2 dm = 1$.

Step 2. Consider the gradient flow of $\mathscr{C}$ starting at $\bar f = g^2$. Since $\bar f \in L^1$ and, since g is bounded, $\bar f \in L^2$ so that $\bar f$ can be taken as initial condition for the heat flow associated with the Cheeger energy

$$\begin{cases} \frac{d}{dt} f_t = \Delta f_t, \\ \lim\limits_{t\to 0^+} f_t = \bar f. \end{cases}$$

Here it is crucial that the problem above has a solution for any $\bar f \in L^2$ as we do not know if $\bar f$ is in the domain of $\mathscr{C}$. Thanks to the regularizing effects (see Section 5), we have $f_t \in \operatorname{dom}\mathscr{C}$ for every $t > 0$ and we are going to prove the same property for $t = 0$.

Step 3. The main point is to estimate the energy dissipation rate of the entropy $\int_X f_t \log f_t \, dm$ in a Eulerian and in a Lagrangian way, respectively, as follows

$$-\frac{d}{dt}\int_X f_t \log f_t \, dm = \int_X \frac{|\nabla f_t|_*^2}{f_t} \, dm, \text{ for a.e. } t, \tag{27}$$

$$-\frac{d}{dt}\int_X f_t \log f_t \, dm \le \frac{1}{2}\int_X \frac{|\nabla f_t|_*^2}{f_t} dm + \frac{1}{2}\int_X \frac{|\nabla f_t|_{BL}^2}{f_t} \, dm. \tag{28}$$

Note that since $|\nabla f_t|_{BL} \le |\nabla f_t|_*$, (28) is sharper than (27). Hence, by comparing the two we get that

$$\int_X \frac{|\nabla f_t|_*^2}{f_t} \, dm = \int_X \frac{|\nabla f_t|_{BL}^2}{f_t} \, dm \text{ for almost every } t.$$

Changing variables in the equality above we deduce

$$4\int_X |\nabla \sqrt{f_t}|_*^2 \, dm = 4\int_X |\nabla \sqrt{f}_t|_{BL}^2 \, dm.$$

Step 4. Recall that $f_t \to \bar f = g^2$ in L^2 as $t \to 0$, whence $\sqrt{f_t} \to g$ as $t \to 0$. The heuristic idea (see Section 5.5) is that passing to limit as $t \to 0^+$ in the previous identity we get $g \in H^{1,2}$ and the inequality between weak gradients. Of course the construction of the approximation sequence by Lipschitz functions for g is indirect, and encoded in the definition of the H-Sobolev space. □

Let us prove (27), whereas we postpone to Section 5.5 the proof of (28), which requires Hopf–Lax formula and superposition principle, and the proof of the final step. To apply the PDE technique outlined above we will need some additional properties of gradient flows.

- If the initial condition $\bar f$ satisfies $c \le \bar f \le C$ m-almost everywhere then also the solution of the heat flow satisfies $c \le f_t \le C$ m-almost everywhere.
- For every initial condition $\bar f \in L^2(X,m)$ one has the mass conservation property $\int_X f_t\,dm = \int_X \bar f\,dm$. Indeed
$$\frac{d}{dt}\int_X f_t\,dm = \int_X \Delta f_t\,dm = 0$$
thanks to the integration by parts formula with $g \equiv 1$.

Proof of (27). By the mass conservation property, $\int f_t\,dm = 1$. Hence, using the heat flow and the integration by parts formula,

$$\begin{aligned} -\frac{d}{dt}\int f_t\log f_t\,dm &= -\frac{d}{dt}\int (f_t\log f_t - f_t)\,dm = -\int \log f_t \frac{d}{dt} f_t\,dm = \\ &= -\int \log f_t \Delta f_t\,dm = \int |\nabla \log f_t|_* |\nabla f_t|_*\,dm = \int \frac{|\nabla f_t|_*^2}{f_t}\,dm. \end{aligned}$$

□

5.1. Hopf–Lax formula. The general setting in which all the statements concerning Hopf–Lax formula hold true is the following: d extended distance on X, $f : X \to \mathbb{R}\cup\{+\infty\}$ lower semicontinuous function with a quadratic growth from below. Here we will only deal with the simpler context of a compact metric space (X,d) and $f \in \mathrm{Lip}_b(X)$.

The *Hopf–Lax semigroup* is defined as

$$\begin{cases} Q_0 f(x) = f(x) \\ Q_t f(x) = \min_{y\in X} F(t,x,y), \end{cases}$$

where $F(t,x,y) = f(y) + \frac{1}{2t}d^2(x,y)$. Using the triangle inequality, it is easy to check that

$$Q_t \circ Q_s \ge Q_{s+t},$$

and equality holds if (X,d) is a length space. This somehow motivates the name of semigroup.

The goal is to study pointwise properties of the Hopf-Lax semigroup. The classical theory in the Euclidean framework $X = \mathbb{R}^n$ tells us that $Q_t f$ is the unique solution in the viscosity sense to the Hamilton–Jacobi equations

$$\begin{cases} \partial_t u + \frac{1}{2}|\nabla u|^2 = 0, \text{ in } \mathbb{R}^n \\ u(x,0) = f(x). \end{cases} \tag{29}$$

Here we do not need viscosity theory, which in a metric framework is quite problematic. Instead, we only look for the following pointwise subsolution properties

$$\begin{aligned} \partial_t Q_t f(x) + \frac{1}{2}|\nabla Q_t f|^2(x) &\leq 0, \\ \partial_t Q_t f(x) + \frac{1}{2}\mathrm{lip}_a^2(Q_t f, x) &\leq 0. \end{aligned}$$

The first inequality above appeared at the same time in the independent works [**15**] and [**35**], whereas the second one (which is stronger) is proved in [**15**].

We first introduce some notation. Given $t > 0$, set

$$\begin{aligned} \mathcal{A}(x,t) &= \{y \in X \mid y \text{ minimizes } F(t,x,\cdot)\}, \\ D^+(x,t) &= \max\{d(x,y) \mid y \in \mathcal{A}(x,t)\}, \\ D^-(x,t) &= \min\{d(x,y) \mid y \in \mathcal{A}(x,t)\}, \end{aligned}$$

and, by convention, $D^\pm(x,0) = 0$.

The compactness assumption ensures $\mathcal{A}(x,t) \neq \emptyset$. Moreover, since X is compact and $f \in \mathrm{Lip}_b(X)$, we have the following properties. The function $(x,t) \mapsto Q_t f(x)$ is continuous on $X \times [0,\infty)$, $Q_t f \in \mathrm{Lip}(X)$ for all $t \geq 0$, and $D^-(x,t) \leq D^+(x,t) \leq \mathrm{lip}(f)$. In particular, the multifunction $(x,t) \mapsto \mathcal{A}(x,t)$ is upper semicontinuous, i.e., if $y_n \in \mathcal{A}(x_n,t_n)$, $(x_n,t_n) \to (x,t)$ and $y_n \to y$ then $y \in \mathcal{A}(x,t)$ (limits of minimizers are minimizers).

Lemma 5.2. *The maps $t \mapsto D^\pm(t,x)$ are non decreasing and coincide out of a countable set. Moreover, $(x,t) \mapsto D^+(x,t)$ is upper semicontinuous and $(x,t) \mapsto D^-(x,t)$ is lower semicontinuous.*

Proof. We are going to use the following fact (easy to check). Assume $g^\pm : (0,\infty) \to \mathbb{R}_+$ are two nondecreasing functions such that $g^+(t) \geq g^-(t)$ for every $t \geq 0$ and $g^+(s) \leq g^-(t)$ for every $s < t$. Then $g^\pm$ have the same jump set and they coincide outside of their jump set.

In order to apply the fact above to $D^\pm$ we need to check that for every (x,t), (x,s) such that $s < t$ one has $D^+(x,s) \leq D^-(x,t)$. Let x_s, $x_t \in X$ be minimizers of $F(s,x,\cdot)$, $F(t,x,\cdot)$, respectively. Then

$$\begin{aligned} f(x_t) + \frac{1}{2t}d^2(x,x_t) &\leq f(x_s) + \frac{1}{2t}d^2(x,x_s), \\ f(x_s) + \frac{1}{2s}d^2(x,x_s) &\leq f(x_t) + \frac{1}{2s}d^2(x,x_t). \end{aligned}$$

Summing up we get

$$\left(\frac{1}{s} - \frac{1}{t}\right)\left(d^2(x_t,x) - d^2(x_s,x)\right) \geq 0,$$

whence $d^2(x_t,x) - d^2(x_s,x) \geq 0$. Now, choosing x_t, x_s such that $d(x_t,x) = D^-(x,t)$, $d(x_s,x) = D^+(x,s)$, respectively, we conclude $D^+(x,s) \leq D^-(x,t)$.

As concerns semicontinuity properties, we only prove the first statement, the proof of the second one being analogous. Let $(x_n,t_n) \to (x,t)$. Choose a subsequence such that

$$\limsup_{n\to\infty} D^+(x_n,t_n) = \lim_{k\to\infty} D^+(x_{n_k},t_{n_k}).$$

Let $y_{n_k} \in \mathcal{A}(x_{n_k}, t_{n_k})$ such that $D^+(x_{n_k}, t_{n_k}) = d(x_{n_k}, y_{n_k})$. Up to subsequences, assume $y_{n_k} \to y$. Then $y \in \mathcal{A}(x,t)$, whence

$$D^+(x,t) \geq d(x,y) = \lim_{k\to\infty} d(x_{n_k}, y_{n_k}) = \lim_{k\to\infty} D^+(x_{n_k}, t_{n_k}) = \limsup_{n\to\infty} D^+(x_n, t_n).$$

□

THEOREM 5.3 (Hamilton–Jacobi subsolution). *Let $x \in X$. Then, out of a countable subset of* $[0,\infty)$, *$\partial_t Q_t f(x)$ exists and*

$$\partial_t Q_t f(x) + \frac{1}{2}\mathrm{lip}_a^2(Q_t f, x) \leq 0, \tag{30}$$

and equality holds if (X,d) is a length space.

PROOF. Fix $\bar{x} \in X$. It is sufficient to prove that

$$\partial_t Q_t f(\bar{x}) = -\frac{1}{2t^2}(D^+(\bar{x},t))^2, \quad \text{with at most countably many exceptions,} \tag{31}$$

$$\mathrm{lip}_a^2(Q_t f, \bar{x}) \leq \frac{(D^+(\bar{x},t))^2}{t^2}, \quad \text{for all } t > 0. \tag{32}$$

The proofs of (31) and (32) are similar. Here we only show (32). To do this, let $x, x' \in B_r(\bar{x})$, $y' \in \mathcal{A}(x', t)$. Then

$$\begin{aligned}
Q_t f(x) - Q_t f(x') &\leq f(y') + \frac{1}{2t}d^2(x,y') - f(y') - \frac{1}{2t}d^2(x',y') = \frac{1}{2t}(d^2(x,y') - d^2(x',y')) \\
&\leq \frac{1}{2t}d(x,x')(d(x,y') + d(x',y')) \leq \frac{1}{2t}d(x,x')(d(x,x') + 2d(x',y')) \\
&\leq \frac{1}{2t}d(x,x')(2r + 2D^+(x',t)).
\end{aligned}$$

Therefore,

$$\frac{Q_t f(x) - Q_t f(x')}{d(x,x')} \leq \frac{1}{t}\left(r + \sup_{z\in B_r(\bar{x})} D^+(z,t)\right),$$

which implies

$$\mathrm{lip}(Q_t, f, B_r(\bar{x})) \leq \frac{1}{t}\left(r + \sup_{z\in B_r(\bar{x})} D^+(z,t)\right).$$

Letting $r \to 0$ and by upper semicontinuity of D^+ we get

$$\mathrm{lip}_a(Q_t f, \bar{x}) \leq \frac{1}{t}D^+(\bar{x},t),$$

which gives (32). □

5.2. Tools from optimal transportation. In this section we give a quick introduction to optimal transportation which will be useful in the sequel.

Let (X,d) be a complete and separable metric space. Set

$$\mathcal{P}_2(X) = \left\{\mu \in \mathcal{P}(X) \mid \forall \bar{x} \in X,\ \int_X d^2(x,\bar{x})d\mu(x) < \infty\right\},$$

i.e., the set of probability measures having finite moments (notice that by the finiteness of μ, the integrability condition needs to be checked for one $\bar{x}$ only). Given μ, $\nu \in \mathcal{P}_2(X)$, the optimal transportation problem [**50**] is to find a transportation map $T : X \to X$ realizing

$$\min \left\{ \int_X d^2(x, T(x)) d\mu(x) \mid T_\# \mu = \nu \right\}. \tag{M}$$

Although in many cases this problem has a solution, it is better to consider the more symmetric and flexible formulation due to Kantorovich [**42**].

A probability measure $\eta \in \mathcal{P}(X \times X)$ is a transport plan for μ, ν if $\eta(A \times X) = \mu(A)$ and $\eta(X \times B) = \nu(B)$, or equivalently, $(\pi^1)_\# \eta = \mu$, $(\pi^2)_\# \eta = \nu$, where π^1, π^2 are the projections. Heuristically, $\eta(A \times B)$ is the amount of mass in A sent to B. The Kantorovich problem is to find a transportation plan realizing

$$\min \left\{ \int_{X \times X} d^2(x, y) d\eta(x, y) \mid (\pi^1)_\# \eta = \mu, (\pi^2)_\# \eta = \nu \right\}. \tag{K}$$

Transport maps induce transport plans through the identity $\eta = (\mathbb{I}_X \times T)_\# \mu$, in which case $\eta(A \times B) = \mu(A \cap T^{-1}(B))$. In general, using transport plans is a less constrained way to move mass than transport maps, since we are allowed to split mass.

To better understand how this theory has to do with moving points, let us give a formulation of the transport problem in the geodesic case. Let (X, d) be a geodesic space. Recall the action functional $\mathscr{A}_2(\gamma) = \int_0^1 |\dot{\gamma}|^2(t) dt$. Define $\mathrm{Geo}(X)$ as the set of curves $\gamma : [0, 1] \to X$ that are constant speed geodesics. To characterize length-minimizers in terms of the action, in general one has

$$\mathscr{A}_2(\gamma) \geq \big(\mathscr{A}_1(\gamma)\big)^2 \geq d^2(\gamma_1, \gamma_0),$$

with the first equality true if and only if the speed is constant, the second one true if and only if γ is length-minimizing. Hence γ is length-minimizing and parameterized by constant speed if and only if $\mathscr{A}_2(\gamma) = d^2(\gamma_1, \gamma_0)$. Hence $\mathrm{Geo}(X) = AC([0, 1], X) \cap \{\gamma \mid \mathscr{A}_2(\gamma) = d^2(\gamma_1, \gamma_0)\}$. Using the action, another formulation of the optimal transport problem is

$$\min \left\{ \int \mathscr{A}_2(\gamma)\, d\Sigma(\gamma) \mid \Sigma \in \mathcal{P}(\mathcal{C}([0, 1], X)), (e_0)_\# \Sigma = \mu, (e_1)_\# \Sigma = \nu \right\} \tag{D}$$

As for transport maps, a dynamic transport plan Σ from μ to ν induces a transport plan η through the identity $\eta = (e_0, e_1)_\# \Sigma$.

In geodesic spaces, the dynamic formulation (D) is equivalent yet somehow richer than (K), as it provides the right measure to consider on the space of curves: having chosen the initial an final points according to Kantorovich's formulation, one has to move following length-minimizing curves at a constant speed.

Theorem 5.4. *Assume (X, d) is a geodesic space. Then (K) and (D) are equivalent. Moreover, Σ is optimal for (D) if and only if $\eta = (e_0, e_1)_\# \Sigma$ is optimal for (K) and Σ is concentrated on* $\mathrm{Geo}(X)$.

Definition 5.1. Given $\mu, \nu \in \mathcal{P}_2(X)$, the *Wasserstein distance* $W_2(\mu, \nu)$ is defined as

$$W_2^2(\mu, \nu) := \min \left\{ \int_{X \times X} d^2(x, y)\, d\eta(x, y) \mid (\pi^1)_\# \eta = \mu, (\pi^2)_\# \eta = \nu \right\}.$$

It turns out that W_2 is indeed a distance in $\mathcal{P}_2(X)$ and that properties of (X,d) lift to properties of $(\mathcal{P}_2(X), W_2)$. For instance $(\mathcal{P}_2(X), W_2)$ is compact, complete, separable when so is (X,d).

THEOREM 5.5 (Dual formulation of the optimal transport problem).

$$\frac{1}{2}W_2^2(\mu,\nu) = \sup_{\varphi,\,\psi\in \mathrm{Lip}_b(X)} \left\{ -\int_X \varphi\, d\mu + \int_X \psi\, d\nu \mid \psi(y)-\varphi(x) \le \frac{1}{2}d^2(x,y)\right\}. \tag{33}$$

PROOF. We just prove the easy inequality $\ge$. Let η be a transport plan from μ to ν and let φ, $\psi \in \mathrm{Lip}_b(X)$ be such that $\frac{1}{2}d^2(x,y) \ge \psi(y)-\varphi(x)$. Integrating with respect to η,

$$\frac{1}{2}\int_{X\times X} d^2(x,y)d\eta(x,y) \ge \int_{X\times X} (\psi(y)-\varphi(x))\, d\eta(x,y) = \int_X \psi d\nu - \int_X \varphi\, d\mu,$$

where the last equality follows by μ, ν being the marginals of η. This gives the inequality

$$\frac{1}{2}W_2^2(\mu,\nu) \ge \sup_{\varphi,\,\psi\in \mathrm{Lip}_b(X)} \left\{ -\int_X \varphi\, d\mu + \int_X \psi\, d\nu \mid \psi(y)-\varphi(x) \le \frac{1}{2}d^2(x,y)\right\}.$$

□

Let us clarify the connection between optimal transportation and the Hopf–Lax formula. Fix $\varphi \in \mathrm{Lip}_b(X)$. Then the largest possible $\psi \in \mathrm{Lip}_b(X)$ compatible with the constraint $\psi(y)-\varphi(x) \le \frac{1}{2}d^2(x,y)$ is precisely $Q_1\varphi$, that is,

$$\sup_{\varphi,\,\psi\in \mathrm{Lip}_b(X)} \left\{ -\int_X \varphi\, d\mu + \int_X \psi\, d\nu \mid \psi(y)-\varphi(x) \le \frac{1}{2}d^2(x,y)\right\} = $$
$$= \sup_{\varphi\in \mathrm{Lip}_b(X)} \left\{ -\int_X \varphi\, d\mu + \int_X Q_1\varphi\, d\nu\right\}. \tag{34}$$

Moreover, an interpolation argument allows to prove that, given an optimal φ, one computes the Wasserstein distance along a geodesic $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(X))$ from μ to ν by

$$\frac{1}{2t}W_2^2(\mu,\mu_t) = -\int_X \varphi d\mu + \int_X Q_t\varphi\, d\mu_t.$$

As a consequence, we get a differential description for geodesics of W_2 as

$$\begin{cases} \partial_t \mu_t + \mathrm{div}\,(V_t\mu_t) = 0, \\ V_t = \nabla\varphi_t, \\ \partial_t\varphi_t + \frac{1}{2}|\nabla\varphi_t|^2 = 0. \end{cases}$$

5.3. Kuwada Lemma. A key connection between the Lagrangian and Eulerian viewpoints, which will be useful to deduce (28), was discovered by Kuwada in [**44**], in his investigation of the relations between the curvature dimension condition and the contractivity of the Wasserstein distance under the heat flow. More precisely, let $f \ge 0$ be such that $\int_X f\, dm = 1$ and denote by $P_t f$ the heat flow starting at f. If one has the mass conservation property for the heat flow (this is surely the case in these notes, since we are assuming that m is finite), one identifies $P_t f$ with the measure $(P_t f)m$ and it makes sense to consider the heat flow on the space of probability measures instead of the space of functions. Then, Kuwada proved in [**44**] that, in a quite general context, the gradient contractivity condition

$$|\nabla P_t f|^2 \le e^{-2Kt}|\nabla f|^2, \tag{CD(K,∞)}$$

is equivalent to the Wasserstein contractivity property

$$W_2((P_t f)m,(P_t g)m) \le e^{-Kt} W_2(fm, gm),\ \forall f,\, g \ge 0, \int_X f\,dm = \int_X g\,dm = 1. \tag{35}$$

Finding a suitable version of the equivalence above in the sub-Riemannian framework is an open problem.

LEMMA 5.6 (Kuwada Lemma). *Let $f_0 \in L^2(X,m)$ be such that $0 < c \le f_0 \le C < \infty$ and $\int_X f_0\,dm = 1$. Let f_t be the gradient flow of $\mathscr{C}$ starting at f_0 and set $\mu_t = f_t m$. Then $\mu_t \in AC^2_{loc}((0,\infty),\mathcal{P}_2(X))$ and*

$$|\dot{\mu}_t|^2 \le \int_X \frac{|\nabla f_t|_*^2}{f_t}\,dm, \ \textit{for almost every } t > 0. \tag{36}$$

PROOF. Fix $t < s$ and set $l = s - t$. Our goal is to estimate $W_2^2(\mu_s,\mu_t)$ from above and eventually let $s \to t$. To do this, we fix $\varphi \in \mathrm{Lip}_b(X)$ and we look for an estimate from above of the quantity $-\int \varphi d\mu_t + \int Q_1\varphi d\mu_s$. By the dual formula in Theorem 5.5, this provides an estimate on $W_2^2(\mu_s,\mu_t)$. Then, by interpolation,

$$\begin{aligned}
-\int_X \varphi\,d\mu_t + \int_X Q_1\varphi\,d\mu_s &= \int_0^l \frac{d}{dr}\left(\int_X Q_{r/l}\varphi d\mu_{t+r}\right)dr \\
&= \int_0^l \frac{d}{dr}\left(\int_X Q_{r/l}\varphi f_{t+r}\,dm\right)dr \\
&\le \int_0^l \int_X \left(-\frac{1}{2l}\mathrm{lip}_a^2(Q_{r/l}\varphi) f_{t+r} + Q_{r/l}\varphi \Delta f_{t+r}\right)dm\,dr,
\end{aligned}$$

where the last inequality follows by Leibniz rule and the subsolution property[2] of Q_t given in Theorem 5.3. Applying the integration by parts formula (see Proposition 4.2), we deduce

$$\begin{aligned}
-\int_X \varphi\,d\mu_t + \int_X Q_1\varphi\,d\mu_s &\le \int_0^l \int_X \left(-\frac{1}{2l}\mathrm{lip}_a^2(Q_{r/l}\varphi) f_{t+r} + |\nabla Q_{r/l}\varphi|_* |\nabla f_{t+r}|_*\right)dm\,dr \\
&\le \int_0^l \int_X \left(-\frac{1}{2l}\mathrm{lip}_a^2(Q_{r/l}\varphi) f_{t+r} + \mathrm{lip}_a(Q_{r/l}\varphi)|\nabla f_{t+r}|_*\right)dm\,dr,
\end{aligned}$$

where the last inequality is a consequence of the very definition of $|\nabla Q_{r/l}\varphi|_*$ and by the fact that $Q_{r/l}\varphi$ is Lipschitz. Apply the Young inequality $-\frac{1}{2l}a^2 \le ab + \frac{l}{2}b^2$ with $a = \mathrm{lip}_a(Q_{r/l}\varphi)\sqrt{f_{t+r}}$ and $b = |\nabla f_{t+r}|_*/\sqrt{f_{t+r}}$, then

$$\int_0^l \int_X \left(-\frac{1}{2l}\mathrm{lip}_a^2(Q_{r/l}\varphi) f_{t+r} + \mathrm{lip}_a(Q_{r/l}\varphi)|\nabla f_{t+r}|_*\right)dm\,dr \le \frac{l}{2}\int_0^l \int_X \frac{|\nabla f_{t+r}|_*^2}{f_{t+r}}\,dm\,dr.$$

[2]Since we are integrating with respect to time, by Fubini's theorem the subsolution property holds m-almost everywhere for almost every t.

Thus, taking the supremum w.r.t. φ, we obtain

$$W_2^2(\mu_s,\mu_t) \le (s-t)\int_t^s \int_X \frac{|\nabla f_r|_*^2}{f_r}\,dm\,dr,$$

and we conclude $\mu_t \in AC^2_{loc}((0,\infty),\mathcal{P}_2(X))$. Finally, dividing by $(s-t)^2$, at Lebesgue points of the map $r \mapsto \int \frac{|\nabla f_r|_*^2}{f_r} dm$ we deduce (36). □

5.4. Superposition principle. Another important connection between curves in $\mathcal{P}_2(X)$ and curves in the metric space X is given by the superposition principle. The origin of this idea goes back to the work of L.C. Young [**57**], see for instance the nice treatment of this topic in [**23**]. More recently it has been adapted to other situations. For instance, Smirnov [**54**] provides an instance of superposition principle for 1-currents, saying essentially that any normal current can be seen as a kind of superposition of currents associated with elementary curves. Another version of this principle for solutions to the continuity equation can be found in the book [**10**]. Here we will use a version of this principle in a metric framework which is due to Lisini, see [**46**].

Let us first give the superposition principle in a classical context, say in $\mathbb{R}^n$. Let b be a smooth vector field that is globally Lipschitz, so that we have global existence of the associated flow. Then we can relate the continuity equation

$$\frac{d}{dt}u_t + \mathrm{div}\,(b_t u_t) = 0,\ \ u_t \ge 0, \tag{37}$$

to the flow

$$\frac{d}{dt}X(t,x) = b_t(X(t,x)), \tag{38}$$

associated with b, by the fact that, for every initial condition $\bar u \in L^1$, the unique distributional solution to (37) satisfying $u_0 = \bar u$ is precisely $u_t = X(t,\cdot)_\# \bar u$ (the verification that this formula provides a solution is elementary, less elementary is the uniqueness part, see [**10**]).

Theorem 5.7 ([**10**]). *Let $\mu_t \in \mathcal{P}_2(\mathbb{R}^n)$ be a solution to*

$$\frac{d}{dt}\mu_t + \mathrm{div}\,(V_t\mu_t) = 0, \tag{39}$$

where $t \mapsto \|V_t\|_{L^2(\mu_t;\mathbb{R}^n)} \in L^1(0,1)$. Then $\mu_t \in AC^2([0,1],\mathcal{P}_2(\mathbb{R}^n))$ and $|\dot\mu_t| \le \|V_t\|^2_{L^2(\mu_t)}$ for almost every $t>0$. Conversely, for any $\mu_t \in AC^2([0,1],\mathcal{P}_2(\mathbb{R}^n))$ there exists V_t with $\|V_t\|_{L^2(\mu_t;\mathbb{R}^n)} \in L^1(0,1)$ for which (39) holds and $|\dot\mu_t| = \|V_t\|^2_{L^2(\mu_t)}$ for almost every $t>0$.

The superposition principle in $\mathbb{R}^n$ says that nonnegative measure-valued solutions of (39) can always be represented as time marginals of a suitable $\pi \in \mathcal{P}(AC^2([0,1],\mathbb{R}^n))$ concentrated on solutions to the ODE, i.e. $(e_t)_\#\pi = \mu_t$ for every $t \in [0,1]$. Going to the metric framework, two main difficulties arise, namely to give a meaning to the continuity equation and to the ODE associated with a vector field. Even though we will see that this is possible, we have seen that in $\mathbb{R}^n$ solutions to (39) can be identified with 2-absolutely continuous curves with values in $\mathcal{P}(X)$. With this replacement, the following result holds, see [**46**].

Theorem 5.8 (Metric superposition principle). *For all* $\mu_t \in AC^2([0,1],\mathcal{P}_2(X))$ *there exists* $\pi \in \mathcal{P}(AC^2([0,1],X))$ *such that*

(1) $(e_t)_\#\pi = \mu_t$ *for every* $t \in [0,1]$*;*
(2) $|\dot\mu_t|^2 = \int |\dot\gamma_t|^2\, d\pi(\gamma)$ *for almost every* t*.*

Any probability measure π satisfying 2 is called a *lifting* of μ_t. Although properties 1 and 2 above do not provide uniqueness, property 2 characterizes the lifting in an optimal sense (if X is a vector space, this means that there is no cancellation of the velocities of different curves passing through the same point at the same time). This is illustrated by the next proposition.

Proposition 5.9. *Let* $\mu_t \in AC^2([0,1],\mathcal{P}_2(X))$ *and let* $\pi \in \mathcal{P}(AC^2([0,1],X))$ *be such that* $(e_t)_\#\pi = \mu_t$ *for every* $t \in [0,1]$*. Then*

$$\int_0^1 |\dot\mu_t|^2 dt \le \int_0^1 \int |\dot\gamma_t|^2 d\pi(\gamma)\, dt.$$

Proof. As for Kuwada Lemma, to estimate $|\dot\mu_t|^2$ we bound $W_2^2(\mu_s,\mu_t)$. Let $s > t$. Thanks to property 1, $\eta = (e_s, e_t)_\#\pi$ is a transport plan from μ_s to μ_t. Hence

$$W_2^2(\mu_s,\mu_t) \le \int d^2(\gamma_s,\gamma_t) d\pi(\gamma) \le \int \left(\int_t^s |\dot\gamma_r| dr\right)^2 d\pi(\gamma) \le (s-t)\int_t^s \left(\int |\dot\gamma_r|^2\, d\pi(\gamma)\right) dr,$$

the last inequality following by Hölder inequality and Fubini's theorem. Dividing by $(s-t)^2$ and taking the limit as $s \to t$, we get that at Lebesgue points of the map $r \mapsto \int |\dot\gamma_r|^2 d\pi(\gamma)$ there holds

$$|\dot\mu_t|^2 \le \int |\dot\gamma_t|^2 d\pi(\gamma).$$

□

Sketch of the proof of Theorem 5.8. After isometrical embedding, since all the statements are isometrically invariant, it is not restrictive to assume (X,d) geodesic. Thanks to Proposition 5.9, to get property 2 it suffices to prove the reverse inequality

$$\int_0^1 |\dot\mu_t|^2 dt \ge \int_0^1 \int |\dot\gamma_t|^2 d\pi(\gamma)\, dt. \tag{40}$$

The idea is to divide $[0,1]$ in n intervals and construct a family of plans $\pi_n \in \mathcal{P}(AC^2([0,1],X))$ such that, for every $j = 1,\dots,n$, $(e_{j/n})_\#\pi_n = \mu_{j/n}$ and

$$\int_0^1 \int |\dot\gamma_t|^2 d\pi_n(\gamma)\, dt \le \int_0^1 |\dot\mu_t|^2 dt.$$

The family of probability measures $\{\pi_n\}_n$ turns out to be tight and thus converges to some measure π satisfying $(e_t)_\#\pi = \mu_t$ and (40). Let us outline the construction of π_n.

For every pair $(\mu_{\frac{j}{n}}, \mu_{\frac{j+1}{n}})$ take an optimal plan for the optimal transportation problem from $\mu_{\frac{j}{n}}$ to $\mu_{\frac{j+1}{n}}$. By a standard gluing procedure, we build a $\Sigma \in \mathcal{P}(X^{n+1})$ such that, for every j, $(\pi^j,\pi^{j+1})_\#\Sigma$ is an optimal plan from $\mu_{\frac{j}{n}}$ to $\mu_{\frac{j+1}{n}}$ (here $(\pi^j,\pi^{j+1}) : X^{n+1} \to X^2$ denotes the projection). The plan π_n is built by pushing forward Σ through a map σ :

$X^{n+1} \to L^2((0,1),X)$ assigning to an $(n+1)$-tuple x of random points in X the piecewise constant map $\sigma_x(t) = x_j, t \in (j/n,(j+1)/n)$. □

5.5. The inclusion $BL^{1,p}(X,d,m) \subset H^{1,p}(X,d,m)$. We end this chapter by concluding the proof of Theorem 5.1 and thus providing equivalence between Beppo Levi space and H-space, in the case $p = 2$. We also assume for simplicity that (X,d) has finite diameter, so that $\mathcal{P}(X) = \mathcal{P}_2(X)$.

LEMMA 5.10 (Lagrangian estimate of entropy decay rate). *Let $g \in BL^{1,2}(X,d,m)$, with $0 < c \le g \le C < \infty$, $\int_X g^2 dm = 1$, let f_t be the solution to the gradient flow associated with Cheeger energy starting at $\bar f = g^2$. Then*

$$\int_X (\bar f \log \bar f - f_t \log f_t) dm \le \frac{1}{2}\int_0^t \left(\int_X \frac{|\nabla f_s|_*^2}{f_s}\, dm\right) ds + \frac{1}{2}\int_0^t \left(\int_X \frac{|\nabla \bar f|_{BL}^2}{\bar f^2} f_s\, dm\right) ds. \quad (41)$$

PROOF. By convexity of $z \mapsto z\log z$, one has

$$\int_X (\bar f \log \bar f - f_t \log f_t)\, dm \le \int_X \log \bar f (\bar f - f_t)\, dm = \int_X \log \bar f (f_0 - f_t)\, dm.$$

Set $\mu_t = f_t m$. Then by Theorem 5.6 $\mu_t \in AC^2_{loc}((0,\infty), \mathcal{P}_2(X))$. Thus we can apply Theorem 5.8 which gives an optimal lifting $\pi \in \mathcal{P}(AC^2([0,1],X))$. Moreover, since $c \le f_t \le C$ for every t, π is also a 2-test plan. Hence

$$\begin{aligned}\int_X \log \bar f (f_0 - f_t)\, dm &= \int \log(\bar f(\gamma_0) - \log \bar f(\gamma_t))\, d\pi(\gamma) \le \int \left(\int_0^t |\nabla \log \bar f|_{BL}(\gamma_s)|\dot\gamma_s| ds\right) d\pi(\gamma) \\ &\le \frac{1}{2}\int_0^t \left(\int |\nabla \log \bar f|_{BL}^2(\gamma_s)\, d\pi(\gamma)\right) ds + \frac{1}{2}\int_0^t \left(\int |\dot\gamma_s|^2\, d\pi(\gamma)\right) ds,\end{aligned}$$

where the equality comes from the lifting property of π, the first inequality from the fact that π is a 2-test plan, and the second is an instance of Young inequality. Eventually, by optimality of π (see property 2 in Theorem 5.8) and by Theorem 5.6

$$\int_0^t \left(\int |\dot\gamma_s|^2 d\pi(\gamma)\right) ds = \int_0^t |\dot\mu_s|^2 ds \le \int_0^t \left(\int_X \frac{|\nabla f_s|_*^2}{f_s} dm\right) ds$$

and by the chain rule,

$$\int |\nabla \log \bar f|_{BL}^2(\gamma_s)\, d\pi(\gamma) = \int_X \frac{|\nabla \bar f|_{BL}^2}{\bar f^2} f_s\, dm,$$

which concludes the proof. □

END OF THE PROOF OF THEOREM 5.1. Recall that, integrating (27) one gets

$$\int_X (\bar f \log \bar f - f_t \log f_t) dm = \int_0^t \left(\int_X \frac{|\nabla f_s|_*^2}{f_s}\, dm\right) ds.$$

Therefore, (41) implies

$$\frac{1}{t}\int_0^t \left(\int_X \frac{|\nabla f_s|_*^2}{f_s}\, dm\right) ds \le \frac{1}{t}\int_0^t \left(\int_X \frac{|\nabla \bar f|_{BL}^2}{\bar f^2} f_s\, dm\right) ds. \quad (42)$$

By definition of the gradient flow, $f_t \to \bar{f}$ as $t \to 0$ strongly in L^2, and since $\{f_t\}_{t>0}$ is bounded in L^∞, $f_t \to \bar{f}$ as $t \to 0$ weakly* in L^∞. Thus, the left-hand side of (42) satisfies

$$\lim_{t\to 0} \frac{1}{t} \int_0^t \left(\int_X \frac{|\nabla \bar{f}|_{BL}^2}{\bar{f}^2} f_s \, dm \right) ds = \int_X \frac{|\nabla \bar{f}|_{BL}^2}{\bar{f}} \, dm = 4 \int_X |\nabla \sqrt{\bar{f}}|_{BL}^2 \, dm = \int_X |\nabla g|_{BL}^2 \, dm.$$

As for the right-hand side of (42),

$$\begin{aligned} \liminf_{t\to 0} \frac{1}{t} \int_0^t \left(\int_X \frac{|\nabla f_s|_*^2}{f_s} \, dm \right) ds &= 4 \liminf_{t\to 0} \frac{1}{t} \int_0^t \left(\int_X \left|\nabla \sqrt{f_s}\right|_*^2 dm \right) ds \geq \\ &\geq 4 \liminf_{i\to\infty} \int_X \left|\nabla \sqrt{f_{t_i}}\right|_*^2 dm \\ &= 4 \liminf_{i\to\infty} \mathscr{C}(\sqrt{f_{t_i}}), \end{aligned}$$

where (t_i) is a sequence converging to 0 whose existence is provided by mean value theorem applied to $t \mapsto \int_0^t \left|\nabla \sqrt{f_s}\right|_*^2 ds$. By lower semicontinuity of $\mathscr{C}$,

$$\limsup_{t\to 0} \frac{1}{t} \int_0^t \left(\int_X \frac{|\nabla f_s|_*^2}{f_s} \, dm \right) ds \geq 4\mathscr{C}(g).$$

Finally, taking the limit as $t \to 0$ in (42), the inequalities above imply $g \in H^{1,2}(X, d, m)$ and

$$\int_X |\nabla g|_*^2 dm \leq \int_X |\nabla g|_{BL}^2 dm.$$

□

CHAPTER 4

Sobolev spaces via integration by parts

In this chapter we introduce the last notion of Sobolev space, denoted $W^{1,p}$, following the recent paper by Di Marino [**26**]. We will show the inclusions

$$H^{1,p} \subset W^{1,p} \subset BL^{1,p},$$

which will provide, thanks to the inclusion $BL^{1,p} \subset H^{1,p}$ of the previous chapter, the coincidence of the three spaces. The first inclusion will be quite easy, once one is aware of the right definitions.

1. Vector fields

In order to provide an integration by parts formula in metric measure spaces we need a notion of vector field. Recall that in differential geometry, vector fields can be seen either as sections of the tangent bundle or as derivations on the module of smooth functions. In particular the latter interpretation has been used independently by Weaver in [**55**] (as an alternative approach to [**24**]) to develop differential calculus in metric measure spaces. Since then, Weaver's point of view has been considered and related to Cheeger's one in [**34, 52, 26, 21**]. Other studies of vector fields in a metric context, based on Weaver's notions and on Γ-calculus, can be found in [**33**], where the author studies for the first time the notion of Sobolev space $H^{2,2}$ and provides a notion of parallel transport in metric measure spaces, and in [**20**] which is concerned with well-posedness of ordinary differential equations.

Definition 1.1. Denote by $L^0(X,m)$ the set of equivalence classes of m-measurable functions on a metric measure space (X,d,m). A *derivation* on (X,d,m) is a linear map $b : \mathrm{Lip}_b(X) \to L^0(X,m)$ with the following properties

(*i*) $b(fg) = b(f)g + fb(g)$ (Leibniz rule);

(*ii*) there exists $g : X \to [0,\infty)$ such that, for every $f \in \mathrm{Lip}_b(X)$,

$$|b(f)| \le g\ \mathrm{lip}_a(f),\ m\text{-almost everywhere,} \tag{43}$$

(weak locality property).

A minimal g satisfying (43) exists and will be denoted by $|b|$.

According to Definition 1.1, the set of derivations over X is an L^∞-module, i.e., one can multiply derivations by functions in $L^\infty(X,m)$.

Remark 1.1 (On the Leibniz and chain rule). The definition of derivation could be sharpened by requiring only (ii) and linearity, and deducing from this the Leibniz rule. To see this, one first proves the chain rule $b(\phi(f)) = \phi'(f)b(f)$ for any $\phi \in C^1$ in an interval containing the image of f; the chain rule can be proved by approximating ϕ by piecewise affine functions, choosing the discontinuity t_i points of ϕ' in such a way that $m(\{f = t_i\}) = 0$ and using the weak locality property. As soon as the chain rule is estabilished, the Leibniz rule for positive functions f, g follows by

$$b(\exp(\log(fg))) = fgb(\log(fg)) = fg(b(\log f) + b(\log g)) = fg\left(\frac{b(f)}{f} + \frac{b(g)}{g}\right)$$

and the general case can be achieved using the invariance under addition of constants.

DEFINITION 1.2. We say that a L^1-valued derivation b on X has a *divergence* if there exists a function in L^1, denoted by $\operatorname{div} b$, such that, for every $f \in \mathrm{Lip}_b(X)$,

$$\int_X f \operatorname{div} b \, dm = -\int_X b(f) \, dm.$$

For $1 < q < \infty$, set

$$\mathrm{Der}^q(X, m) := \{b \text{ derivation } \mid |b| \in L^q(X, m), \operatorname{div} b \in L^q(X, m)\}.$$

The proof of the following proposition is elementary.

PROPOSITION 1.1. $\mathrm{Der}^q(X, m)$ *is a* Lip_b*-module and a Banach space with the norm*

$$\|b\| := \||b|\|_q + \|\operatorname{div} b\|_q.$$

2. $W^{1,p}$-space and the inclusion $H^{1,p}(X, d, m) \subset W^{1,p}(X, d, m)$

We are now in a position to define Sobolev functions via integration by parts formula.

DEFINITION 2.1. Let $1 < p < \infty$ and let $q = p/(p-1)$ be, as usual, the dual exponent.

$$\begin{aligned} W^{1,p}(X, d, m) \quad := \quad & \Big\{ f \in L^p(X, m) \mid \exists\, L_f : \mathrm{Der}^q(X, m) \to L^1(X, m) \text{ linear}, \\ & L^q\text{-continuous, } \mathrm{Lip}_b\text{-linear and such that} \\ & \int_X f \operatorname{div} b \, dm = -\int_X L_f(b) \, dm, \ \forall\, b \in \mathrm{Der}^q(X, m) \Big\}. \end{aligned}$$

Let $f \in W^{1,p}(X, d, m)$. Then f satisfies the following properties (see [**26**] for detailed proofs).

a) Using Lip_b-linearity of L_f, one can show that whenever L_f exists it is unique, see [**26**, Remark 1.6].
b) By the definition of divergence, we have the inclusion $\mathrm{Lip}_b(X) \subset W^{1,p}(X, d, m)$, with $L_f(b) = b(f)$. (This will imply the inclusion $H^{1,p} \subset W^{1,p}$, as $H^{1,p}$ is the closure of smooth functions).
c) By the L^q-continuity of $b \mapsto b(f)$ there exists a function $g \in L^p(X, m)$ such that, for every $b \in \mathrm{Der}^q(X, m)$,

$$|L_f(b)| \le g|b|, \ m\text{-almost everywhere},$$

and the minimal g satisfying the inequality above is denoted by $|\nabla f|_W$. (This is dual to the definition of $|b|$).

In analogy with the smooth (i.e. bounded Lipschitz) case, we will also denote $L_f(b)$ by $df(b)$.

Remark 2.1. Let us point out the role of the distance in the definition of $W^{1,p}$: it is involved first of all because derivations are defined on d-Lipschitz functions and then because by duality with the asymptotic Lipscitz constant we can measure the "length" of vector fields (namely $|b|$) and thus $|\nabla f|_W$.

Remark 2.2 (Non-triviality of $\mathrm{Der}^q(X,m)$). Of course when there are no derivations on X, Definition 2.1 is empty. To see examples where $\mathrm{Der}^q(X,m) \neq \emptyset$, it suffices to consider Lemma 3.1 below, which provides existence of derivations as soon as we have p-test plans π with $\mathcal{A}_q \in L^1(\pi)$. More generally, as the proof of the lemma shows, we need only the existence of p-test plans with parametric barycenter in L^q; according to Theorem 3.2, it is sufficient that $\mathrm{Mod}_p(AC^q) > 0$.

Example 2.1. Let us consider the metric measure structure in $[0,1]$ of Example 2.1. Notice that in this case it is immediate to check that $\mathrm{Mod}_p(AC^q) = 0$, choosing weights vanishing on the rational numbers of $[0,1]$ where m is concentrated. On the other hand, even with extremely concentrated measures it is still possible to build nontrivial derivations. We illustrate the construction in the case when $m = \delta_0$, but building an absolutely convergent series of shifted derivations one can provide the construction in the general case. In the case when $m = \delta_0$ one can simply consider the linear functional

$$L(f) := f'(0) \qquad f \in C^1([0,1])$$

and the seminorm $p(f) := \mathrm{lip}_a(f)(0)$ to obtain via the analytic form of Hahn-Banach theorem an extension $\tilde{L}$ of L to $\mathrm{Lip}_b([0,1])$ satisfying $|\tilde{L}(f)| \leq p(f)$. According to Remark 1.1, this provides a (nontrivial) derivation. However, it is easily seen that the divergence of this derivation is not representable by a function.

Proposition 2.1. *$H^{1,p}(X,d,m) \subset W^{1,p}(X,d,m)$ and, for every $f \in H^{1,p}(X,d,m)$,*

$$|\nabla f|_W \leq |\nabla f|_*.$$

Proof. Let $f \in H^{1,p}$. By Proposition 2.1(i), there exist $f_n \in \mathrm{Lip}_b(X)$ such that f_n converge to f strongly in L^p and $\mathrm{lip}_a(f_n)$ converges strongly to $|\nabla f|_*$ in L^p. Let $b \in \mathrm{Der}^q(X,m)$. Again by definition,

$$\int_X f_n \,\mathrm{div}\, b \, dm = -\int_X b(f_n)\, dm,$$

whence

$$\left|\int_X f_n \,\mathrm{div}\, b\, dm\right| \leq \int_X |b| \,\mathrm{lip}_a(f_n)\, dm.$$

Taking the limit as n goes to ∞, since $f_n \to f$ in L^p, $\mathrm{lip}_a(f_n) \to |\nabla f|_*$ in L^p and $\mathrm{div}\, b, |b| \in L^q$ we have[1]

$$\lim_{n\to\infty} \int_X f_n \,\mathrm{div}\, b\, dm = \int_X f \,\mathrm{div}\, b\, dm,$$

$$\lim_{n\to\infty} \int_X |b| \,\mathrm{lip}_a(f_n)\, dm = \int_X |b|\, |\nabla f|_*\, dm.$$

[1]Note that computing the limit of $\int b(f_n)\, dm$ is not immediate. Indeed, we cannot pass to the limit in $|b(f_n) - b(f_m)| \leq |b|\mathrm{lip}_a(f_n - f_m)$ since we do not know that $\lim_{m,n} \mathrm{lip}_a(f_n - f_m) = 0$ in L^p.

Therefore, for every $b \in \mathrm{Der}^q(X,m)$,

$$\left|\int_X f \operatorname{div} b\, dm\right| \leq \int_X |b|\, |\nabla f|_*\, dm. \tag{44}$$

Using a version of the Hahn–Banach Theorem for Lip_b-modules, (44) provides a linear, L^q-continuous, Lip_b-linear operator $L_f : \mathrm{Der}^q(X,m) \to L^1(X,m)$ such that

$$\int_X f \operatorname{div} b\, dm = -\int_X L_f(b)\, dm, \ \forall\, b \in \mathrm{Der}^q(X,m),$$

and again (44) ensures

$$|\nabla f|_W \leq |\nabla f|_* \quad m\text{-almost everywhere.}$$

□

3. The inclusion $W^{1,p}(X,d,m) \subset BL^{1,p}(X,d,m)$

The inclusion will be a direct consequence of the fact that p-test plans induce canonically derivations.

LEMMA 3.1. *Let π be a p-test plan such that $\mathcal{A}_q \in L^1(\pi)$ and let ρ_π be such that* $\mathrm{Bar}(\pi) = \rho_\pi m$. *Then π induces canonically a derivation b_π such that $|b_\pi| \in L^q(X,m)$,* $\operatorname{div} b_\pi \in L^\infty(X,m)$ *and*

(i) $|b_\pi| \leq \rho_\pi$ *m-almost everywhere,*
(ii) $((e_0)_\#\pi - (e_1)_\#\pi) = \operatorname{div} b_\pi m$.

Recall that given a p-test plan π, the barycenter $\mathrm{Bar}(\pi)$ is a measure on X, see Section 3, defined by

$$\mathrm{Bar}(\pi)(E) = \int_{\mathcal{C}([0,1],X)} \int_{\gamma^{-1}(E)} |\dot\gamma|\, dt\, d\pi(\gamma) \qquad E \subset X \text{ Borel.}$$

Equivalently, passing from characteristic functions to general bounded Borel (or Borel non-negative) functions φ,

$$\int_X \varphi\, d\mathrm{Bar}(\pi) = \int_{\mathcal{C}([0,1],X)} \int_0^1 \varphi \circ \gamma |\dot\gamma|\, dt\, d\pi(\gamma).$$

When $\mathcal{A}_q \in L^1(\pi)$, the barycenter is absolutely continuous with respect to m with Radon–Nykodim derivative in L^q. Indeed, Hölder's inequality gives

$$\begin{aligned}\left|\int_X \varphi\, d\mathrm{Bar}(\pi)\right| &\leq \left(\int_0^1 \int_{\mathcal{C}([0,1],X)} |\varphi|^p(\gamma_t)\, d\pi(\gamma)\, dt\right)^{1/p} \left(\int_{\mathcal{C}([0,1],X)} \mathcal{A}_q(\gamma)\, d\pi(\gamma)\right)^{1/q} \\ &\leq (c(\pi))^{1/p} \left(\int_{\mathcal{C}([0,1],X)} \mathcal{A}_q(\gamma)\, d\pi(\gamma)\right)^{1/q} \|\varphi\|_p\end{aligned}$$

and the claim follows by Riesz theorem.

PROOF. The idea of the construction of b_π is once more based on the superposition principle: we average the "tangent" vector fields to the curves γ, w.r.t. the measure π. More specifically, for all $f \in \mathrm{Lip}_b(X)$ we define $b_\pi(f)$ as the function satisfying

$$\int_X b_\pi(f)\varphi\, dm = \int\int_0^1 \varphi\circ\gamma(t)\ (f\circ\gamma)'(t)\, dt\, d\pi(\gamma)$$

for any bounded Borel function φ. This definition is well posed thanks to the Radon-Nikodym theorem. One checks easily that b_π is linear and that it satisfies the Leibniz rule. Moreover, for every bounded Borel φ one has

$$\left|\int_X b_\pi(f)\varphi\, dm\right| \le \int\int_0^1 |\varphi|\circ\gamma\ |\nabla f|\circ\gamma\, |\dot\gamma|\, dt\, d\pi(\gamma) = \int |\varphi||\nabla f|\rho_\pi\, dm,$$

where the last equality follows by definition of barycenter. Since φ and f are arbitrary, the inequality above implies property (i).

Take now $\varphi = 1$. Then

$$\begin{aligned}\int_X b_\pi(f)\, dm &= \int\int_0^1 (f\circ\gamma)'(t)\, dt\, d\pi(\gamma) = \int (f(\gamma_1) - f(\gamma_0))\, d\pi(\gamma) = \\ &= \int f d((e_1)_\#\pi - (e_0)_\#\pi),\end{aligned}$$

which concludes the proof of (ii). □

THEOREM 3.2. $W^{1,p}(X,d,m) \subset BL^{1,p}(X,d,m)$ *and, for every* $f \in W^{1,p}(X,d,m)$,

$$|\nabla f|_{BL} \le |\nabla f|_W \quad m\text{-almost everywhere.}$$

PROOF. Let $f \in W^{1,p}(X,d,m)$. It is sufficient to prove that, for p-almost every γ,

$$|f(\gamma_1) - f(\gamma_0)| \le \int_\gamma |\nabla f|_W. \tag{45}$$

Possibly changing sign to f, we need only to prove (45) without the absolute value on the left-hand side. By definition of p-almost every curve, it is sufficient to fix a p-test plan π with $\mathcal{A}_q \in L^\infty(\pi)$ (this restricted class of test plans is sufficient to check p-negligibility, see Remark 3.1) and prove (45) for π-almost every γ, namely,

$$f(\gamma_1) - f(\gamma_0) \le \int_\gamma |\nabla f|_W, \text{ for } \pi\text{-almost every } \gamma. \tag{46}$$

Using the derivation b_π provided in Lemma 3.1, one obtains

$$\int (f(\gamma_1) - f(\gamma_0))\, d\pi(\gamma) = \int_X f \operatorname{div} b_\pi\, dm = -\int_X L_f(b_\pi)\, dm.$$

Moreover, by property (i) of Lemma 3.1 and by definition of $|\nabla f|_W$ one can estimate

$$-\int_X L_f(b_\pi)\, dm \le \int_X |\nabla f|_W \rho_\pi\, dm = \int\int_\gamma |\nabla f|_W\, d\pi(\gamma),$$

where the last equality holds by definition of ρ_π. We proved that

$$\int (f(\gamma_1) - f(\gamma_0))\, d\pi(\gamma) \le \int\int_\gamma |\nabla f|_W\, d\pi(\gamma) \tag{47}$$

which is an integral version of (46). To get the pointwise version, let us use a standard localization technique of measure theory. Given a p-test plan π and a Borel set $A \subset AC^q([0,1],X)$ with $\pi(A) > 0$, the measure $\tilde{\pi}(B) = \pi(A \cap B)/\pi(A)$ is still a p-test plan and thus satisfies (47), which writes

$$\int_A (f(\gamma_1) - f(\gamma_0))\, d\pi(\gamma) \le \int_A \int_\gamma |\nabla f|_W \, d\pi(\gamma).$$

Since A is arbitrary, (46) follows. □

Combining Propositions 4.3, 2.1 and Theorems 5.1, 3.2 we get the equivalence of the three notions of Sobolev spaces, as well as the coincidence m-a.e. of the three weak gradients, namely $|\nabla f|_*$, $|\nabla f|_{BL}$ and $|\nabla f|_W$.

CHAPTER 5

Functions of bounded variation

In this chapter we analyze different approaches to define functions of bounded variation in the framework of metric measure spaces, following two among the three main ideas developed for Sobolev functions: approximation by smooth functions and Beppo Levi's point of view. We will skip the approach based on derivations, which is very recent and covered in [**26**].

Throughout the chapter, the reader can think of the theory as a sort of "limit" of the theory in chapters 2,3 when p tends to 1, although this is not completely correct, as, even in the easiest case, $W^{1,1}(\mathbb{R}) \subsetneq BV(\mathbb{R})$.

1. The spaces $BV_*(X,d,m)$ and $BV_{BL}(X,d,m)$

Let (X,d) be a metric space. Since we do not assume any local compactness, let us define precisely what it means for a function to be locally Lipschitz. We set

$$\mathrm{Lip}_{loc}(X,d) := \{f : X \to \mathbb{R} \mid \forall\, x \in X \ \ \exists\, r > 0 \text{ s.t. } f|_{B_r(x)} \in \mathrm{Lip}(X)\}.$$

Obviously, when X is locally compact the set above coincides with the set of functions that are Lipschitz on any compact set.

Definition 1.1 (BV_*-space).

$$BV_*(X,d,m) \ := \ \Big\{ f \in L^1(X,m) \mid \exists\, f_n \in \mathrm{Lip}_{loc}(X,d) \text{ s.t. } \|f_n - f\|_{L^1} \to 0, \\ \limsup_{n\to\infty} \int_X \mathrm{lip}_a(f_n,x)\, dm(x) < \infty \Big\}.$$

Given $f \in BV_*(X,d,m)$ we set

$$|Df|_*(X) := \inf \left\{ \liminf_{n\to\infty} \int_X \mathrm{lip}_a(f_n,x) dm(x) \mid f_n \in \mathrm{Lip}_{loc}(X,d), \|f_n - f\|_{L^1} \to 0 \right\}.$$

This notion appeared in [**49**] in the framework of doubling metric spaces which support a Poincaré inequality and has been recently developed in a more general context in [**6**].

Given an open set $A \subset X$, let us define

$$|Df|_*(A) := \inf \left\{ \liminf_{n\to\infty} \int_A \mathrm{lip}_a(f_n,x) dm(x) \mid f_n \in \mathrm{Lip}_{loc}(X,d), \|f_n - f\|_{L^1} \to 0 \right\}. \quad (48)$$

In analogy with the classical case, where distributional derivatives of functions of bounded variation are Radon measures, the set function $A \mapsto |Df|_*(A)$ can be extended to a Radon measure as stated in the following result.

Theorem 1.1. *For all $f \in BV_*(X,d,m)$ the set function $A \mapsto |Df|_*(A)$ is the restriction to open sets of a unique finite Borel measure in X.*

By virtue of Theorem 1.1 we will denote the unique extension by $|Df|_*(\cdot)$.

Remark 1.1. Because of the generality of our framework, whenever $f \in BV_*(X,d,m)$, it is not immediate to define a vector valued measure Df, but only the positive measure $|Df|_*$. This is in the same spirit as for H- and BL-Sobolev spaces of the previous chapters, where the object ∇f is not defined, but only $|\nabla f|_*$ (see [**26**] for an equivalent definition of BV which provides, to some extent, a vector-valued measure, precisely a *measure-valued* operator L_f on $\mathrm{Der}^q(X,m)$). Of course, this can be done for some classes of metric measure spaces where more structure is available, e.g., Riemannian manifolds, Carnot–Carathéodory spaces, etc. Finally, to see that the language of derivations in a metric context is not very far from the one of derivations in a differential context, let us mention a result proved in [**34**] and [**52**]. With a slightly more restrictive notion of derivation than the one in Definition 1.1, the authors prove that if the metric measure structure is doubling, then Der is finite dimensional as L^∞-module.

Among BV functions characteristic functions play an important role; notice that in the Sobolev theory only trivial characteristic functions belong to the Sobolev class.

Definition 1.2. A Borel set $E \subset X$ has *finite perimeter* if $\mathbf{1}_E \in BV_*(X,d,m)$. In this case, the measure $|D\mathbf{1}_E|_*(\cdot)$ provided by Theorem 1.1 with $f = \mathbf{1}_E$ is called *perimeter measure* and is denoted by $P(E,\cdot)$.

Roughly speaking, the perimeter measure of a subset B is the surface measure of the intersection of B with the boundary of E.

Let us give another very recent notion of BV-space inspired by the point of view of Beppo Levi, see [**6**], based on the notion of "measure upper gradient".

Definition 1.3 (BV_{BL} and measure upper gradients).

$$BV_{BL}(X,d,m) := \Big\{ f \in L^1(X,m) \mid f\circ\gamma \in BV(0,1) \text{ for 1-almost every } \gamma \text{ and there exists a nonnegative finite measure } \mu \text{ on } X \text{ such that } \int \gamma_\#(|D(f\circ\gamma)|)d\pi(\gamma) \le C(\pi)\|\mathrm{lip}\,\gamma\|_{L^\infty(\pi)}\,\mu,\ \forall\, \text{1-test plan } \pi \Big\}.$$

The smallest μ satisfying the property[1] above is called *total variation* of f in the sense of Beppo Levi and it is denoted by $|Df|_{BL}$.

Note that, by definition, 1-test plans are concentrated on Lipschitz functions.

It turns out that the two notions in Definitions 1.1 and 1.3 are equivalent and the corresponding spaces are isometric. An adaption of the argument we used for Sobolev

[1]The inequality means that for every Borel set $E \subset X$,

$$\int \gamma_\#(|D(f\circ\gamma)|)(E)\,d\pi(\gamma) \le C(\pi)\|\mathrm{lip}\gamma\|_{L^\infty(\pi)}\,\mu(E),$$

where $\gamma_\#(|D(f\circ\gamma)|)$ is the measure on X obtained by pushing forward the measure $|D(f\circ\gamma)|$ in $(0,1)$ with γ.

spaces allows to prove rather easily that $BV_*(X,d,m) \subset BV_{BL}(X,d,m)$ and the inequality $|Df|_{BL} \le |Df|_*$ between the corresponding total variation measures, whereas the opposite inclusion and inequality are nontrivial, as in the Sobolev case, and have been proved in [**6**].

Remark 1.2. To get a glimpse of how one adapts the techniques for Sobolev spaces to BV spaces and of the limiting procedure as $p \downarrow 1$, let us see what happens to the Hopf–Lax semigroup as $p \downarrow 1$. For $1 < p < \infty$, we set

$$Q_t f(x) = \inf_{y \in X} \left(f(y) + \frac{d(x,y)^p}{pt^{p-1}} \right).$$

When p tends to 1 the correct version of the above semigroup, suitable to carry out computations, is

$$Q_t f(x) = \inf_{y \in B_t(x)} f(y).$$

With this formula, it is possible to get tools such as Hamilton–Jacobi subsolution properties that eventually allow to prove equivalence of the BV notions, see [**6**] for details.

2. Structure of the perimeter measure

In the sequel, we keep $BV_*(X,d,m)$ as working definition of bounded variation functions and describe more precisely the structure of sets of finite perimeter. Since the notion coincides with the BL-one, we will drop the "star" notation.

Keeping in mind the classical theory, the perimeter measure amounts to a weak notion of surface area.

Proposition 2.1. *Let $E \subset X$ be a set of finite perimeter, i.e., $\mathbf{1}_E \in BV(X,d,m)$. Then the following properties hold.*

(i) *Locality: $P(E,\cdot)$ is local on open sets, i.e., if A is an open set such that $(E\Delta F) \cap A = \emptyset$ then $P(E,A) = P(F,A)$.*
(ii) *Additivity: $P(E, A\cup B) = P(E,A) + P(E,B)$, whenever $A \cap B = \emptyset$.*
(iii) *Stability under complement: $P(E,B) = P(E^c,B)$.*
(iv) *Strong subadditivity: $P(E\cap F,B) + P(E\cup F,B) \le P(E,B) + P(F,B)$.*

The analogy with the classical theory suggests that $P(E,\cdot)$ should be concentrated on a lower dimensional set, and that it should be absolutely continuous with respect to a suitable "codimension 1" measure. The natural guess for such a set is the following.

Definition 2.1. The *essential boundary* of a Borel set E is

$$\partial^* E = \left\{ x \in X \mid \limsup_{\rho\to 0} \frac{\min\{m_E(x,\rho), m_{E^c}(x,\rho)\}}{m(B_\rho(x))} > 0 \right\},$$

where $m_A(x,\rho) = m(A \cap B_\rho(x))$.

Denote by E^0 the set of points where $\lim_{\rho\to 0^+} m_E(x,\rho)/m(B_\rho(x)) = 0$, and by E^1 the set of points where $\lim_{\rho\to 0^+} m_E(x,\rho)/m(B_\rho(x)) = 1$. It is easy to check that $\partial^* E = X \setminus (E^0 \cup E^1)$.

Remark 2.1. If m is doubling then the classical theory of Lebesgue points with respect to doubling measures, applied to the characteristic function of E, gives that $m(\partial^* E) = 0$. Our goal is actually to show more, that is, $\partial^* E$ is a lower dimensional set.

Let us find a natural measure on $\partial^* E$ with respect to which $P(E,\cdot)$ should be absolutely continuous. To this aim, we define the gauge function

$$h(B_\rho(x)) := \frac{m(B_\rho(x))}{\rho}.$$

Note that when (X,d,m) is Ahlfors s-regular, then $h(B_\rho(x)) \sim \rho^{s-1}$.

Denote by $\mathcal{H}^h$ the Hausdorff measure obtained from Carathéodory's construction with h (see for instance [**28**, Section 2.10]). Then $\mathcal{H}^h$ is a σ-additive Borel measure with values in $[0,\infty]$.

Let us recall some tools in theory of functions of bounded variation, which can be found in [**49**].

Proposition 2.2 (Coarea formula). *Let $u \in BV(X,d,m)$ be a nonnegative function. Then*

$$|Du|(B) = \int_0^\infty P(\{u>t\},B)\,dt \qquad \textit{for any Borel set } B \subset X.$$

Proposition 2.3 (Derivative of volume/Surface area). *Let E be a finite perimeter set. Then for all $\rho > 0$ one has*

$$P(E\setminus B_\rho(x), \partial B_\rho(x)) \le \frac{d^+}{dr} m_E(x,r)\Big|_{r=\rho},$$

where $\frac{d^+}{dr}$ denotes the upper right derivative (meaning that finiteness of the right hand side implies finiteness of the perimeter, and the inequality).

From now on we will assume the following condition:

(A1) the metric measure space (X,d,m) is doubling, i.e., there exists $C_D > 0$ such that, for every $x \in X$ and every $\rho > 0$,

$$m(B_{2\rho}(x)) \le C_D m(B_\rho(x)).$$

Let us mention some consequences of assumption (A1).

- The gauge function h is doubling, i.e., $h(B_{2\rho}(x)) \le C_D h(B_\rho(x))$ for every $x \in X$ and for every $\rho \ge 0$.
- There exist c, $s > 0$ depending on C_D such that

  $$\frac{m(B_r(x))}{m(B_R(y))} \ge c\left(\frac{r}{R}\right)^s,\ x \in B_R(y),\ 0 \le r \le R$$

  ($c = \log_2 C_D$). As a consequence, the Hausdorff dimension of X is not bigger than s. Note that this upper bound is of course less sharp than the one provided by curvature dimension inequalities.
- The doubling property of h ensures the validity of Vitali's Covering Theorem, stated below (see for instance [**19**] for a proof).

Proposition 2.4. *Let $K \subset X$ be a compact set and let ν be a finite and nonnegative Borel measure in X. Assume that a family of balls $\mathcal{F}$ is a fine[2] cover of K with the property*

[2]that is, $\forall\, x \in K$, $\forall\, \epsilon > 0$, there exists $B \in \mathcal{F}$ such that $x \in B$ and B has radius $< \epsilon$.

$\nu(B) \geq h(B)$, for every $B \in \mathcal{F}$. Then there exists a disjoint sub-family $\mathcal{F}' \subset \mathcal{F}$ such that

$$\mathcal{H}^h\left(K \setminus \bigcup_{B\in\mathcal{F}'} B\right) = 0.$$

We will apply this result to $\nu = P(E,\cdot)$. Using Proposition 2.4, one can prove the following differentiation result, see for instance [**19**].

COROLLARY 2.5. *Let $A \subset X$ and $t > 0$ be such that*

$$\limsup_{\rho\downarrow 0} \frac{\nu(B_\rho(x))}{h(B_\rho(x))} \geq t \qquad \forall x \in A.$$

Then, for every Borel set $B \subset A$, one has $\nu(B) \geq th(B)$. Moreover, if ν is finite, then

$$\limsup_{\rho\downarrow 0} \frac{\nu(B_\rho(x))}{h(B_\rho(x))} < \infty,\ \mathcal{H}^h\text{-almost everywhere.}$$

THEOREM 2.6. *Let E be a set of finite perimeter. Then*

(i) *$P(E,\cdot) \ll \mathcal{H}^h$,*
(ii) *there exists $C > 0$ depending only on C_D such that, for every open set A with $\mathcal{H}^h(\partial A) < \infty$, there holds $P(A,X) \leq C\mathcal{H}^h(\partial A)$.*

Note that property (i) does not imply the existence of the Radon–Nykodim density of $P(E,\cdot)$ with respect to $\mathcal{H}^h$, since the measure $\mathcal{H}^h$ is not σ-finite. Condition (ii) suggests that $\mathcal{H}^h$ is the good measure to consider on the boundary of A.

PROOF. For the first statement, it is sufficient to prove that $P(E,K) = 0$, for every compact set K with $\mathcal{H}^h(K) = 0$. Let $\epsilon > 0$ and let $B_{r_i}(x_i)$, $i = 1,\dots N_\epsilon$ be a finite family of balls such that $K \subset \cup_i B_{r_i}(x_i)$ and

$$\sum_{i=1}^{N_\epsilon} h(B_{r_i}(x_i)) < \epsilon.$$

Set $u(x) = d(x,x_i)$. Then $u \in \mathrm{Lip}(X)$ and $\mathrm{lip}(u) \leq 1$, which implies $|Du|(B) \leq m(B)$ for every Borel set B. Apply Proposition 2.2 to $u(x) = d(x,x_i)$. Then

$$\int_0^\infty P(B_t(x_i), B_{2r_i}(x_i))dt = \int_0^\infty P(X \setminus B_t(x_i), B_{2r_i}(x_i))dt \leq m(B_{2r_i}(x_i)).$$

Applying the mean value theorem to the function $s \mapsto P(B_s(x_i), B_{2r_i}(x_i))$, there exists $r_i' \in [r_i, 2r_i]$ such that

$$P(B_{r_i'}(x_i), X) \leq \frac{m(B_{2r_i}(x_i))}{r_i} \leq C_D h(B_{r_i}(x_i)),$$

where the last inequality follows by the doubling assumption. Set $A_\epsilon := \cup_{i=1}^{N_\epsilon} B_{r_i'}(x_i)$. Then, since A_ϵ covers K and $E \setminus A_\epsilon$ has no perimeter on K, i.e., $P(E \setminus A_\epsilon, X) = P(E \setminus A_\epsilon, X \setminus K)$.

Thus, the subadditivity of perimeter gives

$$\begin{aligned} P(E\setminus A_\epsilon, X) &= P(E\setminus A_\epsilon, X\setminus K) \le P(E, X\setminus K) + P(A_\epsilon, X\setminus K) \\ &\le P(E, X\setminus K) + P(A_\epsilon, X) \le P(E, X\setminus K) + C_D \sum_{i=1}^{N_\epsilon} h(B_{r_i}(x_i)) \\ &< P(E, X\setminus K) + \epsilon. \end{aligned}$$

Letting $\epsilon \to 0$, clearly $m(A_\epsilon) \to 0$, and by semicontinuity of the perimeter we obtain

$$P(E, X) \le P(E, X\setminus K),$$

whence $P(E, K) = 0$.

Finally, a similar idea without taking the limit leads to property (ii). □

It is now convenient to introduce the $(1,1)$-*Poincaré inequality*, which is the second main assumption on (X, d, m).

(A2) There exist $\lambda > 0$ and $C_P > 0$ such that, for every $u \in \mathrm{Lip}_{loc}(X)$ and $g \in UG(u)$, there holds

$$⨍_{B_\rho(x)} |u - \bar{u}|\, dm \le C_P \rho ⨍_{B_{\lambda\rho}(x)} g\, dm, \ \forall\, x \in X$$

where $\bar{u} = ⨍_{B_\rho(x)} u\, dm$.

Applying the inequality above to a sequence of Lipschitz functions approximating a function of bounded variation, we obtain that

$$⨍_{B_\rho(x)} |u - \bar{u}|\, dm \le C_P \rho \frac{|Du|(B_{\lambda\rho}(x))}{m(B_{\lambda\rho}(x))}, \ \forall\, u \in BV(X, d, m).$$

In the sequel, assumptions (A1), (A2) are sistematically made and a constant C will be called *structural* if it depends only on C_D, C_P, λ.

Hajlasz and Koskela proved (see [**36**]) that assumptions (A1) and (A2) imply the following stronger Poincaré inequality, also called $(1^*,1)$-*Poincaré inequality*, where[3] $1^* = \frac{s}{s-1}$. For every $u \in \mathrm{Lip}_{loc}(X)$ and $g \in UG(u)$,

$$\left(⨍_{B_\rho(x)} |u - \bar{u}|^{1^*}\, dm\right)^{1/1^*} \le C_P \rho ⨍_{B_{\lambda\rho}(x)} g\, dm.$$

Again, approximating by Lipschitz functions, the $(1^*,1)$-Poincaré inequality implies

$$\left(⨍_{B_\rho(x)} |u - \bar{u}|^{1^*}\, dm\right)^{1/1^*} \le C_P \rho \frac{|Du|(B_{\lambda\rho}(x))}{m(B_{\lambda\rho}(x))}, \ \forall\, u \in BV(X, d, m). \tag{49}$$

Applying (49) to characteristic functions, for every finite perimeter set E one obtains

$$\left(⨍_{B_\rho(x)} |\mathbf{1}_E - \bar{\mathbf{1}}_E|^{1^*}\, dm\right)^{1/1^*} \le C_P \frac{P(E, B_{\lambda\rho}(x))}{h(B_{\lambda\rho}(x))}, \tag{50}$$

[3] here $s > 1$ is the structural constant provided in the consequences of (A1).

which highlights the natural role of the Gauge function h. Finally, the inequality above implies the so-called *relative isoperimetric inequality*

$$\min(m_E(x,\rho), m_{E^c}(x,\rho)) \le C_1 \left(\frac{\rho^s}{m(B_\rho(x))}\right)^{1/(s-1)} P(E, B_{\lambda\rho}(x))^{s/(s-1)}, \tag{51}$$

where C_1 is a structural constant.

For every $\gamma > 0$ we set

$$(\partial^* E)_\gamma = \left\{ x \in X \mid \limsup_{\rho\to 0} \frac{\min\{m_E(x,\rho), m_{E^c}(x,\rho)\}}{m(B_\rho(x))} \ge \gamma \right\} \subset \partial^* E.$$

Theorem 2.7 (Structure theorem for sets of finite perimeter [**4**]). *Assume (A1) and (A2). Then there exists a structural constant $\gamma > 0$ such that $P(E,\cdot)$ is concentrated on $(\partial^* E)_\gamma$ and $\mathcal{H}^h((\partial^* E)_\gamma) < \infty$. In addition,*

(i) *$\mathcal{H}^h(\partial^* E \setminus (\partial^* E)_\gamma) = 0$;*

(ii) *there exist a structural constant $\gamma' > 0$ and a Borel function $\vartheta : X \to [\gamma', \infty)$ such that*

$$P(E,B) = \int_{B\cap(\partial^* E)_\gamma} \vartheta \, d\mathcal{H}^h \qquad \textit{for every Borel set } B \subset X.$$

Note that, once $\mathcal{H}^h((\partial^* E)_\gamma) < \infty$ is proved, property (ii) immediately follows by Radon–Nykodim Theorem applied to $\mathcal{H}^h \llcorner_{(\partial^* E)_\gamma}$.

Theorem 2.7 provides a representation of the perimeter measure in terms of the Hausdorff measure $\mathcal{H}^h$, but the density ϑ comes out from a non-constructive result. In practical situations (see [**30**] and [**16**] in Carnot groups and [**9**] in sub-Riemannian manifolds) it is desirable to identify ϑ explicitly. If one is allowed to differentiate $\mathcal{H}^h$ with respect to $P(E,\cdot)$, then it is possible to compute ϑ through a blow-up procedure (see 2.10.17, 2.10.18 in [**28**] and [**47**]).

Theorem 2.8 (Asymptotic doubling property of perimeter). *There exists a structural constant $\Sigma > 0$ such that*

$$\limsup_{\rho\downarrow 0} \frac{P(E, B_{2\rho}(x))}{P(E, B_\rho(x))} \le \Sigma < \infty,$$

at $P(E,\cdot)$-almost every $x \in X$.

This result allows to differentiate with respect to $P(E,\cdot)$ and, in some cases, to apply a blow-up procedure to compute ϑ.

Theorems 2.7 and 2.8 were proved in [**3**] in Ahlfors-regular metric measure spaces satisfying (A2), and in [**4**] in general doubling metric spaces (i.e., (A1) assumption) satisfying (A2). We give now a fairly detailed proof of the first result.

Proof of Theorem 2.7. Without loss of generality, we can reduce to the case where the metric space (X,d) is a length space. Indeed, assumptions (A1) and (A2) imply that (X,d) is quasi-convex and thus there exists a distance d' on X such that (X,d') is a length space and (X,d) and (X,d') are Lipschitz equivalent. In this context, Poincaré inequality (A2) holds with $\lambda = 1$.

We also assume for simplicity that $\rho \mapsto m_E(x,\rho)$ is continuous, i.e., all spheres are m-negligible.

Let $\gamma \in (0,1/2)$. We are going to prove that $P(E,K) = 0$ for every compact set $K \subset X \setminus (\partial^* E)_\gamma$, if γ is chosen sufficiently small. Using Egorov's Theorem we need only to check that if $K \subset X$ is a compact set such that there exists $\rho_0 > 0$ with

$$\min\{m_E(x,\rho), m_{E^c}(x,\rho)\} < \gamma m(B_\rho(x)),\ \forall\, x \in K, \forall \rho \in (0,\rho_0), \tag{52}$$

then $P(E,K) = 0$. The heuristic idea is that, at points $x \in K$, either $E \cap B_\rho(x)$ or $E^c \cap B_\rho(x)$ has small measure at points of K and this should lead to small perimeter of E in K.

We follow the same covering argument as in the proof of property (i) of Theorem 2.6. Let $r < \rho_0/2$. Then

$$K \subset \bigcup_{i=1}^N B_r(x_i),$$

for some N depending on r, where $x_i \in K$ and $d(x_i,x_j) \geq r$. By the doubling assumption, the overlapping of doubled balls is bounded by a structural (in particular not depending on r) constant $\xi > 0$, i.e.,

$$\sum_{i=1}^N \mathbf{1}_{B_{2r}(x_i)}(x) \leq \xi \qquad \forall x \in X.$$

Since $\rho \mapsto m_E(x,\rho)$ is continuous, $\gamma < 1/2$, and $m_E(x,\rho) + m_{E^c}(x,\rho) = m(B_\rho(x))$, inequality (52) implies that, for every $x \in K$, either

$$m_E(x,\rho) < \gamma m(B_\rho(x)),\ \forall \rho \in (0,\rho_0), \tag{53}$$

or

$$m_{E^c}(x,\rho) < \gamma m(B_\rho(x)),\ \forall \rho \in (0,\rho_0).$$

Hence, possibly splitting K in two subsets we can assume that (53) holds (the proof for the case where the second inequality holds can be carried out similarly). Recursively, for $i = 1,\ldots,N$, choose $\rho_i \in [r,2r]$ such that $\mathcal{H}^h(\partial B_{\rho_i}(x_i)) < \infty$, $\mathcal{H}^h(\partial B_{\rho_i}(x_i) \cap \partial B_{\rho_j}(x_j)) = 0$, and

$$r \frac{d}{ds} m_E(x_i,s)|_{s=\rho_i} \leq m_E(x_i,2r). \tag{54}$$

This can be done by the mean value theorem and thanks to the fact that for almost every radius ρ, $\mathcal{H}^h(\partial B_\rho(x)) < \infty$. Then, there exists a structural constant $C > 0$ such that

$$\begin{aligned} P(E \setminus B_{\rho_i}(x_i), \partial B_{\rho_i}(x_i)) &\leq \frac{m_E(x_i,2r)}{r} = m_E(x_i,2r)^{1/s} \frac{m_E(x_i,2r)^{1-1/s}}{r} \\ &\leq C\gamma^{1/s} P(E, B_{2r}(x_i)), \end{aligned} \tag{55}$$

where the first inequality follows by Proposition 2.3 and (54) and the second one by (53) and the relative isoperimetric inequality (51) (with $\lambda = 1$ since (X,d) is a length space). Define the open set $A_r = \cup_{i=1}^N B_{\rho_i}(x_i)$. Then, as $r \to 0$, $m(A_r) \to 0$ and $\mathbf{1}_{E \setminus A_r} \to \mathbf{1}_E$. By the properties of the perimeter (locality, subadditivity), we infer that

$$\begin{aligned}
P(E \setminus A_r, X) &= P(E \setminus A_r, X \setminus A_r) = P(E \setminus A_r, \partial A_r) + P(E \setminus A_r, X \setminus \bar{A}_r) \\
&= P(E \setminus A_r, \partial A_r) + P(E, X \setminus \bar{A}_r) \leq P(E \setminus A_r, \partial A_r) + P(E, X \setminus K) \\
&\leq \sum_{i=1}^{N} P(E \setminus A_r, \partial B_{\rho_i}(x_i)) + P(E, X \setminus K) \\
&\leq \sum_{i=1}^{N} \left(P(E \setminus B_{\rho_i}(x_i), \partial B_{\rho_i}(x_i)) + P\left(\cup_{j \neq i} B_{\rho_j}(x_j), \partial B_{\rho_i}(x_i)\right)\right) + P(E, X \setminus K).
\end{aligned}$$

By our choice, since $\mathcal{H}^h(\partial B_{\rho_i}(x_i) \cap \partial B_{\rho_j}(x_j)) = 0$ and $P(E, \cdot)$ is absolutely continuous with respect to $\mathcal{H}^h$ by Theorem 2.6, for every i, $P\left(\cup_{j \neq i} B_{\rho_j}(x_j), \partial B_{\rho_i}(x_i)\right) = 0$. Thus, using (55),

$$\begin{aligned}
P(E \setminus A_r, X) &\leq \sum_{i=1}^{N} P(E \setminus B_{\rho_i}(x_i), \partial B_{\rho_i}(x_i)) + P(E, X \setminus K) \\
&\leq C\gamma^{1/s}\xi P\left(E, \cup_{i=1}^{N} B_{2r}(x_i)\right) + P(E, X \setminus K).
\end{aligned}$$

Finally, taking the limit as $r \to 0$, by lower semicontinuity of the perimeter, we obtain

$$P(E, X) \leq C\gamma^{1/s}\xi P(E, K) + P(E, X \setminus K),$$

and choosing γ such that $C\gamma^{1/s}\xi < 1$ we conclude $P(E, K) = 0$. This proves that $P(E, \cdot)$ is concentrated on $(\partial^* E)_\gamma$.

Let us prove that, with the previous choice of γ, $\mathcal{H}^h((\partial^* E)_\gamma) < \infty$. By definition of $(\partial^* E)_\gamma$, there exists $\tilde{\gamma} > 0$ such that

$$\limsup_{\rho \downarrow 0} ⨍_{B_\rho(x)} |\mathbf{1}_E - \bar{\mathbf{1}}_E|^{1^*} dm \geq \tilde{\gamma}.$$

On the other hand, by (50),

$$\limsup_{\rho \downarrow 0} ⨍_{B_\rho(x)} |\mathbf{1}_E - \bar{\mathbf{1}}_E|^{1^*} dm \leq C_P^{1^*} \limsup_{\rho \downarrow 0} \frac{P(E, B_\rho(x))^{1^*}}{h(B_\rho(x))^{1^*}}.$$

Thus, for every $x \in (\partial^* E)_\gamma$

$$\limsup_{\rho \downarrow 0} \frac{P(E, B_\rho(x))}{h(B_\rho(x))} \geq \frac{1}{C_P} \tilde{\gamma}^{1/1^*} > 0,$$

and using Corollary 2.5 one gets $\mathcal{H}^h((\partial^* E)_\gamma) < \infty$. Finally an application of Radon–Nykodim theorem allows to prove the last statements. □

Let us give the idea behind the proof of Theorem 2.8. The goal is to estimate $P(E, B_{2\rho})/P(E, B_\rho)$ in terms of the same ratio involving h, which is bounded by the doubling property of h. We know that, at $P(E, \cdot)$-almost every $x \in X$,

$$0 < \limsup_{\rho \downarrow 0} \frac{P(E, B_\rho(x)))}{h(B_\rho(x))} < \infty,$$

(see the conclusion of the proof of Theorem 2.7 and Corollary 2.5). Hence to show Theorem 2.8 it suffices to prove that, at $P(E,\cdot)$-almost every $x \in X$,

$$\liminf_{\rho\downarrow 0} \frac{P(E, B_\rho(x)))}{h(B_\rho(x))} > 0. \tag{56}$$

To do this, the idea is to exploit the fact that

$$\limsup_{\rho\downarrow 0} \frac{\min\{m_E(x,\rho), m_{E^c}(x,\rho)\}}{m(B_\rho(x))} \geq \gamma. \tag{57}$$

in order to deduce the same inequality for the inferior limit and then use the isoperimetric inequality (51) to deduce (56). Thus we reduce to the proof of

$$\liminf_{\rho\downarrow 0} \frac{\min\{m_E(x,\rho), m_{E^c}(x,\rho)\}}{m(B_\rho(x))} \geq \gamma. \tag{58}$$

To this aim, the heuristic idea is to use an ODE argument to prove that volume fractions cannot oscillate too much. Set

$$V(\rho) := \min\{m_E(x,\rho), m_{E^c}(x,\rho)\}^{1/s}.$$

We study the ODE inequality by $V(\cdot)$, which follows by a property of quasi-minimality. To see where quasi-minimality comes from, we exploit an interesting principle, which was used first in [**56**] for currents, then in [**24**] for Sobolev functions and eventually in [**4**] for finite perimeter sets. Roughly speaking, this principle says that any additive and lower semicontinuous energy provides quasi-minimality on small scales for any object with finite energy. The rigorous statement for our purposes is as follows.

PROPOSITION 2.9 (Asymptotic quasi-minimality). *Let $D \in (0, 1/2)$ and $M > 1$. Then, for $P(E,\cdot)$-almost every $x \in X$ there exists $\rho_x > 0$ such that, for almost every $\rho \in (0, \rho_x)$, the inequalities*

$$\frac{1}{2} m(B_\rho(x)) \geq m_E(x,\rho) \geq D m(B_\rho(x))$$

imply

$$P(E, B_\rho(x)) \leq M P(E \setminus B_\rho(x), \partial B_\rho(x)).$$

Proposition 2.9 is essentially a consequence of the relative isoperimetric inequality (51). Thanks to Proposition 2.9 it is possible to deduce a differential inequality for $V(\cdot)$ which, together with (57), allows to prove (58).

Bibliography

[1] A. Agrachev and P. W. Y. Lee. Generalized Ricci curvature bounds for three dimensional contact subriemannian manifolds. *Math. Ann.*, 360(1-2):209–253, 2014.

[2] L. Ahlfors and A. Beurling. Conformal invariants and function-theoretic null-sets. *Acta Math.*, 83:101–129, 1950.

[3] L. Ambrosio. Some fine properties of sets of finite perimeter in Ahlfors regular metric measure spaces. *Adv. Math.*, 159(1):51–67, 2001.

[4] L. Ambrosio. Fine properties of sets of finite perimeter in doubling metric measure spaces. *Set-Valued Anal.*, 10(2-3):111–128, 2002. Calculus of variations, nonsmooth analysis and related topics.

[5] L. Ambrosio, M. Colombo, and S. Di Marino. Sobolev spaces in metric measure spaces: reflexivity and lower semicontinuity of slope. preprint arXiv:1212.3779v1, to appear on Advances studies in Pure Mathematics, 67:1–58, 2015, 2012.

[6] L. Ambrosio and S. Di Marino. Equivalent definitions of BV space and of total variation on metric measure spaces. *J. Funct. Anal.*, 266(7):4150–4188, 2014.

[7] L. Ambrosio, S. Di Marino, and G. Savaré. On the duality between p-Modulus and probability measures. preprint arXiv:1311.1381 to appear on Journal of European Mathematical Society, 2013.

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

[9] L. Ambrosio, R. Ghezzi, and V. Magnani. BV functions and sets of finite perimeter in sub-Riemannian manifolds. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 32(3):489–517, 2015.

[10] 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, 2005.

[11] L. Ambrosio, N. Gigli, and G. Savaré. Heat flow and calculus on metric measure spaces with Ricci curvature bounded below—the compact case. *Boll. Unione Mat. Ital. (9)*, 5(3):575–629, 2012.

[12] L. Ambrosio, N. Gigli, and G. Savaré. Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces. *Rev. Mat. Iberoam.*, 29(3):969–996, 2013.

[13] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.*, 195(2):289–391, 2014.

[14] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7):1405–1490, 2014.

[15] L. Ambrosio, N. Gigli, and G. Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1):339–404, 2015.

[16] L. Ambrosio, B. Kleiner, and E. Le Donne. Rectifiability of sets of finite perimeter in Carnot groups: existence of a tangent hyperplane. *J. Geom. Anal.*, 19(3):509–540, 2009.

[17] L. Ambrosio, A. Mondino, and S. G. Snonlinear diffusion equations and curvature conditions in metric measure spaces. in preparation.

[18] L. Ambrosio, A. Pinamonti, and G. Speight. Weighted sobolev spaces on metric measure spaces. preprint arXiv:1406.3000, 2014.

[19] L. Ambrosio and P. Tilli. *Topics on analysis in metric spaces*, volume 25 of *Oxford Lecture Series in Mathematics and its Applications.* Oxford University Press, Oxford, 2004.

[20] L. Ambrosio and D. Trevisan. Well-posedness of Lagrangian flows and continuity equations in metric measure spaces. *Anal. PDE*, 7(5):1179–1234, 2014.

[21] D. Bate. Structure of measures in Lipschitz differentiability spaces. *Journal of American Mathematical Society*, 28:421–482, 2015.

[22] F. Baudoin and G. Nicola. Curvature-dimension inequalities and ricci lower bounds for sub-riemannian manifolds with transverse symmetries. preprint arXiv:1101.3590, to appear on Journal of European Mathematical Society, 2015.

[23] P. Bernard. Young measures, superposition and transport. *Indiana Univ. Math. J.*, 57(1):247–275, 2008.

[24] J. Cheeger. Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.*, 9(3):428–517, 1999.

[25] J. Cheeger and T. H. Colding. On the structure of spaces with Ricci curvature bounded below. III. *J. Differential Geom.*, 54(1):37–74, 2000.

[26] S. Di Marino. Sobolev and BV spaces on metric measure spaces via derivations and integration by parts. preprint available at http://cvgmt.sns.it/paper/2521, 2014.

[27] M. Erbar, K. Kuwada, and K.-T. Sturm. On the equivalence of the entropic curvature-dimension condition and bochners inequality on metric measure spaces. *Inventiones mathematicae*, pages 1–79, 2014.

[28] H. Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[29] B. Franchi, R. Serapioni, and F. Serra Cassano. Meyers-Serrin type theorems and relaxation of variational integrals depending on vector fields. *Houston J. Math.*, 22(4):859–890, 1996.

[30] B. Franchi, R. Serapioni, and F. Serra Cassano. On the structure of finite perimeter sets in step 2 Carnot groups. *J. Geom. Anal.*, 13(3):421–466, 2003.

[31] B. Fuglede. Extremal length and functional completion. *Acta Math.*, 98:171–219, 1957.

[32] N. Garofalo and D.-M. Nhieu. Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces. *Comm. Pure Appl. Math.*, 49(10):1081–1144, 1996.

[33] N. Gigli. Nonsmooth differential geometry - an approach tailored for spaces with ricci curvature bounded from below. preprint arXiv:1407.0809, submitted, 2014.

[34] J. Gong. Rigidity of derivations in the plane and in metric measure spaces. *Illinois J. Math.*, 56(4):1109–1147, 2012.

[35] N. Gozlan, C. Roberto, and P.-M. Samson. Hamilton Jacobi equations on metric spaces and transport entropy inequalities. *Rev. Mat. Iberoam.*, 30(1):133–163, 2014.

[36] P. Hajłasz and P. Koskela. Sobolev met Poincaré. *Mem. Amer. Math. Soc.*, 145(688):x+101, 2000.

[37] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext, Springer New York, 2001.

[38] J. Heinonen. Nonsmooth calculus. *Bull. Amer. Math. Soc. (N.S.)*, 44(2):163–232, 2007.

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

[40] J. Heinonen, P. Koskela, N. Shanmugalingam, and J. Tyson. Newtonian spaces: an extension of sobolev spaces to metric measure spaces: an approach based on upper gradients. book to appear.

[41] N. Juillet. Geometric inequalities and generalized Ricci bounds in the Heisenberg group. *Int. Math. Res. Not. IMRN*, (13):2347–2373, 2009.

[42] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

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

[44] K. Kuwada. Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.*, 258(11):3758–3774, 2010.

[45] B. Levi. Sul principio di dirichlet. *Rendiconti del Circolo Matematico di Palermo*, 22(1):293–359, 1906.

[46] S. Lisini. Characterization of absolutely continuous curves in Wasserstein spaces. *Calc. Var. Partial Differential Equations*, 28(1):85–120, 2007.

[47] V. Magnani. On a measure theoretic area formula. preprint arXiv:1401.2536, to appear on Proc. Roy. Soc. Edinburgh Sect. A, 2014.

[48] N. G. Meyers and J. Serrin. $H = W$. *Proc. Nat. Acad. Sci. U.S.A.*, 51:1055–1056, 1964.

[49] M. Miranda, Jr. Functions of bounded variation on "good" metric spaces. *J. Math. Pures Appl. (9)*, 82(8):975–1004, 2003.

[50] G. Monge. Mémoire sur la théorie des déblais et de remblais. *Histoire de lAcadémie Royale des Sciences de Paris avec les Mémoires de Mathématique et de Physique pour la même année*, pages pages 666–704, 1781.

[51] L. Rifford. Ricci curvatures in Carnot groups. *Math. Control Relat. Fields*, 3(4):467–487, 2013.

[52] A. Schioppa. On the relationship between derivations and measurable differentiable structures. *Ann. Acad. Sci. Fenn. Math.*, 39(1):275–304, 2014.

[53] N. Shanmugalingam. Newtonian spaces: an extension of Sobolev spaces to metric measure spaces. *Rev. Mat. Iberoamericana*, 16(2):243–279, 2000.

[54] S. K. Smirnov. Decomposition of solenoidal vector charges into elementary solenoids and the structure of normal one-dimensional currents. *St. Petersburg Mathematical Journal*, 5:841–867, 1994.

[55] N. Weaver. Lipschitz algebras and derivations II. Exterior differentiation. *Journal of Functional Analysis*, 178(1):64 – 112, 2000.

[56] B. White. A new proof of the compactness theorem for integral currents. *Comment. Math. Helv.*, 64(2):207–220, 1989.

[57] L. C. Young. *Lectures on the calculus of variations and optimal control theory*. Foreword by Wendell H. Fleming. W. B. Saunders Co., Philadelphia-London-Toronto, Ont., 1969.

[58] W. P. Ziemer. *Weakly differentiable functions*, volume 120 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1989. Sobolev spaces and functions of bounded variation.