Tangent bundles to sub-Riemannian groups

Marius Buliga

Institut Bernoulli
Bâtiment MA
École Polytechnique Fédérale de Lausanne
CH 1015 Lausanne, Switzerland
Marius.Buliga@epfl.ch

and

Institute of Mathematics, Romanian Academy
P.O. BOX 1-764, RO 70700
Bucureşti, Romania
Marius.Buliga@imar.ro

This version: 16.10.2003

1   Introduction

Classical calculus is a basic tool in analysis. We use it so often that we forget that its construction needed considerable time and effort.

Especially in the last decade, the progresses made in the field of analysis in metric spaces make us reconsider this calculus. Along this line of thought, all started with the definition of Pansu derivative [25] and its version of Rademacher theorem in Carnot groups. It is amazing that such a basic notion can still lead to impressive results, like the rigidity of quasi-isometric embeddings.

The Gromov-Hausdorff convergence of metric spaces allows to define the notion of tangent space to a metric space. The tangent space to a metric space at a point is defined only up to isometry. For example, the tangent space to a n-dimensional Riemannian manifold, at a point, is R^n with the Euclidean distance.

For (almost) general metric spaces Cheeger [7] constructed a tangent bundle. The tangent bundle constructed by Cheeger does not have as fiber the metric tangent space, in the case of regular sub-Riemannian manifolds.

The sub-Riemannian manifolds (and improper called sub-Riemannian geometry) form an important class of metric spaces which are not Euclidean at any scale. Basic references for sub-Riemannian geometry are Bellaïche [4] and Gromov [15].

These spaces have more structure than just the metric one.

The extra structure of sub-Riemannian manifolds permitted to Margulis and Mostow [23] to construct a tangent bundle to a sub-Riemannian manifold. Their previous paper [22] contains a study of the differentiability properties of quasi-conformal mappings between regular sub-Riemannian manifolds. The central result is a Stepanov theorem in this setting, that is any quasi-conformal map from a Carnot-Carathéodory manifold to another is a.e. differentiable. Same result for quasi-conformal maps on Carnot groups has been first proved by Koranyi, Reimann [19].

As a motivation of the paper [23] the authors mention a comment of Deligne about the fact that in [22] they refer to "the tangent space" to a point, that is to a tangent bundle structure, which was not constructed in the first paper. They remedy this gap in the second paper [23]; their tangent bundle has fibers isometric with the tangent spaces.

Stepanov theorem is a hard generalisation of Rademacher theorem. A way to prove Stepanov theorem for sub-Riemannian spaces is to prove that Pansu-Rademacher theorem holds Lipschitz functions $f : A \subset M \to N$, where M, N are Carnot groups and A is just measurable. This has been done for the first time in Vodop'yanov, Ukhlov [34], in the case of Carnot groups. Their technique of proof differs from the one used by Pansu in [25]. It is closer to the one used in Margulis, Mostow [22], where it is proven that any quasi-conformal map from a Carnot-Carathéodory manifold to another is a.e. differentiable, without using Rademacher theorem.

This paper is a continuation of [5], where it is argued that sub-Riemannian geometry is in fact non-Euclidean analysis. This is more easy to see when approaching the concept of a sub-Riemannian Lie group. As a starting point we have a real Lie group G which is connected and a vectorspace $D \subset \mathfrak{g} = Lie(G)$ which bracket generates the Lie algebra $\mathfrak{g}$. The vectorspace D provides a left-invariant regular distribution on G which

is completely non-integrable. By an arbitrary choice of an Euclidean norm on D we can endow G with a left invariant Carnot-Caratheodory distance. The machinery of metric spaces comes into action and tells that, as a metric space, G has tangent space at any point and any such tangent space is isomorphic with the nilpotentization of $\mathfrak{g}$ with respect to D, denoted by $N(G)$.

$N(G)$ is a Carnot group, that is a simply connected nilpotent group endowed with a one-parameter group of dilatations. The intrinsic calculus on $N(G)$ is based on the notion of derivative introduced by Pansu. We shall say therefore that a function is smooth if it is Pansu derivable. The same denomination will be used for functions between Carnot groups.

During the process of establishing a non-Euclidean analysis problems begin to appear.

a) Right translations, the group operation and the inverse map are not smooth.

b) If $G \neq N(G)$ then generically there is no atlas of G over $N(G)$ with smooth transition of charts such that the group exponential is smooth. A noticeable exception is the contact case, that is the case where $N(G)$ is a Heisenberg group.

c) Mostow-Margulis (or any other known) construction of a tangent bundle TG of G do not provide a group structure to TG.

d) there is no notion of higher order derivatives (except the horizontal classical derivatives which have no direct intrinsic meaning).

We would like to have a notion of sub-Riemannian Lie group such that a), b), c) d) are solved. This is obviously a problem of good choice of "smoothness".

As a first attempt we present the formalism of uniform groups. We then concentrate on the typical example of a Lie group with left invariant distribution and give several notions of tangent bundle and smoothness, all of them equivalent in the "commutative" case $D = \mathfrak{g}$. We exemplify our constructions in the case of Carnot groups and in the case of compact part of real split groups.

2 Overview of sub-Riemannian manifolds

The name "sub-Riemannian manifold" is a compromise in the actual state of development of non-Euclidean analysis because it contains the word "manifold". This denomination is therefore a bit misleading, because one does not have, up to the moment, a satisfactory notion of manifold, the Euclidean case excepted. Maybe a better name would by "model of sub-Riemannian manifold", which is rather long and equally unclear at the moment.

Definition 2.1 *A sub-Riemannian (SR) manifold is a triple (M, H, g), where M is a connected manifold, H is a subbundle of TM, named horizontal bundle or distribution, and g is a metric (Euclidean inner-product) on the horizontal bundle.*

A horizontal curve is a continuous, almost everywhere differentiable curve, whose tangents lie in the horizontal bundle.

The length of a horizontal curve $c : [a, b] \to M$ is

$$l(c) = \int_a^b \sqrt{g_{c(t)}(\dot{c}(t), \dot{c}(t))} \ dt$$

The SR manifold is called a Carnot-Carathéodory (CC) space if any two points can be joined by a finite length horizontal curve.

A CC space is a path metric space with the Carnot-Carathéodory distance induced by the length l:

$$d(x, y) = \inf \{l(c) : c : [a, b] \to M , \ c(a) = x , \ c(b) = y\}$$

In the class of sub-Riemannian manifolds there is the distinguished subclass of regular ones. Regularity is a notion which concerns only the distribution D. Let us explain it.

Given the distribution D, a point $x \in M$ and a sufficiently small open neighbourhood $x \in U \subset M$, one can define on U a filtration of bundles as follows. Define first the class of horizontal vectorfields on U:

$$\mathcal{X}^1(U, D) = \{X \in \Gamma^\infty(TU) : \forall y \in U , \ X(u) \in D_y\}$$

Next, define inductively for all positive integers k:

$$\mathcal{X}^{k+1}(U, D) = \mathcal{X}^k(U, D) \cup [\mathcal{X}^1(U, D), \mathcal{X}^k(U, D)]$$

Here $[\cdot, \cdot]$ denotes vectorfields bracket. We obtain therefore a filtration $\mathcal{X}^k(U, D) \subset \mathcal{X}^{k+1}(U, D)$. Evaluate now this filtration at x:

$$V^k(x, U, D) = \left\{X(x) : X \in \mathcal{X}^k(U, D)\right\}$$

Because the distribution was supposed to be completely non-integrable, there are $m(x)$, positive integer, and small enough U such that $V^k(x, U, D) = V^k(x, D)$ for all $k \leq m$ and

$$D_x V^1(x, D) \subset V^2(x, D) \subset ... \subset V^{m(x)}(x, D) = T_x M$$

We equally have

$$\nu_1(x) = \dim V^1(x, D) < \nu_2(x) = \dim V^2(x, D) < ... < n = \dim M$$

Off course, $\nu_1(x)$ is constant over M, but generally $m(x)$, $\nu_k(x)$ may vary from a point to another.

Definition 2.2 *The distribution D is regular if $m(x)$, $\nu_k(x)$ are constant on the manifold M.*

For example a contact manifold and a Lie group with left invariant nonintegrable distribution are examples of sub-Riemannian spaces with regular distribution. In contrast, the Grushin plane distribution is non regular.

The particular case that will be interesting for us is the following one. Let G be a real connected Lie group with Lie algebra $\mathfrak{g}$ and $D \subset \mathfrak{g}$ a vector space which generates the algebra. This means that the sequence

$$V_1 \ = \ D \ , \ V^{i+1} \ = \ V^i + [D, V^i]$$

provides a filtration of $\mathfrak{g}$:

$$V^1 \ \subset \ V^2 \ \subset \ ... V^m \ = \ \mathfrak{g} \tag{2.0.1}$$

We set the distribution induced by D to be

$$D_x \ = \ TL_x D$$

where $x \in G$ is arbitrary and $L_x : G \to G$, $L_x y \ = \ xy$ is the left translation by x. We shall use the same notation D for the induced distribution.

This distribution is non-integrable (easy form of Chow theorem). For any (left or right invariant) metric defined on D we have an associated Carnot-Carathéodory distance.

2.1 Carnot groups. Pansu derivative. Mitchell's theorem 2

Carnot groups are particular examples of sub-Riemannian manifolds. They are especially important because they provide infinitesimal models for any sub-Riemannian manifold. Moreover, all the fundamental results of sub-Riemannian geometry are particularly easy to prove and understand in the case of Carnot groups.

Structure of Carnot groups. Mitchell's theorem 2

Definition 2.3 *A Carnot (or stratified nilpotent) group is a connected simply connected group N with a distinguished vectorspace V_1 such that the Lie algebra of the group has the direct sum decomposition:*

$$n \ = \ \sum_{i=1}^{m} V_i \ , \quad V_{i+1} \ = \ [V_1, V_i]$$

The number m is the step of the group. The number

$$Q \; = \; \sum_{i=1}^{m} i \; dim V_i$$

is called the homogeneous dimension of the group.

Because the group is nilpotent and simply connected, the exponential mapping is a diffeomorphism. We shall identify the group with the algebra, if is not locally otherwise stated.

The structure that we obtain is a set N endowed with a Lie bracket and a group multiplication operation given by the Baker-Campbell-Hausdorff formula.

This formula has only a finite number of terms, because the algebra is nilpotent. This shows that the group operation is polynomial. It is easy to see that the Lebesgue measure on the algebra is (by identification of the algebra with the group) a bi-invariant measure.

We give further examples of such groups:

(1.) R^n with addition is the only commutative Carnot group. It generates the classical, or Euclidean, or commutative calculus.

(2.) The Heisenberg group is the first non-trivial example. This is the group $H(n) \; = \; R^{2n} \times R$ with the operation:

$$(x, \bar{x})(y, \bar{y}) \; = \; (x + y, \bar{x} + \bar{y} + \frac{1}{2}\omega(x, y))$$

where ω is the standard symplectic form on R^{2n}. The Lie bracket is

$$[(x, \bar{x}), (y, \bar{y})] \; = \; (0, \omega(x, y))$$

The direct sum decomposition of (the algebra of the) group is:

$$H(n) \; = \; V + Z \; , \quad V \; = \; R^{2n} \times \{0\} \; , \quad Z \; = \; \{0\} \times R$$

Z is the center of the algebra, the group has step 2 and homogeneous dimension $2n+2$. This group generates the symplectic calculus, as shown in section 5.

(3.) H-type groups. These are two step nilpotent Lie groups N endowed with an inner product $(\cdot, \cdot)$, such that the following *orthogonal* direct sum decomposition occurs:

$$N \; = \; V + Z$$

Z is the center of the Lie algebra. Define now the function

$$J : Z \to End(V) \; , \quad (J_z x, x') \; = \; (z, [x, x'])$$

The group N is of H-type if for any $z \in Z$ we have

$$J_z \circ J_z \; = \; - \mid z \mid^2 \; I$$

From the Baker-Campbell-Hausdorff formula we see that the group operation is

$$(x, z)(x', z') \;=\; (x + x', z + z' + \frac{1}{2}[x, x'])$$

These groups appear naturally as the nilpotent part in the Iwasawa decomposition of a semisimple real group of rank one. (see [9])

(**4.**) The last example is the group of $n \times n$ upper triangular matrices, which is nilpotent of step $n - 1$. This example is important because any Carnot group is isomorphic with a subgroup of a group of upper triangular matrices. However, it is not generally true that given a Carnot group N with a compatible familly of dilatations, there is an injective morphism from N to a group of upper triangular matrices which commutes with dilatations (on N we have the chosen dilatations and in the group of upper triangular matrices we have canonical dilatations).

Any Carnot group admits a one-parameter family of dilatations. For any $\varepsilon > 0$, the associated dilatation is:

$$x \;=\; \sum_{i=1}^{m} x_i \;\mapsto\; \delta_\varepsilon x \;=\; \sum_{i=1}^{m} \varepsilon^i x_i$$

Any such dilatation is a group morphism and a Lie algebra morphism.

In fact the class of Carnot groups is characterised by the existence of dilatations.

Proposition 2.4 *Suppose that the Lie algebra $\mathfrak{g}$ admits an one parameter group $\varepsilon \in (0, +\infty) \mapsto \delta_\varepsilon$ of simultaneously diagonalisable Lie algebra isomorphisms. Then $\mathfrak{g}$ is the algebra of a Carnot group.*

We can always find Euclidean inner products on N such that the decomposition $N = \sum_{i=1}^{m} V_i$ is an orthogonal sum. Let us pick such an inner product and denote by $\| \cdot \|$ the Euclidean norm associated to it.

We shall endow the group N with a structure of a sub-Riemannian manifold now. For this take the distribution obtained from left translates of the space V_1. The metric on that distribution is obtained by left translation of the inner product restricted to V_1.

If V_1 Lie generates (the algebra) N then any element $x \in N$ can be written as a product of elements from V_1. A slight reformulation of Lemma 1.40, Folland, Stein [11]:

Lemma 2.5 *Let N be a Carnot group and $X_1, ..., X_p$ an orthonormal basis for V_1. Then there is a a natural number M and a function $g : \{1, ..., M\} \to \{1, ..., p\}$ such that any element $x \in N$ can be written as:*

$$x \;=\; \prod_{i=1}^{M} \exp(t_i X_{g(i)}) \tag{2.1.2}$$

Moreover, if x is sufficiently close (in Euclidean norm) to 0 then each t_i can be chosen such that $\mid t_i \mid \leq C \|x\|^{1/m}$.

This means that there is a horizontal curve joining any two points. Hence the distance

$$d(x, y) \;=\; \inf \left\{ \int_a^b \| c^{-1} \dot{c} \| \; \mathrm{d}t \;\; : \;\; c(a) = x, \; c(b) = y, \; c^{-1} \dot{c} \in V_1 \right\}$$

is finite for any two $x, y \in N$. The distance is obviously left invariant.

Associate to the CC distance d the function:

$$\mid x \mid_d \;=\; d(0, x)$$

The map $\mid \cdot \mid_d$ looks like a norm. However it is intrinsically defined and hard to work with. This is the reason for the introduction of homogeneous norms.

Definition 2.6 *A continuous function from $x \mapsto \mid x \mid$ from G to $[0, +\infty)$ is a homogeneous norm if*

(a) *the set $\{x \in G \; : \; \mid x \mid = 1\}$ does not contain 0.*

(b) *$\mid x^{-1} \mid = \mid x \mid$ for any $x \in G$.*

(c) *$\mid \delta_\varepsilon x \mid = \varepsilon \mid x \mid$ for any $x \in G$ and $\varepsilon > 0$.*

Homogeneous norms exist. $\mid \cdot \mid_d$ is a homogeneous norm. For any $p \in [0, +\infty)$ a homogeneous norm is:

$$\mid x \mid_p \;=\; \left[\sum_{i=1}^m \mid x_i \mid^{p/i} \right]^{1/p}$$

For $p = +\infty$ the corresponding norm is

$$\mid x \mid_\infty \;=\; \max \left\{ \mid x_i \mid^{1/i} \right\}$$

Proposition 2.7 *Any two homogeneous norms are equivalent. Let $\mid \cdot \mid$ be a homogeneous norm. Then the set $\{x \; : \; \mid x \mid = 1\}$ is compact.*

There is a constant $C > 0$ such that for any $x, y \in N$ we have:

$$\mid xy \mid \;\leq\; C \left(\mid x \mid + \mid y \mid \right)$$

There are two important consequences of this simple proposition. The first is a description of the behaviour of any homogeneous norm. The following proposition implies also that the Euclidean topology and uniformity are the same as the topology and uniformity induced by the Carnot-Carathéodory distance.

Proposition 2.8 *Let $\mid \cdot \mid$ be a homogeneous norm. Then there are constants $c, C > 0$ such that for any $x \in N$, $\mid x \mid < 1$, we have*

$$c \| x \| \;\leq\; \mid x \mid \;\leq\; C \| x \|^{1/m}$$

The second consequence of proposition 2.7 says that the balls in CC-distance look roughly like boxes (as it is the case with the Euclidean balls). A box is a set

$$Box(r) \;=\; \left\{ x \;=\; \sum_{i=1}^{m} x_i \;:\; \|x_i\| \le r^i \right\}$$

Proposition 2.9 *("Ball-Box theorem") There are positive constants c, C such that for any $r > 0$ we have:*

$$Box(cr) \;\subset\; B(0,r) \;\subset\; Box(Cr)$$

As a consequence the Lebesgue measure is absolutely continuous with respect to the Hausdorff measure $\mathcal{H}^Q$. Because of the invariance with respect to the group operation it follows that the Lebesgue measure is a multiple of the mentioned Hausdorff measure.

The following theorem is Theorem 2. from Mitchell [24], in the particular case of Carnot groups.

Theorem 2.10 *The ball $B(0,1)$ has Hausdorff dimension Q.*

Proof.   We know that the volume of a ball with radius ε is $c\varepsilon^Q$.

Consider a maximal filling of $B(0,1)$ with balls of radius ε. There are N_ε such balls in the filling; an upper bound for this number is:

$$N_\varepsilon \;\le\; 1/\varepsilon^Q$$

The set of concentric balls of radius 2ε cover $B(0,1)$; each of these balls has diameter smaller than 4ε, so the Hausdorff α measure of $B(0,1)$ is smaller than

$$\lim_{\varepsilon \to 0} N_\varepsilon (2\varepsilon)^\alpha$$

which is 0 if $\alpha > Q$. Therefore the Hausdorff dimension is smaller than Q.

Conversely, given any covering of $B(0,1)$ by sets of diameter $\le \varepsilon$, there is an associated covering with balls of the same diameter; the number M_ε of this balls has a lower bound:

$$M_\varepsilon \;\ge\; 1/\varepsilon^Q$$

thus there is a lower bound

$$\sum_{cover} \varepsilon^\alpha \;\ge\; \varepsilon^\alpha/\varepsilon^Q$$

which shows that if $\alpha < Q$ then $\mathcal{H}^\alpha(B(0,1)) = \infty$. Therefore the Hausdorff dimension of the ball is greater than Q.   ■

We collect the important facts discovered until now: let N be a Carnot group endowed with the left invariant distribution generated by V_1 and with an Euclidean norm on V_1.

(a) If V_1 Lie-generates the whole Lie algebra of N then any two points can be joined by a horizontal path.

(b) The metric topology and uniformity of N are the same as Euclidean topology and uniformity respective.

(c) The ball $B(0, r)$ looks roughly like the box $\left\{ x = \sum_{i=1}^{m} x_i \ : \ \|x_i\| \leq r^i \right\}$.

(d) the Hausdorff measure $\mathcal{H}^Q$ is group invariant and the Hausdorff dimension of a ball is Q.

(e) there is a one-parameter group of dilatations, where a dilatation is an isomorphism δ_ε of N which transforms the distance d in εd.

All these facts are true (under slight and almost obvious modifications) in the case of a sub-Riemannian manifold.

Pansu derivative. Rademacher theorem   A Carnot group has it's own concept of differentiability, introduced by Pansu [25].

In Euclidean spaces, given $f : R^n \to R^m$ and a fixed point $x \in R^n$, one considers the difference function:

$$X \in B(0,1) \subset R^n \ \mapsto \ \frac{f(x + tX) - f(x)}{t} \in R^m$$

The convergence of the difference function as $t \to 0$ in the uniform convergence gives rise to the concept of differentiability in it's classical sense. The same convergence, but in measure, leads to approximate differentiability. This and another topologies might be considered (see Vodop'yanov [30], [31]).

In the frame of Carnot groups the difference function can be written using only dilatations and the group operation. Indeed, for any function between Carnot groups $f : G \to P$, for any fixed point $x \in G$ and $\varepsilon > 0$ the finite difference function is defined by the formula:

$$X \in B(1) \subset G \ \mapsto \ \delta_\varepsilon^{-1} \left(f(x)^{-1} f \left(x \delta_\varepsilon X \right) \right) \in P$$

In the expression of the finite difference function enters δ_ε^{-1} and δ_ε, which are dilatations in P, respectively G.

Pansu's differentiability is obtained from uniform convergence of the difference function when $\varepsilon \to 0$.

The derivative of a function $f : G \to P$ is linear in the sense explained further. For simplicity we shall consider only the case $G = P$. In this way we don't have to use a heavy notation for the dilatations.

Definition 2.11 *Let N be a Carnot group. The function $F : N \to N$ is linear if*

(a) F is a group morphism,

(b) for any $\varepsilon > 0$ $F \circ \delta_\varepsilon = \delta_\varepsilon \circ F$.

We shall denote by $HL(N)$ the group of invertible linear maps of N, called the linear group of N.

The condition (b) means that F, seen as an algebra morphism, preserves the grading of N.

The definition of Pansu differentiability follows:

Definition 2.12 *Let $f : N \to N$ and $x \in N$. We say that f is (Pansu) differentiable in the point x if there is a linear function $Df(x) : N \to N$ such that*

$$\sup \{ d(F_\varepsilon(y), Df(x)y) \ : \ y \in B(0,1) \}$$

converges to 0 when $\varepsilon \to 0$. The functions F_ε are the finite difference functions, defined by

$$F_t(y) \ = \ \delta_t^{-1} \left(f(x)^{-1} f(x \delta_t y) \right)$$

The definition says that f is differentiable at x if the sequence of finite differences F_t uniformly converges to a linear map when t tends to 0.

We are interested to see how this differential looks like. For any $f : N \to N$, $x, y \in N$ $Df(x)y$ means

$$Df(x)y \ = \ \lim_{t \to 0} \delta_t^{-1} \left(f(x)^{-1} f(x \delta_t y) \right)$$

provided that the limit exists.

Proposition 2.13 *Let $f : N \to N$, $y, z \in N$ such that:*

(i) $Df(x)y$, $Df(x)z$ exist for any $x \in N$.

(ii) The map $x \mapsto Df(x)z$ is continuous.

(iii) $x \mapsto \delta_t^{-1} \left(f(x)^{-1} f(x \delta_t z) \right)$ converges uniformly to $Df(x)z$.

Then for any $a, b > 0$ and any $w \ = \ \delta_a x \delta_b y$ the limit $Df(x)w$ exists and we have:

$$Df(x)\delta_a x \delta_b y \ = \ \delta_a Df(x)y \ \delta_b Df(x)z$$

If f is Lipschitz then the previous proposition holds almost everywhere. In Proposition 3.2, Pansu [25] there is a more general statement:

Proposition 2.14 *Let $f : N \to N$ have finite dilatation and suppose that for almost any $x \in N$ the limits $Df(x)y$ and $Df(x)z$ exist. Then for almost any x and for any $w \ = \ \delta_a y \ \delta_b z$ the limit $Df(x)w$ exists and we have:*

$$Df(x)\delta_a x \delta_b y \ = \ \delta_a Df(x)y \ \delta_b Df(x)z$$

Proof. The idea is to use Proposition 2.13. We shall suppose first that f is Lipschitz. Then any finite difference function F_t is also Lipschitz.

Recall the general Egorov theorem:

Lemma 2.15 *(Egorov Theorem) Let μ be a finite measure on X and $(f_n)_n$ a sequence of measurable functions which converges μ almost everywhere (a.e.) to f. Then for any $\varepsilon > 0$ there exists a measurable set X_ε such that $\mu(X \setminus X_\varepsilon) < \varepsilon$ and f_n converges uniformly to f on X_ε.*

Proof. It is not restrictive to suppose that f_n converges pointwise to f. Define, for any $q, p \in \mathbf{N}$, the set:

$$X_{q,p} = \{x \in X : |f_n(x) - f(x)| < 1/p, \forall n \geq q\}$$

For fixed p the sequence of sets $X_{q,p}$ is increasing and at the limit it fills the space:

$$\bigcup_{q \in \mathbf{N}} X_{q,p} = X$$

Therefore for any $\varepsilon > 0$ and any p there is a $q(p)$ such that $\mu(X \setminus X_{q(p)p}) < \varepsilon/2^p$.

Define then

$$X_\varepsilon = \bigcap_{p \in N} X_{q(p)p}$$

and check that it satisfies the conclusion. ■

If we could improve the conclusion of Egorov Theorem by claiming that if μ is a Borel measure then each X_ε can be chosen to be open, then the proof would resume like this.

We can take X to be a ball in the group N and μ to be the Q Hausdorff measure, which is Borel. Then we are able to apply Proposition 2.13 on X_ε. When we tend ε to 0 we obtain the claim (for f Lipschitz).

Unfortunately, such an improvement in the conclusion of Egorov theorem does not hold. In fact, in Pansu paper, during the proof of his Rademacher theorem, there is a missing step, as Vodop'yanov remarked. This missing step describes precisely what happens when X_ε (a Borel measurable set) is replaced by an arbitrarily fine open approximation X'_ε. For this especially subtle step see Vodop'yanov in same proceedings [...]. ■

A consequence of this proposition is Corollary 3.3, Pansu [25].

Corollary 2.16 *If $f : N \to N$ has finite dilatation and $X_1, ..., X_K$ is a basis for the distribution V_1 such that*

$$s \mapsto f(x\delta_s X_i)$$

is differentiable in $s = 0$ for almost any $x \in N$, then f is differentiable almost everywhere and the differential is linear, i.e. if

$$y = \prod_{i=1}^{M} \delta_{t_i} X_{g(i)}$$

then

$$Df(x)y \;=\; \prod_{i=1}^{M} \delta_{t_i} X_{g(i)}(f)(x)$$

A very important result is theorem 2, Pansu [25], which contains the Rademacher theorem for Carnot groups.

Theorem 2.17 *Let $f : M \to N$ be a Lipschitz function between Carnot groups. Then f is differentiable almost everywhere.*

This is the kind of result that motivates the study of sub-Riemannian manifolds as metric spaces endowed with a measure. It is to be noticed that in the Rademacher theorem enters differentiability, measure and distance. This theorem is indeed a concentrate of fundamental facts from analysis and measure theory, hence its importance.

The proof is based on the corollary 2.16 and the technique of development of a curve, Pansu sections 4.3 - 4.6 [25].

Let $c : [0,1] \to N$ be a Lipschitz curve such that $c(0) = 0$. To any division $\Sigma : 0 = t_0 < .... < t_n = 1$ of the interval $[0,1]$ is a associated the element $\sigma_\Sigma \in N$ (the algebra) given by:

$$\sigma_\Sigma \;=\; \sum_{k=0}^{n} c(t_k)^{-1} c(t_{k+1})$$

Lemma (18) Pansu [26] implies the existence of a constant $C > 0$ such that

$$\|\sigma_\Sigma \;-\; c(1)\| \;\leq\; C \left(\sum_{k=0}^{n} \|c(t_k)^{-1} c(t_{k+1})\| \right)^2 \tag{2.1.3}$$

Take now a finer division Σ' and look at the interval $[t_k, t_{k+1}]$ divided further by Σ' like this:

$$t_k = t'_l < ... < t'_m = t_{k+1}$$

The estimate (2.1.3) applied for each interval $[t_k, t_{k+1}]$ lead us to the inequality:

$$\|\sigma_{\Sigma'} \;-\; \sigma_\Sigma\| \;\leq\; \sum_{k=0}^{n} d(c(t_k), c(t_{k+1}))^2$$

The curve has finite length (being Lipschitz) therefore the right hand side of the previous inequality tends to 0 with the norm of the division. Set

$$\sigma(s) \;=\; \lim_{\|\Sigma\| \to 0} \sigma_\Sigma(s)$$

where $\sigma_\Sigma(s)$ is relative to the curve c restricted to the interval $[0, s]$. The curve such defined is called the development of the curve c. It is easy to see that σ has the same length (measured with the Euclidean norm $\| \cdot \|$) as c. If we parametrise c with the length then we have the estimate:

$$\|\sigma(s) \; c(s)\| \;\leq\; Cs^2 \tag{2.1.4}$$

This shows that σ is a Lipschitz curve (with respect to the Euclidean distance). Indeed, prove first that σ has finite dilatation in almost any point, using (2.1.4) and the fact that c is Lipschitz. Then show that the dilatation is majorised by the Lipschitz constant of c. By the classical Rademacher theorem σ is almost everywhere derivable.

Remark 2.18 *The development of a curve can be done in an arbitrary Lie connected group, endowed with a left invariant distribution which generates the algebra. One should add some logarithms, because in the case of Carnot groups, we have identified the group with the algebra. The inequality (2.1.4) still holds.*

Conversely, given a curve σ in the algebra N, we can perform the inverse operation to development (called multiplicative integral by Pansu; we shall call it "lift"). Indeed, to any division Σ of the interval $[0, s]$ we associate the point

$$c_\Sigma(s) = \prod_{k=0}^{n} (\sigma(t_{k+1}) - \sigma(t_k))$$

Define then

$$c(s) = \lim_{\|\Sigma\| \to 0} c_\Sigma(s)$$

Remark that if $\sigma([0, 1]) \subset V_1$ and it is almost everywhere differentiable then c is a horizontal curve.

Fix $s \in [0, 1)$ and parametrise c by the length. Apply the inequality (2.1.4) to the curve $t \mapsto c(s)^{-1}c(s + t)$. We get the fact that the vertical part of $\sigma(s + t) - \sigma(s)$ is controlled by s^2. Therefore, if c is Lipschitz then σ is included in V_1.

Denote the i multiple bracket $[x, [x, ...[x, y]...]$ by $[x, y]_i$. For any division Σ of the interval $[0, s]$ set:

$$A_\Sigma^i(s) = \sum_{k=0}^{n} [\sigma(t_k), \sigma(t_{k+1})]_i$$

As before, one can show that when the norm of the division tends to 0, A_Σ^i converges. Denote by

$$A^i(s) = \int_0^s [\sigma, d\sigma]_i = \lim_{\|\Sigma\| \to 0} A_\Sigma^i(s)$$

the i-area function. The estimate corresponding to (2.1.4) is

$$\|A^i(s)\| \leq C s^{i+1} \tag{2.1.5}$$

What is the significance of $A^i(s)$? The answer is simple, based on the well known formula of derivation of left translations in a group. Define, in a neighbourhood of 0 in the Lie algebra $\mathfrak{g}$ of the Lie group G, the operation

$$X \overset{\mathfrak{g}}{\cdot} Y = \log_G (\exp_G(X) \exp_G(Y))$$

The left translation by X is the function $L_X(Y) = X \overset{\mathfrak{g}}{\cdot} Y$. It is known that

$$DL_X(0)(Z) = \sum_{i=0} \frac{1}{(i+1)!} [X, Z]_i$$

It follows that if the group G is Carnot then $[X, Z]_i$ measures the infinitesimal variation of the V_i component of $L_X(Y)$, for $Y = 0$. Otherwise said,

$$\frac{d}{dt}\sigma_i(t) \;=\; \frac{1}{(i+1)!}\lim_{\varepsilon \to 0}\varepsilon^{-i}\int_t^{t+\varepsilon}[\sigma, d\sigma]_i$$

Because σ is Lipschitz, the left hand side exists for all i for a.e. t. The estimate (2.1.5) tells us that the right hand side equals 0. This implies the following proposition (Pansu, 4.1 [25]).

Proposition 2.19 *If c is Lipschitz then c is differentiable almost everywhere.*

Proof of theorem 2.17.   The proposition 2.19 implies that we are in the hypothesis of corollary 2.16.   ■

In order to prove Stepanov theorem, the Pansu-Rademacher theorem has been improved for Lipschitz functions $f : A \subset M \to N$, where M, N are Carnot groups and A is just measurable, in Vodop'yanov, Ukhlov [34].

Their technique differs from Pansu. It resembles with the one used in Margulis, Mostow [22], where it is proven that any quasi-conformal map from a Carnot-Carathéodory manifold to another is a.e. differentiable, without using Rademacher theorem. Same result for quasi-conformal maps on Carnot groups has been first proved by Koranyi, Reimann [19]. Finally Magnani [21] reproved a.e. differentiability of Lipschitz functions on Carnot groups, defined on measurable sets, continuing Pansu technique.

For a review of other connected results, such as approximate differentiability, differentiability in Sobolev topology, see the excellent Vodop'yanov papers [30], [31].

2.2 Nilpotentisation of a sub-Riemannian manifold

In this section we return to the case of a connected real Lie group endowed with a nonintegrable left invariant distribution. We shall associate to such a sub-Riemannian space a Carnot group, using the procedure of nilpotentisation.

The filtration (2.0.1) has the straightforward property: if $x \in V^i$ and $y \in V^j$ then $[x, y] \in V^{i+j}$ (where $V^k = \mathfrak{g}$ for all $k \geq m$ and $V^0 = \{0\}$). This allows to construct the Lie algebra

$$\mathfrak{n}(\mathfrak{g}, D) \;=\; \oplus_{i=1}^m V_i \;\;,\;\; V_i \;=\; V^i/V^{i-1}$$

with Lie bracket $[\hat{x}, \hat{y}]_n \;=\; \widehat{[x, y]}$, where $\hat{x} = x + V^{i-1}$ if $x \in V^i \setminus V^{i-1}$.

Proposition 2.20 $\mathfrak{n}(\mathfrak{g}, D)$ *with distinguished space* $V_1 \;=\; D$ *is the Lie algebra of a Carnot group of step m. It is called the nilpotentisation of the filtration* (2.0.1).

Remark that $\mathfrak{n}(\mathfrak{g}, D)$ and $\mathfrak{g}$ have the same dimension, therefore they are isomorphic as vector spaces.

Let $\{X_1, ..., X_p\}$ be a basis of the vector space D. We shall build a basis of $\mathfrak{g}$ which will give a vector spaces isomorphism with $\mathfrak{n}(\mathfrak{g}, D)$.

A word with letters $A = \{X_1, ..., X_p\}$ is a string $X_{h(1)}...X_{h(s)}$ where $h : \{1, ..., s\} \rightarrow \{1, ..., p\}$. The set of words forms the dictionary $Dict(A)$, ordered lexicographically. We set the function $Bracket : Dict(A) \rightarrow \mathfrak{g}$ to be

$$Bracket(X_{h(1)}...X_{h(s)}) = [X_{h(1)}, [X_{h(2)}, [..., X_{h(s)}]...]$$

For any $x \in Bracket(Dict(A))$ let $\hat{x} \in Dict(A)$ be the least word such that $Bracket(\hat{x}) = x$. The collection of all these words is denoted by $\hat{g}$.

The length $l(x) = length(\hat{x})$ is well defined. The dictionary $Dict(A)$ admit a filtration made by the length of words function. In the same way the function l gives the filtration

$$V^1 \cap Bracket(Dict(A)) \subset V^2 \cap Bracket(Dict(A)) \subset ... Bracket(Dict(A))$$

Choose now, in the lexicographic order in $\hat{g}$, a set $\hat{B}$ such that $B = Bracket(\hat{B})$ is a basis for $\mathfrak{g}$. Any element X in this basis can be written as

$$X = [X_{h(1)}, [X_{h(2)}, [..., X_{h(s)}]...]$$

such that $l(X) = s$ or equivalently $X \in V^s \setminus V^{s-1}$.

It is obvious that the map

$$X \in B \mapsto \tilde{X} = X + V^{l(X)-1} \in \mathfrak{n}(\mathfrak{g}, D)$$

is a bijection and that $\tilde{B} = \left\{ \tilde{X}_j : j = 1, ..., dim\ \mathfrak{g} \right\}$ is a basis for $\mathfrak{n}(\mathfrak{g}, D)$. We can identify then $\mathfrak{g}$ with $\mathfrak{n}(\mathfrak{g}, D)$ by the identification $X_j = \tilde{X}_j$.

Equivalently we can define the nilpotent Lie bracket $[\cdot, \cdot]_n$ directly on $\mathfrak{g}$, with the use of the dilatations on $\mathfrak{n}(\mathfrak{g}, D)$.

Instead of the filtration (2.0.1) let us start with a direct sum decomposition of $\mathfrak{g}$

$$\mathfrak{g} = \oplus_{i=1}^m W_i , \quad V^i = \oplus_{j=1}^i W_j \tag{2.2.6}$$

such that $[V^i, V^j] \subset V^{i+j}$. The chain V^i form a filtration like (2.0.1).

We set

$$\delta_\varepsilon(x) = \sum_{i=1}^m \varepsilon^i x_i$$

for any $\varepsilon > 0$ and $x \in \mathfrak{g}$, which decomposes according to (2.2.6) as $x = \sum_{i=1}^m x_i$.

Proposition 2.21 *The limit*

$$[x, y]_n = \lim_{\varepsilon \to 0} \delta_\varepsilon^{-1}[\delta_\varepsilon x, \delta_\varepsilon y] \tag{2.2.7}$$

exists for any $x, y \in \mathfrak{g}$ and $(\mathfrak{g}, [\cdot, \cdot]_n)$ is the Lie algebra of a Carnot group with dilatations δ_ε.

Proof. Let $x = \sum_{j=1}^{i} x_j$ and $y = \sum_{k=1}^{l} y_k$. Then

$$[\delta_\varepsilon x, \delta - \varepsilon y] = \sum_{s=2}^{i+l} \varepsilon^s \sum_{j+k=s} \sum_{p=1}^{s} [x_j, y_k]_p$$

We apply δ_ε^{-1} to this equality and we obtain:

$$\delta_\varepsilon^{-1}[\delta_\varepsilon x, \delta_\varepsilon y] = \sum_{s=2}^{i+l} \sum_{j+k=s} \sum_{p=1}^{s} \varepsilon^{s-p}[x_j, y_k]_p$$

When ε tends to 0 the expression converges to the limit

$$[x, y]_n = \sum_{s=2}^{i+l} \sum_{j+k=s} [x_j, y_k]_s \tag{2.2.8}$$

For any $\varepsilon > 0$ the expression

$$[x, y]^\varepsilon = \delta_\varepsilon^{-1}[\delta_\varepsilon x, \delta_\varepsilon y]$$

is a Lie bracket (bilinear, antisymmetric and it satisfies the Jacobi identity). Therefore at the limit $\varepsilon \to 0$ we get a Lie bracket. Moreover, it is straightforward to see from the definition of $[x, y]_n$ that δ_ε is an algebra isomorphism. We conclude that $(\mathfrak{g}, [\cdot, \cdot]_n)$ is the Lie algebra of a Carnot group with dilatations δ_ε. ∎

Proposition 2.22 *Let $\{X_1, ..., X_{dim\ \mathfrak{g}}\}$ be a basis of $\mathfrak{g}$ constructed from a basis $\{X_1, ..., X_p\}$ of D. Set*

$$W_j = span\ \{X_i : l(X_i) = j\}$$

Then the spaces W_j provides a direct sum decomposition (2.2.6). Moreover the identification $\hat{X}_i = X_i$ gives a Lie algebra isomorphism between $(\mathfrak{g}, [\cdot, \cdot]_n)$ and $\mathfrak{n}(\mathfrak{g}, D)$.

Proof. In the basis $\{X_1, ..., X_{dim\ \mathfrak{g}}\}$ the Lie bracket on $\mathfrak{g}$ looks like this:

$$[X_i, X_j] = \sum C_{ijk} X_k$$

where $c_{ijk} = 0$ if $l(X_i) + l(X_j) < l(X_k)$. From here the first part of the proposition is straightforward. The expression of the Lie bracket generated by the decomposition (2.2.6) is obtained from (2.2.8). We have

$$[X_i, X_j]_n = \sum \lambda_{ijk} C_{ijk} X_k$$

where $\lambda_{ikj} = 1$ if $l(X_i) + l(X_j) = l(X_k)$ and 0 otherwise. The Lie algebra isomorphism follows from the expression of the Lie bracket on $\mathfrak{n}(\mathfrak{g}, D)$:

$$[\hat{X}_i, \hat{X}_j] = [X_i, X_j] + V^{l(X_i)+l(X_j)-1}$$

∎

In conclusion the expression of the nilpotent Lie bracket depends on the choice of basis B trough the transport of the dilatations group from $\mathfrak{n}(\mathfrak{g}, D)$ to $\mathfrak{g}$.

Let $N(G, D)$ be the simply connected Lie group with Lie algebra $\mathfrak{n}(\mathfrak{g}, D)$. As previously, we identify $N(G, D)$ with $\mathfrak{n}(\mathfrak{g}, D)$ by the exponential map.

2.3   Gromov-Hausdorff convergence and metric tangent cones

This section contains the general notions and results which allows to say what a tangent space to a metric space is.

Distances between metric spaces   The references for this section are Gromov [16], chapter 3, and Burago & al. [6] section 7.4. There are several definitions of distances between metric spaces. The very fertile idea of introducing such distances belongs to Gromov.

In order to introduce the Hausdorff distance between metric spaces, recall the Hausdorff distance between subsets of a metric space.

Definition 2.23 *For any set $A \subset X$ of a metric space and any $\varepsilon > 0$ set the ε neighbourhood of A to be*

$$A_\varepsilon \; = \; \cup_{x \in A} B(x, \varepsilon)$$

The Hausdorff distance between $A, B \subset X$ is defined as

$$d_H^X(A, B) \; = \; \inf \left\{ \varepsilon > 0 : A \subset B_\varepsilon \; , \; B \subset A_\varepsilon \right\}$$

By considering all isometric embeddings of two metric spaces X, Y into an arbitrary metric space Z we obtain the Hausdorff distance between X, Y (Gromov [16] definition 3.4).

Definition 2.24 *The Hausdorff distance $d_H(X, Y)$ between metric spaces X Y is the infimum of the numbers*

$$d_H^Z(f(X), g(Y))$$

for all isometric embeddings $f : X \to Z$, $g : Y \to Z$ in a metric space Z.

If X, Y are compact then $d_H(X, Y) < +\infty$. Indeed, let Z be the disjoint union of X, Y and $M \; = \; \max \left\{ diam(X), diam(Y) \right\}$. Define the distance on Z to be

$$d^Z(x, y) \; = \; \begin{cases} d^X(x, y) & x, y \in X \\ d^Y(x, y) & x, y \in Y \\ \frac{1}{2} M & \text{otherwise} \end{cases}$$

Then $d_H^Z(X, Y) < +\infty$.

The Hausdorff distance between isometric spaces equals 0. The converse is also true (Gromov *op. cit.* proposition 3.6) in the class of compact metric spaces.

Theorem 2.25 *If X, Y are compact metric spaces such that $d_H(X, Y) = 0$ then X, Y are isometric.*

For the proof of the theorem we need the Lipschitz distance (*op. cit. definition 3.1*) and a criterion for convergence of metric spaces in the Hausdorff distance (*op. cit. proposition 3.5*). We shall give the definition of Gromov for Lipschitz distance and the first part of the mentioned proposition.

Definition 2.26 *The Lipschitz distance $d_L(X, Y)$ between bi-Lipschitz homeomorphic metric spaces X, Y is the infimum of*

$$| \log \ dil(f) \ | + | \log \ dil(f^{-1}) \ |$$

for all $f : X \to Y$, bi-Lipschitz homeomorphisms.

Obviously, if $d_L(X, Y) = 0$ then X, Y are isometric. Indeed, by definition we have a sequence $f_n : X \to Y$ such that $dil(f_n) \to 1$ as $n \to \infty$. Extract an uniformly convergent subsequence; the limit is an isometry.

Definition 2.27 *A ε-net in the metric space X is a set $P \subset X$ such that $P_\varepsilon = X$. The separation of the net P is*

$$sep(P) \ = \ \inf \{d(x, y) : x \neq y \ , \ x, y \in P\}$$

A ε-isometry between X and Y is a function $f : X \to Y$ such that dis $f \ \leq \varepsilon$ and $f(X)$ is a ε net in Y.

The following proposition gives a connection between convergence of metric spaces in the Hausdorff distance and convergence of ε nets in the Lipschitz distance.

Definition 2.28 *A sequence of metric spaces (X_i) converges in the sense of Gromov-Hausdorff to the metric space X if $d_H(X, X_i) \to 0$ as $i \to \infty$. This means that there is a sequence $\eta_i \to 0$ and isometric embeddings $f_i : X \to Z_i$, $g_i : X_i \to Z_i$ such that*

$$d_H^{Z_i}(f_i(X), g_i(X_i)) \ < \ \eta_i$$

The next proposition is corollary 7.3.28 (b), Burago & al. [6]. The proof is adapted from Gromov, proof of proposition 3.5 (b).

Proposition 2.29 *If there exists a ε-isometry between X, Y then $d_H(X, Y) < 2\varepsilon$.*

Proof. Let $f : X \to Y$ be the ε-isometry. On the disjoint union $Z = X \cup Y$ extend the distances d^X, d^Y in the following way. Define the distance between $x \in X$ and $y \in Y$ by

$$d(x, y) \ = \ \inf \left\{ d^X(x, u) + d^Y(f(u), y) + \varepsilon \right\}$$

This gives a distance d^Z on Z. Check that $d_H^Z(X, Y) < 2\varepsilon$. $\blacksquare$

The proof of theorem 2.25 follows as an application of previous propositions.

Metric tangent cones The infinitesimal (not local!) geometry of a path metric space is described with the help of Gromov-Hausdorff convergence of pointed metric spaces.

This is definition 3.14 Gromov [16].

Definition 2.30 *The sequence of pointed metric spaces* (X_n, x_n, d_n) *converges in the sense of Gromov-Hausdorff to the pointed space* (X, x, d_X) *if for any* $r > 0$, $\varepsilon > 0$ *there is* $n_0 \in N$ *such that for all* $n \geq n_0$ *there exists* $f_n : B_n(x_n, r) \subset X_n \to X$ *such that:*

(1) $f_n(x_n) = x$,

(2) the distorsion of f_n *is bounded by* ε: *dis* $f_n < \varepsilon$,

(3) $B_X(x, r) \subset (f_n(B_n(x_n, r)))_\varepsilon$.

The Gromov-Hausdorff limit is defined up to isometry, in the class of compact metric spaces (Proposition 3.6 Gromov [16]) or in the class of locally compact cones. Here it is the definition of a cone.

Definition 2.31 *A pointed metric space* (X, x_0) *is called a cone if there is a one parameter group of dilatations with center* x_0 $\{\delta_\lambda : X \to X : \lambda > 0\}$ *such that* $\delta_\lambda(x_0) = x_0$ *and for any* $x, y \in X$

$$d(\delta_\lambda(x), \delta_\lambda(y)) = \lambda \, d(x, y)$$

The infinitesimal geometry of a metric space X in the neighbourhood of $x_0 \in X$ is described by the tangent space to (X, x_0) (if such object exists).

Definition 2.32 *The tangent space to* (X, x_0) *is the Gromov-Hausdorff limit*

$$(T_{x_0}, 0, d_{x_0}) = \lim_{\lambda \to \infty} (X, x_0, \lambda d)$$

Three remarks are in order:

1. The tangent space is obviously a cone. We shall see that in a large class of situations is also a group, hence a graded nilpotent group, called for short Carnot group.

2. The tangent space comes with a metric inside. This space is path metric (Gromov [16] Proposition 3.8).

3. The tangent cone is defined up to isometry, therefore there is no way to use the metric tangent cone definition to construct a tangent bundle. For a modification of the definitions in this direction see Margulis, Mostow [23], or Vodop'yanov & Greshnov [32], [33].

2.4   Mitchell's theorem 1. Bellaïche's construction

We collect here three key items in the edification of sub-Riemannian geometry. The first is Mitchell [24] theorem 1:

Theorem 2.33 *For a regular sub-Riemannian space* (M, D, g), *the tangent cone of* (M, d_{CC}) *at* $x \in M$ *exists and it is isometric to* $(N(D), d_N)$, *which is a Carnot group with a left invariant distribution and* d_N *is a induced Carnot-Carathéodory distance.*

We shall prove this theorem in the particular case of a Lie group with left-invariant distribution. For a detailed proof of the general case see Vodop'yanov & Greshnov [32] [33].

The Carnot group $N(D)$ is called the nilpotentisation of the regular distribution D and it can be constructed from D exclusively. However, the metric on $N(D)$ depends on the choice of metric g on the sub-Riemannian manifold.

Recall that the limit in the Gromov-Hausdorff sense is defined up to isometry. This means it this case that $N(D)$ is a model for the tangent space at x to (M, d_{CC}). In the Riemannian case $D = TM$ and $N(D) = R^n$, as a group with addition.

This theorem tells us nothing about the tangent bundle. There are however other ways to associate a tangent bundle to a metric measure space (Cheeger [7]) or to a regular sub-Riemannian manifold (Margulis & Mostow [22], [23]). These bundles differs. As Tyson (interpreting Cheeger) asserts (see Tyson paper in these proceedings (??)), Cheeger tangent bundle can be identified with the distribution D and Margulis-Mostow bundle is the same as the usual tangent bundle, but with the fiber isomorphic with $N(D)$, instead of R^n. We shall explain how the Margulis-Mostow tangent bundle is constructed a bit further (again in the particular case considered here).

Let us not, for the moment, be too ambitious and restrict to the question: is there a metric derivation of the group operation on $N(D)$? Bellaïche [4] writes that he asked Gromov this question, who pointed out that the key tool to construct the operation from metric is uniformity. Bellaïche proposed therefore the following construction, which starts from the proof of Mitchell theorem 1, where it can be seen that the Gromov-Hausdorff convergence to the tangent space *is uniform with respect to* $x \in M$. This means that for any $\varepsilon > 0$ there is $R(\varepsilon) > 0$ and map $\phi_{x,\varepsilon} : B_{CC}(x, R(\varepsilon)) \to N(D)$ such that

$$d_N(\phi_{x,\varepsilon}(y), \phi_{x,\varepsilon}(z)) = d_{CC}(y, z) + o(\varepsilon) \ \ \forall y, z \in B_{CC}(x, R(\varepsilon))$$

Let us forget about ε (Bellaïche does not mention anything about it further) and take arbitrary $X, Y \in N(D)$. Pick then $y \in M$ such that $\phi_x(y) = X$. Denote $\phi_{xy} = \phi_y \phi_x^{-1}$. Then the operation in $N(D)$ is defined by:

$$X \overset{n}{\cdot} Y = \lim_{\lambda \to \infty} \phi_{xy}^{-1} \delta_\lambda^{-1} \phi_{xy} \delta_\lambda(Y)$$

It is easier to understand this in the Euclidean case, that is in R^n. We can take for example

$$\phi_x(y) = Q(x)(y - x) = X \ \ , y = x + Q^{-1}X \ \ , x \mapsto Q(x) \in SO(n) \text{ arbitrary}$$

(and we have no dependence on ε) Let us compute the operation. We get

$$\lim_{\lambda \to \infty} \phi_{xy}^{-1} \delta_\lambda^{-1} \phi_{xy} \delta_\lambda(Y) = X + Y$$

as expected. Notice that the arbitrary choice of the rotations $Q(x)$ does not influence the result. The tangent spaces at any point can rotate independently, which is a sign that this construction cannot lead to a tangent bundle.

There are several problems with Bellaïche's construction:

a) when $\lambda \to \infty$ the expression $\phi_{xy} \delta_\lambda(Y)$ might not make sense,

b) is not clear how ε and λ interact.

This is the reason why we introduced (first in [5], then here) the notion of uniform group, which encodes all that one really need to do the construction, again in the case of Lie groups with left invariant distributions.

Another way to transform Bellaïche's construction into an effective one (and more, to obtain a tangent bundle) is proposed by Margulis and Mostow. We shall explain further their construction. However, there are other problems emerging. as mentioned in the introduction.

2.5   Margulis & Mostow tangent bundle

In this section we shall apply Margulis & Mostow [23] construction of the tangent bundle to a SR manifold for the case of a group with left invariant distribution. It will turn that the tangent bundle does not have a group structure, due to the fact that, as previously, the non-smoothness of the right translations is not studied.

The main point in the construction of a tangent bundle is to have a functorial definition of the tangent space. This is achieved by Margulis & Mostow [23] in a very natural way. One of the geometrical definitions of a tangent vector v at a point x, to a manifold M, is the following one: identify v with the class of smooth curves which pass through x and have tangent v. If the manifold M is endowed with a distance then one can define the equivalence relation based in x by: $c_1 \equiv_x c_2$ if $c_1(0) = c_2(0) = x$ and the distance between $c_1(t)$ and $c_2(t)$ is negligible with respect to t for small t. The set of equivalence classes is the tangent space at x. One has to put then some structure on the tangent space (as, for example, the nilpotent multiplication).

To put is practice this idea is not so easy though. This is achieved by the following sequence of definitions and theorems. For commodity we shall explain this construction in the case $M = G$ connected Lie group, endowed with a left invariant distribution D. The general case is the one of a regular sub-Riemannian manifold. We shall denote by d_G the CC distance on G and we identify G with $\mathfrak{g}$, as previously. The CC distance induced by the distribution D^N, generated by left translations of G using nilpotent multiplication $\overset{\text{n}}{\cdot}$, will be denoted by d_N.

Definition 2.34 *A C^∞ curve in G with $x = c(0)$ is called rectifiable at $t = 0$ if $d_G(x, c(t)) \leq Ct$ as $t \to 0$.*

Two C^∞ curves $c', c"$ with $c'(0) = x = c"(0)$ are called equivalent at x if

$$t^{-1} d_G(c'(t), c"(t)) \to 0$$

as $t \to 0$.

The tangent cone to G as x, denoted by $C_x G$ is the set of equivalence classes of all C^∞ paths c with $c(0) = x$, rectifiable at $t = 0$.

Let $c : [-1, 1] \to G$ be a C^∞ rectifiable curve, $x = c(0)$ and

$$v = \lim_{t \to 0} \delta_t^{-1} \left(c(0)^{-1} \overset{\mathfrak{g}}{\cdot} c(t) \right) \tag{2.5.9}$$

The limit v exists because the curve is rectifiable.

Introduce the curve $c_0(t) = x \exp_G(\delta_t v)$. Then

$$d(x, c_0(t)) = d(e, x^{-1} c(t)) < | v | \, t$$

as $t \to 0$ (by the Ball-Box theorem) The curve c is equivalent with c_0. Indeed, we have (for $t > 0$):

$$\frac{1}{t} d_G(c(t), c_0(t)) = \frac{1}{t} d_G(c(t), x \overset{\mathfrak{g}}{\cdot} \delta_t v) = \frac{1}{t} d_G(\delta_t(v^{-1}) \overset{\mathfrak{g}}{\cdot} x^{-1} \overset{\mathfrak{g}}{\cdot} c(t), 0)$$

The latter expression is equivalent (by the Ball-Box Theorem) with

$$\frac{1}{t} d_N(\delta_t(v^{-1}) \overset{\mathfrak{g}}{\cdot} x^{-1} \overset{\mathfrak{g}}{\cdot} c(t), 0) = d_N \left(\delta_t^{-1} \left(\delta_t(v^{-1}) \overset{\mathfrak{g}}{\cdot} \delta_t \left(\delta_t^{-1} \left(x^{-1} \overset{\mathfrak{g}}{\cdot} c(t) \right) \right) \right) \right)$$

The right hand side (RHS) converges to $d_N(v^{-1} \overset{\mathfrak{n}}{\cdot} v, 0)$, as $t \to 0$, as a consequence of the definition of v and theorem 3.14.

Therefore we can identify $C_x G$ with the set of curves $t \mapsto x \exp_G(\delta_t v)$, for all $v \in \mathfrak{g}$. Remark that the equivalence relation between curves c_1, c_2, such that $c_1(0) = c_2(0) = x$ can be redefined as:

$$\lim_{t \to 0} \delta_t^{-1} \left(c_2(t)^{-1} \overset{\mathfrak{g}}{\cdot} c_1(t) \right) = 0 \tag{2.5.10}$$

In order to define the multiplication Margulis & Mostow introduce the families of segments rectifiable at t.

Definition 2.35 *A family of segments rectifiable at $t = 0$ is a C^∞ map*

$$\mathcal{F} : U \to G$$

where U is an open neighbourhood of $G \times 0$ in $G \times R$ satisfying

(a) $\mathcal{F}(\cdot, 0) = id$

(b) *the curve $t \mapsto \mathcal{F}(x, t)$ is rectifiable at $t = 0$ uniformly for all $x \in G$, that is for every compact K in G there is a constant C_K and a compact neighbourhood I of 0 such that $d_G(y, \mathcal{F}(y, t)) < C_K t$ for all $(y, t) \in K \times I$.*

Two families of segments rectifiable at $t = 0$ are called equivalent if

$$t^{-1}d_G(\mathcal{F}_1(x,t), \mathcal{F}_2(x,t)) \to 0$$

as $t \to 0$, uniformly on compact sets in the domain of definition.

Part (b) from the definition of a family of segments rectifiable can be restated as: there exists the limit

$$v(x) \;=\; \lim_{t \to 0} \delta_t^{-1}\left(x^{-1} \stackrel{\mathfrak{g}}{\cdot} \mathcal{F}(x,t)\right) \tag{2.5.11}$$

and the limit is uniform with respect to $x \in K$, K arbitrary compact set.

It follows then, as previously, that $\mathcal{F}$ is equivalent to $\mathcal{F}_0$, defined by:

$$\mathcal{F}_0(x,t) \;=\; x \stackrel{\mathfrak{g}}{\cdot} \delta_t v(x)$$

Also, the equivalence between families of segments rectifiable can be redefined as:

$$\lim_{t \to 0} \delta_t^{-1}\left(\mathcal{F}_2(x,t)^{-1} \stackrel{\mathfrak{g}}{\cdot} \mathcal{F}_1(x,t)\right) \;=\; 0 \tag{2.5.12}$$

uniformly with respect to $x \in K$, K arbitrary compact set.

Definition 2.36 *The product of two families $\mathcal{F}_1$, $\mathcal{F}_2$ of segments rectifiable at $t = 0$ is defined by*

$$(\mathcal{F}_1 \circ \mathcal{F}_2)(x,t) \;=\; \mathcal{F}_1(\mathcal{F}_2(x,t), t)$$

The product is well defined by Lemma 1.2 *op. cit..* One of the main results is then the following theorem (5.5).

Theorem 2.37 *Let c_1, c_2 be C^∞ paths rectifiable at $t = 0$, such that $c_1(0) = x_0 = c_2(0)$. Let $\mathcal{F}_1$, $\mathcal{F}_2$ be two families of segments rectifiable at $t = 0$ with:*

$$\mathcal{F}_1(x_0, t) \;=\; c_1(t) \quad , \quad \mathcal{F}_2(x_0, t) \;=\; c_2(t)$$

Then the equivalence class of

$$t \mapsto \mathcal{F}_1 \circ \mathcal{F}_2(x_0, t)$$

depends only on the equivalence classes of c_1 and c_2. This defines the product of the elements of the tangent cone $C_{x_0} G$.

This theorem is the straightforward consequence of the following facts (5.1(5) and 5.2 in Margulis & Mostow [23]).

We shall denote by $\mathcal{F} \approx \mathcal{F}'$ the equivalence relation of families of segments rectifiable; the equivalence relation of rectifiable curves based at x will be denoted by $c \stackrel{x}{\approx} c'$.

Lemma 2.38 (a) *Let $\mathcal{F}_1 \approx \mathcal{F}_2$ and $\mathcal{G}_1 \approx \mathcal{G}_2$. Then $\mathcal{F}_1 \circ \mathcal{G}_1 \approx \mathcal{F}_2 \circ \mathcal{G}_2$.*

  (b) *The map $\mathcal{F} \mapsto \mathcal{F}_0$ is constant on equivalence classes of families of segments rectifiable.*

Proof.   Let
$$\mathcal{F}_0(x,t) \;=\; x \overset{\mathfrak{g}}{\cdot} \delta_t w_1(x) \;\;,\;\; \mathcal{G}_0(x,t) \;=\; x \overset{\mathfrak{g}}{\cdot} \delta_t w_2(x)$$

For the point (a) it is sufficient to prove that

$$\mathcal{F} \circ \mathcal{G} \approx \mathcal{F}_0 \circ \mathcal{G}_0$$

This is true by the following chain of estimates.

$$\frac{1}{t} d_G(\mathcal{F} \circ \mathcal{G}(x,t), \mathcal{F}_0 \circ \mathcal{G}_0(x,t)) \;=\;$$

$$= \; \frac{1}{t} d_G(\delta_t w_1(\mathcal{G}_0(x,t))^{-1} \overset{\mathfrak{g}}{\cdot} \delta_t w_2(x)^{-1} \overset{\mathfrak{g}}{\cdot} x^{-1} \overset{\mathfrak{g}}{\cdot} \mathcal{F}(\mathcal{G}(x,t),t), 0)$$

The RHS of this equality behaves like

$$d_N(\delta_t^{-1} \left(\delta_t w_1(\mathcal{G}_0(x,t))^{-1} \overset{\mathfrak{g}}{\cdot} \delta_t w_2(x)^{-1} \overset{\mathfrak{g}}{\cdot} \delta_t \left(\delta_t^{-1} \left(x^{-1} \overset{\mathfrak{g}}{\cdot} \mathcal{G}(x,t) \right) \right) \overset{\mathfrak{g}}{\cdot}$$

$$\overset{\mathfrak{g}}{\cdot} \delta_t \left(\delta_t^{-1} \left(\mathcal{G}(x,t)^{-1} \overset{\mathfrak{g}}{\cdot} \mathcal{F}(\mathcal{G}(x,t),t) \right) \right) \right), 0)$$

This quantity converges (uniformly with respect to $x \in K$, K an arbitrary compact) to

$$d_N(w_1(x)^{-1} \overset{\mathfrak{n}}{\cdot} w_2(x)^{-1} \overset{\mathfrak{n}}{\cdot} w_2(x) \overset{\mathfrak{n}}{\cdot} w_1(x), 0) \;=\; 0$$

The point (b) is easier: let $\mathcal{F} \approx \mathcal{G}$ and consider $\mathcal{F}_0$, $\mathcal{G}_0$, as above. We want to prove that $\mathcal{F}_0 \;=\; \mathcal{G}_0$, which is equivalent to $w_1 \;=\; w_2$.

Because $\approx$ is an equivalence relation all we have to prove is that if $\mathcal{F}_0 \approx \mathcal{G}_0$ then $w_1 \;=\; w_2$. We have:

$$\frac{1}{t} d_G(\mathcal{F}_0(x,t), \mathcal{G}_0(x,t)) \;=\; \frac{1}{t} d_G(x \overset{\mathfrak{g}}{\cdot} \delta_t w_1(x), x \overset{\mathfrak{g}}{\cdot} \delta_t w_2(x))$$

We use the $\overset{\mathfrak{g}}{\cdot}$ left invariance of d_G and the Ball-Box theorem to deduce that the RHS behaves like

$$d_N(\delta_t^{-1} \left(\delta_t w_2(x)^{-1} \overset{\mathfrak{g}}{\cdot} \delta_t w_1(x)^{-1} \right), 0)$$

which converges to $d_N(w_1(x), w_2(x))$ as t goes to 0. The two families are equivalent, therefore the limit equals 0, which implies that $w_1(x) = w_2(x)$ for all x.   ∎

We shall apply this theorem. Let $c_i(t) \;=\; x_0 \exp_G \delta_t v_i$, for $i = 1, 2$. It is easy to check that $\mathcal{F}_i(x,t) \;=\; x \exp_G(\delta_t v_i)$ are families of segments rectifiable at $t = 0$ which satisfy the hypothesis of the theorem. But then

$$(\mathcal{F}_1 \circ \mathcal{F}_2)(x,t) \;=\; x_0 \exp_G(\delta_t v_1) \exp_G(\delta_t v_2)$$

which is equivalent with

$$\mathcal{F} \exp_G \left(\delta_t(v_1 \overset{\mathfrak{n}}{\cdot} v_2) \right)$$

Therefore the tangent bundle defined by this procedure is the same as the virtual tangent bundle which we shall define soon, inspired from the construction proposed by Bellaïche.

Maybe I misunderstood the notations, but it seems to me that several times the authors claim that the exponential map which they construct is bi-Lipschitz (as in 5.1(4) and Corollary 4.5). This is false, as explained before. In Bellaïche [4], Theorem 7.32 and also at the beginning of section 7.6 we find that the exponential map is only $1/m$ Hölder continuous (where m is the step of the nilpotentization). However, the final results of Margulis & Mostow hold true, if not entirely proven facts, as statements at least.

2.6   Vodop'yanov & Greshnov definition of the derivability. Rademacher theorem

We choose Vodop'yanov & Greshnov [32], [33] way of defining the derivability in order to explain Margulis & Mostow [22] Rademacher theorem 10.5 (or Vodop'yanov & Greshnov theorem 1). The definition (10.3.1) of derivability in the paper [22] is to be compared with the definition of $\mathcal{P}$ differentiability from [32], first page, which is in my opinion clearer. However, the reader entering for the first time in this subject might find hard to understand why such elementary notions as differentiability need so lengthy discussions. It is, I think, another sign of the fact that the foundations of non-Euclidean analysis are still in construction.

We stay, as previously, in the case of Lie groups with left invariant distributions. We put on such groups Lebesgue measures coming from arbitrary atlases.

Definition 2.39 *A mapping $f : G_1 \to G_2$ is said to be differentiable at $x \in G_1$ if the mapping $\exp_{G_2}^{-1} \circ L_{f(x)} \circ f \circ L_x \circ \exp_{G_1}$ is Pansu derivable at 0, when we identify the algebras $\mathfrak{g}_1$, $\mathfrak{g}_2$ with the nilpotentisations of G_1, G_2 respectively.*

The following theorem then holds. The original (and stronger) versions of this theorem concern quasi-conformal mapping and can be found in Margulis & Mostow [22], Vodop'yanov & Greshnov [32] and the paper by Vodop'yanov in these proceedings.

Theorem 2.40 *Any Lipschitz map $f : E \subset G_1 \to G_2$, E measurable, is derivable almost everywhere.*

3   Uniform groups

We start with the following setting: G is a topological group endowed with an uniformity such that the operation is uniformly continuous. More specifically, we introduce first the double of G, as the group $G^{(2)} = G \times G$ with operation

$$(x, u)(y, v) = (xy, y^{-1}uyv)$$

The operation on the group G, seen as the function

$$op : G^{(2)} \to G , \quad op(x, y) = xy$$

is a group morphism. Also the inclusions:

$$i' : G \to G^{(2)} , \quad i'(x) = (x, e)$$

$$i" : G \to G^{(2)} \ , \quad i"(x) \ = \ (x, x^{-1})$$

are group morphisms.

Definition 3.1   *1. G is an uniform group if we have two uniformity structures, on G and G^2, such that op, i', $i"$ are uniformly continuous.*

2. *A local action of a uniform group G on a uniform pointed space (X, x_0) is a function $\phi \in W \in \mathcal{V}(e) \mapsto \hat{\phi} : U_\phi \in \mathcal{V}(x_0) \to V_\phi \in \mathcal{V}(x_0)$ such that:*

    (a) *the map $(\phi, x) \mapsto \hat{\phi}(x)$ is uniformly continuous from $G \times X$ (with product uniformity) to X,*

    (b) *for any $\phi, \psi \in G$ there is $D \in \mathcal{V}(x_0)$ such that for any $x \in D$ $\widehat{\phi\psi^{-1}}(x)$ and $\hat{\phi}(\hat{\psi}^{-1}(x))$ make sense and $\widehat{\phi\psi^{-1}}(x) = \hat{\phi}(\hat{\psi}^{-1}(x))$.*

3. *Finally, a local group is an uniform space G with an operation defined in a neighbourhood of $(e, e) \subset G \times G$ which satisfies the uniform group axioms locally.*

Remark that a local group acts locally at left (and also by conjugation) on itself.

This definition deserves an explanation. We shall start by recalling what an uniformity is (consult, for example the book James [18], for an introduction to uniform spaces). We need first some notations.

A relation R of a given set X is just a subset $R \subset X \times X$. The inverse of R is

$$R^{-1} \ = \ \{(y, x) \in X \times X : (x, y) \in R\}$$

The composition of two relations R, S is defined by:

$$R \circ S \ = \ \{(x, z) \in X \times X : \exists y \in X \ \ (x, y) \in R \ , \ (y, z) \in S\}$$

The composition is an associative operation.

The diagonal of X is the relation

$$\Delta(X) \ = \ \{(x, x) : x \in x\}$$

A filter on X is a set Ω of subsets of X such that:

- if $A, B \in \Omega$ then $A \cap B \in \Omega$,

- if $A \in \Omega$, $B \subset X$ and $A \subset B$ then $B \in \Omega$.

Definition 3.2 *An uniformity of a given set X is a filter Ω on $X \times X$ such that:*

   i) *for all $D \in \Omega$ $\Delta(X) \subset D$,*

   ii) *if $D \in \Omega$ then $D^{-1} \in \Omega$,*

   iii) *if $D \in \Omega$ then there is $E \in \Omega$ such that $E \circ E \subset D$.*

For example, if (X, d) is a metric space then the filter on $X \times X$ generated by the sets:

$$B(r) \; = \; \{(x, y) \in X \times X : d(x, y) < r\}$$

for all $r > 0$ is an uniformity on X. Indeed, property i) comes from the fact that $d(x, x) = 0$ for any $x \in X$, ii) comes from $d(x, y) = d(y, x)$ for any $x, y \in X$ and iii) comes from the triangle inequality for the distance d.

Let (X, Ω_X) and (Y, Ω_Y) be two uniform spaces. A function $\phi : X \to Y$ is uniformly continuos if

$$(\phi \times \phi)^{-1} (\Omega_Y) \; \subset \; \Omega_X$$

An uniform group, according to the definition (3.1), is a group G such that left translations are uniformly continuous functions and the left action of G on itself is uniformly continuous too. In order two precisely formulate this we need two uniformities: one on G and another on $G \times G$.

These uniformities should be compatible, which is achieved by saying that i', i" are uniformly continuous. The uniformity of the group operation is achieved by saying that the *op* morphism is uniformly continuous.

The particular choice of the operation on $G \times G$ is not essential at this point, but it is justified by the cae of a Lie group endowed with the CC distance induced by a left invariant distribution. We shall construct a natural CC distance on $G \times G$, which is left invariant with respect to the chosen operation on $G \times G$. These distances induce uniformities which transform G into an uniform group according to definition (3.1). The goal, achieved in the next section, is proposition (3.11) which claims that the operation function *op* is derivable, even if right translations are not "smooth", i.e. commutative smooth according to definition (3.6).

We prepare now the path to this result. The "infinitesimal version" of an uniform group is a conical local uniform group.

Definition 3.3 *A conical local uniform group N is a local group with a local action of $(0, +\infty)$ by morphisms δ_ε such that $\lim\limits_{\varepsilon \to 0} \delta_\varepsilon x \; = \; e$ for any x in a neighbourhood of the neutral element e.*

We shall make the following hypotheses on the local uniform group G: there is a local action of $(0, +\infty)$ (denoted by δ), on (G, e) such that

H0. the limit $\lim_{\varepsilon \to 0} \delta_\varepsilon x \; = \; e$ exists and is uniform with respect to x.

H1. the limit

$$\beta(x, y) \; = \; \lim_{\varepsilon \to 0} \delta_\varepsilon^{-1} ((\delta_\varepsilon x)(\delta_\varepsilon y))$$

is well defined in a neighbourhood of e and the limit is uniform.

H2. the following relation holds

$$\lim_{\varepsilon \to 0} \delta_\varepsilon^{-1} \left((\delta_\varepsilon x)^{-1} \right) \; = \; x^{-1}$$

where the limit from the left hand side exists in a neighbourhood of e and is uniform with respect to x.

3.1   Virtual tangent bundle

Proposition 3.4 *Under the hypotheses H0, H1, H2 (G, β) is a conical local uniform group.*

Proof.   All the uniformity assumptions permit to change at will the order of taking limits. We shall not insist on this further and we shall concentrate on the algebraic aspects.

We have to prove the associativity, existence of neutral element, existence of inverse and the property of being conical. The proof is straightforward. For the associativity $\beta(x, \beta(y, z)) = \beta(\beta(x, y), z)$ we compute:

$$\beta(x, \beta(y, z)) = \lim_{\varepsilon \to 0, \eta \to 0} \delta_\varepsilon^{-1} \left\{ (\delta_\varepsilon x) \delta_{\varepsilon/\eta} \left((\delta_\eta y)(\delta_\eta z) \right) \right\}$$

We take $\varepsilon = \eta$ and we get

$$\beta(x, \beta(y, z)) = \lim_{\varepsilon \to 0} \left\{ (\delta_\varepsilon x)(\delta_\varepsilon y)(\delta_\varepsilon z) \right\}$$

In the same way:

$$\beta(\beta(x, y), z) = \lim_{\varepsilon \to 0, \eta \to 0} \delta_\varepsilon^{-1} \left\{ (\delta_{\varepsilon/\eta} x) \left((\delta_\eta x)(\delta_\eta y) \right) (\delta_\varepsilon z) \right\}$$

and again taking $\varepsilon = \eta$ we obtain

$$\beta(\beta(x, y), z) = \lim_{\varepsilon \to 0} \left\{ (\delta_\varepsilon x)(\delta_\varepsilon y)(\delta_\varepsilon z) \right\}$$

The neutral element is e, from H0 (first part): $\beta(x, e) = \beta(e, x) = x$. The inverse of x is x^{-1}, by a similar argument:

$$\beta(x, x^{-1}) = \lim_{\varepsilon \to 0, \eta \to 0} \delta_\varepsilon^{-1} \left\{ (\delta_\varepsilon x) \left(\delta_{\varepsilon/\eta} (\delta_\eta x)^{-1} \right) \right\}$$

and taking $\varepsilon = \eta$ we obtain

$$\beta(x, x^{-1}) = \lim_{\varepsilon \to 0} \delta_\varepsilon^{-1} \left((\delta_\varepsilon x)(\delta_\varepsilon x)^{-1} \right) = \lim_{\varepsilon \to 0} \delta_\varepsilon^{-1}(e) = e$$

Finally, β has the property:

$$\beta(\delta_\eta x, \delta_\eta y) = \delta_\eta \beta(x, y)$$

which comes from the definition of β and commutativity of multiplication in $(0, +\infty)$. This proves that (G, β) is conical.   ∎

We arrive at a natural realization of the tangent space to the neutral element. Let us denote by $[f, g] = f \circ g \circ f^{-1} \circ g^{-1}$ the commutator of two transformations. For the group we shall denote by $L_x^G y = xy$ the left translation and by $L_x^N y = \beta(x, y)$. The preceding proposition tells us that (G, β) acts locally by left translations on G. We shall call the left translations with respect to the group operation β "infinitesimal". Those infinitesimal translations admit the very important representation:

$$\lim_{\lambda \to 0} [L_{(\delta_\lambda x)^{-1}}^G, \delta_\lambda^{-1}] = L_x^N \tag{3.1.1}$$

Definition 3.5 *The group VT_eG formed by all transformations L_x^N is called the virtual tangent space at e to G.*

The virtual tangent space VT_xG at $x \in G$ to G is obtained by translating the group operation and the dilatations from e to x. This means: define a new operation on G by

$$y \overset{x}{\cdot} z = yx^{-1}z$$

The group G with this operation is isomorphic to G with old operation and the left translation $L_x^G y = xy$ is the isomorphism. The neutral element is x. Introduce also the dilatations based at x by

$$\delta_\varepsilon^x y = x\delta_\varepsilon(x^{-1}y)$$

Then $G^x = (G, \overset{x}{\cdot})$ with the group of dilatations δ_ε^x satisfy the axioms Ho, H1, H2. Define then the virtual tangent space VT_xG to be: $VT_xG = VT_xG^x$. A short computation shows that

$$VT_xG = \left\{ L_y^{N,x} = L_x L_{x^{-1}y}^N L_x : y \in U_x \in \mathcal{V}(X) \right\}$$

where

$$L_y^{N,x} = \lim_{\lambda \to 0} \delta_\lambda^{-1,x}[\delta_\lambda^x, L_{(\delta_\lambda x)^x,-1}^G]\delta_\lambda^x$$

We shall introduce the notion of commutative smoothness, which contains a derivative resembling with Pansu derivative. This definition is a little bit stronger than the one given by Vodopyanov & Greshnov [32], because their definition is good for a general CC space, when uniformities are taken according to the distances in CC spaces $G^{(2)}$ and G.

Definition 3.6 *A function $f : G_1 \to G_2$ is commutative smooth at $x \in G_1$, where G_1, G_2 are two groups satisfying H0, H1, H2, if the application*

$$u \in G_1 \mapsto (f(x), Df(x)u) \in G_2^{(2)}$$

exists, where

$$Df(x)u = \lim_{\varepsilon \to 0} \delta_\varepsilon^{-1} \left(f(x)^{-1} f(x\delta_\varepsilon u) \right)$$

and the convergence is uniform with respect to u in compact sets.

For example the left translations L_x are commutative smooth and the derivative equals identity. If we want to see how the derivative moves the virtual tangent spaces we have to give a definition.

Inspired by (3.1.1), we shall introduce the virtual tangent. We proceed as follows: to $f : G \to G$ and $x \in G$ let associate the function:

$$\hat{f}^x : G \times G \to G , \quad \hat{f}^x(y, z) = \hat{f}_y^x(z) = (f(x))^{-1} f(xy)z$$

To this function is associated a flow of left translations

$$\lambda > 0 \mapsto \hat{f}_{\delta_\lambda y}^x : G \to G$$

Definition 3.7 *The function $f : G \to G$ is virtually derivable at $x \in G$ if there is a virtual tangent $VDf(x)$ such that*

$$\lim_{\lambda \to 0} \left[\left(\hat{f}^x_{\delta_\lambda y} \right)^{-1}, \delta_\lambda^{-1} \right] \;=\; VDf(x)y \qquad (3.1.2)$$

and the limit is uniform with respect to y in a compact set.

Proposition 3.8 *Suppose that f is commutative derivable. Then it is also virtually derivable and the virtual tangent to f is given by:*

$$VDf(x)y \;=\; L_{f(x)} L^N_{Df(x)y} L^{-1}_{f(x)}(y)$$

Remark 3.9 *All definitions hold for $f : G_1 \to G_2$ in an obvious way.*

With this definition L_x is commutative smooth and it's virtual tangent in any point y is a group morphism from $VT_y G$ to $VT_{xy} G$. More generally, by standard reasoning, we get again the proposition 2.13, in this more general setting. This time we shall give the proof, for further reference.

Proposition 3.10 *Let $f : G \to G$, $y, z \in G$ such that:*

(i) $Df(x)y$, $Df(x)z$ exist for any $x \in G$.

(ii) The map $x \mapsto Df(x)z$ is continuous.

(iii) $x \mapsto \delta_t^{-1} \left(f(x)^{-1} f(x \delta_t z) \right)$ converges uniformly to $Df(x)z$.

Then for any (sufficiently small) $a, b > 0$ and any $w = \beta(\delta_a x, \delta_b y)$ the limit $Df(x)w$ exists and we have:

$$Df(x) \delta_a x \delta_b y \;=\; \beta(\delta_a Df(x)y, \delta_b Df(x)z)$$

Proof.   The proof is standard. Remark that if $Df(x)y$ exists then for any $a > 0$ the limit $Df(x)\delta_a y$ exists and
$$Df(x)\delta_a y \;=\; \delta_a Df(x)y$$

It is not restrictive therefore to suppose that everything happens in a neighbourhood of the identity. Let $w = \beta(y, z)$. We also use the notation

$$\beta_t(x, y) \;=\; \delta_t^{-1} \left(\delta_t(x) \delta_t(y) \right)$$

We write:
$$\delta_t^{-1} \left(f(x)^{-1} f(x \delta_t w) \right) \;=\; \beta_t \left((1)_t, (2)_t \right)$$

with the notation (useful for further reference in proposition (4.8))

$$(1)_t \;=\; \delta_t^{-1} \left(f(x)^{-1} f(x \delta_t y) \right)$$

$$(2)_t \;=\; \delta_t^{-1}\left(f(x\delta_t y)^{-1} f(x\delta_t y \delta_t z)\right)$$

When t tends to 0 (i) implies that $(1)_t$ tends to $Df(x)y$, (ii) and (iii) imply that $(2)_t$ tends to $Df(x)z$.   ■

Remark that in principle the right translations are not commutative smooth.

Now that we have a model for the tangent space to e at G, we can show that the operation is commutative smooth.

Proposition 3.11 *Let G satisfy H0, H1, H2 and $\delta_\varepsilon^{(2)} : G^{(2)} \to G^{(2)}$ be defined by*

$$\delta_\varepsilon^{(2)}(x, u) \;=\; (\delta_\varepsilon x, \delta_\varepsilon y)$$

Then $G^{(2)}$ satisfies H0, H1, H2, the operation (op function) is commutative smooth and we have the relation:

$$D \ op \ (x, u)(y, v) \;=\; \beta(y, v)$$

Proof.   It is sufficient to use the morphism property of the operation. Indeed, the right hand side of the relation to be proven is

$$RHS \;=\; \lim_{\varepsilon \to 0} \delta_\varepsilon^{-1}\left(op(x, u)^{-1} op(x, u) op\left(\delta_\varepsilon^{(2)}(y, v)\right)\right) \;=$$

$$=\; \lim_{\varepsilon \to 0} \delta_\varepsilon^{-1}\left(op(\delta_\varepsilon^{(2)}(y, v))\right) \;=\; \beta(y, v)$$

The rest is trivial.   ■

3.2   The case of sub-Riemannian groups

The notion of virtual tangent space is not based on the use of distances, but on the use of dilatations. In fact, any manifold has a tangent space to any of its points, not only the Riemannian manifolds. We shall prove further that VT_eG is isomorphic to the nilpotentisation $N(G, D)$.

We start from Euclidean norm on D and we choose an orthonormal basis of D. We can then extend the Euclidean norm to $\mathfrak{g}$ by stating that the basis of $\mathfrak{g}$ constructed, as explained, from the basis on D, is orthonormal. By left translating the Euclidean norm on $\mathfrak{g}$ we endow G with a structure of Riemannian manifold. The induced Riemannian distance d_R will give an uniform structure on G. This distance is left invariant:

$$d_R(xy, xz) \;=\; d_R(y, z)$$

for any $x, y, z \in G$.

Any left invariant distance d is uniquely determined if we set $d(x) \;=\; d(e, x)$.

The following lemma is important (compare with lemma 2.5).

Lemma 3.12 *Let $X_1, ..., X_p$ be a basis of D. Then there are $U \subset G$ and $V \subset N(G, D)$, open neighbourhoods of the neutral elements e_G, e_N respectively, and a surjective function $g : \{1, ..., M\} \to \{1, ..., p\}$ such that any $x \in U$, $y \in V$ can be written as*

$$x = \prod_{i=1}^{M} \exp_G(t_i X_{g(i)}) \ , \ y = \prod_{i=1}^{M} \exp_N(\tau_i X_{g(i)}) \tag{3.2.3}$$

Proof. We shall make the proof for G; the proof for $N(G, D)$ will follow from the identifications explained before.

Denote by n the dimension of $\mathfrak{g}$. We start the proof with the remark that the function

$$(t_1, ..., t_n) \ \mapsto \ \prod_{i=1}^{n} \exp_G(t_i X_i) \tag{3.2.4}$$

is invertible in a neighbourhood of $0 \in R^n$, where the X_i are elements of a basis B constructed as before. Remember that each $X_i \in B$ is a multi-bracket of elements from the basis of D. If we replace a bracket $\exp_G(t[x, y])$ in the expression (3.2.4) by $exp_G(t_1 x) \, exp_G(t_2 y) \, exp_G(t_3 x) \, exp_G(t_4 y)$ and we replace t by $(t_1, ..., t_4)$ then the image of a neighbourhood of 0 by the obtained function still covers a neighbourhood of the neutral element. We repeat this procedure a finite number of times and the thesis is proven. ∎

As a corollary we obtain the Chow theorem for our particular example.

Theorem 3.13 *Any two points $x, y \in G$ can be joined by a horizontal curve.*

Let d_G be the Carnot-Carathéodory distance induced by the distribution D and the metric. This distance is also left invariant. We obviously have $d_R \leq d_G$. We want to show that d_R and d_G induce the same uniformity on G.

Let us introduce another left invariant distance on G

$$d_G^1(x) \ = \ \inf \left\{ \sum | t_i | \ : \ x \ = \ \prod \exp_G(t_i Y_i) \ , \ Y_i \in D \right\}$$

and the auxiliary functions :

$$\Delta_G^1(x) \ = \ \inf \left\{ \sum_{i=1}^{M} | t_i | \ : \ x \ = \ \prod \exp_G(t_i X_{g(i)}) \right\}$$

$$\Delta_G^\infty(x) \ = \ \inf \left\{ \max | t_i | \ : \ x \ = \ \prod \exp_G(t_i X_{g(i)}) \right\}$$

We can prove that $d_G^1 = d_G(e_G, \cdot)$. (Indeed $d_G(e_G, \cdot) \leq d_G^1$. On the other part $d_G^1(x)$ is less equal than the variation of any Lipschitz curve joining e_G with x. Therefore we have equality.)

The functions Δ_G^1, Δ_G^∞ don't induce left invariant distances. Nevertheless they are useful, because of their equivalence:

$$\Delta_G^\infty(x) \ \leq \ \Delta_G^1(x) \ \leq \ M \, \Delta_G^\infty(x) \tag{3.2.5}$$

for any $x \in G$. This is a consequence of the lemma 3.12.

We have therefore the chain of inequalities:

$$d_R \ \leq \ d_G \ \leq \ \Delta_G^1 \ \leq \ M\Delta_G^\infty$$

But from the proof of lemma 3.12 we see that Δ_G^∞ is uniformly continuous. This proves the equivalence of the uniformities.

Because $\exp_G$ does not deform much the Riemannian d_R distances near e, it can be checked that the group G with the dilatations

$$\tilde{\delta}_\varepsilon(\exp_G x) \ = \ \exp_G(\delta_\varepsilon x)$$

satisfies H0, H1, H2.

The same conclusion is true for the local uniform group (with the uniformity induced by the Euclidean distance) $\mathfrak{g}$ with the operation:

$$X \overset{\mathfrak{g}}{\cdot} Y \ = \ \log_G\left(\exp_G(X)\exp_G(Y)\right)$$

for any X, Y in a neighbourhood of $0 \in \mathfrak{g}$. Here the dilatations are δ_ε. We shall denote this group by $\log G$. These two groups are isomorphic as local uniform groups by the map $\exp_G$. Dilatations commute with the isomorphism. They have therefore isomorphic (by $\exp_G$) virtual tangent spaces.

Theorem 3.14 *The virtual tangent space VT_eG is isomorphic to $N(G, D)$. More precisely $N(G, D)$ is equal (as local group) to the virtual tangent space to $\log G$:*

$$N(G, D) \ = \ VT_0\log G$$

Proof.   The product XY in $\log G$ is given by Baker-Campbell-Hausdorff formula

$$X \overset{\mathfrak{g}}{\cdot} Y \ = \ X + Y + \frac{1}{2}[X, Y] + ...$$

Use proposition 2.21 to compute $\beta(X, Y)$ and show that $\beta(X, Y)$ equals the nilpotent multiplication.   ■

3.3   Proof of Mitchell's theorem 1

We shall prove the result of Mitchell [24] theorem 1, that G admits in any point a metric tangent cone, which is isomorphic with the nilpotentisation $N(G, D)$. Mitchell theorem is true for regular sub-Riemannian manifolds. The proof that we give here is based on the lemma 3.12 and Gromov [16] section 1.2.

Because left translations are isometries, it is sufficient to prove that G admits a metric tangent cone in identity and that the tangent cone is isometric with $N(G, D)$.

For this we transport all in the algebra $\mathfrak{g}$, endowed with two brackets $[\cdot, \cdot]_G$, $[\cdot, \cdot]_N$ and with two operations $\overset{\mathfrak{g}}{\cdot}$ and $\overset{\mathfrak{n}}{\cdot}$. Denote by d_G, d_N the Carnot-Carathéodory distances

corresponding to the $\overset{\mathfrak{g}}{\cdot}$, respectively $\overset{\mathfrak{n}}{\cdot}$ left invariant distributions on (a neighbourhood of 0 in) $\mathfrak{g}$. l_G, l_N are the corresponding length functionals. We shall denote by $B_G(x, R)$, $B_N(x, R)$ the balls centered in x with radius R with respect to d_G, d_N.

We can refine lemma 3.12 in order to obtain the Ball-Box theorem in this more general situation.

Theorem 3.15 *(Ball-Box Theorem) Denote by*

$$Box_G^1(\varepsilon) \;=\; \left\{ x \in G \,:\, \Delta_G^1(x) \;<\; \varepsilon \right\}$$

$$Box_N^1(\varepsilon) \;=\; \left\{ x \in N \,:\, \Delta_N^1(x) \;<\; \varepsilon \right\}$$

For small $\varepsilon > 0$ there is a constant $C > 1$ such that

$$\exp_G \left(Box_N^1(\varepsilon) \right) \;\subset\; Box_G^1(C\varepsilon) \;\subset\; \exp_G \left(Box_N^1(C^2\varepsilon) \right)$$

Proof. Reconsider the proof of lemma 3.12. This time, instead of the trick of replacing commutators $\exp_G[X, Y](t)$ with four letters words

$$\exp_G X(t_1) \exp_G Y(t_2) \exp_G X(t_3) \exp_G Y(t_4)$$

we shall use a smarter replacement (which, important! , works equally for the nilpotentisation $N(G, D)$).

We start with a basis $\{Y_1, ..., Y_n\}$ of the algebra $\mathfrak{g}$, constructed from multibrackets of elements $\{X_1, ..., X_p\}$ which form a basis for the distribution D. Introduce an Euclidean distance by declaring the basis of $\mathfrak{g}$ orthonormal. Set

$$[X(t), Y(t)]^\circ \;=\; (tX) \overset{\mathfrak{g}}{\cdot} (tY) \overset{\mathfrak{g}}{\cdot} (-tX) \overset{\mathfrak{g}}{\cdot} (-tY)$$

It is known that for any set of vectors $Z_1, ..., Z_q \in \mathfrak{g}$, if we denote by $\alpha^\circ(Z_1(t), ..., Z_q(t))$ a $[\cdot, \cdot]^\circ$ multibracket and by $\alpha(Z_1, ..., Z_q)$ the same constructed $[\cdot, \cdot]$ multibracket, then we have

$$\alpha^\circ(Z_1(t), ..., Z_q(t)) \;=\; t^q \alpha(Z_1, ..., Z_q) \;+\; o(t^q)$$

with respect to the Euclidean norm.

Remark as previously that the function

$$(t_1, ..., t_n) \;\mapsto\; \prod_{i=1}^{n} (t_i Y_i) \tag{3.3.6}$$

is invertible in a neighbourhood of $0 \in R^n$. Each X_i from the basis of $\mathfrak{g}$ can be written as a multibracket

$$X_i \;=\; \alpha_i(X_{k_1}, ..., X_{k_{j_i}})$$

which has the length $l_i = j_i - 1$. If l_i is odd then replace $(t_i X_i)$ by

$$\alpha^\circ(X_{k_1}(t^{1/l_i}), ..., X_{k_{j_i}}(t^{1/l_i}))$$

If l_i is even then the multibracket α_i can be rewritten as

$$\alpha_i(X_{k_1}, ..., X_{k_{j_i}}) \;=\; \beta_i(X_{k_1}, ..., [X_{k_p}, X_{k_{p+1}}], ..., X_{k_{j_i}})$$

Replace then $(t_i X_i)$ by

$$\begin{cases} \beta_i^\circ(X_{k_1}(\mid t \mid^{1/l_i}), ..., [X_{k_p}(\mid t \mid^{1/l_i}), X_{k_{p+1}}(\mid t \mid^{1/l_i})]^\circ, ..., X_{k_{j_i}}(\mid t \mid^{1/l_i})) & \text{if } t \geq 0 \\ \beta_i^\circ(X_{k_1}(\mid t \mid^{1/l_i}), ..., [X_{k_{p+1}}(\mid t \mid^{1/l_i}), X_{k_p}(\mid t \mid^{1/l_i})]^\circ, ..., X_{k_{j_i}}(\mid t \mid^{1/l_i})) & \text{if } t \leq 0 \end{cases}$$

After this replacements in the expression (3.3.6) one obtains a function E_G which is still invertible in a neighbourhood of 0. We obtain a function E_N with the same algebraic expression as E_G, but with $[\cdot, \cdot]_N^\circ$ brackets instead of $[\cdot, \cdot]_G^\circ$ ones. Use these functions to (obviously) end the proof of the theorem.   ∎

Theorem 3.16 *(Mitchell's theorem 1) The Gromov-Hausdorff limit of pointed metric spaces $(\mathfrak{g}, 0, \lambda d_G)$ as $\lambda \to \infty$ exists and equals $(\mathfrak{g}, d_N)$.*

Proof.   We shall use the proposition 2.29. For this we shall construct ε isometries between $Box_N^1(1)$ and $Box_G^1(C\varepsilon)$. These are provided by the function $E_G \circ E_N^{-1} \circ \delta_\varepsilon$.

The trick consists in the definition of the nets. We shall exemplify the construction for the case of a 3 dimensional algebra $\mathfrak{g}$. The basis of $\mathfrak{g}$ is $X_1, X_2, X_3 \;=\; [X_1, X_2]_G$. Divide the interval $[0, X_1]$ into P equal parts, same for the interval $[0, X_2]$. The interval $[0, X_3]$ though will be divided into P^2 intervals. The net so obtained, seen in $N \equiv \mathfrak{g}$, turn $E_G \circ E_N^{-1} \circ \delta_\varepsilon$ into a ε isometry between $Box_N^1(1)$ and $Box_G^1(C\varepsilon)$. To check this is mostly a matter of smart (but still heavy) notations.   ∎

The proof of this theorem is basically a refinement of Proposition 3.15, Gromov [16], pages 85–86, mentioned in the introduction of these notes. Close examination shows that the theorem is a consequence of the following facts:

(a) the identity map $id \;:\; (\mathfrak{g}, d_G) \to (\mathfrak{g}, d_N)$ has finite dilatation in 0 (equivalently $\exp_G : N(D, G) \to G$ has finite dilatation in 0),

(b) the change from the G-invariant distribution induced by D to the N invariant distribution induced by the same D is classically smooth.

It is therefore natural that the result does not feel the non-derivability of id map in $x \neq 0$. Here one has to think of id map as the exponential map, defined from the group algebra $\mathfrak{g}$, endowed with the nilpotent distribution, to the group G, endowed with initial distribution.

3.4   Constructions of tangent bundles

In the previous sections we had fixed one parameter group of dilatations which satisfy the axioms $H0$, $H1$, $H2$. We shall look now to the class of all dilatation groups (or dilatation flows, because these groups are one dimensional) which satisfy those axioms.

G is a real connected Lie group endowed with a left invariant distribution D, as considered previously. We shall denote by d the Carnot-Carathéodory distance and by

δ_ε the one parameter group of dilatations (which are morphisms of the nilpotentisation, but they are transported first on the Lie algebra $\mathfrak{g}$ and next on a neighbourhood of the neutral element in G). We know now that G endowed with the dilatations δ_ε and with the uniformities

- the d induced uniformity on G,

- the uniformity induced by the Carnot-Carathéodory distance on $G^{(2)}$. This distance correspond to the left-invariant distribution generated by $D \times D \subset Lie(G^{(2)})$,

is an uniform group according to the definition 3.1.

The use of the distance d simplifies the writing of the paper. The constructions described further work for uniform groups.

We begin by considering the class of all dilatation flows $\varepsilon \mapsto \hat{\delta}_\varepsilon : G \to G$ such that:

(a) $\lim\limits_{\varepsilon \to 0} \hat{\delta}_\varepsilon(x) \; = \; e$ uniformly with respect to $x \in K$, K compact neighbourhood of e.

(b) $\beta(x,y) \; = \; \lim\limits_{\varepsilon \to 0} \hat{\delta}_\varepsilon^{-1}\left(\hat{\delta}_\varepsilon(x)\hat{\delta}_\varepsilon(y)\right)$ uniformly with respect to $x, y \in K$, K compact neighbourhood of e.

(c) $x^{-1} \; = \; \lim\limits_{\varepsilon \to 0} \hat{\delta}_\varepsilon^{-1}\left(\hat{\delta}_\varepsilon(x)^{-1}\right)$ uniformly with respect to $x \in K$, K compact neighbourhood of e.

If we replace δ_ε by $\hat{\delta}_\varepsilon$ then we obtain the same infinitesimal objects (same virtual tangent space, same virtual tangent bundle).

Definition 3.17   (i) *Two dilatation flows δ', $\delta"$ are equivalent (we write $\delta' \equiv \delta"$) if*

$$\lim\limits_{\varepsilon \to 0} \frac{1}{\varepsilon} d(\delta'_\varepsilon \circ \delta_\varepsilon^{-1}(x), \delta'_\varepsilon \circ \delta_\varepsilon^{-1}(x)) \; = \; 0$$

uniformly with respect to x. We denote by $[\delta']$ the equivalence class of δ'.

(ii) *$\mathcal{A}$ is the class of all dilatation flows which are equivalent to flows of the following form:*

$$\hat{\delta}_\varepsilon(x) \; = \; y_\varepsilon^{-1} \delta_\varepsilon(x) \quad , \quad \lim\limits_{\varepsilon \to 0} \delta_\varepsilon^{-1} y_\varepsilon \; = \; 0$$

It is easy to check that any dilatation flow $\delta'_\varepsilon \in \mathcal{A}$ satisfies conditions (a), (b), (c).

Further we shall look to "finite curves" which are kind of "rectifiable" curves (in Margulis, Mostow terminology) with respect to a dilatation flow in the class $\mathcal{A}$.

Definition 3.18 *The class of finite curves is the collection of all curves $c : [a, +\infty) \to G$ such that $\delta_\varepsilon c(\frac{1}{\varepsilon})$ converges to a point x_c, as ε goes to 0.*

On the class of finite curves we define two equivalence relations (weak and strong):

- *the curves c_1, c_2 are weakly equivalent*

$$c_1 \overset{w}{\approx} c_2 \;\; \Leftrightarrow \;\; \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} d(\delta_\varepsilon c_1(\frac{1}{\varepsilon}), \delta_\varepsilon c_2(\frac{1}{\varepsilon})) \;=\; 0$$

- *the curves c_1, c_2 are strongly equivalent*

$$c_1 \overset{s}{\approx} c_2 \;\; \Leftrightarrow \;\; \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} d(\; \delta_\varepsilon c_1(\frac{1}{\varepsilon})\, x, \delta_\varepsilon c_2(\frac{1}{\varepsilon})\, x) \;=\; 0$$

uniform with respect to x (in a neighbourhood of the neutral element).

It is an obvious remark that $c_1 \overset{s}{\approx} c_2$ implies $c_1 \overset{w}{\approx} c_2$. If we denote by $[c]_w$, $[c]_s$, the weak, strong respectively, equivalence class of c, then $[c]_w$ decomposes in disjoint union of strong classes.

Proposition 3.19 *Let c_1, c_2 be finite curves. Define*

$$\delta_{(c_1,c_2)_\varepsilon} \;=\; \delta_\varepsilon c_1(\frac{1}{\varepsilon})^{-1} \, \delta_\varepsilon c_2(\frac{1}{\varepsilon}) \, \delta_\varepsilon$$

We have then:

(a) $c_1 \overset{w}{\approx} c_2$ is equivalent to $\delta_{(c_1,c_2)} \in \mathcal{A}$.

(b) $c_1 \overset{s}{\approx} c_2$ is equivalent to $\delta_{(c_1,c_2)} \equiv \delta$.

(c) Let $\Sigma[c]$ be the spectrum of the finite curve c, that is the collection of all $[\delta_{(c,c')}]$, with $c' \overset{w}{\approx} c$. Then

$$[c]_w \;=\; \bigcup_{[\delta_{(c,c')}] \in \Sigma[c]} [c']_s$$

(disjoint union). In other words the weak class of c decomposes into strong classes enumerated by the spectrum of c.

Proof. The points (a), (b) by direct check of definitions. Point (c) follows from (a), (b). ∎

The definition of spectrum in proposition (3.19) (c) deserves attention. We may choose first a class of curves $\mathcal{C}$. Then weak, strong classes and spectra are defined with respect to the class $\mathcal{C}$. For the moment the class $\mathcal{C}$ is the class of finite curves, but in the future we shall restrict this class.

The composition of curves is defined by:

$$c_1 \circ c_2(\frac{1}{\varepsilon}) \;=\; \delta_\varepsilon^{-1} \left(\delta_\varepsilon c_1(\frac{1}{\varepsilon}) \, \delta_\varepsilon c_2(\frac{1}{\varepsilon}) \right)$$

The following proposition justifies the introduction of strong equivalence classes.

Proposition 3.20 *The composition operation induces an operation on strong equivalence classes.*
Moreover, if $c_1 \overset{s}{\approx} c_1'$ and $c_2 \overset{w}{\approx} c_2'$ then $c_1 \circ c_2 \overset{w}{\approx} c_1' \circ c_2'$.
However, composition does not induce an operation on weak classes, generally.

Proof.   We shall prove the second assertion: if $c_1 \overset{s}{\approx} c_1'$ and $c_2 \overset{w}{\approx} c_2'$ then $c_1 \circ c_2 \overset{w}{\approx} c_1' \circ c_2'$. The first and third parts of the proposition have a similar proof.

Let $\varepsilon > 0$. We have then

$$\frac{1}{\varepsilon} d \left(\delta_\varepsilon^{-1}(c_1 \circ c_2) \left(\frac{1}{\varepsilon} \right), \delta_\varepsilon^{-1}(c_1' \circ c_2') \left(\frac{1}{\varepsilon} \right) \right) =$$

$$= \frac{1}{\varepsilon} d \left(\delta_\varepsilon c_1 \left(\frac{1}{\varepsilon} \right) \delta_\varepsilon c_2 \left(\frac{1}{\varepsilon} \right), \delta_\varepsilon c_1' \left(\frac{1}{\varepsilon} \right) \delta_\varepsilon c_2' \left(\frac{1}{\varepsilon} \right) \right) \le$$

$$\le \frac{1}{\varepsilon} d \left(\delta_\varepsilon c_2' \left(\frac{1}{\varepsilon} \right), \delta_\varepsilon c_1' \left(\frac{1}{\varepsilon} \right) \right) +$$

$$+ \le \frac{1}{\varepsilon} d \left(\delta_\varepsilon c_1' \left(\frac{1}{\varepsilon} \right) \delta_\varepsilon c_2 \left(\frac{1}{\varepsilon} \right), \delta_\varepsilon c_1 \left(\frac{1}{\varepsilon} \right) \delta_\varepsilon c_2 \left(\frac{1}{\varepsilon} \right) \right)$$

But both quantities from the RHS tend to 0 as $\varepsilon \to 0$, because of the hypothesis.   ∎

On $\mathcal{A}/\equiv$ we have the following operation:

$$[y_\varepsilon^{-1} \delta_\varepsilon] \circ [u_\varepsilon^{-1} \delta_\varepsilon] \;=\; [y_\varepsilon^{-1} u_\varepsilon^{-1} \delta_\varepsilon]$$

(Check that the operation is well defined). This operation makes $\mathcal{A}/\equiv$ into a group. This group admits an one parameter family of isomorphisms:

$$\underline{\delta}_\mu [[y_\varepsilon^{-1} \delta_\varepsilon] \;=\; [y_{\varepsilon\mu}^{-1} \delta_{\varepsilon\mu}]$$

Therefore any Lie subgroup of $\mathcal{A}/\equiv$ is Carnot. A particular subgroup of $\mathcal{A}/\equiv$ is the spectrum of a curve.

Another observation is that a weak class can be seen as a bundle. The bijection between fibers (strong classes) is as follows: take in the strong class $[c_1]_s$ a representant $c_1 \overset{w}{\approx} c$. Then the function:

$$c' \in [c]_s \;\mapsto\; c_1 \circ c^{-1} \circ c' \; \in [c_1]_s$$

is a bijection.

However, $\mathcal{A}/\equiv$, or the spectrum of a curve are too big to be finite dimensional groups. That is why we shall restrict to smaller classes.

A particular class of finite curves is given by words. A word is a string $(x_1 y_1 .... x_p y_p)$ with $x_i, y_i \in G$. To any word it is associated the curve:

$$c = \begin{pmatrix} x_1 & & x_2 & ... & x_p & \\ & y_1 & & y_2 & ... & y_p \end{pmatrix} \; , \; c(\frac{1}{\varepsilon}) \;=\; \delta_\varepsilon^{-1} \left(x_1 \, \delta_\varepsilon y_1 \, ... x_p \delta_\varepsilon y_p \right)$$

Word concatenation correspond to curves composition.

For any $\mu > 0$ we define $(\delta_\mu \circ c) \left(\frac{1}{\varepsilon} \right) \;=\; \delta_\mu c(\frac{1}{\varepsilon\mu})$.

Proposition 3.21 *We have the following relations:*

(a) $\begin{pmatrix} x_1 & & x_2 & ... & x_p & \\ & y_1 & & y_2 & ... & y_p \end{pmatrix} \circ \begin{pmatrix} 0 & \\ & y \end{pmatrix} \overset{w}{\approx} \begin{pmatrix} x_1 & & x_2 & ... & x_p & \\ & y_1 & & y_2 & ... & \beta(y_p, y) \end{pmatrix}$

(b) $\begin{pmatrix} x & \\ & 0 \end{pmatrix} \circ \begin{pmatrix} x_1 & x_2 & ... & x_p \\ & y_1 & y_2 & ... & y_p \end{pmatrix} = \begin{pmatrix} x\ x_1 & x_2 & ... & x_p \\ & y_1 & y_2 & ... & y_p \end{pmatrix}$

(c) $\delta_\mu \circ \begin{pmatrix} x_1 & x_2 & ... & x_p \\ & y_1 & y_2 & ... & y_p \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & ... & x_p \\ & \delta_\mu y_1 & \delta_\mu y_2 & ... & \delta_\mu y_p \end{pmatrix}$

This proposition gives an interesting interpretation of the Margulis & Mostow tangent bundle.

Corollary 3.22 *The virtual tangent bundle (or the Margulis & Mostow tangent bundle) is the bundle of two-letters words, with basis the group and bundle projection on the first letter.*

The definition of the tangent bundle can be naturally given in at least two ways. Here is the first.

Definition 3.23 *The word tangent bundle WG is the group of strong equivalence classes of words. The bundle structure of WG is given by the map*

$$[c]_s \mapsto x_c = \lim_{\varepsilon \to 0} \delta_\varepsilon c \left(\frac{1}{\varepsilon} \right)$$

$W_0 G$ *is the fiber of the neutral element and it is a group.*

We shall explain further the structure of WG. For this we shall start from the class $\mathcal{C}^m([0, +\infty), \mathfrak{g})$ of all functions in $\mathfrak{g}$ which admit a development of order m around 0. To any $f \in \mathcal{C}^m([0, +\infty), \mathfrak{g})$ we associate the finite curve

$$c_f(\frac{1}{\varepsilon}) = \delta_\varepsilon^{-1}(f(\varepsilon))$$

Conversely, to any finite curve c we associate the function

$$f_c(\varepsilon) = \delta_\varepsilon \left(c \left(\frac{1}{\varepsilon} \right) \right)$$

This function is prolongated by continuity to $\varepsilon = 0$ (which is possible, because the curve is finite).

For any curve c associated to a word we have $f_c \in \mathcal{C}^m([0, +\infty), \mathfrak{g})$. We shall use further the following notations:

$$C(G) = \{c_f : f \in \mathcal{C}^m([0, +\infty), \mathfrak{g})\} / \overset{s}{\approx}$$

$$WG = \{c_f : f \text{ is a word } \} / \overset{s}{\approx}$$

The word tangent bundle WG can be seen as a subset of $C(G)$.

The tangent bundle WG is a group with the multiplication operation of strong classes of curves. We want to show that in fact is a Lie group.

Let $f \in \mathcal{C}^m([0, +\infty), \mathfrak{g})$ and

$$f(\varepsilon) \;=\; P(\varepsilon) + S(\varepsilon)$$

with $P \in \mathfrak{g}[\varepsilon]$ is a polynomial of degree at most m and S is of order $o(\varepsilon^m)$. We want to check if c_f and c_P are strongly equivalent. This is equivalent with $c_P^{-1} \circ c_f \overset{s}{\approx} 0$, which is straightforward.

We conclude therefore that

$$C(G) \;=\; \left(\mathfrak{g}[\varepsilon]/\varepsilon^{m+1}\mathfrak{g}[\varepsilon] \right) / \overset{s}{\approx}$$

in the sense that the sides of the previous equality can be identified. Let us denote

$$\mathfrak{g}^{(m)}[\varepsilon] \;=\; \mathfrak{g}[\varepsilon]/\varepsilon^{m+1}\mathfrak{g}[\varepsilon]$$

This is a factorisation of a Lie algebra by an ideal, hence it is itself a Lie algebra.

Recall that $\mathfrak{g}$ admits a direct sum decomposition and a filtration:

$$\mathfrak{g} \;=\; V_1 + ... + V_m \;, \quad V^i \;=\; V_1 + ... + V_i$$

such that $[V^i, V^j] \subset V^{i+j}$ (here $V^k = \mathfrak{g}$ for any $k \geq m$).

Any element of $\mathfrak{g}^{(m)}[\varepsilon]$ has a representant $P \in \mathfrak{g}[\varepsilon]$, $\deg P \leq m$,

$$P(\varepsilon) \;=\; P_1(\varepsilon) + ... + P_m(\varepsilon)$$

such that $P_i \in V_i[\varepsilon]$ for all i. From the degree condition we see that we can write (for any $i = 1, ..., m$):

$$P_i(\varepsilon) \;=\; \sum_{k=0}^{m} a_{ik}\varepsilon^k \;, \quad a_{ik} \in V_i$$

We can do the following decomposition:

$$P(\varepsilon) \;=\; P^W(\varepsilon) + P^S(\varepsilon)$$

$$P^W(\varepsilon)_i \;=\; \sum_{k=0}^{i} a_{ik}\varepsilon^k$$

Let c^1, c^2 two arbitrary finite curves and $f_i = f_{c^i}$, $i = 1, 2$. Then $c^1 \overset{w}{\approx} c^2$ if and only if

$$\delta_\varepsilon^{-1}\left(f_1^{-1}(\varepsilon)f_2(\varepsilon) \right) \to 0$$

as $\varepsilon \to 0$. From here (and the filtration of $\mathfrak{g}$) we deduce that $c_P \overset{w}{\approx} c_{PW}$, therefore the weak class of f is P^W, where P is the development of order m of f around 0.

We have to identify now the strong classes which compose a weak class. For this consider the sets:

$$Weak(\mathfrak{g}) \;=\; \left\{ P^W : P \in \mathfrak{g}[\varepsilon] \;, \quad \deg P \leq m \right\}$$

$$Strong(\mathfrak{g}) \;=\; \left\{ P^S : P \in \mathfrak{g}[\varepsilon] \;, \;\; \deg P \le m \right\}$$

$$O(\mathfrak{g}) \;=\; \left\{ P \in Strong(\mathfrak{g}) : c_P \overset{s}{\approx} 0 \right\}$$

The set $O(\mathfrak{g})$ can be seen as a normal subgroup of $\mathfrak{g}^{(m)}[\varepsilon]$. Indeed, $P \in O(\mathfrak{g})$ if and only if for any $x \in G$ we have

$$\delta_\varepsilon^{-1} Ad_x^G P(\varepsilon) \to 0$$

as $\varepsilon \to 0$. This implies that $O(\mathfrak{g})$ is Ad^G invariant, hence the claim.

Because of proposition 3.20 it follows that if $P \in O(\mathfrak{g})$ and $Q \in Strong(\mathfrak{g})$ then $PQ \in Strong(\mathfrak{g})$. Putting all together, we have obtained the characterisation of $C(G)$.

Theorem 3.24 $C(G)$ *is a Lie group with Lie algebra isomorphic to*

$$Weak(\mathfrak{g}) + (Strong(\mathfrak{g})/O(\mathfrak{g}))$$

For any $f \in \mathcal{C}^m([0,+\infty), G)$ with associated P, development of order m around 0, the weak class $[c_f]_w$ can be identified with P^W. The spectrum of c_f (in the same class of curves as c_f) does not depend on f and equals $Strong(\mathfrak{g})/O(\mathfrak{g})$.

WG is a closed subgroup of $C(G)$, hence it is a Lie group. $W_0 G$ is a subgroup of WG.

It will be useful further to work with $C(G)$. We shall denote by $T_0 G$ the fiber of 0 in $C(G)$, namely the set:

$$T_0 G \;=\; \{ f \in \mathcal{C}^m([0,+\infty), G) : f(0) = 0 \} / \overset{s}{\approx}$$

The notion which later will lead to derivability is transport.

Definition 3.25 *The continuous function $f : N \to M$ transports finite curves into finite curves according to the definition:*

$$f * c(\frac{1}{\varepsilon}) \;=\; \delta_\varepsilon^{-1} f(\delta_\varepsilon c(\frac{1}{\varepsilon}))$$

Let us see how the translations and dilatations act on the tangent bundle.

We start from the observation: translations and dilatations preserve strong classes. Indeed, let $f : G \to G$ be a left or right translation and $c_{f_1} \overset{s}{\approx} c_{f_2}$, with $f_1, f_2 \in \mathcal{C}^m([0,+\infty), G)$. We want to prove that $f * c_1 \overset{s}{\approx} f * c_2$. Suppose first that $f(x) = yx$. The claim is then true because the distance (or uniformity) in G is left invariant. If f is a right translation ($f(x) = xy$), take $z \in G$ arbitrary. Then

$$\frac{1}{\varepsilon} d(\delta_\varepsilon \left(f * c_1(\frac{1}{\varepsilon}) \right) z, \delta_\varepsilon \left(f * c_2(\frac{1}{\varepsilon}) \right) z) \;=\; \frac{1}{\varepsilon} d(f_1(\varepsilon) yz, f_2(\varepsilon) yz)$$

which converges to 0 as $\varepsilon \to 0$ because $c_1 \overset{s}{\approx} c_2$.

Recall that we defined a dilatation on the class of curves (and we used the concatenation notation)

$$\delta_\mu(c)(\frac{1}{\varepsilon}) = (\delta_\mu \circ c)(\frac{1}{\varepsilon}) = \delta_\mu c(\frac{1}{\varepsilon\mu})$$

with $\mu > 0$. This function preserves strong classes. Indeed:

$$\frac{1}{\varepsilon}d(\delta_\varepsilon\left(\delta_\mu \circ c_1(\frac{1}{\varepsilon})\right)z, \delta_\varepsilon\left(\delta_\mu \circ c_2(\frac{1}{\varepsilon})\right)z) \ = $$

$$= \ \mu\frac{1}{\varepsilon\mu}d(f_1(\varepsilon\mu)z, f_2(\varepsilon\mu)z)$$

which converges to 0 as $\varepsilon \to 0$ because $c_1 \overset{s}{\approx} c_2$.

We prove now that classical dilatations transport strong classes into strong classes. For this we shall use the following characterisation of strong equivalence: $c_1 \overset{s}{\approx} c_2$ if and only if for any z (in a neighbourhood of the neutral element) we have

$$d_N(\delta_\varepsilon^{-1}(f_1(\varepsilon) \overset{n}{\cdot} z), \delta_\varepsilon^{-1}(f_1(\varepsilon) \overset{n}{\cdot} z)) \to 0$$

as $\varepsilon \to 0$, uniformly with respect to z. Here d_N is the distance on the tangent cone at the neutral element (remember that, as in Vodop'yanov & Greshnov [32], we have identified a neighbourhood of the neutral element of the tangent cone with a neighbourhood of the neutral element of the group G; on this neighbourhood we have also the distance d_N).

With this characterisation $\delta_\mu * c_{f_1} \overset{s}{\approx} \delta_\mu * c_{f_2}$ if and only if

$$\frac{\mu}{\varepsilon}d_N(f_1(\varepsilon) \overset{n}{\cdot} \delta_\mu^{-1}z, f_1(\varepsilon) \overset{n}{\cdot} \delta_\mu^{-1}z) \to 0$$

as $\varepsilon \to 0$, uniformly with respect to z, which is true by hypothesis.

It is easy to see that translation and dilatations (in both senses: $\delta_\mu \circ ...$ and $\delta_\mu * ...$) preserve the class of words and the class of curved induced from functions in $\mathcal{C}^m([0, +\infty), G)$.

Coming back to transport, we notice that any Lipschitz maps f transports weak classes into weak classes. We cannot use this to define a tangent map the tangent bundles in the sense introduced here, because such tangent bundles are constructed with strong classes.

The following lemma (with straightforward proof) shows that Pansu derivative gives only a partial description about how f transports finite curves. Here by Pansu derivative we mean a derivative in the sense of definition 2.39.

Lemma 3.26 *If $f : N \to L$ is Pansu derivable in x then it transports any two letter word of the form* $\begin{pmatrix} x \\ y \end{pmatrix}$ *to a curve weakly equivalent to the word* $f(x)\begin{pmatrix} 0 \\ Pf(x)y \end{pmatrix}$, *where $Pf(x)$ denotes the Pansu derivative of f at x.*

From the point of view developed in this section, the meaning of Rademacher theorem is that any Lipschitz function transports two letter word weak classes into two letter word weak classes, almost everywhere with respect to the first letter.

The problem that we have in order to introduce a derivative notion is that we would like the derivative of f at x to act on strong classes, but (the Pansu derivative) provide a transformation of (a particular family of) weak classes.

Before introducing the derivative notion in this setting, we shall look closer to Carnot groups.

4 Calculus on Carnot groups

Let N be a Carnot group. We shall develop the notions introduced in the previous section. But let us first justify the approach by looking at the right translations in the Heisenberg group.

Consider therefore the right translation $f : H(n) \to H(n)$,

$$f(x, \bar{x}) = (x + y, \bar{x} + \bar{y} + \frac{1}{2}\omega(x, y))$$

We look at $f(\tilde{x}\delta_\varepsilon\tilde{z})$, with $\tilde{x}, \tilde{z} \in H(n)$, $\tilde{x} = (x, \bar{x})$, $\tilde{z} = (z, \bar{z})$. Then

$$f(\tilde{x}\delta_\varepsilon\tilde{z}) = f(x + \varepsilon z, \bar{x} + \varepsilon\frac{1}{2}\omega(x, z) + \varepsilon^2\bar{z}) =$$

$$= (x + y + \varepsilon z, \bar{x} + \bar{y} + \frac{1}{2}\omega(x, y) + \varepsilon\left(\frac{1}{2}\omega(z, y) + \frac{1}{2}\omega(x, z)\right) + \varepsilon^2\bar{z})$$

We compute now the expression

$$(f(\tilde{x}))^{-1} f(\tilde{x}\delta_\varepsilon\tilde{z}) = (\varepsilon z, \varepsilon\omega(z, y) + \varepsilon^2\bar{z})$$

Because of the term $\varepsilon\omega(z, y)$, the right translation is not derivable, unless $y = 0$. Nevertheless, what is to be remarked is that the polynomial expression $(\varepsilon z, \varepsilon^2\bar{z})$ has been transported to $(\varepsilon z, \varepsilon\omega(z, y) + \varepsilon^2\bar{z})$, which is not so bad.

We go back now to general Carnot groups. We shall see that in this case we can pass from weak classes to strong classes by choosing representantives in a canonical way.

Definition 4.1 *Let $C(N)$ the group of polynomial curves*

$$\varepsilon > 0 \mapsto (P_i(\varepsilon))_{i=1,\ldots,n} \ , \ deg P_i(\varepsilon) \le i \ , \ P_i(\varepsilon) \in V_i[\varepsilon]$$

This is a group with respect to the pointwise multiplication in N.

To any variational word

$$c = \begin{pmatrix} x_1 & & x_2 & \ldots & x_p & \\ & y_1 & & y_2 & \ldots & y_p \end{pmatrix}$$

we associate the element $P_c(\varepsilon) = \delta_\varepsilon c(\frac{1}{\varepsilon})$. The class of all such polynomials, denoted by WN, forms a subgroup of $C(N)$.

Conversely, to any $P \in C(N)$ we associate the finite curve

$$c_P(\frac{1}{\varepsilon}) = \delta_\varepsilon^{-1} P(\varepsilon)$$

A curve c is *good* if $c \overset{w}{\approx} c_P$ for some $P \in C(N)$. The curve c is *very good* if $c \overset{w}{\approx} c_P$ with $P \in WN$.

The tangent bundle of N is by definition WN. We collect in a proposition some basic facts about weak and strong classes of good and very good curves.

Proposition 4.2   *(a) Let $P, Q \in C(N)$. $c_P \overset{w}{\approx} c_Q$ if and only if $P = Q$. Therefore on the class of good curves the weak classes are equal with strong classes. Otherwise said: the spectrum of a good curve is trivial (in the class of good curves).*

  (b) The composition operation between strong classes of good and very good curves descends on the group operation on $C(N)$, respectively WN.

To any left or right translation in N there is an associated map on the tangent bundle. Indeed, let $x \in N$; we look at the transport given by left translations of variational words:

$$L_x * \begin{pmatrix} x_1 & & x_2 & ... & x_p & \\ & y_1 & & y_2 & ... & y_p \end{pmatrix} = \begin{pmatrix} xx_1 & & x_2 & ... & x_p & \\ & y_1 & & y_2 & ... & y_p \end{pmatrix}$$

In the same way

$$R_x * \begin{pmatrix} x_1 & & x_2 & ... & x_p & \\ & y_1 & & y_2 & ... & y_p \end{pmatrix} = \begin{pmatrix} x_1 & & x_2 & ... & x_p & & x \\ & y_1 & & y_2 & ... & y_p & & 0 \end{pmatrix}$$

For that dilation δ_μ we have:

$$\delta_\mu * \begin{pmatrix} x_1 & & x_2 & ... & x_p & \\ & y_1 & & y_2 & ... & y_p \end{pmatrix} = \begin{pmatrix} \delta_\mu x_1 & & \delta_\mu x_2 & ... & \delta_\mu x_p & \\ & \delta_\mu y_1 & & \delta_\mu y_2 & ... & \delta_\mu y_p \end{pmatrix}$$

We also have the formula

$$\delta_\mu \circ \begin{pmatrix} x_1 & & x_2 & ... & x_p & \\ & y_1 & & y_2 & ... & y_p \end{pmatrix} = \begin{pmatrix} x_1 & & x_2 & ... & x_p & \\ & \delta_\mu y_1 & & \delta_\mu y_2 & ... & \delta_\mu y_p \end{pmatrix}$$

Remark that the tangent bundle is a Carnot group. We consider the generating distribution

$$D^2 = \left\{ \begin{pmatrix} x & \\ & y \end{pmatrix} : x, y \in D \right\}$$

and the Euclidean norm

$$\left| \begin{pmatrix} x & \\ & y \end{pmatrix} \right| = \left(| x |^2 + | y |^2 \right)^{\frac{1}{2}}$$

(although any L_p norm is good as well). This gives us a distance on WN, denoted by d_W.

Remark 4.3 *The fact that the tangent bundle is a Carnot group is a key property, because we stay in the same category after derivation. Hence one can do an indefinite number of derivatives.*

The tangent space at the neutral element is

$$T_0 N = \{ P \in C(N) : P(0) = 0 \}$$

and it is a Carnot group too. The dimension of $T_0 N$ equals the homogeneous dimension of N. The tangent bundle WN is a semidirect product between N and $W_0 N$.

The dilatations in $T_0 N$ are defined by:

$$\delta_\mu(P)(\varepsilon) \; = \; P(\mu \varepsilon)$$

for any $\mu > 0$.

Definition 4.4 *The function $f : N \to L$ is weakly derivable at $x \in N$ if there is a function $Tf(x) : T_0 N \to WL$ with the following property:*

$$\frac{1}{\varepsilon} d \left(f(xP(\varepsilon)), Tf(x)(P)(\varepsilon) \right) \to 0 \tag{4.0.1}$$

as $\varepsilon \to 0$, uniformly with respect to $P \in B(0,1) \subset T_0 N$.

It is straightforward from this definition that there exist $Df(x) : T_0 N \to T_0 L$ such that

$$Tf(x)(P) \; = \; f(x) Df(x)(P)$$

and $Df(x)$ commutes with dilatations.

The following is a simple, but important observation.

Proposition 4.5 *Let $f : N \to N$ be a function, $f = (f_1, ..., f_m)$, where m is the step of N. Suppose that f_i admits a development of order i around x. Then f is weakly derivable.*

The equivalent of proposition 3.10 is not true for weak derivatives, because the quantities from the proof, denoted by $(1)_t$, $(2)_t$ might converge to infinity. That is why we state two more definitions.

Definition 4.6 *The function $f : N \to L$ is derivable at $x \in N$ if there exists **a group morphism** $Df(x) : T_0 N \to T_0 L$ such that*

$$\frac{1}{\varepsilon} d \left(f(xP(\varepsilon)), f(x) Df(x)(P)(\varepsilon) \right) \to 0$$

as $\varepsilon \to 0$, uniformly with respect to $P \in B(\bar{0}, 1) \subset T_0 N$.

Definition 4.7 *The function $f : N \to L$ is strongly derivable at $x \in N$ if there exists $Df(x) : T_0 N \to T_0 L$ such that*

$$\frac{1}{\varepsilon} d \left(f(xP(\varepsilon))y, f(x) Df(x)(P)(\varepsilon)y \right) \to 0$$

as $\varepsilon \to 0$, for any $y \in N$, uniformly with respect to $P \in B(\bar{0}, 1) \subset T_0 N$.

Now comes the equivalent of proposition 3.10.

Proposition 4.8 *Let $f : N \to L$ be a function.*

a) *If f is strongly derivable at x then it is derivable.*

b) *If f is weakly derivable and*

- *the convergence in the definition of the derivative is uniform with respect to x,*
- *f satisfies the following condition (A) (from "additive"): for any $P, Q \in B(0, 1) \subset T_0 N$ and any $x \in N$*

   (A) $\dfrac{1}{\varepsilon} d\left(f(xP(\varepsilon)Q(\varepsilon)), f(xP(\varepsilon))f(x)^{-1}f(xQ(\varepsilon))\right) \to 0$ (4.0.2)

   goes to 0 as $\varepsilon \to 0$, uniformly with respect to P, Q,

then f is derivable.

One can imagine many sufficient conditions for (A). One of them is the following: there exists $C > 0$ such that for any $u, x, z \in N$ we have

$$d(f(u)^{-1}f(uz), f(x)^{-1}f(xz)) \leq Cd(u, x)d(0, z)^m \qquad (4.0.3)$$

Let us see what (4.0.3) means in the commutative case $N = R^n$. Take $f : R^n \to R^k$ and $x, y \in R^n$. Construct the function $g_{xy}(z) = f(x + z) - f(y + z)$. If f has Lipschitz derivatives then it satisfies condition (4.0.3), which in this case reads:

$$\|g_{xy}(z) - g_{xy}(0)\| \leq C\|x - y\|\|z\|$$

The constant C equals $Lip(\nabla f)$.

Proof.   (of proposition 4.8) We shall prove only the point (b), because (a) has an easy proof. We have to prove the morphism property of $Df(x)$.

Let $P, Q \in T_0 N$. Because $Df(y)$ commutes with dilatations for any $y \in U$, we can suppose that $P, Q, \in B(\bar{0}, 1) \subset T_0 N$.

We start from the inequality:

$$\frac{1}{\varepsilon} d\left(f(x(P)(\varepsilon)Q(\varepsilon)), f(x)Df(x)(P)(\varepsilon)Df(x)(Q)(\varepsilon)\right) \leq A(\varepsilon) + B(\varepsilon)$$

with $A(\varepsilon) \to 0$ because of condition (A). We used the notations:

$$A(\varepsilon) \;=\; \frac{1}{\varepsilon} d\left(f(xP(\varepsilon)Q(\varepsilon)), f(xP(\varepsilon))f(x)^{-1}f(xQ(\varepsilon))\right)$$

$$B(\varepsilon) \;=\; \frac{1}{\varepsilon} d\left(f(xP(\varepsilon))f(x)^{-1}f(xQ(\varepsilon)), f(x)Df(x)(P)(\varepsilon)Df(x)(Q)(\varepsilon)\right)$$

We continue:

$$B(\varepsilon) \;\leq\; C(\varepsilon) + D(\varepsilon)$$

with the notations

$$C(\varepsilon) \;=\; \frac{1}{\varepsilon} d\left(f(xP(\varepsilon))f(x)^{-1}f(xQ(\varepsilon)), f(xP(\varepsilon))Df(x)(Q)(\varepsilon)\right)$$

$$D(\varepsilon) \;=\; \frac{1}{\varepsilon} d\left(f(xP(\varepsilon))Df(x)(Q)(\varepsilon), f(x)Df(x)(P)(\varepsilon)Df(x)(Q)(\varepsilon)\right) \qquad (4.0.4)$$

Remark 4.9 *The quantity $D(\varepsilon)$ will be important in the section concerning general groups, in relation to the definition 6.1 of mild equivalence.*

The quantity $C(\varepsilon)$ converges to 0 as $\varepsilon \to 0$ because f is weakly derivable. The second quantity $D(\varepsilon)$ behaves in the same way again because f is weakly derivable, because of the uniformity assumption in part (b) of the hypothesis and because of proposition 4.2 (a). We have proved that

$$\frac{1}{\varepsilon} d\left(f(x(P)(\varepsilon)Q(\varepsilon)), f(x)Df(x)(P)(\varepsilon)Df(x)(Q)(\varepsilon)\right)$$

converges to 0 as $\varepsilon \to 0$, uniformly with respect to $P, Q, \in B(\bar{0}, 1) \subset T_0 N$.

The one thing left to do is to use once again the fact that f is weakly derivable, this time along $P(\varepsilon)Q(\varepsilon)$. This fact, together with the convergence previously established, end the proof of the morphism property. ∎

It seems hard to believe that these three definitions of derivability are really different. This is a purely non-Euclidean property. We shall see in the section dedicated to the Heisenberg group that the differences are really big and that the good notion of derivability seems to be 4.6.

The Rademacher theorem for our notion of derivative is:

Theorem 4.10 *Let $f : N \to L$ be a Lipschitz function which satisfies condition (A). Then f is derivable a.e. and the derivative of f has the property:*

$$Df(x)\begin{pmatrix} x \\ & y \end{pmatrix} = \begin{pmatrix} 0 \\ & Pf(x)y \end{pmatrix}$$

The proof is a direct consequence of Pansu-Rademacher theorem, propositions 4.5, 4.8 and the observation that the Margulis-Mostow tangent bundle (bundle of two-letter words) generates the word tangent bundle.

We end this section with the remark that the derivative satisfies the chain rule by the very definition.

5 Case of the Heisenberg group

In this case $CH(n) = WH(n)$, which comes from the fact that two letter words generate $CH(n)$. $T_0 H(n)$ is the group of all polynomials with the form:

$$P(\varepsilon) = (u, u', u")(\varepsilon) = (\varepsilon u, \varepsilon u' + \varepsilon^2 u")$$

with $u \in R^{2n}, u', u" \in R$. This group is isomorphic with $H(n) \times R$ and the isomorphism is given by:

$$(u, u', u") \in T_0 H(n) \mapsto ((u, u"), u') \in H(n) \times R$$

The morphisms from R to $T_0 H(n)$ which commute with dilatations forms a set called $HL(R, T_0 H(n))$ ("HL" means "horizontal linear"). Any element F of this set can be seen as an element $(a, a') \in R^{2n} \times R$, meaning that $F \in HL(R, T_0 H(n))$ has the form:

$$F(u) = (ua, u'a', 0)$$

Derivable curves Let $\tilde{f} : R \times H(n)$, $f(t) = (f(t), \bar{f}(t))$ derivable. This means that there is $(a(t), a'(t))$ such that for any $u \in R, | u | < 1$ we have:

$$\delta_\varepsilon^{-1} \left((-f(t + \varepsilon u), -\bar{f}(t + \varepsilon u))(f(t), \bar{f}(t))(\varepsilon u a(t), \varepsilon u a' t) \right) \to 0$$

as $\varepsilon \to 0$, uniformly with respect to u. We obtain the following characterisation of $\tilde{f}$:

   a) f is derivable with respect to t and $a(t) = \dot{f}(t)$,

   b) the map $s \mapsto \bar{f}(t + s) - \dfrac{1}{2}\omega(f(t), f(t + s))$ has a second order development at
$s = 0$ of the form:

$$\bar{f}(t + s) - \frac{1}{2}\omega(f(t), f(t + s)) = \bar{f}(t) + sa'(t) + o(s^2)$$

Putting a) and b) together we get: $\tilde{f}$ is classically derivable and

$$\begin{cases} a(t) = \dot{f}(t) \\ a'(t) = \dot{\bar{f}} - \frac{1}{2}\omega(f(t), \dot{f}(t)) \\ \frac{d}{dt}a'(t) = 0 \end{cases}$$

So the curve $\tilde{f}$ is not horizontal generally, but the difference from horizontality, measured by a', is constant along the curve. A straightforward computation says that when f is parametrised by arclength, the 2 Hausdorff measure of $\tilde{f}$ is $a'L(f)$, where $L(f)$ is the length of f.

    An analoguous computation shows that $\tilde{f}$ is strongly derivable if and only if it is linear.

Derivative of the right translation The right translations are derivable. We continue with the motivating example from the beginning of the previous section.

    We saw there that the right translation $f : H(n) \to H(n)$, $f(\tilde{x}) = \tilde{x}\tilde{y}$ transports the polynomial $(z, 0, \bar{z})$ to $(z, \omega(z, y), \bar{z})$. The same computation shows that

$$Df(\tilde{x})(u, u', u") \;=\; (u, u' + \omega(u, y), u")$$

We can write this derivative as a matrix:

$$Df(\tilde{x}) \;=\; \begin{pmatrix} 1 & 0 & 0 \\ -Jy & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and check that $Df(\tilde{x})$ is indeed a morphism of $T_0 H(n)$ which commute with dilatations. That means $Df(\tilde{x}) \in HL(T_0 H(n))$, where

$$HL(T_0 H(n)) \;=\; \{F \in GL(T_0 H(n)) : \forall x, y \in T_0 H(n) \; [Fx, Fy] = F[x, y] \;,\;\; \forall \varepsilon > 0 \; \delta_\varepsilon F = F\delta_\varepsilon\}$$

Remark 5.1 *It has been shown in Ambrosio, Rigot [2] that the problem of optimal transportation in the Heisenberg group has solutions. In classical situations the optimal transport map ψ satisfies the Monge-Ampère equation, which contains $\det \nabla\psi$. In the case of the Heisenberg group one can prove that right translations are optimal transport maps (for a particular choice of the measures to be transported). But right translations are not Pansu differentiable, hence the problem of understanding the Monge-Ampère equation in this setting. We don't deal in this paper with the calculus of variations associated to the notions of derivative just introduced. This will be done in a future paper. I think that it is encouraging though the fact that right translations are derivable, which might lead to an interpretation of the Monge-Ampère equation in the Heisenberg group.*

Derivable homeomorphisms   We look first for the class of isomorphisms of $T_0H(n)$ which commute with dilatations. This class is denoted by $HL(T_0H(n))$. We get $\mathcal{A} \in HL(T_0H(n))$ if and only if $\mathcal{A}$ is represented by a matrix:

$$\begin{pmatrix} A & 0 & 0 \\ c & d & 0 \\ 0 & 0 & h \end{pmatrix}$$

such that A is conformally symplectic and $h = (\det A)^{\frac{1}{n}}$.

   We proceed like in the case of curves: let $f : H(n) \to H(n)$, $f(x,\bar{x}) = (\phi(x,\bar{x}), \bar{\phi}(x,\bar{x}))$. If f is derivable at $(x,\bar{x})$ then the first conclusion is that $\phi = \phi(x)$. An interesting consequence is obtained if f is also volume preserving. In this case ϕ is symplectic and moreover

$$\bar{\phi}(x,\bar{x}) = \bar{x} + c \cdot x + F(x)$$

where F is the generating function of ϕ and c does not depend on $(x,\bar{x})$.

   We paste here the section 3.2, 3.3 of [5], for completeness of the picture of derivability in the Heisenberg group. Before this we state Theorem 5.2, which is correct version of theorem 3.8. from [5].

Theorem 5.2 *Take any $\tilde{\phi}$ locally bi-Lipschitz volume preserving homeomorphism of $H(n)$, **which is derivable***. *Then $\tilde{\phi}$ has the form:*

$$\tilde{\phi}(x,\bar{x}) = (\phi(x), \bar{x} + F(x))$$

Moreover

$$\phi \in Sympl(R^{2n}, Lip)$$

$F : R^{2n} \to R$ is Lipschitz and for almost any point $(x,\bar{x}) \in H(n)$ we have:

$$DF(x)y = \frac{1}{2}\omega(\phi(x), D\phi(x)y) - \frac{1}{2}\omega(x,y)$$

Proof.   Let $(u, u', u") \in T_0 H(n)$ and $(x, \bar{x}) \in H(n)$. The condition on f to be derivable has the form:

$$E(\varepsilon) = \delta_\varepsilon^{-1} \left\{ (-\phi((x, \bar{x})(u, u', u")(\varepsilon)), -\bar{\phi}((x, \bar{x})(u, u', u")(\varepsilon)))(\phi(x, \bar{x}), \bar{\phi}(x, \bar{x})) \right.$$

$$\left. (\varepsilon A u, \varepsilon c u + \varepsilon d u' + \varepsilon^2 h u") \right\} \to 0$$

as $\varepsilon \to 0$, uniformly with respect to $(u, u', u")$ such that

$$\mid u \mid + \mid u' \mid + \mid u" \mid^{\frac{1}{2}} \leq 1$$

Here A is conformally symplectic and $h = (\det A)^{\frac{1}{n}}$. The R^{2n} component of the $E(\varepsilon)$ is

$$\frac{1}{\varepsilon} \left(\phi(x, \bar{x}) - \phi(x + \varepsilon u, \bar{x} + \varepsilon u' + \varepsilon \frac{1}{2} \omega(x, u) + \varepsilon^2 u") + \varepsilon A u \right)$$

and it has to converge to 0 as $\varepsilon \to 0$. Take now $u = 0$, $u' = 0$ and get that ϕ is derivable with respect to $\bar{x}$ and the derivative equals 0. Therefore $\phi = \phi(x)$. From here we get that ϕ is classically derivable with respect to x and

$$A = \frac{\partial \phi}{\partial x}$$

The second component of $E(\varepsilon)$ admits a similar treatment. We leave the details to the reader. The conclusions are:

1) $\bar{\phi}$ is derivable with respect to x and $\bar{x}$,

2) we have $\dfrac{\partial \bar{\phi}}{\partial \bar{x}} = d = h$,

3) we have also the equality

$$cu = \frac{\partial \bar{\phi}}{\partial x} u - \frac{1}{2} \omega(\phi, \frac{\partial \phi}{\partial x} u) + \frac{\partial \bar{\phi}}{\partial \bar{x}} \frac{1}{2} \omega(x, u)$$

for any $u \in R^{2n}$.

4) the coefficient c has to be constant with respect to $(x, \bar{x})$.

The function f is Lipschitz therefore Pansu derivable almost everywhere. Because f is also volume preserving, the area formula (change of variables formula, we have not touched this subject in this paper, see for example Vodop'yanov, Ukhlov [30] [34] or Pauls [27] and Magnani [21]) implies that a.e. A have to be symplectic, hence $h = d = 1$.
   Use 2), 3), 4) to get

$$\frac{\partial \bar{\phi}}{\partial x} = \frac{1}{2} \omega(\phi, \frac{\partial \phi}{\partial x} u) - \frac{1}{2} \omega(x, u) + c$$

But the generating function F of ϕ is defined up to an additive constant by

$$DF(x)u = \frac{1}{2} \omega(\phi, \frac{\partial \phi}{\partial x} u) - \frac{1}{2} \omega(x, u)$$

(This function exists, as a consequence of the fact that ϕ is symplectic a.e.) The proof is finished.   ∎

5.1   Hamiltonian dynamics in a nutshell

The purpose of this section is to use the sub-Riemannian calculus on the Heisenberg group to recover the basic notions in Hamiltonian dynamics. We show that Hamiltonian dynamics is just the study of flows of volume preserving derivable homeomorphisms of the Heisenberg group. Notions of generating function, hamiltonian, Hamilton equation, Hofer distance, appear without premeditation. One might wonder what is the result of the study of volume preserving derivable homeomorphisms in a general Carnot group. The case of H-type Carnot groups may as well be linked to Clifford analysis (because there is a natural correspondence between Clifford algebras and H-type groups, see Reimann [29] or Barbano [3]).

Classically smooth maps and the derivative: lifts of symplectic diffeomorphisms   In this section we are interested in the group of volume preserving classically diffeomorphisms of $H(n)$, which are also derivable in our new sense. Recall a previous proposition which roughly says that a sufficiently regular function is weakly derivable. In order that the function to be derivable some supplementary conditions upon higher order derivatives have to be satisfied. What is interesting in this subsection is the proof that classically smooth enough **volume preserving** diffeomorphisms are also derivable, not only weakly derivable.

Volume preserving diffeomorphisms   Let $A \subset R^{2n}$, open, such that $0 \notin A$. This condition does not restrict the generality; it leads only to a simplified notation.

Definition 5.3 $Diff^2(A, vol)$ *is the group of volume preserving diffeomorphisms $\tilde{\phi}$ of $H(n)$ such that:*

a) *$\tilde{\phi}$ has horizontal compact support in $A \times R \subset H(n)$, in the sense that it differs from a map $(x, \bar{x}) \mapsto (x, \bar{x} + c)$ ($c \in R$, arbitrar), only in a compact set included in $A \times R$,*

b) *$\tilde{\phi}$ and it's inverse have (classical) regularity C^2.*

In the same way we define $Sympl^2(A)$ to be the group of C^2 symplectomorphisms of R^{2n} with compact support in A.

For the regularity chosen in the definition (5.3), the result equivalent with theorem (5.2) is given further. We denote further by λ a one-form on R^{2n} such that $2 \, d\lambda = \omega$.

Theorem 5.4 *We have the isomorphism of groups*

$$Diff^2(A, vol) \approx Sympl^2(A) \times R$$

given by the mapping

$$\tilde{f} = (f, \bar{f}) \in Diff^2(A, vol) \mapsto \left(f \in Sympl^2(A), \bar{f}(0, 0) \right)$$

The inverse of this isomorphism has the expression

$$\left(f \in Sympl^2(A), a \in R\right) \;\mapsto\; \tilde{f} = (f, \bar{f}) \;\in\; Diff^2(A, vol)$$

$$\tilde{f}(x, \bar{x}) \;=\; (f(x), \; \bar{x} + F(x))$$

where $F(0) = a$ and $dF \;=\; f^*\lambda \;-\; \lambda$.

Remark 5.5 *We need the assumption $0 \notin A$ only for the morphism property.*

Proof.   Let $\tilde{f} = (f, \bar{f}) : H(n) \to H(n)$ be an element of the group
$Diff^2(A, vol)$. We shall compute:

$$D\tilde{f}((x, \bar{x}))(y, \bar{y}) \;=\; \lim_{\varepsilon \to 0} \delta_{\varepsilon^{-1}} \left(\left(\tilde{f}(x, \bar{x})\right)^{-1} \tilde{f}\left((x, \bar{x})\delta_\varepsilon(y, \bar{y})\right) \right)$$

We know that $D\tilde{f}(x, \bar{x})$ has to be a linear mapping.

After a short computation we see that we have to pass to the limit $\varepsilon \to 0$ in the
following expressions (representing the two components of $D\tilde{f}((x, \bar{x}))(y, \bar{y})$):

$$\frac{1}{\varepsilon} \left(f\left(x + \varepsilon y, \bar{x} + \varepsilon^2 \bar{y} + \frac{\varepsilon}{2}\omega(x, y)\right) - f(x, \bar{x}) \right) \tag{5.1.1}$$

$$\frac{1}{\varepsilon^2} \left(\bar{f}\left(x + \varepsilon y, \bar{x} + \varepsilon^2 \bar{y} + \frac{\varepsilon}{2}\omega(x, y)\right) - \bar{f}(x, \bar{x}) - \tag{5.1.2}$$

$$-\frac{1}{2}\omega\left(f(x, \bar{x}), f\left(x + \varepsilon y, \bar{x} + \varepsilon^2 \bar{y} + \frac{\varepsilon}{2}\omega(x, y)\right)\right) \right)$$

The first component (5.1.1) tends to

$$\frac{\partial f}{\partial x}(x, \bar{x})y + \frac{1}{2}\frac{\partial f}{\partial \bar{x}}(x, \bar{x})\omega(x, y)$$

The terms of order ε must cancel in the second component (5.1.2). We obtain the cancellation condition (we shall omit from now on the argument $(x, \bar{x})$ from all functions):

$$\frac{1}{2}\omega(x, y)\frac{\partial \bar{f}}{\partial \bar{x}} - \frac{1}{2}\omega(f, \frac{\partial f}{\partial x}y) - \frac{1}{4}\omega(x, y)\omega(f, \frac{\partial f}{\partial \bar{x}}) + \frac{\partial \bar{f}}{\partial x} \cdot y \;=\; 0 \tag{5.1.3}$$

The second component tends to

$$\frac{\partial \bar{f}}{\partial \bar{x}}\bar{y} - \frac{1}{2}\omega(f, \frac{\partial f}{\partial \bar{x}})\bar{y}$$

The group morphism $D\tilde{f}(x, \bar{x})$ is represented by the matrix:

$$d\tilde{f}(x, \bar{x}) \;=\; \begin{pmatrix} \frac{\partial f}{\partial x} + \frac{1}{2}\frac{\partial f}{\partial \bar{x}} \otimes Jx & 0 \\ 0 & \frac{\partial \bar{f}}{\partial \bar{x}} - \frac{1}{2}\omega(f, \frac{\partial f}{\partial \bar{x}}) \end{pmatrix} \tag{5.1.4}$$

We shall remember now that $\tilde{f}$ is volume preserving. This implies:

$$\frac{\partial f}{\partial x} + \frac{1}{2}\frac{\partial f}{\partial \bar{x}} \otimes Jx \ \in Sp(n) \tag{5.1.5}$$

$$\frac{\partial \bar{f}}{\partial \bar{x}} - \frac{1}{2}\omega(f, \frac{\partial f}{\partial \bar{x}}) = 1 \tag{5.1.6}$$

The cancellation condition (5.1.3) and relation (5.1.6) give

$$\frac{\partial \bar{f}}{\partial x}y \ = \ \frac{1}{2}\omega(f, \frac{\partial f}{\partial x}y) \ - \ \frac{1}{2}\omega(x,y) \tag{5.1.7}$$

These conditions describe completely the class of volume preserving diffeomorphisms of $H(n)$. Conditions (5.1.6) and (5.1.7) are in fact differential equations for the function $\bar{f}$ when f is given. However, there is a compatibility condition in terms of f which has to be fulfilled for (5.1.7) to have a solution $\bar{f}$. Let us look closer to (5.1.7). We can see the symplectic form ω as a closed 2-form. Let λ be a 1-form such that $2d\lambda = \omega$. If we take the (regular) differential with respect to x in (5.1.7) we quickly obtain the compatibility condition

$$\frac{\partial f}{\partial x} \ \in \ Sp(n) \tag{5.1.8}$$

and (5.1.7) takes the form:

$$d\bar{f} \ = \ f^*\lambda \ - \ \lambda \tag{5.1.9}$$

(all functions seen as functions of x only).

Conditions (5.1.8) and (5.1.5) imply: there is a scalar function $\mu = \mu(x, \bar{x})$ such that

$$\frac{\partial f}{\partial \bar{x}} \ = \ \mu \ Jx$$

Let us see what we have until now:

$$\frac{\partial f}{\partial x} \ \in \ Sp(n) \tag{5.1.10}$$

$$\frac{\partial \bar{f}}{\partial x} \ = \ \frac{1}{2}\left[\left(\frac{\partial f}{\partial x}\right)^T Jf \ - \ Jx\right] \tag{5.1.11}$$

$$\frac{\partial \bar{f}}{\partial \bar{x}} \ = \ 1 + \frac{1}{2}\omega(f, \frac{\partial f}{\partial \bar{x}}) \tag{5.1.12}$$

$$\frac{\partial f}{\partial \bar{x}} \ = \ \mu \ Jx \tag{5.1.13}$$

Now, differentiate (5.1.11) with respect to $\bar{x}$ and use (5.1.13). In the same time differentiate (5.1.12) with respect to x. From the equality

$$\frac{\partial^2 \bar{f}}{\partial x \partial \bar{x}} \ = \ \frac{\partial^2 \bar{f}}{\partial \bar{x} \partial x}$$

we shall obtain by straightforward computation $\mu = 0$.

The morphism property is easy to check.   ∎

Hamilton's equations   For any element $\phi \in Sympl^2(A)$ we define the lift $\tilde{\phi}$ to be the image of $(\phi, 0)$ by the isomorphism described in the theorem 5.4.

Definition 5.6 *For any flow* $t \mapsto \phi_t \in Sympl^2(A)$ *denote by* $\phi^h(\cdot, x))$ *the horizontal flow in* $H(n)$ *obtained by the lift of all curves* $t \mapsto \phi(t, x)$ *and by* $\tilde{\phi}(\cdot, t)$ *the flow obtained by the lift of all* ϕ_t. *The vertical flow is defined by the expression*

$$\phi^v \;=\; \tilde{\phi}^{-1} \circ \phi^h \tag{5.1.14}$$

Relation (5.1.14) can be seen as Hamilton equation, with known quantity ϕ^v and unknown quantities $\phi^h, \tilde{\phi}$. This is explained further.

Proposition 5.7 *Let* $t \in [0, 1] \mapsto \phi_t^v$ *be a curve of diffeomorphisms of* $H(n)$ *satisfying the equation:*

$$\frac{d}{dt}\phi_t^v(x, \bar{x}) \;=\; (0, H(t, x)) \;\;, \;\; \phi_0^v \;=\; id_{H(n)} \tag{5.1.15}$$

Consider the Hamiltonian flow $t \mapsto \phi_t$ *generated by* H *and lift it using theorem 5.4 to the flow* $t \mapsto \tilde{\phi}_t$; *lift it also horizontally, as explained in definition 5.6, to the flow* $t \mapsto \phi_t^h$. *Then equation (5.1.14) is satisfied.*

Conversely, for any Hamiltonian flow $t \mapsto \phi_t$, *generated by* H, *the vertical flow* $t \mapsto \phi_t^v$ *(defined this time by the relation (5.1.14)) satisfies the equation (5.1.15).*

Proof.   Write the lifts $\tilde{\phi}_t$ and ϕ_t^h, compute then the differential of the quantity $\dot{\tilde{\phi}}_t - \dot{\phi}_t^h$ and show that it equals the differential of H. Use finally the fact that all quantities in the proof have compact support.   ■

Flows of volume preserving diffeomorphisms   We want to know if there is any nontrivial smooth (according to Pansu differentiability) flow of volume preserving diffeomorphisms.

Proposition 5.8 *Suppose that* $t \mapsto \tilde{\phi}_t \in Diff^2(A, vol)$ *is a flow such that*

- *is* C^2 *in the classical sense with respect to* (x, t),

- *is horizontal, that is* $t \mapsto \tilde{\phi}_t(x)$ *is a horizontal curve for any* x.

Then the flow is constant.

Proof.   By direct computation, involving second order derivatives. Indeed, let $\tilde{\phi}_t(x, \bar{x}) = (\phi_t(x), \bar{x} + F_t(x))$. From the condition $\tilde{\phi}_t \in Diff^2(A, vol)$ we obtain

$$\frac{\partial F_t}{\partial x}y \;=\; \frac{1}{2}\omega(\phi_t(x), \frac{\partial \phi_t}{\partial x}(x)y) \;-\; \frac{1}{2}\omega(x, y)$$

and from the hypothesis that $t \mapsto \tilde{\phi}_t(x)$ is a horizontal curve for any x we get

$$\frac{dF_t}{dt}(x) \;=\; \frac{1}{2}\omega(\phi_t(x), \dot{\phi}_t(x))$$

Equal now the derivative of the RHS of the first relation with respect to t with the derivative of the RHS of the second relation with respect to x. We get the equality, for any $y \in R^{2n}$:

$$0 \;=\; \frac{1}{2}\omega(\frac{\partial \phi_t}{\partial x}(x)y, \dot{\phi}_t(x))$$

therefore $\dot{\phi}_t(x) \;=\; 0$.   ∎

One should expect such a result to be true, based on two remarks. The first, general remark: take a flow of left translations in a Carnot group, that is a flow $t \mapsto \phi_t(x) \;=\; x_t x$. We can see directly that each ϕ_t is smooth, because the distribution is left invariant. But the flow is not horizontal, because the distribution is not right invariant. The second, particular remark: any flow which satisfies the hypothesis of proposition 5.8 corresponds to a Hamiltonian flow with null Hamiltonian function, hence the flow is constant.

5.2   Symplectomorphisms, capacities and Hofer distance

Symplectic capacities are invariants under the action of the symplectomorphisms group. Hofer geometry is the geometry of the group of Hamiltonian diffeomorphisms, with respect to the Hofer distance. For an introduction into the subject see Hofer, Zehnder [17] chapters 2,3 and 5, and Polterovich [28], chapters 1,2.

A symplectic capacity is a map which associates to any symplectic manifold (M, ω) a number $c(M, \omega) \in [0, +\infty]$. Symplectic capacities are special cases of conformal symplectic invariants, described by:

A1. Monotonicity: $c(M, \omega) \leq c(N, \tau)$ if there is a symplectic embedding from M to N,

A2. Conformality: $c(M, \varepsilon\omega) =\mid \varepsilon \mid c(M, \omega)$ for any $\alpha \in R$, $\alpha \neq 0$.

We can see a conformal symplectic invariant from another point of view. Take a symplectic manifold (M, ω) and consider the invariant defined over the class of Borel sets $B(M)$, (seen as embedded submanifolds). In the particular case of R^{2n} with the standard symplectic form, an invariant is a function $c : B(R^{2n}) \rightarrow [0, +\infty]$ such that:

B1. Monotonicity: $c(M) \leq c(N)$ if there is a symplectomorphism ϕ such that $\phi(M) \subset N$,

B2. Conformality: $c(\varepsilon M) = \varepsilon^2 c(M)$ for any $\varepsilon \in R$.

An invariant is nontrivial if it takes finite values on sets with infinite volume, like cylinders:

$$Z(R) = \left\{ x \in R^{2n} \,:\, x_1^2 + x_2^2 < R \right\}$$

There exist highly nontrivial invariants, as the following theorem shows:

Theorem 5.9 *(Gromov's squeezing theorem) The ball $B(r)$ can be symplectically embedded in the cylinder $Z(R)$ if and only if $r \leq R$.*

This theorem permits to define the invariant:

$$c(A) \ = \ \sup \left\{ R^2 : \exists \phi(B(R)) \subset A \right\}$$

called Gromov's capacity.

Another important invariant is Hofer-Zehnder capacity. In order to introduce this we need the notion of a Hamiltonian flow.

A flow of symplectomorphisms $t \mapsto \phi_t$ is Hamiltonian if there is a function $H : M \times R \to R$ such that for any time t and place x we have

$$\omega(\dot{\phi}_t(x), v) \ = \ dH(\phi_t(x), t)v$$

for any $v \in T_{\phi_t(x)} M$.

Let $H(R^{2n})$ be the set of compactly supported Hamiltonians. Given a set $A \subset R^{2n}$, the class of admissible Hamiltonians is $H(A)$, made by all compactly supported maps in A such that the generated Hamiltonian flow does not have closed orbits of periods smaller than 1. Then the Hofer-Zehnder capacity is defined by:

$$hz(A) \ = \ \sup \left\{ \|H\|_\infty : H \in H(A) \right\}$$

Let us denote by $Ham(A)$ the class of Hamiltonian diffeomorphisms compactly supported in A. A Hamiltonian diffeomorphism is the time one value of a Hamiltonian flow. In the case which interest us, that is R^{2n}, $Ham(A)$ is the connected component of the identity in the group of compactly supported symplectomorphisms.

A curve of Hamiltonian diffeomorphisms (with compact support) is a Hamiltonian flow. For any such curve $t \mapsto c(t)$ we shall denote by $t \mapsto H_c(t, \cdot)$ the associated Hamiltonian function (with compact support).

On the group of Hamiltonian diffeomorphisms there is a bi-invariant distance introduced by Hofer. This is given by the expression:

$$d_H(\phi, \psi) \ = \ \inf \left\{ \int_0^1 \|H_c(t)\|_{\infty, R^{2n}} \, \mathrm{d}t : c : [0, 1] \to Ham(R^{2n}) \right\} \qquad (5.2.16)$$

It is easy to check that d is indeed bi-invariant and it satisfies the triangle property. It is a deep result that d is non-degenerate, that is $d(id, \phi) \ = \ 0$ implies $\phi = \ id$.

With the help of the Hofer distance one can define another symplectic invariant, called displacement energy. For a set $A \subset R^{2n}$ the displacement energy is:

$$de(A) \ = \ \inf \left\{ d_H(id, \phi) \ : \phi \in Ham(R^{2n}) \ , \ \phi(A) \cap A = \emptyset \right\}$$

A displacement energy can be associated to any group of transformations endowed with a bi-invariant distance (see Eliashberg & Polterovich [10], section 1.3). The fact that the displacement energy is a nontrivial invariant is equivalent with the non-degeneracy of the Hofer distance. Section 2. Hofer & Zehnder [17] is dedicated to the implications of the existence of a non-trivial capacity. All in all the non-degeneracy of the Hofer distance (proved for example in the Section 5. Hofer & Zehnder [17]) is the cornerstone of symplectic rigidity theory.

5.3   Hausdorff measure 2 and Hofer distance

We shall give in this section a metric proof of the non-degeneracy of the Hofer distance.

A flow of symplectomorphisms $t \mapsto \phi_t$ with compact support in A is Hamiltonian if it is the projection onto R^{2n} of a flow in $Hom(H(n), vol, Lip)(A)$.

Consider such a flow which joins identity id with ϕ. Take the lift of the flow $t \in [0,1] \mapsto \tilde{\phi}_t \in Hom(H(n), vol, Lip)(A)$.

Proposition 5.10 *The curve* $t \mapsto \tilde{\phi}_t(x,0)$ *has Hausdorff dimension* 2 *and measure (up to a multiplicative constant)*

$$\mathcal{H}^2\left(t \mapsto \tilde{\phi}_t(x,0)\right) \;=\; \int_0^1 \mid H_t(\phi_t(x)) \mid \; dt \qquad\qquad (5.3.17)$$

Proof.   The curve is not horizontal. The tangent has a vertical part (equal to the Hamiltonian). Therefore the curve has Hausdorff measure greater than one. We shall compute its Hausdorff measure 2.

By definition this is:

$$\mathcal{H}^2\left(t \mapsto \tilde{\phi}_t(x,0)\right) \;=\; \sup\left\{Q_\delta : \delta > 0\right\}$$

$$Q_\delta \;=\; \inf\left\{\sum_{i=1}^n d^2(\tilde{\phi}_{t_i}(x,0), \tilde{\phi}_{t_{i+1}}(x,0)) : 0 = t_1 < ... < t_N = 1 \;,\; \forall i \; d(\tilde{\phi}_{t_i}(x,0), \tilde{\phi}_{t_{i+1}}(x,0)) < \delta\right\}$$

The Ball-Box theorem 2.9 and the (classical) derivability of the curve ensures us that approximatively

$$d^2(\tilde{\phi}_{t_i}(x,0), \tilde{\phi}_{t_{i+1}}(x,0)) \approx \mid H(\phi_{t_i}(x)) \mid (t_{i+1} - t_i) + \mid \dot{\phi}_{t_i}(x) \mid^2 (t_{i+1} - t_i)^2$$

This implies (almost) the desired inequality (5.3.17). If we would use instead of the CC distance another distance which is more computable, for example the distance induced by the homogeneous norm

$$\|(x, \bar{x})\|^2 \;=\; \mid x \mid^2 + \frac{1}{4} \mid \bar{x} \mid$$

then we would have exactly the relation (5.3.17), modulus a multiplicative constant $C = \frac{1}{4}$, namely:

$$\mathcal{H}^2\left(t \mapsto \tilde{\phi}_t(x,0)\right) \;=\; C \int_0^1 \mid H_t(\phi_t(x)) \mid \; dt \qquad\qquad (5.3.18)$$

The Hausdorff measure correspond to the chosen homogeneous norm.

In our case, that is for the Hausdorff measure which comes from the CC distance, the constant C is simply

$$C \;=\; \lim_{t \to 0} \frac{d^2((0,0),(0,t))}{t}$$

∎

We shall prove now that the Hofer distance (5.2.16) is non-degenerate. For given ϕ with compact support in A and generating function F, look at the one parameter family:

$$\tilde{\phi}^a(x, \bar{x}) \;=\; (\phi(x), \bar{x} + F(x) + a)$$

The volume of the set

$$\{(x, z) : x \in A, \; z \text{ between } 0 \text{ and } F(x) + a\}$$

attains the minimum $V(\phi, A)$ for an $a_0 \in R$. Let us take an arbitrary flow $t \in [0, 1] \mapsto \tilde{\phi}_t \in Hom(H(n), vol, Lip)(A)$ such that $\tilde{\phi}_0 = id$ and $\tilde{\phi}_1 = \tilde{\phi}^a$. The family of curves $t \mapsto \tilde{\phi}_t(x, 0)$ provides a foliation of the set

$$B \;=\; \left\{ \tilde{\phi}_t(x, 0) : x \in A \right\}$$

We want to prove the following proposition:

Proposition 5.11 *There is a constant $C > 0$ such that*

$$\mathcal{H}^Q(B) \;\leq\; C \, vol(A) \int_0^1 \|H_t\|_{\infty, A} \, dt \tag{5.3.19}$$

Suppose that (5.3.19) is true. From the the definition of the Hofer distance (5.2.16) and from the obvious inequality $\mathcal{H}^Q(B) \geq V(\phi, A)$ we get

$$V(\phi, A) \;\leq\; vol(A) \, C \, d_H(id, \phi) \tag{5.3.20}$$

This proves the non-degeneracy of the Hofer distance, because if the RHS of (5.3.20) equals 0 then $V(\phi, A)$ is 0, which means that the generating function of ϕ is almost everywhere constant, therefore ϕ is the identity everywhere in R^{2n}.

Proof of proposition 5.11   We shall use proposition 5.10 and the definition of Hausdorff measures. For $\varepsilon > 0$ pick a ε-dense (in Euclidean distance) set

$$\{x_i : i \in I_\varepsilon\}$$

in A and a ε division $0 = t_0 < ... < t_{N(\varepsilon)} = 1$. It is clear that one can (almost) cover the set B with sets

$$B_{ij}(\varepsilon) \;=\; L_{\tilde{\phi}_{t_j}(x_i, 0)} \left(D\phi_{t_j}(x_i) B_{R^{2n}}^{eucl}(0, \frac{3}{4}\varepsilon) \;\times\; [-\varepsilon, \varepsilon] \right)$$

Denote by $D(\varepsilon)$ this cover and define a $D(\varepsilon)$ nested cover to be an arbitrary (almost) cover

$$\mathcal{D}(\varepsilon) \left\{ B^{CC}(\tilde{x}_\alpha, r_\alpha) : \alpha \in K_\varepsilon \right\}$$

of the set B with CC balls such that any ball $B^{CC}(\tilde{x}_\alpha, r_\alpha)$ from the nested cover $\mathcal{D}(\varepsilon)$ is included in exactly one set B_{ij} and moreover the center $\tilde{x}_\alpha$ of the ball has the form

$$L_{\tilde{\phi}_{t_j}(x_i, 0)}(0, z_\alpha)$$

We shall denote the fact that $\mathcal{D}(\varepsilon)$ is $D(\varepsilon)$ nested by $\mathcal{D}(\varepsilon) < D(\varepsilon)$. Let $P(\varepsilon)$ be the number:

$$P(\varepsilon) \ = \ \inf\left\{\sum r_\alpha^Q : \mathcal{D}(\varepsilon) < D(\varepsilon)\right\}$$

and let

$$S \ = \ \limsup_{\varepsilon \to 0} P(\varepsilon)$$

Use proposition 5.10 to estimate S, by first "counting" the balls in $\mathcal{D}(\varepsilon) < D(\varepsilon)$ which cover the fiber

$$\left\{\tilde{\phi}_t(x_i, 0) : t \in [0, 1]\right\}$$

and then "counting" by the elements x_i of the ε net of A. We get the upper bound

$$C \ vol(A) \ \int_0^1 \|H_t\|_{\infty, A} \ \mathrm{d}t$$

On the other part, not any cover of B with CC balls is nested, therefore from the definition of the Hausdorff measure $\mathcal{H}^Q$ we get

$$S \geq \mathcal{H}^Q(B)$$

These two inequalities provide the desired result. ∎

We close with the translation of the inequality (5.3.20) in symplectic terms.

Proposition 5.12 *Let ϕ be a Hamiltonian diffeomorphism with compact support in A and F its generating function, that is $dF \ = \ \phi^*\lambda - \lambda$, where $d\lambda \ = \ \omega$. Consider a Hamiltonian flow $t \mapsto H_t$, with compact support in A, such that the time one map equals ϕ. Then the following inequality holds:*

$$C \ \inf\left\{\int_A | F(x) - c | \ dx : c \in R\right\} \ \leq \ vol(A) \int_0^1 \|H_t\|_{\infty, A} \ dt$$

with $C > 0$ an universal constant.

5.4   Submanifolds in the Heisenberg group

We can imagine that a submanifold in the Heisenberg group should be a set which is locally connected by derivable curves. More than this, suppose that we have a coordinate system made of derivable curves. This means that a submanifold is locally a set:

$$\tilde{x} \ = \ \tilde{\phi}(t_1, ..., t_k)$$

such that all curves

$$\tau \mapsto \tilde{f}_j(\tau) = \tilde{\phi}(...t_j + \tau...)$$

(and all other components fixed) are derivable. Recall that a curve $\tilde{f}$ is derivable if it is classically derivable and moreover the quantity

$$\dot{\tilde{f}}(t) - \frac{1}{2}\omega(f(t), \dot{f}(t))$$

is constant along the curve. This quantity has an interpretation: it is (up to a constant) the constant density of the Hausdorff 2 measure of the curve.

We can imagine therefore (at least) three types of submanifolds. The type I and II submanifolds are (regular versions of) rectifiable sets:

I in the sense of Franchi, Serapioni, Serra Cassano [12], as sets which locally are level sets of Lipschitz maps from $H(n)$ to R,

II in the sense of Pauls [27], as sets which locally are images of Lipschitz maps from a nilpotent group to $H(n)$.

For the notion of rectifiable sets see Garofalo, Nhieu [13].

We have to comment on type II submanifolds. There is, up to now, no satisfactory theory of currents in Carnot groups. The Ambrosio, Kirchheim [1] theory of currents in metric spaces does not apply for Carnot groups. It is the belief of the author that a meaningful theory of currents in Carnot groups can be obtained as a theory of representations of the group of "smooth" homeomorphisms of the Carnot group. (This proposition will be the subject of a forthcoming paper; the axioms of this theory, in the case of the Heisenberg group, have been presented in a talk at the Mathematical Institute, University of Bern, in 2002. I use the opportunity to thank Martin Reimann for the kind invitation). It seems reasonable to think that:

- in such a theory we should have, instead of a gradation in the space of currents, a more complex relation connected to the ways in in which one can split the Carnot group in smaller Carnot groups, and

- that submanifolds are (associated to) a kind of currents, even in the setting of Carnot groups.

The Ambrosio-Kirchheim theory can be described as a theory associated to R^n, which is a Carnot group which split successively in one dimensional subgroups. Hence the usual gradation of currents which lead to the usual classification of manifolds according to their dimension.

In general Carnot group one should consider also the splitting given by the well-known Kirillov lemma (see, for example [8] lemma 1.1.12):

Lemma 5.13 *Let $\mathfrak{g}$ be a noncommutative nilpotent Lie algebra with one dimensional center $Z(\mathfrak{g}) = RZ$. Then $\mathfrak{g}$ can written as*

$$\mathfrak{g} \;=\; RZ \oplus RY \oplus RX \oplus \mathfrak{w} \;=\; RX \oplus \mathfrak{g}_0$$

where $[X, Y] = Z$ and $\mathfrak{g}_0$ is the centralizer of Y and an ideal.

One can split a Carnot group N in two ways: if $Z(N)$ has dimension greater than 1 then it might be possible to split N using an ideal in $Z(N)$, or, if $Z(N)$ is one dimensional, we can use Kirillov lemma.

Now I am getting to the point: type II submanifolds in the group N should then be associated to factorisations by ideals of N which appear in a splitting of N. For example, in the Heisenberg group one can have as type II submanifolds Heisenberg submanifolds or M submanifolds, where $M \subset R^{2n}$ is a Lagrangian subspace (the symplectic form is identically 0 on M).

We shall introduce here yet another type of submanifolds.

Definition 5.14 *A type III k-submanifold in the Heisenberg group is a set $M \subset H(n)$ such that for any $\tilde{x} \in M$ there is an open neighbourhood $U_{\tilde{x}} \cap M$ which has the form*

$$U_{\tilde{x}} \cap M \ = \ \left\{ \tilde{\phi}_t(\tilde{x}) \, : \, t \in T^k \right\} \cap M$$

where T^k is the k real torus and $t \in T^k \mapsto \tilde{\phi}_t$ is the lift (up to multiplication by a constant vertical element of $H(n)$) of a Hamiltonian action of T^k on R^{2n}.

Let $\tilde{x} = (x, \bar{x}) \in M$ and U_x be as in the previous definition. Let also $V_x \subset R^{2n}$ be the projection of U_x on R^{2n}. The moment map of the torus action which defines the manifold around $\tilde{x}$ is denoted by $J : V_x \to R^k$. The coordinate curve

$$t_j \mapsto \tilde{f}_j(t_j) = \tilde{\phi}_{(...t_j...)}(x)$$

(all other components are 0) is then derivable and the derivative of this curve in $t_j = 0$ is

$$(\frac{\partial \phi}{\partial t_j}(0), J_j(x), 0)$$

Therefore a submanifold in the Heisenberg group can be seen as a lift of a symplectic sub-manifold in R^{2n} endowed with a Hamiltonian torus action. The moment map of the action is visible in the tangent space of the lift.

Definition 5.15 *The tangent space at M in $\tilde{x} \in M$ is the class of all $v \in T_0 H(n)$ which are tangent to derivable curves in M, at $\tilde{x}$. This space has the description:*

$$T_{\tilde{x}} M \ = \ span \left\{ (\frac{\partial \phi}{\partial t_j}(0), J_j(x), 0) \, : \, j = 1, ..., k \right\}$$

A general submanifold should be a combination of these types. The difficult task is to classify these combinations, maybe in the same way irreducible representations of nilpotent Lie groups are classified.

6   Calculus on general groups

In this section we come back to a general group G and we try to find the good notion of derivative. The problem that we have is that the Pansu derivative of a function transforms weak classes of two letter words into similar weak classes. We defined the

word tangent bundle to be a set of strong classes. In the case of Carnot groups we had the possibility to choose polynomial representantives of strong classes, one in each weak class. We cannot do the same in the general case.

Until now we have a Pansu type derivative, which is good for Lipschitz functions. We also remarked that left, right translations preserve the word tangent bundle, but right translations are not generally Pansu derivable. We need to find a notion of derivative in between these situations.

6.1 Mild equivalence

Our starting point is the proposition 3.20, more specifically the following useful fact: strong classes act at left on weak classes. This means the function

$$([c]_s, c']_w \;\; \mapsto \;\; [c]_s * [c']_w = [c \circ c']_w$$

is well defined and it is an action:

$$[c^1]_s * \big([c^2]_s * [c]_w\big) \;\; = \;\; \big([c^1]_s \circ [c^2]_s\big) * [c]_w$$

Instead of working with strong classes (elements of the word tangent bundle), we shall work with the action of these on weak classes. This is the same idea as in the case of the virtual derivative. If we were in an Euclidean situation, these choices and differences are not at all significant. Here they are.

We introduced weak and strong equivalence relations. Now we want to identify strong classes which acts the same on weak classes of finite curves c_q, $q \in C_0^m([0,+\infty), G)$, where

$$C_0^m([0,+\infty), G) \;\; = \;\; \{s \in C^m([0,+\infty), G) : s(0) = 0\}$$

This is the same as introducing yet another equivalence relation on finite curves, which is between the weak and strong relations. We shall call it "mild".

Definition 6.1 *Two finite curves c_f, c_p, $f, p \in C^m([0,+\infty), G)$, are mildly equivalent $c_f \stackrel{m}{\approx} c_p$ if for any $q \in C_0^m([0,+\infty), G)$, we have $[c_f]_s * [c_q]_w = [c_p]_s * [c_q]_w$.*

This equivalence relation can be written as $c_f \circ c_q \stackrel{w}{\approx} c_p \circ c_q$, for any $q \in C_0^m([0,+\infty), G)$.

Otherwise said, two elements $P, R \in \mathfrak{g}^{(m)}[\varepsilon]$, are mildly equivalent if and only if for any $Q \in Weak(\mathfrak{g})$, $Q(0) = 0$ we have

$$\frac{1}{\varepsilon} d(P(\varepsilon)Q(\varepsilon), R(\varepsilon)Q(\varepsilon)) \to 0$$

as $\varepsilon \to 0$.

Here comes the second way to define a nice tangent bundle.

Definition 6.2 *The mild tangent bundle MG is the group of mild equivalence classes of words, with the bundle structure given by the projection on the first letter.*

To understand the structure of this bundle we introduce

$$MC(G) \;=\; \{c_f \,:\, f \in C^m([0,+\infty),G)\} \,/ \stackrel{m}{\approx}$$

As previously, $MC(G)$ is a Lie group, with the algebra identified with

$$(Weak(\mathfrak{g}) + Strong(\mathfrak{g})) \,/ M(\mathfrak{g})$$

where $M(\mathfrak{g})$ is the subalgebra of $Strong(\mathfrak{g})$ of elements mildly equivalent with 0.

Remark 6.3 *This algebra is nilpotent (see remarks about the spectrum), like $O(\mathfrak{g})$, but it is not an ideal. Therefore the operation on $MC(G)$ cannot be read directly from the factorisation. The reason for this is that* **mild equivalence is not invariant under conjugation**.

The fiber of the neutral element in $MC(G)$ will be denoted by $MC_0(G)$. Similarly, M_0G denotes the fiber of the neutral element in MG.

We can take only the factorisation

$$MC^0(G) \;=\; \{c_f \,:\, f \in C_0^m([0,+\infty),G)\} \,/ \stackrel{m}{\approx}$$

We shall call this the true tangent space of G at the neutral element. This is also a group and the Lie algebra structure of it is obtained by a factorisation. Indeed, consider the class

$$\mathfrak{g}_0^{(m)}[\varepsilon] \;=\; \left\{ P \in \mathfrak{g}^{(m)} \,:\, P(0) = 0 \right\}$$

Then $MC(G) \cap \mathfrak{g}_0^{(m)}[\varepsilon]$ is an ideal in $\mathfrak{g}_0^{(m)}[\varepsilon]$ and $MC^0(G)$ is a Lie group with Lie algebra obtained by factorisation.

We arrive now to the definition of the semi-derivative.

Definition 6.4 *The function f is semi-derivable at x if there is a function*

$$Df(x) : \mathfrak{g}_0^{(m)}[\varepsilon] \to Weak(\mathfrak{g})$$

such that

$$\frac{1}{\varepsilon} d(f(xP(\varepsilon)), f(x)Df(x)(P)(\varepsilon)) \to 0$$

as $\varepsilon \to 0$, uniformly with respect to P.

The notion of weak derivative follows.

Definition 6.5 *The function $f : G \to G'$ is weakly derivable if there is a function $Df(x) : \mathfrak{g}_0^{(m)}[\varepsilon] \to M^0G'$ such that for any $Q \in Weak(\mathfrak{g}')$, $Q(0) = 0$, we have*

$$\frac{1}{\varepsilon} d(f(xP(\varepsilon))Q(\varepsilon), Df(x)(P)(\varepsilon)Q(\varepsilon)) \to 0$$

uniformly with respect to P.

All these algebraic construction seem complicated. There are a lot of differences (like the one between $MC_0(G)$ and $MC^0(G)$) which are unseen in the Euclidean case, and even in the nilpotent case. That is why in the next section we shall take an example and we shall compute everything.

We close this section with the Rademacher theorem. This will be the identical with the theorem 4.10, but from slightly different reasons. Namely, instead of proposition 4.8 we have:

Proposition 6.6 *If f satisfies the condition (A) and it is weakly derivable then $Df(x)$ is a group morphism. Moreover, for any $P, R \in \mathfrak{g}_0^{(m)}[\varepsilon]$, if $P \overset{m}{\approx} R$ then $Df(x)(P) = Df(x)(Q)$.*

Proof.   The proof is the same as the one of proposition 4.8, with a single difference. The reason for which the quantity $D(\varepsilon)$, defined by the relation (4.0.4), converges to 0 as $\varepsilon \to 0$ is the derivability of f.   ∎

This suggest the good definition of derivative.

Definition 6.7 *The function $f : G \to G'$ is derivable if there is a group morphism $Df(x) : MC^0(G) \to MC^0(G')$ such that for any $P \in MC^0(G)$ we have*

$$\frac{1}{\varepsilon} d(f(xP(\varepsilon))Q(\varepsilon), Df(x)(P)(\varepsilon)Q(\varepsilon)) \to 0$$

uniformly with respect to $Q \in Weak(\mathfrak{g}')$, $Q(0) = 0$.

The Rademacher theorem in this context is simply this.

Theorem 6.8 *Let $f : G \to H$ be a Lipschitz function which satisfies condition (A). Then f is a.e. derivable in the sense of definition 6.7.*

If $f : G \to G'$ is derivable then Df is not a function from $MC(G)$ to $MC(G')$. This poses a problem concerning upper order derivatives.

The chain rule for this type of derivatives holds. Moreover, if f preserves the word tangent bundles and g is just derivable then $f \circ g$ is derivable. We leave to the reader to check this fact.

All in all, we have obtained reasonable notions for calculus in general groups. But the real hard work of understanding this calculus is still ahead. Several phenomena which are non-Euclidean have been discovered. We have to learn how to control better these new behaviours and this can be done only after deep study of many particular cases, in order to make the difference between important and marginal features of non-Euclidean analysis. Only after this we are free to forget this intricate mechanism and use it with same ease as we do with the classical calculus.

6.2   An example: $SO(3)$

In this section we shall look to the group $G = SO(3)$ of orthogonal matrices in R^3. The Lie algebra $\mathfrak{so}(3)$ is spanned by the matrices

$$X_1 \;=\; e_1 \otimes e_2 - e_2 \otimes e_1 \quad X_1 \;=\; e_2 \otimes e_3 - e_3 \otimes e_2 \quad X_1 \;=\; e_1 \otimes e_3 - e_3 \otimes e_1$$

with the bracket relations

$$[X_1, X_2] \;=\; X_3 \quad [X_2, X_3] \;=\; X_1 \quad [X_3, X_1] \;=\; X_2$$

The space spanned by X_1, X_2 generate the algebra. We shall consider therefore the dilatation flow $\delta_\varepsilon : \mathfrak{so}(3) \to \mathfrak{so}(3)$

$$\delta_\varepsilon X_1 \;=\; \varepsilon X_1 \quad \delta_\varepsilon X_2 \;=\; \varepsilon X_2 \quad \delta_\varepsilon X_3 \;=\; \varepsilon^2 X_3$$

Then $\mathfrak{so}(3)$ decomposes as

$$\mathfrak{so}(3) \;=\; V_1 + V_2 \quad V_1 \;=\; span\{X_1, X_2\} \quad V_2 \;=\; RX_3$$

The nilpotentisation N of $(SO(3), V_1)$ is the Heisenberg group. The step m of the group is equal to 2. The algebra $\mathfrak{so}(3)^{(2)}$ is 12 dimensional and it is spanned by X_i, εX_i, $\varepsilon^2 X_i$, $i = 1, 2, 3$, with bracket relations modulo $o(\varepsilon^2)$. This is the algebra of the word tangent bundle (check that $WSO(3) = CSO(3)$). It contains as subalgebras

$$\mathfrak{so}(3) \;=\; span\{X_1, X_2, X_3\}$$

$$H(1) \;=\; N(SO(3)) \;=\; span\{\varepsilon X_1, \varepsilon X_2, \varepsilon^2 X_3\}$$

(Check that the space $O(\mathfrak{so}(3))$ is trivial.)

The tangent at 0, called $T_0 SO(3)$, is 6 dimensional, spanned by εX_i, $\varepsilon^2 X_i$, $i = 1, 2, 3$.

If we want to look to $MC^0 SO(3)$, we find that it is 4 dimensional, isomorphic to the algebra $T_0 H(1)$. This can be done by knowing that for any $q \in SO(3)$ Ad_q is an orthogonal matrix, obtained as a permutation of q. This allows us to compute $M(\mathfrak{so}(3))$ to be the group with Lie algebra spanned by $\varepsilon^2 X_1$, $\varepsilon^2 X_2$, which is abelian (hence Carnot, as forecasted).

Finally, the group $MC(SO(3))$, the mild tangent bundle, can be recovered by looking to the actions of Ad^G, Ad^N, on $Weak_0(SO(3))$, spanned by εX_i, $\varepsilon^2 X_3$, $i = 1, 2, 3$. Attention, this action is done by applying Ad^G or Ad^N and then taking projection on $Weak_0(SO(3))$. We notice that $\varepsilon^2 X_3$ is a fixed point of these actions and that the group generated by Ad^G, Ad^N is simply $SL(3, R)$, which is 8 dimensional. We infer that the mild tangent bundle is $SL(3, R)$.

References

[1] L. Ambrosio; B. Kirchheim, Currents in metric spaces, *Acta Math.*, **185** (2000), no. 1, 1–80

[2] L. Ambrosio, S. Rigot, Optimal Mass Transportation in the Heisenberg Group, to appear in *J. of Funct. Analysis*

[3] P. E. Barbano, Automorphisms and quasi-conformal mappings of Heisenberg-type groups, *J. Lie Theory*, **8** (1998), no. 2, 255–277

[4] A. Bellaïche, The tangent space in sub-Riemannian geometry, in: *Sub-Riemannian Geometry*, A. Bellaïche, J.-J. Risler eds., Progress in Mathematics, **144**, Birchuser, (1996), 4 – 78

[5] M. Buliga, Sub-Riemannian geometry and Lie groups. Part I, arXiv:math.MG/0210189, (2002)

[6] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, *Graduate Studies in Mathematics*, **33**, AMS Providence, Rhode Island, (2000)

[7] J. Cheeger, Differentiability of Lipschitz functions on metric measure spaces,*Geom. Funct. Anal.*,**9** (1999), no. 3, 428–517

[8] L. Corwin, F.P. Greenleaf, Representations of nilpotent Lie groups and their applications. Part 1: Basic theory and examples, Cambridge Univ. Press, 1990

[9] M. Cowling, A Dooley, A. Koranyi, F. Ricci, H-type groups and Iwasawa decompositions, *Adv. in Math.*, **87** (1991), 1–41

[10] Y. Eliashberg, L. Polterovich, Biinvariant metrics on the group of Hamiltonian diffeomorphisms, *International J. of Math.*, **4**, no. 5, (1993), 727–738

[11] G.B. Folland, E.M. Stein, Hardy spaces on homogeneous groups, *Mathematical notes*, **28**, Princeton - N.J.: Princeton University Press; Tokyo: University of Tokyo Press, (1982)

[12] B. Franchi, R. Serapioni, F. Serra Cassano, Rectifiability and Perimeter in the Heisenberg Group, *Math. Ann.*, **321**, (2001), 479–531

[13] N. Garofalo, D.M. Nhieu, Isoperimetric and Sobolev inequalitites for Carnot-Carathéodory spaces and the existence of minimal surfaces, *Comm. Pure Appl. Math.*, **49** (1996), no. 10, 1081–1144

[14] R. Goodman, Nilpotent Lie groups: structure and applications to analysis, *Lecture Notes in Math.* , **53** (1981)

[15] M. Gromov, Carnot-Caratheodory spaces seen from within, in: *Sub-Riemannian Geometry*, A. Bellaïche, J.-J. Risler eds., Progress in Mathematics, **144**, Birchuser, (1996), 79 – 323

[16] M. Gromov, Metric structures for Riemannian and non-Riemannian spaces, *Progress in Math.*, **152**, Birchäuser (1999), chapter 1.

[17] H. Hofer, E. Zehnder, Symplectic invariants and Hamiltonian dynamics, *Birkhuser Advanced Texts: Basler Lehrbcher*, Birkhuser Verlag, Basel, (1994)

[18] I.M. James, Introduction to uniform spaces, *LMS Lecture Notes Series*, **144**, Cambridge University Press, 1990

[19] A. Koranyi, M. Reimann, Foundations for the theory of quasi-conformal mappings of the Heisenberg group, *Adv. in Math.*, **III:1** (1995), 1–87

[20] B. Kostant, On convexity, the Weyl group and the Iwasawa decomposition, *Ann. Scient. Ec. Norm. Sup.*, 4e série, **t. 6** (1973), 413–455

[21] V. Magnani, Differentiability and area formula on stratified Lie groups, preprint http:
www.cvgmt.it, (2000)

[22] G.A. Margulis, G.D. Mostow, The differential of a quasi-conformal mapping of a Carnot-Carathéodory space, *Geom. Funct. Analysis*, **8** (1995), 2, 402–433

[23] G.A: Margulis, G.D. Mostow, Some remarks on the definition of tangent cones in a Carnot-Carathéodory space, *J. D'Analyse Math.*, **80** (2000), 299–317

[24] J. Mitchell, On Carnot-Carathéodory metrics, *J. Diff. Geometry*, **21** (1985), 35–45

[25] P. Pansu, Métriques de Carnot-Carathéodory at quasiisometries des espaces symetriques de rang un, *Ann. of Math.*, **129** (1989), 1–60

[26] P. Pansu, Croissance des boules et des géodésiques fermées dans les nilvariétés, *Ergod. Th. & Dynam. Sys.*, **3** (1983), 415–445

[27] S. Pauls, A Notion of Rectifiability Modeled on Carnot Groups, Submitted, available at //hilbert.dartmouth.edu/~pauls/preprints.html

[28] L. Polterovich, The geometry of the group of symplectic diffeomorphisms, *Lectures in Mathematics ETH Zrich*, Birkhuser Verlag, Basel, (2001)

[29] H. M. Reimann, Rigidity of H-type groups, *Math. Z.*, **237** (2001), no. 4, 697–725

[30] S. K. Vodop'yanov, $\mathcal{P}$-Differentiability on Carnot Groups in Different Topologies and Related Topics, *Proc. on Analysis ans Geometry*, Novosibirsk: Sobolev Institute Press, (2000), 603 – 670

[31] S. K. Vodop'yanov, On differentiability of mappings of Sobolev classes on Carnot groups, *Matematicheskiy Sbornik* **194** (2003), 6, 67–86.

[32] S. K. Vodop'yanov, A. V. Greshnov, On differentiability of mappings of Carnot-Carathéodory spaces, *Dokl. R.A.S.*, t. **389**, 5, (2003)

[33] S. K. Vodop'yanov, A. V. Greshnov, Geometry of Carnot-Carathéodory spaces and differentiability of weakly contact mappings, *Izvestiya R.A.S., Matematicheskaya seriya*, to appear

[34] S. K. Vodop'yanov, A. D. Ukhlov, Approximately differentiable transformations and change of variables on nilpotent groups, *Sib. Math. J.*, **37** (1996), 1, 62–78