Optimal control in sub-Riemannian geometry

Luca Tamanini

November 29, 2016

Contents

Part I

Summer School CIRM

Chapter 1

The geometry of subelliptic diffusion

This section is based on the course held by A. Thalmaier at the CIRM in Marseille during the CIRM Summer School *Sub-Riemannian manifolds: from geodesics to hypoelliptic diffusion*. We discuss hypoelliptic and subelliptic diffusions; the lectures include the following topics:

- Malliavin calculus;
- Hörmander's theorem;
- smoothness of transition probabilities under Hormander's brackets condition;
- control theory and Stroock-Varadhan's support theorems;
- hypoelliptic heat kernel estimates;
- gradient estimates and Harnack type inequalities for subelliptic diffusion semi-groups;
- notions of curvature related to sub-Riemannian diffusions.

1.1 Lecture 1 - 1 September

1.2 Lecture 2 - 2 September

1.3 Lecture 3 - 4 September

1.4 Lecture 4 - 5 September

Chapter 2

Geometric control and sub-Riemannian geodesics

This section is based on the course held by L. Rifford at the CIRM in Marseille during the CIRM Summer School *Sub-Riemannian manifolds: from geodesics to hypoelliptic diffusion*. This will be an introduction to sub-Riemannian geometry from the point of view of control theory. We will define sub-Riemannian structures and prove the Chow Theorem. We will also describe normal and abnormal geodesics and discuss the completeness of the Carnot-Carathéodory distance (Hopf-Rinow Theorem). Several examples will be given (Heisenberg group, Martinet distribution, Grušin plane).

2.1 Lecture 1 - 1 September

2.2 Lecture 2 - 2 September

2.3 Lecture 3 - 3 September

2.4 Lecture 4 - 4 September

Chapter 3

Differential forms and the Hölder equivalence problem

This section is based on the course held by P. Pansu at the CIRM in Marseille during the CIRM Summer School *Sub-Riemannian manifolds: from geodesics to hypoelliptic diffusion*. In 1993, Gromov asked for which exponents α there exist C^α-Hölder continuous (local) homeomorphisms of Euclidean spaces to sub-Riemannian Carnot groups. We will explain some of Gromov's partial results on this question. They rely on Rumin's theory of differential forms adapted to sub-Riemannian spaces. If time permits, other approaches to the Hölder equivalence problem will be discussed.

3.1 Lecture 1 - 1 September

3.2 Lecture 2 - 2 September

3.3 Lecture 3 - 4 September

3.4 Lecture 4 - 5 September

Chapter 4

Hypoelliptic operators and analysis on Carnot-Carathéodory spaces

This section is based on the course held by NĠarofalo at the CIRM in Marseille during the CIRM Summer School *Sub-Riemannian manifolds: from geodesics to hypoelliptic diffusion.* In this course, we will define the sub-Laplacian associated with a sub-Riemannian structure and we will describe its hypoellipticity under the Hörmander condition. We will also introduce the main tools for the study of sub-elliptic PDEs.

4.1 Lecture 1 - 1 September

4.2 Lecture 2 - 2 September

4.3 Lecture 3 - 3 September

4.4 Lecture 4 - 4 September

Part II

IHP Trimester

Chapter 5

Geometric measure theory

This section is based on the course held by Francesco Serra Cassano and Luigi Ambrosio at the IHP in Paris during the IHP Trimester *Geometry, Analysis and Dynamics on Sub-Riemannian Manifolds*. The course is divided in two parts and the structure of the first one is the following:

(i) introduction to Carnot groups;

(ii) differential calculus between Carnot groups;

(iii) differential calculus within Carnot groups;

(iv) sets of finite perimeter and minimal surfaces.

Main aim is to provide the reader with the most important notions and results in the theory of Carnot groups; hence many proofs will be omitted.

In the second part, we will focus our attention on Sobolev and BV functions on metric measure spaces. In more details, we will illustrate some recent developments of calculus on metric measure spaces $(X, d, \mathfrak{m})$. These developments have been motivated by recent works in collaboration with N. Gigli and G. Savaré on the theory of metric measure spaces with Ricci curvature bounded from below and provided (at least in the infinitesimally Hilbertian case) a complete equivalence between the Eulerian theory of Bakry-Émery and the Lagrangian theory of Lott-Villani and Sturm. Independently of these motivations, we will focus on the most conceptual aspects of the theory, following as a guideline the theory of Sobolev and BV spaces in metric measure spaces, where already this Eulerian-Lagrangian duality appears. We will also draw some links with optimal transport and with the seminal theory of Cheeger, showing how some density and reflexivity results of the Sobolev theory can be estabilished even dropping the doubling assumption on the reference measure $\mathfrak{m}$ and the validity of the Poincaré inequality.

5.1 Lecture 1 - 8 September

Abstract

Aim of the first lecture is to introduce the basic definitions of sub-Riemannian geometry, as for instance the notion of Carnot group, horizontal subbundle and invariant distance. Some examples and very elementary results will also be given.

Let us begin by providing the definition of a Carnot group.

Definition 5.1.1. *A* Carnot group $\mathbb{G}$ *(also called* stratified group *by Gromov) is a connected, simply connected Lie group of finite dimension whose Lie algebra* $\mathfrak{g}$ *admits a* k*-step stratification, that is* $\mathfrak{g}$ *can be written as*

$$\mathfrak{g} = V_1 \oplus ... \oplus V_k \tag{5.1.1}$$

where, for $i = 1, ..., k$*,* V_i *is a linear subspace such that*

$$[V_1, V_i] = V_{i+1}, \quad V_k \neq \{0\}, \quad V_{i+1} = \{0\} \tag{5.1.2}$$

Condition (5.1.1) + (5.1.2) is also called Hörmander condition and notice that it means that the Lie algebra is generated by V_1. Let us also fix the notation: we will denote by m_i the dimension of V_i and

$$h_i := m_1 + m_2 + ... + m_i$$

with the convention $h_0 := 0$; notice that $h_k = n$. We also provide the notion of adapted basis for $\mathfrak{g}$.

Definition 5.1.2. *A basis* $X_1, ..., X_n$ *of* $\mathfrak{g}$ *is said to be* adapted to (5.1.1) *if, for every* $j = 1, ..., k$*, the family* $X_{h_{j-1}+1}, ..., X_{h_j}$ *is a basis of* V_j*. In an adapted basis, the family* $X_1, ..., X_{m_1}$ *is called a* generating system of vector fields.

As a further step, let us introduce a privileged system of coordinates on $\mathbb{G}$, namely the so-called exponential coordinates; they are constructed in the following way. First, consider the exponential map $\exp : \mathfrak{g} \to \mathbb{G}$; it is well-known that it is 1-to-1 and onto (indeed, it is a diffeomorphism). Thus, any point $p \in \mathbb{G}$ can be written in a unique way as

$$p = \exp(p_1 X_1 + ... + p_n X_n)$$

and therefore any point $p \in \mathbb{G}$ can be identified with its coordinates $(p_1, ..., p_n) \in \mathbb{R}^n$. By means of this identification, the original group $\mathbb{G}$ can be seen as $(\mathbb{R}^n, \cdot)$, where $\cdot$ is a suitable group law. Such law can be explicitly determinated by means of the Campbell-Hausdorff formula, but this will be done in the following Proposition 5.1.6. Now let us provide a different way to represent $\mathbb{G}$ in coordinates, a way which is more linked to the decomposition (5.1.1). If $p \in \mathbb{G}$, we put

$$p^i = (p_{h_{i-1}+1}, ..., p_{h_i}) \in \mathbb{R}^{m_i}$$

In this way we can identify

$$p = [p^1, ..., p^k] \in \mathbb{R}^{m_1} \times ... \times \mathbb{R}^{m_k} \equiv \mathbb{R}^n \tag{5.1.3}$$

and the relation with (5.1.1) is due to the fact that, via the exponential map, $\mathbb{G}$ decomposes as

$$\mathbb{G} = \mathbb{V}_1 \oplus ... \oplus \mathbb{V}_k \tag{5.1.4}$$

where $\mathbb{V}_i = \exp(V_i)$ for $i = 1, ..., k$. At this point, let us introduce the so-called horizontal subbundle and the sub-Riemannian structure.

Definition 5.1.3. *Let $X_1, ..., X_{m-1}$ be a system of generating vector fields. The* horizontal subbundle*, denoted by $H\mathbb{G}$, is defined by*

$$H\mathbb{G}_x := \text{span}(X_1(x), ..., X_{m_1}(x)), \quad \forall x \in \mathbb{G}$$

and it is endowed with a scalar product $\langle\cdot,\cdot\rangle_x$ and an associated norm $|\cdot|_x$ making $X_1(x), ..., X_{m_1}(x)$ an orthonormal basis of the fiber $H\mathbb{G}_x$. The sections of $H\mathbb{G}$ are called horizontal sections*, a vector of $H\mathbb{G}_x$ is a* horizontal vector *while any vector in $T\mathbb{G}_x$ that is not horizontal is a* vertical vector.

In this framework, a *horizontal section* is represented by its canonical (exponential) coordinates w.r.t. the moving frame $X_1, ..., X_{m_1}$, that is if $\psi : \Omega \to H\mathbb{G}$ (where Ω is an open subset of $\mathbb{G}$) is an horizontal section, then in coordinates

$$\psi(x) = \psi_1(x)X_1(x) + ... + \psi_{m_1}X_{m_1}(x), \quad \forall x \in \Omega$$

and so we can identify

$$\psi \equiv (\psi_1, ..., \psi_{m_1}) : \Omega \subset \mathbb{R}^n \to \mathbb{R}^{m_1}$$

We will refer to V_1 as *horizontal layer*. For what concerns the differential structure of Carnot groups the discussion is (for the moment) over; on the other hand, as Carnot groups are in particular Lie groups, they have an algebaric structure that we are going to endeavor immediately by talking about two important families of diffeomorphisms extremely linked to the algebraic structure of $(\mathbb{R}^n, \cdot)$. As a first step, let us begin with the left-translations: for any $x \in \mathbb{G}$, the *left-translation* $\tau_x : \mathbb{G} \to \mathbb{G}$ is defined as

$$\tau_x(z) := x \cdot z, \quad \forall z \in \mathbb{G}$$

It is well-known that τ_x is a diffeomorphism. The second family is given by dilations: for $\lambda > 0$, the *dilation* $\delta_\lambda : \mathbb{G} \to \mathbb{G}$ is defined as

$$\delta_\lambda(x_1, ..., x_n) := (\lambda^{\alpha_1} x_1, ..., \lambda^{\alpha_n} x_n) \tag{5.1.5}$$

where $\alpha_i \in \mathbb{N}$ is called *homogeneity of the variable x_i* in $\mathbb{G}$ (in the following lectures of F. Jean, it will be called *weight*) and it is defined as $\alpha_j = i$ whenever $h_{i-1} + 1 \leq j \leq h_i$. As for left-translations, also dilations are diffeomorphisms. Let us now provide two simple but relevant examples of Carnot groups.

Example 5.1.4 (Euclidean spaces)**.** The Euclidean space is a trivial step 1 Carnot group. In more details, $\mathbb{G} = (\mathbb{R}^n, +)$, where $+$ is the usual sum, satisfies the conditions of Definition 5.1.1 and this means that the Euclidean geometry is a particular subclass of sub-Riemannian geometry. $\diamondsuit$

Example 5.1.5 (Heisenberg groups)**.** Let us denote by $\mathbb{H}^n = (\mathbb{R}^{2n+1}, \cdot)$ the n-Heisenberg group, where the group law $\cdot$ is defined coordinatewise as follows: for $p, q \in \mathbb{R}^{2n+1}$,

$$p \cdot q := \left(p_1 + q_1, ..., p_{2n} + q_{2n}, p_{2n+1} + q_{2n+1} + \frac{1}{2}\sum_{i=1}^{n}(p_i q_{i+n} - p_{i+n} q_i) \right)$$

Let us also introduce, for any $\lambda > 0$, the family of (non-isotropic) dilations by

$$\delta_\lambda(p) := (\lambda p_1, ..., \lambda p_{2n}, \lambda^2 p_{2n+1}), \quad \forall p \in \mathbb{H}^n$$

The standard basis of the Lie algebra $\mathfrak{h}_n$ of $\mathbb{H}^n$ is given by the left-invariant vector fields

$$\begin{aligned} X_j &= \partial_j - \frac{p_{n+j}}{2}\partial_{2n+1}, \quad j = 1, ..., n \\ Y_j &= \partial_{n+j} + \frac{p_j}{2}\partial_{2n+1}, \quad j = 1, ..., n \\ T &= \partial_{2n+1} \end{aligned}$$

It is well-known that the only non-trivial commutator is given by $[X_j, Y_j] = T$

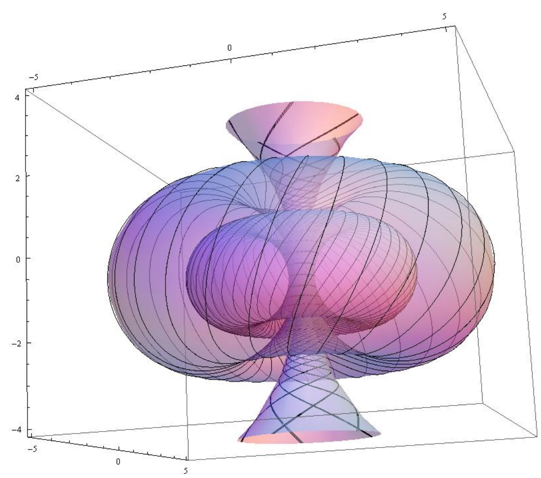


Figure 5.1: The Heisenberg group viewed as a Hopf fibration

for $j = 1, ..., n$. Therefore, $\mathbb{H}^n$ turns out to be a step 2 Carnot group and, as a consequence, if we define

$$V_1 := \text{span}(X_1, ..., X_n, Y_1, ..., Y_n), \quad V_2 := \text{span}(T)$$

we can see that the Lie algebra $\mathfrak{h}_n$ is stratified as $\mathfrak{h}_n = V_1 \oplus V_2$. $\diamondsuit$

Now let us talk about the group law in $(\mathbb{R}^n, \cdot)$ in a more detailed way. Indeed, as already anticipated above, by means of the Campbell-Hausdorff formula it is possible to give an explicit representation for it.

Proposition 5.1.6. *The group product has the form*

$$x \cdot y = x + y + \mathscr{Q}(x, y), \quad \forall x, y \in \mathbb{R}^n$$

where $\mathscr{Q} = (\mathscr{Q}_1, ..., \mathscr{Q}_n) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ *and each* $\mathscr{Q}_i$ *is a homogeneous polynomial of degree* α_i *w.r.t. the intrinsic dilations of* $\mathbb{G}$ *defined in (5.1.5), that is*

$$\mathscr{Q}_i(\delta_\lambda x, \delta_\lambda y) = \lambda^{\alpha_i} \mathscr{Q}_i(x, y), \quad \forall x, y \in \mathbb{G}$$

Moreover

(i) $\mathscr{Q}$ *is antisymmetric, that is* $\mathscr{Q}(p, q) = -\mathscr{Q}(-p, -q)$ *for all* $p, q \in \mathbb{G}$*;*

(ii) for every $x, y \in \mathbb{G}$ *it holds*

$$\begin{aligned}
&\mathscr{Q}_1(x, y) = ... = \mathscr{Q}_{m_1}(x, y) = 0 \\
&\mathscr{Q}_j(x, 0) = \mathscr{Q}_j(0, y) = 0, \ \mathscr{Q}_j(x, x) = \mathscr{Q}_j(x, -x) = 0, \quad \text{for } m_1 < j \le n \\
&\mathscr{Q}_j(x, y) = \mathscr{Q}_j(x_1, ..., x_{h_{i-1}}, y_1, ..., y_{h_{i-1}}), \quad \text{if } 1 < i \le k \text{ and } j \le h_i
\end{aligned}$$

The first condition is due to the fact that there are no commutators in the first layer, whereas the third one tells us that $\mathscr{Q}_j$ *depends only on the first* h_{i-1} *coordinates provided that* $j \le h_i$*.*

(iii) it holds

$$\mathscr{Q}_i(p, q) = \sum_{k,h} \mathcal{R}^i_{k,h}(p, q)(p_k q_h - p_h q_k)$$

where the functions $\mathcal{R}^i_{k,h}$ *are homogeneous polynomials of degree* $\alpha_i - \alpha_k - \alpha_h$ *w.r.t. group dilations and the sum is extended to all* h, k *such that* $\alpha_h + \alpha_k \le \alpha_i$*.*

Proof. See [6]. □

As already said, the proof of this result relies on the Campbell-Hausdorff formula. Note that, in particular, from this proposition it follows that the dilations δ_λ are not only diffeomorphisms but also automorphisms of $\mathbb{G}$, namely

$$\delta_\lambda(x \cdot y) = \delta_\lambda x \cdot \delta_\lambda y$$

the inverse x^{-1} of an element $x = (x_1, ..., x_n) \in \mathbb{G}$ is given by $-x = (-x_1, ..., -x_n)$, thanks to the fact that $\mathscr{Q}_j(x, -x) = 0$, and the unit element of the group is $e = (0, ..., 0)$. In addition, from the structure of the group law we can also deduce the structure of the basis $X_1, ..., X_n$, as we are going to see immediately.

Proposition 5.1.7. *The vector fields $X_1, ..., X_n$ have polynomial coefficients and if $h_{\ell-1} < j \le h_\ell$, $1 \le \ell \le k$, then*

$$X_j(x) = \partial_j + \sum_{i>h_\ell}^{n} q_{i,j}(x)\partial_i$$

where

$$q_{i,j}(x) = \frac{\partial \mathcal{Q}_i}{\partial y_j}(x,y)\bigg|_{y=0}$$

so that if $h_{\ell-1} < j \le h_\ell$, then $q_{i,j}(x) = q_{i,j}(x_1, ..., x_{h_{\ell-1}})$, i.e. $q_{i,j}$ depends only on the first $h_{\ell-1}$ coordinates, and $q_{i,j}(0) = 0$. In particular, we can infer that $X_1, ..., X_n$ are self-adjoint w.r.t. the L^2-scalar product.

To sum up the environment we described, by a Carnot group $\mathbb{G}$ we will mean a couple $(\mathbb{R}^m, \cdot)$, in exponential coordinates, equipped with a basis $X_1, ..., X_n$ of its Lie algebra $\mathfrak{g}$ satisfying the properties of Proposition 5.1.7. After a quick introduction on differential and algebraic properties of Carnot groups, let us now consider their metric aspects. Indeed, a Lie group is not *a priori* endowed with a distance (no metric notions are involved in the definition), but metrics can be defined on it; thus, let us define the metrics we are interested in, namely those which are compatible with the topological and algebraic structure.

Definition 5.1.8. *Given a Carnot group $\mathbb{G} = (\mathbb{R}^n, \cdot)$, a distance d on it is said to be* invariant *if*

(i) $d : \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty[$ is continuous w.r.t. the Euclidean topology;

(ii) d is left-invariant w.r.t. the family of left-translations and 1-homogeneous w.r.t. the one of dilations, namely

$$d(g \cdot p, g \cdot q) = d(p,q), \quad d(\delta_\lambda p, \delta_\lambda q) = \lambda d(p,q)$$

for all $p, q, g \in \mathbb{G}$ and all $\lambda > 0$.

It is well-known that an invariant distance d can be equivalently represented by a *homogeneous norm* $\|\cdot\|$, i.e. a continuous function $\|\cdot\| : G \equiv \mathbb{R}^n \to [0, +\infty[$ w.r.t. the Euclidean topology such that

(i) $\|p\| = 0$ if and only if $p = 0$;

(ii) $\|p^{-1}\| = \|p\|$, $\|\delta_\lambda p\| = \lambda\|p\|$ for every $p \in \mathbb{G}$ and $\lambda > 0$;

(iii) $\|p \cdot q\| \le \|p\| + \|q\|$ for every $p, q \in \mathbb{G}$.

Indeed, given a homogeneous norm $\|\cdot\|$, we can define an invariant distance d by setting

$$d(p,q) = d(q^{-1} \cdot p, 0) := \|q^{-1} \cdot p\|, \quad \forall p, q \in \mathbb{G}$$

Viceversa, if d is an invariant distance on $\mathbb{G}$, a homogeneous norm can be simply defined by $\|p\| := d(p, 0)$ for any $p \in \mathbb{G}$. The following proposition is a classical result on invariant distances in $\mathbb{R}^n$.

Proposition 5.1.9. *Let $\mathbb{G}$ be a Carnot group. Then two invariant distances d_1 and d_2 on $\mathbb{G}$ are equivalent, that is there exists $C \geq 1$ such that for all $x, y \in \mathbb{G}$ it holds*

$$C^{-1}d_1(x,y) \leq d_2(x,y) \leq Cd_1(x,y)$$

Indeed, in any finite-dimensional vector space two (homogeneous) norms are equivalent and then also the associated invariant distances. As a further step, let us fix the notation for open and closed balls centered at p and with radius $r > 0$ for a given invariant metric d; they will be denoted by $U_d(p,r)$ and $B_d(p,r)$ respectively and a first important fact is the following: their diameter is exactly equal to twice the radius (in general sub-Riemannian manifolds, we can only say that the diameter is smaller than or equal to twice the radius).

Proposition 5.1.10. *Let $\mathbb{G}$ be a Carnot group endowed with an invariant distance d. Then*

$$\operatorname{diam} B_d(p,r) = 2r, \quad \forall p \in \mathbb{G}, \forall r > 0$$

Moreover, if μ is a Radon measure on $\mathbb{G}$, s-homogeneous w.r.t. the family of dilations for some $s > 0$, then

$$\mu(\partial B_d(0,r)) = 0, \quad \forall r > 0$$

Proof. See [8]. □

5.2 Lecture 2 - 9 September

Abstract

In this lecture we first deepen our knowledge on invariant distances on Carnot groups, by providing both an abstract approach (via Carnot-Carathéodory distances) and a constructive one (the d_∞ distance. Then we move to the more general framework of metric spaces, in order to provide three diffent types of Hausdorff measure and the fundamental notion of Ahlfors regularity.

In the previous lecture we introduced the notion of invariant distance on a Carnot group and we said that two invariant distances on the same Carnot group are always equivalent. Let us begin this lecture by deepening this topic and let us provide a couple of relevant invariant metrics on a Carnot group:

(i) the Carnot-Carathéodory distance;

(ii) the d_∞ distance.

For the first one, given a Carnot group $\mathbb{G}$ and a generating system of vector fields $X_1, ..., X_{m_1}$, we say that an absolutely continuous curve $\gamma : [0,T] \to \mathbb{G}$

is a *sub-unit curve* w.r.t. $X_1, ..., X_{m_1}$ if there exist measurable real functions $c_1, ..., c_{m_1}$ on $[0, T]$ such that $\sum_{j=1}^{m_1} c_j^2 \leq 1$ and

$$\dot{\gamma}(s) = \sum_{j=1}^{m_1} c_j(s) X_j(\gamma(s)), \quad \text{for a.e. } s \in [0, T]$$

The Carnot-Carathéodory distance is then defined as follows.

Definition 5.2.1. *If $p, q \in \mathbb{G}$, their Carnot-Carathéodory distance is defined as*

$$d_C(p, q) := \inf\{T > 0 \,:\, \gamma \text{ sub-unit}, \gamma(0) = p, \gamma(T) = q\}$$

By Chow-Rashevskii's theorem (see Theorem 7.1.8) we know that the set of sub-unit curves joining p and q is not empty, so that $d_C(p, q) < +\infty$ always. Furthermore, d_C induces the Euclidean topology on $\mathbb{G}$ but let us point out that in general Carnot-Carathéodory distances are not Euclidean at any scale, and hence not Riemannian.

Theorem 5.2.2. *Let $\mathbb{G} = (\mathbb{R}^n, \cdot)$ be a step k Carnot group. Let d_C and d_E denote, respectively, the Carnot-Carathéodory distance on $\mathbb{G} \equiv \mathbb{R}^n$ and the Euclidean one on $\mathbb{R}^n$. Then, the following facts hold:*

(i) $A \subset \mathbb{R}^n$ is bounded w.r.t. d_C if and only if it is bounded w.r.t. d_E;

(ii) for each compact set $K \subset (\mathbb{R}^n, d_E)$ there exists a positive constant C_K such that

$$\frac{1}{C_K} d_E(x, y) \leq d_C(x, y) \leq C_K d_E(x, y)^{1/k} \tag{5.2.1}$$

for all $x, y \in K$;

(iii) the identity map $\mathrm{Id} : (\mathbb{R}^n, d_C) \to (\mathbb{R}^n, d_E)$ is a homeomorphism.

Proof. See [6]. □

Actually, the left inequality in (5.2.1) holds globally, i.e. there exists a constant $C > 0$ such that $C^{-1} d_E(x, y) \leq d_C(x, y)$ for every $x, y \in \mathbb{R}^n$, and the sharp constant is easily seen to be $C = 1$. Notice also that if $k > 1$, then d_C is not a Riemannian distance (it is rather a "fractal" distance) and this is coherent with what we have just said above: Carnot-Carathéodory distances are not Riemannian ones in general. In other words, the metric geometry of $(\mathbb{R}^n, d_E)$ may be different from the metric geometry of $(\mathbb{R}^n, d_C)$ In order to understand this claim, let us consider the following example, but let us first explain the notion of *total variation* that we will use.

Definition 5.2.3. *Let $\mathbb{G}$ be a Carnot group. The* total variation *of a curve $\gamma : [0, 1] \to \mathbb{G}$ is defined as*

$$\mathrm{Var}(\gamma) := \sup_{0 \leq t_1 \leq \ldots \leq t_k \leq 1} \sum_{i=1}^{k-1} d_C(\gamma(t_i), \gamma(t_{i+1}))$$

where the supremum is taken over all finite partitions of $[0, 1]$. If $\mathrm{Var}(\gamma) < +\infty$, we say that γ is rectifiable.

Example 5.2.4. Let $X = (\mathbb{R}^n, \sqrt{\|\cdot\|_{\mathbb{R}^n}})$; a rough look allows us to say that X is a step 2 Carnot group, but we can actually say something more. The metric space X is of "fractal" type, in the sense that for every C^1 curve $\gamma : [a, b] \to \mathbb{R}^n$, the total variaton of γ w.r.t. the distance $\sqrt{\|\cdot\|_{\mathbb{R}^n}}$ is infinite. $\Diamond$

Another important feature of Carnot-Carathéodory distances is that they are geodesic, in the sense that for all $x, y \in \mathbb{G}$ there exists a geodesic connecting them such that the distance between x and y coincides with the length of the geodesic. Let us recall that a continuous rectifiable curve $\gamma : [0, 1] \to (\mathbb{R}^n, d_C)$ is said to be a *geodesic* provided that $\mathrm{Var}(\gamma) = d_C(\gamma(0), \gamma(1))$. However, there is a flip side: d_C is not explicit and so from an analytical point of view it is difficult to carry out estimates; several attempts have been pursued in order to introduce explicit invariant distances and the case of d_∞ is one of them. In order to define it, let $\mathbb{G} = (\mathbb{R}^n, \cdot)$ be a Carnot group, let $p \in \mathbb{G}$ and, recalling the layer coordinates (5.1.3), let us write $p = [p^1, ..., p^k]$. We can first introduce the ∞-norm as follows

$$\|p\|_\infty := \max_{j=1,\dots,k} \{\varepsilon_j \|p^j\|^{1/j}_{\mathbb{R}^{m_j}}\}$$

where $\varepsilon_1, ..., \varepsilon_k$ are suitable constants in $]0, 1]$ (in [8] it has been proved that there actually exist $\varepsilon_1, ..., \varepsilon_k$ such that $\|\cdot\|_\infty$ is a homogeneous norm on $\mathbb{G}$ continuous w.r.t. the Euclidean distance and moreover $\varepsilon_1 = 1$). Having such a norm, the ∞-distance can be defined as

$$d_\infty(x, y) = d_\infty(y^{-1} \cdot x, 0) := \|y^{-1} \cdot x\|_\infty$$

for any $x, y \in \mathbb{G}$ and it is clear that d_∞ is left-invariant. Hence, thanks to Proposition 5.1.9, d_∞ is equivalent to any Carnot-Carathéodory distance on $\mathbb{G}$.

Remark 5.2.5. In the case of the Heisenberg group $\mathbb{G} = \mathbb{H}^n$, we can choose $\varepsilon_1 = \varepsilon_2 = 1$, so that if $p = [p^1, p^2] = [(p_1, ..., p_{2n}), p_{2n+1}]$, then $\|p\|_\infty = \max\{\|p^1\|_{\mathbb{R}^{2n}}, |p_{2n+1}|^{1/2}\}$. $\Diamond$

Up to now we have described the differential, algebraic and metric structure of Carnot groups, but in order to carry out analysis on them we need also to single out a volume measure. As a first tool, let us introduce the integer

$$Q := \sum_{i=1}^{k} i \dim(V_i) \tag{5.2.2}$$

recalling that V_i is the i-th layer in the decomposition (5.1.1); such integer is called the *homogeneous dimension* of $\mathbb{G}$ and the reason is explained in the next result.

Proposition 5.2.6. *Let $(\mathbb{G}, \cdot)$ be a Carnot group equipped with an invariant metric d. Then the n-dimensional Lebesgue measure $\mathscr{L}^n$ is the Haar measure of the group $\mathbb{G}$, that is if $E \subset \mathbb{R}^n$ is measurable, then*

$$\mathscr{L}^n(\tau_x(E)) = \mathscr{L}^n(E), \quad \forall x \in \mathbb{G}$$

Moreover, $\mathscr{L}^n$ is Q-homogeneous w.r.t. dilations, namely for any $\lambda > 0$ it holds

$$\mathscr{L}^n(\delta_\lambda(E)) = \lambda^Q \mathscr{L}^n(E)$$

An easy consequence of the left-invariance and Q-homogeneity of $\mathscr{L}^n$, together with the second part of Proposition 5.1.10, is the following

$$\begin{aligned} \mathscr{L}^n(B_d(p,r)) &= \mathscr{L}^n(U_d(p,r)) = r^Q \mathscr{L}^n(B_d(p,1)) \\ &= r^Q \mathscr{L}^n(B_d(0,1)) \end{aligned} \tag{5.2.3}$$

Actually, $\mathscr{L}^n$ is not only left- but also right-invariant. From this fact we infer that the metric measure space $(\mathbb{R}^n, d_\infty, \mathscr{L}^n)$ is an Ahlfors space of dimension Q and this is still true if we equip $\mathbb{R}^n$ with any other Carnot-Carathéodory distance. The definition of Ahlfors space will be provided soon in the general framework of metric measure spaces. On the contrary, let us now sketch the proof of the proposition above.

Proof. First notice that, by property (i) of Theorem 5.2.2, $\mathscr{L}^n(B_d(p,r)) < +\infty$ for any $p \in \mathbb{G}$ and $r > 0$. Secondly, observe that translations and dilations $\tau_x, \delta_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ are diffeomorphisms and, because of the group structure, it is easy to see that

$$\det J(\tau_x) = 1, \quad \det J(\delta_\lambda) = \lambda^Q$$

where $J(\psi)$ denotes the Jacobian matrix associated to a regular map $\psi : \mathbb{R}^n \to \mathbb{R}^n$. The conclusion then follows. □

As a consequence of (5.2.3) we also deduce that the Hausdorff dimension of a Carnot group equals its homogeneous dimension, but we will better investigate these aspects, as the Ahlfors regularity, in the more general framework of metric measure spaces. For the moment, let us only say that we have another evidence of the fact that sub-Riemannian geometry is very different from Riemannian one: in fact, if the step k of a Carnot group $\mathbb{G}$ is strictly bigger than 1, then the topological dimension n of $\mathbb{G}$ is strictly smaller than its homogeneous dimension (in the Riemannian case, they are always equal). Actually, $n = Q$ if and only if $k = 1$. Thus, in sub-Riemannian geometry we have to leave the Euclidean beliefs we are used to: for instance, regular (i.e. embedded) submanifolds do not keep both topological and metric dimension, because of the gap $n < Q$, and so one of the obstacles we shall overcome is the definition of regular submanifold in sub-Riemannian geometry.

Now let us move to a general metric space, in order to talk about the Hausdorff measure. Let (X, d) denote a separable metric space (separability assumption is required in order to assure the existence of countable coverings) and, for sake of simplicity, let us omit d in the notation for open and closed balls in (X, d). Furthermore, let us assume that the following diameter regularity condition (at small scales) holds: there exist constants $0 < \rho_0 \leq 2$ and $\delta_0 > 0$ such that, for all $0 < r < \delta_0$ and for all $x \in X$, we have

$$\operatorname{diam} B(x,r) = \rho_0 r$$

Indeed, at least in Carnot groups, such hypothesis is crucial in order to perform the differentiation of the Hausdorff measure and in Carnot groups we already know that this condition is satisfied with $\rho_0 = 2$, thanks to Proposition 5.1.10. The upper bound $\rho_0 \leq 2$ is not restrictive at all, because by the triangle inequality $d(y,z) \leq 2r$ for any $y, z \in B(x,r)$. For $m > 0$, let us denote

$$\alpha_m := \frac{\pi^{m/2}}{\Gamma(m/2+1)}$$

being Γ the Euler function and

$$\beta_m := \alpha_m \rho_0^{-m} \tag{5.2.4}$$

After this introduction, we can now define the Hausdorff measure or, more precisely, the Hausdorff measures, since three definitions are possible.

Definition 5.2.7. *Let (X,d) be a metric space, $m \in [0,+\infty[$, $A \subset X$ and β_m be the constant defined in (5.2.4). Then:*

*(i) the m-*dimensional Hausdorff measure *of A is defined as*

$$\mathscr{H}^m(A) := \lim_{\delta \to 0} \mathscr{H}^m_\delta(A)$$

where $\mathscr{H}^m_\delta(A)$ is the δ-approximating m-dimensional Hausdorff measure of A *(it is a premeasure), given by*

$$\mathscr{H}^m_\delta(A) := \inf\left\{ \sum_{i=1}^\infty \beta_m (\operatorname{diam} E_i)^m \,:\, \bigcup_{i=1}^\infty E_i \supset A,\ \operatorname{diam} E_i \leq \delta \right\}$$

*(ii) the m-*dimensional spherical Hausdorff measure *of A is defined as*

$$\mathscr{S}^m(A) := \lim_{\delta \to 0} \mathscr{S}^m_\delta(A)$$

where $\mathscr{S}^m_\delta(A)$ is the δ-approximating m-dimensional spherical Hausdorff measure of A *(it is a premeasure), given by*

$$\mathscr{S}^m_\delta(A) := \inf\left\{ \sum_{i=1}^\infty \beta_m (\operatorname{diam} B(x_i,r_i))^m \,: \right.$$
$$\left. \bigcup_{i=1}^\infty B(x_i,r_i) \supset A,\ \operatorname{diam} B(x_i,r_i) \leq \delta \right\}$$

*(iii) the m-*dimensional centered Hausdorff measure *of A is defined as*

$$\mathscr{C}^m(A) := \sup_{B \subset A} \mathscr{C}^m_0(B)$$

where

$$\mathscr{C}^m_0(B) := \lim_{\delta \to 0} \mathscr{C}^m_\delta(B)$$

and in turn $\mathscr{C}^m_\delta(B)$ *is the* δ-approximating m-dimensional centered Hausdorff measure of B *(it is a premeasure), given by* $\mathscr{C}^m_\delta(B) = 0$ *if* $B = \emptyset$ *and by*

$$\mathscr{C}^m_\delta(B) := \inf\Bigg\{ \sum_{i=1}^\infty \beta_m (\operatorname{diam} B(x_i, r_i))^m \ : \ \bigcup_{i=1}^\infty B(x_i, r_i) \supset B,\, x_i \in B,\, \operatorname{diam} B(x_i, r_i) \le \delta \Bigg\}$$

otherwise.

The first two definitions, even if called Hausdorff measures, are obtained via a classical procedure already outlined by Carathéodory in 1910 and indeed few years later Hausdorff acknowledged that the original definitions were due to the Greek mathematician. Hence it is well known that $\mathscr{H}^m$ and $\mathscr{S}^m$ are both metric (outer) measures and, in particular, also Borel measures in any metric space, since this is true for any metric measure, as proved by Carathéodory still in 1910; let us recall that an outer measure $\mu : \mathcal{P}(X) \to [0, +\infty]$ is said to be *metric* provided that

$$\mu(A \cup B) = \mu(A) + \mu(B), \quad \text{if } d(A, B) > 0$$

On the contrary, the third and maybe less known definition is really due to Hausdorff. Also $\mathscr{C}^m$ is a metric measure in any metric space, but this fact is not as immediate as for $\mathscr{H}^m$ and $\mathscr{S}^m$.

Remark 5.2.8. Let us point out that $\mathscr{C}^m_0 : \mathcal{P}(X) \to [0, +\infty]$ is not increasing. This means that if $A \subset B$, then it is not always true that $\mathscr{C}^m_0(A) \le \mathscr{C}^m_0(B)$. $\Diamond$

Notice that $\mathscr{H}^m$, $\mathscr{S}^m$ and $\mathscr{C}^m$ are mutually absolutely continuous, because

$$\mathscr{H}^m \le \mathscr{S}^m \le \mathscr{C}^m \le 2^m \mathscr{H}^m \tag{5.2.5}$$

The first and the second inequalities are trivial, since from left to right the infimum is taken on smaller and smaller sets; for the third one, it suffices to observe that, thanks to our separability assumption, it is possible to choose $B \subset A$ such that a subset of diameter at most δ can be covered by a ball of radius equal to the diameter of the subset and centered at a point of B. As a digression, let us recall one more time that in a general metric setting we can only say that the diameter of a ball is smaller than twice the radius; in Carnot-Carathéodory spaces the diameter is exactly equal to twice the radius (as already stated in Proposition 5.1.10). It is also worth saying that, even in the Euclidean case, $\mathscr{H}^m$ and $\mathscr{S}^m$ may be different: a counterexample is given by the Sierpinski carpet (see Figure 5.2). On the contrary, an important result due to Federer states that for rectifiable sets $\mathscr{H}^m$ and $\mathscr{S}^m$ agree. As a next step, let us provide the notion of Hausdorff dimension.

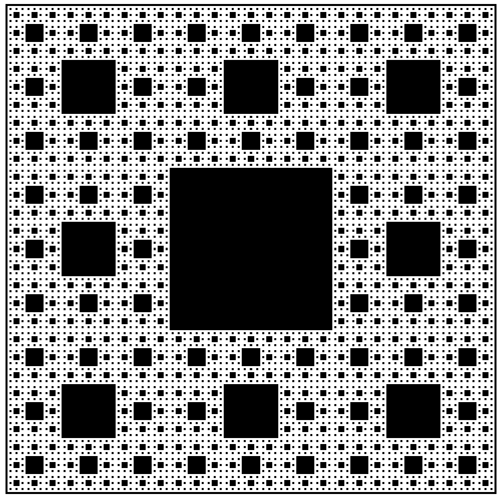


Figure 5.2: Sierpinski carpet

Definition 5.2.9. *Let (X, d) be a separable metric space and let $A \subset X$. The* Hausdorff dimension *(also called* metric dimension*) of A is defined as*

$$\mathrm{Hdim}(A) := \inf\{m \in [0, +\infty[: \mathscr{H}^m(A) = 0\}$$

The Hausdorff dimension has the natural properties of monotonicity and stability w.r.t. the countable union, that is for $A, B, A_i \subset X$ we have

$$A \subset B \quad \Rightarrow \quad \mathrm{Hdim}(A) \leq \mathrm{Hdim}(B)$$
$$\mathrm{Hdim}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sup_{i\in\mathbb{N}} \mathrm{Hdim}(A_i) \tag{5.2.6}$$

Since $\mathscr{H}^m$, $\mathscr{S}^m$ and $\mathscr{C}^m$ are equivalent, the notion of Hausdorff dimension can be equivalently restated w.r.t. $\mathscr{S}^m$ or $\mathscr{C}^m$. As well, the notion of metric dimension is stable w.r.t. equivalent metrics on X. However, one of the main difficulties in geometric measure theory is exactly the computation of the Hausdorff dimension: for instance, the graph of the Weierstrass function is a fractal set whose Hausdorff dimension is still unknown (we only know that it is between 1 and 2). In this setting, the notion of Ahlfors regularity turns out to be particularly important, because under this assumption the Hausdorff dimension coincides with the Ahlfors one; thus, let us provide the following definition.

Definition 5.2.10. *A metric measure space (X, d, μ), where μ is a Radon measure on X, is said to be* Ahlfors regular *of dimension Q (with $Q \in [0, +\infty[$) if there exist positive constants a_1, a_2 such that*

$$a_1 r^Q \leq \mu(B(x, r)) \leq a_2 r^Q, \quad \forall x \in X, \forall r \in]0, \mathrm{diam}\, X[$$

The number Q is called the Ahlfors dimension *of X.*

With this definition, let us now state a result just mentioned above.

Theorem 5.2.11. *Let (X, d, μ) be an Ahlfors regular metric measure space of dimension Q. Then* $\mathrm{Hdim}(X) = Q$.

Proof. As the three Hausdorff measures we presented previously are equivalent, it is sufficient to prove that

$$0 < \mathscr{S}^Q(B(x_0, r)) < +\infty \tag{5.2.7}$$

for all $x_0 \in X$ and for all $r > 0$, because this fact implies that the Hausdorff dimension of any ball is exactly Q. Indeed, since

$$X = \bigcup_{h=1}^{\infty} B(x_0, h)$$

by virtue of (5.2.6) we infer that

$$\mathrm{Hdim}(X) = \sup_{h \in \mathbb{N}} \mathrm{Hdim}(B(x_0, h)) = Q$$

The proof of (5.2.7) relies now on a classical covering theorem in metric spaces (the forthcoming Theorem 5.2.12), which is an extension of the well-known Vitali covering theorem; the details can be found in [22], Theorem 2.23. □

Theorem 5.2.12. *Let (X, d) be a separable metric space and let $\mathcal{F}$ be an arbitrary family of balls in X such that*

$$\sup_{B \in \mathcal{F}} \operatorname{diam} B < +\infty$$

Then there exists a countable disjoint subfamily $\mathcal{F}' \subset \mathcal{F}$ such that

$$\bigcup_{B \in \mathcal{F}} B \subset \bigcup_{B \in \mathcal{F}'} 5B$$

where $5B$ denotes the ball with same centre as B and radius five times the radius of B.

As a consequence of Theorem 5.2.11 and of (5.2.3) we deduce the following corollary, first proved by Mitchell in [16] in 1985.

Corollary 5.2.13. *Let $\mathbb{G}$ be a Carnot group of homogeneous dimension Q. Then* $\mathrm{Hdim}(\mathbb{G}) = Q$.

In what follows we will be particularly interested in understand the structure of the “useful” submanifolds of a Carnot group and it is worth stressing that in any metric space there always exist 1-dimensional regular submanifolds (even if by now “regular submanifold” has no meaning in a metric setting), namely the Lipschitz curves. For this reason, let us estimate the metric dimension of their graphs.

Theorem 5.2.14. *Let (X, d) be a separable metric space, $\gamma : [a, b] \to X$ be a Lipschitz curve and $\Gamma = \gamma([a, b])$. Then*

$$\mathscr{H}^1(\Gamma) \le \mathscr{S}^1(\Gamma) \le \mathscr{C}^1(\Gamma) \le \mathrm{Var}(\gamma) < +\infty$$

and equality holds if γ is injective.

Also in this case, the proof can be found in [22], Theorem 2.26.

5.3 Lecture 3 - 10 September

Abstract

The first (and main) part of this lecture is devoted to the relationships between the different types of Hausdorff measure and the Lebesgue one in the framework of Carnot groups; in particular, we will point out the importance of the isodiametric problem. In the second part, we will approach a new subject: differential calculus between Carnot groups.

In the previous lecture we talked about Hausdorff measures and we provided a threefold definition; in particular, even if less studied, the third version is really useful in the study of very irregular sets. Now let us turn our attention to some applications of Hausdorff measures and metric dimension issues to Carnot groups; to this aim, let us fix our framework: $X = (\mathbb{G}, \cdot)$ is a Carnot group and d an invariant distance on it.

As a first remark, in this environment all the three Hausdorff measures $\mathscr{H}^m$, $\mathscr{S}^m$, $\mathscr{C}^m$ are left-invariant and m-homogeneous, but more important is the following problem: when do $\mathscr{H}^m$, $\mathscr{S}^m$ and $\mathscr{C}^m$ agree? The relevance of the problem is due to the fact that if the three Hausdorff measures coincide, then important geometric features follow (as, for instance, the isodiametric inequality). For this reason, let us investigate the agreement between $\mathscr{H}^Q$, $\mathscr{S}^Q$, $\mathscr{C}^Q$ and $\mathscr{L}^n$, where Q is the homogeneous dimension of $\mathbb{G}$ and $\mathscr{L}^n$ is the n-dimensional Lebesgue measure. We already know that the four measures are Haar measures of the group, because they are left-invariant Radon measures. Therefore, it is well-known that they differ only for a multiplicative constant. By fixing $\mathscr{S}^Q$ as reference measure, this implies that

$$\mathscr{H}^Q = c_1 \mathscr{S}^Q, \quad \mathscr{C}^Q = c_2 \mathscr{S}^Q, \quad \mathscr{L}^n = c_3 \mathscr{S}^Q \tag{5.3.1}$$

and the problem now becomes: can c_1, c_2, c_3 be equal to 1? Our attention will be firstly focus on c_2, starting with the following theorem.

Theorem 5.3.1. *Let $\mathbb{G}$ be a Carnot group of homogeneous dimension Q, equipped with an invariant distance d. Then the following facts hold:*

(i) for each ball $B \subset \mathbb{G}$, $\mathscr{S}^Q(B) = \beta_Q (\mathrm{diam}\, B)^Q$, where β_Q is the normalization constant defined in (5.2.4);

(ii) $\mathscr{S}^Q = C_d \mathscr{H}^Q$*, where* C_d *is the* isodiametric constant*, namely*

$$C_d := \sup \left\{ \frac{\mathscr{S}^Q(A)}{\beta_Q (\operatorname{diam} A)^Q} \; : \; 0 < \operatorname{diam} A < +\infty \right\}$$

(iii) for each ball $B \subset \mathbb{G}$*,* $\mathscr{C}^Q(B) = \beta_Q(\operatorname{diam} B)^Q$*.*

Proof. For (i) and (ii) see respectively Proposition 2.1 and 2.3 of [21]. The only comment we make on these two properties is the following: in (i), the fact that $\mathscr{S}^Q(B) \leq \beta_Q(\operatorname{diam} B)^Q$ is trivial by definition, whereas the opposite inequality follows by a covering argument (similar to the one of Vitali's theorem). For (iii), just observe that from (i) and the very definition of $\mathscr{C}^Q$ we infer that

$$\beta_Q(\operatorname{diam} B)^Q = \mathscr{S}^Q(B) \leq \mathscr{C}^Q(B) \leq \beta_Q(\operatorname{diam} B)^Q$$

and so the conclusion. □

As a consequence of this theorem, we get that $c_2 = 1$ in (5.3.1) and thus $\mathscr{S}^Q$ and $\mathscr{C}^Q$ coincide. Indeed, the fact that $\mathbb{G}$ is a Carnot group allows us to apply Lebesgue differentiation theorem, as we will better see in the next corollary. Moreover, $c_1 = C_d^{-1}$ but the agreement of $\mathscr{S}^Q$ and $\mathscr{H}^Q$ is much more delicate.

Corollary 5.3.2. *Under the same assumptions of Theorem 5.3.1, it holds that* $\mathscr{S}^Q = \mathscr{C}^Q$*.*

Proof. From Theorem 5.3.1 we know that $\mathscr{S}^Q$ and $\mathscr{C}^Q$ agree on balls. On the other hand, $\mathscr{S}^Q$ and $\mathscr{C}^Q$ are doubling Radon measures; thus, by fundamental results in harmonic analysis, Lebesgue differentiation theorem holds for both measures. This means that for any $f \in L^1_{loc}(\mathbb{R}^n, \mathscr{C}^Q)$ we have

$$\exists \lim_{r \to 0} \fint_{B(x,r)} f(y) d\mathscr{C}^Q(y) = f(x), \quad \text{for } \mathscr{C}^Q\text{-a.e. } x \in \mathbb{R}^n$$

and an analogous statement is true for $\mathscr{S}^Q$. By choosing $f \in L^1_{loc}(\mathbb{R}^n, \mathscr{C}^Q)$ as the Radon-Nikodym derivative of $\mathscr{S}^Q$ w.r.t. $\mathscr{C}^Q$, we entail that

$$\frac{d\mathscr{S}^Q}{d\mathscr{C}^Q}(x) = \lim_{r \to 0} \frac{\mathscr{S}^Q(B(x,r))}{\mathscr{C}^Q(B(x,r))}, \quad \text{for } \mathscr{C}^Q\text{-a.e. } x \in \mathbb{R}^n$$

and from property (iii) of Theorem 5.3.1 we have that

$$\frac{\mathscr{S}^Q(B(x,r))}{\mathscr{C}^Q(B(x,r))} = 1$$

Therefore

$$\frac{d\mathscr{S}^Q}{d\mathscr{C}^Q}(x) = 1, \quad \text{for } \mathscr{C}^Q\text{-a.e. } x \in \mathbb{R}^n$$

and this implies that $\mathscr{S}^Q = \mathscr{C}^Q$ on Borel sets. Being $\mathscr{S}^Q$ and $\mathscr{C}^Q$ Borel regular, through a standard approximation argument we can extend the equality to all subsets of $\mathbb{G}$ and the proof is thus accomplished. □

Let us stress that at this level the value of β_Q is not relevant, because aim of the previous theorem and corollary was to establish the agreement between $\mathscr{S}^Q$ and $\mathscr{C}^Q$. Hence for the moment β_Q is just a constant. Its degree of freedom is essential in order to make $\mathscr{L}^n$ and $\mathscr{S}^Q$ coincide, because if we define

$$\beta_Q := \frac{\mathscr{L}^n(B_d(0,1))}{2^Q}$$

then $c_3 = 1$ and $\mathscr{S}^Q = \mathscr{L}^n$; notice that this definition of β_Q is different from the one given in (5.2.4): indeed the latter is meaningful in the Euclidean case, as we will better see later, but in the sub-Riemannian case there is no particular reason to choose (5.2.4). On the contrary, the agreement between $\mathscr{S}^Q$ and $\mathscr{H}^Q$ is very delicate and now we will deepen this topic, also known as *isodiametric problem*, in the framework of Carnot groups. Let us first recall that in the Euclidean case

$$\begin{aligned}&\sup\left\{\frac{\mathscr{L}^n(A)}{(\operatorname{diam} A)^n} \,:\, 0 < \operatorname{diam} A < +\infty\right\}\\ = \quad &\max\left\{\frac{\mathscr{L}^n(A)}{(\operatorname{diam} A)^n} \,:\, 0 < \operatorname{diam} A < +\infty\right\} = \beta_n\end{aligned}$$

with β_n defined as in (5.2.4) with $\rho_0 = 2$, which is equivalent to say that balls maximize the so-called *sharp isodiametric inequality*, i.e.

$$\mathscr{L}^n(A) \le \beta_n (\operatorname{diam} A)^n, \quad \forall A \subset \mathbb{R}^n$$

and the geometric meaning of this fact is clear: balls maximize the volume among all sets with prescribed diameter. Such inequality has many interesting consequences; one of them is the fact that

$$\mathscr{L}^n = \mathscr{H}^n$$

where $\mathscr{H}^n$ is the n-dimensional Hausdorff measure defined w.r.t. the Euclidean distance (thus, this identity may be false w.r.t. a sub-Riemannian distance). The proof of the sharp isodiametric inequality is based on Stein's symmetrization argument and even in the Euclidean case an alternative procedure is not knwon yet; this is one of the main difficulties in dealing with the isodiametric problem in sub-Riemannian geometry, since there Stein's argument cannot be applied. In more details, in Carnot groups the sharp isodiametric inequality can be written as

$$\mathscr{S}^Q(A) \le \beta_Q (\operatorname{diam} A)^Q, \quad \forall A \subset \mathbb{R}^n \tag{5.3.2}$$

because by property (i) of Theorem 5.3.1 we already know that in the case of balls equality holds in (5.3.2). In particular, this implies that $C_d = 1$ and, for this reason, we can observe that the inequality above is equivalent to say that

$$\mathscr{S}^Q = \mathscr{H}^Q$$

because from (5.2.5), from property (ii) of Theorem 5.3.1 and from what we have just said, we get that

$$\mathscr{H}^Q \leq \mathscr{S}^Q = C_d \mathscr{H}^Q = \mathscr{H}^Q$$

Unfortunately, in general the sharp isodiametric inequality does not hold in Carnot groups. Indeed, in [21] it has been proved that given a non-abelian Carnot group $\mathbb{G}$ one can always find some invariant distance on $\mathbb{G}$ for which the isodiametric inequality (5.3.2) does not hold, even if C_d is always a maximum. This means that the maximum is attained by a suitable compact set $K \subset \mathbb{R}^n$ which is not a ball and therefore

$$C_d = \frac{\mathscr{S}^Q(K)}{\beta_Q (\operatorname{diam} K)^Q} > 1$$

The sets attaining the maximum in the definition of C_d are called *isodiametric sets*; their characterization and regularity is still an open problem, also in the Heisenberg group (in this case, some partial results have been proved under symmetry assumption on the isodiametric sets).

To sum up, we have proved that

$$\mathscr{S}^Q = \mathscr{C}^Q = \mathscr{L}^n$$

whereas in (almost) every Carnot group $c_1 > 1$ and thus $\mathscr{H}^Q$ differs from the other Hausdorff measures and from $\mathscr{L}^n$.

Remark 5.3.3. In [21] it has been proved that if $C_d > 1$, then a Carnot group $(\mathbb{G}, d)$ of step $k > 1$ is purely Q-unrectifiable, that is for any Lipschitz function $f : A \subset (\mathbb{R}^Q, \|\cdot\|_{\mathbb{R}^Q}) \to (\mathbb{G}, d)$ it holds

$$\mathscr{H}^Q_d(f(A)) = 0$$

The notion of pure Q-unrectifiability is due to Federer. In the case of the first Heisenberg group $\mathbb{H}^1$, whose homogeneous dimension is $Q = 4$, the pure Q-unrectifiability was already known thanks to the work of Ambrosio and Kirchheim [4]. $\diamondsuit$

This concludes the introductory part on Carnot groups, so that now we can turn our attention to the differential calculus between them. Let us begin this new part with a consideration: taking into account Cheeger's paper [7], one of the most important results in calculus, among those linked to the geometric structure of the spaces involved, is Rademacher's theorem, because it is really from this particular result that calculus follows, e.g. Sobolev spaces and chain rules. Thus, in this perspective the key result of Pansu is a cornerstone: roughly speaking, it states that any Lipschitz function between Carnot groups can be "linearized". But what is a linear function between Carnot groups? Let us begin with the following definition, also due to Pansu (actually, all the forthcoming notions and results are due to or inspired by him).

Definition 5.3.4. *Let $\mathbb{G}_1, \mathbb{G}_2$ be Carnot groups with homogeneous norms $\|\cdot\|_1$, $\|\cdot\|_2$ and dilations $\delta^1_\lambda, \delta^2_\lambda$. A map $L : \mathbb{G}_1 \to \mathbb{G}_2$ is said to be H-*linear *("H" stands for "horizontal"), or a* homogeneous homomorphism*, if L is a group homomorphism such that*

$$L(\delta^1_\lambda x) = \delta^2_\lambda L(x), \quad \forall x \in \mathbb{G}_1, \forall \lambda > 0$$

Now the crucial step consists in providing a characterization of H-linear maps between Carnot groups, but first let us fix the notations. Given a Carnot group $(\mathbb{G}, \cdot)$ and its Lie algebra $\mathfrak{g}$, we shall denote by

$$\mathfrak{g} = V_1 \oplus ... \oplus V_k, \quad \mathbb{G} = \mathbb{V}_1 \oplus ... \oplus \mathbb{V}_k$$

their stratifications, according to (5.1.1) and (5.1.4).

Theorem 5.3.5. *Let $(\mathbb{G}_1, \|\cdot\|_1)$ and $(\mathbb{G}_2, \|\cdot\|_2)$ be two Carnot groups with Lie algebras respectively given by $\mathfrak{g}_1$ and $\mathfrak{g}_2$ and stratifications*

$$\begin{aligned} \mathfrak{g}_1 &= V_1 \oplus ... \oplus V_k, & \mathfrak{g}_2 &= W_1 \oplus ... \oplus W_\ell \\ \mathbb{G}_1 &= \mathbb{V}_1 \oplus ... \oplus \mathbb{V}_k, & \mathbb{G}_2 &= \mathbb{W}_1 \oplus ... \oplus \mathbb{W}_\ell \end{aligned}$$

Let $L : \mathbb{G}_1 \to \mathbb{G}_2$ be a group homomorphism. Then the following are equivalent:

(i) L is H-linear;

(ii) $L : (\mathbb{G}_1, \|\cdot\|_1) \to (\mathbb{G}_2, \|\cdot\|_2)$ is Lipschitz and satisfies the contact property

$$L(\mathbb{V}_j) \subset \mathbb{W}_j, \quad j = 1, ..., k \tag{5.3.3}$$

where if $\ell < k$ we mean that $L(\mathbb{V}_j) = \{0\}$ for $\ell < j \leq k$.

Moreover, if L is H-linear, then L can be read between the Lie algebras as $\tilde{L} = \exp^{-1} \circ L \circ \exp : \mathfrak{g}_1 \to \mathfrak{g}_2$ (see the diagram below)

$$\begin{array}{ccc} \mathbb{G}_1 & \xrightarrow{L} & \mathbb{G}_2 \\ \exp\uparrow & & \uparrow\exp \\ \mathfrak{g}_1 & \xrightarrow{\tilde{L}} & \mathfrak{g}_2 \end{array}$$

and $\tilde{L}$ is a Lie algebra homomorphism: this means that $\tilde{L}$ is linear and

$$\tilde{L}([X, Y]) = [\tilde{L}(X), \tilde{L}(Y)], \quad \forall X, Y \in \mathfrak{g}_1$$

Hence to any H-linear map between Carnot groups we are able to associate a linear map between the corresponding Lie algebras: this is the sort of linearization we were looking for. Let us now introduce the notion of differentiability *à la Pansu*, also called P-differentiability.

Definition 5.3.6. *Let* $(\mathbb{G}_1, d_1)$ *and* $(\mathbb{G}_2, d_2)$ *be two Carnot groups,* $\mathcal{A} \subset \mathbb{G}_1$ *be a subset,* $f : \mathcal{A} \to \mathbb{G}_2$ *be a map and* $g_0 \in \mathcal{A}$*. We say that* f *is* P-differentiable at g_0 *if there exists an* H*-linear map* $L : \mathbb{G}_1 \to \mathbb{G}_2$ *such that*

$$\lim_{g \to g_0} \frac{d_2(f(g_0)^{-1} \cdot f(g), L(g_0^{-1} \cdot g))}{d_1(g, g_0)} = 0$$

The H*-linear function* L *is called the* P-differential of f in g_0 *and is denoted by* $d_P f(g_0)$.

Let us point out that, as in the Euclidean case, if f is P-differentiable at a given point g_0, then $d_P f(g_0) : \mathbb{G}_1 \to \mathbb{G}_2$ is unique and f is continuous at g_0. Moreover, P-differentiability applies to a fundamental result, namely the so-called Pansu-Rademacher theorem for Lipschitz functions between Carnot groups.

Theorem 5.3.7 (Pansu-Rademacher)**.** *Let* $(\mathbb{G}_1, d_1)$ *and* $(\mathbb{G}_2, d_2)$ *be two Carnot groups,* $\mathcal{A} \subset \mathbb{G}_1$ *be an open set and* $f : \mathcal{A} \to \mathbb{G}_2$ *be a Lipschitz function. Then* f *is* P*-differentiable a.e. in* $\mathcal{A}$ *w.r.t. the Haar measure of* $\mathbb{G}_1$*.*

Proof. See [19]. □

For sake of information, an extension of this result is possible; indeed, $\mathcal{A}$ can be assumed to be only measurable, but we will not need this level of generality. Now, thanks to Pansu-Rademacher we can finally carry out calculus on Carnot groups. As a first step, a notational digression: let us denote by $\mathcal{L}_H(\mathbb{G}_1, \mathbb{G}_2)$ the set of H-linear functions from $(\mathbb{G}_1, \|\cdot\|_1)$ to $(\mathbb{G}_2, \|\cdot\|_2)$ and let us endow it with the norm $\|\cdot\|_{\mathcal{L}_H(\mathbb{G}_1,\mathbb{G}_2)}$ defined as follows, for $L : \mathbb{G}_1 \to \mathbb{G}_2$

$$\|L\|_{\mathcal{L}_H(\mathbb{G}_1,\mathbb{G}_2)} := \sup\{\|L(p)\|_2 \ : \ p \in \mathbb{G}_1,\ \|p\|_1 \leq 1\}$$

With this definition, if Ω is an open set in $\mathbb{G}_1$, then we shall denote by $C^1_H(\Omega, \mathbb{G}_2)$ the set of P-differentiable functions $f : \Omega \to \mathbb{G}_2$ such that $d_P f : \Omega \to \mathcal{L}_H(\mathbb{G}_1, \mathbb{G}_2)$ is continuous.

In general a complete characterization of H-linear maps between Carnot groups is not easy, but there is a particularly simple case. In fact, when $\mathbb{G}_1 \equiv \mathbb{G}$ and $\mathbb{G}_2 \equiv \mathbb{R}$, we have the forthcoming result. Before the statement, let us just point out that previous notations can be simplified a little: $\mathcal{L}_H(\mathbb{G}, \mathbb{R})$ will be also denoted by $\mathcal{L}_\mathbb{G}$, $C^1_H(\mathbb{G}_1, \mathbb{G}_2)$ by $C^1_\mathbb{G}(\Omega)$ and a map $L \in \mathcal{L}_\mathbb{G}$ will be simply called $\mathbb{G}$-linear; moreover, a map $L : \mathbb{G} \to \mathbb{R}$ is $\mathbb{G}$-linear if and only if it is a group homomorphism from $(\mathbb{R}^n, \cdot)$ to $(\mathbb{R}, +)$ and it is positively homogeneous of degree 1 w.r.t. the dilations of $\mathbb{G}$, i.e.

$$L(\delta_\lambda x) = \lambda L(x), \quad \forall x \in \mathbb{G}, \forall \lambda > 0$$

If we fix a basis $X_1, ..., X_n$ for $\mathfrak{g}$, then all $\mathbb{G}$-linear maps can be represented as follows.

Proposition 5.3.8. *A map $L : \mathbb{G} \to \mathbb{R}$ is H-linear if and only if there exists $a = (a_1, ..., a_{m_1}) \in \mathbb{R}^{m_1}$ (recall that m_1 is the dimension of the first layer in (5.1.1)) such that, if $x = (x_1, ..., x_n) \in \mathbb{G}$, then $L(x) = \sum_{i=1}^{m_1} a_i x_i$.*

Thus, in this case $\mathcal{L}_{\mathbb{G}}$ coincides with the set of linear functions from $\mathbb{G}$ to $\mathbb{R}$. Moreover, the notion of P-differentiability (at a given point x_0) is particularly simple, because it reduces to require the existence of a suitable H-linear map $L : \mathbb{G} \to \mathbb{R}$ such that

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - L(x_0^{-1} \cdot x)}{d(x, x_0)} = 0$$

It is also simple to see that Pansu differentiability implies the existence of the horizontal derivatives of f at x_0 and so in particular

$$(X_j f)(x_0) = a_j, \quad j = 1, ..., m_1$$

As a last remark, notice that $C^1(\Omega) \subset C^1_{\mathbb{G}}(\Omega)$, where $C^1(\Omega)$ denotes the set of C^1-functions in the Euclidean sense, and the inclusion is strict, because in general the functions in $C^1_{\mathbb{G}}(\Omega)$ are very irregular: not only they need not be Lipschitz, but they may even fail to be BV. Thus, from a Euclidean point of view we can essentially think to $C^1_{\mathbb{G}}$-functions as 1/2-Hölder continuous functions. For this reason, when we require existence for horizontal derivatives, yet we do not know anything on existence for Euclidean derivatives.

5.4 Lecture 4 - 11 September

Abstract

We present the sub-Riemannian counterparts of well-known objects and results of Geometric Measure Theory on $\mathbb{R}^n$, as for instance the horizontal gradient, the horizontal divergence, the area formula and Whitney's extension theorem. By means of these tools, we can begin the study of BV functions.

In the last part of the previous lecture we introduced most of the fundamental notations and definitions in order to handle differential calculus on Carnot groups. In particular, we introduced the set $C^1_H(\Omega, \mathbb{G}_2)$ of P-differentiable functions with continuous P-differential and it would be very pleasant for us to characterize such functions; a general result is due to Valentino Magnani, but here we restrict ourselves to the particular case $(\mathbb{G}_2, \cdot) = (\mathbb{R}, +)$: in this situation, the $C^1_{\mathbb{G}}(\Omega)$-functions are, roughly speaking, those functions admitting continuous horizontal derivatives. In order to make this statement precise, let us provide a further definition.

Definition 5.4.1. *A function $f : \mathbb{G} \to \mathbb{R}$ is said to be* differentiable along X_j *(for $j = 1, ..., m_1$) at x_0 if the map $\lambda \mapsto f(\tau_{x_0}(\delta_\lambda e_j))$ is differentiable at $\lambda = 0$,* *where e_j is the j-th vector of the canonical basis of $\mathbb{R}^n$. In this case, we define*

$$(X_j f)(x_0) := \lim_{\lambda \to 0} \frac{f(\tau_{x_0}(\delta_\lambda e_j)) - f(x_0)}{\lambda}$$

As in the smooth case, at the side of the definition of (strong) derivative we can find a notion of distributional derivative: in this case, assuming that $f \in L^1_{loc}(\mathbb{G})$, we say that $X_j f$ is the *distributional derivative of f along X_j* (also called *weak derivative*) provided that, for all $\psi \in C_0^\infty(\mathbb{G})$,

$$\int_{\mathbb{R}^n} f X_j \psi \, dx = - \int_{\mathbb{R}^n} (X_j f) \psi \, dx \tag{5.4.1}$$

Now fix a system of generating vector fields $X_1, ..., X_{m_1}$ of the Lie algebra $\mathfrak{g}$ and let $f : \mathbb{G} \to \mathbb{R}$ be a function for which the partial derivatives $X_j f$ for $j = 1, ..., m_1$ exist in the sense of Definition 5.4.1. The *horizontal gradient of f*, denoted by $\nabla_{\mathbb{G}} f$, is defined as

$$\nabla_{\mathbb{G}} f := \sum_{i=1}^{m_1} (X_i f) X_i$$

and its canonical coordinates are given by $(X_1 f, ..., X_{m_1} f)$; the adjective *horizontal* is motivated by the fact that $\nabla_{\mathbb{G}} f : \Omega \to H\mathbb{G}$. If we drop off the existence of the horizontal derivatives $X_j f$ in the sense of Definition 5.4.1 and we only assume their existence as distributional derivatives, then we will talk about *distributional gradient of f*. Similarly, if $\varphi = (\varphi_1, ..., \varphi_{m_1})$ is an horizontal section such that $X_j \varphi_j$ exists in the sense of Definition 5.4.1 and $X_j \varphi_j \in L^1_{loc}(\mathbb{G})$ for every $j = 1, ..., m_1$, then we can define the *horizontal divergence of φ* as the real-valued function

$$\text{div}_{\mathbb{G}}(\varphi) := - \sum_{j=1}^{m_1} X_j^* \varphi_j = \sum_{j=1}^{m_1} X_j \varphi_j \tag{5.4.2}$$

where X_j^* is the operator formally adjoint to X_j in $L^2(\mathbb{R}^n)$, that is the operator satisfying

$$\int_{\mathbb{R}^n} \varphi(x) X_j \psi(x) dx = \int_{\mathbb{R}^n} \psi(x) X_j^* \varphi(x) dx$$

for all $\varphi, \psi \in C_0^\infty(\mathbb{R}^n)$. The assumption $X_j \varphi_j \in L^1_{loc}(\mathbb{G})$ is technical in order to give a distributional meaning to (5.4.2). Also in this case, by dropping off the strong regularity of $X_j \varphi_j$ we can talk about *distributional divergence of φ*. Obviously, the definitions of horizontal gradient and divergence can be adapted to functions f and φ defined on an open set $\Omega \subset \mathbb{G}$. Let us stress that the notation we have used for the horizontal gradient is partially imprecise, because $\nabla_{\mathbb{G}}$ really depends on the choice of the generating family $X_1, ..., X_{m_1}$; on the contrary, the horizontal divergence is an intrinsic notion and for this reason the notation $\text{div}_{\mathbb{G}}$ is completely correct. Now let us come back to the characterization of $C^1_{\mathbb{G}}(\Omega)$ and let us talk about projections. More precisely, if $x_0 \in \mathbb{G}$ is given, then we define the map $\pi_{x_0} : \mathbb{G} \to H\mathbb{G}_{x_0}$ by

$$x = (x_1, ..., x_n) \mapsto \pi_{x_0}(x) := \sum_{j=1}^{m_1} x_j X_j(x_0)$$

and we can observe that the map $x_0 \mapsto \pi_{x_0}(x)$ with x fixed is a smooth section of $H\mathbb{G}$. With all this machinery, we can now state a first result.

Proposition 5.4.2. *If $f : \Omega \to \mathbb{G}$ is P-differentiable at x_0, then it is differentiable along X_j at x_0 for $j = 1, ..., m_1$ and*

$$d_P f(x_0)(v) = \langle \nabla_{\mathbb{G}} f, \pi_{x_0}(v) \rangle_{x_0}$$

Proof. See [18], Remark 3.3. □

By means of this proposition, we can finally state in a rigorous way the characterization of continuously P-differentiable real-valued functions on a Carnot group $\mathbb{G}$.

Proposition 5.4.3. *A continuous function $f : \Omega \to \mathbb{R}$ belongs to $C^1_{\mathbb{G}}(\Omega)$ if and only if its distributional derivatives $X_j f$ are continuous in Ω for $j = 1, ..., m_1$.*

Proof. See [9], Proposition 5.8. □

As already anticipated, a generalization of this proposition to the case $f \in C^1_H(\mathbb{G}_1, \mathbb{G}_2)$ is due to Valentino Magnani and it has been recently proved in [13]. Roughly speaking, he proved that $f \in C^1_H(\mathbb{G}_1, \mathbb{G}_2)$ if and only if the horizontal Jacobian of f exists and preserves the contact property (5.3.3). As a byproduct of Magnani's characterization, it follows the next result.

Theorem 5.4.4. *Let $(\mathbb{G}_1, d_1)$ and $(\mathbb{G}_2, d_2)$ be two Carnot groups, let $\Omega \subset \mathbb{G}_1$ be an open set and let $f \in C^1_H(\Omega, \mathbb{G}_2)$. Then f is locally Lipschitz.*

Now let us focus our attention on another important tool in Euclidean geometric measure theory, which turns out to be true also in the setting of Carnot groups: the so-called area formula for Lipschitz maps. It enables us to calculate the area of Lipschitz surfaces and in order to generalize it from the Euclidean setting to the sub-Riemannian one, we first need to introduce the concept of horizontal Jacobian; hence, we still go on with differential calculus on Carnot groups.

Definition 5.4.5. *Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two Carnot groups equipped with the invariant distances d_1 and d_2 respectively, let $L : \mathbb{G}_1 \to \mathbb{G}_2$ be a H-linear map and let Q denote the homogeneous dimension of $\mathbb{G}_1$. The* H-Jacobian of L *is defined by*

$$J_Q(L) := \frac{\mathscr{H}^Q_{d_2}(L(B_1))}{\mathscr{H}^Q_{d_1}(B_1)} \tag{5.4.3}$$

where B_1 denotes the unit open ball in $(\mathbb{G}_1, d_1)$.

Notice that in the Euclidean case (5.4.3) corresponds to the determinant of the Jacobian matrix of L, since this quantity describes the distortion rate of the volume. This definition was the only one we further needed. Now we can state the area formula in the case of Carnot groups.

Theorem 5.4.6 (Area formula)**.** *Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two Carnot groups equipped with the invariant distances d_1 and d_2 respectively, let $\mathcal{A} \subset \mathbb{G}_1$ be a measurable set and let $f : \mathcal{A} \to \mathbb{G}_2$ be a Lipschitz continuous function. Then*

$$\int_{\mathcal{A}} J_Q(d_P f(x)) d\mathscr{H}^Q_{d_1}(x) = \int_{f(\mathcal{A}} N(f, \mathcal{A}, y) d\mathscr{H}^Q_{d_2}(y)$$

where $N(f, \mathcal{A}, y)$ denotes the multiplicity function of f, that is the number of elements of $f^{-1}(\{y\}) \cap \mathcal{A}$.

Also in this case we will skip the proof, but let us point out that this is only a partial result, since we do not know what happens to the submanifolds of $\mathbb{G}_1$ and $\mathbb{G}_2$: indeed, this result involves only the Q-dimensional Hausdorff measure and not the lower dimensional ones. Then, let us mention another important tool, whose applications range from analysis to differential geometry: Whitney's extension theorem. Even in the Euclidean case and still nowadays, such theorem poses many different problems, studied for instance by Charles Feffermann. In the setting of Carnot groups, our aim is to extend a function $f : F \subset \mathbb{G}_1 \to \mathbb{G}_2$ to a continuously P-differentiable map $\tilde{f} : \mathbb{G}_1 \to \mathbb{G}_2$ provided that F is a closed set and a suitable notion of differentiability on F is expressed. Up to now, a positive answer is known only if $\mathbb{G}_2 = \mathbb{R}^k$; in the statement of the next theorem, however, for sake of simplicity we are going to consider $\mathbb{G}_2 = \mathbb{R}$, since the generalization from the 1-dimensional case to the k-dimensional one is immediate.

Theorem 5.4.7 (Whitney's extension theorem)**.** *Let $F \subset \mathbb{G}$ be a closed set and let $f : F \to \mathbb{R}$, $\psi : F \to H\mathbb{G}$ be, respectively, a continuous real function and a continuous horizontal section. Let us set*

$$R(x, y) := \frac{f(x) - f(y) - \langle \psi(y), \pi_y(y^{-1} \cdot x) \rangle_y}{d(y, x)}$$

and, if $K \subset F$ is a compact set,

$$\rho_K(\delta) := \sup\{|R(x, y)| \, : \, x, y \in K, \, 0 < d(x, y) < \delta\}$$

if $\delta > 0$. Let us assume that

$$\lim_{\delta \to 0} \rho_K(\delta) = 0$$

for every compact set $K \subset F$ (this is the analogous of the so-called Whitney's condition*). Then there exists a function $\tilde{f} : \mathbb{G} \to \mathbb{R}$ with $\tilde{f} \in C^1_{\mathbb{G}}(\mathbb{G})$ such that*

$$\tilde{f}|_F = f, \quad \nabla_{\mathbb{G}} \tilde{f}|_F = \psi$$

Proof. See [8], Theorem 5.2. □

Whitney's extension theorem in more general cases, as already said, is still an open problem: even if $\mathbb{G}_1 = \mathbb{R}$ and $\mathbb{G}_2 = \mathbb{H}^1$, we do not know how to extend a curve $\gamma : F \subset \mathbb{R} \to \mathbb{H}^1$ keeping the horizontal derivatives. So for instance in the Heisenberg group we can not say that the Lipschitz rectifiability is equivalent to the C^1 rectifiability, as in the Euclidean case, because we do not have an analogous of Whitney's extension theorem yet.

The discussion on Sobolev spaces and Poincaré inequality will now be almost entirely skipped, since it can be found in the course *Hypoelliptic operators* held by Nicola Garofalo and Fabrice Baudoin at the IHP in Paris during the IHP Trimester *Geometry, Analysis and Dynamics on Sub-Riemannian Manifolds.* Let us spend only few words on this topic. Given a Carnot group $\mathbb{G} = (\mathbb{R}^n, \cdot)$ and a system of generating vector fields $X_1, ..., X_{m_1}$, it is possible to introduce the notion of weak derivative w.r.t. X_j (for $j = 1, ..., m_1$) for functions $f : \Omega \subset \mathbb{G} \to \mathbb{R}$ as already done in (5.4.1) and so, in particular, it is possible to talk about horizontal weak gradients. In addition, we can also introduce Sobolev spaces. Thus, up to now in the sub-Riemannian case we have been able to introduce the C^k spaces and the Sobolev ones, but as we are going to show immediately we can also talk about BV functions and this topic is strongly linked to sets of finite perimeter in Carnot groups. In order to define the notion of bounded variation for a function $f : \Omega \subset \mathbb{G} \to \mathbb{R}$, roughly speaking the idea is the following: just replace the Euclidean divergence by the horizontal one $\mathrm{div}_{\mathbb{G}}$. More precisely, for an open set $\Omega \subset \mathbb{R}^n$ let us denote by $C_c^\infty(\Omega, H\mathbb{G})$ the space of compactly supported smooth sections of $H\mathbb{G}$; if $k \in \mathbb{N}$, $C_c^k(\Omega, H\mathbb{G})$ is defined analogously. With this precisation, the space $BV_{\mathbb{G}}(\Omega)$ is defined as follows.

Definition 5.4.8. *The set $BV_{\mathbb{G}}(\Omega)$ is the family of functions $f \in L^1(\Omega)$ having finite horizontal variation on Ω, denoted by $\|\nabla_{\mathbb{G}} f\|(\Omega)$ and defined by*

$$\|\nabla_{\mathbb{G}} f\|(\Omega) := \sup\left\{ \int_\Omega f(x) \,\mathrm{div}_{\mathbb{G}}\, \varphi(x) dx \;:\; \varphi \in C_c^1(\Omega, H\mathbb{G}),\ |\varphi(x)|_x \le 1 \right\}$$

where dx is the Haar measure on $\mathbb{G}$ and $|\cdot|_x$ is the intrinsic fiber norm. The set $BV_{\mathbb{G},loc}(\Omega)$ is made up by the functions $f \in L^1_{loc}(\Omega)$ satisfying $\|\nabla_{\mathbb{G}} f\|(U) < +\infty$ for any open set U compactly supported in Ω.

In order to better understand this definition, let Ω be a bounded set and let $f \in C^1_{\mathbb{G}}(\Omega)$. Then, by linearity of the integral and integration by parts, we have

$$\begin{aligned} \int_\Omega f \,\mathrm{div}_{\mathbb{G}}\, \varphi \, dx &= \int_\Omega f \sum_{j=1}^{m_1} X_j \varphi_j \, dx = -\int_\Omega \sum_{j=1}^{m_1} (X_j f) \varphi_j \, dx \\ &= -\int_\Omega \sum_{j=1}^{m_1} \frac{X_j f}{|\nabla_{\mathbb{G}} f|} \varphi_j |\nabla_{\mathbb{G}} f| dx \end{aligned}$$

and by taking the supremum on φ we obtain

$$\|\nabla_{\mathbb{G}} f\|(\Omega) = \int_\Omega |\nabla_{\mathbb{G}} f| dx \tag{5.4.4}$$

namely the classical variational representation. Now, recalling the classical theory of BV functions, one of the main result in measure theory is given by Riesz' representation theorem, which holds on any locally compact Hausdorff space, because it is a bridge between this field and functional analysis. In the situation we have described, the measure is given by $\|\nabla_{\mathbb{G}} f\|$, even if for the moment it does not seem to be a measure, because it is defined only on open sets. Nevertheless, by means of the classical Carathéodory procedure this problem can be overcome. In a more detailed way, given $f \in BV_{\mathbb{G},loc}(\Omega)$ we can define an (outer) measure $\mu_f : \mathcal{P}(\Omega) \to [0, +\infty]$ by

$$\mu_f(E) := \inf\{\|\nabla_{\mathbb{G}} f\|(U) \,:\, E \subset U \subset\subset \Omega,\, U \text{ open}\}$$

It is well-known that μ_f, that from now on will be denoted by $\|\nabla_{\mathbb{G}} f\|$, is a Radon measure, that is a Borel regular measure finite on compact sets. By Riesz' representation theorem, the structure theorem for BV functions in the setting of Carnot groups follows; in this case, BV functions will be also called $BV_{\mathbb{G}}$ functions.

Theorem 5.4.9. *Let $f \in BV_{\mathbb{G},loc}(\Omega)$. Then there exists a unique $\|\nabla_{\mathbb{G}} f\|$-measurable horizontal section $\sigma_f : \Omega \to H\mathbb{G}$ such that $|\sigma_f(x)| = 1$ for $\|\nabla_{\mathbb{G}} f\|$-a.e. $x \in \Omega$ and integration by parts formula holds, i.e.*

$$\int_\Omega f(x) \operatorname{div}_{\mathbb{G}} \varphi(x) dx = \int_\Omega \langle \varphi, \sigma_f \rangle d\|\nabla_{\mathbb{G}} f\| \tag{5.4.5}$$

for all $\varphi \in C_c^1(\Omega, H\mathbb{G})$, where $\langle \cdot, \cdot \rangle$ is the scalar product of the sub-Riemannian structure. Moreover, the notion of gradient $\nabla_{\mathbb{G}}$ can be extended from regular functions to functions $f \in BV_{\mathbb{G}}$ defining $\nabla_{\mathbb{G}} f$ as the vector valued measure

$$\nabla_{\mathbb{G}} f := -\sigma_f \llcorner \|\nabla_{\mathbb{G}} f\| = (-\sigma_f^{(1)} \llcorner \|\nabla_{\mathbb{G}} f\|, ..., -\sigma_f^{(m_1)} \llcorner \|\nabla_{\mathbb{G}} f\|)$$

where $\sigma_f^{(j)}$ is the j-th component of σ_f w.r.t. the base $X_1, ..., X_{m_1}$.

Proof. See [11]. □

Note that $-\sigma_f^{(j)} \llcorner \|\nabla_{\mathbb{G}} f\|$ is the distributional derivative of f along X_j in the sense of (5.4.1), because (5.4.5) can be explicitly written as

$$\int_\Omega \sum_{j=1}^{m_1} f X_j \varphi_j \, dx = \int_\Omega \sum_{j=1}^{m_1} \varphi_j \sigma_f^{(j)} d\|\nabla_{\mathbb{G}} f\|$$

and so, for any $j = 1, ..., m_1$, if we choose φ identically zero on all coordinates except for the j-th one, the integration by parts formula (5.4.1) follows, showing that $X_j f = -\sigma_f^{(j)} \llcorner \|\nabla_{\mathbb{G}} f\|$ in distributional sense. As a particular case, observe that in (5.4.4) $\|\nabla_{\mathbb{G}} f\|$ is absolutely continuous w.r.t. the n-dimensional Lebesgue measure, but in general this is not true. The importance of the space $BV_{\mathbb{G}}(\Omega)$ for the calculus of variations is mainly motivated by the following two results. The first one concerns compactness.

Theorem 5.4.10 (Compactness). *The space $BV_{\mathbb{G},loc}(\mathbb{G})$ is compactly embedded in $L^1_{loc}(\mathbb{G})$.*

More generally, the statement above holds also for $L^p_{loc}(\mathbb{G})$, where

$$1 \leq p < \frac{Q}{Q-1}$$

and Q is the homogeneous dimension of $\mathbb{G}$. The second result shows that the perimeter (rather than the Hausdorff measure) is the right notion to have in mind, because it is lower semicontinuous w.r.t. the L^1 topology and so this information, coupled with the previous theorem, allows us to apply the direct methods of calculus of variations. On the contrary, the Hausdorff measure is lower semicontinuous w.r.t. the Hausdorff distance (see (7.3.6) for the definition) and this is very strong.

Theorem 5.4.11 (Lower semicontinuity). *Let $f, f_k \in L^1(\Omega)$, $k \in \mathbb{N}$, be such that $f_k \to f$ in $L^1(\Omega)$. Then*

$$\|\nabla_{\mathbb{G}} f\|(\Omega) \leq \liminf_{k\to\infty} \|\nabla_{\mathbb{G}} f_k\|(\Omega)$$

These two results are essential in order to solve Plateau's problem (existence of minimal surfaces). Indeed, while the problem in dimension 2 was independently solved by Tibor Radó and Jesse Douglas in 1930 for the Euclidean case, the generalization of the problem to higher dimensions (still in the Euclidean case) turns out to be much more difficult to study. Nevertheless, in the case of codimension 1 Ennio De Giorgi developed a general solution, based on the theory of sets of locally finite perimeter, which was in turn introduced by Renato Caccioppoli in 1929; this notion will be investigated at the very beginning of the next lecture in the setting of Carnot groups. Roughly speaking, the idea is the following: a minimizing sequence of sets of finite perimeter is compactly contained in L^1, hence it admits a limit point and by lower semicontinuity such limit point is a minimum of our problem. The flip side of De Giorgi's technique is the loss of regularity, because minima can be very irregular sets.

5.5 Lecture 5 - 15 September

Abstract

The topic of this lecture will be the sets of finite perimeter in Carnot groups and by the following discussion we will see that differential calculus *between* Carnot groups is not sufficient; indeed, we will see that the Euclidean notion of rectifiability is useless for Carnot groups, because Lebesgue-Besicovitch differentiation theorem fails to hold in general. Hence we will further need differential calculus *within* Carnot groups.

Let us begin with the notion of set of locally finite perimeter.

Definition 5.5.1. *A measurable set $E \subset \mathbb{G} \equiv (\mathbb{R}^n, \cdot)$ is of* locally finite $\mathbb{G}$-perimeter *in Ω (also called a* $\mathbb{G}$-Caccioppoli set*) if $\mathbb{1}_E \in BV_{\mathbb{G},loc}(\Omega)$. In this case, we call* perimeter of E *the measure*

$$|\partial E|_{\mathbb{G}} := \|\nabla_{\mathbb{G}} \mathbb{1}_E\|$$

and we call generalized (inward) $\mathbb{G}$-normal to ∂E in Ω *the horizontal section*

$$\nu_E(x) := -\sigma_{\mathbb{1}_E}(x)$$

As a first remark, ν_E is essentially an analytical tool, rather than a geometric one, since it comes from Riesz' representation theorem. Secondly, notice that for ν_E the following integration by parts formula holds

$$\int_E \operatorname{div}_{\mathbb{G}} \varphi \, dx = -\int_\Omega \langle \nu_E, \varphi \rangle d|\partial E|_{\mathbb{G}} \tag{5.5.1}$$

for every $\varphi \in C^1_c(\Omega, H\mathbb{G})$ and, recalling Theorem 5.4.9, this means that ν_E is the distributional gradient of $\mathbb{1}_E$. The identity above can also be seen as Gauss-Green theorem in the setting of Carnot groups and for this reason we call ν_E generalized $\mathbb{G}$-normal. As already stressed for $\nabla_{\mathbb{G}}$, let us point out that the notation $|\partial E|_{\mathbb{G}}$ is slightly imprecise, because it depends on the choice of the generating system $X_1, ..., X_{m_1}$ through the bound $|\varphi|_x \leq 1$ in the definition of horizontal variation. Nevertheless, perimeters induced by different families are equivalent as measures and, as a consequence, the notion of being a $\mathbb{G}$-Caccioppoli set is intrinsic and depends only on the group $\mathbb{G}$.

The previous definition is formulated for any measurable set $E \subset \mathbb{R}^n$, hence also for very irregular ones. Let us see what happens in the case E has a regular boundary.

Proposition 5.5.2. *If E is a $\mathbb{G}$-Caccioppoli set with (Euclidean) C^1 boundary in a given open set $\Omega \subset \mathbb{G}$, then the $\mathbb{G}$-perimeter has the following representation*

$$|\partial E|_{\mathbb{G}}(\Omega) = \int_{\partial E \cap \Omega} \left(\sum_{j=1}^{m_1} \langle X_j, n_E \rangle^2 \right)^{1/2} d\mathscr{H}^{n-1}$$

where $X_1, ..., X_{m_1}$ is a generating system, n_E is the Euclidean unit outward normal to E and $\mathscr{H}^s$ is the Euclidean s-dimensional Hausdorff measure.

Thus the density of the $\mathbb{G}$-perimeter w.r.t. the classical $(n-1)$-dimensional Hausdorff measure is expressed in terms of the projections of the Euclidean unit normal on the horizontal vectors $X_1, ..., X_{m_1}$. In this framework, a very difficult problem to treat is given by the characteristic points of ∂E, that is

$$\operatorname{Char}(\partial E) := \{x \in \partial E \, : \, \langle X_j, n_E \rangle = 0, \, j = 1, ..., m_1\}$$

From PDEs point of view, characteristic points behave like cusps, but from geometric measure theory point of view we are lucky, since we know they are few, in the sense of the following theorem.

Theorem 5.5.3 (Magnani)**.** *If E is a set with Euclidean C^1 boundary in a Carnot group $\mathbb{G}$ equipped by an invariant distance d and with homogeneous dimension Q, then*

$$\mathscr{H}_d^{Q-1}(\mathrm{Char}(\partial E)) = 0$$

Let us now spend few words on the Euclidean setting, since the notions we have seen in this lecture were introduced by De Giorgi in order to attack the problem of existence of minimal surfaces in dimension higher than 2. In this situation, our Carnot group is $\mathbb{G} = (\mathbb{R}^n, \cdot)$, whose stratification can be simply expressed as

$$\mathfrak{g} = \mathrm{span}(\partial_1, ..., \partial_n)$$

and for brevity we will write $|\partial E|$ in place of $|\partial E|_{\mathbb{G}}$. Roughly speaking, we have to think about hypersurfaces (1-codimensional topological surfaces in $\mathbb{R}^n$) as boundaries of suitable sets. Before making this statement more rigorous, let us assume, for sake of simplicity, that $|\partial E|(\Omega) < +\infty$ for any open bounded set $\Omega \subset \mathbb{R}^n$. Now, the fundamental step is the transfer of the problem from geometry to measure theory, by associating to $E \subset \mathbb{R}^n$ the Radon measure $|\partial E|$. Once Plateau's problem has passed from geometry to geometric measure theory, the questions we are interested in become the following: the determination of the structure of the measure $|\partial E|$ and the regularity of its support. Let us briefly recall that, given a metric space (X, d) and an outer measure μ on X, the support of μ is defined by

$$\mathrm{supp}(\mu) := \{x \in X \,:\, \mu(B(x, r)) > 0, \forall r > 0\}$$

We guess that the $\mathrm{supp}(|\partial E|)$ is contained in the topological boundary of E and indeed it is so, for any Carnot group, as we are going to see in the next result.

Lemma 5.5.4. *Given a Carnot group $\mathbb{G}$, we have that* $\mathrm{supp}(|\partial E|_{\mathbb{G}}) \subset \partial E$*.*

Proof. If $x_0 \notin \partial E$, then either $x_0 \in \mathbb{R}^n \setminus \overline{E}$ or $x_0 \in \mathring{E}$. In the first case there exists a neighbourhood U of x_0 such that $U \subset \mathbb{R}^n \setminus \overline{E}$ and so, for any $\varphi \in C_c^1(U, H\mathbb{G})$, it holds

$$\int_E \mathrm{div}_{\mathbb{G}}\, \varphi \, dx = 0$$

because the support of φ is disjoint from E. Taking into account (5.5.1), we get that $|\partial E|_{\mathbb{G}}(U) = 0$ and so, by definition, $x_0 \in \mathbb{R}^n \setminus \mathrm{supp}(|\partial E|_{\mathbb{G}})$. In the second case, the argument is exactly the same. □

Thus, we have been able to obtain a first information on the measure $|\partial E|_{\mathbb{G}}$, but such information is very weak, because it can be shown that in $\mathbb{G} = (\mathbb{R}^n, +)$ there exists a measurable set E_0 such that $|\partial E_0|(\mathbb{R}^n) < +\infty$ but its boundary is veri fat, that is $\mathscr{L}^n(\partial E_0) > 0$. Therefore, De Giorgi's idea was to look for a suitable subset $\partial^* E$ of the topological boundary ∂E (the so-called *reduced boundary* of E) such that

(DG1) $|\partial E|(\mathbb{R}^n \setminus \partial^* E) = 0$, i.e. $|\partial E|$ is concentrated on the reduced boundary;

(DG2) $|\partial E|(A) = \mathscr{H}^{n-1}(A \cap \partial^* E)$ for any $A \in \mathscr{B}(\mathbb{R}^n)$, where $\mathscr{H}^{n-1}$ is the Euclidean $(n-1)$-dimensional Hausdorff measure;

(DG3) $\partial^* E$ is $\mathscr{H}^{n-1}$-rectifiable, namely there exists a sequence of C^1 hypersurfaces $(S_j)_{j\in\mathbb{N}} \subset \mathbb{R}^n$ such that

$$\mathscr{H}^{n-1}\left(\partial^* E \setminus \bigcup_{j=1}^{\infty} S_j\right) = 0$$

How can we find such reduced boundary? As already pointed out, the generalized normal ν_E has most and foremost an analytical meaning; thus, from De Giorgi's point of view it is necessary to provide it also a geometric sense and this is achieved in the next definition, where we define the reduced boundary via three properties that entail the three ones above (at least in the Euclidean setting).

Definition 5.5.5. *Let $E \subset \mathbb{R}^n$ be a set of locally finite perimeter. Then a point x_0 belongs to $\partial^* E$ if*

(i) $|\partial E|(B(x_0, r)) > 0$ *for any* $r > 0$*;*

(ii) x_0 *is a Lebesgue point, that is*

$$\exists \lim_{r\to 0} \fint_{B(x_0,r)} \nu_E(x) d|\partial E|(x) = \nu_E(x_0)$$

(iii) $\|\nu_E(x_0)\|_{\mathbb{R}^n} = 1$.

In the setting of Carnot groups, this definition can be simply translated as follows.

Definition 5.5.6. *Let $\mathbb{G}$ be a Carnot group endowed with an invariant distance d and let $E \subset \mathbb{G}$ be a set of locally finite $\mathbb{G}$-perimeter. Then a point x_0 belongs to $\partial^*_{\mathbb{G}} E$ if*

(i) $|\partial E|(B_d(x_0, r)) > 0$ *for any* $r > 0$*;*

(ii) x_0 *is a Lebesgue point, that is*

$$\exists \lim_{r\to 0} \fint_{B_d(x_0,r)} \nu_E(x) d|\partial E|(x) = \nu_E(x_0)$$

(iii) $|\nu_E(x_0)|_{x_0} = 1$.

where the limit has to be understood as a convergence of the averages of the coordinates of ν_E w.r.t. the chosen moving base of the fibers.

However, the utility of this definition is not obvious. In fact, in the Euclidean setting, it is immediate to show that the perimeter measure is concentrated on the reduced boundary, because, by virtue of Lebesgue-Besicovitch differentiation lemma, given a Radon measure μ, for any $f \in L^1_{loc}(\mu)$ we have

$$\exists \lim_{r\to 0} \fint_{|y-x|<r} f(y) d\mu_E = f(x), \quad \text{for } \mu\text{-a.e. } x$$

and this implies that $|\partial E| = |\partial E| \llcorner \partial^* E$. However, in the case of Carnot groups Vitali covering lemma can not be applied because in general μ is not doubling (e.g. the Hausdorff measure restricted to a set). Even Besicovitch covering lemma fails to hold (we know it is true only in the Riemannian setting) and therefore still nowadays the generalization of Lebesgue-Besicovitch differentiation lemma is an open problem. As a particular result, it has been recently proved that for Heisenberg groups there exists an invariant distance for which Besicovitch covering lemma holds and, thus, also Lebesgue-Besicovitch differentiation lemma. Fortunately, for our purposes such a strong result is not required in the greatest generality, because thanks to a deep asymptotic estimate due to Ambrosio we know that it holds when μ is the perimeter measure.

Lemma 5.5.7. *Let $\mathbb{G}$ be a Carnot group equipped with an invariant distance d and let $E \subset \mathbb{G}$ be a set of locally finite $\mathbb{G}$-perimeter. Then*

$$\nu_E(x) = \lim_{r\to 0} \fint_{B_d(x,r)} \nu_E(y) d|\partial E|_{\mathbb{G}}(y)$$

for $|\partial E|_{\mathbb{G}}$-a.e. $x \in \mathbb{G}$. This means that $|\partial E|_{\mathbb{G}}$-a.e. $x \in \mathbb{G}$ belongs to the reduced boundary $\partial^*_{\mathbb{G}} E$.

Proof. See [2]. □

Thanks to this result, property (DG1) is satisfied, but what about the other properties? For (DG2), a partial result has been achieved by Ambrosio and it is the following theorem.

Theorem 5.5.8. *Given a Carnot group $(\mathbb{G}, \cdot)$ endowed with an invariant distance d, there exists a Borel function $\vartheta : \partial^*_{\mathbb{G}} E \to [c_1, c_2]$ with $0 < c_1 < c_2 < +\infty$ such that*

$$|\partial E|_{\mathbb{G}}(A) = \int_{A\cap\partial^*_{\mathbb{G}} E} \vartheta(x) d\mathscr{H}^{Q-1}_d(x) \tag{5.5.2}$$

for any $A \in \mathscr{B}(\mathbb{G})$, where $\mathscr{H}^{Q-1}_d$ is the $(Q-1)$-dimensional Hausdorff measure w.r.t. d.

Proof. See [2]. □

Notice that in general ϑ is not constant and for this reason the theorem just stated differs from (DG2). Moreover, we said that this result is partial because we would like to know more about ϑ; indeed, this function comes from the Radon-Nikodym theorem and so we do not know whether it can be obtained via a blow-up procedure, namely whether it can be written as

$$\vartheta(x) = \vartheta_0 \lim_{r\to 0} \frac{|\partial E|_{\mathbb{G}}(B_d(x,r))}{r^{Q-1}} \tag{5.5.3}$$

for $|\partial E|_{\mathbb{G}}$-a.e. $x \in \mathbb{G}$. To get this representation, we should be able to prove the uniqueness of the blow-up, but this is still an open problem. Finally, for (DG3) we have to introduce a suitable notion of regular submanifold, because the classical notions are useless; more precisely, we have to give a correct notion of rectifiability and also of regular hypersurfaces. Indeed, the Euclidean rectifiability does not apply, because, as a first remark, if we assume that $\partial^*_{\mathbb{G}} E \neq \emptyset$ and we pick $x_0 \in \partial^*_{\mathbb{G}} E$, then by Theorem 5.5.8 we get that

$$0 < \mathscr{H}_d^{Q-1}(\partial^*_{\mathbb{G}} \cap B(x_0, r)) < +\infty$$

provided that $r > 0$; thus, the right dimension for a possible rectifiability result has to be $Q-1$. As a second remark, in the case of the first Heisenberg group $\mathbb{H}^1$, whose homogeneous dimension is $Q = 4$, we have the following lemma.

Lemma 5.5.9. *There exists a set $N \subset \mathbb{H}^1$ such that $\mathscr{H}_d^3(N) = 0$ and $0 < \mathscr{H}^{2+\varepsilon}(N) < +\infty$ for an appropriate $\varepsilon > 0$, where $\mathscr{H}^s$ is the Euclidean s-dimensional Hausdorff measure and $\mathscr{H}_d^s$ is the s-dimensional Hausdorff measure w.r.t. the invariant distance d.*

This lemma shows that N is $\mathscr{H}_d^{Q-1}$-rectifiable (actually, it is $(Q-1)$-dimensional $\mathbb{H}^1$-rectifiable in the sense of the forthcoming Definition 5.5.11), but it is not $\mathscr{H}^{n-1}$-rectifiable (note that $n-1 = 2$), because in this case its Euclidean Hausdorff dimension would be 2 and this is impossible, since it is $2+\varepsilon$. Having this motivational example in mind, we can now provide the notion of $\mathbb{G}$-regular hypersurface. Roughly speaking, in the Euclidean space a C^1 regular hypersurface is a non-critical level set of a regular function; thus, in the case of Carnot groups it will be sufficient to replace the Euclidean regularity by Pansu differentiability and once this is done, we shall restate property (DG3) in the setting of Carnot groups.

Definition 5.5.10. *Let $\mathbb{G} = (\mathbb{R}^n, \cdot)$ be a Carnot group. We say that $S \subset \mathbb{G}$ is a* $\mathbb{G}$-regular hypersurface *if for each $x \in S$ there exist a neighbourhood U of x and a function $f \in C^1_{\mathbb{G}}(U)$ such that*

(i) $S \cap U = \{y \in U \,:\, f(y) = 0\}$;

(ii) $\nabla_{\mathbb{G}} f(y) \neq 0$ *for any* $y \in U$.

Definition 5.5.11. *A set $E \subset \mathbb{G}$ is said to be* $(Q-1)$-dimensional $\mathbb{G}$-rectifiable *if there exists a sequence $(S_j)_{j\in\mathbb{N}}$ of $\mathbb{G}$-regular hypersurfaces such that*

$$\mathscr{H}_d^{Q-1}\left(E \setminus \bigcup_{j=1}^{\infty} S_j\right) = 0$$

With these definitions, we can now state an important theorem giving us a partial answer to the question whether the version of (DG3) adapted to Definition 5.5.11 is satisfied for Carnot groups. The great limitation comes from the fact that we restrict ourselves to step 2 Carnot groups.

Theorem 5.5.12. *Let $\mathbb{G} = (\mathbb{R}^n, \cdot)$ be a step 2 Carnot group and let $E \subset \mathbb{G}$ be a set of locally finite $\mathbb{G}$-perimeter. Then the following facts hold:*

(i) *$\partial^*_{\mathbb{G}} E$ is $(Q-1)$-dimensional $\mathbb{G}$-rectifiable;*

(ii) *for any $A \in \mathscr{B}(\mathbb{R}^n)$,*

$$|\partial E|_{\mathbb{G}}(A) = \vartheta_0 \mathscr{S}_\infty^{Q-1}(A \cap \partial^*_{\mathbb{G}} E)$$

where ϑ_0 is a constant independent of A such that

$$\vartheta_0 = \lim_{r \to 0} \frac{|\partial E|_{\mathbb{G}}(B(x,r))}{r^{Q-1}}$$

for $|\partial E|_{\mathbb{G}}$-a.e. $x \in \mathbb{G}$. Pay attention to the fact that $\mathscr{S}_\infty^{Q-1}$ is considered w.r.t. the distance d_∞ and this distance can not be replaced by any Carnot-Carathéodory one.

Proof. See [8]. □

Thus, with the previous definitions and this result we have dealt with 1-codimensional submanifolds; in the next lecture we will provide more suitable and intrinsic definitions, so that, even if we do not deal with extension of the theory to higher codimensions, the interested reader will have the required tools for the comprehension.

5.6 Lecture 6 - 16 September

Abstract

In this last lecture (of the first part of the course), the main topic is differential calculus within Carnot groups; motivated by what we have seen in the last part of the previous lecture, i.e. the impossibility to use the Euclidean notion of rectifiability, we will try to define intrinsic C^1 surfaces as well as intrinsic Lipschitz manifolds within Carnot groups.

In this last lecture (of the first part of the course), the main topic is differential calculus within Carnot groups; motivated by what we have seen in the last part of the previous lecture, i.e. the impossibility to use the Euclidean notion of rectifiability, we will try to define intrinsic C^1 surfaces as well as intrinsic Lipschitz manifolds within Carnot groups. By *intrinsic* we mean that the property only depends on the algebraic structure of the group or its Lie algebra. About the definition of a suitable notion of intrinsic C^1 surface, give a look at

Section 4.5 in [22]. Essentially, such definition, together with the one of intrinsic Lipschitz manifold, relies on the notion of graph and for this reason we will now talk about graphs and also complementary subgroups in Carnot groups, because their comprehension is required. As a trivial remark, it is well-known that a Carnot group can not be viewed as a cartesian product of its subgroups.

Definition 5.6.1. *Given a Carnot group* $\mathbb{G}$*, a* homogeneous subgroup $\mathbb{H} \subset \mathbb{G}$ *is a Lie subgroup of* $\mathbb{G}$ *such that it is stable by dilations, that is* $\delta_\lambda(g) \in \mathbb{H}$ *for any* $g \in \mathbb{H}$ *and for any* $\lambda > 0$*.*

Exploiting the usual identification $\mathbb{G} = (\mathbb{R}^n, \cdot)$, any homogeneous subgroup of $\mathbb{G}$ turns out to be a vector subspace of $\mathbb{R}^n$. As a second step, let us provide the definition of topological and metric dimension of a homogeneous subgroup.

Definition 5.6.2. *Let* $\mathbb{H} \subset \mathbb{G}$ *be a homogeneous subgroup with associated Lie algebra* $\mathfrak{h}$*. Then the* topological dimension d_t *of* $\mathbb{H}$ *is defined by*

$$d_t := \dim \mathfrak{h}$$

The metric dimension d_m *is defined by*

$$d_m := \mathrm{Hdim}_d(\mathbb{H})$$

w.r.t. any invariant distance d *on* $\mathbb{G}$*.*

Recall that $d_t \leq d_m$. In order to encode both pieces of information in a single notation, we will say that $\mathbb{H}$ is a (d_t, d_m)-homogeneous subgroup. Finally, as already anticipated, let us recall the notion of complementary subgroup.

Definition 5.6.3. *Given a Carnot group* $\mathbb{G}$*, two homogeneous subgroups* $\mathbb{M}, \mathbb{H} \subset \mathbb{G}$ *are said to be* complementary *if*

(i) $\mathbb{M} \cap \mathbb{H} = \{e\}$*;*

(ii) $\mathbb{G} = \mathbb{M} \cdot \mathbb{H}$*, that is for any* $g \in \mathbb{G}$ *there exist uniquely* $m \in \mathbb{M}$ *and* $h \in \mathbb{H}$ *such that* $g = m \cdot h$

To a couple $\mathbb{M}, \mathbb{H}$ of complementary subgroups we can associate in a natural way two projections:

$$P_{\mathbb{M}} : \mathbb{G} \to \mathbb{M}, \quad P_{\mathbb{M}}(g) := m$$
$$P_{\mathbb{H}} : \mathbb{G} \to \mathbb{H}, \quad P_{\mathbb{H}}(g) := h$$

Given two complementary subgroups $\mathbb{M}$ and $\mathbb{H}$, if one of them is normal in $\mathbb{G}$, then $\mathbb{G}$ is said to be the *semi-direct product* of $\mathbb{M}$ and $\mathbb{H}$; if both are normal, then $\mathbb{G}$ is called the *direct product* of $\mathbb{M}$ and $\mathbb{H}$. Another algebraic remark is the following: if $\mathbb{M}$ and $\mathbb{H}$ are complementary, then so are $\mathbb{H}$ and $\mathbb{M}$, namely if $\mathbb{G} = \mathbb{M} \cdot \mathbb{H}$, then $\mathbb{G} = \mathbb{H} \cdot \mathbb{M}$. In order to better understand these notions, let us begin by investigating the homogeneous and complementary subgroups of the Heisenberg group $\mathbb{H}^n$.

Example 5.6.4. Let $\mathbb{G} = \mathbb{H}^n = (\mathbb{R}^{2n+1}, \cdot)$ be the n-dimensional Heisenberg group, where the group law $\cdot$ is defined as in Example 5.1.5, and let $\mathbb{W} \subset \mathbb{H}^n$ be a homogeneous subgroup. Then only two cases are possible:

(i) $\mathbb{W}$ is *horizontal*, i.e. it is contained in the horizontal fiber $H\mathbb{H}^n_e$;

(ii) $\mathbb{W}$ is *vertical*, namely it contains the subgroup $\mathbb{T}$, given by

$$\mathbb{T} := \{\exp(sT) \,:\, s \in \mathbb{R}\} \equiv \{se_{2n+1} \,:\, s \in \mathbb{R}\}$$

where $T = \partial_{2n+1}$.

In the case of $\mathbb{H}^1$, the homogeneous subgroups are thus given by the lines contained in the plane $x_3 = 0$ (horizontal subgroups) and by the planes containing the x_3-axis (vertical subgroups). As a second remark, let us point out that a horizontal subgroup in $\mathbb{H}^n$ is always a (k,k)-homogeneous subgroup (that is, topological and metric dimension agree) with $1 \le k \le n$, whereas a vertical subgroup is always a $(m, m+1)$-homogeneous subgroup with $1 \le m \le 2n+1$. With these considerations, it is easy to see that $\mathbb{V}, \mathbb{W}$ are complementary subgroups in $\mathbb{H}^n$ if and only if $\mathbb{V}$ is a horizontal (k,k)-homogeneous subgroup with $1 \le k \le n$ and $\mathbb{W}$ is a vertical and normal $(2n+1-k, 2n+2-k)$-homogeneous subgroup. In particular, we see that the vertical axis $\mathbb{T}$ does not admit any complement. $\Diamond$

The characterization of all complementary subgroups in general Carnot groups, however, is not trivial at all, even if we can prove some nice splittings, as we are going to see immediately in the next example.

Example 5.6.5. Let $\mathbb{G} = (\mathbb{R}^n, \cdot)$ be a Carnot group and let $X_1, ..., X_n$ be an adapted basis of its Lie algebra $\mathfrak{g}$ (i.e. adapted to the stratification), with $X_1, ..., X_{m_1}$ as system of generating vector fields. As a first step, let

$$\mathbb{H} := \{\exp(sX_{j_0}) \,:\, s \in \mathbb{R}\} \equiv \{se_{j_0} \,:\, s \in \mathbb{R}\} \equiv \mathbb{R}$$

for a given $1 \le j_0 \le m_1$; choose for instance $j_0 = 1$. Then $\mathbb{H}$ is a $(1,1)$-homogeneous subgroup of $\mathbb{G}$. Secondly, let $\mathfrak{m} := \mathrm{span}(X_2, ..., X_n)$, which is easy seen to be a Lie subalgebra of $\mathfrak{g}$, and let us define

$$\mathbb{M} := \exp(\mathfrak{m}) = \{(0, x_2, ..., x_n) \,:\, x_i \in \mathbb{R}\} \equiv \mathbb{R}^{n-1}$$

The dimension of $\mathfrak{m}$ as Lie subalgebra of $\mathfrak{g}$ is $n-1$, since $[\mathfrak{m}, \mathfrak{m}] \subset \mathfrak{m}$; moreover, $\mathfrak{m}$ is also a (left) ideal of $\mathfrak{g}$, that is $[\mathfrak{g}, \mathfrak{m}] \subset \mathfrak{m}$. Thus, $\mathbb{M}$ turns out to be a normal subgroup of $\mathbb{G}$ of homogeneous dimension $Q-1$. Indeed, one can show that $\mathbb{M} \equiv \mathbb{R}^{n-1}$ can be considered as a Lie group with Haar measure $\mathscr{L}^{n-1}$, satisfying (5.2.3) with $Q-1$ in place of Q and closed w.r.t. the family of dilations of $\mathbb{G}$. Therefore, $(\mathbb{M}, d, \mathscr{L}^{n-1})$ is a metric measure Ahlfors regular space of dimension $Q-1$: as a byproduct, its metric dimension is $Q-1$. In conclusion, $\mathbb{M}$ is a $(n-1, Q-1)x$-homogeneous subgroup and we get the splitting $\mathbb{G} = \mathbb{M} \cdot \mathbb{H}$. $\Diamond$

After these examples, let us stress some differences w.r.t. the Euclidean case. First of all, the projections maps $P_{\mathbb{M}}$ and $P_{\mathbb{H}}$ may be not Lipschitz, even though they are still continuous and even smooth, as told by the next proposition.

Proposition 5.6.6. *Let $\mathbb{G}$ be a Carnot group and let $\mathbb{M}, \mathbb{H} \subset \mathbb{G}$ be complementary subgroups. Then the projection maps $P_{\mathbb{M}}$ and $P_{\mathbb{H}}$ are polynomial functions.*

The maps $P_{\mathbb{M}}$ and $P_{\mathbb{H}}$ may fail to be Lipschitz when $\mathbb{M}$ and $\mathbb{H}$ are equipped with the restrictions of the distance d of $\mathbb{G}$ (see Example 4.9 in [22]). However, the projection on the complement of a normal subgroupis always metric Lipschitz continuous, that is:

- if $\mathbb{H} \subset \mathbb{G}$ is a normal subgroup, then $P_{\mathbb{M}}$ is Lipschitz;
- if $\mathbb{M} \subset \mathbb{G}$ is a normal subgroup, then $P_{\mathbb{H}}$ is Lipschitz.

Now let us give the notion of intrinsic graph.

Definition 5.6.7. *Let $\mathbb{H} \subset \mathbb{G}$ be a homogeneous subgroup. A set $S \subset \mathbb{G}$ is said to be a* (left) $\mathbb{H}$-graph *(or also a* left graph in direction $\mathbb{H}$ *or, simply, an* intrinsic graph*) if S intersects each left coset of $\mathbb{H}$ in one point, at most; that is, for any $p \in \mathbb{G}$, $S \cap p \cdot \mathbb{H}$ is either a singleton or the empty set.*

Observe that this definition is completely general, because we are not asking $\mathbb{H}$ to admit a complementary subgroup. Secondly, an immediate characterization can be proved. Indeed, if $A \subset \mathbb{G}$ parametrizes the left cosets of $\mathbb{H}$, that is if A itself intersects each left coset of $\mathbb{H}$ exactly one time, and if S is an $\mathbb{H}$-graph, then there is a unique function $\varphi : \mathcal{E} \subset A \to \mathbb{H}$ such that

$$S = \mathrm{graph}(\varphi) = \{z \cdot \varphi(z) \,:\, z \in \mathcal{E}\}$$

Conversely, if A still parametrizes the left cosets of $\mathbb{H}$, then for any $\psi : \mathcal{D} \subset A \to \mathbb{H}$ the set $\mathrm{graph}(\psi)$ is an $\mathbb{H}$-graph of $\mathbb{G}$. Now let us see what happens if $\mathbb{H}$ admits a complementary subgroup $\mathbb{M}$. In this case, $\mathbb{M}$ parametrizes the left cosets of $\mathbb{H}$ and therefore we can say that S is an $\mathbb{H}$-graph if and only if $S = \mathrm{graph}(\varphi)$ for a unique function $\varphi : \mathcal{E} \subset \mathbb{M} \to \mathbb{H}$. However, it is worth stressing that examples of $\mathbb{H}$-graphs that are not, in general, graphs of functions acting between complementary subgroups have been considered in the Heisenberg groups: for instance, if $\mathbb{H}$ is the horizontal graph $\mathbb{T}$, then $A = H\mathbb{H}_e$. As we are going to see immediately, the importance of intrinsic graphs is due to the fact that on them we can carry out intrinsic calculus, because the notion of $\mathbb{H}$-graph is stable under left translations and group dilations. More precisely, we have the following result.

Proposition 5.6.8. *Let $\mathbb{H}$ be a homogeneous subgroup of $\mathbb{G}$. If S is an $\mathbb{H}$-graph, then so are $\delta_\lambda S$ and $q \cdot S$ for every $\lambda > 0$ and for every $q \in \mathbb{G}$. If, in particular, $\mathbb{M}, \mathbb{H} \subset \mathbb{G}$ are complementary subgroups and $S = \mathrm{graph}(\varphi)$ with $\varphi : \mathcal{E} \subset \mathbb{M} \to \mathbb{H}$, then for any $q \in \mathbb{G}$ and for any $\lambda > 0$*

$$q \cdot S = \mathrm{graph}(\varphi_q), \quad \varphi_q : \mathcal{E}_q \subset \mathbb{M} \to \mathbb{H}$$

$$\delta_\lambda S = \text{graph}(\varphi_\lambda), \quad \varphi_\lambda : \mathcal{E}_\lambda \subset \mathbb{M} \to \mathbb{H}$$

where

$$\mathcal{E}_q = \{m \in \mathbb{M} \,:\, P_{\mathbb{M}}(q^{-1}m) \in \mathcal{E}\}, \quad \mathcal{E}_\lambda = \delta_\lambda \mathcal{E}$$

and

$$\varphi_q(m) = (P_{\mathbb{H}}(q^{-1}m))^{-1} \cdot \varphi(P_{\mathbb{M}}(q^{-1}m)), \quad \varphi_\lambda(m) = \delta_\lambda \varphi(\delta_{1/\lambda} m)$$

Proof. See Proposition 4.13 of [22]. □

Now, having Definition 5.5.10 in mind, we can state an implicit function theorem for $C^1_{\mathbb{G}}$ functions within Carnot groups, thus motivating such definition. But first, a natural question: why do we need to formulate an adapted version of the implicit function theorem? Does not the classical statement apply in the setting of Carnot groups? The answer is given by [14], where B. Kirchheim and F. Serra Cassano proved that in $\mathbb{G} = \mathbb{H}^1 \equiv \mathbb{R}^3$ there exists an $\mathbb{H}^1$-regular hypersurface S such that

$$\text{Hdim}(S) = 2.5$$

where the Hausdorff dimension is considered w.r.t. the Euclidean norm $\| \cdot \|_{\mathbb{R}^3}$. So, from a Euclidean point of view the surface is purely 2-unrectifiable. Nevertheless, from a sub-Riemannian point of view we are able to prove the following statement.

Theorem 5.6.9. *Let $\mathbb{G} = (\mathbb{R}^n, \cdot)$ be a Carnot group, let $\Omega \subset \mathbb{G}$ be an open set with $0 \in \Omega$ and let $f \in C^1_{\mathbb{G}}(\Omega)$ be such that $f(0) = 0$ and $X_1 f(0) > 0$. Let us define*

$$E := \{x \in \Omega \,:\, f(x) < 0\}, \quad S := \{x \in \Omega \,:\, f(x) = 0\}$$

Then there exists an open neighbourhood $U \subset \mathbb{G}$ of 0 such that:

(i) E has finite $\mathbb{G}$-perimeter in U, that is $|\partial E|_{\mathbb{G}}(U) < +\infty$;

(ii) $\partial E \cap U = S \cap U$;

(iii) for all $x \in S \cap U$ we have

$$\nu_E(x) = -\frac{\nabla_{\mathbb{G}} f(x)}{|\nabla_{\mathbb{G}} f(x)|_x}$$

and ν_E can be identified with a section of $H\mathbb{G}$; in particular, the map $x \mapsto \nu_E(x)$ is continuous.

Moreover, if

$$\mathbb{H} := \exp(\text{span}(X_1)), \quad \mathbb{M} := \exp(\text{span}(X_2, ..., X_n))$$

and we identify them with $\mathbb{R}$ and $\mathbb{R}^{n-1}$, then there exists a unique function $\varphi : \mathcal{E} \subset \mathbb{M} \to \mathbb{H}$ such that:

(iv) $S \cap U = \mathrm{graph}(\varphi)$ *and* $E \cap U = U \cap \{z \cdot \delta_s(e_1) : z \in \mathcal{E},\ s < \varphi(z)\}$*, that is* S *is an intrinsic graph and* E *is the associated intrinsic sublevel;*

(v) φ *is continuous;*

(vi) the $\mathbb{G}$*-perimeter has an integral representation, that is the following area formula holds*

$$|\partial E|_{\mathbb{G}}(U) = \int_{\mathcal{E}} \frac{|\nabla_{\mathbb{G}} f|}{X_1 f} \circ \Phi \, d\mathscr{L}^{n-1}$$

where $\mathscr{L}^{n-1}$ *is nothing but the Haar measure of* $\mathbb{M}$ *and* $\Phi : \mathcal{E} \to U$ *is given by* $\Phi(z) := z \cdot \varphi(z)$.

Proof. See [10]. □

Notice that in (iii) we are allowed to consider the generalized $\mathbb{G}$-normal to ∂E in U because of (i); in (iv) $s < \varphi(z)$ makes sense because we are identifying $\mathbb{H}$ with $\mathbb{R}$. This theorem has many interesting consequences and the first one is the following: a priori, it is not clear which the topological dimension of S is, but since Φ is continuous and injective, then Φ is a homeomorphism onto its image, so that we are allowed to state the next result.

Corollary 5.6.10. *Each* $\mathbb{G}$*-regular hypersurface has topological dimension* $n-1$.

The second consequence involves the metric dimension and follows by Ambrosio's representation result (Theorem 5.5.8).

Corollary 5.6.11. *Each* $\mathbb{G}$*-regular hypersurface has metric dimension (in* $\mathbb{G}$*)* $Q-1$.

The discussion could now go on in many different ways, since calculus within Carnot groups is a very rich topic; for instance, we could introduce the notion of tangent space as a suitable vertical subgroup and expand the subject, but unfortunately we do not have enough time. For this reason, we suggest some selected developments, more linked with what we have already discussed:

(1) study of intrinsic regular surfaces in Heisenberg groups (see Section 4.5.1 of [22]) and some extensions to general Carnot groups (see Section 4.5.2 of [22]); in this case the $\mathbb{G}$-regularity is extended to higher codimensions;

(2) for an intrinsic graph $S = \mathrm{graph}(\varphi)$ with $\varphi : \mathcal{E} \subset \mathbb{M} \to \mathbb{H}$, a notion of intrinsic Lipschitz graph can be introduced (see Section 4.6 of [22]), even though it is worth stressing that φ need not be Lipschitz (and viceversa, if φ is Lipschitz, then its graph need not be intrinsic Lipschitz); nevertheless, a good feature is the following: the Hausdorff dimension of (S, d) is the same of $(\mathbb{M}, d)$, where d is the invariant distance endowing $\mathbb{G}$;

(3) a notion of intrinsic differentiability can be introduced for functions $\varphi : \mathcal{E} \subset \mathbb{M} \to \mathbb{H}$, that is for functions acting between complementary subgroups; such notion agrees with Pansu differentiability only if $\mathbb{G} = \mathbb{M} \times \mathbb{H}$, that is $\mathbb{G}$ is the direct product of $\mathbb{M}$ and $\mathbb{H}$ (i.e. $\mathbb{M}, \mathbb{H}$ are normal in $\mathbb{G}$);

(4) in the case of Heisenberg groups, it has been proved that any regular surface is locally an intrinsic Lipschitz graph and those graphs are intrinsically differentiable (see Section 4.7 of [22]);

(5) in the case of Heisenberg groups, it has also been shown that a Rademacher-type result holds for intrinsic Lipschitz functions $\varphi : \mathcal{E} \subset \mathbb{W} \to \mathbb{V}$ when $\mathbb{V}$ is a $(1,1)$-homogeneous subgroup; however, the result is partial because the problem is still open when $\mathbb{V}$ has higher dimension.

5.7 Lecture 7 - 17 September

Abstract

We begin the second part of the course by providing three different (but equivalent) definitions of Sobolev space in $\mathbb{R}^n$: via integration by parts, closure of test functions and Beppo Levi condition. These notions will be then generalized to several situations, as for instance sub-Riemannian and weighted Riemannian manifolds. On a metric measure space, the situation is much more complicated and we will try to build the $H^{1,p}$ Sobolev space by means of relaxed slopes.

In this second part of the course, the subjects we will consider are completely different from those of the first part. As already anticipated, we will focus our attention on Sobolev and BV functions on metric measure spaces, starting with a bit of history, the motivations and some examples. As a first fact, it is well-known that we have several definitions of Sobolev space in $\mathbb{R}^n$ and that they are equivalent. Indeed, on the one hand we can define for $1 < p < +\infty$ the space $W^{1,p}(\mathbb{R}^n)$ as the set of functions $f \in L^p(\mathbb{R}^n)$ for which there exists a vector field $g \in L^p(\mathbb{R}^n, \mathbb{R}^n)$ such that the integration by parts formula holds coordinatewise for any test function φ at least C^1, that is

$$\int_{\mathbb{R}^n} f \frac{\partial \varphi}{\partial x_i}\, dx = -\int_{\mathbb{R}^n} g_i \varphi \, dx, \quad i = 1, ..., n \tag{5.7.1}$$

The vector field g is called *weak gradient* and it is denoted by ∇f. In the case the test functions φ have to be smooth and compactly supported, i.e. they belong to $C_c^\infty(\mathbb{R}^n)$, we see that this definition is coherent with the one coming from the theory of distributions. On the other hand, still for $1 < p < +\infty$, we can define $H^{1,p}(\mathbb{R}^n)$ as the closure of $C_c^\infty(\mathbb{R}^n)$ in the $W^{1,p}$ norm. These two definitions turn out to be equivalent, so that

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

Indeed, for the inclusion $\supset$ it is sufficient to observe that condition (5.7.1) passes to the limit in the $W^{1,p}$ norm. For the opposite inclusion, take $f \in W^{1,p}(\mathbb{R}^n)$ and consider the convolution of f with a smooth mollifier, i.e. $f * \rho_\varepsilon \in C^\infty \cap W^{1,p}$; up to multiply with a cut function, we can assume that $f * \rho_\varepsilon$ has compact support and thus belongs to $H^{1,p}(\mathbb{R}^n)$. By well-known properties of the convolution, $f * \rho_\varepsilon \to f$ in the $W^{1,p}$ norm and thus the conclusion. The modification procedure (convolution and truncation) we oulined first appeared in the paper [15] of Meyers and Serrin; for sake of information, let us acknowledge the existence of a Meyers-Serrin theorem in sub-Riemannian geometry, proved by Franchi, Serapioni and Serra Cassano.

A part from $W^{1,p}$ and $H^{1,p}$, many other possible definitions of Sobolev spaces are available. In particular, we would like to present a third approach, which is maybe less known than the previous ones. Such approach is due to Beppo Levi, who formulated it in 1901 for the case $p = 2$ and $n = 2$, hence much earlier than the definitions of $W^{1,p}$ and $H^{1,p}$. He was motivated by the search of a solution to a Dirichlet problem in $\mathbb{R}^2$ and, after him, these Sobolev spaces are also called *Beppo Levi spaces*. Nowadays, their definition has been generalized to any $1 < p < +\infty$ and $n \in \mathbb{N}$, so that we can talk about $BL^{1,p}(\mathbb{R}^n)$ spaces. More precisely, we will say that a function f belongs to $BL^{1,p}(\mathbb{R}^n)$ if for any $i \in \{1, ..., n\}$ and for $\mathscr{L}^{n-1}$-a.e. $\hat{x}_i := (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) \in \mathbb{R}^{n-1}$ the map

$$t \mapsto f(x_1, ..., x_{i-1}, t, x_{i+1}, ..., x_n) \tag{5.7.2}$$

is absolutely continuous in $\mathbb{R}$ and, moreover,

$$\int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \left(|f|^p + \left| \frac{\partial f}{\partial x_i} \right|^p \right) dt d\hat{x}_i < +\infty$$

Indeed, it is well-known that absolutely continuous functions are a.e. differentiable and so the condition above makes sense. In order to understand the spirit of this definition, we have to take into account the fact that at the beginning of the 20$^{\text{th}}$ century the theory of absolutely continuous functions was already well developed. However, Beppo Levi's definition has been long ignored, since at that time there was a lack of mathematical tools in order to overcome some natural problems linked to such definition, for instance its invariance under a change of coordinates. Fortunately, in the years when the classical theory of Sobolev spaces $W^{1,p}$ and $H^{1,p}$ was being developed and more precisely in 1957, Fuglede answered to the objection about frame change in Beppo Levi's definition, proving that this notion is not affected by them.

After this historical introduction, let us clearly point out our goal: the definition of analogous spaces H, W and BL in metric measure spaces and the proof of their agreement. To this aim, let us first analyze the previous three definitions and stress that, even in the Euclidean case, they are very different, although they induce the same Sobolev spaces. In fact, in the definition of $W^{1,p}$ vector fields are involved, in the case of $H^{1,p}$ we are interested in approximations by smooth and compactly supported functions and, finally, for $BL^{1,p}$ we look at the behaviour of functions along curves. Another slight difference between $W^{1,p}$, $H^{1,p}$ on the one hand and $BL^{1,p}$ on the other one is the following: the last definition is pointwise, because absolute continuity is a pointwise property and, indeed, when we ask (5.7.2) to be absolutely continuous, we can not expect that the same holds for its Lebesgue equivalence class; on the contrary, the first two definitions involve Lebesgue equivalence classes and are therefore insensitive to pointwise modifications.

The motivation for the study of BV functions on metric measure spaces comes from several joint works of Ambrosio with Gigli, Savaré, Mondino, Colombo and Di Marino. In particular, in the papers with Gigli and Savaré the attention was focused on Ricci lower bounds in metric measure spaces $(X, d, \mathfrak{m})$, in order to analyze two parallel theories:

(1) Bakry-Émery theory: it is an Eulerian theory, based, in the case of $CD(K, N)$ spaces, on the differential inequality

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

which encodes the upper bound N on the dimension and the lower bound K on the Ricci curvature. By means of the tools of Γ-calculus and Dirichlet forms, we can give a precise meaning to every term appearing in this inequality;

(2) Lott-Sturm-Villani theory: it is a Lagrangian theory, based on optimal transport theory, where a key role is played by the K-convexity of the entropy, defined by

$$\rho\mathfrak{m} \mapsto \int_X \rho \log \rho \, d\mathfrak{m}$$

in the case $N = +\infty$ (notice that this does not imply that we are working in an infinite-dimensional space, but only that we are not asking any upper bound on the dimension) and by

$$\rho\mathfrak{m} \mapsto -\int_X \rho^{1-1/N} \, d\mathfrak{m}$$

in the case N is finite.

It has been proved that Lott-Sturm-Villani theory, coupled with the additional assumption of infinitesimal Hilbertianity, is equivalent to Bakry-Émery theory, so that there are really two sides of the same theory and this fact takes advantage to both theories: indeed, theory (1) is very useful in order to get functional inequalities, even in sharp form, whereas theory (2) is optimal for stability properties in the sense of Gromov-Hausdorff convergence. In the case $N < +\infty$, this result has been very recently (2013) generalized by Erbar, Kuwada and Sturm and further developments are being studied by Ambrosio, Mondino and Savaré.

In the sub-Riemannian case the Eulerian theory has been deply studied by Badouin, Garofalo, Agrachev and Lee; on the contrary, the Lagrangian part is unsatisfactory and its comprehension is still an open problem. Indeed, Juillet showed that the standard Lott-Villani approach does not work in the sub-Riemannian case.

In order to understand the reason why the equivalence between Eulerian and Lagrangian theory is not trivial at all, let us consider the following example, which is an immediate corollary of the equivalence.

Example 5.7.1. Let $(X, d, \mathfrak{m})$ be a metric measure space and assume that it contains no nonconstant rectifiable curves i.e. all the rectifiable curves in X are constant (think for instance to $(\mathbb{R}, \sqrt{|\cdot|})$). Then $BL^{1,p}(X, d, \mathfrak{m})$ is essentially equivalent to $L^p(X, \mathfrak{m})$, because Beppo Levi's definition concerns oscillations along curves and in this case all curves are constant. Thus, coinciding with $BL^{1,p}$, also the other Sobolev spaces are equivalent to $L^p(X, \mathfrak{m})$; in particular, this is true for $H^{1,p}(X, d, \mathfrak{m})$ and this means that for any $f \in L^p(X, \mathfrak{m})$ we are able to find a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}(X, d)$ (in metric setting, Lipschitz functions play the role of smooth ones) with the property that

$$\lim_{n\to\infty} \int_X \Big(|f_n - f|^p + |\nabla f_n|^p\Big) d\mathfrak{m} = 0 \tag{5.7.3}$$

and this means that we can approximate any function $f \in L^p(X, \mathfrak{m})$ by means of Lipschitz functions with slope as small as we want. This is in some sense surprising, because if we know something about pathwise behaviour, then we also know something about global approximation by means of Lipschitz functions. $\diamondsuit$

Clearly, in this example the discussion is rough and intuitive, because the spaces $H^{1,p}(X, d, \mathfrak{m})$ and $BL^{1,p}(X, d, \mathfrak{m})$ have not been defined yet. The quantity $|\nabla f|$ in (5.7.3) is called *slope* or *local Lipschitz constant* and is defined as

$$|\nabla f|(x) := \limsup_{y\to x} \frac{|f(y) - f(x)|}{d(y, x)} \tag{5.7.4}$$

For this reason we have just said that the approximating sequence $(f_n)_{n\in\mathbb{N}}$ is made up by functions with arbitrarily small slope. After this remark, let us provide further examples motivating the study of BV functions in metric setting. Obviously we have Euclidean spaces endowed with the Lebesgue measure, but we have also more refined examples. For instance, as a second case we can still consider a Euclidean space, but modify the measure it is equipped with.

Example 5.7.2. We are going to work on $\mathbb{R}^n$. Let us replace the usual Lebesgue measure $\mathscr{L}^n$ with $\mathfrak{m} = \omega\mathscr{L}^n$, where ω is a suitable density. In most of the cases, we will consider $\omega = e^{-V}$ for a certain potential function V (note that when $V(x) = |x|^2/2$, we recover the Gaussian case) and, if V is sufficiently regular, an integration by parts formula w.r.t. $\mathfrak{m}$ holds, namely

$$\int_{\mathbb{R}^n} \frac{\partial \psi}{\partial x_i}\, d\mathfrak{m} = \int_{\mathbb{R}^n} \frac{\partial V}{\partial x_i} \psi\, d\mathfrak{m} \tag{5.7.5}$$

for any test function ψ (say smooth and compactly supported). By means of this relation, we can introduce a notion of weak derivative w.r.t. the structure given by $\mathfrak{m}$ as follows: $\frac{\partial f}{\partial x_i}$ is said to be the weak derivative of $f \in L^p(\mathbb{R}^n, \mathfrak{m})$ in the i-th direction if, for any $\varphi \in C_c^\infty(\mathbb{R}^n)$

$$\int_{\mathbb{R}^n} \frac{\partial f}{\partial x_i} \varphi\, d\mathfrak{m} = -\int_{\mathbb{R}^n} f \frac{\partial \varphi}{\partial x_i}\, d\mathfrak{m} + \int_{\mathbb{R}^n} \frac{\partial V}{\partial x_i} f \varphi\, d\mathfrak{m}$$

What happens if ω is not smooth (this is the case in the theory of degenerate PDEs)? Can we still get an integration by parts formula? As we will see, the answer is positive but we have to give up the idea of coordinate directions. In this situation, i.e. $\mathfrak{m} = \omega\mathscr{L}^n$, we can define the so-called *weighted Sobolev spaces* $W^{1,p}_\omega(\mathbb{R}^n)$ by saying that $f \in W^{1,p}_\omega(\mathbb{R}^n)$ if and only if $f \in W^{1,1}_{loc}(\mathbb{R}^n)$ and

$$\int_{\mathbb{R}^n} \Big(|f|^p + |\nabla f|^p\Big)\omega dx < +\infty$$

This definition is extrinsic or hybrid, in some sense, because the definition of weighted weak derivative relies on the classical one and the weight ω appears only in the integrability condition. $\Diamond$

As a third example, we can move from the Euclidean setting to the Riemannian one. In this case we can point out some interesting facts which will be useful also in the following abstract setting.

Example 5.7.3. Let us consider a weighted Riemannian manifold $(M, g, \mathfrak{m})$, where $\mathfrak{m} = e^{-V}\,\mathrm{vol}_M$ and vol_M is the volume form on M associated to the Riemannian metric g. First, the Laplacian is defined as

$$\Delta f := \mathrm{div}_{\mathfrak{m}}(\nabla f)$$

and we can already see that there is an interaction between the metric and the measure, because $\mathrm{div}_{\mathfrak{m}}$ depends on $\mathfrak{m}$, whereas ∇ depends on g. Indeed, on the one hand ∇f is the vector field satisfying the relation

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

for any vector field v; if we think of vector fields as derivations, then this condition becomes $v(f) = g(\nabla f, v)$. On the other hand, if b is a vector field on M, then its divergence is defined as the function $\mathrm{div}_{\mathfrak{m}}\, b$ for which the following integration by parts formula holds

$$\int_M (\mathrm{div}_{\mathfrak{m}}\, b)\varphi \, d\mathfrak{m} = -\int_M g(b, \nabla\varphi)\, d\mathfrak{m} \tag{5.7.6}$$

for every $\varphi \in C^\infty_c(M)$. This definition plays the same role of (5.7.5) in the previous example, because if φ is not smooth but b is regular, then we can take (5.7.6) as definition of the weak derivative of φ. Such definition could seem implicit, but as we are working in a Riemannian setting, it becomes very explicit by taking into account the fact that

$$\mathrm{div}_{\mathfrak{m}}\, b = \mathrm{div}\, b - g(b, \nabla V)$$

Anyway, we can already see that we have to give up the idea of coordinate directions, even in a smooth setting such as the Riemannian one: it is much better to define Sobolev spaces via the integration by parts formula (5.7.6), testing it against all smooth vector fields. $\Diamond$

The same procedure applies to Carnot-Carathéodory spaces, as we are going to see immediately in the next example.

Example 5.7.4. Let $X_1, ..., X_m$ be a family of vector fields on $\mathbb{R}^n$ and let us assume that

$$X_i, \operatorname{div} X_i \in L^\infty(\mathbb{R}^n), \quad i = 1, ..., m$$

Then we can define weak derivatives in the usual way and the Sobolev space $W_X^{1,p}(\mathbb{R}^n)$ is given by

$$W_X^{1,p}(\mathbb{R}^n) := \{f \in L^p(\mathbb{R}^n) \,:\, X_i f \in L^p(\mathbb{R}^n),\, i = 1, ..., m\}$$

where $X_i f$ has to be meant in the distributional sense and this is possible, since we are assuming $\operatorname{div} X_i \in L^\infty(\mathbb{R}^n)$. It will be interesting to compare this definition with the metric one, which will be better discussed later in great generality. For the moment, let us only say that to the vector fields $X_1, ..., X_m$ we can associate a Carnot-Carathéodory distance d_{CC} or, better, an *extended* Carnot-Carathéodory distance, because we are not assuming any bracket-generating condition, and by means of this distance we can introduce a metric Sobolev space $W^{1,p}(\mathbb{R}^n, d_{CC}, \mathscr{L}^n)$. In the case the vector fields X_i are smooth, we have that

$$W^{1,p}(\mathbb{R}^n, d_{CC}, \mathscr{L}^n) = W_X^{1,p}(\mathbb{R}^n) \tag{5.7.7}$$

When $X_1, ..., X_m$ are not smooth the relationship between the two spaces is not clear yet, because in (5.7.7) smoothness is used in the proof of both inclusions. $\Diamond$

In all the examples above we have considered finite-dimensional spaces, but calculus can be carried out also in the infinite-dimensional case. For this reason, in the next example we will consider infinite-dimensional Gaussian spaces, as for instance the Wiener space.

Example 5.7.5. In order to build an infinite-dimensional Gaussian space, let γ_1 be the Gaussian with mean 0 and variance 1 in $\mathbb{R}$, namely

$$\gamma_1 = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathscr{L}^1$$

and let us equip $\mathbb{R}^\infty$ with $\mathfrak{m} := \pi_{i=1}^\infty \gamma_1$. Clearly $\mathfrak{m}$ is a probability measure on $\mathbb{R}^\infty$, which is a locally convex space. For people more at ease with the theory of Banach spaces, let us fix a sequence $c \in \ell_1$ and let us define

$$H := \left\{ x \in \mathbb{R}^\infty \,:\, \sum_{i=1}^\infty c_i x_i^2 < +\infty \right\}$$

which turns out to be a Hilbert space when endowed with the scalar product

$$\langle x, y \rangle_H := \sum_{i=1}^\infty c_i x_i y_i$$

Taking into account the fact that the variance of γ_1 is 1, we easily see that

$$\int \Big(\sum_{i=1}^{\infty} c_i x_i^2 \Big) d\mathfrak{m} = \sum_{i=1}^{\infty} c_i < +\infty$$

and therefore $\mathfrak{m}$ is also a probability measure on H. The distance we put on H, even if it is a Hilbert space, is the extended distance defined by

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

and also known in the literature as Cameron-Martin distance. From the point of view of the connectivity, the situation is clear: H can be factorized in parallel copies of ℓ_2 (in the sense that the distance between any two different copies is infinite), because $x + \ell_2$ and $y + \ell_2$ are either coincident or disjoint and, in the second case, the distance between any two points (not taken on the same line) is infinite. In other words, movement is allowed only along directions parallel to ℓ_2 and this behaviour can be compared to what happens in the sub-Riemannian case, where movements are constrained by the horizontal vector fields. Thus, from the point of view of single points it is impossible to move from a copy of ℓ_2 to a different one. However, what makes this example so interesting is the fact that the situation completely changes if we move from single points to random points, i.e. if we consider probability distributions on H.

In more details, let $\mu_1 = \rho_1 \mathfrak{m}$ and $\mu_2 = \rho_2 \mathfrak{m}$ be two probability measures on H and let us assume that they have finite entropy, that is

$$\int_H \rho_i \log \rho_i \, d\mathfrak{m} < +\infty, \quad i = 1,2$$

Then, by virtue of an *a priori* estimate of Talagrand and results in optimal transportation due to Feyel and Ustunel, we can affirm that there exists a displacement map $D : H \to \ell_2$ such that the transport map T defined by

$$T(x) := x + D(x)$$

transports μ_1 on μ_2, in mathematical language $T_\# \mu_1 = \mu_2$. Hence, at level of random points we recover connectivity, because, under the finite entropy hypothesis, we can move from every distribution to any other one. Moreover, we have also the following quantitative estimate on the displacement:

$$\int_H d^2(x, x + D(x)) \rho_1(x) d\mathfrak{m}(x) \le 2 \Big(\int_H \rho_1 \log \rho_1 \, d\mathfrak{m} + \int_H \rho_2 \log \rho_2 \, d\mathfrak{m} \Big)$$

For sake of information, this inequality can be seen as a generalization of Brenier's theorem. Finally, the map T is also optimal, so that the left-hand side in the inequality above coincides with the squared Wasserstein distance between $\rho_1 \mathfrak{m}$ and $\rho_2 \mathfrak{m}$. $\Diamond$

Let us briefly recall the definition of push-forward of measures we used in the previous example. Let $(X,\mathscr{E})$ and $(Y,\mathscr{F})$ be two spaces with σ-algebra, $T : X \to Y$ a measurable map and $\mu \in \mathscr{P}(X)$. The *push-forward of* μ *through* T is the measure $T_{\#}\mu \in \mathscr{P}(Y)$ defined by

$$T_{\#}\mu(E) = \mu(T^{-1}(E)), \quad \forall E \in \mathscr{F}$$

An immediate consequence of this definition is the following change of variable formula.

$$\int_Y \varphi \, dT_{\#}\mu = \int_X (\varphi \circ T)\, d\mu \tag{5.7.8}$$

The last example we would like to mention is given by measured Gromov-Hausdorff limits and also in this case extended distances play an important role. In the seminal paper of Cheeger and Colding it has been studied the case where $(X,d,\mathfrak{m})$ is the measured Gromov-Hausdorff limit of a sequence $((M_n,g_n,\mathfrak{m}_n))_{n\in\mathbb{N}}$ of Riemannian manifolds under Ricci curvature, dimension or diameter bounds. The aim was to establish fundamental properties such as doubling and local Poincaré inequality in order to carry out calculus.

After this detailed motivating discussion, we can now begin to study the theory of BV functions with an initial part devoted to the H-Sobolev space and the first tools of differential calculus. Let $(X,d,\mathfrak{m})$ be a metric measure space. The first step relies on the introduction of the *asymptotic Lipschitz constant* $\mathrm{lip}_a(f,x)$, which is defined by

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

where $\mathrm{Lip}(f,B(x,r))$ denotes the Lipschitz constant of f on $B(x,r)$; by monotonicity, the limit exists and $x \mapsto \mathrm{lip}_a(f,x)$ is an upper semicontinuous function. It is easy to check that

$$\mathrm{lip}_a(f,x) \geq |\nabla f|(x), \quad \forall x \in X \tag{5.7.9}$$

where $|\nabla f|$ is the slope defined at (5.7.4); for technical reasons, it will be better to work with $\mathrm{lip}_a(f,x)$ rather than with $|\nabla f|$. By virtue of the upper semicontinuity of $x \mapsto \mathrm{lip}_a(f,x)$, (5.7.9) easily yields that

$$\mathrm{lip}_a(f,x) \geq |\nabla f|^*(x), \quad \forall x \in X \tag{5.7.10}$$

where $|\nabla f|^*$ is the upper semicontinuous relaxation of $|\nabla f|$, that is

$$|\nabla f|^*(x) := \sup\left\{ \limsup_{n\to\infty} |\nabla f|(y_n) \,:\, y_n \to x \right\}$$

Note that $\mathrm{lip}_a(f,\cdot)$ and $|\nabla f|^*$ agree everywhere if (X,d) is a *length space*, namely for every $x,y \in X$

$$d(x,y) = \inf\{\ell(\gamma) \,:\, \gamma(0) = x,\, \gamma(1) = y\}$$

Pay attention that the infimum needs not be a minimum, but in the case it is always attained, (X, d) is said to be a *geodesic space*. In what follows we will not assume that (X, d) is a length space and for this reason we have to keep in mind that the inequality in (5.7.10) may be strict.

With these preliminary definitions we are ready to define the H-Sobolev space $H^{1,p}(X, d, \mathfrak{m})$, still assuming $1 < p < +\infty$, as the set of functions $f \in L^p(X, \mathfrak{m})$ for which there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}_0(X, d) \cap L^p(X, \mathfrak{m})$ such that:

(i) $f_n \to f$ as $n \to \infty$ in $L^p(X, \mathfrak{m})$;

(ii) it holds that

$$\limsup_{n\to\infty} \int_X \mathrm{lip}_a^p(f_n, x) d\mathfrak{m}(x) < +\infty$$

where $\mathrm{Lip}_0(X, d)$ denotes the family of Lipschitz functions with bounded support. Let us point out that this definition is a variant of Cheeger's original one, due to Gafa in 2000; this choice is more restrictive than Cheeger's one, because we use a stronger notion of gradient and a smaller class of approximating functions. After this precisation, for any function $f \in H^{1,p}(X, d, \mathfrak{m})$ we would like to define the gradient or at least its modulus. To this aim, observe that to f a sequence $(f_n)_{n\in\mathbb{N}}$ is associated, whose asymptotic Lipschitz constants are bounded in $L^p(X, \mathfrak{m})$; thus, by the reflexivity of $L^p(X, \mathfrak{m})$ a weak limit exists and this motivates the following definition.

Definition 5.7.6. *We say that $g \in L^p(X, \mathfrak{m})$ is a* relaxed slope *of $f \in L^p(X, \mathfrak{m})$ if there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}_0(X, d) \cap L^p(X, \mathfrak{m})$ such that:*

(a) $f_n \to f$ in $L^p(X, \mathfrak{m})$ and $\mathrm{lip}_a(f_n) \rightharpoonup \tilde{g}$ in $L^p(X, \mathfrak{m})$;

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

The set of relaxed slopes of f will be denoted by $RS(f)$.

Thanks to the previous discussion, $H^{1,p}(X, d, \mathfrak{m}) = \{f \in L^p(X, \mathfrak{m}) : RS(f) \neq \emptyset\}$. As a next step, let us try to single out the "best" approximation, the "optimal" relaxed slope in a sense to be precised. In order to do that, a technical remark is required: by means of Mazur lemma, it is always possible to turn weak limits into strong ones, so that property (a) of Definition 5.7.6 is equivalent to

(a') $f_n \to f$ in $L^p(X, \mathfrak{m})$ and $g_n \to g$ in $L^p(X, \mathfrak{m})$, for suitable functions $g_n \in L^p(X, \mathfrak{m})$ such that $g_n \geq \mathrm{lip}_a(f_n)$.

The importance of this characterization is due to the fact that strong limits allow diagonal arguments, whereas this is impossible with weak limits, and by diagonal arguments we get that if $f_n \to f$ in $L^p(X, \mathfrak{m})$, $g_n \in RS(f_n)$ and $g_n \rightharpoonup g$ in $L^p(X, \mathfrak{m})$, then $g \in RS(f)$. This fact has many important implications.

(i) $RS(f)$ is weakly closed and convex; indeed, on the one hand the weak closure follows by taking $f_n = f$ for any $n \in \mathbb{N}$ and on the other hand the convexity is a direct consequence of the convexity of $f \mapsto \mathrm{lip}_a(f)$. Therefore, there exists a unique $g \in RS(f)$ with least L^p norm; we shall denote it by

$$g =: |\nabla f|_*$$

because it is a sort of weak notion of modulus of the gradient and we will refer to it as *minimal relaxed slope*. The associated approximating sequence in $\mathrm{Lip}_0(X,d)$ will be called *optimal*.

(ii) Let us define Cheeger's energy as the lower semicontinuous relaxation of the energy functionsl $\int_X \mathrm{lip}_a^p(f_n) d\mathfrak{m}$ starting from Lipschitz functions, namely

$$\mathrm{Ch}(f) := \inf \left\{ \liminf_{n\to\infty} \frac{1}{p} \int_X \mathrm{lip}_a^p(f_n, x) d\mathfrak{m}(x) \right\}$$

where the infimum is taken on all the sequences $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}_0(X,d) \cap L^p(X,\mathfrak{m})$ such that $f_n \to f$ in $L^p(X,\mathfrak{m})$. Then we can represent it as the integral of the minimal relaxed slope, i.e.

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

(iii) The functional $f \mapsto \mathrm{Ch}(f)$ is convex and lower semicontinuous in $L^p(X,\mathfrak{m})$. Moreover, its domain is dense because we know that Lipschitz functions with bounded support have finite Cheeger energy.

In particular, the third property will play a crucial role, because we will use gradient flows techniques. Now let us see some elements of differential calculus. First of all, we are able to perform strong approximations, because for every $f \in H^{1,p}(X,d,\mathfrak{m})$ there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}_0(X,d)$ with $\mathrm{lip}_a(f_n) \to |\nabla f|_*$ strongly in $L^p(X,\mathfrak{m})$. Indeed, note that the Sobolev space $H^{1,p}$ coincides with the domain of Cheeger's energy and, in addition, weak convergence in $L^p(X,\mathfrak{m})$ coupled with convergence of the L^p norms implies strong convergence in $L^p(X,\mathfrak{m})$, so that in correspondence of the optimal sequence it is not necessary to apply Mazur lemma. Another important property, emphasized by Cheeger, is the pointwise minimality of the minimal relaxed slope; in fact, we know that $|\nabla f|_*$ minimizes the L^p norm in $RS(f)$ but this property can actually be refined, because it can be proved that

$$|\nabla f|_* \le g \ \mathfrak{m}\text{-a.e. in } X, \quad \forall g \in RS(f)$$

Thirdly, Cheeger's energy is a local operator, in the sense that for any $f, g \in H^{1,p}(X,d,\mathfrak{m})$

$$|\nabla f|_* = |\nabla g|_* \quad \mathfrak{m}\text{-a.e. on } \{f = g\}$$

Finally, the chain rule holds in the following sense: for any $\varphi \in \mathrm{Lip}(\mathbb{R})$ with $\varphi(0) = 0$

$$|\nabla(\varphi \circ f)|_* = |\varphi' \circ f| |\nabla f|_* \quad \mathfrak{m}\text{-a.e. in } X \tag{5.7.11}$$

where the condition $\varphi(0) = 0$ is required to make $\varphi \circ f$ p-integrable. Let us briefly sketch the proof of these properties.

Proof. The strong approximation property has already been sketched above. For the pointwise minimality, let us first point out a property of the asymptotic Lipschitz constant, namely the fact that it is a pseudo-gradient; indeed, let $\chi \in \mathrm{Lip}(X,[0,1])$, $f, \tilde{f} \in \mathrm{Lip}(X,d)$ and observe that

$$\mathrm{lip}_a((1-\chi)f + \chi\tilde{f}) \le (1-\chi)\,\mathrm{lip}_a(f) + \chi\,\mathrm{lip}_a(\tilde{f}) + \mathrm{Lip}(\chi)|f-\tilde{f}|$$

In the case of gradients, $\mathrm{Lip}(\chi)$ is replaced by $|\nabla\chi|$. Now, let $f \in H^{1,p}(X,d,\mathfrak{m})$, $g, \tilde{g} \in RS(f)$ and let $(f_n)_{n\in\mathbb{N}}$ and $(\tilde{f}_n)_{n\in\mathbb{N}}$ be the sequences associated to g and $\tilde{g}$ respectively as in Definition 5.7.6. If we replace $f, \tilde{f}$ by $f_n, \tilde{f}_n$ in the previous inequality and we pass to the limit as $n \to \infty$, then we get that

$$(1-\chi)g + \chi\tilde{g} \in RS(f), \quad \forall g, \tilde{g} \in RS(f) \tag{5.7.12}$$

because f_n and $\tilde{f}_n$ converge in $L^p(X,\mathfrak{m})$ to the same limit, so that $\mathrm{Lip}(\chi)|f_n - \tilde{f}_n| \to 0$ in $L^p(X,\mathfrak{m})$; (5.7.12) can be seen as a strong convexity property, because the convexity coefficients depend on x. Recall that for the moment $\chi \in \mathrm{Lip}(X,[0,1])$, but since $RS(f)$ is a closed set, by approximation we can extend (5.7.12) to any $\chi \in L^\infty(X,\mathfrak{m})$ with $0 \le \chi \le 1$. In order to conclude the proof of the pointwise minimality, let us argue by contradiction and assume that

$$B := \{x \in X \,:\, g(x) < |\nabla f|_*(x)\}$$

has positive $\mathfrak{m}$-measure. Then, applying (5.7.12) with $\tilde{g} = |\nabla f|_*$ and $\chi = \mathbb{1}_{X\setminus B}$, we infer that

$$\hat{g} := \begin{cases} |\nabla f|_* & \text{on } X \setminus B \\ g & \text{on } B \end{cases}$$

is a relaxed slope of f with L^p norm strictly smaller than the one of $|\nabla f|_*$ and this is impossible. Now let us show that Cheeger's energy is a local functional; to this aim, using the subadditivity of $f \mapsto |\nabla f|_*$ we can reduce to the case $g = 0$, i.e.

$$|\nabla f|_* = 0 \ \mathfrak{m}\text{-a.e. on } \{f = 0\}$$

In order to prove this fact we can not use an integration by parts formula. Hence, we will exploit the lower semicontinuity of Ch. Let $\varphi_n \in C^1$ with $0 \le \varphi_n' \le 1$, $\varphi_n(t) \to t$ and $\varphi_n'(0) = 0$ and observe that, for any function φ satisfying the same conditions of φ_n's, we have

$$\mathrm{lip}_a(\varphi \circ f) \le (\varphi_n' \circ f)\,\mathrm{lip}_a(f)$$

for any Lipschitz function f; by standard approximation procedures, this inequality entails a weak form of the chain rule (5.7.11), that is

$$|\nabla(\varphi \circ f)|_* \le (\varphi' \circ f)|\nabla f|_* \quad \mathfrak{m}\text{-a.e. in } X \tag{5.7.13}$$

where φ still satisfies the previous assumptions. By integration and by the fact that $0 \le \varphi_n' \le 1$, we easily infer that

$$\int_X |\nabla(\varphi_n \circ f)|_*^p d\mathfrak{m} \le \int_{\{f\neq 0\}} |\nabla f|_*^p d\mathfrak{m}$$

so that if we pass to the limit as $n \to \infty$ we conclude that

$$\int_X |\nabla f|_*^p d\mathfrak{m} \le \int_{\{f\neq 0\}} |\nabla f|_*^p d\mathfrak{m}$$

and this proves the desired locality. Now we can finally demonstrate the chain rule, taking into account that we have already proved (5.7.13) for $\varphi \in C^1 \cap \mathrm{Lip}$. We have to prove that equality actually holds for $\varphi \in Lip(\mathbb{R})$, assuming for sake of simplicity that $\varphi' \ge 0$ (in the general case, the proof becomes more technical). However, observe that by dropping the previous regularity assumption $\varphi \in C^1(\mathbb{R})$ a first problem arises in (5.7.11): what is the meaning of $\varphi' \circ f$? Indeed, if φ is only Lipschitz, then it is differentiable only $\mathfrak{m}$-a.e. and therefore $\varphi' \circ f$ is undefined on $f^{-1}(N)$, where N denotes the set of points of nondifferentiability of φ. A priori, the size of $f^{-1}(N)$ could be large, but a classical result in the theory of Sobolev spaces tells us that $|\nabla f|_* = 0$ $\mathfrak{m}$-a.e. in $f^{-1}(N)$, whenever $\mathscr{L}^1(N) = 0$ (as it is in this case). Observe that we have just proved this property when N is a singleton (in order to prove that Ch is local); in this case the argument is more sophisticated, but the basic ideas are the same: we fix $K \subset N$ compact, we pick a sequence $(\varphi_n)_{n\in\mathbb{N}}$ of C^1 functions converging to the identity and with derivative identically zero on K and we perform the same procedure.

Once we are able to give a meaning to (5.7.11) in the general case, we can notice that still by approximation (5.7.13) implies (5.7.11); in fact, given a Lipschitz function φ with $\varphi' \ge 0$, we can approximate it by means of $\varphi * \rho_\varepsilon - \varphi * \rho_\varepsilon(0)$ as $\varepsilon \downarrow 0$ and to these regularized functions (5.7.13) applies. Hence, passing to the limit we exactly obtain

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

and the difference w.r.t. (5.7.13) is the fact that here φ is Lipschitz with $\varphi' \ge 0$ (hence, without C^1 assumption). The conclusion is now close, because by subadditivity the inequality turns into equality and the trick is the following. Let $M := \|\varphi'\|_\infty$ and $\psi(t) := Mt - \varphi(t)$; by construction, ψ is non-decreasing, Lipschitz and $\psi(0) = 0$, so that (5.7.13) applies and, consequently,

$$\begin{aligned} |\nabla(Mf)|_* &= |\nabla(\psi \circ f + \varphi \circ f)|_* \le |\nabla(\psi \circ f)|_* + |\nabla(\varphi \circ f)|_* \\ &\le (\psi' \circ f + \varphi' \circ f)|\nabla f|_* = M|\nabla f|_* \end{aligned}$$

Since $|\nabla(Mf)|_* = M|\nabla f|_*$, all the inequalities above are equalities and thus the chain rule (5.7.11) is established. □

This complete the first part on differential calculus on metric measure spaces.

5.8 Lecture 8 - 18 September

Abstract

After the definition of the minimal relaxed slope in the previous lecture, we are now interested in the introduction of a Laplacian, in order to talk about the heat flow even in metric measure spaces. Several links to the classical theory in the Hilbertian case will be given. As a second purpose of the lecture, we provide some tools on absolutely continuous curves and a way to measure them, due to Fuglede and later generalized by Koskela and MacManus. This will give raise to the notion of Newtonian space.

Our next goal is the study of the Laplacian in a purely metric setting, but first some considerations are required. As a first warning, since we are not asking anything to $(X, d, \mathfrak{m})$ besides separability, completeness and finiteness of $\mathfrak{m}$ on compact sets, in general the Sobolev space $H^{1,p}(X, d, \mathfrak{m})$ may be trivial, i.e. $H^{1,p}(X, d, \mathfrak{m}) = L^p(X, \mathfrak{m})$, which corresponds to $\mathrm{Ch} \equiv 0$. This can be shown by working with singular measures, rather than with singular distances, as we are going to see immediately.

Example 5.8.1. Let us consider $(\mathbb{R}, d, \mathfrak{m})$, where $d(x, y) := |x - y|$ and, if $(q_n)_{n\in\mathbb{N}}$ is an enumeration of the rationals, $\mathfrak{m}$ is defined by

$$\mathfrak{m} := \sum_{n=1}^{\infty} 2^{-n} \delta_{q_n}$$

Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of open sets such that $A_n \supset \mathbb{Q}$ and $\mathscr{L}^1(A_n) \downarrow 0$ as $n \to \infty$. Also, let $(\varphi_n)_{n\in\mathbb{N}}$ be a sequence of Lipschitz functions such that $\varphi_n(0) = 0$, $\varphi_n(t) \to t$, $0 \le \varphi_n' \le 1$ and $\varphi_n' \equiv 0$ in A_n. As a final ingredient, let $f \in \mathrm{Lip}_0(\mathbb{R})$. Now consider $f \circ \varphi_n$ and observe that

$$\mathrm{lip}_a(f \circ \varphi_n) \equiv 0 \quad \text{in } A_n$$

so that $\mathrm{Ch}(f \circ \varphi_n) = 0$, because $\mathfrak{m}$ is concentrated on a subset of A_n. Since f is Lipschitz, $f \circ \varphi_n$ converges to f in $L^p(X, \mathfrak{m})$ and, by lower semicontinuity of Ch, we get that $\mathrm{Ch}(f) = 0$ for every $f \in \mathrm{Lip}_0(\mathbb{R})$. By density of Lipschitz functions and again by lower semicontinuity of Ch, we conclude that $\mathrm{Ch}(f) = 0$ for every $f \in L^p(X, \mathfrak{m})$, as desired. $\Diamond$

Note that the procedure we have just adopted in this example is in some sense dual to the one we used in the proof of the locality of Ch, because we have considered $f \circ \varphi_n$ instead of $\varphi_n \circ f$.

As a second warning, keep in mind that even for $p = 2$ Cheeger's energy may not be quadratic, although it is always p-homogeneous, and this motivates the following definition.

Definition 5.8.2. *A metric measure space* $(X, d, \mathfrak{m})$ *is said to be* asymptotically Hilbertian *(or* infinitesimal Hilbertian*) if* Ch_2 *is a quadratic form in* $L^2(X, \mathfrak{m})$*, where* Ch_2 *denotes Cheeger's energy for* $p = 2$.

By the way, sub-Riemannian spaces are asymptotically Hilbertian. An open problem linked to this definition is the characterization of the asymptotic Hilbertianity and the interest in this problem comes from the fact that, as we have already pointed out, the asymptotic Hilbertianity is the required hypothesis in order to make Lott-Villani-Sturm theory and Bakry-Émery's one equivalent. Now let us provide a simple example, showing that Ch_2 is not quadratic in general.

Example 5.8.3. Let us consider $(\mathbb{R}^2, d, \mathscr{L}^2)$, where $d(x, y) := \|x - y\|_\infty$. In this situation, for every $f \in C^1(\mathbb{R}^2)$ we have that

$$|\nabla f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|}{\|y - x\|_\infty} = \|\nabla f\|_1$$

and, by standard approximation techniques (e.g. convolution), we can pass to any function in $H^{1,2}(\mathbb{R}^2, d, \mathscr{L}^2)$, finally obtaining

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

It is evident that Ch_2 is 2-homogeneous, but it is not a quadratic form, since the parallelogram identity does not hold. $\Diamond$

In order to motivate in a deeper way the possible non-quadraticity of Ch_2 and explain where this lack comes from, let us say that what we have depicted in the previous example is common in Finsler geometry. Indeed, if we consider $(E, \|\cdot\|_E, \mathscr{L}^n)$, where $E = \mathbb{R}^n$ and $\|\cdot\|_E$ is any non-Hilbertian norm, then the map $f \mapsto df$ (associating to the function f its differential) is always linear and so there is no problem at the level of differentials. On the contrary, for the definition of the gradient we need the duality map $J : E^* \to E$, defined as follows: for any $v \in E^*$, $J(v)$ is made up by the elements of E such that

$$\langle v, J(v) \rangle_{E^* \times E} = \|v\|_{E^*} \|J(v)\|_E, \quad \|v\|_{E^*} = \|J(v)\|_E$$

where $\|\cdot\|_{E^*}$ is the dual norm, naturally induced by $\|\cdot\|_E$ on E^*. Observe that the conditions above do not characterize a unique point of E and for this reason J is possibly multi-valued. By means of this duality map, the gradient of f is defined by

$$\nabla f := J(df)$$

Since $\|\cdot\|_E$ is not Hilbertian, J is not linear (actually, J is linear if and only if $\|\cdot\|_E$ is Hilbertian) and thus the map $f \mapsto \nabla f$ is not linear too. As a consequence, Cheeger's energy is not quadratic. The lack of linearity of J does not depend on the fact that J is possibly multi-valued, because the problem is still present even if we ask $\|\cdot\|_E$ to be strictly convex (so that J is single-valued).

Keeping in mind these two relevant warnings, we can now approach the problem of a reasonable definition of Laplacian. To this aim, we start with some reminders of convex analysis. Let E be a Banach space, let $\phi : E \to \mathbb{R} \cup \{+\infty\}$ be convex and lower semicontinuous and let $x \in \{\phi < +\infty\}$; the *subdifferential of ϕ at x* is given by

$$\partial\phi(x) := \{p \in E^* \,:\, \phi(y) \geq \phi(x) + \langle p, y - x\rangle_{E^* \times E},\, \forall y \in E\}$$

and it is a convex and closed set, possibly empty. In the case E is reflexive and $\|\cdot\|_{E^*}$ is strictly convex, we define the *gradient of ϕ at x* as the unique element in $\partial\phi(x)$ with minimal norm and we shall denote it by $\nabla\phi(x)$. These classical notions are related to the *descending slope*, i.e. the one-sided slope defined by

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

because

$$|\nabla^- \phi|(x) \leq \|p\|_{E^*}, \quad \forall p \in \partial\phi(x) \tag{5.8.1}$$

Indeed, by the very definition of subdifferential it is easy to see that, for any $p \in \partial\phi(x)$,

$$\phi(x) - \phi(y) \leq \langle p, y - x\rangle_{E^* \times E} \leq \|p\|_{E^*}\|y - x\|_E$$

whence, taking the positive part, dividing by $\|y-x\|_E$ and passing to the limsup, (5.8.1) follows. In particular, this implies that

$$|\nabla^- \phi|(x) \leq \inf\{\|p\|_{E^*} \,:\, p \in \partial\phi(x)\}$$

and if E is reflexive and $\|\cdot\|_{E^*}$ is strictly convex, then the infimum is actually a (unique) minimum and $|\nabla^-\phi|(x) = \|\nabla\phi(x)\|_E$ (the fact that equality holds is a direct consequence of Hahn-Banach theorem). As a last remark, let us notice that $|\nabla^- f|$ is zero at the minima of f, even if the slope is not zero in general; thus, $|\nabla^- f|$ really represents how much f can decrease and it is not affected by how much it can increase.

This is all what we need to know on convex analysis. By means of these notions, we can finally provide a metric definition of Laplacian that we will handle in the case $p = 2$ for sake of simplicity.

Definition 5.8.4. *Let $f \in L^2(X, \mathfrak{m})$ with $\partial\,\mathrm{Ch}(f) \neq \emptyset$; in this case we say that f belongs to the domain of the Laplacian and we write $f \in D(\Delta)$. The* Laplacian *of f is denoted by Δf and is defined by $-\Delta f := \nabla\,\mathrm{Ch}(f)$, i.e. $-\Delta f$ is the unique element with minimal L^2 norm in $\partial\,\mathrm{Ch}(f)$.*

As a first trivial remark, note that $D(\Delta) \subset D(\mathrm{Ch})$; secondly, pay attention to the fact that we are not making any assumption of local Hilbertianity, so that Δ is not linear in general, even if 1-homogeneous. As a third fact, let us investigate the motivation behind this definition. On $\mathbb{R}^n$, if $\partial\,\mathrm{Ch}(f) \neq \emptyset$ and $p \in \partial\,\mathrm{Ch}(f)$, then for every $\varepsilon > 0$ and for every $g \in C_c^\infty(\mathbb{R}^n)$ we have that

$$\frac{1}{2}\int_{\mathbb{R}^n} |\nabla(f + \varepsilon g)|^2 dx - \frac{1}{2}\int_{\mathbb{R}^n} |\nabla f|^2 dx \geq \int_{\mathbb{R}^n} \langle p, \varepsilon g\rangle dx \tag{5.8.2}$$

Dividing by ε and letting $\varepsilon \downarrow 0$ we obtain

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

If we replace g by $-g$, the opposite inequality follows and therefore we can say that

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

that is $p = -\Delta f$ in the sense of distributions. Conversely, if $p = -\Delta f$ in $\mathscr{D}'(\mathbb{R}^n)$, then by convexity of Dirichlet's energy we recover (5.8.2). Thus the subdifferentiability at f of Dirichlet's energy holds if and only if $\Delta f \in L^2(\mathbb{R}^n, \mathscr{L}^n)$. This concludes the motivating part, so that we can now turn our attention to a first relevant property: the integration by parts formula or, better, the integration by parts inequality.

Proposition 5.8.5. *For every $f \in D(\Delta)$ and $g \in D(\mathrm{Ch})$ it holds*

$$-\int_X g \Delta f \, d\mathfrak{m} \leq \int_X |\nabla g|_* |\nabla f|_* \, d\mathfrak{m} \tag{5.8.3}$$

If $g = \varphi \circ f$ with φ Lipschitz and $\varphi(0) = 0$, then equality holds.

Proof. First of all, by the very definition of subdifferential and by subadditivity of the minimal relaxed slope, we have that for every $\varepsilon > 0$

$$\begin{aligned} -\int_X \varepsilon g \Delta f \, d\mathfrak{m} &\leq \frac{1}{2} \int_X |\nabla(f + \varepsilon g)|_*^2 \, d\mathfrak{m} - \frac{1}{2} \int_X |\nabla f|_*^2 \, d\mathfrak{m} \\ &\leq \frac{1}{2} \int_X \Big(|\nabla f|_* + \varepsilon |\nabla g|_* \Big)^2 d\mathfrak{m} - \frac{1}{2} \int_X |\nabla f|_*^2 \, d\mathfrak{m} \\ &= \varepsilon \int_X |\nabla g|_* |\nabla f|_* \, d\mathfrak{m} + \frac{\varepsilon^2}{2} \int_X |\nabla g|_*^2 \, d\mathfrak{m} \end{aligned}$$

If we divide this inequality by ε and we let $\varepsilon \downarrow 0$, then (5.8.3) follows. The proof of the equality when $g = \varphi \circ f$ is carried out by analogous arguments. □

Pay attention to the fact that, as usual, we ask $\varphi(0) = 0$ because we are not assuming $\mathfrak{m}$ to be finite. If on the contrary this is the case, then the condition $\varphi(0) = 0$ becomes irrelevant. This integration by parts formula will be exploited later. Now, we can talk about the heat flow in a metric measure space $(X, d, \mathfrak{m})$, but to this aim we require some reminders about the Komura-Brezis theory, i.e. the theory of evolution equations for maximal monotone operators; the ideal framework for this topic is that of Hilbert spaces (it is not clear yet whether it is possible or not to extend such theory to Banach spaces) and for this reason we are only considering the case $p = 2$. Let H be a Hilbert space and let $\Phi : H \to \mathbb{R} \cup \{+\infty\}$ be convex and lower semicontinuous; then for every $\overline{u} \in \overline{D(\Phi)}$

the Cauchy problem

$$\begin{cases} u'(t) \in -\partial\Phi(u(t)) & \text{for a.e. } t \\ u \in AC^2_{loc}(]0,+\infty[, H) \\ \lim_{t\downarrow 0} u(t) = \overline{u} \end{cases}$$

has a unique solution and it generates a contraction semigroup on $\overline{D(\Phi)}$, that is if $u(t), v(t)$ are the solutions associated to the initial conditions $\overline{u}, \overline{v}$, then $\|u(t) - v(t)\| \leq \|\overline{u} - \overline{v}\|$. Trivially, contractivity entails uniqueness and in order to prove contractivity we will exploit monotonicity. This can be seen quite easily, because the subdifferential inequality for $x \in D(\Phi)$ and $p \in \partial\Phi(x)$ reads as

$$\Phi(y) - \Phi(x) \geq \langle p, y - x\rangle, \quad \forall y$$

whereas the same inequality for $y \in D(\Phi)$ and $q \in \partial\Phi(y)$ is

$$\Phi(x) - \Phi(y) \geq \langle q, x - y\rangle, \quad \forall x$$

Hence, if we sum up these two inequalities, we obtain the so-called *monotonicity inequality*, namely

$$\langle p - q, x - y\rangle \geq 0 \tag{5.8.4}$$

whenever $p \in \partial\Phi(x)$ and $q \in \partial\Phi(y)$, and by means of it we can prove contractivity, because (5.8.4) motivates the inequality below (indeed $-u'(t) \in \partial\Phi(u(t))$ and $-v'(t) \in \partial\Phi(v(t))$ by the fact that u, v are solutions).

$$\frac{d}{dt}\frac{1}{2}|u(t) - v(t)|^2 = \langle u'(t) - v'(t), u(t) - v(t)\rangle \leq 0$$

The notation AC^2_{loc} in the previous Cauchy problem denotes those functions which are locally absolutely continuous with locally square-integrable derivative. Notice that we ask u to be only locally absolutely continuous because, *a priori*, the initial condition could have infinite energy and therefore there might be blow-ups at $t = 0$. The solution of the previous Cauchy problem enjoys interesting regularizing effects:

(i) $\Phi(u(t)) < +\infty$ for any $t > 0$, even if the initial condition has infinite energy; we can also give an estimate of the blow-up rate as $t \downarrow 0$, because

$$\Phi(u(t)) \leq \inf_{v\in H}\left\{\Phi(v) + \frac{1}{2t}\|v - \overline{u}\|^2\right\}$$

(ii) even if *a priori* $\partial\Phi(u(t))$ is multi-valued and therefore it is not clear that we have a unique solution, *a posteriori* there is no freedom in the choice of $u'(t) \in \partial\Phi(u(t))$, because

$$u'(t) = -\nabla\Phi(u(t)) \quad \text{for a.e. } t$$

and this means that $u'(t)$ always minimizes the norm. This shows that in some sense we have a least action principle;

(iii) the function $t \mapsto \Phi(u(t))$ is locally absolutely continuous on $]0,+\infty[$ and

$$\frac{d}{dt}(\Phi \circ u)(t) = -\|u'(t)\|\|\nabla\Phi(u(t))\| = -\|u'(t)\||\nabla^-\Phi|(u(t))$$

If we apply this general discussion to Cheeger's energy, then for any $\overline{u} \in L^2(X,\mathfrak{m})$ the Cauchy problem

$$\begin{cases} \partial_t u = \Delta u & \text{for a.e. } t \\ u \in AC^2_{loc}(]0,+\infty[, L^2(X,\mathfrak{m})) \\ \lim_{t\downarrow 0} u(t) = \overline{u} \end{cases} \tag{5.8.5}$$

admits a unique solution; observe that the initial condition can be any function in $L^2(X,\mathfrak{m})$ because the domain of Cheeger's energy is dense.

Remark 5.8.6. Even if the notation $H^{1,p}(X,d,\mathfrak{m})$ might suggest the existence of higher-order Sobolev spaces, by now the apex 1 is really useless, because we know very little in a general setting and so the theory of Sobolev spaces in metric measure spaces is still an open problem. Indeed, as a first step for the definition of (at least) $H^{2,p}(X,d,\mathfrak{m})$, we need a good notion of (gradient) vector field ∇f as section of the tangent bundle and then we need to differentiate it; for the moment, such a satisfactory notion has been found only in the asymptotically Hilbertian case. In a recent paper of Nicola Gigli, second order calculus and the definition of the Sobolev space $H^{2,2}(X,d,\mathfrak{m})$ have been carried out in a very satisfactory way, but curvature-dimension conditions in the sense of Lott-Villani-Sturm are required. $\Diamond$

Now let us turn our attention to the Lagrangian (or Beppo Levi's) approach with the following purpose: the proof of the fact that $BL^{1,p} = H^{1,p}$ also in a metric setting, even if for the moment the definition of $BL^{1,p}(X,d,\mathfrak{m})$ has not been precised yet. The inclusion $BL^{1,p} \supset H^{1,p}$ is the easier one, whereas for $BL^{1,p} \subset H^{1,p}$ much more efforts are required, but in both cases we first need some preliminary tools:

- metric derivative;
- the spaces $AC([0,1],X)$, $AC^p([0,1],X)$ and the notion of p-action on a curve $\mathcal{A}_p(\gamma)$;
- upper gradients and curvilinear integrals;
- p-modulus for families of curves.

Let us first deal with absolutely continuous curves and metric derivative. The classical definition of absolute continuity in the metric framework is the following.

Definition 5.8.7. *Let* $(X, d, \mathfrak{m})$ *be a metric measure space and let* $1 \le q \le +\infty$. *A curve* $\gamma : [0,1] \to X$ *is said to be* q-absolutely continuous *(and we will write* $\gamma \in AC^q([0,1], X)$*) if there exists* $g \in L^q(0,1)$ *such that*

$$d(\gamma(s), \gamma(t)) \le \int_s^t g(r)dr, \quad \forall 0 \le s \le t \le 1 \tag{5.8.6}$$

As a trivial remark, it is easy to see that $AC^\infty([0,1], X) = \mathrm{Lip}([0,1], X)$. A more interesting topic is the following question: does there exists a minimal g (in the sense of a.e. minimality)? The answer is affirmative and relies on the notion of metric derivative.

Theorem 5.8.8. *If* $\gamma \in AC([0,1], X)$, *then for* $\mathscr{L}^1$*-a.e.* $t \in [0,1]$ *the limit*

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

exists, is called the metric derivative of γ at t *and is the minimal function for which (5.8.6) holds, namely for every* g *satisfying (5.8.6) we have* $|\gamma'|(t) \le g(t)$ *for* $\mathscr{L}^1$*-a.e.* $t \in [0,1]$.

The metric derivative we have just introduced allows us to define the q-action on a curve γ as

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

and it is clear that $\mathcal{A}_q(\gamma)$ is finite if and only if $\gamma \in AC^q([0,1], X)$. Typically, and we will better see this fact in the next computations, $q = p'$; this means that for the study of $H^{1,p}$ we have to work with p'-absolutely continuous curves and with $\mathcal{A}_{p'}$, for the study of $H^{1,1}$ we have to consider Lipschitz curves. At this point, the curvilinear integral of a Borel function $g : X \to [0, +\infty]$ can be defined by

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

It does not depend on the parametrization of γ and it is useful to represent it in terms of push-forward, i.e.

$$\int_\gamma g = \int_X g \, dJ\gamma \tag{5.8.7}$$

where $J\gamma := \gamma_\#(|\gamma'|\mathscr{L}^1)$ (pay attention to the fact that this needs not be a probability measure); the validity of this representation is due to the change of variable formula (5.7.8). Let us set aside for a moment the notion of upper gradient and let us immediately investigate that of p-modulus. Such notion arose in the context of convex analysis and was later studied by Fuglede in 1957. Let $\Gamma \subset AC([0,1], X)$; the *p-modulus of* Γ is defined by

$$\mathrm{Mod}_p(\Gamma) := \inf \left\{ \int_X \rho^p \, d\mathfrak{m} \right\}$$

where the infimum is taken among all the non-negative Borel functions $\rho : X \to [0, +\infty]$ satisfying

$$\int_\gamma \rho \geq 1, \quad \forall \gamma \in \Gamma$$

and notice that this definition is not affected by reparametrization.

Hence, Mod_p is an outer measure on the set of curves from $[0,1]$ to X concentrated on $AC([0,1], X)$ and, by means of this notion, Fuglede was able to overcome the objection to Beppo Levi's definition of $BL^{1,p}$ space, i.e. the frame invariance of the absolute continuity condition. Indeed, Fuglede proved that $H^{1,p}(\mathbb{R}^n)$ is given by those functions $f \in L^p(\mathbb{R}^n)$ admitting a representative $\tilde{f}$ such that, for a suitable function $F \in L^p(\mathbb{R}^n, \mathbb{R}^n)$,

$$\tilde{f}(\gamma_1) - \tilde{f}(\gamma_0) = \int_\gamma F, \quad \text{for } \mathrm{Mod}_p\text{-a.e. } \gamma \tag{5.8.8}$$

where $\int_\gamma F$ is the standard notation for the Euclidean (and not metric) curvilinear integral. As in Beppo Levi's definition, condition (5.8.8) is not asked for every curve γ, because it would be too restrictive. For this reason, we allow the existence of exceptional curves and their number is controlled by Mod_p, which is clearly frame invariant. In this way, Fuglede bypassed the problem of Beppo Levi's definition, because for any fixed $i \in \{1, ..., n\}$ the family of all the lines parallel to the i-th axis in $\mathbb{R}^n$ has positive Mod_p-measure and therefore (5.8.8) holds along all such lines. Let us stress that the vector field F is unique and so we call it the *weak gradient* of f.

Fuglede's procedure was conceived for $\mathbb{R}^n$, but Koskela and MacManus on the one hand and Shanmugalingam on the other one generalized it to the metric setting in the following way: condition (5.8.8) is replaced by

$$|f(\gamma_1) - f(\gamma_0)| \leq \int_\gamma g \tag{5.8.9}$$

and if this inequality holds for every curve γ, then we say that g is an *upper gradient* of f and we write $g \in UG(f)$. For instance, in the case of Lipschitz functions the slope is an upper gradient and, naturally, also the asymptotic Lipschitz constant. The proof of this claim is very quick, because it suffices to observe that $t \mapsto f(\gamma(t))$ is absolutely continuous (thanks to the fact that f is Lipschitz) and therefore

$$|f(\gamma_1) - f(\gamma_0)| \leq \int_0^1 \left| \frac{d}{dt}(f \circ \gamma) \right| dt \leq \int_0^1 |\nabla f||\gamma'| dt = \int_\gamma |\nabla f| \tag{5.8.10}$$

where the second inequality is nothing but an application of the pointwise inequality

$$|\nabla(f \circ \gamma)| \leq |\nabla f||\gamma'|$$

At this point we can adapt Fuglede's characterization of Sobolev spaces to the metric framework. More precisely, inspired by Fuglede's work we shall define the *Newtonian space* $N^{1,p}(X, d, \mathfrak{m})$ as the set of functions f admitting a representative $\tilde{f}$ such that, for a suitable function $g \in L^p(X, \mathfrak{m})$ for Mod_p-a.e. γ.

Why is this definition meaningful? In fact, a lot of questions naturally arises on its correctness. For instance, notice that if $\int_\gamma g$ is infinite, then condition (5.8.9) is useless, so we have to prove that $\int_\gamma g < +\infty$ holds for sufficiently many curves. Fortunately, this property is true for Mod_p-a.e. curve γ provided that $g \in L^p(X, \mathfrak{m})$. The proof is straightforward, because if we define

$$\Gamma_M := \left\{ \gamma \in AC([0,1], X) \, : \, \int_\gamma g \geq M \right\}$$

then, observing that $\int_\gamma (g/M) \geq 1$, by the very definition of p-modulus we have that

$$\mathrm{Mod}_p(\Gamma_M) \leq \int_X \left(\frac{g}{M} \right)^p d\mathfrak{m} = M^{-p} \|g\|^p_{L^p(X,\mathfrak{m})}$$

The set Γ of curves along which the curvilinear integral of g is infinite can be written as

$$\Gamma = \bigcap_{M>0} \Gamma_M$$

so that, for any $M > 0$,

$$\mathrm{Mod}_p(\Gamma) \leq \mathrm{Mod}_p(\Gamma_M) \leq M^{-p} \|g\|^p_{L^p(X,\mathfrak{m})} \to 0, \quad \text{as } M \to +\infty$$

and this exactly means that Γ is Mod_p-negligible.

5.9 Lecture 9 - 22 September

Abstract

In this lecture we provide a different way to measure absolutely continuous curves and, as a byproduct, we introduce the metric analogue of the Beppo Levi space, where the notion of *weak upper gradient* plays the same crucial role of the relaxed slope for the $H^{1,p}$ Sobolev space. The relationship between $BL^{1,p}$ and $H^{1,p}$ is investigated and an isometric inclusion is established. For the opposite one the work is much more harder and several tools are needed; one of them, the Hopf-Lax formula, is introduced at the end of the lecture.

In the previous lecture we have introduced the Newtonian space $N^{1,p}(X, d, \mathfrak{m})$ and in this definition a fundamental role is played by the outer measure Mod_p, because it controls the number of exceptional curves, i.e. those curves where condition (5.8.9) does not hold. However, this way to measure curves is not the best one. For this reason, we now propose a different and more probabilistic approach due to Ambrosio, Gigli and Savaré, which first requires the following definition.

Definition 5.9.1. *A probability measure $\pi \in \mathscr{P}(C([0,1],X))$ is said to be a* p-test plan *if:*

(i) π is concentrated on $AC^q([0,1],X)$ with $q = p/(p-1)$, i.e. q is the Hölder conjugate of p;

(ii) there exists a non-negative constant $C = C(\pi)$ such that

$$(e_t)_\# \pi \le C\mathfrak{m}, \quad \forall t \in [0,1] \tag{5.9.1}$$

From a technical point of view, we will mostly exploit condition (ii) above in the following form: if ψ is a non-negative function, then

$$\int \psi(\gamma_t) d\pi(\gamma) \le C \int_X \psi \, d\mathfrak{m}$$

From a heuristic point of view, property (ii) is a non-concentration condition, because if $\mathfrak{m}$ is a non-atomic measure, then $(e_t)_\#\pi$ can not be atomic too; for this reason, $C(\pi)$ is also called *compression constant* and it is easy to see that the larger is $C(\pi)$, the smaller is the $\mathfrak{m}$-measure of the set where $(e_t)_\#\pi$ is concentrated. In this way, we have a family of probability measures, the p-test plans, and by duality we can provide the notion of negligible set.

Definition 5.9.2. *A set $\Gamma \subset AC([0,1],X)$ is said to be* p-negligible *if $\pi(\Gamma) = 0$ for any p-test plan π. A property holds* p-a.e. *if the set of curves for which the property is false is p-negligible.*

Duality enters in this definition because p-test plans are concentrated on $AC^{p'}([0,1],X)$.

Remark 5.9.3. To check p-negligibility it is sufficient to consider only those test plans $\tilde{\pi}$ such that $\mathcal{A}_q \in L^\infty(\tilde{\pi})$, namely the q-action is essentially bounded. Indeed, if π is any test plan, then we can consider

$$\pi_M := \frac{\pi \llcorner \{\gamma : \mathcal{A}_q(\gamma) \le M\}}{\pi(\{\gamma : \mathcal{A}_q(\gamma) \le M\})}$$

for any $M > 0$ and if $\pi_M(\Gamma) = 0$ for any $M > 0$ (for a given family Γ), then it is easy to see that $\pi(\Gamma) = 0$ too. $\diamondsuit$

By means of this last notion, we can finally introduce the Beppo Levi space $BL^{1,p}(X,d,\mathfrak{m})$ as the set of functions $f \in L^p(X,\mathfrak{m})$ for which there exists a function $g \in L^p(X,\mathfrak{m})$ (depending on f) such that (5.8.9) holds for p-a.e. γ. The function g is called a *weak upper gradient* of f and we will write $g \in WUG(f)$, analogously to what we have already done for relaxed slopes; the set $WUG(f)$ and its properties will be investigated later. Now, let us compare the notions of Mod_p-a.e. and p-a.e.

(i) As a first remark, condition (5.9.1) is not invariant under reparametrization, because if the supports of two curves $\gamma, \tilde{\gamma}$ intersect, then for suitable parametrizations there exists a time t when $\gamma_t = \tilde{\gamma}_t$. Therefore, at a given time t we can concentrate the mass at single points by choosing suitable parametrizations. On the contrary, we have already pointed out that Mod_p is parameter-free.

(ii) Given a family of curves Γ, if $\mathrm{Mod}_p(\Gamma) = 0$ then Γ is p-negligible. In order to prove this claim, let π be a test plan with $\mathcal{A}_q \in L^\infty(\pi)$ (this is not restrictive, thanks to Remark 5.9.3), let us denote

$$\tilde{C}(\pi) := \int \mathcal{A}_q(\gamma) d\pi(\gamma)$$

and let ρ be an admissible function in the definition of Mod_p, namely

$$\int_\gamma \rho \geq 1, \quad \forall \gamma \in \Gamma$$

By integrating this inequality w.r.t. π we obtain

$$\int \int_0^1 \rho(\gamma(t))|\gamma'|(t) dt d\pi(\gamma) \geq \pi(\Gamma)$$

and by applying Hölder's inequality to the left-hand side we see that

$$\begin{aligned}
&\int \int_0^1 \rho(\gamma(t))|\gamma'|(t) dt d\pi(\gamma) \\
\leq\ & \left(\int \int_0^1 \rho^p(\gamma(t)) dt d\pi(\gamma) \right)^{1/p} \left(\int \mathcal{A}_q(\gamma) d\pi(\gamma) \right)^{1/q} \\
\leq\ & C(\pi)^{1/p} \|\rho\|_{L^p(X,\mathfrak{m})} \tilde{C}(\pi)^{1/q}
\end{aligned}$$

Hence

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

and since $\mathrm{Mod}_p(\Gamma) = 0$, we can find admissible ρ's with arbitrarily small L^p norm, so that $\pi(\Gamma) = 0$ and by the arbitrariness of the test plan π, the conclusion follows. This fact tells us that $N^{1,p} \subset BL^{1,p}$, even if *a posteriori* the two spaces coincide.

(iii) The opposite implication does not hold in general and it is easy to find a counter-example, because in $\mathbb{R}^n$ the family of lines parallel to a given coordinate axis and parametrized with constant speed is p-negligible, whereas, as we have already said, its Mod_p-measure is strictly positive.

Another aspect that clarifies the difference between Mod_p-a.e. and p-a.e. is the following. A recent result of Ambrosio, Di Marino and Savaré provides a duality formula for Mod_p, namely a way to represent Mod_p as a supremum. In order to state it, let $\eta \in \mathscr{P}(C([0,1],X))$ be concentrated on $AC([0,1],X)$ and let us consider the map $\gamma \mapsto J\gamma$ from $AC([0,1],X)$ to $\mathscr{M}_+(X)$, where $J\gamma$ has already been introduced at (5.8.7) and $\mathscr{M}_+(X)$ denotes the family of non-negative finite measures over X. By means of this function we can define the baricenter of η as the average of all the measures $J\gamma$, that is

$$\mathrm{Bar}(\eta) := \int J\gamma \, d\eta(\gamma) \tag{5.9.2}$$

and we can also define its q-norm by setting

$$\|\mathrm{Bar}(\eta)\|_q := \begin{cases} \|g\|_{L^q(X,\mathfrak{m})} & \text{if } \mathrm{Bar}(\eta) = g\mathfrak{m} \\ +\infty & \text{otherwise} \end{cases}$$

Thus, notice that $\mathrm{Bar}(\eta)$ is a non-negative measure on X. In this setting, we can state an analogous version of condition (ii) of Definition 5.9.1 by asking $\mathrm{Bar}(\eta)$ to be not too much concentrated, more precisely $\|\mathrm{Bar}(\eta)\|_q < +\infty$. After these preliminary considerations, the duality formula reads as follows:

$$\mathrm{Mod}_p(\Gamma) = \sup\left\{ \frac{1}{\|\mathrm{Bar}(\eta)\|_q} \ : \ \eta \in \mathscr{P}(\Gamma) \right\} \tag{5.9.3}$$

The "$\leq$" inequality is very easy to prove, as usual in duality arguments; the procedure is direct and resembles a lot what we have already done to show that Mod_p-negligibility entails p-negligiblity. The opposite inequality requires more work, but essentially relies on Hahn-Banach theorem, which is a classical tool in duality results. As a final remark, let us stress that (5.9.3) is true also for family of measures and not only for family of curves Γ; in fact, Fuglede had already observed that the notion of p-modulus could be extended to family of hypersurfaces (so that we can deal with exceptionality at higher dimensions), but more generally this notion and the related results hold for family of measures.

Now let us come back to the space $BL^{1,p}(X,d,\mathfrak{m})$ and observe that its definition is invariant under modifications of f and g in $\mathfrak{m}$-negligible sets, which is not true (*a priori*) for the Newtonian space $N^{1,p}(X,d,\mathfrak{m})$, because its definition is given in terms of representatives. For this reason, it will be easier to compare $BL^{1,p}(X,d,\mathfrak{m})$ with $H^{1,p}(X,d,\mathfrak{m})$. Let us quickly prove the claim.

Proof. Let $f \in BL^{1,p}(X,d,\mathfrak{m})$ and let $\tilde{f}$ be a function such that $\mathfrak{m}(\{f \neq \tilde{f}\}) = 0$. Our aim is to show that

$$|\tilde{f}(\gamma_1) - \tilde{f}(\gamma_0)| \leq \int_\gamma g, \quad \text{for } p\text{-a.e. curve } \gamma \tag{5.9.4}$$

and this property follows if we are able to prove that, for any given $t \in [0,1]$,

$$f(\gamma_t) = \tilde{f}(\gamma_t), \quad \text{for } p\text{-a.e. curve } \gamma \tag{5.9.5}$$

Indeed, it is sufficient to apply (5.9.5) with $t = 0$ and $t = 1$ to get (5.9.4). Hence, let us show (5.9.5). To this aim, pick a test plan π and recall that, by definition, $(e_t)_\#\pi \ll \mathfrak{m}$; this implies that $(e_t)_\#\pi(\{f \neq \tilde{f}\}) = 0$ and this means that

$$f(\gamma_t) = \tilde{f}(\gamma_t), \quad \text{for } \pi\text{-a.e. curve } \gamma$$

By the arbitrariness of π, (5.9.5) follows. The same procedure applies in the case of modifications of g. □

After this introductory part, let us consider the set $WUG(f)$ and let us mimic the procedure we performed for $RS(f)$ in order to get the minimal relaxed slope. In that case, the key property was the closure of $RS(f)$ in the following sense: if $f_n \to f$ in $L^p(X, \mathfrak{m})$, $g_n \in RS(f_n)$ and $g_n \rightharpoonup g$ in $L^p(X, \mathfrak{m})$, then $g \in RS(f)$. This property easily entailed the weak closure of $RS(f)$, which is also convex, and at that point it was immediate to infer the existence (and uniqueness) of a minimal object $|\nabla f|_*$ that we called minimal relaxed slope. In the present situation, the same procedure is possible, because $WUG(f)$ is convex and the following closure property holds: if $f_n \to f$ in $L^p(X, \mathfrak{m})$, $g_n \in WUG(f_n)$ and $g_n \rightharpoonup g$ in $L^p(X, \mathfrak{m})$, then $g \in WUG(f)$. Therefore, also in this case there exists a unique element $g \in WUG(f)$ with minimal L^p norm; we shall denote it by

$$g =: |\nabla f|_{BL}$$

and we will refer to it as *minimal weak upper gradient*. As in the case of the minimal relaxed slope, the minimality of $|\nabla f|_{BL}$ is not only global, but also pointwise, namely

$$|\nabla f|_{BL} \leq g \ \ \mathfrak{m}\text{-a.e. in } X, \quad \forall g \in WUG(f) \tag{5.9.6}$$

Eventually, we will prove not only that $H^{1,p}(X, d, \mathfrak{m})$ and $BL^{1,p}(X, d, \mathfrak{m})$ coincide, but also that the same is true for $|\nabla f|_*$ and $|\nabla f|_{BL}$. Thus, the identification of the two Sobolev spaces is very sharp. Now let us prove the closure property of $WUG(f)$ and let us point out that in this case the situation is different from the one of $RS(f)$, because no diagonal arguments are possible.

Proof. By Mazur's lemma we can assume that $g_n \to g$ in $L^p(X, \mathfrak{m})$. We can also suppose without loss of generality that

$$\sum_{n=1}^{\infty} \Big(\|f_n - f\|_{L^p(X,\mathfrak{m})} + \|g_n - g\|_{L^p(X,\mathfrak{m})} \Big) < +\infty$$

In addition, let π be a test plan with $\mathcal{A}_q \in L^\infty(\pi)$, which is not restrictive by Remark 5.9.3. By construction, we know that for any $n \in \mathbb{N}$

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

and the aim is to prove that the inequality passes to the limit. If we manage to do it, then by the arbitrariness of π we have that $g \in WUG(f)$. On the one hand, the left-hand side can be handled by means of the same techniques we used in the previous proof, so that

$$|f_n(\gamma_1) - f_n(\gamma_0)| \to |f(\gamma_1) - f(\gamma_0)|, \quad \text{for } \pi\text{-a.e. curve } \gamma$$

On the other hand, we have to prove that

$$\int_\gamma g_n \to \int_\gamma g, \quad \text{for } \pi\text{-a.e. curve } \gamma \tag{5.9.7}$$

and this fact is an immediate consequence of a technical result in Fuglede's theory, which tells us that the convergence holds for Mod_p-a.e. curve and, *a posteriori*, for p-a.e. curve. But we can also prove (5.9.7) directly. Indeed

$$\begin{aligned}
\int\int_\gamma \sum_{n=1}^\infty |g_n - g| d\pi(\gamma) &= \int\int_0^1 \sum_{n=1}^\infty |g_n - g|(\gamma(t))|\gamma'|(t) dt d\pi(\gamma) \\
\leq \quad & \left(\int\int_0^1 \Big(\sum_{n=1}^\infty |g_n - g|(\gamma(t))\Big)^p dt d\pi(\gamma)\right)^{1/p} \left(\int \mathcal{A}_q(\gamma) d\pi(\gamma)\right)^{1/q} \\
= \quad & \left(\int_0^1 \int \Big(\sum_{n=1}^\infty |g_n - g|(x)\Big)^p d(e_t)_\# \pi dt\right)^{1/p} \left(\int \mathcal{A}_q(\gamma) d\pi(\gamma)\right)^{1/q} \\
\leq \quad & C(\pi)^{1/p} \Big\| \sum_{n=1}^\infty |g_n - g| \Big\|_{L^p(X,\mathfrak{m})} \left(\int \mathcal{A}_q(\gamma) d\pi(\gamma)\right)^{1/q} \\
\leq \quad & C(\pi)^{1/p} \sum_{n=1}^\infty \|g_n - g\|_{L^p(X,\mathfrak{m})} \left(\int \mathcal{A}_q(\gamma) d\pi(\gamma)\right)^{1/q} < +\infty
\end{aligned}$$

where the first inequality follows by Hölder's inequality, the second one by the fact that π is a test plan and the third one by the subadditivity of the L^p norm. The chain of inequalities we have just shown clearly entails (5.9.7) and by the arbitrariness of the test plan π the conclusion follows. ☐

At this point we can prove the following two facts at the same time:

$$\begin{aligned}
&H^{1,p}(X, d, \mathfrak{m}) \subset BL^{1,p}(X, d, \mathfrak{m}) \\
&|\nabla f|_{BL} \leq |\nabla f|_*, \quad \mathfrak{m}\text{-a.e. in } X
\end{aligned}$$

In the classical theory of Sobolev spaces, this inclusion is the easy one.

Proof. Let $f \in H^{1,p}(X, d, \mathfrak{m})$; by definition, there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}_0(X)$ such that $f_n \to f$ and $\mathrm{lip}_a(f_n) \to |\nabla f|_*$ in $L^p(X, \mathfrak{m})$. As we have already seen in (5.8.10), recall that $|\nabla f_n| \in UG(f_n)$, because f_n is Lipschitz continuous, and so in particular $|\nabla f_n| \in WUG(f_n)$. Since $\mathrm{lip}_a(f_n) \geq |\nabla f_n|$, we conclude that $\mathrm{lip}_a(f_n) \in WUG(f_n)$ and passing to the limit we get that $|\nabla f|_* \in WUG(f)$. By definition, this means that $f \in BL^{1,p}(X, d, \mathfrak{m})$ and the inequality follows from (5.9.6). ☐

As a next step, we will reverse both the inclusion and the inequality (for sake of simplicity, we will only deal with the case $p = 2$); then we will introduce a $W^{1,p}(X, d, \mathfrak{m})$ Sobolev space and we will prove that

$$H^{1,p}(X, d, \mathfrak{m}) \subset W^{1,p}(X, d, \mathfrak{m}) \subset BL^{1,p}(X, d, \mathfrak{m})$$

so that, *a posteriori*, all the three spaces are equal. Let us first focus our attention on the problem of the opposite inclusion and inequality. The tools we will need are the Hopf-Lax formula, reminders of optimal transport theory and the superposition principle, but before the study of these objects, let us briefly describe the strategy: we have to look at the energy dissipation rate of the entropy along the heat flow and this is firstly motivated by the following reason. Roughly speaking, we can say that the heat flow replace the convolution, which plays a fundamental role in the classical theory of Sobolev spaces. In more details, let $g \in BL^{1,2}(X, d, \mathfrak{m})$ and let us show how we can approximate it in an optimal way by Lipschitz functions; without loss of generality, we can assume that g is bounded away from 0 and $+\infty$, i.e.

$$0 < c \leq g \leq C < +\infty \quad \mathfrak{m}\text{-a.e. in } X$$

and in addition $\|g\|_{L^2(X,\mathfrak{m})} = 1$. Take $\overline{f} = g^2$ as initial condition for the heat flow of Cheeger's energy (f_t), i.e. for the Cauchy problem (5.8.5), and observe that this is possible since g is bounded, so that $\overline{f} \in L^2(X, \mathfrak{m})$; secondly, the fact that we are able to solve (5.8.5) for any initial condition in $L^2(X, \mathfrak{m})$ is crucial, because we do not know yet if $\overline{f}$ belongs to $D(\mathrm{Ch})$ (actually, this is precisely what we want to show). By the regularizing property of the heat flow (this is nothing but an immediate consequence of the classical theory of gradient flows) we have that $f_t \in D(\mathrm{Ch}) = H^{1,2}(X, d, \mathfrak{m})$ for any $t > 0$ and so we can really look at f_t as regularized approximations of the initial datum $\overline{f}$. In concrete terms, this means that we are allowed to write $|\nabla f_t|_*$.

Let us come back to the energy dissipation rate of the entropy along the heat flow. We will compute it in a Eulerian way and also in a Lagrangian way; it is exactly for the second technique that we need first to develop further tools of optimal transportation, because this is a Lagrangian theory. By performing these computations, we will find two important formulas: first of all, for a.e. $t > 0$ it holds

$$-\frac{d}{dt}\int_X f_t \log f_t \, d\mathfrak{m} = \int_X \frac{|\nabla f_t|_*^2}{f_t} \, d\mathfrak{m} \tag{5.9.8}$$

whence its integrated version at $t = 0$

$$\int_X (\overline{f}\log\overline{f} - f_t \log f_t) d\mathfrak{m} = \int_0^t \int_X \frac{|\nabla f_s|_*^2}{f_s} \, d\mathfrak{m} ds$$

and secondly

$$\int_X (\overline{f}\log\overline{f} - f_t \log f_t) d\mathfrak{m} \quad \leq \quad \frac{1}{2}\int_0^t \int_X \frac{|\nabla \overline{f}|_{BL}^2}{\overline{f}^2} f_s \, d\mathfrak{m} ds$$

$$+\frac{1}{2}\int_0^t\int_X \frac{|\nabla f_s|_*^2}{f_s}\,d\mathfrak{m}ds \tag{5.9.9}$$

The functional that appears at the right-hand side in (5.9.8) is called *Fisher information.* A straightforward comparison of the two integrated formulas provides us with the relation

$$\frac{1}{2}\int_0^t\int_X \frac{|\nabla f_s|_*^2}{f_s}\,d\mathfrak{m}ds \le \frac{1}{2}\int_0^t\int_X \frac{|\nabla \overline{f}|_{BL}^2}{\overline{f}^2} f_s\,d\mathfrak{m}ds$$

Notice that the left-hand side can be rewritten as $4\int_0^t \mathrm{Ch}(\sqrt{f_s})ds$ since via a change of variable we see that

$$\int_X \frac{|\nabla f_s|_*^2}{f_s}\,d\mathfrak{m} = 4\int_X |\nabla\sqrt{f_s}|_*^2\,d\mathfrak{m}$$

and so, by averaging in time, we have

$$\frac{4}{t}\int_0^t \mathrm{Ch}(\sqrt{f_s})ds \le \frac{1}{2}\frac{1}{t}\int_0^t\int_X \frac{|\nabla \overline{f}|_{BL}^2}{\overline{f}^2} f_s\,d\mathfrak{m}ds \tag{5.9.10}$$

Now, taking into account the L^2-lower semicontinuity of Ch and the fact that $\sqrt{f_s} \to \sqrt{\overline{f}}$ in $L^2(X,\mathfrak{m})$ as $s \downarrow 0$ (this is easy to check by virtue of the maximum principle stated in Proposition 5.9.4 below), we get

$$\mathrm{Ch}(\sqrt{\overline{f}}) \le \liminf_{t\downarrow 0}\frac{1}{t}\int_0^t \mathrm{Ch}(\sqrt{f_s})ds$$

and this allows us to estimate the left-hand side in (5.9.10). For the right-hand side, the bound $\overline{f} \ge c > 0$ ensures that $|\nabla \overline{f}|_{BL}^2/\overline{f} \in L^1(X,\mathfrak{m})$ and again the maximum principle, coupled with the convergence of f_s to $\overline{f}$ in $L^2(X,\mathfrak{m})$, grants that the convergence is also weak* in $L^\infty(X,\mathfrak{m})$; therefore

$$\int_X \frac{|\nabla \overline{f}|_{BL}^2}{\overline{f}}\,d\mathfrak{m} = \frac{1}{t}\lim_{t\downarrow 0}\int_0^t\int_X \frac{|\nabla \overline{f}|_{BL}^2}{\overline{f}^2} f_s\,d\mathfrak{m}ds$$

In conclusion, taking into account the definition of $\overline{f} := g^2$ with $g \in BL^{1,2}(X,d,\mathfrak{m})$, we proved

$$4\,\mathrm{Ch}(g) = 4\,\mathrm{Ch}(\sqrt{\overline{f}}) \le \frac{1}{2}\int_X \frac{|\nabla \overline{f}|_{BL}^2}{\overline{f}}\,d\mathfrak{m} = 4\int_X |\nabla g|_{BL}^2\,d\mathfrak{m} < +\infty$$

This shows that $g \in D(\mathrm{Ch}) = H^{1,2}(X,d,\mathfrak{m})$, so that we are now allowed to write $|\nabla g|_*$ and, together with the inequality $|\nabla g|_{BL} \le |\nabla g|_*$ $\mathfrak{m}$-a.e. in X, this gives us the conclusion, that is

$$|\nabla g|_{BL} = |\nabla g|_*, \quad \mathfrak{m}\text{-a.e. in } X$$

After this rough introduction, let us turn our attention to what remains to prove, namely (5.9.8) and (5.9.9). The latter requires a huge amount of auxiliary results that will be described in the next lecture, but (5.9.8) only needs the forthcoming proposition, where some additional properties of the gradient flows are emphasized. Hence, immediately after we will be able to demonstrate it.

Proposition 5.9.4. *Let* $\overline{f} \in L^2(X, \mathfrak{m})$ *and let* (f_t) *be the gradient flow of* Ch *starting from* $\overline{f}$*. Then the following properties hold:*

(i) mass preservation: if in addition $\overline{f} \in L^1(X, \mathfrak{m})$*, then for any* $t \geq 0$ *it holds* $f_t \in L^1(X, \mathfrak{m}) \cap L^2(X, \mathfrak{m})$ *and*

$$\int_X f_t \, d\mathfrak{m} = \int_X f_0 \, d\mathfrak{m}$$

(ii) maximum principle: if $f_0 \leq C$ *(resp.* $f_0 \geq c$*)* $\mathfrak{m}$*-a.e. in* X*, then* $f_t \leq C$ *(resp.* $f_t \geq c$*)* $\mathfrak{m}$*-a.e. in* X *for any* $t \geq 0$*.*

In our framework, the mass preservation corresponds to the stochastic completeness. In a general metric framework, stochastic completeness holds if, for suitable non-negative constants α, β independent of $x \in X$ we have

$$\mathfrak{m}(B(x, r)) \leq \alpha e^{\beta r^2}, \quad \forall x \in X, \, r > 0$$

Now we can prove (5.9.8).

$$\begin{aligned} -\frac{d}{dt} \int_X f_t \log f_t \, d\mathfrak{m} &= -\frac{d}{dt} \int_X (f_t \log f_t - f_t) d\mathfrak{m} = -\int_X \log f_t \partial_t f_t \, d\mathfrak{m} \\ &= -\int_X \log f_t \Delta f_t \, d\mathfrak{m} = \int_X |\nabla \log f_t|_* |\nabla f_t|_* \, d\mathfrak{m} \\ &= \int_X \frac{|\nabla f_t|_*^2}{f_t} \, d\mathfrak{m} \end{aligned}$$

In this chain of equalities, the first one is due to the mass preservation just stated above, the third one to the fact that f_t solves the heat equation, the fourth one follows by the integration by parts formula stated in Proposition 5.8.5 (in this case, equality holds) and finally the last one comes from the chain rule. Thus, (5.9.8) has been proved. On the contrary, for (5.9.9) we now need the additional tools we mentioned before.

The first one is the Hopf-Lax formula, whose natural frame is a metric space; this means that for the moment we are going to drop off the measure. Thus, let (X, d) be a metric space (the distance could even be extended) and let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function with quadratic growth from below. For sake of simplicity, in the following proofs we will assume that (X, d) is compact and f is Lipschitz continuous, but this assumptions are not required for the validity of the statements. In this setting, the *Hopf-Lax formula* reads as

$$Q_t f(x) := \min_{y \in X} F(t, x, y), \quad Q_0 f(x) := f(x) \tag{5.9.11}$$

where

$$F(t, x, y) := f(y) + \frac{1}{2t} d^2(x, y)$$

Let us stress that in the non compact setting we are not allowed to define Q_t as a minimum, but only as an infimum; therefore we should work with minimizing sequences. After this digression, note that by the triangle inequality it is clear that $Q_t \circ Q_s \geq Q_{s+t}$; if in addition (X, d) is a length space, then the opposite inequality also holds and in this case we can refer to Q_t as the *Hopf-Lax semigroup*. In the smooth case, it is well-known that formula (5.9.11) is strictly related to the Hamilton-Jacobi equation

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

because $u(t, x) := Q_t f(x)$ is the unique solution of the problem above in the viscosity sense. In the metric setting we are not particularly interested in viscosity theory, because it becomes problematic; we only need a pointwise property: more precisely, our aim is to show that in general metric spaces (5.9.11) provides a pointwise subsolution to the Hamilton-Jacobi equation. This result was proved independently by Ambrosio-Gigli-Savaré and by Gozlan-Samson-Roberto in 2011. It amounts to show that

$$\frac{d}{dt} Q_t f(x) + \frac{1}{2}|\nabla Q_t f|^2(x) \leq 0, \quad \forall x \in X$$

with at most countably many exceptions in t, but actually we will perform a more refined analysis, by proving that the same inequality is true if we replace the weak gradient by the asymptotic Lipschitz constant, i.e.

$$\frac{d}{dt} Q_t f(x) + \frac{1}{2} \operatorname{lip}_a^2(Q_t f, x) \leq 0, \quad \forall x \in X \tag{5.9.12}$$

To this aim, we have to introduce a bit more of notation: for $t > 0$ and $x \in X$ let

$$\begin{aligned} A(x, t) \quad &:= \quad \operatorname{argmin} F(t, x, \cdot) \\ D^+(x, t) := \max_{y \in A(x,t)} d(x, y) \qquad & D^-(x, t) := \min_{y \in A(x,t)} d(x, y) \end{aligned}$$

and let us also set $D^{\pm}(x, 0) := 0$.

5.10 Lecture 10 - 23 September

Abstract

This lecture is extremely technical. First aim of the forthcoming discussion is the investigation of Hopf-Lax formula's properties and the fact that it provides a pointwise subsolution of an Hamilton-Jacobi equation.

Secondly, after a quick introduction to optimal transport, we point out the link between the dual Monge-Kantorovich problem and Hopf-Lax formula. Thirdly, a powerful result due to Kuwada is proved. Finally, the last technical tool we need is given: the superposition principle, in the form of Lisini's theorem; it will enable us to pass from Wasserstein geodesics to geodesics on the underlying metric space.

In the previous lecture we have introduced all the required notations, so that now we can prove (5.9.12); let us recall that we are assuming (X, d) to be compact and $f \in \mathrm{Lip}(X)$. As a first step, there are some easy properties:

- $Q_t f$ is continuous in $X \times [0, +\infty[$;
- for any $t > 0$, $x \mapsto Q_t f(x)$ is Lipschitz continuous;
- $D^-(x, t) \leq D^+(x, t) \leq 2t \,\mathrm{Lip}(f)$;
- $(x, t) \mapsto A(x, t)$ is upper semicontinuous in the sense of multifunctions, that is if $y_n \in A(x_n, t_n)$, $(x_n, t_n) \to (x, t)$ and $y_n \to y$, then $y \in A(x, t)$; roughly speaking, this means that limits of minimizers are minimizers.

As a second step, let us point out the following fact.

Remark 5.10.1. Let $g^+, g^- : \mathbb{R}_+ \to \mathbb{R}_+$ be such that $g^+ \geq g^-$ and $g^+(s) \leq g^-(t)$ when $s < t$. Then $g^\pm$ are non-decreasing, they have the same jump set (which is countable) and they coincides outside it. $\Diamond$

We are going to apply this remark to $D^\pm(x, \cdot)$, so that we deduce that $D^\pm(x, \cdot)$ are non-decreasing and coincide out of a countable set. The only thing we have to check in order to invoke the remark is the following: $D^+(x, s) \leq D^-(x, t)$ when $s < t$.

Proof. Let $s < t$ and let x_s, x_t be minimizers of $F(s, x, \cdot)$ and $F(t, x, \cdot)$ respectively. As a consequence

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

If we add these two inequalities, then we obtain

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

and since $1/s > 1/t$, this entails that

$$d^2(x_t, x) - d^2(x_s, x) \geq 0$$

Hence, if we choose x_t and x_s in such a way that $d(x_t, x) = D^-(x, t)$ and $d(x_s, x) = D^+(x, s)$, then the conclusion follows. $\square$

Another important property of $D^{\pm}$ is stated in the following lemma.

Lemma 5.10.2. *The map D^- (resp. D^+) is lower (resp. upper) semicontinuous in $X\times\,]0,+\infty[$.*

Proof. Let us prove that D^+ is upper semicontinuous in $X\times\,]0,+\infty[$ (the proof of the lower semicontinuity of D^- is analogous). To this aim, fix any $(x,t)\in X\times\,]0,+\infty[$ and consider an arbitrary sequence $((x_n,t_n))_{n\in\mathbb{N}}\subset X\times\,]0,+\infty[$ convergent to (x,t). Then we have to show that

$$\limsup_{n\to\infty} D^+(x_n,t_n)\le D^+(x,t)$$

For any $n\in\mathbb{N}$, let $y_n\in X$ be a minimum of $F(t_n,x_n,\cdot)$ for which $D^+(x_n,t_n)=d(x_n,y_n)$ and, up to extract a suitable subsequence, let us assume that $(y_n)_{n\in\mathbb{N}}$ is convergent and let us denote by y its limit. By the fact that $(x,t)\mapsto A(x,t)$ is upper semicontinuous, we can say that $y\in A(x,t)$ and therefore

$$\lim_{n\to\infty} d(x_n,y_n)=d(x,y)\le D^+(x,t)$$

whence the conclusion, since

$$\limsup_{n\to\infty} D^+(x_n,t_n)=\limsup_{n\to\infty} d(x_n,y_n)=\lim_{n\to\infty} d(x_n,y_n)$$

□

Now let us pass to the proof of (5.9.12) and observe that part of the statement consists in the existence of the time derivative of Q_tf. The claim follows immediately by the next two propositions.

Proposition 5.10.3. *The map $t\mapsto Q_tf$ is Lipschitz continuous from $[0,+\infty[$ to $C(X)$ equipped with the sup norm and, for all $x\in X$, its derivative satisfies*

$$\frac{d}{dt}Q_tf(x)=-\frac{(D^{\pm}(x,t))^2}{2t^2} \tag{5.10.1}$$

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

Proof. For the moment, let us prove that $t\mapsto Q_tf(x)$ is Lipschitz continuous on $]\varepsilon,+\infty[$ as real valued function for any $\varepsilon>0$ and $x\in X$ fixed. For this purpose, let $t<s$ and choose x_t,x_s minima of $F(t,x,\cdot)$ and $F(s,x,\cdot)$ respectively. Then we have

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

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

whence the Lipschitz continuity of $t \mapsto Q_t f(x)$ on $]\varepsilon, +\infty[$ for any $\varepsilon > 0$ and $x \in X$. The first step is thus proved. Now, dividing by $s - t$ the expressions above, by definition of $D^{\pm}$ we infer that

$$-\frac{(D^-(x,s))^2}{2ts} \leq \frac{Q_s f(x) - Q_t f(x)}{s-t} \leq -\frac{(D^+(x,t))^2}{2ts}$$

and since $D^-(x,\cdot)$ and $D^+(x,\cdot)$ are both non-decreasing, they have at most countably many discontinuity points. Moreover, $D^-(x,t) = D^+(x,t)$ for any $t > 0$ with at most countably many exceptions. Hence, taking the limit as $s \to t$ we get

$$\frac{d}{dt} Q_t f(x) = -\frac{(D^{\pm}(x,t))^2}{2t^2}$$

provided that $D^-(x,\cdot)$ is continuous at t and $D^-(x,t) = D^+(x,t)$. So (5.10.1) holds. It remains to prove that $t \mapsto Q_t f$ is Lipschitz continuous from the whole $[0, +\infty[$ to $C(X)$, but from the fact that $D^-(x,t) \leq D^+(x,t) \leq 2t \operatorname{Lip}(f)$, we deduce that

$$\left| \frac{d}{dt} Q_t f(x) \right| \leq 2 \operatorname{Lip}^2(f)$$

for every $x \in X$ and for $\mathscr{L}^1$-a.e. $t \in]0, +\infty[$ and since in addition $Q_t f$ converges pointwise to f as $t \downarrow 0$, then $t \mapsto Q_t f$ is Lipschitz from $[0, +\infty[$ to $C(X)$. $\square$

Proposition 5.10.4. *For any* $(x,t) \in X \times]0, +\infty[$ *it holds*

$$\operatorname{lip}_a(Q_t f, x) \leq \frac{D^+(x,t)}{t} \tag{5.10.2}$$

Proof. As a first remark, let us point out that if we are able to prove that

$$\operatorname{Lip}(Q_t f, B(\overline{x}, r)) \leq \frac{1}{t}\left(r + \sup_{B(\overline{x},r)} D^+(\cdot, t) \right)$$

then (5.10.2) follows by taking the limit as $r \downarrow 0$, because of the upper semicontinuity of D^+. Pick $x, x' \in B(\overline{x}, r)$, $y \in A(x,t)$ and $y' \in A(x',t)$. In order to get the conclusion we have to prove two inequalities; the first one is the following

$$\frac{Q_t f(x) - Q_t f(x')}{d(x,x')} \leq \frac{1}{t}\left(r + \sup_{B(\overline{x},r)} D^+(\cdot, t) \right) \tag{5.10.3}$$

and it is an immediate consequence of the following chain of inequalities

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

$$
\begin{aligned}
&\leq \frac{d(x,x')}{2t}(d(x,x')+2d(x',y')) \\
&\leq \frac{d(x,x')}{2t}(2r+2D^+(x',t)) \\
&\leq \frac{d(x,x')}{t}\left(r+\sup_{B(\overline{x},r)} D^+(\cdot,t)\right)
\end{aligned}
$$

The second inequality we need is

$$
\frac{Q_tf(x')-Q_tf(x)}{d(x,x')} \leq \frac{1}{t}\left(r+\sup_{B(\overline{x},r)} D^+(\cdot,t)\right) \tag{5.10.4}
$$

and in this case observe that

$$
\begin{aligned}
Q_tf(x')-Q_tf(x) &\leq F(t,x',y)-F(t,x,y) \\
&= f(y)+\frac{d^2(x',y)}{2t}-f(y)-\frac{d^2(x,y)}{2t} \\
&\vdots \\
&\leq \frac{d(x,x')}{t}\left(r+\sup_{B(\overline{x},r)} D^+(x',t)\right)
\end{aligned}
$$

Then the conclusion follows in the same manner depicted above. Clearly (5.10.3) and (5.10.4) give (5.10.2). ☐

Hence (5.9.12) is proved and this is all we need to know about the Hopf-Lax formula.

Now we can provide a quick overview of the second topic we require, namely optimal transport theory. As usual, let (X,d) be a complete separable metric space and let $\mu,\nu \in \mathscr{P}_2(X)$, where $\mathscr{P}_2(X)$ denotes the family of probability measures on X with finite second momentum. The original optimal transport problem was formulated by Monge in 1781 and consisted in finding a map $T: X \to X$ that realizes the infimum

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

If a map T satisfies the constraint $T_\#\mu=\nu$, then it is called a *transport map from μ to ν*; if in addition T attains the infimum, then it is called an *optimal transport map from μ to ν*. Even if it is known that in some cases this problem admits solutions, it is much better to work with a different problem, the so-called Kantorovich's optimal transport problem, formulated in 1940; in this case, we are looking for a probability measure $\eta \in \mathscr{P}(X\times X)$ that realizes the infimum

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

A measure η satisfying the constraints $\pi^1_\#\eta=\mu$ and $\pi^2_\#\eta=\nu$ is called a *transport plan from μ to ν*; if in addition η attains the infimum, then we will refer to it as an *optimal transport map from μ to ν*. Heuristically speaking, $\eta(A\times B)$ denotes the amount of mass in A which is moved to B.

The link between the two problems is easy to state: transport maps induce transport plans by defining

$$\eta := (\mathrm{Id}, T)_{\#}\mu$$

that is $\eta(A \times B) := \mu(A \cap T^{-1}(B))$. As already said, optimal transport is a Lagrangian theory, because it is based on the way we move points and this aspect can be appreciated in an even clearer way if we assume that (X, d) is geodesic, because under this particular assumption we can introduce a third problem, namely the dynamical optimal transport problem. In order to formulate it, let us begin by defining the space $\mathrm{Geo}(X)$ of constant speed geodesics $\gamma : [0, 1] \to X$ and recall the definition of the two action $\mathcal{A}_2$, i.e.

$$\mathcal{A}_2(\gamma) := \int_0^1 |\dot{\gamma}_t|^2 \, dt$$

By means of it, it is possible to characterize constant speed geodesics, because for any $\gamma \in AC^2([0, 1], X)$ it holds

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

but if $\mathcal{A}_2(\gamma) = d^2(\gamma_0, \gamma_1)$, then the two inequalities above are actually equalities and this means that γ is length minimizing (from the fact that $\mathcal{A}_1(\gamma) = d(\gamma_0, \gamma_1)$) and has constant speed (from the fact that $\mathcal{A}_2(\gamma) = (\mathcal{A}_1(\gamma))^2$). Hence, given $\gamma \in AC^2([0, 1], X)$ we have that

$$\gamma \in \mathrm{Geo}(X) \quad \Leftrightarrow \quad \mathcal{A}_2(\gamma) = d^2(\gamma_0, \gamma_1)$$

As already anticipated, these considerations suggest a dynamical optimal transport problem, where we are looking for a probability measure $\Sigma \in \mathscr{P}(C([0, 1], X))$ realizing the infimum

$$\inf \left\{ \int \mathcal{A}_2(\gamma) d\Sigma(\gamma) \,:\, (e_0)_{\#}\Sigma = \mu,\, (e_1)_{\#}\Sigma = \nu \right\} \tag{D}$$

We can see a strong analogy between this problem and the notion of test plan we introduced in the study of metric Beppo Levi spaces. Also in this case, it is easy to recover Kantorovich's optimal transport problem, because given a solution Σ to the dynamical optimal transport problem, the probability measure $\eta := (e_0, e_1)_{\#}\Sigma$ is an optimal transport plan. Although completely equivalent to Kantorovich's one, the dynamical optimal transport problem is richer in terms of information, because of the following theorem, which also explains why optimal transportation is a Lagrangian theory.

Theorem 5.10.5. *Let (X, d) be a geodesic space. Then (D) and (K) are equivalent and the following two conditions are equivalent:*

(i) Σ is optimal for (D);

(ii) $\eta = (e_0, e_1)_\# \Sigma$ *is optimal for (K) and* Σ *is concentrated on* $\mathrm{Geo}(X)$.

Now let us investigate the connection between optimal transport and the Hopf-Lax formula. As a first step, let us define the Wasserstein distance between $\mu, \nu \in \mathscr{P}_2(X)$ as

$$W_2(\mu, \nu) := \sqrt{\min(K)}$$

Let us stress that it is actually a distance and that $(\mathscr{P}_2(X), W_2)$ inherits many properties of (X, d); for instance, it is complete and separable as well and if we further assume that (X, d) is compact or geodesic, then the same is true for $(\mathscr{P}_2(X), W_2)$. As a second step, if we consider the dual formulation of (K), then we have

$$\frac{1}{2} W_2^2(\mu, \nu) = \sup_{\varphi, \psi \in \mathrm{Lip}_b(X)} \left\{ -\int_X \varphi \, d\mu + \int_X \psi \, d\nu \, : \, \psi(y) - \varphi(x) \leq \frac{1}{2} d^2(x, y) \right\} \tag{5.10.5}$$

Indeed, the proof of the $\geq$ inequality (the easy one) is the following: let η be a transport plan from μ to ν and let $\varphi, \psi \in \mathrm{Lip}_b(X)$ be such that

$$\frac{1}{2} d^2(x, y) \geq \psi(y) - \varphi(x), \quad \forall x, y \in X \tag{5.10.6}$$

If we integrate this inequality w.r.t. η and take into account the fact that $\pi^1_\# \eta = \mu$ and $\pi^2_\# \eta = \nu$, we see that

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

whence the conclusion, by taking the minimum over η and the supremum over φ, ψ. The opposite inequality is more technical and relies on the min-max principle. With these ingredients, we can finally establish the relation between optimal transport theory and Hopf-Lax formula, because if we fix φ, then in the dual problem we have to maximize over ψ and this means that we have to consider the largest function ψ compatible with the constraint (5.10.6). Hence, if we rewrite (5.10.6) as

$$\psi(y) \leq \varphi(x) + \frac{1}{2} d^2(x, y), \quad \forall x, y \in X$$

and we minimize the right-hand side in x, it is clear that the largest ψ is precisely $Q_1 \varphi$. Therefore, (5.10.5) can be rephrased as

$$\frac{1}{2} W_2^2(\mu, \nu) = \sup_{\varphi \in \mathrm{Lip}_b(X)} \left\{ -\int_X \varphi \, d\mu + \int_X Q_1 \varphi \, d\nu \right\}$$

and the function φ is called *Kantorovich* or *optimal potential*. By interpolation, it is possible to show that if φ is optimal for the dual problem relative to μ and ν and $(\mu_t)_{t \in [0,1]}$ is a constant speed geodesic from μ to ν with $\mu_t \in \mathscr{P}_2(X)$, then

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

and this shows that along Wasserstein geodesics the optimal potential evolves according to the Hopf-Lax formula. This is extremely relevant, because it allows a completely differential description of geodesics, namely the following system of equations

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0 \\ v_t = \nabla \varphi_t \\ \partial_t \varphi_t + \dfrac{1}{2} |\nabla \varphi_t|^2 = 0 \end{cases}$$

After the Hopf-Lax formula and the quick overview on optimal transportation, we need a third ingredient: Kuwada's lemma. In a paper appeared in 2010 Kuwada proved, roughly speaking, that the $CD(K,\infty)$ condition, which can be equivalently formulated as

$$|\nabla \mathrm{P}_t f|^2 \le e^{-2Kt} |\nabla f|^2$$

where P_t is the heat semigroup, that condition is equivalent to the contraction of W_2 under the heat flow, namely if $f, g \ge 0$ with

$$\int_X f \, d\mathfrak{m} = \int g \, d\mathfrak{m} = 1$$

then

$$W_2((\mathrm{P}_t f)\mathfrak{m}, (\mathrm{P}_t g)\mathfrak{m}) \le e^{-Kt} W_2(f\mathfrak{m}, g\mathfrak{m})$$

For one of the two implications a key role is played by the so-called Kuwada's lemma, whose statement is the following.

Lemma 5.10.6 (Kuwada)**.** *Let $(X, d, \mathfrak{m})$ be a metric measure space with $\mathfrak{m}(X) < +\infty$, let $f_0 \in L^2(X, \mathfrak{m})$ with $0 < c \le f_0 \le C < +\infty$ $\mathfrak{m}$-a.e. in X for suitable constants c, C and*

$$\int_X f_0 \, d\mathfrak{m} = 1$$

Let (f_t) be the gradient flow of Ch *starting from f_0 and let us define $\mu_t := f_t \mathfrak{m}$ for any $t \ge 0$. Then $(\mu_t) \in AC^2_{loc}([0, +\infty[, \mathscr{P}_2(X))$ and its metric derivative can be estimated from above by the Fisher information, i.e.*

$$|\dot{\mu}_t|^2 \le \int_X \frac{|\nabla f_t|_*^2}{f_t} \, d\mathfrak{m}, \quad \textit{for a.e. } t \tag{5.10.7}$$

Let us point out that this result connects the Eulerian and the Lagrangian points of view, because the left-hand side in (5.10.7) is clearly Lagrangian as it involves the dynamics, whereas the right-hand side is Eulerian. Let us also stress that for any $t > 0$ it holds

$$0 < c \le f_t \le C < +\infty$$

because of the mass preservation (see property (i) of Proposition 5.9.4) and

$$\int_X f_t \, d\mathfrak{m} = 1$$

thanks to the maximum principle (see property (ii) of Proposition 5.9.4).

Proof. Let us fix $t < s$, let us set $\ell := s - t$ and let us estimate $W_2^2(\mu_s, \mu_t)$ using the dual formula (5.10.5). To this aim, choose $\varphi \in \mathrm{Lip}_b(X)$ and observe that

$$\begin{aligned}
-\int_X \varphi\, d\mu_t + \int_X Q_1\varphi\, d\mu_s &= \int_0^\ell \frac{d}{dr}\Big(\int_X Q_{r/\ell}\varphi\, d\mu_{t+r}\Big) dr \\
&= \int_0^\ell \int_X \Big(\frac{d}{dr}(Q_{r/\ell}\varphi) f_{t+r} + Q_{r/\ell}\varphi \frac{d}{dr} f_{t+r}\Big) d\mathfrak{m} dr \\
&\leq \int_0^\ell \int_X \Big(-\frac{1}{2\ell}\,\mathrm{lip}_a^2(Q_{r/\ell}\varphi) f_{t+r} + Q_{r/\ell}\varphi \Delta f_{t+r}\Big) d\mathfrak{m} dr \\
&\leq \int_0^\ell \int_X \Big(-\frac{1}{2\ell}\,\mathrm{lip}_a^2(Q_{r/\ell}\varphi) f_{t+r} + |\nabla f_{t+r}|_* |\nabla Q_{r/\ell}\varphi|_*\Big) d\mathfrak{m} dr \\
&\leq \int_0^\ell \int_X \Big(-\frac{1}{2\ell}\,\mathrm{lip}_a^2(Q_{r/\ell}\varphi) f_{t+r} + |\nabla f_{t+r}|_*\,\mathrm{lip}_a(Q_{r/\ell}\varphi)\Big) d\mathfrak{m} dr \\
&\leq \frac{\ell}{2}\int_0^\ell \int_X \frac{|\nabla f_{t+r}|_*^2}{f_{t+r}}\, d\mathfrak{m} dr
\end{aligned}$$

The technical aspects motivating this chain of equalities and inequalities will be skipped, but let us mention the most important ones. The first inequality is due to the subsolution property (5.9.12) and we can apply it because by Fubini's theorem it holds a.e. for a.e. t; the second inequality is motivated by the integration by parts formula (Proposition 5.8.5); the third inequality comes from the fact that $|\nabla Q_{r/\ell}\varphi|_*$ is the relaxation of $\mathrm{lip}_a(Q_{r/\ell}\varphi)$ and therefore it is smaller than the latter; finally, the fourth inequality follows by a slightly modified Young's inequality, i.e.

$$ab \leq \frac{1}{2\ell}a^2 + \frac{\ell}{2}b^2$$

with $a = \mathrm{lip}_a(Q_{r/\ell}\varphi)\sqrt{f_{t+r}}$ and $b = |\nabla f_{t+r}|_*/\sqrt{f_{t+r}}$. Thus, we have obtained

$$-\int_X \varphi\, d\mu_t + \int_X Q_1\varphi\, d\mu_s \leq \frac{\ell}{2}\int_0^\ell \int_X \frac{|\nabla f_{t+r}|_*^2}{f_{t+r}}\, d\mathfrak{m} dr$$

and note that the right-hand side is independent of φ. Hence, if we take the supremum over all $\varphi \in \mathrm{Lip}_b(X)$, by the duality formula we get

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

and since $t \mapsto |\nabla f_t|_*^2/f_t$ is locally absolutely integrable, we infer that $(\mu_t) \in AC^2_{loc}([0,+\infty[, \mathscr{P}_2(X))$. If we divide the inequality above by $(s-t)^2$ and we take the limit as $t \to s$ at the Lebesgue points of the right-hand side, then (5.10.7) follows. The proof is completed. □

Now let us endeavor the last tool we need: the so-called *superposition principle*; it will enable us to pass from curves in $\mathscr{P}_2(X)$ to curves in X and this is exactly what we need, because after Kuwada's lemma we know that the heat flow can be seen as a curve in $\mathscr{P}_2(X)$. This technique was firstly introduced by L.C. Young in the original setting of Calculus of Variations and Optimal Transport, but in more recent times it has been adapted to more general situations; in particular, Smirnov formulated it for currents and it proved that any normal current can be viewed as a kind of superposition of currents associated to elementary curves; in 2005, Ambrosio, Gigli and Savaré also considered the superposition principle for solutions to the continuity equation.

For the general metric theory we need Lisini's theorem, but for the moment let us explain the case of solutions to the continuity equation in $\mathbb{R}^n$ (or in any Riemannian manifold). Let b be a smooth vector field (so that the usual existence and uniqueness results are guaranteed) and let X be the flow associated to b; this means that X solves the following ODE

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

By well-known results in the theory of first order PDEs, this equation is strictly related to the continuity equation

$$\frac{d}{dt}u_t + \mathrm{div}(b_t u_t) = 0 \tag{5.10.8}$$

and if for instance we work with non-negative solutions u_t, then u_t can be expressed in terms of the push-forward under the flow map X of the initial datum $\overline{u}$, namely the unique solution to (5.10.8) for the initial condition $\overline{u}$ is given by

$$u_t = X(t,\cdot)_\# \overline{u}$$

Theorem 5.10.7 (Ambrosio-Gigli-Savaré)**.** *Let (μ_t) be a solution to the continuity equation*

$$\frac{d}{dt}\mu_t + \mathrm{div}(v_t \mu_t) = 0$$

where the only regularity condition we ask v_t is the following:

$$\|v_t\|_{L^2(\mu_t)} \in L^1(0,1)$$

Then $(\mu_t) \in AC^2([0,1], \mathscr{P}_2(\mathbb{R}^n))$ and there exists $\boldsymbol{\pi} \in \mathscr{P}(AC^2([0,1], \mathbb{R}^n)$ such that:

(i) $\boldsymbol{\pi}$ is concentrated on solutions to the ODE $\gamma' = v_t(\gamma)$;

(ii) $\|v_t\|_{L^2(\mu_t)} = |\dot{\mu}_t|$ for a.e. $t \in [0,1]$;

(iii) $(e_t)_\# \boldsymbol{\pi} = \mu_t$ for any $t \in [0,1]$.

Let us stress that, since the continuity equation is written in conservative form, no particular regularity assumptions on v_t are required; as a consequence, there may exist multiple solutions to the ODE $\gamma' = v_t(\gamma)$, but this does not affect condition (i). Now let us move to the metric framework and note that there are several difficulties in adapting the differential objects that appear in the previous statement: for instance, the continuity equation and the ODE $\gamma' = v_t(\gamma)$. We will investigate these problems later. For the moment, let us first state Lisini's theorem.

Theorem 5.10.8 (Lisini)**.** *Let (X, d) be a separable metric space. For any $(\mu_t) \in AC^2([0,1], \mathscr{P}_2(X))$ there exists $\boldsymbol{\pi} \in \mathscr{P}(AC^2([0,1], X))$ satisfying*

(i) for a.e. $t \in [0,1]$ it holds

$$\int |\dot{\gamma}_t|^2 d\boldsymbol{\pi}(\gamma) = |\dot{\mu}_t|^2 \tag{5.10.9}$$

(ii) $(e_t)_\# \boldsymbol{\pi} = \mu_t$ for any $t \in [0,1]$.

A plan $\boldsymbol{\pi}$ satisfying condition (i) is called a *lifting* of μ_t.

5.11 Lecture 11 - 24 September

Abstract

Summing up all the ingredients we have previously presented, in the first part of the lecture we prove that $H^{1,p} = BL^{1,p}$ and the two notions of weak gradient agree. Then we provide the third possible definition of metric Sobolev space and it turns out that it is equivalent to the previous two; the same holds for the associated weak gradient, so that in the end a threefold approach to the theory of Sobolev spaces is possible in the metric case too. In the third and last part of the lecture we focus our attention on BV functions and we move to the metric framework some well-known notions of Geometric Measure Theory.

The last topic we presented in the previous lecture was the superposition principle and, more precisely, Lisini's theorem. By the statement of the theorem, the existence of a lifting $\boldsymbol{\pi}$ for any curve $(\mu_t) \in AC^2([0,1], \mathscr{P}_2(X))$ is guaranteed, but nothing is said about uniqueness; indeed, condition (i) and (ii) are not sufficient to infer it, but nevertheless (5.10.9) tells us that the lifting we are considering is optimal, because for any lifting $\boldsymbol{\pi}$ we always have

$$|\dot{\mu}_t|^2 \leq \int |\dot{\gamma}_t|^2 d\boldsymbol{\pi}(\gamma) \tag{5.11.1}$$

Hence (5.10.9) represents the best situation we can have and it is precisely for this reason that we call the plan $\boldsymbol{\pi}$ provided by Lisini's theorem an optimal lifting. This also motivates the proof of Theorem 5.10.8, because we are looking for a lifting $\boldsymbol{\pi}$ for which the inequality opposite to (5.11.1) holds. After this digression, let us quickly prove (5.11.1); to this aim observe that, given $s > t$, $\eta := (e_s, e_t)_\# \boldsymbol{\pi}$ is a transport plan from μ_s to μ_t and therefore

$$W_2^2(\mu_s, \mu_t) \quad \leq \quad \int_{X^2} d^2(x, y) d\eta(x, y) = \int d^2(\gamma_s, \gamma_t) d\boldsymbol{\pi}(\gamma)$$

$$
\begin{aligned}
&\leq \int \left(\int_s^t |\dot\gamma_r| \right)^2 d\boldsymbol{\pi}(\gamma) \\
&\leq (s-t) \int_s^t \int |\dot\gamma_r|^2 d\boldsymbol{\pi}(\gamma) dr
\end{aligned}
$$

Thus, if we divide by $(s-t)^2$ and take the limit as $t \to s$ at the Lebesgue points of the right-hand side, we precisely obtain (5.11.1). As a last remark, let us quickly recall the notion of tightness and its relation with convergence in the sense of measures: a family of measures $\mathcal{K} \subset \mathscr{P}(X)$ is said to be *tight* provided for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset X$ such that

$$\mu(X \setminus K_\varepsilon) \leq \varepsilon, \quad \forall \mu \in \mathcal{K}$$

and in the case of a complete and separable metric space (X, d) a family $\mathcal{K} \subset \mathscr{P}(X)$ is relatively compact w.r.t. convergence in the sense of measures if and only if it is tight. Now we can sketch the proof of Lisini's theorem.

Proof. For sake of simplicity, let us assume that (X, d) is geodesic (this condition is not particularly restrictive, because (X, d) can be isometrically embedded into ℓ^∞ and all the forthcoming arguments are invariant under isometric embeddings). Let us begin by dividing the unit interval $[0, 1]$ into n parts of equal length and let us try to find a sort of approximated plan $\boldsymbol{\pi}^n$ which is a lifting of (μ_t) at $t = j/n$ for $j = 0, ..., n$, i.e.

$$(e_{j/n})_\# \boldsymbol{\pi}^n = \mu_{j/n}, \quad j = 0, ..., n \tag{5.11.2}$$

and such that

$$\int \int_0^1 |\dot\gamma_t|^2 d\boldsymbol{\pi}^n(\gamma) dt \leq \int_0^1 |\dot\mu_t|^2 dt \tag{5.11.3}$$

The construction of $\boldsymbol{\pi}^n$ will be only sketch. Roughly speaking, we have to glue together the optimal plans we get by solving the Monge-Kantorovich problems relative to the couples $(\mu_{(j-1)/n}, \mu_{j/n})$ for $j = 1, ..., n$, so that we obtain a measure $\Sigma \in \mathscr{P}(X^{n+1})$ such that $(\pi^{j-1}, \pi^j)_\# \Sigma$ is an optimal plan from $\mu_{(j-1)/n}$ to $\mu_{j/n}$ for every $j = 1, ..., n$. At this point, if we choose $n+1$ random points $x_0, ..., x_n$, by assumption we know that for any couple (x_{j-1}, x_j) there exists at least one constant speed geodesic connecting them; hence, if we select the geodesics connecting the random points $x_0, ..., x_n$ and we perform this procedure for any $(n+1)$-tuple of random points, we can push-forward Σ to a measure $\boldsymbol{\pi}^n$ on $C([0, 1], X)$. The fact that (5.11.2) holds is clear by definition, as well as (5.11.3), because $\boldsymbol{\pi}^n$ is concentrated on piecewise geodesic curves (piecewise w.r.t. the given partition of $[0, 1]$).

On the contrary, let us completely omit the proof of the fact that $(\boldsymbol{\pi}^n)_{n\in\mathbb{N}}$ is tight and let us immediately observe that by tightness we can infer that $(\boldsymbol{\pi}^n)_{n\in\mathbb{N}}$ admits a convergent subsequence, whose limit point will be denoted by $\boldsymbol{\pi}$. The measure $\boldsymbol{\pi}$ is now a lifting of (μ_t), i.e. $(e_t)_\# \boldsymbol{\pi} = \mu_t$ for any $t \in [0, 1]$, and in addition it holds

$$\int \int_0^1 |\dot\gamma_t|^2 d\boldsymbol{\pi}(\gamma) dt \leq \int_0^1 |\dot\mu_t|^2 dt$$

Since by (5.11.1) the opposite inequality holds for a.e. $t \in [0,1]$, we conclude that (5.10.9) is satisfied for a.e. $t \in [0,1]$. □

Now we are finally ready to prove that $H^{1,p}(X,d,\mathfrak{m}) = BL^{1,p}(X,d,\mathfrak{m})$; let us recall that we have already proved the "$\subset$" inclusion, so that it only remains to prove the opposite one and we will do it in the case $p=2$ under the assumption of finite mass, i.e. $\mathfrak{m}(X) < +\infty$. However, let us stress that the identity between the two Sobolev spaces is true for any p and without growth assumptions on $\mathfrak{m}$.

Let $g \in BL^{1,2}(X,d,\mathfrak{m})$ and let us recall that, without loss of generality, we can assume that g is bounded away from 0 and $+\infty$ and $\|g\|_{L^2(X,\mathfrak{m})} = 1$. Then, let us define $\overline{f} := g^2$, so that we can regard $\overline{f}$ as a probability density, and let (f_t) be the gradient flow of Ch starting from $\overline{f}$. Let us also recall that we have already proved the energy dissipation rate of the entropy (5.9.8), but it still remains to prove (5.9.9) and this is exactly the content of the next key lemma.

Lemma 5.11.1. *In the framework we described above, it holds*

$$\begin{aligned}\int_X (\overline{f}\log\overline{f} - f_t\log f_t)d\mathfrak{m} \quad &\leq \quad \frac{1}{2}\int_0^t\int_X \frac{|\nabla\overline{f}|^2_{BL}}{\overline{f}^2} f_s\, d\mathfrak{m}ds \\ &\quad + \frac{1}{2}\int_0^t\int_X \frac{|\nabla f_s|^2_*}{f_s}\, d\mathfrak{m}ds\end{aligned}$$

Proof. First of all, by convexity of $z \mapsto z\log z$

$$\int_X (\overline{f}\log\overline{f} - f_t\log f_t)d\mathfrak{m} \leq \int_X (\overline{f} - f_t)\log\overline{f}\, d\mathfrak{m}$$

Secondly, by Kuwada's lemma we know that the curve (μ_t) with $\mu_t := f_t\mathfrak{m}$ is absolutely continuous, so that by the superposition principle we can pick an optimal lifting $\boldsymbol{\pi}$ of it. Notice that thanks to the maximum principle all the f_t's are bounded away from 0 and $+\infty$, which means that $\boldsymbol{\pi}$ is also a test plan. Hence, we obtain

$$\begin{aligned}\int_X \log\overline{f}(\overline{f} - f_t)d\mathfrak{m} \quad &= \quad \int(\log\overline{f}(\gamma_0) - \log\overline{f}(\gamma_t))d\boldsymbol{\pi}(\gamma) \\ &\leq \quad \int\int_0^t |\nabla\log\overline{f}|_{BL}(\gamma_s)|\dot{\gamma}_s|ds d\boldsymbol{\pi}(\gamma) \\ &\leq \quad \frac{1}{2}\int_0^t\int |\nabla\log\overline{f}|^2_{BL}(\gamma_s)d\boldsymbol{\pi}(\gamma)ds + \frac{1}{2}\int_0^t\int |\dot{\gamma}_s|^2 d\boldsymbol{\pi}(\gamma)ds\end{aligned}$$

where the identity simply comes from the lifting property of $\boldsymbol{\pi}$ and the first inequality is due to the fact that $\boldsymbol{\pi}$ is a test plan and therefore $\log\overline{f}$ can be estimated by means of $|\nabla\log\overline{f}|_{BL}$ along $\boldsymbol{\pi}$-a.e. curve γ; finally, the last inequality is given by the Cauchy-Schwarz inequality. Now observe that, by the optimality condition (5.10.9) and by Kuwada's lemma, the second term at the right-hand side above can be estimated as

$$\int_0^t\int |\dot{\gamma}_s|^2 d\boldsymbol{\pi}(\gamma)ds \leq \int_0^t\int_X \frac{|\nabla f_s|^2_*}{f_s}\, d\mathfrak{m}ds$$

On the other hand, by chain rule the first term at the right-hand side can be rewritten as

$$\begin{aligned}\int_0^t \int |\nabla \log \overline{f}|_{BL}^2(\gamma_s) d\boldsymbol{\pi}(\gamma) ds &= \int_0^t \int \frac{|\nabla \overline{f}|_{BL}^2}{\overline{f}^2}(\gamma_s) d\boldsymbol{\pi}(\gamma) ds \\ &= \int_0^t \int_X \frac{|\nabla \overline{f}|_{BL}^2}{\overline{f}} f_s \, d\mathfrak{m} ds\end{aligned}$$

The proof is thus accomplished. □

As we have already depicted in the previous lecture, by means of this result we get the desired conclusion:

$$\begin{aligned}&H^{1,p}(X, d, \mathfrak{m}) = BL^{1,p}(X, d, \mathfrak{m}) \\ &|\nabla f|_{BL} = |\nabla f|_*, \quad \mathfrak{m}\text{-a.e. in } X\end{aligned}$$

Hence, not only the two Sobolev spaces coincide, but also the two arising notions of gradient are the same.

Now, let us talk about the last topic of this series of lectures, that is Sobolev spaces via integration by parts. Our aim will be the introduction of a new Sobolev space $W^{1,p}$ and thanks to the available tools it will not be too difficult to prove that $H^{1,p} \subset W^{1,p} \subset BL^{1,p}$, so that *a fortiori* all the three spaces coincide. The difficult part is precisely the definition of $W^{1,p}(X, d, \mathfrak{m})$, because integration by parts relies on vector fields and this notion has to be formalized in a purely metric setting. Our approach is based on a classical aspect coming from differential geometry, that is vector fields seen as derivations. The history of this approach is interesting, because it goes back to a paper of Weaver, published in 2000. Aim of the article was the development of differential calculus in metric measure spaces, but unfortunately the same year Cheeger published its celebrated paper and for this reason Weaver's approach was not considered. Only ten years later some mathematicians started using Weaver's ideas; let us recall some of them:

- Di Marino (from the point of view of Sobolev spaces)
- Bate, Gong, Schioppa (analysis of the structure of derivations and their relation with Cheeger's energy)
- Gigli (the Sobolev space $H^{2,2}$)
- Ambrosio, Trevisan (well-posedness of ODEs)

In all these articles, the definition of derivation is not always the same, there are subtle differences; for sake of information, we will follow Di Marino's approach.

Definition 5.11.2. *Let $L^0(\mathfrak{m})$ be the set of equivalence classes of $\mathfrak{m}$-measurable functions. A linear map $b : \mathrm{Lip}_b(X) \to L^0(\mathfrak{m})$ is called a* derivation *if the following properties are satisfied:*

(i) *Leibniz rule: for any $f, g \in \mathrm{Lip}_b(X)$ it holds*

$$b(fg) = gb(f) + fb(g)$$

(ii) *weak locality: there exists $g \geq 0$ such that, for any $f \in \mathrm{Lip}_b(X)$, $|b(f)| \leq g\,\mathrm{lip}_a(f)$ $\mathfrak{m}$-a.e. in X.*

The minimal function g for which property (ii) is satisfied, if it exists, will be denoted by $|b|$. In addition, sometimes we will use the appealing notation $df(b) := b(f)$.

We will omit the proof, but it is actually possible to show that $|b|$ always exists for any derivation b. Let us also point out that if we replace $\mathrm{Lip}_b(X)$ by $C^\infty(X)$ in the case $X = \mathbb{R}^n$, we recover the standard notion of derivation in the smooth case. Furthermore, on Riemannian manifolds the family of derivations is a C^∞-module; in metric measure spaces it is a L^∞-module and this means that we can multiply a derivation b by an L^∞ function χ getting χb, where the product has to be meant pointwise. As a next step, we can introduce the notion of divergence for derivations by using an integration by parts formula (this had been already anticipated in Lecture 6); more precisely, for any derivation b such that $|b| \in L^1(X, \mathfrak{m})$, the divergence of b is the function $\mathrm{div}\, b$ for which the following holds:

$$\int_X (\mathrm{div}\, b) f \, d\mathfrak{m} = -\int_X b(f)\, d\mathfrak{m}, \quad \forall f \in \mathrm{Lip}_b(X) \tag{5.11.4}$$

With this definition we can introduce the space

$$\mathrm{Der}^q(X, \mathfrak{m}) := \{b \,:\, |b| \in L^q(X, \mathfrak{m}),\, \mathrm{div}\, b \in L^q(X, \mathfrak{m})\}$$

which turns out to be a Banach space when equipped with the norm $\|b\|_{L^q(X,\mathfrak{m})} + \|\mathrm{div}\, b\|_{L^q(X,\mathfrak{m})}$; in addition, $\mathrm{Der}^q(X, \mathfrak{m})$ is a Lip_b-module but it is not an L^∞-module. Now, the introduction of a Sobolev space $W^{1,p}$ for $1 < p < +\infty$ is immediate: by $W^{1,p}(X, d, \mathfrak{m})$ we will denote the collection of functions $f \in L^p(X, \mathfrak{m})$ for which there exists a Lip_b-linear continuous map $L_f : \mathrm{Der}^q(X, \mathfrak{m}) \to L^1(X, \mathfrak{m})$ such that

$$\int_X f \,\mathrm{div}\, b \, d\mathfrak{m} = -\int_X L_f(b)\, d\mathfrak{m}, \quad \forall b \in \mathrm{Der}^q(X, \mathfrak{m}) \tag{5.11.5}$$

Let us stress that this definition is in the same spirit of the one we gave in Example 5.7.3 in the case of a weighted Riemannian manifold, which was based on the integration by parts formula (5.7.6). In addition, (5.11.4) and (5.11.5) are strongly related by duality, since on the one hand the definition of divergence is tested against Lipschitz functions and on the other hand the definition of Sobolev function is tested against derivations admitting divergence. Let us also point out some interesting remarks:

(i) by Lip_b-linearity of L_f, it is not difficult to show that condition (5.11.5) actually characterizes L_f and therefore $L_f(b) = b(f)$ for any $b \in \mathrm{Der}^q(X, \mathfrak{m})$;

(ii) by definition of divergence, $\mathrm{Lip}_b(X) \subset W^{1,p}(X,d,\mathfrak{m})$ and this property will be fundamental in the proof of the fact that $H^{1,p}(X,d,\mathfrak{m}) \subset W^{1,p}(X,d,\mathfrak{m})$, since $H^{1,p}(X,d,\mathfrak{m})$ is obtained by closure of $\mathrm{Lip}_b(X)$;

(iii) given $f \in W^{1,p}(X,d,\mathfrak{m})$ there exists $g \in L^p(X,\mathfrak{m})$ such that, for every $b \in \mathrm{Der}^q(X,\mathfrak{m})$,

$$|b(f)| \leq g|b|, \quad \mathfrak{m}\text{-a.e. in } X$$

The minimal g satisfying this condition will be denoted by $|\nabla f|_w$.

Notice that, again, the definition of $|\nabla f|_w$ is linked to $|b|$ by duality. As $\mathrm{Der}^q(X,\mathfrak{m})$ may be the emptyset, $W^{1,p}(X,d,\mathfrak{m})$ may be trivially equivalent to $L^p(X,\mathfrak{m})$, but this is not surprising, since the same happens to $BL^{1,p}(X,d,\mathfrak{m})$ when no test plans exist. With all these definitions, we can now show that $H^{1,p} \subset W^{1,p} \subset BL^{1,p}$; let us begin with the first inclusion, i.e.

$$H^{1,p}(X,d,\mathfrak{m}) \subset W^{1,p}(X,d,\mathfrak{m})$$

Proof. Let $f \in H^{1,p}(X,d,\mathfrak{m})$. By definition, this means that there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}_0(X)$ such that $f_n \to f$ in $L^p(X,\mathfrak{m})$ and $\mathrm{lip}_a(f_n) \to |\nabla f|_*$ in $L^p(X,\mathfrak{m})$. Secondly, let $b \in \mathrm{Der}^q(X,\mathfrak{m})$ and notice that by definition we have

$$\int_X f_n \,\mathrm{div}\, b \, d\mathfrak{m} = -\int_X b(f_n)\, d\mathfrak{m}$$

Since $\mathrm{div}\, b \in L^q(X,\mathfrak{m})$, no problems arise when we pass to the limit as $n \to \infty$ at the left-hand side and so we have

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

On the contrary, the right-hand side is harder to handle in the general case (the discussion would be much easier under some reflexivity assumption, but our aim is a general treatise), because in the standard estimate

$$|b(f_n) - b(f_m)| \leq |b|\, \mathrm{lip}_a(f_n - f_m)$$

we can not say that $\mathrm{lip}_a(f_n - f_m) \to 0$ and so we can not directly pass to the limit as we have just done for the left-hand side. The only thing we are allowed to claim is that $|b(f_n)| \leq |b|\, \mathrm{lip}_a(f_n)$ and now, passing to the limit as $n \to \infty$ and taking into account the fact that $\mathrm{lip}_a(f_n) \to |\nabla f|_*$, we get

$$\left| \int_X f \,\mathrm{div}\, b\, d\mathfrak{m} \right| \leq \int_X |b||\nabla f|_*\, d\mathfrak{m}, \quad \forall b \in \mathrm{Der}^q(X,\mathfrak{m}) \tag{5.11.6}$$

This inequality allows us to invoke a version of Hahn-Banach theorem for Lip_b-modulus proved by Di Marino, so that the following representation formula holds:

$$\int_X f \,\mathrm{div}\, b\, d\mathfrak{m} = -\int_X L_f(b)\, d\mathfrak{m} \tag{5.11.7}$$

This shows that $f \in W^{1,p}(X,d,\mathfrak{m})$; putting together (5.11.6) and (5.11.7) we also get that $|\nabla f|_w \leq |\nabla f|_*$. □

For the second inclusion, we first need a crucial lemma.

Lemma 5.11.3. *Any p-test plan* $\boldsymbol{\pi}$ *with* $\mathcal{A}_q \in L^1(\boldsymbol{\pi})$ *(as already said, this assumption is not restrictive) canonically induces a derivation* $b_{\boldsymbol{\pi}}$ *with* $|b\boldsymbol{\pi}| \in L^q(X, \mathfrak{m})$ *satisfying:*

(i) $|b_{\boldsymbol{\pi}}| \leq \rho_{\boldsymbol{\pi}}$ $\mathfrak{m}$*-a.e., where* $\rho_{\boldsymbol{\pi}}$ *is the Radon-Nikodym derivative of* $\mathrm{Bar}(\boldsymbol{\pi})$ *(see (5.9.2 for the definition);*

(ii) $\operatorname{div} b_{\boldsymbol{\pi}} \in L^\infty(X, \mathfrak{m})$ *and we can also provide an explicit formula for* $\operatorname{div} b_{\boldsymbol{\pi}}$*, namely*

$$\operatorname{div} b_{\boldsymbol{\pi}} = \frac{(e_0)_\# \boldsymbol{\pi} - (e_1)_\# \boldsymbol{\pi}}{\mathfrak{m}} \tag{5.11.8}$$

Note that by definition (5.9.2) of baricenter, it follows immediately that

$$\int \varphi \, d\,\mathrm{Bar}(\boldsymbol{\pi}) = \int \int_\gamma \varphi \, d\boldsymbol{\pi}(\gamma)$$

and actually we can also regard this formula as definition of baricenter. In the case of a test plan, as it is our present framework, under the further assumption $\mathcal{A}_q \in L^1(\boldsymbol{\pi})$ one can show that $\mathrm{Bar}(\boldsymbol{\pi})$ is absolutely continuous w.r.t. $\mathfrak{m}$ and if we denote by $\rho_{\boldsymbol{\pi}}$ the Radon-Nikodym derivative, then $\rho_{\boldsymbol{\pi}} \in L^q(X, \mathfrak{m})$. Now let us pass to the proof of the previous lemma.

Proof. Given $f \in \mathrm{Lip}_b(X)$, let us first define $b_{\boldsymbol{\pi}}$ as a derivation in the following way:

$$\langle b_{\boldsymbol{\pi}}(f), \varphi \rangle := \int \int_0^1 \frac{d}{dt}(f \circ \gamma)\varphi \circ \gamma \, dt d\boldsymbol{\pi}(\gamma) \tag{5.11.9}$$

From a heuristic point of view, the superposition principle is again fundamental, because we associate to any curve γ the tangent vector field, given by the time derivative of $f \circ \gamma$ regarded as a derivation, and then we average over $\boldsymbol{\pi}$. With this definition it is easy to check linearity and Leibniz rule, but for the moment we do not know yet whether $b_{\boldsymbol{\pi}}$ is a function (it is only a distribution); furthermore, the other properties are more technical and delicate to handle. As a first step, keeping f fixed observe that, by virtue of Radon-Nikodym theorem, we infer

$$\begin{aligned} |\langle b_{\boldsymbol{\pi}}(f), \varphi \rangle| &\leq \int \int_0^1 |\nabla f| \circ \gamma \varphi \circ \gamma |\dot{\gamma}| \, dt d\boldsymbol{\pi}(\gamma) \\ &= \int_X |\nabla f| \varphi \rho_{\boldsymbol{\pi}} \, d\mathfrak{m} \end{aligned}$$

and this *a priori* estimate shows that $b_{\boldsymbol{\pi}}$ is actually a function, so that we are allowed to write

$$\langle b_{\boldsymbol{\pi}}(f), \varphi \rangle = \int_X b_{\boldsymbol{\pi}} \varphi \, d\mathfrak{m}$$

Hence we have

$$\left| \int_X b_{\boldsymbol{\pi}}(f) \varphi \, d\mathfrak{m} \right| \leq \int_X |\nabla f| \varphi \rho_{\boldsymbol{\pi}} \, d\mathfrak{m}$$

and by localization in φ and in f, (i) follows. Now let us check the formula (5.11.8) for the divergence and note that once this is proved, we automatically get that $\operatorname{div} b_{\boldsymbol{\pi}} \in L^\infty(X, \mathfrak{m})$, since the right-hand side in (5.11.8) is a bounded function (indeed, $\boldsymbol{\pi}$ is a test plan). To this aim, note that

$$\begin{aligned}\int_X f \operatorname{div} b_{\boldsymbol{\pi}}\, d\mathfrak{m} &= -\int_X b_{\boldsymbol{\pi}}(f)\, d\mathfrak{m} = -\int\int_0^1 \frac{d}{dt}(f\circ\gamma) dt d\boldsymbol{\pi}(\gamma) \\ &= \int (f(\gamma_0) - f(\gamma_1)) d\boldsymbol{\pi}(\gamma) = \int_X f\, d((e_0)_\# \boldsymbol{\pi} - (e_1)_\# \boldsymbol{\pi})\end{aligned}$$

where the second identity follows by taking $\varphi \equiv 1$ in (5.11.9); by the arbitrariness of f we get the conclusion. □

Now let us explain where this lemma comes into play in the proof of the inclusion

$$W^{1,p}(X, d, \mathfrak{m}) \subset BL^{1,p}(X, d, \mathfrak{m})$$

Proof. Let $f \in W^{1,p}(X, d, \mathfrak{m})$ and let us prove that

$$|f(\gamma_1) - f(\gamma_0)| \le \int_\gamma |\nabla f|_w, \quad \text{for } p\text{-a.e. } \gamma$$

because this fact implies that $f \in BL^{1,p}(X, d, \mathfrak{m})$ and $|\nabla f|_{BL} \le |\nabla f|_w$. Fix a test plan $\boldsymbol{\pi}$ and let $b_{\boldsymbol{\pi}}$ be the derivation canonically induced by $\boldsymbol{\pi}$, as provided by the previous lemma; if we integrate $f(\gamma_1) - f(\gamma_0)$ w.r.t. $\boldsymbol{\pi}$, we get

$$\begin{aligned}\int (f(\gamma_1) - f(\gamma_0)) d\boldsymbol{\pi}(\gamma) &= -\int_X f \operatorname{div} b_{\boldsymbol{\pi}}\, d\mathfrak{m} = \int_X b_{\boldsymbol{\pi}}(f)\, d\mathfrak{m} \\ &\le \int_X \rho_{\boldsymbol{\pi}} |\nabla f|_w\, d\mathfrak{m} = \int\int_\gamma |\nabla f|_w\, d\boldsymbol{\pi}(\gamma)\end{aligned}$$

and then by localization we conclude. Indeed, if we choose any Borel set $A \subset C([0,1], X)$ with $\boldsymbol{\pi}(A) > 0$, we define

$$\boldsymbol{\pi}_A := \frac{1}{\boldsymbol{\pi}(A)} \boldsymbol{\pi} \llcorner A$$

and we replace $\boldsymbol{\pi}$ by $\boldsymbol{\pi}_A$ in the argument above, then we have

$$\int_A (f(\gamma_1) - f(\gamma_0)) d\boldsymbol{\pi}(\gamma) \le \int_A \int_\gamma |\nabla f|_w d\boldsymbol{\pi}(\gamma)$$

whence, by the arbitrariness of A,

$$f(\gamma_1) - f(\gamma_0) \le \int_\gamma |\nabla f|_w, \quad \text{for } \boldsymbol{\pi}\text{-a.e. } \gamma$$

The inequality for $f(\gamma_0) - f(\gamma_1)$ follows immediately by replacing f with $-f$ and, by the arbitrariness of the test plan $\boldsymbol{\pi}$, we get the conclusion. □

Hence the three definitions of Sobolev space we gave in the metric setting coincide and so do the relative notions of weak gradient. Now we can approach the last topic of the course, namely the theory of BV functions and sets of finite perimeter, so that a connection with the first part of the course, held by prof. Serra Cassano, will be established. As for the theory of Sobolev spaces, a threefold approach is possible even for the theory of BV functions; this means that we have three *a priori* different BV spaces: a space BV_* in the sense of approximations, a space BV_{BL} in the sense of Beppo Levi and a space BV_w in the sense of integration by parts; however, as proved in Di Marino's paper, the three definitions are equivalent. We will skip the definition of BV_w, but roughly speaking we can imagine what happens: the map L_f maps $\mathrm{Der}^q(X, \mathfrak{m})$ into $\mathscr{M}_+(X)$ instead of $L^1(X, \mathfrak{m})$, where $\mathscr{M}_+(X)$ denotes the family of non-negative finite measures on X.

Now let us turn our attention to the BV_* point of view and let us start by providing the definition of $\mathrm{Lip}_{loc}(X, d)$, which may be a little bit ambiguous, since we are not assuming any local compactness on (X, d). In more details, we set

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

and in the case (X, d) is locally compact, this definition amounts to say that $\mathrm{Lip}_{loc}(X, d)$ coincides with the functions which are Lipschitz continuous on compact sets. With this premise, we can define $BV_*(X, d, \mathfrak{m})$, because $\mathrm{Lip}_{loc}(X, d)$ is the class of functions involved in the approximation technique. More precisely, we define $BV_*(X, d, \mathfrak{m})$ as the set of functions $f \in L^1(X, \mathfrak{m})$ for which there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}_{loc}(X, d)$ such that $f_n \to f$ in $L^1(X, \mathfrak{m})$ and

$$\limsup_{n\to\infty} \int_X \mathrm{lip}_a(f_n) \, d\mathfrak{m} < +\infty$$

Notice that this definition has exactly the same form of the one of $H^{1,p}(X, d, \mathfrak{m})$ we provided in Lecture 7, but that definition made sense only for $1 < p < +\infty$. Thus, $BV_*(X, d, \mathfrak{m})$ can be seen as a limiting case of $H^{1,p}(X, d, \mathfrak{m})$. In connection with this notion of BV space we can also introduce a sort of Cheeger energy, which will be denoted by $|Df|_*(X)$, because in the classical setting this object coincides with the total variation of the distributional derivative of f; the definition is the following

$$|Df|_*(X) := \inf \left\{ \liminf_{n\to\infty} \int_X \mathrm{lip}_a(f_n, x) d\mathfrak{m}(x) \right\} \tag{5.11.10}$$

where the infimum is taken on all the sequences $(f_n)_{n\in\mathbb{N}} \subset \mathrm{Lip}_{loc}(X, d)$ such that $f_n \to f$ in $L^1(X, \mathfrak{m})$, and it corresponds to the lower semicontinuous relaxation of the energy functional $\int_X \mathrm{lip}_a(f_n) d\mathfrak{m}$. It is worth noticing that, in a metric setting, as it is not always possible to speak about gradients ∇f but it is still possible to introduce its modulus $|\nabla f|_*$, in the same way we are not allowed to handle distributional derivatives Df but nevertheless we can talk about total variations $|Df|_*$. All the metric definitions we provided in the previous and in the present lecture are perfectly consistent with the classical ones in a smooth framework. For instance, in order to perceive the proximity of the metric theory to the smooth one, let us consider the following fact: Gong and Schioppa proved that if the space $(X, d, \mathfrak{m})$ is doubling, then $\mathrm{Der}^q(X, \mathfrak{m})$ is finite-dimensional as L^∞-module (however, for sake of completeness, Gong and Schioppa consider a more restrictive class of derivations, namely the continuous ones).

However, in order to perform computations we need to single out a minimal object; in the case of Sobolev spaces, this was the minimal relaxed slope, whereas in this case, as we are going to see, the object we are looking for is a measure and this was first stressed by Miranda in 2004; thanks to his contribution, it has been possible to develop the Sobolev theory of BV functions. For the construction of such a measure, we simply have to "localize" (5.11.10); more precisely, for any open set $A \subset X$, we define

$$|Df|_*(A) := \inf \left\{ \liminf_{n\to\infty} \int_A \mathrm{lip}_a(f_n, x) d\mathfrak{m}(x) \right\}$$

where the infimum is taken on the same family of approximating sequences. In this way, we have a finite set function $A \mapsto |Df|_*(A)$ defined on the open subsets of X and the fact that it is finite follows by $|Df|_*(X) < +\infty$.

Theorem 5.11.4. *The set function $A \mapsto |Df|_*(A)$ is the restriction to the family of open sets of a (unique) finite Borel measure, for which we retain the same notation.*

Thanks to this fact and to the availability of the measure $|Df|_*$ we can introduce several further definitions; in particular, we will say that a set $E \subset X$ has *finite perimeter* if $\chi_E \in BV_*(X, d, \mathfrak{m})$ and, given another set $B \subset X$, we will define the *perimeter of E in B* as

$$P(E, B) := |D\chi_E|_*(B) \tag{5.11.11}$$

Observe that these definitions coincide with the ones we gave in the first part of the course, but now the situation is fairly more general. In the next (and last) lecture we will deepen the study of the measure $|D\chi_E|$ and its link with the geometry of the space; on the contrary, we conclude the present lecture with the presentation of the Beppo Levi theory of BV functions, carried out by Ambrosio and Di Marino in 2013. Roughly speaking, we will ask a given function to be BV along almost every curve and we will use this information to single out a different definition of total variation. More precisely, $BV_{BL}(X, d, \mathfrak{m})$ is the space of functions $f \in L^1(X, \mathfrak{m})$ with the following properties:

(i) $f \circ \gamma \in BV(0, 1)$ for 1-a.e. curve γ;

(ii) there exists a measure μ on X such that, for any test plan $\boldsymbol{\pi}$,

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

where $C(\boldsymbol{\pi})$ is the compression constant of $\boldsymbol{\pi}$.

Let us stress few comments. About property (i), since p-test plans are concentrated on absolutely continuous curves with $L^{p'}$ metric derivative, 1-test plans are concentrated on Lipschitz curves. About property (ii), since $f \circ \gamma \in BV(0,1)$ 1-almost surely, $|D(f \circ \gamma)|$ exists in the classical sense for 1-a.e. curve γ; in addition, we can look for the minimal measure μ satisfying the inequality above: such a measure will be called the *total variation of f in the sense of Beppo Levi* and will be denoted by $|Df|_{BL}$. As in the case of Sobolev spaces (and indeed, by miming the same approach), it is easy to show that

$$\begin{aligned} &BV_*(X, d, \mathfrak{m}) \leq BV_{BL}(X, d, \mathfrak{m}) \\ &|Df|_{BL} \leq |Df|_* \end{aligned}$$

On the contrary, it is much more difficult to prove the converse inclusion and inequality and this was achieved by Ambrosio and Di Marino in 2013. Hence, a posteriori the two BV spaces are equal and the two notions of total variation are equivalent. The main idea behind the work of Ambrosio and Di Marino relies on considering the limiting case of the tools we exploited for Sobolev spaces. For instance, as we have just pointed out, 1-test plans are concentrated on Lipschitz curves (since they have bounded metric derivative); Hopf-Lax formula has been described only for $p = 2$, but for general p it reads as

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

and if we take the limit as $p \downarrow 1$, then we get

$$Q_t f(x) := \inf_{y \in B(x,t)} f(y)$$

and we still have that $Q_t f$ is a subsolution to the Hamilton-Jacobi equation.

5.12 Lecture 12 - 25 September

Abstract

The present lecture is strongly related to the last lecture of the first part of the course, because we will investigate the structure of the perimeter measure, which corresponds to a weak notion of surface area. In more details, we will see that such measure is concentrated on the so-called essential boundary and it is absolutely continuous w.r.t. a "1-codimensional measure", so that the essential boundary can be seen as a sort of 1-codimensional subspace.

Let us recall that in the previous lecture we have introduced two weak notions of total variation and, after all, they turned out to be equivalent; hence, from now on the metric total variation of a given function f (more precisely of its distributional derivative) will be simply denoted by $|Df|$ and note that it is a finite measure on X. Let us also recall the definition of the perimeter of E in B, given at (5.11.11), and let us point out some properties which will be particularly useful for the forthcoming discussion:

(i) $P(E, B)$ is local on open sets, i.e. if A is an open set, then for any $E, F \subset X$ it holds $P(E, A) = P(F, A)$ provided that $(E \Delta F) \cap A = \emptyset$, where Δ denotes the set-theoretic symmetric difference; from an intuitive point of view, this means that $P(\cdot, A)$ is not affected by modifications outside A;

(ii) $P(E, B)$ is additive in B since it is a measure (thus, it is also σ-additive);

(iii) $P(E, B)$ is stable under complement, that is $P(E^c, B) = P(E, B)$;

(iv) by approximation it is possible to check that

$$P(E \cap F, B) + P(E \cup F, B) \leq P(E, B) + P(F, B)$$

whence, in particular, we get that $P(E, B)$ is subadditive in E, i.e.

$$P(E \cup F, B) \leq P(E, B) + P(F, B)$$

(v) given a set $B \subset X$, the perimeter measure in B is lower semicontinuous in E w.r.t. the $L^1_{loc}(B)$ topology. For our purposes, this means that if $\mathfrak{m}(E_n) \to 0$ as $n \to \infty$, then

$$P(X, B) \leq \liminf_{n \to \infty} P(X \setminus E_n, B)$$

Now, in analogy with the classical setting, we expect $|D\chi_E|$ to be concentrated on a sort of "lower" dimensional set and also to be absolutely continuous w.r.t. a "1-codimensional" measure, but what are the natural guess for such a set and measure? For the first one, the answer is given by the *essential boundary*, defined by

$$\partial^* E := \left\{ x \in E \,:\, \limsup_{\rho \downarrow 0} \min \left\{ \frac{\mathfrak{m}_E(x, \rho)}{\mathfrak{m}(B(x, \rho))}, \frac{\mathfrak{m}_{E^c}(x, \rho)}{\mathfrak{m}(B(x, \rho))} \right\} > 0 \right\}$$

where $\mathfrak{m}_E(x, \rho)$ is the measure fraction and is given by

$$\mathfrak{m}_E(x, \rho) := \mathfrak{m}(E \cap B(x, \rho)) \tag{5.12.1}$$

An equivalent definition of the essential boundary is the following: it is the complement of the set of points with density either 0 or 1; however, the definition we provided fits better to our purposes and it is the one we will exploit for computations. Notice that if $\mathfrak{m}$ is doubling, then the essential boundary is "small", that is $\mathfrak{m}(\partial^* E) = 0$, but out goal is much stronger: we would like to prove that $\partial^* E$ is lower dimensional and not only negligible. To this aim we need a slightly more precise version of the essential boundary, namely

$$(\partial^* E)_\gamma := \left\{ x \in E \,:\, \limsup_{\rho \downarrow 0} \min \left\{ \frac{\mathfrak{m}_E(x, \rho)}{\mathfrak{m}(B(x, \rho))}, \frac{\mathfrak{m}_{E^c}(x, \rho)}{\mathfrak{m}(B(x, \rho))} \right\} \geq \gamma \right\}$$

where γ is a positive constant. It is then immediate to check that $(\partial^* E)_\gamma \subset \partial^* E$. Thus, we have described the natural candidate to be the lower dimensional set supporting $\partial^* E$. For the 1-codimensional measure, let us define

$$h(B(x,\rho)) := \frac{\mathfrak{m}(B(x,\rho))}{\rho}$$

and observe, for sake of information, that if $(X,d,\mathfrak{m})$ is Ahlfors s-regular (cf. Definition 5.2.10), then $h(B(x,\rho))$ behaves like ρ^{s-1} as $\rho \downarrow 0$. However, in what follows we will not work under Ahlfors regularity hypothesis; we will assume instead $\mathfrak{m}$ to be doubling. With this definition, we can now consider the Hausdorff measure $\mathscr{H}^h$, obtained via Carathéodory's construction from h; it is a σ-additive (but not σ-finite) Borel measure with values in $[0,+\infty]$ and it is the natural candidate we were looking for. Thus, we have singled out the natural guesses for the set and the measure; now we are going to introduce gradually some tools, starting with the ones that hold even without doubling assumption (they can be found in the paper of Miranda):

- the coarea formula, which relates BV functions to sets of finite perimeter; more precisely if $u \in BV(X,d,\mathfrak{m})$ is a non-negative function, then

$$|Du|(B) = \int_0^\infty P(\{u > t\}, B)dt$$

- the derivative of the volume-surface area, namely for every $x \in X$ and for a.e. $\rho > 0$ it holds

$$P(E \setminus B(x,\rho), \partial B(x,\rho)) \le \frac{d}{dr}\mathfrak{m}_E(x,r)\Big|_{r=\rho}$$

  In order to make this relation pointwise we have to consider the upper right derivative; this amounts to say that for every $x \in X$ and for every $\rho > 0$ we have

$$P(E \setminus B(x,\rho), \partial B(x,\rho)) \le \frac{d^+}{dr}\mathfrak{m}_E(x,r)\Big|_{r=\rho} \tag{5.12.2}$$

These formulas are completely general, but let us now investigate some consequences of the doubling assumption, which will be used from now on; for sake of information, the doubling constant will be sometimes denoted by C_D and let us recall that w.r.t. this constant the following fact holds: for every $x \in X$ and for every $r > 0$

$$\mathfrak{m}(B(x,2r)) \le C_D \mathfrak{m}(B(x,r))$$

First of all, notice that the function h is doubling as well and secondly we have an upper bound on the dimension of $(X,d,\mathfrak{m})$, because given $0 < r \le R$, $y \in X$ and $x \in B(y,R)$ we have

$$\frac{\mathfrak{m}(B(x,r))}{\mathfrak{m}(B(y,R))} \ge C(r/R)^s$$

with C, s depending only on C_D; more precisely we can say that $s = \log_2 C_D$. However, the estimate above is rough w.r.t. the ones usually asked in the CD conditions. A further consequence of the doubling assumption is Vitali's covering theorem, whose statement is the following.

Theorem 5.12.1. *Let ν be a finite measure on X, $K \subset X$ be a compact set and $\mathscr{F}$ be a fine cover of K by balls (this means that for any $x \in K$ and for any $\varepsilon > 0$ there exists a ball in $\mathscr{F}$ centered at x with radius less than ε) such that*

$$\nu(B) \geq h(B), \quad \forall B \in \mathscr{F}$$

Then there exists a disjoint subfamily $\mathscr{F}'$ such that

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

Hence the subfamily $\mathscr{F}'$ does not cover everything but $\mathscr{H}^h$-almost everything. In addition, we do not need ν to be doubling (in the applications, ν will be the perimeter measure): we only need h to be so. After these considerations, let us immediately see the implications of Vitali's covering theorem, even if they will be exploited in the last part of the lecture:

- if we have, for a given Borel set B,

$$\limsup_{\rho \downarrow 0} \frac{\nu(B(x,\rho))}{h(B(x,\rho))} \geq t, \quad \forall x \in B$$

  then $\nu \geq t\mathscr{H}^h$ on the Borel subsets of B;

- if ν is finite, then

$$\limsup_{\rho \downarrow 0} \frac{\nu(B(x,\rho))}{h(B(x,\rho))} < +\infty, \quad \text{for } \mathscr{H}^h\text{-a.e. } x$$

  and this property follows from the previous one if we let $t \uparrow +\infty$.

Now, using only the doubling property, we can show a first interesting result, namely a comparison between the perimeter measure $P(E, \cdot)$ and $\mathscr{H}^h$.

Theorem 5.12.2. *$P(E, \cdot) \ll \mathscr{H}^h$.*

Pay attention to the fact that $\mathscr{H}^h$ is not σ-finite, so we are not allowed to invoke Radon-Nikodym theorem and state that $P(E, \cdot)$ admits a density w.r.t. $\mathscr{H}^h$; the only thing we can say is that $\mathscr{H}^h$-negligible sets are also $P(E, \cdot)$-negligible.

Proof. Let K be a $\mathscr{H}^h$-negligible set and let us assume, for sake of simplicity, K to be compact; we have to show that $P(E, K) = 0$. To this aim, observe first that, by definition of $\mathscr{H}^h$-Hausdorff measure, for any $\varepsilon > 0$ we can find a finite family of balls $B(x_i, r_i)$ for $i = 1, ..., n$ such that

$$K \subset \bigcup_{i=1}^{n} B(x_i, r_i), \quad \sum_{i=1}^{n} h(B(x_i, r_i)) < \varepsilon \tag{5.12.3}$$

Secondly, for any $i = 1, ..., n$ we can find $r_i' \in]r_i, 2r_i[$ with

$$P(B(x_i, r_i'), X) \leq \frac{C_D}{2} h(B(x_i, r_i)) \tag{5.12.4}$$

Indeed, by the coarea formula applied with $u(x) = d(x, x_i)$ and $B = B(x_i, 2r_i)$ and divided by $2r_i$ we first get

$$\frac{1}{2r_i} \int_0^\infty P(\{d(\cdot, x_i) > t\}, B(x_i, 2r_i))dt = \frac{1}{2r_i} |Dd(\cdot, x_i)|(B(x_i, 2r_i))$$

The right-hand side can be easily estimated since $d(\cdot, x_i)$ is 1-Lipschitz and therefore $|Dd(\cdot, x_i)| \leq \mathfrak{m}$, whence we obtain (taking also into account the fact that $\mathfrak{m}$ is doubling)

$$\begin{aligned} \frac{1}{2r_i} |Dd(\cdot, x_i)|(B(x_i, 2r_i)) &\leq \frac{\mathfrak{m}(B(x_i, 2r_i))}{2r_i} \leq \frac{C_D}{2} \frac{\mathfrak{m}(B(x_i, r_i))}{r_i} \\ &= \frac{C_D}{2} h(B(x_i, r_i)) \end{aligned}$$

On the other hand, the left-hand side can be handled in the following way:

$$\begin{aligned} &\frac{1}{2r_i} \int_0^\infty P(\{d(\cdot, x_i) > t\}, B(x_i, 2r_i))dt \\ &= \frac{1}{2r_i} \int_0^{2r_i} P(\{d(\cdot, x_i) > t\}, B(x_i, 2r_i))dt \\ &= P(\{d(\cdot, x_i) > r_i'\}, B(x_i, 2r_i)) = P(X \setminus \overline{B(x_i, r_i')}, B(x_i, 2r_i)) \\ &= P(\overline{B(x_i, r_i')}, B(x_i, 2r_i)) = P(\overline{B(x_i, r_i')}, X) \\ &= P(B(x_i, r_i'), X) \end{aligned}$$

The first identity is trivial, the second one is due to the mean value theorem, whereas what follows is a consequence of the properties of the perimeter measure, e.g. the stability under complement. Now, let us define

$$A_\varepsilon := \bigcup_{i=1}^{n} B(x_i, r_i') \supset K$$

and observe that

$$\begin{aligned} P(E \setminus A_\varepsilon, X) &= P(E \setminus A_\varepsilon, X \setminus K) \\ &\leq P(E, X \setminus K) + P(A_\varepsilon, X \setminus K) \end{aligned}$$

$$
\begin{aligned}
&= P(E, X \setminus K) + P(A_\varepsilon, X)\\
&\leq P(E, X \setminus K) + \sum_{i=1}^{n} P(B(x_i, r_i'), X)\\
&\leq P(E, X \setminus K) + \frac{C_D}{2} \sum_{i=1}^{n} h(B(x_i, r_i))\\
&\leq P(E, X \setminus K) + \frac{C_D \varepsilon}{2}
\end{aligned}
$$

The two identities are simply motivated by the fact that K is far away from the boundary of $E \setminus A_\varepsilon$ and A_ε. For the inequalities we used the subadditivity of the perimeter measure, (5.12.4) and the second part of (5.12.3). Thus, if we pass to the limit as $\varepsilon \downarrow 0$, by the lower semicontinuity of the perimeter measure (this follows from the lower semicontinuity of the total variation) we infer that

$$
\begin{aligned}
P(E, X) &\leq \liminf_{\varepsilon \downarrow 0} P(E \setminus A_\varepsilon, X) \leq \liminf_{\varepsilon \downarrow 0} \left(P(E, X \setminus K) + \frac{C_D \varepsilon}{2} \right)\\
&= P(E, X \setminus K)
\end{aligned}
$$

but on the other hand $P(E, X) = P(E, K) + P(E, X \setminus K)$, so that $P(E, K) = 0$, which is the desired conclusion. □

By a similar argument, it is possible to prove that if $\mathscr{H}^h(\partial A) < +\infty$ for any open set A, then $P(A, X) \leq c\mathscr{H}^h(\partial A) < +\infty$, where the constant c only depends on C_D, and thus the perimeter of A is finite. Actually, we will use several times this kind of argument.

After having seen the consequences of the doubling assumption and in order to go beyond the already cited results, let us present the last hypothesis we need: we ask $(X, d, \mathfrak{m})$ to support a $(1,1)$-Poincaré inequality. This assumption can be stated in many equivalent ways, we prefer the following one: there exist constants $C_I > 0$ and $\lambda \geq 1$ such that, for any $x \in X$, for any $r > 0$, for any $u \in \mathrm{Lip}_{loc}(X)$ and for any $g \in UG(u)$, it holds

$$
\fint_{B(x,\rho)} |u - \overline{u}| \, d\mathfrak{m} \leq C_I \rho \fint_{B(x,\lambda\rho)} g \, d\mathfrak{m} \tag{5.12.5}
$$

where

$$
\overline{u} := \fint_{B(x,\rho)} u \, d\mathfrak{m}
$$

In other cases (i.e. in other equivalent formulations), at the right-hand side the upper gradient g can be replaced by the pointwise Lipschitz constant or by the minimal weak upper gradient, but *a posteriori* the inequality has the same strength, thanks to the identification result between Sobolev spaces we proved. A different form of the $(1,1)$-Poincaré inequality involves BV functions instead of locally Lipschitz ones and it can be easily deduced from (5.12.5) by passing to the limit at both sides:

$$
\fint_{B(x,\rho)} |u - \overline{u}| \, d\mathfrak{m} \leq C_I \rho \frac{|Du|(B(x, \lambda\rho))}{\mathfrak{m}(B(x,\rho))}
$$

Note that at the right-hand side we exactly get the total variation, because by definition $|Du|$ is obtained via this approximation procedure. *A priori*, doubling and Poincaré inequality are independent assumptions, but from now on we will assume them simultaneously. In addition, a constant C will be called *structural* if it only depends on C_D, C_I and λ. However, Poincaré inequality as written in (5.12.5) is not enough yet; we need a stronger sort of non-linear isoperimetric inequality and, fortunately, this comes out from the combination of the doubling and the $(1,1)$-Poincaré inequality assumptions, as proved by Hajlasz and Koskela. In their paper it is shown that the coupling of the two hypotheses entails the $(1^*,1)$-Poincaré inequality, where $1^* := s/(s-1)$ is a bound on the Sobolev exponent; this means that (5.12.5) can be improved to

$$\left(\fint_{B(x,\rho)} |u-\overline{u}|^{1^*}\, d\mathfrak{m} \right)^{1/1^*} \leq C_I \rho \fint_{B(x,\lambda\rho)} g\, d\mathfrak{m}$$

and again, by a limiting procedure, this inequality entails the following one for BV functions:

$$\left(\fint_{B(x,\rho)} |u-\overline{u}|^{1^*}\, d\mathfrak{m} \right)^{1/1^*} \leq C_I \rho \frac{|Du|(B(x,\lambda\rho))}{\mathfrak{m}(B(x,\rho))} \tag{5.12.6}$$

In the case of characteristic functions, i.e. $u = \chi_E$, we infer that

$$\left(\fint_{B(x,\rho)} |\chi_E-\overline{\chi_E}|^{1^*}\, d\mathfrak{m} \right)^{1/1^*} \leq C \frac{P(E,B(x,\lambda\rho))}{h(B(x,\lambda\rho))} \tag{5.12.7}$$

where C is structural, because (5.12.7) is obtained from (5.12.6) by using the doubling assumption, the definition of h and of the perimeter measure. By computing the left-hand side of (5.12.7), the previous inequality can actually be refined, so that out of a Sobolev-type inequality we eventually get a sort of relative isoperimetric inequality, given by

$$\begin{aligned} &\min\{\mathfrak{m}_E(x,\rho), \mathfrak{m}_{E^c}(x,\rho)\} \\ &\quad \leq \ C\left(\frac{\rho^s}{\mathfrak{m}(B(x,\rho))} \right)^{1/(s-1)} P(E,B(x,\lambda\rho))^{s/(s-1)} \end{aligned} \tag{5.12.8}$$

where the definition of $\mathfrak{m}_E(x,\rho)$ was given at (5.12.1). After this discussion, we can finally state the first main result.

Theorem 5.12.3. *Let us assume that the measure $\mathfrak{m}$ is doubling and that $(X,d,\mathfrak{m})$ supports a $(1,1)$-Poincaré inequality. Then there exists a structural constant $\gamma > 0$ such that $P(E,\cdot)$ is concentrated on $(\partial^* E)_\gamma$ and $(\partial^* E)_\gamma$ has finite $\mathscr{H}^h$-measure. In addition:*

(i) $\partial^* E \setminus (\partial^* E)_\gamma$ is $\mathscr{H}^h$-negligible;

(ii) there exist a structural constant $\gamma' > 0$ and a density function $\vartheta : X \to [\gamma', +\infty[$ such that the perimeter measure can be represented as

$$P(E,B) = \int_{B\cap\partial^* E} \vartheta\, d\mathscr{H}^h \tag{5.12.9}$$

Befor the proof, let us spend few words for a couple of interesting remarks. The core of the theorem is the first part and claim (i), since (ii) is a direct consequence of the Radon-Nikodym theorem applied to $\mathscr{H}^h \llcorner \partial^* E$, which is a finite measure once we know that $(\partial^* E)_\gamma$ has finite $\mathscr{H}^h$-measure. However, the representation formula (5.12.9) is not satisfactory yet, because ϑ comes out of a completely abstract result (indeed, the Radon-Nikodym theorem is not constructive) and thus we do not know anything about it in a general setting. Some partial results have been achieved in particular frameworks, as for instance:

- by Franchi-Serapioni-Serra Cassano and Ambrosio-Kleiner-Le Donne in Carnot groups;
- by Ambrosio-Ghezzi-Magnani in Carnot-Carathéodory spaces.

In the first case, what one expects is an explicit formula for ϑ related to the group structure of the space, while in the second case ϑ should be linked to the vector fields inducing the Carnot-Carathéodory structure. However, these problems are still open. Note that this fact has been already stressed for (5.5.2) in Lecture 5.

In practical situations, our aim is to identify ϑ by a blow-up procedure, as outlined at (5.5.3), but there would be no hope for that if we did not have a differentiation result; more precisely, we need to differentiate w.r.t. $P(E,\cdot)$ (observe that differentiation w.r.t. $\mathfrak{m}$ is trivial by the doubling assumption). Fortunately, the forthcoming theorem tells us that the differentiation w.r.t. the perimeter measure is actually possible and therefore one can always try to perform the blow-up procedure, looking for an explicit representation of ϑ.

Theorem 5.12.4. *There exists a structural constant* Σ *such that*

$$\limsup_{\rho\to 0} \frac{P(E, B(x,2\rho))}{P(E, B(x,\rho))} \le \Sigma < +\infty$$

for $P(E,\cdot)$*-a.e.* $x \in X$*.*

Hence, $P(E,\cdot)$ satisfies what is called *asymptotic doubling property* in the literature and this is enough for differentiation purposes, because if a measure is asymptotically doubling, then we can differentiate w.r.t. it. Now, let us provide the demonstration of Theorem 5.12.3.

Proof. In order to simplify the proof as much as possible and remove some totally unnecessary constants, let us assume first that (X,d) is a length space; this is not particularly restrictive, because the doubling assumption and the $(1,1)$-Poincaré inequality entail the quasi-convexity of the distance d and, by virtue of quasi-convexity, the passage from the distance to the length distance is bi-Lipschitz, so that (X,d) is bi-Lipschitz equivalent to the same space X equipped with the length distance. Thanks to this hypothesis, we are allowed to take $\lambda = 1$ in the Poincaré inequality and so remove one of the structural constants. Secondly, let us assume that

$$\rho \mapsto \mathfrak{m}_E(x,\rho)$$

is continuous, which means that the spheres are $\mathfrak{m}$-negligible.

After this preliminary part, let us show that the perimeter measure is concentrated on $(\partial^* E)_\gamma$. To this aim, fix $0 < \gamma < 1/2$ and take into account Egorov's theorem; recall that this theorem allows to change pointwise convergence into uniform convergence, up to possibly remove a set of arbitrarily small measure. What we should prove is that the perimeter of E in any compact set $K \subset X \setminus (\partial^* E)_\gamma$ is zero, but thanks to Egorov's theorem we only need to show that $P(E, K) = 0$ with K compact such that, for some $\rho_0 > 0$ independent of K,

$$\min\{\mathfrak{m}_E(x,\rho), \mathfrak{m}_{E^c}(x,\rho)\} < \gamma \mathfrak{m}(B(x,\rho)) \tag{5.12.10}$$

for every $x \in K$ and for every $\rho \in]0, \rho_0[$. The importance of Egorov's theorem is due to the existence of the constant ρ and to its independence on K, because we know that (5.12.10) holds pointwise outside $(\partial^* E)_\gamma$ for sufficiently small ρ (hence also on every compact set $K \subset X \setminus (\partial^* E)_\gamma$), but from Egorov's theorem we can make the estimate uniform in ρ, up to eliminate a subset of $X \setminus (\partial^* E)_\gamma$ of arbitrarily small $\mathfrak{m}$-measure.

From a heuristic point of view, (5.12.10) tells us that, for any $x \in K$, $B(x, \rho)$ has a very small $\mathfrak{m}$-measure either in E or in E^c and this should lead to the fact that E has a very small perimeter in K: this is the rough idea that we would like to make formal. In other words, a good amount of mass is required both in E and in E^c to have some perimeter of E in K (at least under doubling and $(1, 1)$-Poincaré inequality assumption). Now, let us try to prove concretely what we have just said by using the same covering argument we developed for the proof of Theorem 5.12.2, but first let us briefly explain a fundamental remark. On the one hand

$$\mathfrak{m}_E(x,\rho) + \mathfrak{m}_{E^c}(x,\rho) = \mathfrak{m}(B(x,\rho))$$

and on the other hand (5.12.10) holds with $\gamma < 1/2$; thus, by the continuity of $\rho \mapsto \mathfrak{m}_E(x, \rho)$, only two cases are possible: either

$$\mathfrak{m}_E(x,\rho) < \gamma \mathfrak{m}(B(x,\rho)), \quad \forall \rho \in]0, \rho_0[\tag{5.12.11}$$

or

$$\mathfrak{m}_{E^c}(x,\rho) < \gamma \mathfrak{m}(B(x,\rho)), \quad \forall \rho \in]0, \rho_0[$$

For this reason K splits in two precise subsets. From now on we will restrict ourselves to the subset where (5.12.11) holds; for the other case the procedure will be completely analogous and for this reason we will omit it. After this consideration, let us come back where we previously stopped. By the compactness of K, given $r < \rho_0/2$ (thanks to this hypothesis, (5.12.11) is satisfied for any ball $B(x, 2r)$) we can choose $x_1, ..., x_{N_r} \in K$ such that $d(x_i, x_j) \geq r$ for any $i \neq j$ and

$$K \subset \bigcup_{i=1}^{N_r} B(x_i, r)$$

As a second step, by virtue of the doubling assumption the overlapping of the doubled balls $B(x_i, 2r)$ is controlled by a constant C' depending only on C_D

(and thus independent of r). By overlapping at a given point we mean the number of balls the point belongs to and by overlapping *tout court* we mean its supremum over K. Thus, we know that

$$\sum_{i=1}^{N_r} \chi_{B(x_i,2r)}(x) \leq C', \quad \forall x \in K$$

and this is a well-known fact in doubling metric measure spaces. Roughly speaking, this uniform bound is due to the fact that the centres x_i's are sufficiently far away the ones from the others. Again as in the proof of Theorem 5.12.2, the balls $B(x_i, r)$ will be enlarged by choosing, for every $i = 1, ..., N_r$, a suitable radius $\rho_i \in]r, 2r[$ such that

$$r \frac{d}{ds} \mathfrak{m}_E(x,s)\Big|_{s=\rho_i} \leq \mathfrak{m}_E(x,2r) \tag{5.12.12}$$

$$\mathscr{H}^h(\partial B(x,\rho_i)) < +\infty$$

$$\mathscr{H}^h(\partial B(x,\rho_i) \cap \partial B(x,\rho_j)) = 0, \quad \forall i \neq j \tag{5.12.13}$$

The first requirement can be satisfied by virtue of the mean value theorem and the second one holds for a.e. ρ, so that the two conditions can be asked at the same time; by a recursive procedure in the choice of ρ_i also the third condition is true. Now we can put together all the ingredients we have pointed out. First, by the derivative of the volume-surface area (5.12.2) and by condition (5.12.12) we have

$$P(E \setminus B(x_i,\rho_i), \partial B(x_i,\rho_i)) \leq \frac{\mathfrak{m}_E(x_i,2r)}{r}$$

where the right-hand side can be handled in the following way, thanks to a straightforward application of (5.12.11) and (5.12.8),

$$\begin{aligned} \frac{\mathfrak{m}_E(x_i,2r)}{r} &= \mathfrak{m}_E(x_i,2r)^{1/s} \frac{\mathfrak{m}_E(x_i,2r)^{1-1/s}}{r} \\ &\leq C\gamma^{1/s} P(E, B(x_i,2r)) \end{aligned}$$

so that

$$P(E \setminus B(x_i,\rho_i), \partial B(x_i,\rho_i)) \leq C\gamma^{1/s} P(E, B(x_i,2r)) \tag{5.12.14}$$

Now we can basically repeat the same argument of Theorem 5.12.2, paying more attention to the estimates. To this aim, let us define

$$A_r := \bigcup_{i=1}^{N_r} B(x_i,\rho_i)$$

and observe that on the one hand, by the previous volume estimates, $\mathfrak{m}(A_r) \to 0$ so that $E \setminus A_r$ converges to E. On the other hand, the perimeter of $E \setminus A_r$ can be bounded from above as follows:

$$P(E \setminus A_r, X) = P(E \setminus A_r, X \setminus A_r)$$

$$\begin{aligned} &= P(E \setminus A_r, \partial A_r) + P(E \setminus A_r, X \setminus \overline{A}_r) \\ &\leq P(E \setminus A_r, \partial A_r) + P(E, X \setminus K) \end{aligned}$$

For the first identity, just observe that A_r is open and thus there is no perimeter of $E \setminus A_r$ in A_r. The second identity simply comes from the fact that $P(E \setminus A_r, \cdot)$ is a measure and then additive. Finally, the inequality is motivated by the following remark: firstly $P(E \setminus A_r, X \setminus \overline{A}_r) = P(E, X \setminus \overline{A}_r)$ by the locality of the perimeter measure and secondly by monotonicity of $P(E, \cdot)$ we obtain the upper bound. By lower semicontinuity and the fact that $E \setminus A_r$ converges to E (as previously said), we infer that

$$P(E, X) \leq \liminf_{r \downarrow 0} P(E \setminus A_r, \partial A_r) + P(E, X \setminus K) \tag{5.12.15}$$

Hence, the proof is completed if we are able to show that $P(E \setminus A_r, \partial A_r)$ is infinitesimal as $r \downarrow 0$. To this aim, just note that

$$\begin{aligned} P(E \setminus A_r, \partial A_r) &\leq \sum_{i=1}^{N_r} P(E \setminus A_r, \partial B(x_i, \rho_i)) \\ &\leq \sum_{i=1}^{N_r} \Bigg(P(E \setminus B(x_i, \rho_i), \partial B(x_i, \rho_i)) \\ &\quad + \sum_{j \neq i} P(B(x_j, \rho_j), \partial B(x_i, \rho_i)) \Bigg) \\ &= \sum_{i=1}^{N_r} P(E \setminus B(x_i, \rho_i), \partial B(x_i, \rho_i)) \\ &\leq C' C \gamma^{1/s} P\left(E, \bigcup_{i=1}^{N_r} B(x_i, 2r)\right) \end{aligned}$$

where the first inequality simply follows by subadditivity from

$$\partial A_r \subset \bigcup_{i=1}^{N_r} B(x_i, \rho_i)$$

and in an analogous way we obtain the second inequality, taking into account the fact that

$$E \setminus A_r = (E \setminus B(x_i, \rho_i)) \setminus \bigcup_{j \neq i} B(x_j, \rho_j)$$

and arguing again by subadditivity. Then, recall the absolute continuity of the perimeter measure w.r.t. $\mathscr{H}^h$, because this fact coupled with (5.12.13) provides the identity. Finally, estimate (5.12.14) and the definition of the overlapping constant C' entail the last inequality. Now, let us pass to the limit as $r \downarrow 0$ and observe that $\bigcup_{i=1}^{N_r} B(x_i, 2r)$ collapses on K, so that

$$\liminf_{r \downarrow 0} P(E \setminus A_r, \partial A_r) \leq C' C \gamma^{1/s} P(E, K)$$

Putting this result into (5.12.15) we get

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

and by the arbitrariness of γ we can ask $C'C\gamma^{1/s} < 1$. Therefore, if $P(E,K) > 0$,

$$P(E,X) < P(E,K) + P(E,X\setminus K)$$

and this is impossible. Thus, $P(E,K) = 0$ and this proves the main part of the theorem, that is the fact that $P(E,\cdot)$ is concentrated on $(\partial^* E)_\gamma$. Now, let us briefly see how to show what remains.

First of all, in order to prove that $\mathscr{H}^h((\partial^* E)_\gamma) < +\infty$ note that on the one hand, by the very definition of γ-essential boundary, in $(\partial^* E)_\gamma$

$$\limsup_{\rho\downarrow 0} \min\left\{\frac{\mathfrak{m}_E(x,\rho)}{\mathfrak{m}(B(x,\rho))}, \frac{\mathfrak{m}_{E^c}(x,\rho)}{\mathfrak{m}(B(x,\rho))}\right\} \geq \gamma > 0 \tag{5.12.16}$$

and therefore it is possible to deduce that

$$\limsup_{\rho\downarrow 0} ⨍_{B(x,\rho)} |\chi_E - \overline{\chi_E}|^{1^*}\, d\mathfrak{m} \geq \tilde{\gamma} > 0$$

for a suitable constant $\tilde{\gamma}$; on the other hand, by (5.12.7)

$$\limsup_{\rho\downarrow 0} ⨍_{B(x,\rho)} |\chi_E - \overline{\chi_E}|^{1^*}\, d\mathfrak{m} \leq \limsup_{\rho\downarrow 0} \left(C\frac{P(E,B(x,\rho))}{h(B(x,\rho))}\right)^{1^*}$$

where C is a structural constant (recall that we are allowed to take $\lambda = 1$, thanks to what we have already said at the beginning of the proof). Thus, we have

$$\limsup_{\rho\downarrow 0} \frac{P(E,B(x,\rho))}{h(B(x,\rho))} \geq \frac{\tilde{\gamma}^{1/1^*}}{C} \tag{5.12.17}$$

for any $x \in (\partial^* E)_\gamma$ and by general results of measure theory concerning Hausdorff measures we deduce that $\mathscr{H}^h((\partial^* E)_\gamma) < +\infty$. Combining this information with the previous one, i.e. the concentration of the perimeter measure on $(\partial^* E)_\gamma$, we immediately infer that

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

and property (ii) is a straightforward application of the Radon-Nikodym theorem, as already noticed. □

After the proof of Theorem 5.12.3, let us spend few words on Theorem 5.12.4 too.

Proof. Roughly speaking, in order to prove that $P(E,B(x,\rho))$ is asymptotically doubling the main idea is to compare it with a quantity we already know to be doubling, that is h. On the one hand, by general facts which do not even require the $(1,1)$-Poincaré inequality we know that

$$\limsup_{\rho\downarrow 0} \frac{P(E,B(x,\rho))}{h(B(x,\rho))} < +\infty, \quad \text{for } P(E,\cdot)\text{-a.e. x } \in X$$

and on the other hand, as we have just shown in the proof of Theorem 5.12.3 and more precisely at (5.12.17),

$$\limsup_{\rho\downarrow 0} \frac{P(E,B(x,\rho))}{h(B(x,\rho))} > 0$$

If we are able to refine the last inequality by replacing the limsup with the liminf, then the conclusion follows by comparison. This amounts to pass from (5.12.16) to

$$\liminf_{\rho\downarrow 0} \min\left\{\frac{\mathfrak{m}_E(x,\rho)}{\mathfrak{m}(B(x,\rho))}, \frac{\mathfrak{m}_{E^c}(x,\rho)}{\mathfrak{m}(B(x,\rho))}\right\} \geq \gamma$$

because if we do that, then it will be sufficient to apply the relative isoperimetric inequality (5.12.8) to get the same conclusion. This last improvement relies on a sort of ODE argument, which first appeared in a paper of White in the 90's; later, in 2000 Cheeger applied the same fundamental ideas to Sobolev functions and in 2002 Ambrosio generalized them to sets of finite perimeter.

The heuristic idea is the following: if the limsup is strictly positive and the liminf is zero, then large oscillations of the volume fraction happen at any scale. Let us briefly explain why this is impossible. As a first step, let us define

$$\nu(\rho) := \Big(\min\{\mathfrak{m}_E(x,\rho), \mathfrak{m}_{E^c}(x,\rho)\}\Big)^{1/s}$$

and say that ν satisfies an ODE; such an ODE comes out of the quasi-minimality of E (thus, a connection with minimal surfaces is present) and this is surprising, because we are dealing with a set E only assumed to be of finite perimeter. Nevertheless, roughly speaking we can say that an additive and lower semicontinuous energy implies quasi-minimality on small scales for objects with finite energy; let us restate this claim in a more precise and rigorous way in the case of sets of finite perimeter. Let $d \in]0,1/2[$ and $M > 1$ (the first controlling the volume fraction and the latter the quasi-minimality, in the sense that the more M is near to 1, the more E is near to minimality); then, for $P(E,\cdot)$-a.e. $x \in X$ there exists a critical scale $\rho(x) > 0$ such that, for every $\rho \in]0, \rho(x)[$, if

$$\frac{1}{2}\mathfrak{m}(B(x,\rho)) \geq \mathfrak{m}_E(x,\rho) \geq d\mathfrak{m}(B(x,\rho))$$

then

$$P(E,B(x,\rho)) \leq MP(E\setminus B(x,\rho), \partial B(x,\rho))$$

and in this sense we have to interpret quasi-minimality: the perimeter of E is minimal at small scales, up to the constant M, w.r.t. modifications of the form

$E \setminus B(x,\rho)$ (actually, more general statements with more general modifications are possible, but for our purposes this is enough). In order to prove this principle we can argue by contradiction, but we will not give the details. If we know that

$$\frac{1}{M} P(E, B(x,\rho)) > P(E \setminus B(x,\rho), \partial B(x,\rho))$$

on many balls, then we can use these balls to cut E in a new set $\tilde{E}$ very close to E (here a fundamental role is played by the additivity of the energy) but with small perimeter; then, passing to the limit and using the lower semicontinuity of the energy, a contradiction follows. □

Unfortunately, our time is over; thus, for a rigorous proof of this theorem, the interested reader is addressed to [3]. The argument we used in the last part of the proof can be easily rephrased in Sobolev spaces, because we can say that any Sobolev function is almost minimal if, for $\mathfrak{m}$-a.e. point, we look at sufficiently small scales; *a posteriori* this is clear by Rademacher theorem.

Chapter 6

Hypoelliptic operators

This section is based on the course held by Nicola Garofalo and Fabrice Baudoin at the IHP in Paris during the IHP Trimester *Geometry, Analysis and Dynamics on Sub-Riemannian manifolds*. The course is divided in two parts.

In this second part we focus on the geometric framework in which the generalized curvature-dimension estimate is available. In particular, we will prove that if a Riemannian foliation is totally geodesic, then under natural geometric conditions, the horizontal Laplacian satisfies the generalized curvature dimension inequality. In the last lecture, we will study some problems related to Kolmogorov type operators and discuss methods to prove hypocoercive estimates.

6.1 Lecture 1 - 8 September

6.2 Lecture 2 - 9 September

6.3 Lecture 3 - 10 September

6.4 Lecture 4 - 11 September

6.5 Lecture 5 - 15 September

6.6 Lecture 6 - 16 September

6.7 Lecture 7 - 17 September

6.8 Lecture 8 - 18 September

6.9 Lecture 9 - 22 September

6.10 Lecture 10 - 23 September

6.11 Lecture 11 - 24 September

6.12 Lecture 12 - 26 September

Chapter 7

Geodesics in sub-Riemannian manifolds

This section is based on the course held by Frédéric Jean and Andrei Agrachev at the IHP in Paris during the IHP Trimester *Geometry, Analysis and Dynamics on Sub-Riemannian Manifolds*. However, the course title is not completely fair, as a more appropriate one would be *Sub-Riemannian geometry from a control theory viewpoint*. The course is divided in two parts and the structure of the first one is the following:

(i) sub-Riemannian structures;

(ii) first order theory (nilpotent approximation);

(iii) metric tangent space;

(iv) desingularization;

(v) Hausdorff measures.

For such part, main reference is [12], but sometimes we will also refer to [1], [5], [17] and [20].

7.1 Lecture 1 - 14 October

Abstract

Aim of this introductory lecture is to provide the reader with the basic notions of sub-Riemannian geometry, but the approach we chose is mostly tailored on optimal control theory. We first introduce the notion of sub-Riemannian manifold in three different ways and we prove a key tool: Chow-Rashevskii's theorem. As last part, we start to point out some basic distance estimates.

In sub-Riemannian geometry there are several ways how sub-Riemannian distances arise: for instance, when studying hypoelliptic operators, such distances arise from vector fields, but there are many other (geometric) approaches. We will focus our attention on three different types. Before starting, let us point out that the framework we will work in is given by an n-dimensional manifold M.

A: control systems

In this first case, the data is an m-tuple $(X_1, ..., X_m)$ of smooth vector fields on M and we consider the following ODE

$$\dot{q} = \sum_{i=1}^{m} u_i X_i(q) \tag{7.1.1}$$

where $q \in M$ and $u = (u_1, ..., u_m) \in \mathbb{R}^m$; we will address to (7.1.1) as *control system.* In particular, such control system is called *non-holonomic control system.*

Definition 7.1.1. *A curve $\gamma : [0, T] \to M$ is said to be a* trajectory *of (7.1.1) if it is absolutely continuous and there exists a function $u : [0, T] \to \mathbb{R}^m$ belonging to L^1 such that γ solves (7.1.1), that is*

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t) X_i(\gamma(t)) \quad \textit{for a.e. } t$$

The function u is called a control of the system.

It is important to stress that even if a curve γ solves (7.1.1) for a suitable control u, nevertheless u need not to be absolutely continuous and for this reason in the above definition absolute continuity is explicitly requested. As a counter-example, it is sufficient to consider Cantor function, whose derivative is zero (and thus solves the system $\dot{q} \equiv 0$) but it is not absolutely continuous. Furthermore, if γ is a trajectory, then $\dot{\gamma}(t) \in \mathscr{D}(\gamma(t))$ for a.e. t, where $\mathscr{D}$ is the C^∞-module generated by $X_1, ..., X_m$, that is

$$\mathscr{D} = C^\infty - \text{span}(X_1, ..., X_m)$$

and so

$$\mathscr{D}(\gamma(t)) = C^\infty - \text{span}(X_1(\gamma(t)), ..., X_m(\gamma(t)))$$

However, the converse is not true; that is, an absolutely continuous curve γ such that $\dot{\gamma}(t) \in \mathscr{D}(\gamma(t))$ need not to be a trajectory. After these preliminary considerations, we can introduce a sub-Riemannian metric, a notion of length and a sub-Riemannian distance in the following way.

sub-Riemannian metric: we define the map $g : TM \to [0, +\infty]$ by

$$g(q, v) := \inf\{|u|^2 : \sum_{i=1}^{m} u_i X_i(q) = v\} \tag{7.1.2}$$

where $q \in M$ and $v \in T_qM$. Notice that $+\infty$ can actually be attained, since if $v \notin \mathscr{D}(q)$, then $\sum_{i=1}^m u_i X_i(q) = v$ cannot be satisfied and thus $g(q,v)$ equals the infimum of $|u|^2$ taken over an empty set and by definition this is $+\infty$. On the contrary, if $v \in \mathscr{D}(q)$, then the infimum is actually a minimum and there exists a unique minimizer, say u^*, so that $g(q,v) = |u^*|^2$.

length: for an absolutely continuous path $\gamma : [0,T] \to M$, its length is defined by

$$\ell(\gamma) := \int_0^T \sqrt{g(\gamma(t), \dot{\gamma}(t))}dt$$

The definition is well posed, since $\sqrt{g(\gamma(t), \dot{\gamma}(t))}$ is measurable.

sub-Riemannian distance: given two points $p, q \in M$, we set their distance as

$$d(p,q) := \inf\{\ell(\gamma) \,:\, \gamma \text{ is a.c.}, \gamma(0) = p, \gamma(T) = q\}$$

but it is not clear yet that d actually defines a distance. Nevertheless, we can observe that if $\ell(\gamma) < +\infty$, then γ is a trajectory of (7.1.1) and moreover

$$\ell(\gamma) = \min\{\|u\|_{L^1} \,:\, u \text{ is a control associated to } \gamma\} \tag{7.1.3}$$

B: distributions

In this second case our data is (Δ, g^Δ), where Δ is a distribution (i.e. a subbundle of TM) and g^Δ is a Riemannian metric on Δ, so that for each $q \in M$, $g_q^\Delta : \Delta_q \to \mathbb{R}$ is a quadratic form. In this environment, the following definition is provided.

Definition 7.1.2. *A curve $\gamma : [0,T] \to M$ is called an* horizontal path *if it is absolutely continuous and $\dot{\gamma}(t) \in \Delta_{\gamma(t)}$ for a.e. t, that is the curve is tangent to the distribution.*

The sub-Riemannian metric on TM is given by g^Δ, in the sense that

$$g(q,v) := \begin{cases} g^\Delta(q,v) & \text{if } v \in \Delta_q \\ +\infty & \text{otherwise} \end{cases}$$

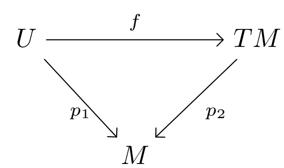
and the length can be defined as in the previous case (but w.r.t. g^Δ, of course). Finally, we can introduce the sub-Riemannian distance as follows.

$$d(p,q) := \inf\left\{ \int_0^T \sqrt{g^\Delta(\gamma(t), \dot{\gamma}(t))}dt \,:\, \gamma \text{ horizontal}, \gamma(0) = p, \gamma(T) = q \right\}$$

By now we have two different approaches for introducing sub-Riemannian objects, but if $\mathrm{rk}(X_1, ..., X_m) = m$, i.e. $\mathrm{rk}(X_1(q), ..., X_m(q)) = m$ for every $q \in M$ (so that $\mathscr{D}$ can be identified with the distribution Δ), then approaches **A** and **B** coincide. However, this further condition does not always hold: one can think about $\mathbb{S}^2$.

C: sub-Riemannian structures

In the third case, our data is the triple (U, g^U, f), where U is a vector bundle on M, g^U is a Riemannian metric on U, so that $g^U_q : U_q \to \mathbb{R}$ is a quadratic form (the couple (U, g^U) is also called *Riemannian* or *Euclidean bundle*), and $f : U \to TM$ is a morphism of vector bundles, that is the diagram below commutes (where π_1 and π_2 are the canonical projections) and f is linear on fibers, i.e. $f : U_q \to T_qM$ is linear for every $q \in M$.



More briefly, the commutativity condition can be written as

$$p_1 = p_2 \circ f$$

In this framework, the sub-Riemannian metric on TM is given by

$$g(q, v) = \inf\{g^U(q, u) \,:\, u \in U_q,\ f(q, u) = (q, v)\}$$

The length and the sub-Riemannian distance can be defined as in **A**.

The approach **C** is particularly relevant because it is a generalization of the previous ones, as we are going to see in the first two examples, but it is also a generalization of the smooth case.

Example 7.1.3. Let $U = M \times \mathbb{R}^m$, $g^U(q, u) = |u|^2$ the Euclidean norm and $f(q, u) = (q, \sum_{i=1}^m u_i X_i(q))$ for a given m-tuple $(X_1, ..., X_m)$ of smooth vector fields. Then for this triple (U, g^U, f) the sub-Riemannian metric, the length and the sub-Riemannian distance we get coincide with those of the approach **A**. $\Diamond$

Example 7.1.4. Let $U = \Delta$ for a given distribution Δ, $g^U = g^\Delta$ and $f : \Delta \hookrightarrow TM$ the inclusion. In this case the sub-Riemannian metric g is not an infimum, but a minimum, and the objects we get coincide with those of the approach **B**. $\Diamond$

Example 7.1.5. Let $U = TM$, $g^U = g$ for any Riemannian metric g on M and $f = \mathrm{Id}$. Then the sub-Riemannian framework coincides with the Riemannian one. $\Diamond$

Thus we can informally write **A** $\subset$ **C** and **B** $\subset$ **C**, but actually approaches **A** and **C** are equivalent in a global (and not only local) sense.

Theorem 7.1.6. *Let (U, g^U, f) be a sub-Riemannian structure on M, let g and d be respectively the sub-Riemannian metric and distance associated to the sub-Riemannian structure provided and let*

$$\mathscr{D} := \{f(Z) \,:\, Z \text{ smooth section on } U\} \subset VF(M)$$

where $VF(M)$ denotes the set of vector fields on M. Then there exist smooth vector fields $X_1, ..., X_m$ such that $\mathscr{D} = C^\infty - \operatorname{span}(X_1, ..., X_m)$ and g^U coincides with the right-hand side in (7.1.2). As the length and the sub-Riemannian distance of (U, g^U, f) coincide already with the counterparts associated to $(X_1, ..., X_m)$, then in conclusion we have $\boldsymbol{A} = \boldsymbol{C}$.

Actually, if $\operatorname{rk} U = k$, then the following bound on the number of vector fields $X_1, ..., X_m$ provided by the theorem above holds.

$$m \leq 2(n + k)$$

Proof. First of all, by classical results on vector bundles (proved in the 70's), there exists a vector bundle U' on M such that

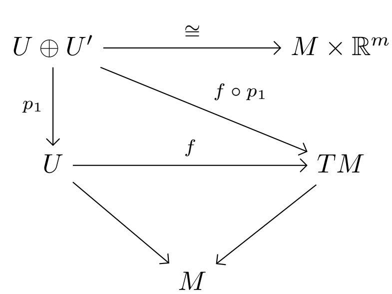
$$\tilde{U} := U \oplus U' \cong M \times \mathbb{R}^m$$

where $\tilde{U} := U \oplus U'$ means that $\tilde{U}_q = (U_q, U'_q)$. The basic idea in order to prove such classical result is the following: if $\operatorname{rk} U = k$, then U is a smooth manifold of dimension $n + k$ and then it can be embedded in $\mathbb{R}^m$ for m large enough; similarly, TU can be embedded in $\mathbb{R}^m \times \mathbb{R}^m$. Hence, such classical result relies on Whitney embedding theorem.

In this way we have produced the first element of a new sub-Riemannian structure. The second one can be simply provided by

$$g^{\tilde{U}}(q, (y, y')) := g^U(q, y) + g^{U'}(q, y')$$

for a chosen Riemannian metric $g^{U'}$ on U'. In this way we get a sub-Riemannian metric on $M \times \mathbb{R}^m$, which is isomorphic to $\tilde{U}$. Finally, the third element of the new sub-Riemannian structure is given by $f \circ p_1$, where $p_1 : U \oplus U' \to U$ is the canonical projection on the first factor. The situation is depicted in the diagram below.



Now we observe that there exists a global orthonormal frame $Y_1, ..., Y_m$ for $\tilde{U}$; this means that for every $q \in M$, any $z \in \tilde{U}_q$ can be written as

$$z = \sum_{i=1}^{m} u_i Y_i(q)$$

for suitable functions $u_1, ..., u_m$ (i.e. $Y_1, ..., Y_m$ generate the fibers) and, moreover, $g^{\tilde{U}}(q, z) = |u|^2$. By means of $f \circ p_1$, this global orthonormal frame can be pushed into the family $X_1, ..., X_m$ of smooth vector fields on M defined as

$$X_i := f \circ p_1(Y_i), \quad i = 1, ..., m$$

and the conclusion follows. Indeed, if Z is a smooth section of U, then $Z = p_1(W)$ for a suitable smooth section W of $\tilde{U}$; as $W \in \text{span}(Y_1, ..., Y_m)$, by definition of $\mathscr{D}$ we get that

$$\mathscr{D} = C^\infty - \text{span}(f \circ p_1(Y_i)\, ;\, i = 1, ..., m) = C^\infty - \text{span}(X_1, ..., X_m)$$

Similarly one can prove that g^U coincides with the right-hand side in (7.1.2). □

Hence having a family $(X_1, ..., X_m)$ of smooth vector fields on M is equivalent to having a sub-Riemannian structure (U, g^U, f). The first approach is easier to handle for computations, whereas the second one fits better from a theoretical point of view, since it is intrinsic. Now let us show that the sub-Riemannian distance defined in **A** (or equivalently in **C**) is actually a distance. Symmetry and the triangle inequality are very easy properties to prove. Indeed, for the first one if γ is a trajectory of the control system (7.1.1) from p to q on the time interval $[0, T]$, then $\gamma(T - \cdot)$ is a trajectory of (7.1.1) from q to p and the length is the same; thus $d(p, q) = d(q, p)$. For the trinagle inequality, just notice that the concatenation of a trajectory from p to q with a second trajectory from q to r provides a trajectory from p to r. Moreover, if $p = q$, then trivially $d(p, q) = 0$. What is not trivial at all to prove are the following two facts:

(i) if $d(p, q) = 0$, then $p = q$;

(ii) $d < +\infty$.

Both are consequences of Chow-Rashevskii's theorem, whose statement is going to be immediately recalled. We are also going to provide a sketch of Chow's original proof, since some aspects will be needed later. But before, a further definition is required.

Definition 7.1.7. *Let $\mathscr{F}$ be a family of vector fields. We say that $\mathscr{F}$ is* bracket-generating *if*

$$\dim \mathfrak{Lie}(\mathscr{F}) = n \tag{7.1.4}$$

where $\mathfrak{Lie}(\mathscr{F})$ is the Lie algebra generated by $\mathscr{F}$, that is the span of all iterated Lie brackets between elements of $\mathscr{F}$ or in an explicit way

$$\mathfrak{Lie}(\mathscr{F}) := \text{span}(\{[Y_i, [...[Y_{k-1}, Y_k],]...] \,:\, Y_i \in \mathscr{F},\, k \in \mathbb{N}\})$$

As $\mathfrak{Lie}_q(\mathscr{F}) \subset T_qM$, where $\mathfrak{Lie}_q(\mathscr{F})$ is the Lie algebra at $q \in M$ defined as follows

$$\mathfrak{Lie}_q(\mathscr{F}) := \{Y(q) \,:\, Y \in \mathfrak{Lie}(\mathscr{F})\}$$

then condition (7.1.4) can be equivalently restated as

$$\mathfrak{Lie}_q(\mathscr{F}) = T_qM, \quad \forall q \in M$$

The bracket-generating condition is also known as *Chow's condition* or *Lie algebra rank condition* (LARC) in control theory and *Hörmander condition* in the context of PDEs. It can also be formulated in erms of distributions: more precisely, a distribution is said to be bracket generating if the vector fields $X_1, ..., X_m$ generating it are bracket-generating. In order to state Chow-Rashevskii's theorem, we also need to introduce, for a given point $p \in M$, the *attainable set starting from* p by

$$\mathscr{A}_p := \{\gamma(T) \,:\, \gamma \text{ is a trajectory starting from } p\}$$

and notice that if $\mathscr{A}_p = M$ for every $p \in M$, then clearly $d < +\infty$ and so (ii).

Theorem 7.1.8 (Chow-Rashevskii). *If M is connected and $(X_1, ..., X_m)$ is bracket-generating, then any two points of M can be joined by a trajectory of (7.1.1), so that $\mathscr{A}_p = M$ for every $p \in M$.*

Proof. Let $p \in M$ and notice that we just have to prove that $\mathscr{A}_p$ is a neighbourhood of p. Indeed, if it is this case, then the equivalence classes of the equivalence relation $d(p, q) < +\infty$ are open sets. Since M is connected there is only one such class and thus $\mathscr{A}_p = M$. For sake of simplicity assume $M = \mathbb{R}^3$ and $\mathscr{F} = \{X_1, X_2\}$. As $\mathscr{F}$ is bracket-generating, then

$$\mathrm{rk}(X_1(p), X_2(p), [X_1, X_2](p)) = 3$$

at any $p \in M$. Now let ϕ_t^{12} be the commutator of the flows e^{tX_1} and e^{tX_2}, namely

$$\phi_t^{12} := [e^{tX_1}, e^{tX_2}] = e^{-tX_2} \circ e^{-tX_1} \circ e^{tX_2} \circ e^{tX_1}$$

It is a well known fact that p and ϕ_t^{12} differ by $t^2[X_1, X_2](p)$ more or less or, which is the same, the tangent to the flow ϕ_t^{12} is nearly $[X_1, X_2]$; in a rigorous way,

$$\phi_t^{12} = \mathrm{Id} + t^2[X_1, X_2] + o(t^2)$$

To obtain a C^1-diffeomorphism whose time-derivative is exactly $[X_1, X_2]$ we set

$$\psi_t^{12} := \begin{cases} \phi_{\sqrt{t}}^{12} & \text{if } t \geq 0 \\ \phi_{\sqrt{-t}}^{21} & \text{if } t < 0 \end{cases}$$

In this way

$$\psi_t^{12} = \mathrm{Id} + t[X_1, X_2] + o(t) \tag{7.1.5}$$

Finally, let us introduce the map $\Phi_p : \mathbb{R}^3 \to M = \mathbb{R}^3$ by

$$\Phi_p(t_1, t_2, t_3) := \psi^{12}_{t_3} \circ e^{t_2 X_2} \circ e^{t_1 X_1}(p)$$

and notice that $\Phi_p(0) = p$. Moreover, by (7.1.5) this map is C^1 near 0 and its derivative in this point, given by

$$D\Phi_p(0) = (X_1(p), X_2(p), [X_1, X_2](p))$$

is invertible. Hence Φ_p is a local C^1-diffeomorphism and the image under Φ_p of a neighbourhood Ω of 0 contains a neighbourhood of p. As $\mathscr{A}_p \subset \Phi_p(\Omega)$, then the conclusion follows. $\square$

From this result, it follows that d is a distance (also called *Carnot-Carathéodory distance*) and so (M, d) is a metric space, also known as *Carnot-Carathéodory space.*

In order to generalize the proof to higher dimension and to an m-distribution generated by $\mathscr{F}$, for a multi-index $I = iJ$ let us define by induction the local diffeomorphism $\phi^I_t := [e^{tX_i}, \phi^J_t]$ and let us introduce the notation

$$X_I := [X_{i_1}, [X_{i_2}, ..., [X_{i_{k-1}}, X_{i_k}],]...]$$

where $I = (i_1, ..., i_k)$. By Campbell-Hausdorff formula we have

$$\phi^I_t = \mathrm{Id} + t^{|I|} X_I + o(t^{|I|})$$

and thus the same argument applies.

Actually, the proof of Chow-Rashevskii's theorem gives a little bit more than accessibility. Assume first that the vector fields X_1, X_2 (the ones of $\mathscr{F}$) are an orthonormal frame of (Δ, g), our sub-Riemannian structure. For ε small enough, any $e^{tX_i}(x)$ for $t \in [0, \varepsilon]$ is a trajectory of length ε and so, thinking to the picture above, $\phi^{12}_t(p)$ for $t \in [0, \varepsilon]$ is a curve of length $4|t|$; thus, $\Phi_x(t_1, t_2, t_3)$ is the endpoint of a trajectory of length less than $4(|t_1|^{1/2} + |t_2|^{1/2} + |t_3|^{1/2})$, which is in turn less than $C|t|^{1/2}$ for a suitable constant C. Therefore, we have un upper bound for the distance:

$$d(x, \Phi_x(t)) \leq C|t|^{1/2}$$

This kind of estimate of the distance in function of local coordinates plays an important role in sub-Riemannian geometry and since we are working in a simplified framework, than we can say that

$$d(x, y) \leq C|x - y|^{1/2} \tag{7.1.6}$$

However, in a general setting the argument is a bit more tricky, because (t_1, t_2, t_3) are not local coordinates and so we have to choose a suitable coordinate system. Moreover, the exponent in the above estimate depends on the dimension. Therefore, the conclusion we get is the following

$$d(p, q^y) \leq \tilde{C}|y|^{1/r} \tag{7.1.7}$$

where q^y denotes the point of coordinates y and r is a sufficiently large integer. Now, (7.1.6) and more generally (7.1.7) allow us to compare d to a Riemannian distance. Indeed, if g^R is a Riemannian metric compatible with g, which means that $g^R|_\Delta = g$, and d_R is the associated Riemannian distance, then by construction $d_R(p,q) \leq d(p,q)$. Moreover, near p, $d_R(p,q^y) \geq \hat{C}|y|$. In this way we obtain a first estimate of the sub-Riemannian distance d, that can be stressed in the following statement.

Theorem 7.1.9. *Assume that $(X_1,...,X_m)$ is bracket-generating. For any Riemannian metric g^R compatible with g we have, for q close enough to p,*

$$d_R(p,q) \leq d(p,q) \leq c d_R(p,q)^{1/r}$$

where the constant c depends on p and g^R, whereas r depends only on p and not on g^R.

As a consequence, the sub-Riemannian distance is continuous.

Corollary 7.1.10. *If $(X_1,...,X_m)$ is bracket generating, then the topology of the metric space (M,d) is the original topology of M.*

7.2 Lecture 2 - 16 October

Abstract

In the second lecture we are mostly interested in the infinitesimal structure of sub-Riemannian manifolds. For this reason we first provide the notion of non-holonomic order for functions and vector fields, together with the basic calculus rules. Motivated by this definition, we introduce privileged coordinates and w.r.t. a fixed privileged frame we construct the nilpotent approximation of a sub-Riemannian manifold.

In the previous lecture we have seen that the good way to introduce a sub-Riemannian structure is via (U,g^U,f). Now let us talk about the infinitesimal point of view, also called *first order theory*; in the whole section we will assume that the bracket-generating assumption holds, we will denote by d the sub-Riemannian distance associated to our sub-Riemannian structure, we fix a point $p \in M$ and we also fix a local orthonormal frame $X_1,...,X_m$ of the sub-Riemannian structure in a neighbourhood of p. The basic idea in first order theory is the notion of linearization. In order to introduce it, let us begin with a concrete example: in $\mathbb{R}^n$, if f is a sufficiently regular function, then

$$f(x) - f(x_0) = \text{terms in } O(|x|) + \text{higher order terms}$$

and this is well known, but if we consider the Heisenberg group $\mathbb{H}^3$, then the distance is not Euclidean and the situation is completely different; indeed, movements along the z-axis behave like $\sqrt{z}$ in distance and so the naïf idea of linearization (as meant in the Euclidean case) does not fit this sub-Riemannian framework. Hence, first of all we need to provide a suitable definition of order.

1: Non-holonomic order

Roughly speaking, the notion of non-holonomic order of a given function f relies on the following consideration: we want that in a neighbourhood of p the function f behaves like the sub-Riemannian distance d to a suitable exponent. Hence, we formulate the next definition.

Definition 7.2.1. *Let $f : M \to \mathbb{R}$ be a smooth function. The* non-holonomic order of f at p *is defined as*

$$\mathrm{ord}_p(f) := \sup\{s \,:\, f(q) = O(d(p,q))^s\}$$

In the Euclidean case, the order of a function is characterized in terms of algebraic conditions on its partial derivatives. More precisely, it is well known that $\mathrm{ord}_p(f) \geq s$ if and only if

$$\frac{\partial^k f}{\partial x_{i_1} ... \partial x_{i_k}} = 0, \quad \forall k < s$$

We would like to do the same in our sub-Riemannian framework; however, here we do not have partial derivatives, but only horizontal directions $X_1, ..., X_m$ (i.e. vector fields tangent to the distribution describing the sub-Riemannian structure). Therefore, we need a further definition.

Definition 7.2.2. *Let $f : M \to \mathbb{R}$ be a smooth function. The* non-holonomic derivative of f of order 0 *is set as $f(p)$ at any $p \in M$. We call* non-holonomic derivatives of f of order 1 *the Lie derivatives $X_1 f, ..., X_m f$. Furthermore, we call* non-holonomic derivatives of f of order 2 *all the $X_i(X_j f)$ for $i, j = 1, ..., m$. In an iterated way, all the non-holonomic derivatives of f of order s can be introduced.*

Now we can prove that even in a sub-Riemannian setting a characterization of the order analogous to the Euclidean one holds.

Lemma 7.2.3. *Let $s \geq 0$ be an integer. We have that $\mathrm{ord}_p(f) \geq s$ if and only if all the non-holonomic derivatives of f of order strictly smaller than s vanish at p.*

Proof. Let us show the two implications.

$\Rightarrow$ We have to prove that for any $k < \mathrm{ord}_p(f)$, $X_{i_1} ... X_{i_k} f(p) = 0$ for any multi-index $(i_1, ..., i_k)$. Note first that a non-holonomic derivative of f of order k can be written as

$$(X_{i_1} ... X_{i_k} f)(p) = \frac{\partial^k}{\partial t_1 ... \partial t_k} f(e^{t_k X_{i_k}} \circ ... \circ e^{t_1 X_{i_1}}(p))\bigg|_{t=0}$$

The point $q(t) := e^{t_k X_{i_k}} \circ ... \circ e^{t_1 X_{i_1}}(p)$ is the endpoint of a trajectory of length $|t_1| + ... + |t_k|$, so that

$$d(p, q(t)) \leq |t_1| + ... + |t_k| \leq C|t|$$

and since by assumption $\mathrm{ord}_p(f) \geq s$, then

$$f(q(t)) \leq \tilde{C} d(p, q(t))^{\mathrm{ord}_p(f)} \leq \hat{C} |t|^{\mathrm{ord}_p(f)}$$

whence

$$\frac{\partial^k f(q(t))}{\partial t_1 ... \partial t_k} = 0, \quad \forall k < \mathrm{ord}_p(f)$$

and so the conclusion.

$\Leftarrow$ The proof goes by induction on s. For $s = 0$ there is nothing to prove. Assume that the implication holds for a given $s \geq 0$ (induction hypothesis) and take a function f such that all its non-holonomic derivatives of order strictly smaller than $s+1$ vanish at p. Observe that, for $i = 1, ..., m$, all the non-holonomic derivatives of $X_i f$ of order strictly smaller than s vanish at p: indeed, $X_{i_1} ... X_{i_k}(X_i f) = X_{i_1} ... X_{i_k} X_i f$. Applying the induction hypothesis to $X_i f$ yields $\mathrm{ord}_p(X_i f) \geq s$. In other terms, there exist positive constants $C_1, ..., C_m$ such that, for q close enough to p,

$$X_i f(q) \leq C_i d(p, q)^s$$

Fix now a point q near p and let γ be a minimizing curve joining p to q (the existence will be proved later; we can also assume it has velocity one). This means that γ satisfies

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t) X_i(\gamma(t)) \quad \text{for a.e. } t \in [0, T], \quad \gamma(0) = p,\ \gamma(T) = q$$

with $\sum_{i=1}^{m} u_i^2(t) = 1$ a.e. and $d(p, q) = \ell(\gamma) = T$. Actually, every sub-arc of γ is still length-minimizing, so that $d(p, \gamma(t)) = t$ for any $t \in [0, T]$. In order to estimate $f(\gamma(T))$, notice that

$$\frac{d}{dt} f(\gamma(t)) = \sum_{i=1}^{m} u_i(t) X_i f(\gamma(t))$$

whence

$$\left| \frac{d}{dt} f(\gamma(t)) \right| \leq \sum_{i=1}^{m} |u_i(t)| C_i d(p, \gamma(t))^s \leq C t^s$$

where $C := C_1 + ... + C_m$. By integration on $[0, t]$, this inequality provides

$$|f(\gamma(t))| \leq |f(p)| + \frac{C}{s+1} t^{s+1}$$

Since $f(p) = 0$ (it is the non-holonomic derivative of f of order 0), then at $t = T = d(p, q)$ we finally get

$$|f(q)| \leq \frac{C}{s+1} T^{s+1}$$

and so the conclusion. □

Notice that the order does not depend on the chosen frame $X_1, ..., X_m$. Actually, the order is a non-holonomic notion rather than a sub-Riemannian one, as it depends on the fiber bundle U and not on the Riemannian metric g. Moreover, we have the classical properties of orders.

$$\begin{aligned}
&\mathrm{ord}_p(fg) \geq \mathrm{ord}_p(f) + \mathrm{ord}_p(g) \\
&\mathrm{ord}_p(\lambda f) = \mathrm{ord}_p(f), \quad \forall \lambda \in \mathbb{R}^* \\
&\mathrm{ord}_p(f+g) \geq \min\{\mathrm{ord}_p(f), \mathrm{ord}_p(g)\}
\end{aligned}$$

even if the first inequality is actually an equality. In a similar way we can also define order for vector fields.

Definition 7.2.4. *Let X be a vector field on M. Its order at p is defined as*

$$\mathrm{ord}_p(X) := \sup\{\sigma \in \mathbb{Z} \,:\, \mathrm{ord}_p(Xf) \geq \mathrm{ord}_p(f) + \sigma \,\forall f \in C^\infty(p)\}$$

where $C^\infty(p)$ denotes the family of smooth functions defined near p.

Pay attention to the fact that, with this definition, the i-th Euclidean derivative $\frac{\partial}{\partial x_i}$ has order -1. Moreover, for any vector field X, we have $\mathrm{ord}_p(X) \geq -1$ and equality holds if and only if $X(p) \neq 0$; from this fact, it follows that $\mathrm{ord}_p([X_i, X_j]) \geq -2$ and, more generally, $\mathrm{ord}_p(X_I) \geq -|I|$. Finally, the following properties hold.

$$\begin{aligned}
&\mathrm{ord}_p([X,Y]) \geq \mathrm{ord}_p(X) + \mathrm{ord}_p(Y) \\
&\mathrm{ord}_p(fX) \geq \mathrm{ord}_p(f) + \mathrm{ord}_p(X), \quad \forall f \in C^\infty(p) \\
&\mathrm{ord}_p(X+Y) \geq \min\{\mathrm{ord}_p(X), \mathrm{ord}_p(Y)\}
\end{aligned}$$

Also in this case, the second inequality is in fact an equality. We are able now to precise the meaning of approximation of a family $X_1, ..., X_m$ of vector fields.

Definition 7.2.5. *A system of vector fields $\hat{X}_1, ..., \hat{X}_m$ defined near p is called a* first-order approximation *of $X_1, ..., X_m$ at p if $\mathrm{ord}_p(X_i - \hat{X}_i) \geq 0$ for any $i = 1, ..., m$.*

This is the same to say that $X_1, ..., X_m$ and $\hat{X}_1, ..., \hat{X}_m$ define the same notion of order at p.

2: Privileged coordinates

As a first step, let us introduce the following notation:

$$\mathscr{D}^s := \mathrm{span}(\{X_I \,:\, |I| \leq s\})$$

Since we are still assuming the bracket-generating hypothesis, the values of these sets at p form a flag of subspaces of T_pM, that is

$$\mathscr{D}_p^1 \subset \mathscr{D}_p^2 \subset ... \subset \mathscr{D}_p^{r(p)} = T_pM \tag{7.2.1}$$

The integer $r(p)$ is the smallest one the flag stops at and it is called the *degree of non-holonomy at p* or also *step at p*. Set $n_i(p) := \dim \mathscr{D}^i_p$; the integer list $(n_1(p), ..., n_{r(p)}(p))$ is called the *growth vector at p*. Notice that the set $\mathscr{D}^i$ is a distribution if and only if $n_i(\cdot)$ is constant on M. We then distinguish two kinds of points.

Definition 7.2.6. *The point p is a* regular point *if the growth vector is constant in a neighbourhood of p. Otherwise, p is a singular point.*

Thus, near a regular point, all sets $\mathscr{D}^i$ are locally distributions.

Example 7.2.7. Let us consider $M = \mathbb{R}^3$ and the vector fields $X_1 = \partial_x$, $X_2 = \partial_y + \frac{x^2}{2}\partial_z$. As

$$[X_1, X_2] = x\partial_z, \quad [X_1, [X_1, X_2]] = \partial_z$$

then the growth vector is (2,3) if $x \neq 0$ and (2,2,3) otherwise. Hence, all the points with $x = 0$ are singular. $\Diamond$

Two important remarks are the following:

- the regular points form an open and dense subset of M; the singular ones form then a closed set with empty interior, but nevertheless they may have positive measure;

- at a regular point, the growth vector is a strictly increasing sequence, because otherwise if $n_i(q) = n_{i+1}(q)$ in a neighbourhood of p, then $\mathscr{D}^i$ would be an involutive distribution and so $i = r(p) < n$, which contradicts the bracket-generating assumption.

A different way to represent the flag (7.2.1) is the following: set $w_1 = ... = w_{n_1} = 1$, $w_{n_1+1} = ... = w_{n_2} = 2$ until $w_{n_{r-1}+1} = ... = w_{n_r} = r$, where $r = r(p)$. In this way, $w(p) \in \mathbb{R}^n$ and the integers $w_i = w_i(p)$ for $i = 1, ..., n$ are called the *weights at p*. The meaning of this sequence is best understood in terms of basis of T_pM. Choose first vector fields $Y_1, .., Y_{n_1}$ in $\mathscr{D}^1$ whose values at p form a basis of $\mathscr{D}^1_p$ (for instance $Y_i = X_i$); then choose vector fields $Y_{n_1+1}, ..., Y_{n_2}$ in $\mathscr{D}^2$ such that $Y_1(p), ..., Y_{n_2}(p)$ form a basis of $\mathscr{D}^2_p$ (for instance $Y_i = X_{I_i}$ with $|I_i| = 2$ for $i = n_1+1, ..., n_2$). For each s, choose $Y_{n_{s-1}+1}, ..., Y_{n_s}$ in $\mathscr{D}^s$ such that $Y_1, ..., Y_{n_s}$ form a basis of $\mathscr{D}^s_p$ (for instance $Y_i = X_{I_i}$ with $|I_i| = s$ for $i = n_{s-1}+1, ..., n_s$. In this way we have a sequence of vector fields $Y_1, ..., Y_n$ whose values at p form a basis of T_pM and such that

$$Y_i \in \mathscr{D}^{w_i}, \quad i = 1, ..., n$$

Such a sequence of vector fields is called an *adapted frame at p*. Let us relate now the weights to the orders. Write first the tangent space as a direct sum

$$T_pM = \mathscr{D}^1_p \oplus \mathscr{D}^2_p/\mathscr{D}^1_p \oplus ... \oplus \mathscr{D}^r_p/\mathscr{D}^{r-1}_p$$

where $\mathscr{D}_p^i/\mathscr{D}_p^{i-1}$ denotes a supplementary of $\mathscr{D}_p^{i-1}$ in $\mathscr{D}_p^i$ and take a local system of coordinates $(y_1, ..., y_n)$. The dimension of each $\mathscr{D}_p^i/\mathscr{D}_p^{i-1}$ is equal to $n_i - n_{i-1}$, so we can assume that, up to a reordering, $dy_j(\mathscr{D}_p^i/\mathscr{D}_p^{i-1}) \neq 0$ for $n_{i-1} < j \leq n_i$. Thus, for $0 < j \leq n_1$ we have $dy_j(\mathscr{D}_p^1) \neq 0$; therefore there exists X_k such that $dy_j(X_k(p)) \neq 0$. Since $dy_j(X_k) = X_k y_j$ is a first order non-holonomic derivative of y_j, we have $\mathrm{ord}_p(y_j) \leq 1 = w_j$. In the same way, for $n_{i-1} < j \leq n_i$ there exists a multi-index I of length i such that $dy_j(X_I(p)) = (X_I y_j)(p) \neq 0$ and so $\mathrm{ord}_p(y_j) \leq w_j$. Summing up, for any system of local coordinates $(y_1, ..., y_n)$ we have, up to a reordering, $\mathrm{ord}_p(y_j) \leq w_j$ or, without reordering,

$$\sum_{i=1}^{n} \mathrm{ord}_p(y_i) \leq \sum_{i=1}^{n} w_i(p) = Q(p)$$

and we will see later that $Q(p)$ is the Hausdorff dimension of the manifold. These considerations motivate then the next definition.

Definition 7.2.8. *A* system of privileged coordinates at p *is a system of local coordinates* $(z_1, ..., z_n)$ *such that* $\mathrm{ord}_p(z_j) = w_j$ *for* $j = 1, ..., n$.

Lemma 7.2.9. *There exists at least a system of privileged coordinates, hence the definition above is not empty.*

An example of privileged coordinates is provided by the so-called *exponential coordinates*. Choose an adapted frame $Y_1, ..., Y_n$ at p; then the inverse of the local diffeomorphism

$$(z_1, ..., z_n) \mapsto e^{z_1 Y_1 + ... + z_n Y_n}(p) \tag{7.2.2}$$

defines a system of local privileged coordinates at p, called *canonical coordinates of the first kind*; the inverse of the local diffeomorphism

$$(z_1, ..., z_n) \mapsto e^{z_n Y_n} \circ ... \circ e^{z_1 Y_1}(p)$$

also defines privileged coordinates at p, called *canonical coordinates of the second kind*. They are easier to work with than the ones of the first kind, since for instance in these coordinates the vector field Y_n reads as ∂_{z_n}. One can also exchange the order of the flows in the definition to obtain any of the Y_i as ∂_{z_i}. In addition any mix between first and second kind canonical coordinates is still a system of privileged coordinates, that is the inverse of the following local diffeomorphism defines a system of privileged coordinates.

$$(z_1, ..., z_n) \mapsto e^{z_n Y_n + ... + z_{s+1} Y_{s+1}} \circ e^{z_s Y_s} \circ ... \circ e^{z_1 Y_1}(p)$$

An explicit (algebraic) construction can be found in the work of Bellaïche and Risler: in their book [5], they show how to get a system of privileged coordinates from an arbitrary system of coordinates. However, if we start with privileged coordinates, the procedure they describe provides different privileged coordinates. They are the same if we suppose to start with nilpotent privileged coordinates (the notion will be given in the next section).

After this digression, let us better explain why privileged coordinates are so important: they are an essential tool to compute non-holonomic orders, characterize first-order approximations and estimate the distance. Indeed, as a first remark they satisfy the following conditions

$$dz_i(\mathscr{D}_p^{w_i}) \neq 0,\ dz_i(\mathscr{D}_p^{w_i-1}) = 0, \quad i = 1, ..., n$$

meaning that $\partial_{z_i}|_p \in \mathscr{D}_p^{w_i} \setminus \mathscr{D}_p^{w_i-1}$. Furthermore, given a system of privileged coordinates $(z_1, ..., z_n)$ and a sequence of integers $\alpha = (\alpha_1, ..., \alpha_n)$, we can define the weighted degree of the monomial $z^\alpha := z_1^{\alpha_1} ... z_n^{\alpha_n}$ as

$$\mathrm{ord}_p(z^\alpha) := \langle w, \alpha \rangle = w_1\alpha_1 + ... + w_n\alpha_n$$

and in a similar fashion we can define the weighted degree of the monomial vector field $z^\alpha \partial_{z_j}$ as

$$\mathrm{ord}_p(z^\alpha \partial_{z_j}) := \langle w, \alpha \rangle - w_j$$

By means of weighted degrees, orders of functions and vector fields can be computed in a purely algebraic way, because if f and X are a smooth function and a vector field respectively with Taylor expansions

$$f(z) \sim \sum_\alpha a_\alpha z^\alpha, \quad X(z) \sim \sum_{\alpha, j} a_{\alpha,j} z^\alpha \partial_{z_j}$$

then their orders are given by

$$\mathrm{ord}_p(f) = \min\{\langle w, \alpha \rangle \ : \ a_\alpha \neq 0\}$$
$$\mathrm{ord}_p(X) = \min\{\langle w, \alpha \rangle - w_j \ : \ a_{\alpha,j} \neq 0\}$$

Furthermore, if we introduce the function

$$\|z\|_p := |z_1|^{1/w_1} + ... + |z_n|^{1/w_n}$$

which is a pseudo-norm near p, then $\|z\|_p \leq C d(p, q^z)$ (q^z is the point of coordinates z), because by definition of orders we have $z_i = O(d(p, q^z)^{w_i})$ and so $\|z\|_p = O(d(p, q^z))$. Finally, privileged coordinates allow us to introduce a notion of homogeneity in a natural way. Define first the 1-parameter group of dilations

$$\delta_\lambda : (z_1, ..., z_n) \mapsto (\lambda^{w_1} z_1, ..., \lambda^{w_n} z_n)$$

The dilation δ_λ acts on functions and vector fields by pull-back: $\delta_\lambda^* f = f \circ \delta_\lambda$ and $\delta_\lambda^* X = X \circ \delta_\lambda$, so that $(\delta_\lambda^* X)(\delta_\lambda^* f) = \delta_\lambda^*(Xf)$.

Definition 7.2.10. *A function f is* homogeneous of degree s *if $f \circ \delta_\lambda = \lambda^s f$; the same definition fits for vector fields.*

Notice that for a smooth function, the definition above can be equivalently restated as

$$f(z) = \sum_{\langle w, \alpha \rangle = s} a_\alpha z^\alpha$$

Furthermore, the pseudo-norm $\|z\|_p$ is homogeneous oF degree 1.

3: Nilpotent approximations

Fix now a system of privileged coordinates $(z_1, ..., z_n)$ at p. We already know that each vector field X_i in $X_1, ..., X_m$ is of order greater than or equal to -1; moreover, for at least one coordinate z_j among $z_1, ..., z_m$ the derivative $(X_i z_j)(p)$ is nonzero as $dz_j(\mathscr{D}_p^1) \neq 0$. Hence $X_1, ..., X_m$ are of order -1. In z coordinates, X_i has a Taylor expansion

$$X_i(z) \sim \sum_{\alpha,j} a_{\alpha,j} z^\alpha \partial_{z_j}$$

and thanks to the remark above we can rewrite this series as

$$X_i = X_i^{(-1)} + X_i^{(0)} + X_i^{(1)} + ...$$

where $X_i^{(s)}$ is a homogeneous vector field of degree s. Set then $\hat{X}_i := X_i^{(-1)}$, that is

$$\hat{X}_i = \sum_{j=1}^{m} \sum_{\langle w,\alpha\rangle - w_j = -1} a_{\alpha,j} z^\alpha \partial_{z_j}$$

By construction, $\hat{X}_1, ..., \hat{X}_m$ is a first-order approximation of $X_1, ..., X_m$ at p, but there are two other nice properties:

- $\mathfrak{Lie}(\hat{X}_1, ..., \hat{X}_m)$ is a nilpotent Lie algebra of step $r = w_n$, that is every Lie bracket between more than r vector fields among $\hat{X}_1, ..., \hat{X}_m$ is zero; indeed, if X, Y are homogeneous vector fields of degree k, l respectively, then $[X, Y]$ is still homogeneous of degree $k + l$;

- $\hat{X}_1, ..., \hat{X}_m$ are bracket-generating on $\mathbb{R}^n$, hence they define a sub-Riemannian structure.

The second property motivates the following definition.

Definition 7.2.11. *The sub-Riemannian manifold* $(\mathbb{R}^n, \hat{X}_1, ..., \hat{X}_m)$ *is the* (homogeneous) nilpotent approximation *of* $(M, X_1, ..., X_m)$ *at* p *associated to the coordinates* z.

The nilpotent approximation is not unique and it is not intrinsic at all, since it depends on the chosen system of privileged coordinates. Nevertheless, if $\hat{X}_1, ..., \hat{X}_m$ and $\hat{X}'_1, ..., \hat{X}'_m$ are the nilpotent approximations associated to two different systems of coordinates, the corresponding Lie algebras $\mathfrak{Lie}(\hat{X}_1, ..., \hat{X}_m)$ and $\mathfrak{Lie}(\hat{X}'_1, ..., \hat{X}'_m)$ are isomorphic. If moreover p is a regular point, then $\mathfrak{Lie}(\hat{X}_1, ..., \hat{X}_m)$ is isomorphic to the graded nilpotent Lie algebra

$$\mathfrak{gr}_p = \mathscr{D}_p^1 \oplus \mathscr{D}_p^2/\mathscr{D}_p^1 \oplus ... \oplus \mathscr{D}_p^r/\mathscr{D}_p^{r-1}$$

which is equipped with a weighted Lie bracket defined as follows: for any $v \in \mathscr{D}_p^k/\mathscr{D}_p^{k-1}$ and $w \in \mathscr{D}_p^s/\mathscr{D}_p^{s-1}$, $[v,w] := [V,W]$, where V, W are such that

$$V \in \mathscr{D}^k,\, V(p) = v, \quad W \in \mathscr{D}^s,\, W(p) = w$$

As $\mathfrak{gr}_p$ has a proper Lie algebra structure, we can consider the Lie group associated to it: $\mathrm{Gr}_p := \exp(\mathfrak{gr}_p)$. Such Lie group is simply connected and it is then a Carnot group. Notice that the vector fields $\hat{X}_1, ..., \hat{X}_m$ can be seen in a threefold way: as vector fields on $\mathbb{R}^n$, as representations in z coordinates of the vector fields $z^*\hat{X}_1, ..., z^*\hat{X}_m$ or as vector fields on Gr_p. In this last meaning, $\hat{X}_1, ..., \hat{X}_m$ generate a left-invariant distribution $\mathscr{D}_G$ on Gr_p with associated sub-Riemannian metric g_G and sub-Riemannian distance d_G. As we will better see in Lecture 4, we then have

$$(\mathrm{Gr}_p, \mathscr{D}_G, g_G) \cong (\mathbb{R}^n, \hat{X}_1, ..., \hat{X}_m)$$

and therefore $(\mathbb{R}^n, \hat{d})$ is a Carnot-Carathéodory space, where $\hat{d}$ is the sub-Riemannian distance associated with $\hat{X}_1, ..., \hat{X}_m$. By now, let us state the following result.

Lemma 7.2.12. *The following facts hold:*

(i) $\hat{\mathscr{D}}$ *is a bracket-generating distribution on* $\mathbb{R}^n$*;*

(ii) the distance $\hat{d}$ *is homogeneous of degree 1;*

(iii) there exists a constant $C > 0$ *such that, for all* $z \in \mathbb{R}^n$*,*

$$\frac{1}{C}\|z\|_p \le \hat{d}(0,z) \le C\|z\|_p$$

In particular, property (ii) is due to the fact that if γ is a trajectory of the control system $\dot{z} = \sum_{i=1}^m u_i \hat{X}_i(z)$, then $\delta_\lambda \gamma$ is a trajectory of the control system with controls λu_i.

For sake of completeness, an intrinsic definition of nilpotent approximation has been recently proposed by Agrachev and Marigo (algebraic approach) and also by Falbel and Jean (metric approach). In the first one it is possible to generalize to singolar points the procedure we describe above for the regular ones.

7.3 Lecture 3 - 20 October

What does it really mean to be a nilpotent approximation? In this lecture we will see the role of the sub-Riemannian manifold $(\mathbb{R}^n, \hat{X}_1, ..., \hat{X}_m)$, but before entering into the details, let us begin with the following example.

Example 7.3.1. Let $M = \mathbb{R}^2 \times \mathbb{S}^1$ with coordinates (x, y, ϑ): this manifold represents the human locomotion, since coordinates x, y describe the position of a person on the plane, whereas the angle ϑ denotes the direction such person is looking at. In this framework, let us consider the following vector fields

$$X_1 = \cos\vartheta\partial_x + \sin\vartheta\partial_y, \quad X_2 = \partial_\vartheta$$

and observe that

$$[X_1, X_2] = -\sin\vartheta\partial_x + \cos\vartheta\partial_y$$

At $q = 0$ the growth vector is given by (2,3) and the weight vector by (1,1,2). Notice in addition that $X_1(0) = \partial_x$, $X_2(0) = \partial_\vartheta$ and $[X_1, X_2](0) = \partial_y$, so that

$$\mathrm{ord}_0(x) = 1, \quad \mathrm{ord}_0(\vartheta) = 1, \quad \mathrm{ord}_0(y) = 2$$

As a consequence, (x, ϑ, y) are privileged coordinates at $q = 0$. Following the procedure described in the previous lecture, let us compute the nilpotent approximation $(\mathbb{R}^3, \hat{X}_1, \hat{X}_2)$ of (M, X_1, X_2). In order to do it, we have to write X_1, X_2 in Taylor series and take the homogeneous parts of order -1; for X_2 there is nothing to do, as it is already homogeneous of degree -1 and so $\hat{X}_2 = X_2$. For the other vector field, observe that

$$\begin{aligned} X_1 &= \left(1 - \frac{\vartheta^2}{2} + ...\right)\partial_x + \left(\vartheta - \frac{\vartheta^3}{3!} + ...\right)\partial_y \\ &= (\partial_x + \vartheta\partial_y) + \text{higher order terms} \end{aligned}$$

In conclusion, the nilpotent approximation is described by

$$\hat{X}_1 = \partial_x + \vartheta\partial_y, \quad \hat{X}_2 = \partial_x$$

and these vector fields are the same generating the Heisenberg group $\mathbb{H}^1$. Hence $\mathbb{H}^1$ is the nilpotent approximation of (M, X_1, X_2). $\Diamond$

Let us now investigate the metric meaning of nilpotent approximations.

Applications to distance estimates

Fix $p \in M$, a system of privileged coordinates z at p and vector fields $\hat{X}_1, ..., \hat{X}_m$. From now on we will identify a neighbourhood of p in M with the corresponding (via z) neighbourhood of 0 in $\mathbb{R}^n$, that is we will work directly in $\mathbb{R}^n$. Thus, let z^0 be near 0 and let $u \in L^1$ be a control. We can then consider two kinds of trajectories γ_u and $\hat{\gamma}_u$, solutions of the following control systems respectively

$$\dot{z} = \sum_{i=1}^{m} u_i X_i(z) \tag{7.3.1}$$

$$\dot{z} = \sum_{i=1}^{m} u_i \hat{X}_i(z)$$

and with initial condition z^0. Up to reparametrization, we can assume $|u(t)| \leq 1$ a.e. so that both $\ell(\gamma_u([0,t])) \leq t$ and $\ell(\hat{\gamma}_u([0,t])) \leq t$. Aim of the following result is to express how much $\hat{\gamma}_u$ differs from γ_u.

Lemma 7.3.2. *There exist constant C and $\varepsilon > 0$ such that, for any $z^0 \in \mathbb{R}^n$ and $t \geq 0$ with $\tau := \max\{\|z^0\|_p, t\} < \varepsilon$, we have*

$$\|\gamma_u(t) - \hat{\gamma}_u(t)\|_p \leq C\tau t^{1/r}$$

Proof. The first step is to prove that there exists a constant c such that $\|\gamma_u(t)\|_p$ and $\|\hat{\gamma}_u(t)\|_p$ are both bounded by $c\tau$ for τ small enough. Let us do it for $\gamma_u(t)$ (the proof is exactly the same for $\hat{\gamma}_u(t)$. First of all, γ_u is a solution of the ODE (7.3.1), which reads in coordinates as follows

$$\dot{z}_j = \sum_{i=1}^{m} u_i X_i^j(z), \quad j = 1, ..., n$$

By assumption $|u_i| \leq 1$; in addition, X_i^j is of order greater than or equal to $w_j - 1$, hence there exists a constant c' such that, when $\|z\|_p$ is small enough, $|X_i^j(z)| \leq c'\|z\|_p^{w_j-1}$ for any $j = 1, ..., n$ and $i = 1, ..., m$. From these considerations we infer that

$$|\dot{z}_j| \leq c' m \|z\|_p^{w_j-1} \tag{7.3.2}$$

Now let us go on with a formal computation, namely

$$\frac{d}{dt}\|z\|_p \leq c'' \sum_{j=1}^{n} |z_j|^{\frac{1}{w_j}-1} |\dot{z}_j|$$

Plugging (7.3.2) into this inequality and observing that

$$\sum_{j=1}^{n} |z_j|^{\frac{1}{w_j}-1} = \|z\|_p^{1-w_j}$$

we finally get $\frac{d}{dt}\|z\|_p \leq \tilde{c}$. By integration, this inequality provides us

$$\|\gamma_u(t)\|_p \leq \tilde{c}t + \|z^0\|_p \leq c\tau$$

and so the first step is accomplished. For the second step, let us show that

$$|z_j(t) - \hat{z}_j(t)| \leq c\tau^{w_j} t \tag{7.3.3}$$

for a suitable constant c. To this aim, notice that the function $z_j - \hat{z}_j$ satisfies the differential equation

$$\dot{z_j} - \dot{\hat{z}}_j = \sum_{i=1}^{m} u_i (X_i^j(z) - \hat{X}_i^j(z))$$

and $X_i^j(z) = \hat{X}_i^j(z) +$higher order terms with $\hat{X}_i^j(z) = \hat{X}_i^j(z_1, ..., z_{j-1})$, arguing on the order. Hence, the ODEs system has a sort of triangular structure, which enables an integration by induction and as a consequence (7.3.3) follows. Finally,

$$\|\gamma_u(t) - \hat{\gamma}_u(t)\|_p \leq C'\tau(t^{1/w_1} + ... + t^{1/w_n}) \leq C\tau t^{1/r}$$

which completes the proof of the lemma. □

As a direct consequence, we are now able to give estimates of the sub-Riemannian distance in terms of privileged coordinates.

Theorem 7.3.3. *A coordinate system $(z_1, ..., z_n)$ is privileged at p if and only if there exists constants C_p and $\varepsilon_p > 0$ such that, if $\|z\|_p < \varepsilon_p$, then*

$$\frac{1}{C_p}\|z\|_p \leq d(0, z) \leq C_p\|z\|_p$$

(recall that $z(p) = 0$).

Proof. Observe first that, by definition of order, a coordinate system z is privileged if and only if $d(0, z) \geq c_p\|z\|_p$. Hence, the only thing to prove is the following: if z are privileged coordinates, then $d(0, z) \leq C_p\|z\|_p$. To this aim, we will show that, for $\|z\|_p$ small enough,

$$d(0, z) \leq 2\hat{d}(0, z)$$

so that $d(0, z) \leq C\|z\|_p$ follows by the previous lemma.

Fix $z \in \mathbb{R}^n$ with $\|z\|_p < \varepsilon$. Let $\hat{\gamma}_{u,1}(t)$ for $t \in [0, T_1]$ be a minimizing curve for $\hat{d}$ joining z to 0 with velocity one (thus, $T_1 = \hat{d}(0, z) = \ell(\gamma_{u,1})$); let $\gamma_{u,1}(t)$ for $t \in [0, T_1]$ be the trajectory of the control system (7.3.1) starting at z. Set $z^2 := \gamma_{u,1}(T_1)$. By Lemma 7.3.2 we have

$$\|z^2\|_p = \|\gamma_{u,1}(T_1) - \hat{\gamma}_{u,1}(T_1)\|_p \leq C\tau T_1^{1/r}$$

where $\tau = \max\{\|z\|_p, T_1\}$ and by Lemma 7.2.12 (iii) $T_1 = \hat{d}(0, z)$ satisfies $T_1 \geq \|z\|_p/C'$, so that $\tau \leq C'T_1$ and

$$\hat{d}(0, z^2) \leq C'\|z^2\|_p \leq C''\hat{d}(0, z)^{1+1/r}$$

with $C'' = C'^2C$. Let now $\hat{\gamma}_{u,2}(t)$ for $t \in [0, T_2]$ be a minimizing curve for $\hat{d}$ with velocity one joining z^2 to 0 and let $\gamma_{u,2}(t)$ for $t \in [0, T_2]$ be the trajectory of the control system (7.3.1) starting at z^2; set $z^3 := \gamma_{u,2}(T_2)$. As previously, $\ell(\gamma_{u,2}) = \hat{d}(0, z^2)$ and $\hat{d}(0, z^3) \leq C''\hat{d}(0, z^2)^{1+1/r}$. In this way we get a sequence $(z^k)_{k\geq 1}$ of points (with $z^1 := z$) such that $\hat{d}(0, z^{k+1}) \leq C''\hat{d}(0, z^k)^{1+1/r}$ and a sequence of trajectories $\gamma_{u,k}$ of the control system (7.3.1) joining z^k to z^{k+1} of length equal to $\hat{d}(0, z^k)$. By taking $\|z\|_p$ small enough we can assume $C''\hat{d}(0, z)^{1/r} \leq 1/2$, so that

$$\hat{d}(0, z^{k+1}) \leq \frac{1}{2^k}\hat{d}(0, z)$$

As a consequence, $z^k \to 0$ and gluing together the curves $\gamma_{u,k}$ we get a curve joining z to 0 which is still a trajectory of the control system (7.3.1) and has length smaller than or equal to $2\hat{d}(0,z)$. This implies $d(0,z) \leq 2\hat{d}(0,z)$ and so the theorem. □

Corollary 7.3.4 (Ball-Box Theorem)**.** *Expressed in a given system of privileged coordinates, the sub-Riemannian balls* $B(p,\varepsilon)$ *satisfy, for* $\varepsilon < \varepsilon_p$,

$$\mathrm{Box}(C_p^{-1}\varepsilon) \subset B(p,\varepsilon) \subset \mathrm{Box}(C_p\varepsilon)$$

where

$$\mathrm{Box}(\varepsilon) := [-\varepsilon^{w_1}, \varepsilon^{w_1}] \times ... \times [-\varepsilon^{w_n}, \varepsilon^{w_n}]$$

Corollary 7.3.5. *Near* p *the distances* d *and* $\hat{d}$ *are related by*

$$d(p,q) = \hat{d}(p,q)\Big(1 + O(\hat{d}(p,q))\Big)$$

Actually, it is possible to prove that near p a more refined formula holds, namely

$$|d(q,q') - \hat{d}(q,q')| \leq C(\hat{d}(p,q) + \hat{d}(p,q'))d(q,q')^{1/r} \tag{7.3.4}$$

but it requires a lot of work. It was proved by Bellaïche.

Metric tangent space to a Carnot-Carathéodory space

In this section we will see that a notion of tangent space can be defined for a general metric space. Indeed, in describing the tangent space to a manifold we usually imagine looking at smaller and smaller neighbourhoods of a given point, the manifold being fixed (in order to introduce the notion of germs). Equivalently, we can imagine looking at a fixed neighbourhood, but expanding the manifold and, as noticed by Gromov, this idea can be used in a purely metric setting. Before we start, a comment: recall that by Carnot-Carathéodory space we simply mean a sub-Riemannian manifold seen as metric space.

Let us now introduce a bit of notation. If $X = (M,d)$ is a metric space, then we define λX for $\lambda > 0$ as $(M, \lambda d)$, i.e. we dilate the distance. A *pointed metric space* (X,p) is a metric space with a distinguished point p. Then, a still rough definition of metric tangent space is the following.

Definition 7.3.6. *The* metric tangent space *to the metric space* X *at* p *is defined as*

$$C_pX := \lim_{\lambda\to+\infty}(\lambda X, p) \tag{7.3.5}$$

provided the limit exists. When it exists, the metric tangent space is a pointed metric space.

In order to give a meaning to this definition, let us first introduce Hausdorff distance between two sets as follows: let $A, B \subset X$ and, for any $\delta > 0$, let $N_\delta(A)$ and $N_\delta(B)$ be the δ-enlarged sets of A and B respectively; then

$$d_H(A,B) := \inf\{\delta \,:\, B \subset N_\delta(A),\, A \subset N_\delta(B)\} \tag{7.3.6}$$

However, in order to give a meaning to the limit of pointed metric spaces, we have to define a distance between metric spaces and not only between subsets of the same metric space. Actually, the distance we are looking for is a generalization of the Hausdorff one and it is the Gromov-Hausdorff distance. For two metric spaces X, Y, it is defined as

$$d_{GH}(X, Y) := \inf\{d_H(i(X), j(Y))\}$$

where the infimum is taken among all metric spaces Z and all isometric embeddings $i : X \to Z$ and $j : Y \to Z$. If X, Y are compact metric spaces and $d_{GH}(X, Y) = 0$, then the two spaces are isometric (hence, d_{GH} is a distance on isometry classes); roughly speaking, when the Gromov-Hausdorff distance between two spaces is small, then such spaces are almost isometric. Moreover, thanks to Gromov-Hausdorff distance, we can define the notion of limit of a sequence of pointed metric spaces: (X_n, p_n) converges to (X, P) if

$$\lim_{n\to\infty} d_{GH}(B^{X_n}(p_n, R), B^X(p, R)) = 0, \quad \forall R > 0$$

where $B^{X_n}(p_n, R)$ and $B^X(p, R)$ are seen as metric spaces. Notice that the limit of a sequence of metric spaces is unique provided the closed balls around the distinguished point p are compact. As shown by the next example, the notion of metric tangent space we introduced is a generalization of the usual notion of tangent space in a smooth setting.

Example 7.3.7. Let (M, g) be a Riemannian manifold with distance d_R. In this case the metric tangent space at a point p exists and is the Euclidean tangent space (T_pM, g_p), that is its standard tangent space endowed with the scalar product defined by g_p. $\diamondsuit$

For a sub-Riemannian manifold, the metric tangent space is given by the nilpotent approximation, as shown in the next theorem. Mitchell proposed firstly a proof in the 80's but it was wrong; the correct proof arrived in 1996 thanks to Bellaïche.

Theorem 7.3.8 (Bellaïche-Mitchell)**.** *A Carnot-Carathéodory space (M, d) admits a unique (up to isometries) metric tangent space $(C_pM, 0)$ at every point $p \in M$. The space C_pM is a sub-Riemannian manifold isometric to $(\mathbb{R}^n, \hat{d}_p)$, where $\hat{d}_p$ is the sub-Riemannian distance associated to a homogeneous nilpotent approximation at p.*

Proof. Let us take $\lambda = 1/\varepsilon$ in (7.3.5), $X = (\mathbb{R}^n, \hat{d})$ and $X_\varepsilon = (M, \varepsilon^{-1}d)$; the thesis will follow if we are able to prove that

$$\lim_{n\to\infty} d_{GH}(B^{X_\varepsilon}(p, R), B^X(0, R)) = 0, \quad \forall R > 0 \tag{7.3.7}$$

To this aim, let us introduce the dilation $\delta_\varepsilon : \hat{B}(0, R) \to \mathbb{R}^n$, where $\hat{B}(0, R)$ is the ball of $\hat{d}$-radius R centered at 0, and claim that for a suitable function g

continuous at 0 such that $g(0) = 0$ it holds

$$\left| \frac{1}{\varepsilon} d(\delta_\varepsilon x, \delta_\varepsilon y) - \hat{d}(x, y) \right| \le g(\varepsilon) \tag{7.3.8}$$

for every $x, y \in \hat{B}(0, R)$. In [17] Montgomery proved that this fact entails (7.3.7) and so the conclusion, but now let us prove the claim; this can be done in two different ways:

- the first one has been proposed by Bellaïche and relies on (7.3.4);
- the second one is due to Agrachev-Barilari-Boscain and Ambrosio-Ghezzi-Magnani.

We will perform the second procedure, observing first that

$$d_\varepsilon(x, y) := \frac{1}{\varepsilon} d(\delta_\varepsilon x, \delta_\varepsilon y)$$

defines a sub-Riemannian distance associated to the vector fields $X_i^\varepsilon := \varepsilon \delta_{1/\varepsilon}^* X_i$. Secondly, (7.3.8) is equivalent to the fact that $d_\varepsilon \to \hat{d}$ uniformly on $\hat{B}(0, R) \times \hat{B}(0, R)$. In order to prove the uniform convergence, we will adopt a sort of Ascoli-Arzelà argument, showing that:

(i) $\{d_\varepsilon\}_{\varepsilon > 0}$ is equicontinuous;

(ii) pointwise convergence: $d_\varepsilon(x, y) \to \hat{d}(x, y)$ for every $x, y \in \hat{B}(0, R)$.

For property (i), notice that we have already proved it in Lecture 1, since it is a consequence of Chow-Rashevskii's theorem. Indeed, we know that

$$d(x, y) \le C|x - y|^{1/r} \quad \text{on } \hat{B}(0, R) \tag{7.3.9}$$

for any Carnot-Carathéodory distance, hence in particular also for d_ε, and it is not difficult to prove that C can be chosen independent of ε, provided that ε is small enough.

For (ii), we need the following two remarks. First of all, (7.1.3) can actually be strengthened as follows: if γ is a trajectory of finite length, then there exists a control $u \in L^2([0, 1], \mathbb{R}^m)$ such that $\ell(\gamma) = \|u\|_{L^2}$. Secondly, near p there always exist length-minimizing curves, so that, together with the previous remark, for any q, q' in a suitable neighbourhood of p we have

$$d(q, q') = \min\{\|u\|_{L^2}\}$$

where the minimum is taken among all the controls u associated to any curve γ joining q to q' in time 1, that is $\gamma : [0, 1] \to M$ with $\gamma(0) = q$ and $\gamma(1) = q'$. Now we claim that $X_i^\varepsilon \to \hat{X}_i$ uniformly on compact sets; to this aim notice that

$$X_i(z) = \hat{X}_i(z) + \sum_{j=1}^{m} R_{ij}(z) \partial_{z_j}$$

with

$$\mathrm{ord}_0(R_{ij}) \geq w_j \tag{7.3.10}$$

and

$$X_i^\varepsilon(z) = \hat{X}_i(z) + \varepsilon \sum_{j=1}^m R_{ij}(\delta_\varepsilon z)\delta_{1/\varepsilon}^* \partial_{z_j} = \hat{X}_i(z) + \varepsilon \sum_{j=1}^m \frac{R_{ij}(\delta_\varepsilon z)}{\varepsilon^{w_j}} \partial_{z_j}$$

Thanks to (7.3.10), each term $R_{ij}(\delta_\varepsilon z)/\varepsilon^{w_j}$ is bounded on every compact set by definition of order; thus the claim is proved. As a consequence, if $(u^\varepsilon)_{\varepsilon>0} \subset L^2([0,1], \mathbb{R}^m)$ admits a weak limit u for $\varepsilon \downarrow 0$, the solutions γ^ε to

$$\begin{cases} \dot{q} = \sum_{i=1}^m u_i^\varepsilon X_i^\varepsilon(q) \\ q(0) = x \end{cases} \tag{7.3.11}$$

for a fixed initial condition x uniformly converge to $\hat{\gamma}$, solution to

$$\begin{cases} \dot{q} = \sum_{i=1}^m u_i \hat{X}_i(q) \\ q(0) = x \end{cases} \tag{7.3.12}$$

Now fix x, y and apply the argument above to $u^\varepsilon \equiv u$ for any $\varepsilon > 0$ with $\|u\|_{L^2} = \hat{d}(x,y)$ (the existence of such a function u is guaranteed by the first remark above). Let γ_u^ε be the solutions to (7.3.11) and $\hat{\gamma}_u$ be the solution to (7.3.12), connecting x to y and minimizing the length. Then notice that

$$\limsup_{\varepsilon\downarrow 0} d_\varepsilon(x,y) \leq \limsup_{\varepsilon\downarrow 0} d_\varepsilon(x, \gamma_u^\varepsilon(1)) + \limsup_{\varepsilon\downarrow 0} d_\varepsilon(\gamma_u^\varepsilon(1), y)$$

Since $\gamma_u^\varepsilon(1) \to y$ by the uniform convergence of γ_u^ε to $\hat{\gamma}_u$ and $d_\varepsilon(x, \gamma_u^\varepsilon(1)) \leq \hat{d}(x,y)$ by construction, we infer that

$$\limsup_{\varepsilon\downarrow 0} d_\varepsilon(x,y) \leq \hat{d}(x,y)$$

For the liminf inequality, from (7.3.9) we deduce that $\liminf_{\varepsilon\downarrow 0} d_\varepsilon(x,y)$ is finite. Following a minimizing sequence argument, take u^n such that $\|u^n\|_{L^2} = d_{\varepsilon_n}(x,y)$ with $\varepsilon_n \downarrow 0$ and

$$d_{\varepsilon_n}(x,y) \to \liminf_{\varepsilon\downarrow 0} d_\varepsilon(x,y)$$

Up to extract a suitable subsequence, we can assume that u^n weakly converges in $L^2([0,1], \mathbb{R}^m)$ to a limit function u. Therefore,

$$\begin{aligned} \hat{d}(x,y) &\leq \|u\|_{L^2} \leq \liminf_{n\to\infty} \|u^n\|_{L^2} = \liminf_{n\to\infty} d_{\varepsilon_n}(x,y) \\ &= \liminf_{\varepsilon\downarrow 0} d_\varepsilon(x,y) \end{aligned}$$

Coupling this inequality with the limsup one we got above, we infer property (ii) and thus the thesis follows. □

7.4 Lecture 4 - 21 October

In this lecture we are going to complete the description of the metric tangent space associated to a Carnot-Carathéodory space and we are going to investigate its algebraic properties. First of all, as seen in Theorem 7.3.8 the metric tangent space C_pM to a Carnot-Carathéodory space (M, d) at any point p is isomorphic to $(\mathbb{R}^n, \hat{d}_p)$. However, in general this is not a linear space, as $\hat{d}_p$ is homogeneous of degree 1 but with respect to dilations δ_λ and not to the usual Euclidean ones. In general $(\mathbb{R}^n, \hat{d}_p)$ can be seen as a quotient of groups; if we further assume p to be regular, then, as we are going to see, $(\mathbb{R}^n, \hat{d}_p)$ has a natural intrinsic structure of Carnot group, but this is not obtained in the most naïf way. Indeed, one could be tempted to consider the group generated by the flows of $X_1, ..., X_m$ acting on the left on $\mathbb{R}^n$, which will be denoted by $e^{tX_1}, ..., e^{tX_m}$. Unfortunately, $e^{tX_1}, ..., e^{tX_m}$ are only local diffeomorphisms and thus the group generated by their flows is not a Lie group. On the contrary $e^{t\hat{X}_1}, ..., e^{t\hat{X}_m}$ are global diffeomorphisms on $\mathbb{R}^n$, so that the group generated by their flows acting on the left, namely

$$G_p := \{e^{t_1\hat{X}_{i_1}} \circ ... \circ e^{t_k\hat{X}_{i_k}}(p) \,:\, k \in \mathbb{N},\, i_j \in \{1, ..., m\} \text{ for } j = 1, ..., k\}$$

turns out to be a Lie group. Since $\mathfrak{g}_p := \mathfrak{Lie}(\hat{X}_1, ..., \hat{X}_m)$ is a nilpotent Lie algebra, then G_p is a simply connected nilpotent Lie group, because $G_p = \exp(\mathfrak{g}_p)$. Furthermore, $\mathfrak{g}_p$ splits as

$$\mathfrak{g}_p = \mathfrak{g}^{(-1)} \oplus ... \oplus \mathfrak{g}^{(-r)}$$

where $r = r(p)$ is exactly the step at p and $\mathfrak{g}^{(-i)}$ is the family of homogeneous vector fields of degree $-i$. Notice that

$$\mathfrak{g}^{(-i)} = [\mathfrak{g}^{(-1)}, \mathfrak{g}^{(-s+1)}]$$

so that $\mathfrak{g}_p$ is also graded, because $\mathfrak{g}^{(-1)} = \text{span}(\hat{X}_1, ..., \hat{X}_m)$ generates $\mathfrak{g}_p$ as Lie algebra. All these properties imply that G_p is a Carnot group and this fact does not depend on the regularity of p.

Definition 7.4.1. *A* Carnot group *is a simply connected Lie group whose Lie algebra is graded, nilpotent and generated by its first component.*

Before the forthcoming discussion, let us recall what a right-invariant vector field is (of course an analogous definition is possible for left-invariance). First of all, the notion can be stated only on Lie groups, as the invariance is understood w.r.t. the derivative of the group action. Thus, if by R_g we denote the right translation by g (i.e. $R_g(h) := hg$) and by r_g its derivative, then a vector field X acting on the Lie group G is said to be *right-invariant* if

$$r_g \circ X = X \circ R_g$$

so that, in particular, it holds $X_g = r_g X_0$. To the vector fields $\hat{X}_1, ..., \hat{X}_m$ we can thus associate right-invariant vector fields $\hat{\xi}_1, ..., \hat{\xi}_m$ on G_p by putting

$$\hat{\xi}_i(g) := \frac{d}{dt} e^{t\hat{X}_i} g \Big|_{t=0} \tag{7.4.1}$$

and they define a right-invariant sub-Riemannian structure on G_p (we refer to this structure as *right-invariant* simply because it is generated by right-invariant vector fields); moreover, $(\hat{\xi}_1, ..., \hat{\xi}_m)$ is an orthonormal frame of such sub-Riemannian structure. Finally, the associated sub-Riemannian distance d_G is homogeneous (in the sense of Carnot groups, namely it respects the group law).

About the action of G_p on $\mathbb{R}^n$, it is smooth and transitive since $\hat{X}_1, ..., \hat{X}_m$ is bracket-generating. This means that the orbit of 0 under the action of G_p is the whole $\mathbb{R}^n$; recall that the orbit of a point x under the action of a group G is defined as

$$\mathrm{Orb}_x(G) := \{g(x) \,:\, g \in G\}$$

Therefore, the map $\phi_p : G_p \to \mathbb{R}^n$ defined by $\phi_p(g) := g(0)$ is surjective, but what about the injectivity? As a next step, let us introduce the isotropy subgroup of 0 as follows

$$H_p := \{g \in G_p \,:\, g(0) = 0\}$$

that is the subgroup consisting of all the elements of G_p whose action has 0 as fixed point. In general G_p/H_p is not a quotient group, since H_p need not be normal in G_p: it is only a coset space, but nevertheless it is a manifold of dimension n and the map $\psi_p : G_p/H_p \to \mathbb{R}^n$ defined by $\psi_p(gH_p) := g(0)$ is a diffeomorphism. Notice that H_p is invariant under dilations, since $\delta_t g(\delta_t x) = \delta_t(g(x))$, and therefore it is connected and simply connected; this implies that $H_p = \exp(\mathfrak{h}_p)$ where $\mathfrak{h}_p$ is the Lie sub-algebra of $\mathfrak{g}_p$ containing the vector fields vanishing at 0, i.e.

$$\mathfrak{h}_p = \{Z \in \mathfrak{g}_p \,:\, Z(0) = 0\}$$

In other words, $\mathfrak{h}_p$ is the annihilator. Furthermore, as $\mathfrak{g}_p$, $\mathfrak{h}_p$ is invariant under dilations and splits into homogeneous components. Now, similarly to what we did in (7.4.1) on G_p, on G_p/H_p we can introduce vector fields $\overline{\xi}_1, ..., \overline{\xi}_m$ by

$$\overline{\xi}_i(gH_p) := \frac{d}{dt} e^{t\hat{X}_i} gH_p \Big|_{t=0}$$

and this simply means that we are looking at $\hat{X}_1, ..., \hat{X}_m \in \mathfrak{g}_p$ acting on the left on G_p/H_p. Also in this case, such vector fields are an orthonormal basis of the sub-Riemannian structure associated to them. In addition, we have $(\psi_p)_* \overline{\xi}_i = \hat{X}_i$, so that ψ_p maps isomorphically $(G_p/H_p, \overline{\xi}_1, ..., \overline{\xi}_m)$ to $(\mathbb{R}^n, \hat{X}_1, ..., \hat{X}_m)$. The following theorem then follows.

Theorem 7.4.2. *Given a Carnot-Carathéodory space (M, d) and a point $p \in M$, the metric tangent space C_pM and $(\mathbb{R}^n, \hat{d}_p)$ are isometric to the homogeneous space $(G_p/H_p, \overline{d})$, where $\overline{d}$ is the sub-Riemannian distance associated to $\overline{\xi}_1, ..., \overline{\xi}_m$.*

Thus we have proved that the metric tangent space has a structure of coset space of Lie groups. As a corollary, the metric tangent space at a regular point has a natural structure of Carnot group, because in this case $H_p = \{e\}$.

Corollary 7.4.3. *If p is a regular point, then $H_p = \{e\}$ and the metric tangent space C_pM (and also $(\mathbb{R}^n, \hat{d}_p)$, of course) is isometric to the Carnot group (G_p, d_G), where d_G is the homogeneous distance associated to the right-invariant structure $\hat{\xi}_1, ..., \hat{\xi}_m$.*

Proof. Let $X_{I_1}, ..., X_{I_n}$ be an adapted basis at p. Since p is regular, the growth vector is constant in the near of it and $X_{I_1}, ..., X_{I_n}$ is still adapted near p. Thus, near p, if $|J| \leq r(p)$ we have also that (in local coordinates)

$$X_J(z) = \sum_{|I_i| \leq |J|} a_i(z) X_{I_i}(z)$$

with a_i of non-negative order and X_{I_i} of order greater than or equal to $-|I_i|$. As a consequence,

$$\hat{X}_J(z) = X_J^{(-|J|)} = \sum_{|I_i| = |J|} a_i(0) \hat{X}_{I_i}(z)$$

and this shows that $\hat{X}_J \in \operatorname{span}(\hat{X}_{I_1}, ..., \hat{X}_{I_n})$. Therefore, $\hat{X}_{I_1}, ..., \hat{X}_{I_n}$ is a basis of $\mathfrak{g}_p$, whence $\dim G_p = \dim \mathfrak{g}_p = n$. ☐

The situation is summarized in the following diagram.

$$\begin{array}{ccc} (G_p, d_G) & & \\ \Big\downarrow \pi & \searrow^{\,g \mapsto g(0)} & \\ (G_p/H_p, \overline{d}) & \xrightarrow{\;\cong\;} & (\mathbb{R}^n, \hat{d}) \end{array}$$

Let us conclude the discussion on the algebraic properties of the metric tangent space to a Carnot-Carathéodory space with an example. As we are going to see, the conclusion of Corollary 7.4.3 can be false, since the dimension of G_p may be strictly bigger than the dimension of the manifold we are considering.

Example 7.4.4 (Grušin plane)**.** Let us consider $\mathbb{R}^2$ and the vector fields $X_1 = \partial_x$, $X_2 = x\partial_y$. As $[X_1, X_2] = \partial_y$, the distribution is bracket-generating. Outside the line $x = 0$, the sub-Riemannian structure is actually Riemannian, but at $p = (0, y)$ the growth vector is (1,2), so that the weights at the same point are also (1,2), and

$$\operatorname{ord}_p(x) = 1, \quad \operatorname{ord}_p(y) = 2$$

This means that (x, y) are privileged coordinates at $p = (0, y)$ and the nilpotent approximation of $(\mathbb{R}^2, X_1, X_2)$ is given by $\hat{X}_1 = X_1$, $\hat{X}_2 = X_2$, since the homogeneous parts of X_1, X_2 of order -1 coincide with themselves. The tangent space at the same point is generated by $\hat{Y}_1 = X_1$ and $\hat{Y}_2 = [X_1, X_2]$, but to get all of $\mathfrak{g}_p$ we must add $\hat{Y}_3 = X_2 = x\hat{Y}_2$, because by definition

$$\mathfrak{g}_p = \operatorname{span}(X_1, X_2, [X_1, X_2]) = \operatorname{span}(\partial_x, x\partial_y, \partial_y)$$

Notice that if we look at the values of $X_1, X_2, [X_1, X_2]$ at particular points, then these vectors are linearly dependent, but as functions they are linearly independent and for this reason $\mathfrak{g}_p$ has dimension 3, even if $p = (0, y)$. On the other hand

$$\mathfrak{h}_p = \operatorname{span}(X_2) = \{e^{tX_2} \,:\, t \in \mathbb{R}\}$$

Notice then that $\mathfrak{g}_p$ has a dimension strictly bigger than the dimension of the manifold; indeed it has dimension 3 and so does G_p, which is isomorphic to $\mathbb{H}^3$: in fact, we are asking all the brackets to be zero, except for $[X_1, X_2]$. Therefore

$$G_p/H_p \cong \mathbb{H}^3/\{e^{tX_2} \,:\, t \in \mathbb{R}\}$$

and it has dimension 2. On the contrary, outside the line $x = 0$ the tangent space to the Grušin plane is isometric to the Euclidean plane $\mathbb{R}^2$ (hence, it has also dimension 2). $\diamondsuit$

Now let us see how we can eliminate singular points by constructing a sort of lift of the Carnot-Carathéodory space we are working with. As a consequence, we will be able to state a uniform version of the Ball-Box theorem.

Desingularization

In order to understand the aim behind desingularization, let us begin with a remark on the Ball-Box Theorem. We know that for every $p \in M$ there exist a system of privileged coordinates z at p and suitable constants $\varepsilon_p, C_p > 0$ such that if $d(p, q) < \varepsilon_p$, then

$$\frac{1}{C_p}\|z(q)\|_p \leq d(p, q) \leq C_p\|z(q)\|_p$$

What happens when p is moving? The statement is “continuous” in some sense, or better uniform; let us explain the reason. If p is a regular point, then there exists a smooth map $\Phi : N \to \mathbb{R}^n$ (where $N \subset M \times M$ is a neighbourhood of (p, p)) such that $\Phi(q, \cdot)$ is a system of privileged coordinates at q for any q. Such a map always exists because if $X_{I_1}, ..., X_{I_n}$ is an adapted frame at p, then it is sufficient to set $\Phi(q, q') = z$ with $q = \exp(z_1 X_{I_1} + ... + z_n X_{I_n})(q)$ and the behaviour is smooth. Thus, $\varepsilon_{(\cdot)}$ and $C_{(\varepsilon)}$ can be chosen in a continuous way or, in a better way, constant on compact sets. On the contrary, near a singular point the situation is completely different, because in general we do not know how to construct privileged coordinates. Moreover, if p is a singular point and $(p_n)_{n\in\mathbb{N}}$ is a sequence of regular points converging to p, then $\varepsilon_{p_n} \to 0$ as $n \to \infty$ but $\varepsilon_p > 0$: thus, a uniform estimate in the Ball-Box Theorem is impossible. In this framework, desingularization appears in order to eliminate in a proper way the singularities of the space we are working with. To understand the philosophy behind desingularization, just think about the following fact: by projecting a regular object, singularities may appear. Therefore, we have to perform a sort of inverse procedure and find a “lifting” of $(M, X_1, ..., X_m)$; more precisely, but still roughly speaking, we have to find

- $\tilde{M} = M \times \mathbb{R}^k$ or better $\tilde{M} \subset M \times \mathbb{R}^k$;
- vector fields $\xi_1, ..., \xi_m$ on $\tilde{M}$ which are bracket-generating and such that $\pi_* \xi_i = X_i$, where $\pi : \tilde{M} \to M$ is the natural projection;

such that $(\tilde{M}, \xi_1, ..., \xi_m)$ is *equiregular*, i.e. it has no singular points. In this setting the projection $\pi : \tilde{M} \to M$ plays a fundamental role, because via π_* the vector fields $\xi_1, ..., \xi_m$ on $\tilde{M}$ are sent on $X_1, ..., X_m$ and via π the trajectories $\tilde{\gamma}$ of the control system

$$\dot{\tilde{q}} = \sum_{i=1}^{m} u_i \xi_i(\tilde{q})$$

are sent into curves $\gamma = \pi(\tilde{\gamma})$ which are trajectories of

$$\dot{q} = \sum_{i=1}^{m} u_i X_i(q)$$

In addition, $\ell(\tilde{\gamma}) \geq \ell(\pi(\tilde{\gamma}))$ and if we denote by $\tilde{d}$ the distance induced by $\xi_1, ..., \xi_m$ on $\tilde{M}$, then from what we have just said we get

$$B(q, \varepsilon) = \pi\Big(B_{\tilde{d}}((q, 0), \varepsilon)\Big) \tag{7.4.2}$$

or equivalently

$$d(p, q) = \inf_{\pi(\tilde{q}) = q} \tilde{d}((p, 0), \tilde{q}) \tag{7.4.3}$$

Is it possible to find the desired lifting? Actually, we have already seen an example of desingularization, as pointed out in the following example.

Example 7.4.5. Let (M, d) be a Carnot-Carathéodory space and fix a point $p \in M$. Then, as shown in the diagram above, $(G_p, \hat{\xi}_1, ..., \hat{\xi}_m)$ is the desingularization of $(\mathbb{R}^n, \hat{d})$ (so we have not desingularized (M, d)!). $\diamondsuit$

If we want to desingularize a sub-Riemannian manifold $(M, X_1, ..., X_m)$, let us first take a singular point p and then let us choose privileged coordinates x at p, so that $X_i(x) = \hat{X}_i(x) + R_i(x)$ with $\mathrm{ord}_0(R_i) > -1$, as usual. Then, complete the coordinates x to coordinates (x, y) on G_p and take $\tilde{M} := M \times \mathbb{R}^k$ with $k \in \mathbb{N}$ such that $\dim G_p = n + k$. The vector fields $\hat{\xi}_1, ..., \hat{\xi}_m$ on (G_p, d_G) can thus be represented as

$$\hat{\xi}_i(x, y) = \hat{X}_i(x) + \sum_{j=n+1}^{n+k} v_{ij}(x, y) \partial_{y_j} \tag{7.4.4}$$

namely as a sum of a horizontal part $\hat{X}_i(x)$ and a vertical one, given by the sum (horizontal and vertical w.r.t. π). With this observation, if we set now

$$\xi_i(x,y) := X_i(x) + \sum_{j=n+1}^{n+k} v_{ij}(x,y)\partial_{y_j}$$

with the same functions v_{ij} than in (7.4.4). Clearly $\pi_*\xi_i = X_i$ and in this way we define a sub-Riemannian structure on an open set $\tilde{U} \subset \tilde{M}$ (indeed, we are working only locally) whose nilpotent approximation at $(p,0)$ is given by $(\hat{\xi}_1,...,\hat{\xi}_m)$ by construction. Is $(\tilde{M},\xi_1,...,\xi_m)$ equiregular? Unfortunately, $(p,0)$ can be itself a singular point for $\xi_1,...,\xi_m$ and so the answer is negative. Indeed, a point can be singular for a sub-Riemannian structure (in this case $(\tilde{M},\xi_1,...,\xi_m)$) and regular for the nilpotent approximation taken at this point (i.e. $(\tilde{M},\hat{\xi}_1,...,\hat{\xi}_m)$), just consider the following example.

Example 7.4.6. On $\mathbb{R}^3$ consider the distribution generated by

$$X_1 = \partial_x, \quad X_2 = \partial_y + x\partial_z, \quad X_3 = x^{100}\partial_z$$

Then, the nilpotent approximation at 0 is given by

$$\hat{X}_1 = X_1, \quad \hat{X}_2 = X_2, \quad \hat{X}_3 \equiv 0$$

With respect to it, the point 0 is regular, but the same point is singular for the initial distribution X_1, X_2, X_3. $\Diamond$

Thus, we need a group bigger than G_p in order to avoid this problem, namely the free nilpotent group N_r of step r mith m generators. It is a Carnot group and its Lie algebra $\mathfrak{n}_r$ is the free nilpotent Lie algebra of step r with m generators, say $\alpha_1,...,\alpha_m$; they define a right-invariant sub-Riemannian structure on N_r. Moreover, the group N_r can be thought as a group of diffeomorphisms and so define a left action on $\mathbb{R}^n$; more explicitly, an element $g \in N_r$ can be written as

$$g = \exp(\sum a_I \alpha_I)$$

and therefore, for $x \in \mathbb{R}^n$, we can define the left action of g on x as

$$g \cdot x := (\exp(\sum a_I \hat{X}_I))(x)$$

If we denote by J the isotropy subgroup of 0 for this action, then we obtain that $(\mathbb{R}^n, \hat{X}_1,...,\hat{X}_m)$ is isometric to N_r/J endowed with the restrictions of the right-invariant sub-Riemannian structure associated to $\alpha_1,...,\alpha_m$. Now, arguing as we previously did, we are able to lift locally the sub-Riemannian structure on M to a sub-Riemannian structure in $\tilde{M}$, here defined as $M \times \mathbb{R}^{\tilde{n}-n}$ with $\tilde{n} = \dim N_r$. Its nilpotent approximation at $(p,0)$ is the orthonormal basis of the sub-Riemannian structure defined by $\alpha_1,...,\alpha_m$ and, since N_r is free up to step r, we obtain that $(p,0)$ is a regular point for that structure in $\tilde{M}$. Thus we can summarize our results in the following lemma.

Lemma 7.4.7. *Let (M, d) be a Carnot-Carathéodory space, $p \in M$, r be the degree of nonholonomy at p, $\tilde{n}$ the dimension of the free Lie algebra of step r with m generators and $\tilde{M} := M \times \mathbb{R}^{\tilde{n}-n}$. Then there exist a neighbourhood $\tilde{U} \subset \tilde{M}$ of $(p, 0)$, a neighbourhood $U \subset M$ of p with $U \times \{0\} \subset \tilde{U}$, local coordinates (x, y) on $\tilde{U}$ and smooth vector fields on $\tilde{U}$*

$$\xi_i(x, y) = X_i(x) + \sum_{j=n+1}^{\tilde{n}} v_{ij}(x, y)\partial_{y_j}$$

such that:

(i) the distribution generated by $\xi_1, ..., \xi_m$ is bracket-generating and its non-holonomy degree is r everywhere (so its Lie algebra is free up to step r);

(ii) every $\tilde{q} \in \tilde{U}$ is regular;

(iii) denoting by $\pi : \tilde{M} \to M$ the canonical projection and $\tilde{d}$ the sub-Riemannian distance defined by $\xi_1, ..., \xi_m$ on $\tilde{U}$, we have that (7.4.2) and (7.4.3) hold for $q \in U$ and $\varepsilon > 0$ small enough.

All these properties are still true if r is any integer greater than the nonholonomy degree at p.

Hence, any sub-Riemannian manifold is locally the projection of an equiregular sub-Riemannian manifold and the projection preserves the trajectories, the minimizing curves and the distance.

Example 7.4.8. Let us consider the Heisenberg group $\mathbb{H}^3$ equipped with the distribution generated by

$$\xi_1 = \partial_x, \quad \xi_2 = \partial_y + x\partial_z$$

Then, via the projection associated to y we exactly get the Grušin plane, since

$$(\pi_y)_*\xi_1 = \partial_x, \quad (\pi_y)_*\xi_2 = x\partial_z$$

and thus, for instance, we have

$$d_{\text{Grušin}}(0, (x, z)) = \inf_y d_{\mathbb{H}^3}(0, (x, y, z))$$

as in (7.4.3). $\Diamond$

The importance of regular points is linked to uniformity:

- uniformity of the flag (7.2.1);
- uniformity w.r.t. p of the Gromov-Hausdorff convergence of $(\lambda M, p)$ to C_pM;
- uniformity of the distance estimates.

The last property is fundamental for several reasons (to compute Hausdorff dimension and prove global convergence of approximate motion planning algorithms, for instance), but at singular points we loose it. However, using the desingularization procedure we have just described, it is possible to give a uniform version of the Ball-Box Theorem and of the distance estimates. In order to do that, let us assume M to be oriented (we will need a volume form ω), let $K \subset M$ be a compact subset and let

$$r_{\max} := \max_{p \in K} r(p)$$

Given $q \in K$ and $\varepsilon > 0$ we consider the families $\mathcal{X} = \{X_{I_1}, ..., X_{I_n}\}$ of brackets of length $|I_i| \leq r_{\max}$. On the finite set of these families we define a function

$$\begin{aligned} f_{q,\varepsilon}(\mathcal{X}) &:= |\omega_q(\varepsilon^{|I_1|} X_{I_1}(q), ..., \varepsilon^{|I_n|} X_{I_n}(q))| \\ &= |\det(\varepsilon^{|I_1|} X_{I_1}(q), ..., \varepsilon^{|I_n|} X_{I_n}(q))| \end{aligned}$$

and we say that $\mathcal{X}$ is an *adapted frame at* (q, ε) if it achieves the maximum of $f_{q,\varepsilon}$. Notice that in this case $X_{I_1}(q), ..., X_{I_n}(q)$ is a basis of T_qM and, q being fixed, the adapted frames at (q, ε) are adapted frame at q for ε small enough. Now we can state the uniform version of the Ball-Box Theorem; the idea of the proof is due to Rothschild and Stein.

Theorem 7.4.9 (Uniform Ball-Box Theorem)**.** *There exist positive constant C, ε_0 such that, for all $q \in K$ and for any $\varepsilon < \varepsilon_0$, if $\mathcal{X}$ is an adapted frame at (q, ε) it holds*

$$\mathrm{Box}_{\mathcal{X}}(q, C^{-1}\varepsilon) \subset B(q, \varepsilon) \subset \mathrm{Box}_{\mathcal{X}}(q, C\varepsilon)$$

where $\mathrm{Box}_{\mathcal{X}}(q, \varepsilon) = \{e^{x_1 X_{I_1}} \circ ... \circ e^{x_n X_{I_n}} \ : \ |x_i| \leq \varepsilon^{|I_i|}, \ 1 \leq i \leq n\}$.

As a consequence, we entail that

$$\mathrm{vol}_\omega B(q, \varepsilon) \asymp \max_{\mathcal{X}} f_{q,\varepsilon}(\mathcal{X})$$

where $\asymp$ means that

$$c \max_{\mathcal{X}} f_{q,\varepsilon}(\mathcal{X}) \leq \mathrm{vol}_\omega B(q, \varepsilon) \leq C \max_{\mathcal{X}} f_{q,\varepsilon}(\mathcal{X})$$

for suitable constants c, C. As a final remark, let us consider again the Grušin plane and let us describe explicitly balls and boxes in this sub-Riemannian environment.

Example 7.4.10. In the Grušin plane there are two kinds of boxes, which are the following ones

$$\mathrm{Box}(\varepsilon) = [x - \varepsilon, x + \varepsilon] \times \begin{cases} [-x\varepsilon, x\varepsilon] & \text{if } \varepsilon \leq x \\ [-\varepsilon^2, \varepsilon^2] & \text{if } \varepsilon \geq x \end{cases}$$

We can also represent balls as follows.
In the picture above, the inner and the outer balls are respectively given by

$$\{e^{x_1X_1+x_2X_2}(q)\,:\,|x_1|<\varepsilon,\,|x_2|<\varepsilon\}$$
$$\{e^{z_1X_1+z_2[X_1,X_2]}(q)\,:\,|z_1|<\varepsilon,\,|z_2|<\varepsilon^2\}$$

and so we can see that the Ball-Box Theorem holds. $\Diamond$

7.5 Lecture 5 - 27 October

Up to now we have treated sub-Riemannian geometry from an infinitesimal point of view: first order approximations, nilpotent approximations, metric tangent spaces and desingularization. In this lecture we will use some of these notions in order to treat Hausdorff measure and dimension in sub-Riemannian manifolds or, more precisely, in Carnot-Carathéodory spaces. First, let us briefly recall the notions of Hausdorff dimension and measure in a general metric space, in a way which is slightly different from Definition 5.2.7 because no normalization constants are involved.

Definition 7.5.1. *Let (M,d) be a metric space, $\alpha\in[0,+\infty[$ and $A\subset M$. The* α-dimensional Hausdorff measure of A *is defined as*

$$\mathscr{H}^\alpha(A):=\lim_{\delta\to 0}\mathscr{H}^\alpha_\delta(A)$$

where $\mathscr{H}^\alpha_\delta(A)$ is the δ-approximating α-dimensional Hausdorff measure of A, *given by*

$$\mathscr{H}^\alpha_\delta(A):=\inf\left\{\sum_i(\operatorname{diam}S_i)^\alpha\,:\,S_i\subset M\ \textit{closed},\ \bigcup_i S_i\supset A,\ \operatorname{diam}S_i\leq\delta\right\}$$

In an analogous way we can also introduce the α-dimensional spherical Hausdorff measure $\mathscr{S}^\alpha$ *by considering only closed balls as S_i. In this case it is possible to introduce a normalization constant, since if $M=\mathbb{R}^n$ some authors would prefer to have $\mathscr{S}^n=\mathscr{L}^n$, where $\mathscr{L}^n$ is the n-dimensional Lebesgue measure. However, we will not introduce the normalization constant.*

Thanks to (5.2.5), we know that $\mathscr{H}^\alpha$ and $\mathscr{S}^\alpha$ are mutually absolutely continuous; hence, from now on we will only consider $\mathscr{S}^\alpha$. The behaviour of this measure as a function of α for a fixed set $A\subset M$ is described in Figure 7.1. Thus, $\mathscr{S}^\alpha(A)$ can be finite and not zero only for at most one value of α and we will refer to such value as the *Hausdorff dimension of* A, denoted by $\dim_H(A)$. In a more explicit way, we can say that

$$\dim_H(A):=\sup\{\alpha\,:\,\mathscr{S}^\alpha(A)=+\infty\}=\inf\{\alpha\,:\,\mathscr{S}^\alpha(A)=0\}$$

For this reason, the Hausdorff volume on M is defined as the spherical Hausdorff measure of dimension $\dim_H(M)$ and will be also denoted by vol_H. As a first remark, let us point out that the definitions we provided are consistent with the classical ones in a smooth setting.

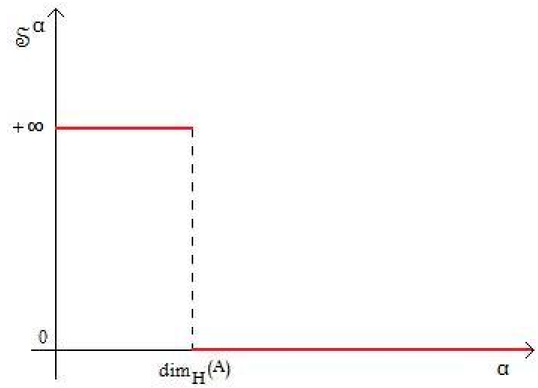


Figure 7.1: Hausdorff dimension

Example 7.5.2. Let (M, g) be a Riemannian manifold. In this case the Hausdorff dimension of M coincides with its topological dimension; moreover, vol_H and the Lebesgue measure on M are mutually absolutely continuous. We can also be more precise: in coordinates, the link between vol_H and the Lebesgue measure is given by

$$\mathrm{vol}_H = c\sqrt{\det g}dx_1...dx_n$$

where c is the normalization constant. $\Diamond$

Now let us see what happens in sub-Riemannian manifolds. Let us consider a sub-Riemannian manifold $(M, X_1, ..., X_m)$ (recall that it can be equivalently represented by a suitable sub-Riemannian structure (U, g^U, f)) and assume for the moment that M is equiregular, that is the flag (7.2.1) has constant dimension. Indeed, a point is regular if the flag dimension is locally constant near it and since we are assuming M to be equiregular, every point is regular; as in addition M is connected, we infer that the flag (7.2.1) has constant dimension *tout court* on M. We have already seen that we can associate weights $w_1, ..., w_n$ to such flag; as a consequence of the equiregularity assumption, also the weights are constant on M. As a further step, let us introduce the nilpotent integer Q (in the course *Geometric measure theory* by Serra Cassano and Ambrosio we called it *homogeneous dimension*) as

$$Q := \sum_{i=1}^{n} w_i$$

and notice that it can be equivalently written as

$$Q = \sum_{j\geq 1} j(\dim \mathscr{D}_p^j - \dim \mathscr{D}_p^{j-1})$$

being $\mathscr{D}_p^j$ independent of p. This number plays a fundamental role, as it turns out to be the Hausdorff dimension of the sub-Riemannian manifold, as we are going to see in the forthcoming theorem; the proof can be found in [17]. But first let us recall that a measure μ on M is *smooth* provided there exists an n-form ω on M with ω nowhere equal to zero such that

$$\mu(A) = \int_A \omega, \quad \forall A \in \mathscr{B}(M)$$

where $\mathscr{B}(M)$ denotes the Borel σ-algebra over M.

Theorem 7.5.3. *Let $(M, X_1, ..., X_m)$ be a equiregular sub-Riemannian manifold. Then the following facts hold:*

(i) $\dim_H M = Q$ *and Q is strictly bigger than the topological dimension of M;*

(ii) $\mathscr{S}^Q$ *is a Radon measure*

(iii) given any smooth measure μ on M, $\mathscr{S}^Q$ and μ are mutually absolutely continuous.

Thus we can perceive the difference between Example 7.5.2 and the sub-Riemannian case; however, property (iii) tells us that such difference is not so big and $\mathscr{S}^Q$ has more or less the same behaviour of a smooth measure. Before the proof, let us spend few more words on smooth measures: they always exist at least locally, but for the global existence the orientability of M is required, because otherwise there does not exist a nowhere-vanishing n-dimensional form. Furthermore, we will need the following tool.

Lemma 7.5.4. *Let $p \in M$ and μ be a smooth measure on M (actually, the smoothness near p is sufficient). Let us assume that there exist $\varepsilon_0, \mu_-, \mu_+ > 0$ such that*

$$\mu_- \le \frac{\mu(B(q,\varepsilon))}{(2\varepsilon)^Q} \le \mu_+ \tag{7.5.1}$$

for any $q \in B(p, \varepsilon_0)$ and for any $\varepsilon < \varepsilon_0$. Then

$$\frac{1}{\mu_+} \le \frac{\mathscr{S}^Q(B(p,\varepsilon))}{\mu(B(p,\varepsilon))} \le \frac{1}{\mu_-} \tag{7.5.2}$$

Proof. In order to prove the thesis we have to estimate the Q-dimensional spherical Hausdorff measure of $B(p, \varepsilon)$ and $\mu(B(p, \varepsilon))$. Thus, let $\{B(q_i, r_i)\}_{i\in\mathbb{N}}$ be a covering of $B(p, \varepsilon)$ and let us assume that $\operatorname{diam} B(q_i, r_i) < \delta$ for any $i \in \mathbb{N}$ for a given δ. In the forthcoming argument, we will let $\delta \to 0$, so we can assume without loss of generality that $\delta \ll \varepsilon_0$. As a first step, observe that

$$\mu(B(p,\varepsilon)) \le \sum_{i\in\mathbb{N}} \mu(B(q_i, r_i))$$

and now we can apply assumption (7.5.1) to each ball $B(q_i, r_i)$, so that

$$\sum_{i\in\mathbb{N}} \mu(B(q_i, r_i)) \leq \mu_+ \sum_{i\in\mathbb{N}} (2r_i)^Q = \mu_+ \sum_{i\in\mathbb{N}} (\operatorname{diam} B(q_i, r_i))^Q$$

Indeed, even in sub-Riemannian geometry, $\operatorname{diam} B(q, \varepsilon) = 2\varepsilon$ for ε sufficiently small (if ε is large, this is false even in the Riemannian case). This is due to the fact for any $q \in M$ there exists at least one normal geodesic $\gamma : [0, \delta[\to M$ starting from this point and since normal geodesics are solutions of an ODE, they can be extended also at negative times, so that actually γ is defined on $]-\delta', \delta'[$ (for a suitable $\delta' < \delta$) and it is also length minimizing, i.e. $d(\gamma(-\delta'), \gamma(\delta')) = 2\delta'$. Clearly, the parameter ε is not constant, but it depends smoothly on q. Thus, we have proved that

$$\mu(B(p, \varepsilon)) \leq \mu_+ \sum_{i\in\mathbb{N}} (\operatorname{diam} B(q_i, r_i))^Q$$

and since this inequality holds for any covering $\{B(q_i, r_i)\}_{i\in\mathbb{N}}$ of $B(p, \varepsilon)$, by taking the infimum over all the possible covering at the right-hand side we get the δ-approximating Q-dimensional spherical Hausdorff measure. This means that $\mu(B(p, \varepsilon)) \leq \mu_+ \mathscr{S}^Q_\delta(B(p, \varepsilon))$ and letting $\delta \to 0$ we obtain

$$\mu(B(p, \varepsilon)) \leq \mu_+ \mathscr{S}^Q(B(p, \varepsilon))$$

i.e. the left inequality in (7.5.2). For the other one, the argument is exactly the same, except for the fact that we consider an "almost" optimal covering, that is a covering of $B(p, \varepsilon)$ such that $\mathscr{S}^Q(B(p, \varepsilon))$ differs from $\sum_{i\in\mathbb{N}} (\operatorname{diam} B(q_i, r_i))^Q$ by a small error (as small as we want); then, by virtue of Vitali's covering lemma the conclusion follows. □

Let us recall that by the Ball-Box Theorem we have (in privileged coordinates)

$$B(p, \varepsilon) \asymp \{z \in \mathbb{R}^n \,:\, |z_i| \leq \varepsilon^{w_i}\}$$

and

$$\mathscr{L}^n(\{z \in \mathbb{R}^n \,:\, |z_i| \leq \varepsilon^{w_i}\}) \asymp \varepsilon^Q$$

In the equiregular case, all the things above behave smoothly and the Ball-Box Theorem holds in a uniform sense, so that the assumption (7.5.1) of the previous lemma holds on every compact subset and, thus, also the estimate (7.5.2). As a consequence, using covering by balls relation (7.5.2) can be extended to any compact subset and so we deduce that

(i) $\mathscr{S}^Q$ is finite on every compact subset (this follows by the right-hand side of (7.5.2)), i.e. it is a Radon measure;

(ii) $\mathscr{S}^Q$ and μ are mutually absolutely continuous;

(iii) $\dim_H(M) = Q$

whence Theorem 7.5.3 follows. In particular, for (iii) notice that if $\alpha > Q$, then $\mathscr{S}^\alpha(K) = 0$ for any compact subset $K \subset M$, whence by taking a covering of M by compact subsets we get

$$\mathscr{S}^\alpha(M) = \mathscr{S}^\alpha\left(\bigcup_{n\in\mathbb{N}} K_n\right) \leq \sum_{n\in\mathbb{N}} \mathscr{S}^\alpha(K_n) = 0$$

Conversely, if $\alpha < Q$ then $\mathscr{S}^\alpha(B(p,\varepsilon)) = +\infty$ because $\mathscr{S}^Q(B(p,\varepsilon))$ is finite and Q is the only value for which this can happen. Thus Theorem 7.5.3 is completely proved.

Now observe that from property (ii) there exists the Radon-Nikodym derivative of $\mathscr{S}^Q$ w.r.t. any smooth measure μ and, by metric properties of our space, it can be proved that

$$\frac{d\mathscr{S}^Q}{d\mu}(p) = \lim_{\varepsilon\to 0} \frac{\mathscr{S}^Q(B(p,\varepsilon))}{\mu(B(p,\varepsilon))} \tag{7.5.3}$$

If assumption (7.5.1) is satisfied, then the term $\frac{\mathscr{S}^Q(B(p,\varepsilon))}{\mu(B(p,\varepsilon))}$ belongs to $[1/\mu_+, 1/\mu_-]$ by virtue of the previous lemma. Hence if we are able to prove that (7.5.1) holds as $\varepsilon_0 \to 0$ and we are also able to estimate μ_+ and μ_- as $\varepsilon_0 \to 0$, then we can estimate the Radon-Nikodym derivative and this is exactly what we are going to do. Let us recall that we are working in an equiregular manifold and let us assume that it is oriented, since in what follows we will need globally defined smooth measures; thus let ω be a volume form and let μ be the associated smooth measure, namely

$$\mu = \int \omega \tag{7.5.4}$$

Let us also recall from the previous lecture that the nilpotent approximation at p is a Carnot group G_p with right-invariant sub-Riemannian structure which can be expressed as $G_p = \exp(\mathfrak{g}_p)$, where

$$\mathfrak{g}_p = \mathscr{D}_p^1 \oplus \mathscr{D}_p^2/\mathscr{D}_p^1 \oplus ... \oplus \mathscr{D}_p^r/\mathscr{D}_p^{r-1}$$

This means that at regular points an intrinsic representation is possible. Now, starting from the smooth measure μ on M let us introduce a measure $\hat{\mu}_p$ on G_p. The first thing we need is the following result.

Lemma 7.5.5. *There exists a canonical isomorphism between n-forms on T_p^*M and n-forms on $\mathfrak{g}_p^*$, i.e.*

$$\bigwedge^n(T_p^*M) \cong \bigwedge^n(\mathfrak{g}_p)$$

Proof. As a first remark, an element of $\bigwedge^n(\mathfrak{g}_p^*)$ can be written as $\overline{\vartheta}_1 \wedge ... \wedge \overline{\vartheta}_n$, where every $\overline{\vartheta}_i$ is a 1-form on $\mathscr{D}_p^{w_i}/\mathscr{D}_p^{w_{i-1}}$. From this expression we can find 1-form $\vartheta_1, ..., \vartheta_n$ with ϑ_i a 1-form on $\mathscr{D}_p^{w_i}$ such that

$$\overline{\vartheta}_i = \vartheta_i + (\mathscr{D}_p^{w_{i-1}})^*$$

and thus we deduce that

$$\vartheta_1 \wedge ... \wedge \vartheta_n \in \bigwedge^n (T_p^* M)$$

This procedure is independent of the choice of ϑ_i, because if for every $i = 1, ..., n$ we choose another representative ϑ_i', then we know that

$$\vartheta_i' = \vartheta_i + \alpha_i$$

for a suitable $\alpha_i \in (\mathscr{D}^{w_{i-1}})^*$. Hence

$$\begin{aligned} (\vartheta_1 + \alpha_1) \wedge ... \wedge (\vartheta_n + \alpha_n) &= (\vartheta_1 + \alpha_1) \wedge ... \wedge \vartheta_n + (\vartheta_1 + \alpha_1) \wedge ... \wedge \alpha_n \\ &= (\vartheta_1 + \alpha_1) \wedge ... \wedge \vartheta_n \end{aligned}$$

because the second term at the right-hand side in the first equality is a wedge product of $n_{r-1}+1$ terms belonging to $(\mathscr{D}^{r-1})^*$ and so it must be equal to zero. By iterating this procedure we infer that

$$(\vartheta_1 + \alpha_1) \wedge ... \wedge (\vartheta_n + \alpha_n) = \vartheta_1 \wedge ... \wedge \vartheta_n$$

and so the procedure depicted above is really independent of the choice of $\vartheta_1, ..., \vartheta_n$. As a consequence, we have the canonical isomorphism. □

Thanks to this lemma, $\omega_p \in \bigwedge^n (T_p^* M)$ can be seen as $\tilde{\omega} \in \bigwedge^n (\mathfrak{g}_p^*)$ and to this form we can easily associate a further right-invariant n-form $\hat{\omega}_p$ on G_p, which will play the role of volume form. Finally, let us denote by $\hat{\mu}_p$ the smooth measure associated to $\hat{\omega}_p$ via relation (7.5.4). In this way we have been able to produce a measure $\hat{\mu}_p$ on G_p starting from a smooth measure μ on M, as desired, and we have found an intrinsic notion of infinitesimal measure. Now we will see that $\hat{\mu}_p$ is the correct candidate (up to suitable constants) to be μ_+ and μ_- in (7.5.2), so that, as declared above, we will be able to estimate the Radon-Nikodym derivative of $\mathscr{S}^Q$ w.r.t. μ.

Theorem 7.5.6 (Agrachev-Barilari-Boscain, Ghezzi-Jean)**.** *In our framework we have that, for any smooth measure μ,*

$$\frac{d\mathscr{S}^Q}{d\mu}(p) = \frac{2^Q}{\hat{\mu}_p(\hat{B}_p)} \tag{7.5.5}$$

where $\hat{B}_p$ is the unit ball centered at p in the nilpotent approximation, i.e. $\hat{B}(p, 1)$.

Notice that $\operatorname{diam} \hat{B}_p = 2$ w.r.t. the distance $\hat{d}$ and this motivates the term 2^Q at the right-hand side in (7.5.5).

Proof. By Lemma 7.5.4 we know that if assumption (7.5.1) is satisfied, then the estimate (7.5.2) holds and we also know that for the Radon-Nikodym derivative of $\mathscr{S}^Q$ w.r.t. μ (7.5.3) is true. Hence, the only thing we have to show is that for suitable $\mu_+(\varepsilon), \mu_-(\varepsilon)$ such that (7.5.1) holds we have

$$\lim_{\varepsilon\to 0} \mu_+(\varepsilon) = \frac{\hat{\mu}_p(\hat{B}_p)}{2^Q}, \quad \lim_{\varepsilon\to 0} \mu_-(\varepsilon) = \frac{\hat{\mu}_p(\hat{B}_p)}{2^Q}$$

Thus, let us prove that (7.5.1) is satisfied with

$$\mu_\pm(\varepsilon) = \frac{\hat{\mu}_p(\hat{B}_p)}{2^Q} + O(\varepsilon)$$

This immediately follows if we prove that

$$\frac{\mu(B(p,\varepsilon))}{\varepsilon^Q} \to \hat{\mu}_p(\hat{B}_p) \tag{7.5.6}$$

uniformly on compact subsets. Let us fix a point $p \in M$; even if we prove the simple pointwise convergence w.r.t. p in (7.5.6), the procedure we are going to describe is uniform in p and so the conclusion follows. Let $X_{I_1}, ..., X_{I_n}$ be an adapted basis at p, which means that $|I_i| = w_i$ for any $i = 1, ..., n$, and let us consider exponential coordinates of the first type, defined as the inverse of the local diffeomorphism (7.2.2). Within this frame, let us describe the passage from ω to $\hat{\omega}_p$ we explained before. As a first step, denoting by $X^*_{I_1}, ..., X^*_{I_n}$ the basis of the cotangent bundle dual to $X_{I_1}, ..., X_{I_n}$, we have that

$$\omega = \omega(X_{I_1}, ..., X_{I_n}) dX^*_{I_1} \wedge ... \wedge dX^*_{I_n}$$

and so

$$\omega_p = \omega(X_{I_1}, ..., X_{I_n})(p) dX^*_{I_1}|_p \wedge ... \wedge dX^*_{I_n}|_p$$

Via the canonical isomorphism of Lemma 7.5.5 we have that

$$\tilde{\omega} = \omega(X_{I_1}, ..., X_{I_n})(p) d\hat{X}^*_{I_1}|_p \wedge ... \wedge d\hat{X}^*_{I_n}|_p$$

and finally

$$\hat{\omega}_p = \omega(X_{I_1}, ..., X_{I_n})(p) d\hat{X}_{I_1} \wedge ... \wedge d\hat{X}_{I_n}$$

Now observe that $\hat{X}_{I_i}$ is a homogeneous vector field of weighted degree $-w_i$ and its value at 0 coincides with ∂_{z_i}; hence

$$\hat{X}_{I_i} = \partial_{z_i} + \text{terms in } \partial_{z_k}$$

with $k > j$. As a second remark, by evaluating at p the relation above we get

$$\hat{X}_{I_i}(p) = \partial_{z_i}(p)$$

As a consequence, we entail that

$$\hat{\omega}_p = \omega(\partial_{z_1}, ..., \partial z_n)(p) dz_1 \wedge ... \wedge dz_n \tag{7.5.7}$$

In order to prove (7.5.6), notice that

$$\mu(B(p,\varepsilon)) = \int_{B(p,\varepsilon)} \omega$$

By Corollary 7.3.5 we have that $B(p,\varepsilon) \asymp \hat{B}(0,\varepsilon)$ and so

$$\mu(B(p,\varepsilon)) = (1+O(\varepsilon)) \int_{\hat{B}(0,\varepsilon)} \omega(\partial_{z_1}, ..., \partial_{z_n}) dz_1 \wedge ... \wedge dz_n$$

Using the fact that $\hat{B}(0,\varepsilon) = \delta_\varepsilon \hat{B}_p$ and changing the variables z_i into $\delta_\varepsilon z_i$ we obtain

$$\begin{aligned} \mu(B(p,\varepsilon)) &= (1+O(\varepsilon)) \int_{\hat{B}_p} \omega(\partial_{z_1}, ..., \partial_{z_n})(\delta_\varepsilon z) \varepsilon^{w_1} dz_1 ... \varepsilon^{w_n} dz_n \\ &= (1+O(\varepsilon)) \int_{\hat{B}_p} \varepsilon^Q (\omega(\partial_{z_1}, ..., \partial_{z_n})(p) + O(\varepsilon)) dz \\ &= \varepsilon^Q \Big(\hat{\mu}_p(\hat{B}_p) + O(\varepsilon) \Big) \end{aligned}$$

where the last identity is due to (7.5.7) and this exactly proves (7.5.6). □

Hence we have been able to compute explicitly the Radon-Nikodym derivative of $\mathscr{S}^Q$ w.r.t. any smooth measure μ and formula (7.5.5) also summarizes the conclusions of Theorem 7.5.3. However, further questions naturally arise.

(i) What is the regularity of $\frac{d\mathscr{S}^Q}{d\mu}$? For this first question, Agrachev, Barilari and Boscain proved that it is always continuous and sometimes also smooth, but this last condition sometimes may fail. Indeed they proved that, in a general sub-Riemannian structure (M, Δ, g) with the distribution Δ having corank 1 and $\dim M \geq 5$, the Radon-Nikodym derivative of $\mathscr{S}^Q$ w.r.t. is always C^3, sometimes C^4, but it can not be C^5. This is a bad news, because the Hausdorff measure $\mathscr{S}^Q$ is fundamental for the definition of an intrinsic volume, which is in turn necessary for the introduction of the Laplace-Beltrami operator.

(ii) Taking into account what we have just said, are there smooth intrinsic measures? The answer is affirmative and Popp's measure is one of them. For instance, if we take a distribution Δ of rank 2 on $\mathbb{R}^3$ and X_1, X_2 form an orthonormal frame for Δ, then it is not difficult to see that

$$dX_1 \wedge dX_2 \wedge d[X_1, X_2]$$

is a smooth volume independent of the choice of X_1, X_2. The associated smooth measure is Popp's measure.

(iii) Let us no more assume M to be equiregular and let $N \subset M$ be a strongly equiregular submanifold. This means that the growth vector is constant on N and, moreover, also

$$q \mapsto (\dim(\mathscr{D}_q^1 \cap T_qN), ..., \dim(\mathscr{D}_q^r \cap T_qN))$$

is constant. In this case we can prove for N all the results we just proved on M under the equiregularity assumption (pay attention to the fact that N is a closed subset of M, so a point regular w.r.t. N need not be regular in M). We will not give the details, but let us mention that in this case Q is replaced by

$$Q_N := \sum_{j\geq 1} j(\dim(\mathscr{D}_p^j \cap T_pN) - \dim(\mathscr{D}_p^{j-1} \cap T_pN))$$

and it has been proved by R. Ghezzi and F. Jean that $\dim_H N = Q_N$ and

$$\frac{d\mathscr{S}^{Q_N}|_N}{d\mu}(p) = \frac{\operatorname{diam}(\hat{B}_p \cap T_pN)^{Q_N}}{\hat{\mu}(\hat{B}_p \cap T_pN)}$$

for any smooth measure μ on N. Notice that in general $\operatorname{diam}(\hat{B}_p \cap T_pN)$ is not equal to 2 and need not be constant on N.

7.6 Lecture 6 - 3 November

7.7 Lecture 7 - 4 November

7.8 Lecture 8 - 5 November

7.9 Lecture 9 - 6 November

7.10 Lecture 10 - 17 November

7.11 Lecture 11 - 18 November

7.12 Lecture 12 - 20 November

Part III

Thematic days

Chapter 8

Optimal transport and sub-Riemannian manifolds

8.1 Monge's transport problem for distance cost in metric spaces - Séverine Rigot

Abstract: In this talk we will address the problem of the existence of solutions to Monge's transport problem in metric spaces. We will present a strategy that has been successfully used in the case of the sub-Riemannian Heisenberg group (joint work with L. De Pascale). This strategy does not involve any Sudakov-type dimension reduction argument nor disintegration of measures and is by many aspects much more simplier. We will also discuss extension of this strategy to other metric spaces focusing on Carnot groups and more generally sub-Riemannian manifolds.

8.2 On measure contraction properties in sub-Riemannian geometry - Luca Rizzi

Abstract: We discuss how the volume of measurable sets evolves upon geodesic contraction. A recent results by Juillet suggests that the behaviour is dramatically different when compared with the Riemannian one. For example, the volume of the metric ball of the Heisenberg group contracts to zero with an exponent equal to 5, whereas one would expect the topological (equal to 3) or at least the Hausdorff dimension (equal to 4). We show that the same behaviour occurs for any sub-Riemannian manifold, and the critical exponent is a new dimensional invariant (the geodesic dimension). A recent result by Rifford states that any (ideal) Carnot group satisfies the classical $MCP(0, N)$ for some exponent N. It is unknown which is the optimal N. We conjecture that, at least for step 2 Carnot groups, the optimal exponent is the geodesic dimension. We

prove this fact for the (3,6) Carnot group (which has geodesic dimension 14), and we discuss how the same techniques work for the general step 2, rank k, free Carnot group (whose geodesic dimension grows as k^3, while the topological and Hausdorff dimensions both grow as k^2). This is a work in progress, in collaboration with D. Barilari.

8.3 On the Brunn-Minkowski inequality - Nicolas Juillet

Abstract: We will recall different versions of the Brunn-Minkowski inequality (interpolation of sets) and their connection to the isoperimetric problem and the theory of optimal transportation (interpolation of measures). We will examine the interpolation of sets on some sub-Riemannian manifolds, including the Heisenberg group and the Grushin plane.

8.4 BV functions and sets of finite perimeter in sub-Riemannian manifolds - Roberta Ghezzi

Abstract: We give a notion of BV function on an oriented manifold where a volume form and a family of lower semicontinuous quadratic forms are given. Using this notion, we generalize the structure theorem for BV functions that holds in the Euclidean case. When we consider sub-Riemannian manifolds, our definition coincide with the one given in the more general context of metric measure spaces which are doubling and support a Poincar inequality. We study finite perimeter sets of sub-Riemannian manifold, i.e. sets whose characteristic function is BV, and we prove a blowup theorem, generalizing the one obtained for step-2 Carnot groups in [Franchi-Serapioni-Serra Cassano JGA 2003]. This is a joint work with L. Ambrosio and V. Magnani.

8.5 Optimal transport maps on non-branching metric measure spaces - Martin Huesmann

Abstract: Let (X, d, m) be a proper, non-branching, metric measure space. We show existence and uniqueness of optimal transport maps for cost written as non-decreasing and strictly convex functions of the distance, provided (X, d, m) satisfies a new weak property concerning the behavior of m under the shrinking of sets to points. This in particular covers spaces satisfying the measure contraction property. This is joint work with Fabio Cavalletti.

Chapter 9

Hypoelliptic diffusion: analysis and control

9.1 Heat kernel small time asymptotics at the cut locus in sub-Riemannian geometry - Davide Barilari

9.2 Intrinsic hypoelliptic diffusion in sub-Riemannian and almost-Riemannian geometry - Dario Prandi

Bibliography

[1] A. Agrachev, D. Barilari, U. Boscain, *Introduction to Riemannian and Sub-Riemannian geometry*, lecture notes (2014)

[2] L. Ambrosio, *Fine properties of sets of finite perimeter in Ahlfors regular metric measure spaces*, Adv. in Math. 159 (2001) pp. 51-67

[3] L. Ambrosio, *Fine properties of sets of finite perimeter in doubling metric measure spaces*, Set Valued Analysis 10 (2002) pp. 111-128

[4] L. Ambrosio, B. Kirchheim, *Rectifiable sets in metric and Banach spaces*, Math. Ann. 318 (2000) pp. 527-555

[5] A. Bellaïche, Jean-Jacques Risler, *Sub-Riemannian Geometry*, Progress in Mathematics 144, Birkhäuser Verlag (1996)

[6] A. Bonfiglioli, E. Lanconelli, F. Uguzzoni, *Stratified Lie Groups and Potential Theory for their Sub-Laplacians*, Springer Monographs in Mathematics, Springer-Verlag Berlin Heidelberg, New York (2007)

[7] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geometry and Functional Analysis 9 (1999) pp. 428-517

[8] B. Franchi, R. Serapioni, F. Serra Cassano, *On the structure of finite perimeter sets in step 2 Carnot groups*, Journal of Geometric Analysis 13, n. 3 (2003) pp. 421-466

[9] B. Franchi, R. Serapioni, F. Serra Cassano, *Rectifiability and perimeter in the Heisenberg group*, Math. Ann. 321 (2001) pp. 479-531

[10] B. Franchi, R. Serapioni, F. Serra Cassano, *Regular hypersurfaces, Intrinsic Perimeter and Implicit Function Theorem in Carnot groups*, Comm. Anal. Geom. 11, no 5. (2003) pp. 909-944

[11] N. Garofalo, D. M. Nhieu, *Lipschitz continuity, global smooth approximations and extension theorems for Sobolev functions in Carnot Carathéodory spaces*, J. Anal. Math. 74 (1998) pp. 67-97

[12] F. Jean, *Sub-Riemannian Geometry*, lecture notes

[13] V. Magnani, *Towards differential calculus in stratified groups*, J. Aust. Math. Soc. 95 (2013) pp. 76-128

[14] B. Kirchheim, F. Serra Cassano, *Rectifiability and parametrizations of intrinsically regular surfaces in the Heisenberg group*, Ann. Scuola Norm. Sup. Pisa, Cl. Sc. 5 III (2004) pp. 871-896

[15] N. G. Meyers, J. Serrin, $H = W$, Proc. Nat. Acad. Sci. USA, vol. 51 (1964) pp. 1055-1056

[16] J. Mitchell, *On Carnot-Carath'eodory metrics*, J. Differ. Geom. 21 (1985) pp. 35-45

[17] R. Montgomery, *A Tour of Subriemannian Geometries, Their Geodesics and Applications*, Providence American Mathematical Society (2002)

[18] R. Monti, F. Serra Cassano, *Surface measures in Carnot-Carathéodory spaces*, Calc. Var. Partial Diff. Eq. 13 (2001) pp. 339-376

[19] P. Pansu, *Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un*, Ann. of Math. 129 (1989) pp. 1-60

[20] L. Rifford, *Sub-Riemannian Geometry and Optimal transport*, Springerbriefs in Mathematics (2014)

[21] S. Rigot, *Mass transportation in groups of type H*, Commun. Contemp. Math. 7 (2005) pp. 509-537

[22] F. Serra Cassano, *Some topics of geometric measure theory in Carnot groups*, lecture notes