Control of Nonholonomic Systems and Sub-Riemannian Geometry

Frédéric Jean

Control of Nonholonomic Systems and Sub-Riemannian Geometry

Frédéric Jean*†

Lectures given at the CIMPA School "Géométrie sous-riemannienne", Beirut, Lebanon, 2012

Contents

*ENSTA ParisTech, UMA, 32, bd Victor, 75015 Paris, France and Team GECO, INRIA Saclay – Île-de-France, frederic.jean@ensta-paristech.fr

†This work was supported by the ANR project *GCM*, program "Blanche", project number NT09_504490, and by the Commission of the European Communities under the 7th Framework Programme Marie Curie Initial Training Network (FP7-PEOPLE-2010-ITN), project SADCO, contract number 264735.

Nonholonomic systems are control systems which depend linearly on the control. Their underlying geometry is the sub-Riemannian geometry, which plays for these systems the same role as Euclidean geometry does for linear systems. In particular the usual notions of approximations at the first order, that are essential for control purposes, have to be defined in terms of this geometry. The aim of these notes is to present these notions of approximation and their link with the metric tangent structure in sub-Riemannian geometry.

The notes are organized as follows. In Section 1 we introduce the basic definitions on nonholonomic systems and sub-Riemannian geometry. Section 2 is devoted to the study of the controllability of nonholonomic systems, and to the topological properties of sub-Riemannian distances. Section 3 provides a detailed exposition of the notions of first-order approximation, including nonholonomic orders, privileged coordinates, nilpotent approximations, and distance estimates such as the Ball-Box Theorem. We then see in Section 4 how these notions allow us to describe the tangent structure to a Carnot-Carathéodory space (the metric space defined by a sub-Riemannian distance). Finally, we present in the appendix some results on flows in connection with the Hausdorff formula (Section A), and some proofs on privileged coordinates (Section B).

1 Geometry of nonholonomic systems

Throughout these notes we work in a smooth n-dimensional manifold M. However most of our considerations are local, so M can also be thought of as an open subset of $\mathbb{R}^n$.

1.1 Nonholonomic systems

A *nonholonomic system* on M is a control system which is of the form

$$\dot{q} = \sum_{i=1}^{m} u_i X_i(q),\ q \in M, \quad u = (u_1, \dots, u_m) \in \mathbb{R}^m, \tag{Σ}$$

where $X_1, \dots, X_m$ are C^∞ vector fields on M. To give a meaning to such a control system, we have to define what are its solutions, that is, its trajectories.

Definition 1.1. A *trajectory of* (Σ) is a path $\gamma : [0, T] \to M$ for which there exists a function $u(\cdot) \in L^1([0, T], \mathbb{R}^m)$ such that γ is a solution of the

ordinary differential equation:

$$\dot{q}(t) = \sum_{i=1}^{m} u_i(t) X_i(q(t)), \quad \text{for a.e. } t \in [0, T].$$

Such a function $u(\cdot)$ is called a *control* associated with γ.

Equivalently, a trajectory is an absolutely continuous path γ on M such that $\dot{\gamma}(t) \in \Delta(\gamma(t))$ for almost every $t \in [0, T]$, where we have set, for every $q \in M$,

$$\Delta(q) = \text{span}\,\{X_1(q), \ldots, X_m(q)\} \subset T_q M. \tag{1}$$

Note that the rank of the vector spaces $\Delta(q)$ is a function of q, which may be non constant. If it is constant, Δ defines a distribution on M, that is, a subbundle of TM.

Example 1.1 (unicycle). The most typical example of nonholonomic system is the simplified kinematic model of a unicycle. In this model, a configuration $q = (x, y, \theta)$ of the unicycle is described by the planar coordinates (x, y) of the contact point of the wheel with the ground, and by the angle θ of orientation of the wheel with respect to the x-axis. The space of configurations is then the manifold $\mathbb{R}^2 \times S^1$.

The wheel is subject to the constraint of rolling without slipping, which writes as $\dot{x} \sin\theta - \dot{y} \cos\theta = 0$, or, equivalently as $\dot{q} \in \ker \omega(q)$, where ω is the one-form $\sin\theta dx - \cos\theta dy$. Hence the set Δ of (1) is $\ker \omega$.

Choosing as controls the tangential velocity u_1 and the angular velocity u_2, we obtain the nonholonomic system $\dot{q} = u_1 X_1(q) + u_2 X_2(q)$ on $\mathbb{R}^2 \times S^1$, where $X_1 = \cos\theta \partial_x + \sin\theta \partial_y$, and $X_2 = \partial_\theta$.

Let us mention here a few properties of the trajectories of (Σ) (for more details, see [Rif]).

- Fix $p \in M$ and $T > 0$. For every control $u(\cdot) \in L^1([0, T], \mathbb{R}^m)$, there exists $\tau \in (0, T]$ such that the Cauchy problem

$$\begin{cases} \dot{q}(t) = \sum_{i=1}^{m} u_i(t) X_i(q(t)) & \text{for a.e. } t \in [0, \tau], \\ q(0) = p, \end{cases} \tag{2}$$

  has a unique solution denoted by γ_u or $\gamma(\cdot; p, u)$. It is called the *trajectory issued from p associated with u.*

- If the rank of $X_1, \ldots, X_m$ is constant and equal to m on M, every trajectory is associated with a unique control. Otherwise different controls can be associated with the same trajectory. In this case it will sometimes be useful to consider among these controls only the ones which minimize the L^1 norm $\int \|u(t)\| dt$. By convexity, this defines a unique control with which the trajectory is associated.

- Any time-reparameterization of a trajectory is still a trajectory: if $\gamma : [0,T] \to M$ is a trajectory associated with a control u, and $\alpha : [0,S] \to [0,T]$ is a C^1-diffeomorphism, then $\gamma \circ \alpha : [0,S] \to M$ is a trajectory associated with the control $\alpha'(s)u(\alpha(s))$. In particular, one can reverse time along γ: the resulting path $\gamma(T-s)$, $s \in [0,T]$, is a trajectory associated with the control $-u(T-s)$.

In this context, the first question is the one of the controllability: can we join any two points by a trajectory? This suggests to introduce the following definition.

Definition 1.2. The *attainable set from* $p \in M$ is defined to be the set $\mathcal{A}_p$ of points attained by a trajectory of (Σ) issued from p.

The question above then becomes: is the attainable set from any point equal to the whole manifold M? We will answer this question in Section 2.

In the case where the answer is positive, next issues are notably the motion planning (i.e. find a trajectory joining two given points) and the stabilization (i.e. design the control as a function $u(q)$ of the state in such a way that the resulting differential equation is stable). The usual way to deal with these problems is to use a first-order approximation of the system. The underlying idea is the following. Consider a nonlinear control system in $\mathbb{R}^n$,

$$\dot{x} = f(x,u), \quad x \in \mathbb{R}^n,\ u \in \mathbb{R}^m,$$

and a pair $(\bar{x},\bar{u}) \in \mathbb{R}^{m+n}$ such that $f(\bar{x},\bar{u}) = 0$. The linearized system around this equilibrium pair is defined to be the linear control system:

$$\dot{\delta x} = \frac{\partial f}{\partial x}(\bar{x},\bar{u})\delta x + \frac{\partial f}{\partial u}(\bar{x},\bar{u})\delta u, \quad \delta x \in \mathbb{R}^n,\ \delta u \in \mathbb{R}^m.$$

If this linearized system is controllable, so is the nonlinear one near $\bar{x}$. In this case the solutions to the motion planning and stabilization problems for the linearized system may be used to construct solutions of the corresponding problems for the nonlinear system (see for instance [Kha01]). Thus, locally, the study of the control system amounts to the one of the linearized system.

Does this strategy apply to nonholonomic systems? Consider a nonholonomic system (Σ) defined on an open subset M of $\mathbb{R}^n$. For every $\bar{q} \in M$, the pair $(\bar{q},0)$ is an equilibrium pair and the corresponding linearized system is

$$\dot{\delta x} = \sum_{i=1}^{m} \delta u_i X_i(\bar{q}), \quad \delta x \in \mathbb{R}^n,\ \delta u \in \mathbb{R}^m.$$

For this linearized system, the attainable set from a point δq is obviously the affine subset

$$\delta q + \Delta(\bar{q}) = \delta q + \text{span}\left\{X_1(\bar{q}), \ldots, X_m(\bar{q})\right\}.$$

Thus, except in the very special case where $\text{rank}\Delta(\bar{q}) = n$, the linearized system is not controllable and the strategy above does not apply, whereas nonholonomic systems may be controllable (and generically they are), as we will see Section 2.

This may be explained as follows. The linearization is a first-order approximation with respect to a Euclidean (or a Riemannian) distance. However for nonholonomic systems the underlying distance is a sub-Riemannian one and it behaves very differently from a Euclidean one. Thus, the local behaviour should be understood through the study of a first-order approximation with respect to this sub-Riemannian distance, not through the linearized system.

We will introduce now the sub-Riemannian distances. In Section 3 we will see how to construct first-order approximations with respect to this kind of distances, and how to use them for motion planning for instance.

1.2 Sub-Riemannian distance

A nonholonomic system induces a distance on M in the following way. We first define the *sub-Riemannian metric* associated with (Σ) to be the function $g : TM \to \overline{\mathbb{R}}$ given by

$$g(q, v) = \inf\left\{u_1^2 + \cdots + u_m^2 \ : \ \sum_{i=1}^{m} u_i X_i(q) = v\right\}, \tag{3}$$

for $q \in M$ and $v \in T_qM$, where we adopt the convention that $\inf \emptyset = +\infty$. This function g is smooth and satisfies:

- $g(q, v) = +\infty$ if $v \notin \Delta(q)$,
- g restricted to $\Delta(q)$ is a positive definite quadratic form.

Such a metric allows to define a distance in the same way as in Riemannian geometry.

Definition 1.3. The *length* of an absolutely continuous path $\gamma(t)$, $t \in [0, T]$, is

$$\text{length}(\gamma) = \int_0^T \sqrt{g\left(\gamma(t), \dot{\gamma}(t)\right)} dt,$$

and the *sub-Riemannian distance* on M associated with the nonholonomic system (Σ) is defined by

$$d(p,q) = \inf \text{length}(\gamma),$$

where the infimum is taken over all absolutely continuous paths γ joining p to q.

Note that only trajectories of (Σ) may have a finite length. In particular, if no trajectory joins p to q, then $d(p,q) = +\infty$. We will see below in Corollary 2.4 that, under an extra assumption on the nonholonomic system, d is actually a distance function.

Remark 1.1. When γ is a trajectory, its length is also equal to

$$\min \int_0^T \|u(t)\| dt,$$

the minimum being taken over all control $u(\cdot)$ associated with γ. As already noticed, this minimum is attained at a unique control which could be defined as *the* control associated with γ.

An important feature of the length of a path is that it is independent of the parametrization of the path. As a consequence, the sub-Riemannian distance $d(p,q)$ may also be understood as the minimal time needed for the nonholonomic system to go from p to q with bounded controls, that is,

$$d(p,q) = \inf \left\{ T \geq 0 \ : \begin{array}{c} \exists \text{ a trajectory } \gamma_u : [0,T] \to M \text{ s.t.} \\ \gamma_u(0) = p,\ \gamma_u(T) = q, \\ \text{and } \|u(t)\| \leq 1 \text{ for a.e. } t \in [0,T] \end{array} \right\}. \tag{4}$$

This formulation justifies the assertion made in Section 1.1: for nonholonomic systems, first-order approximations with respect to the time should be understood as first-order approximations with respect to the sub-Riemannian distance.

Another consequence of (4) is that $d(p,q)$ is the solution of a time-optimal control problem. It then results from standard existence theorems (see for instance [LM67] or [Rif]) that, when p and q are sufficiently close and $d(p,q) < \infty$, there exists a trajectory γ joining p to q such that

$$\text{length}(\gamma) = d(p,q).$$

Such a trajectory is called a *minimizing trajectory*.

Remark 1.2. Any reparameterization of a minimizing trajectory is also minimizing. Therefore any pair of close enough points can be joined by a minimizing trajectory *of velocity one*, that is, a trajectory γ such that $g(\gamma(t), \dot{\gamma}(t)) = 1$ for a.e. t. As a consequence, there exists a control $u(\cdot)$ associated with γ such that $\|u(t)\| = 1$ a.e. Every sub-arc of such a trajectory γ is also clearly minimizing, hence the equality $d(p, \gamma(t)) = t$ holds along γ.

1.3 Sub-Riemannian manifolds

The distance d defined in Section 1.2 does not always meet the classical notion of sub-Riemannian distance arising from a sub-Riemannian manifold. Let us recall the latter definition.

A *sub-Riemannian manifold* (M, D, g_R) is a smooth manifold M endowed with a *sub-Riemannian structure* (D, g_R), where:

- D is a distribution on M, that is a subbundle of TM;
- g_R is a Riemannian metric on D, that is a smooth function $g_R : D \to \mathbb{R}$ whose restrictions to $D(q)$ are positive definite quadratic forms.

The *sub-Riemannian metric* associated with (D, g_R) is the function $g_{SR} : TM \to \overline{\mathbb{R}}$ given by

$$g_{SR}(q, v) = \begin{cases} g_R(q, v) & \text{if } v \in D(q), \\ +\infty & \text{otherwise.} \end{cases} \tag{5}$$

The *sub-Riemannian distance* d_{SR} on M is then defined from the metric g_{SR} as d is defined from the metric g in Section 1.2.

What is the difference between the two constructions, that is, between the definitions (3) and (5) of a sub-Riemannian metric?

Consider a *sub-Riemannian structure* (D, g_R). Locally, on some open subset U, there exist vector fields $X_1, \ldots, X_m$ whose values at each point $q \in U$ form an orthonormal basis of $D(q)$ for the quadratic form g_R; the metric g_{SR} associated with (D, g_R) then coincides with the metric g associated with $X_1, \ldots, X_m$. Thus, locally, there is a one-to-one correspondence between sub-Riemannian structures and nonholonomic systems for which the rank of $\Delta(q) = \operatorname{span}\{X_1(q), \ldots, X_m(q)\}$ is constant.

However this correspondence does not hold globally since, for topological reasons, a distribution of rank m may not always be generated by m vector fields on the whole M. Conversely, the vector fields $X_1, \ldots, X_m$ of a nonholonomic system do not always generate a linear space $\Delta(q)$ of constant rank equal to m. It may even be impossible, again for topological reasons (for instance, on an even dimensional sphere).

A way to conciliate both notions is to generalize the definition of sub-Riemannian structure.

Definition 1.4. A *generalized sub-Riemannian structure* on M is a triple (E, σ, g_R) where

- E is a vector bundle over M;
- $\sigma : E \to TM$ is a morphism of vector bundles;
- g_R is a Riemannian metric on E.

With a generalized sub-Riemannian structure a metric is associated which is defined by

$$g_{SR}(q, v) = \inf\{g(q, u) \ : \ u \in E(q),\ \sigma(u) = v\}, \quad \text{for } q \in M,\ v \in T_qM.$$

The generalized sub-Riemannian distance d_{SR} on M is then defined from this metric g_{SR} as d is defined from the metric g.

This definition of sub-Riemannian distance actually contains the two notions of distance we have introduced before.

- Take $E = M \times \mathbb{R}^m$, $\sigma : E \to TM$, $\sigma(q, u) = \sum_{i=1}^m u_i X_i(q)$ and g_R the Euclidean metric on $\mathbb{R}^m$. The resulting generalized sub-Riemannian distance is the distance associated with the nonholonomic system (Σ).
- Take $E = D$, where D is a distribution on M, $\sigma : D \hookrightarrow TM$ the inclusion, and g_R a Riemannian metric on D. We recover the distance associated with the sub-Riemannian structure (D, g_R).

Locally, a generalized sub-Riemannian structure can always be defined by a single finite family $X_1, \dots, X_m$ of vector fields, and so by a nonholonomic system (without rank condition). It actually appears that this is also true globally (see [ABB12], or [DLPR12] for the fact that a submodule of TM is finitely generated): any generalized sub-Riemannian distance may be associated with a nonholonomic system.

In these notes, we will always consider a sub-Riemannian distance d associated with a nonholonomic system. However, as we just noticed, all the results actually hold for a generalized sub-Riemannian distance.

2 Controllability

Consider a nonholonomic system

$$\dot{q} = \sum_{i=1}^{m} u_i X_i(q), \tag{Σ}$$

on a smooth n-dimensional manifold M. This section is concerned with the question of *controllability*: is the attainable set $\mathcal{A}_p$ from any point p equal to the whole manifold M? We will see next the implications on the sub-Riemannian distance d and on the topology of the metric space (M, d).

2.1 The Chow-Rashevsky Theorem

The controllability of (Σ) is mainly characterized by the properties of the Lie algebra generated by $X_1, \dots, X_m$. We first introduce notions and definitions on this subject.

Let $VF(M)$ denote the set of smooth vector fields on M. We define Δ^1 to be the linear subspace of $VF(M)$ generated by $X_1, \dots, X_m$,

$$\Delta^1 = \text{span}\{X_1, \dots, X_m\}.$$

For $s \geq 1$, define $\Delta^{s+1} = \Delta^s + [\Delta^1, \Delta^s]$, where we have set $[\Delta^1, \Delta^s] = \text{span}\{[X, Y] \,:\, X \in \Delta^1,\ Y \in \Delta^s\}$. The *Lie algebra generated by* $X_1, \dots, X_m$ is defined to be $Lie(X_1, \dots, X_m) = \bigcup_{s \geq 1} \Delta^s$. Due to the Jacobi identity, $Lie(X_1, \dots, X_m)$ is the smallest linear subspace of $VF(M)$ which both contains $X_1, \dots, X_m$ and is invariant by Lie brackets.

Let us denote by $I = i_1 \cdots i_k$ a multi-index of $\{1, \dots, m\}$, and by $|I| = k$ the length of I. We set

$$X_I = [X_{i_1}, [\dots, [X_{i_{k-1}}, X_{i_k}] \dots].$$

With these notations, $\Delta^s = \text{span}\{X_I \,:\, |I| \leq s\}$.

For $q \in M$, we set $Lie(X_1, \dots, X_m)(q) = \{X(q) : X \in Lie(X_1, \dots, X_m)\}$, and, for $s \geq 1$, $\Delta^s(q) = \{X(q) \,:\, X \in \Delta^s\}$. By definition these sets are linear subspaces of T_qM.

Definition 2.1. We say that (Σ) (or the vector fields $X_1, \dots, X_m$) satisfies *Chow's Condition* if

$$Lie(X_1, \dots, X_m)(q) = T_qM, \qquad \forall q \in M.$$

Equivalently, for any $q \in M$, there exists an integer $r = r(q)$ such that $\dim \Delta^r(q) = n$.

This property is also known as the Lie algebra rank condition (LARC), and as the Hörmander condition (in the context of PDE).

Lemma 2.1. *If (Σ) satisfies Chow's Condition, then for every $p \in M$, the set $\mathcal{A}_p$ is a neighbourhood of p.*

Proof. We work in a small neighbourhood $U \subset M$ of p that we identify with a neighbourhood of 0 in $\mathbb{R}^n$.

Let $\phi_t^i = \exp(tX_i)$ be the flow of the vector field X_i, $i = 1, \dots, m$. Every curve $t \mapsto \phi_t^i(q)$ is a trajectory of (Σ) and we have

$$\phi_t^i = \mathrm{id} + tX_i + o(t).$$

For every multi-index I of $\{1, \dots, m\}$, we define the local diffeomorphisms ϕ_t^I on U by induction on the length $|I|$ of I: if $I = iJ$, then

$$\phi_t^{iJ} = [\phi_t^i, \phi_t^J] := \phi_{-t}^J \circ \phi_{-t}^i \circ \phi_t^J \circ \phi_t^i.$$

By construction, $\phi_t^I(q)$ is the endpoint of a trajectory of (Σ) issued from q. Moreover, on a neighbourhood of p there holds

$$\phi_t^I = \mathrm{id} + t^{|I|} X_I + o(t^{|I|}). \tag{6}$$

We postpone the proof of this formula to the Appendix (Proposition A.4).

To obtain a diffeomorphism whose derivative with respect to the time is exactly X_I, we set

$$\psi_t^I = \begin{cases} \phi^I_{t^{1/|I|}} & \text{if } t \geq 0, \\ \phi^I_{-|t|^{1/|I|}} & \text{if } t < 0 \text{ and } |I| \text{ is odd}, \\ [\phi^J_{|t|^{1/|I|}}, \phi^i_{|t|^{1/|I|}}] & \text{if } t < 0 \text{ and } |I| \text{ is even}, \end{cases}$$

where $I = iJ$. Thus

$$\psi_t^I = \mathrm{id} + tX_I + o(t), \tag{7}$$

and $\psi_t^I(q)$ is the endpoint of a trajectory of (Σ) issued from q.

Let us choose now commutators $X_{I_1}, \dots, X_{I_n}$ whose values at p span T_pM. This is possible thanks to Chow's Condition. We introduce the map φ defined on a small neighbourhood Ω of 0 in $\mathbb{R}^n$ by

$$\varphi(t_1, \dots, t_n) = \psi_{t_n}^{I_n} \circ \cdots \circ \psi_{t_1}^{I_1}(p) \in M.$$

We conclude from (7) that this map is C^1 near 0 and has an invertible derivative at 0, which implies that it is a local C^1-diffeomorphism. Therefore $\varphi(\Omega)$ contains a neighbourhood of p.

Now, for every $t \in \Omega$, $\varphi(t)$ is the endpoint of a concatenation of trajectories of (Σ), the first one being issued from p. It is then the endpoint of a trajectory starting from p. Therefore $\varphi(\Omega) \subset \mathcal{A}_p$, which implies that $\mathcal{A}_p$ is a neighbourhood of p. □

Theorem 2.2 (Chow-Rashevsky's theorem)**.** *If M is connected and if (Σ) satisfies Chow's Condition, then any two points of M can be joined by a trajectory of (Σ).*

Proof. Let $p \in M$. If $q \in \mathcal{A}_p$, then $p \in \mathcal{A}_q$. As a consequence, $\mathcal{A}_p = \mathcal{A}_q$ for any $q \in M$ and the lemma above implies that $\mathcal{A}_p$ is an open set. Hence the manifold M is covered by the union of the sets $\mathcal{A}_p$ that are pairwise disjointed. Since M is connected, there is only one such open set. □

Remark 2.1. This theorem appears also as a consequence of the Orbit Theorem (Sussmann, Stefan [Ste74, Sus73]): each set $\mathcal{A}_p$ is a connected immersed submanifold of M and, at each point $q \in \mathcal{A}_p$, $Lie(X_1, \ldots, X_m)(q) \subset T_q\mathcal{A}_p$. Moreover, when the rank of the Lie algebra is constant on M, both spaces are equal, i.e. $Lie(X_1, \ldots, X_m)(q) = T_q\mathcal{A}_p$.
Thus, when the Lie algebra generated by $X_1, \ldots, X_m$ has constant rank, Chow's Condition is not restrictive: it is indeed satisfied on each $\mathcal{A}_p$ by the restriction of the vector fields $X_1, \ldots, X_m$ to the manifold $\mathcal{A}_p$.

Remark 2.2. The converse of Chow's theorem is false in general. Consider for instance the nonholonomic system in $\mathbb{R}^3$ defined by $X_1 = \partial_x$, $X_2 = \partial_y + f(x)\partial_z$ where $f(x) = e^{-1/x^2}$ for positive x and $f(x) = 0$ otherwise. The associated sub-Riemannian distance is finite whereas $X_1, \ldots, X_m$ do not satisfy Chow's Condition.
However, for an analytic nonholonomic system (i.e. when M and the vector fields $X_1, \ldots, X_m$ are in the analytic category), Chow's Condition is equivalent to the controllability of (Σ) (see [Nag66, Sus74]).

Remark 2.3. Our proof of Theorem 2.2 also shows that, under the assumptions of the theorem, for every point $p \in M$ the set

$$\left\{\exp(t_{i_1}X_{i_1}) \circ \cdots \circ \exp(t_{i_k}X_{i_k})(p) \ : \ k \in \mathbb{N},\ t_{i_j} \in \mathbb{R},\ i_j \in \{1, \ldots, m\}\right\}$$

is equal to the whole M. This set is often called the *orbit* at p of the vector fields $X_1, \ldots, X_m$.

2.2 Topological structure of (M, d)

The proof of Lemma 2.1 gives a little bit more than the openness of $\mathcal{A}_p$. For ε small enough, any $\phi_t^i(q)$, $0 \leq t \leq \varepsilon$, is a trajectory of length ε. Thus $\varphi(t_1, \ldots, t_n)$ is the endpoint of a trajectory of length less than $N\big(|t_1|^{1/|I_1|} + \cdots + |t_n|^{1/|I_n|}\big)$, where N counts the maximal number of concatenations involved in the $\psi_t^{I_i}$'s. This gives an upper bound for the distance,

$$d\big(p, \varphi(t)\big) \leq N\big(|t_1|^{1/|I_1|} + \cdots + |t_n|^{1/|I_n|}\big). \tag{8}$$

This kind of estimates of the distance in terms of local coordinates plays an important role in sub-Riemannian geometry, as we will see in Section 3.4. However here $(t_1, \ldots, t_n)$ are not smooth local coordinates, as φ is only a C^1-diffeomorphism, not a smooth diffeomorphism.

Let us try to replace $(t_1, \ldots, t_n)$ by smooth local coordinates. Choose local coordinates $(y_1, \ldots, y_n)$ centered at p such that $\frac{\partial}{\partial y_i}|_p = X_{I_i}(p)$. The map $\varphi^y = y \circ \varphi$ is a C^1-diffeomorphism between neighbourhoods of 0 in $\mathbb{R}^n$, and its differential at 0 is $d\varphi_0^y = \mathrm{Id}_{\mathbb{R}^n}$.

Denoting by $\|\cdot\|_{\mathbb{R}^n}$ the Euclidean norm on $\mathbb{R}^n$, we obtain, for $\|t\|_{\mathbb{R}^n}$ small enough, $y_i(t) = t_i + o(\|t\|_{\mathbb{R}^n})$. The inequality (8) becomes

$$d(p, q^y) \leq N' \|y\|_{\mathbb{R}^n}^{1/r},$$

where q^y denotes the point of coordinates y, and $r = \max_i |I_i|$. This inequality allows to compare d to a Riemannian distance.

Let g_R be a Riemannian metric on M, and d_R the associated Riemannian distance. On a compact neighbourhood of p, there exists a constant $c > 0$ such that $g(X_i, X_i)(q) \leq c^{-1}$, which implies $c d_R(p, q) \leq d(p, q)$. Moreover we have $d_R(p, q^y) \geq Cst \|y\|_{\mathbb{R}^n}$. We have then obtained a first estimate to the sub-Riemannian distance.

Theorem 2.3. *Assume (Σ) satisfies Chow's Condition. For any Riemannian metric g_R, we have, for q close enough to p,*

$$c d_R(p, q) \leq d(p, q) \leq C d_R(p, q)^{1/r},$$

where c, C are positive constants and r is an integer such that $\Delta_p^r = T_pM$.

Remark 2.4. If we choose for g_R a Riemannian metric which is compatible with g, that is, which satisfies $g_R|_\Delta = g$, then by construction $d_R(p, q) \leq d(p, q)$.

Corollary 2.4. *Under the hypotheses of Theorem 2.2, d is a distance function on M, i.e.,*

(i) d *is a function from* $M \times M$ *to* $[0, \infty)$*;*

(ii) $d(p, q) = d(q, p)$ *(symmetry);*

(iii) $d(p, q) = 0$ *if and only if* $p = q$*;*

(iv) $d(p, q) + d(p, q') \leq d(p, q')$ *(triangle inequality).*

Proof. By Chow-Rashevsky's theorem (Theorem 2.2), the distance between any pair of points is finite, which gives *(i)*. The symmetry of the distance results from the fact that, if $\gamma(s)$, $s \in [0, T]$, is a trajectory joining p to q, then $s \mapsto \gamma(T - s)$ is a trajectory of same length joining q to p. Point *(iii)* follows directly from Theorem 2.3. Finally, the triangle inequality is a consequence of the following remark. If $\gamma(s)$, $s \in [0, T]$, is a trajectory joining p to q and $\gamma'(s)$, $s \in [0, T']$, is a trajectory joining q to q', then the concatenation $\gamma * \gamma'$, defined by

$$\gamma * \gamma'(s) = \begin{cases} \gamma(s) & \text{if } s \in [0, T], \\ \gamma'(s - T) & \text{if } s \in [T, T + T'], \end{cases}$$

is a trajectory joining p to q' whose length satisfies

$$\text{length}(\gamma * \gamma') = \text{length}(\gamma) + \text{length}(\gamma').$$

□

A second consequence of Theorem 2.3 is that the sub-Riemannian distance d is $1/r$-Hölder with respect to any Riemannian distance, and so continuous.

Corollary 2.5. *If* (Σ) *satisfies Chow's Condition, then the topology of the metric space* (M, d) *coincides with the topology of* M *as a smooth manifold.*

3 First-order approximations

Consider a nonholonomic system (Σ): $\dot{q} = \sum_{i=1}^{m} u_i X_i(q)$ on a manifold M satisfying Chow's Condition, and denote by d the induced sub-Riemannian distance. As we have seen in Section 1.1, the infinitesimal behaviour of this system should be captured by an approximation to the first-order with respect to d. In this section we will then provide notion of first-order approximation and construct the basis of an infinitesimal calculus adapted to nonholonomic systems. To this aim, a fundamental role will be played by the concept of noholonomic order of a function at a point. We will then see that approximations to the first-order appear as nilpotent approximations, in the sense that $X_1, \ldots, X_m$ are approximated by vector fields that generate a nilpotent Lie algebra.

The whole section is concerned with local objects. Henceforth, throughout the section we fix a point $p \in M$ and an open neighbourhood U of p that we identify with a neighbourhood of 0 in $\mathbb{R}^n$ through some local coordinates.

3.1 Nonholonomic orders

Definition 3.1. Let $f : M \to \mathbb{R}$ be a continuous function. The *nonholonomic order of f at p*, denoted by $\mathrm{ord}_p(f)$, is the real number defined by

$$\mathrm{ord}_p(f) = \sup\left\{ s \in \mathbb{R} \ : \ f(q) = O\big(d(p,q)^s\big) \right\}.$$

This order is always nonnegative. Moreover $\mathrm{ord}_p(f) = 0$ if $f(p) \neq 0$, and $\mathrm{ord}_p(f) = +\infty$ if $f(p) \equiv 0$.

Example 3.1 (Euclidean case). When $M = \mathbb{R}^n$, $m = n$, and $X_i = \partial_{x_i}$, the sub-Riemannian distance is simply the Euclidean distance on $\mathbb{R}^n$. In this case, nonholonomic orders coincide with the standard ones. Namely, $\mathrm{ord}_0(f)$ is the smallest degree of monomials having nonzero coefficient in the Taylor series

$$f(x) \sim \sum c_\alpha x_1^{\alpha_1} \dots x_n^{\alpha_n}$$

of f at 0. We will see below that there exists in general an analogous characterization of nonholonomic orders.

Let $C^\infty(p)$ denote the set of germs of smooth functions at p. For $f \in C^\infty(p)$, we call *nonholonomic derivatives of order 1 of f* the Lie derivatives $X_1 f, \dots, X_m f$. We call further $X_i(X_j f)$, $X_i(X_j(X_k f))$,... the *nonholonomic derivatives of f of order 2, 3,...* The nonholonomic derivative of order 0 of f at p is $f(p)$.

Proposition 3.1. *Let $f \in C^\infty(p)$. Then $\mathrm{ord}_p(f)$ is equal to the biggest integer k such that all nonholonomic derivatives of f of order smaller than k vanish at p. Moreover,*

$$f(q) = O\big(d(p,q)^{\mathrm{ord}_p(f)}\big).$$

Proof. The proposition results from the following two assertions:

(i) if ℓ is an integer such that $\ell < \mathrm{ord}_p(f)$, then all nonholonomic derivatives of f of order $\leq \ell$ vanish at p;

(ii) if ℓ is an integer such that all nonholonomic derivatives of f of order $\leq \ell$ vanish at p, then $f(q) = O\big(d(p,q)^{\ell+1}\big)$.

Let us first prove point *(i)*. Let ℓ be an integer such that $\ell < \mathrm{ord}_p(f)$. We write a nonholonomic derivative of f of order $k \leq \ell$ as

$$(X_{i_1} \dots X_{i_k} f)(p) = \frac{\partial^k}{\partial t_1 \cdots \partial t_k} f\big(\exp(t_k X_{i_k}) \circ \cdots \circ \exp(t_1 X_{i_1})(p)\big)\Big|_{t=0}.$$

The point $q = \exp(t_k X_{i_k}) \circ \cdots \circ \exp(t_1 X_{i_1})(p)$ is the endpoint of a trajectory of length $|t_1| + \cdots + |t_n|$. Therefore, $d(p,q) \leq |t_1| + \cdots + |t_n|$.

Since $k \leq \ell < \mathrm{ord}_p(f)$, there exists a real number $s > 0$ such that $f(q) = O\big((|t_1| + \cdots + |t_n|)^{k+s}\big)$. This implies that

$$(X_{i_1} \dots X_{i_k} f)(p) = \frac{\partial^k}{\partial t_1 \cdots \partial t_k} f(q)\Big|_{t=0} = 0.$$

Thus point *(i)* is proved.

The proof of point *(ii)* goes by induction on ℓ. For $\ell = 0$, assume that all nonholonomic derivatives of f of order ≤ 0 vanish at p, that is $f(p) = 0$. Choose any Riemannian metric on M and denote by d_R the associated Riemannian distance on M. Since f is smooth, there holds $f(q) \leq Cst\ d_R(p,q)$ near p. By Theorem 2.3, this implies $f(q) \leq Cst\ d(p,q)$, and so property *(ii)* for $\ell = 0$.

Assume that, for a given $\ell \geq 0$, *(ii)* holds for any function f (induction hypothesis) and take a function f such that all its nonholonomic derivatives of order $< \ell + 1$ vanish at p.

Observe that, for $i = 1, \dots, m$, all the nonholonomic derivatives of $X_i f$ of order $< \ell$ vanish at p. Indeed, $X_{i_1} \dots X_{i_k}(X_i f) = X_{i_1} \dots X_{i_k} X_i f$. Applying the induction hypothesis to $X_i f$ leads to $X_i f(q) = O\big(d(p,q)^\ell\big)$. In other words, there exist positive constants $C_1, \dots, C_m$ such that, for q close enough to p,

$$X_i f(q) \leq C_i d(p,q)^\ell.$$

Fix now a point q near p. By Remark 1.2, there exists a minimizing curve $\gamma(\cdot)$ of velocity one joining p to q. Therefore γ satisfies

$$\dot\gamma(t) = \sum_{i=1}^m u_i(t) X_i\big(\gamma(t)\big) \quad \text{for a.e. } t \in [0,T], \qquad \gamma(0) = p,\ \gamma(T) = q,$$

with $\sum_i u_i^2(t) = 1$ a.e. and $d\big(p, \gamma(t)\big) = t$ for any $t \in [0,T]$. In particular $d(p,q) = T$.

To estimate $f(q) = f\big(\gamma(T)\big)$, we compute the derivative of $f\big(\gamma(t)\big)$ with respect to t,

$$\begin{aligned} \frac{d}{dt} f\big(\gamma(t)\big) &= \sum_{i=1}^{m} u_i(t) X_i f\big(\gamma(t)\big), \\ \Rightarrow \left| \frac{d}{dt} f\big(\gamma(t)\big) \right| &\leq \sum_{i=1}^{m} |u_i(t)| C_i d\big(p, \gamma(t)\big)^{\ell} \leq C t^{\ell}, \end{aligned}$$

where $C = C_1 + \cdots + C_m$. Integrating this inequality between 0 and t gives

$$\big|f\big(\gamma(t)\big)\big| \leq |f(p)| + \frac{C}{\ell+1} t^{\ell+1}.$$

Note that $f(p) = 0$, since the nonholonomic derivative of f of order 0 at p vanishes. Finally, at $t = T = d(p, q)$, we obtain

$$|f(q)| \leq \frac{C}{\ell+1} T^{\ell+1},$$

which concludes the proof of *(ii)*. □

As a consequence, the nonholonomic order of a smooth (germ of) function is given by the formula

$$\mathrm{ord}_p(f) = \min\big\{ s \in \mathbb{N} \ : \ \exists\, i_1, \ldots, i_s \in \{1, \ldots, m\} \text{ s.t. } (X_{i_1} \ldots X_{i_s} f)(p) \neq 0 \big\},$$

where as usual we adopt the convention that $\min \emptyset = +\infty$.

It is clear now that any function in $C^\infty(p)$ vanishing at p is of order ≥ 1. Moreover, the following basic computation rules are satisfied: for every f, g in $C^\infty(p)$ and every $\lambda \in \mathbb{R} \setminus \{0\}$,

$$\begin{aligned} \mathrm{ord}_p(fg) &\geq \mathrm{ord}_p(f) + \mathrm{ord}_p(g), \\ \mathrm{ord}_p(\lambda f) &= \mathrm{ord}_p(f), \\ \mathrm{ord}_p(f+g) &\geq \min\big(\mathrm{ord}_p(f), \mathrm{ord}_p(g)\big). \end{aligned}$$

Note that the first inequality is actually an equality. However the proof of this fact requires an additional result (see Proposition 3.2).

The notion of nonholonomic order extends to vector fields. Let $VF(p)$ denote the set of germs of smooth vector fields at p.

Definition 3.2. Let $X \in VF(p)$. The *nonholonomic order of X at p*, denoted by $\mathrm{ord}_p(X)$, is the real number defined by:

$$\mathrm{ord}_p(X) = \sup\big\{ \sigma \in \mathbb{R} \ : \ \mathrm{ord}_p(Xf) \geq \sigma + \mathrm{ord}_p(f), \quad \forall f \in C^\infty(p) \big\}.$$

The order of a differential operator is defined in the same way.

Note that $\mathrm{ord}_p(X) \in \mathbb{Z}$ since the order of a smooth function is an integer. Moreover the null vector field $X \equiv 0$ has infinite order, $\mathrm{ord}_p(0) = +\infty$.

Since the order of a function coincides with its order as a differential operator acting by multiplication, we have the following properties. For every $X, Y \in VF(p)$ and every $f \in C^\infty(p)$,

$$\begin{array}{rcl} \mathrm{ord}_p([X,Y]) & \geq & \mathrm{ord}_p(X) + \mathrm{ord}_p(Y), \\ \mathrm{ord}_p(fX) & \geq & \mathrm{ord}_p(f) + \mathrm{ord}_p(X), \\ \mathrm{ord}_p(X) & \leq & \mathrm{ord}_p(Xf) - \mathrm{ord}_p(f), \\ \mathrm{ord}_p(X+Y) & \geq & \min\big(\mathrm{ord}_p(X), \mathrm{ord}_p(Y)\big). \end{array} \tag{9}$$

As already noticed for functions, the second inequality is in fact an equality. This is not the case for the first inequality (take for instance $X = Y$).

As a consequence of (9), $X_1, \ldots, X_m$ are of order ≥ -1, $[X_i, X_j]$ of order ≥ -2, and more generally, every X in the set Δ^k is of order $\geq -k$.

Example 3.2 (Euclidean case). In the Euclidean case (see example 3.1), the nonholonomic order of a constant differential operator is the negative of its usual order. For instance ∂_{x_i} is of nonholonomic order -1. Actually, in this case, every vector field that does not vanish at p is of nonholonomic order -1.

Example 3.3 (Heisenberg case). Consider the following vector fields on $\mathbb{R}^3$:

$$X_1 = \partial_x - \frac{y}{2}\partial_z \quad \text{and} \quad X_2 = \partial_y + \frac{x}{2}\partial_z.$$

The coordinate functions x and y have order 1 at 0, whereas z has order 2 at 0, since $X_1 x(0) = X_2 y(0) = 1$, $X_1 z(0) = X_2 z(0) = 0$, and $X_1 X_2 z(0) = 1/2$. These relations also imply $\mathrm{ord}_0(X_1) = \mathrm{ord}_0(X_2) = -1$. Finally, the Lie bracket $[X_1, X_2] = \partial_z$ is of order -2 at 0 since $[X_1, X_2]z = 1$.

We are now in a position to give a meaning to first-order approximation.

Definition 3.3. A family of m vector fields $\widehat{X}_1, \ldots, \widehat{X}_m$ defined near p is called a *first-order approximation of* $X_1, \ldots, X_m$ *at* p if the vector fields $X_i - \widehat{X}_i$, $i = 1, \ldots, m$, are of order ≥ 0 at p.

A consequence of this definition is that the order at p defined by the vector fields $\widehat{X}_1, \ldots, \widehat{X}_m$ coincides with the one defined by $X_1, \ldots, X_m$. Hence for any $f \in C^\infty(p)$ of order greater than $s-1$,

$$(X_{i_1} \ldots X_{i_s} f)(q) = (\widehat{X}_{i_1} \ldots \widehat{X}_{i_s} f)(q) + O\left(d(p,q)^{\mathrm{ord}_p(f)-s+1}\right).$$

To go further in the characterization of orders and approximations, we need suitable systems of coordinates.

3.2 Privileged coordinates

We have introduced in Section 2.1 the sets of vector fields Δ^s, defined by $\Delta^s = \text{span}\{X_I \ : \ |I| \leq s\}$. Since $X_1, \ldots, X_m$ satisfy Chow's Condition, the values of these sets at p form a flag of subspaces of T_pM, that is,

$$\Delta^1(p) \subset \Delta^2(p) \subset \cdots \subset \Delta^{r-1}(p) \subsetneq \Delta^r(p) = T_pM, \tag{10}$$

where $r = r(p)$ is called the *degree of nonholonomy at p*.

Set $n_i(p) = \dim \Delta^i(p)$. The r-tuple of integers $(n_1(p), \ldots, n_r(p))$ is called the *growth vector at p*. The first integer $n_1(p) \leq m$ is the rank of the family $X_1(p), \ldots, X_m(p)$, and the last one $n_r(p) = n$ is the dimension of the manifold M.

Let $s \geq 1$. By abuse of notations, we continue to write Δ^s for the map $q \mapsto \Delta^s(q)$. This map Δ^s is a distribution if and only if $n_s(q)$ is constant on M. We then distinguish two kind of points.

Definition 3.4. The point p is a *regular point* if the growth vector is constant in a neighbourhood of p. Otherwise, p is a *singular point*.

Thus, near a regular point, all maps Δ^s are locally distributions.

The structure of flag (10) may also be described by another sequence of integers. We define the *weights at p*, $w_i = w_i(p)$, $i = 1, \ldots, n$, by setting $w_j = s$ if $n_{s-1}(p) < j \leq n_s(p)$, where $n_0 = 0$. In other words, we have

$$w_1 = \cdots = w_{n_1} = 1, \ w_{n_1+1} = \cdots = w_{n_2} = 2, \ldots, \\ w_{n_{r-1}+1} = \cdots = w_{n_r} = r.$$

The weights at p form an increasing sequence $w_1(p) \leq \cdots \leq w_n(p)$ which is constant near p if and only if p is a regular point.

Example 3.4 (Heisenberg case). The Heisenberg case in $\mathbb{R}^3$ given in example 3.3 has a growth vector which is equal to $(2, 3)$ at every point. Therefore all points of $\mathbb{R}^3$ are regular. The weights at any point are $w_1 = w_2 = 1$, $w_3 = 2$.

Example 3.5 (Martinet case). Consider the following vector fields on $\mathbb{R}^3$,

$$X_1 = \partial_x \quad \text{and} \quad X_2 = \partial_y + \frac{x^2}{2}\partial_z.$$

The only nonzero brackets are

$$X_{12} = [X_1, X_2] = x\partial_z \quad \text{and} \quad X_{112} = [X_1, [X_1, X_2]] = \partial_z.$$

Thus the growth vector is equal to $(2, 2, 3)$ on the plane $\{x = 0\}$, and to $(2, 3)$ elsewhere. As a consequence, the set of singular points is the plane $\{x = 0\}$. The weights are $w_1 = w_2 = 1$, $w_3 = 2$ at regular points, and $w_1 = w_2 = 1$, $w_3 = 3$ at singular ones.

Example 3.6. Consider the vector fields on $\mathbb{R}^3$

$$X_1 = \partial_x \quad \text{and} \quad X_2 = \partial_y + f(x)\partial_z,$$

where f is a smooth function on $\mathbb{R}$ which admits every positive integer $n \in \mathbb{N}$ as a zero with multiplicity n (such a function exists and can even be chosen in the analytic class thanks to the Weierstrass factorization theorem [Rud70, Th. 15.9]). Every point (n, y, z) is singular and the weights at this point are $w_1 = w_2 = 1$, $w_3 = n + 1$. As a consequence the degree of nonholonomy w_3 is unbounded on $\mathbb{R}^3$.

Let us give some basic properties of the growth vector and of the weights.

- At a regular point, the growth vector is a strictly increasing sequence: $n_1(p) < \cdots < n_r(p)$. Indeed, if $n_s(q) = n_{s+1}(q)$ in a neighbourhood of p, then Δ^s is locally an involutive distribution and so $s = r$. As a consequence, at a regular point p, the jump between two successive weights is never greater than 1, $w_{i+1} - w_i \leq 1$, and there holds $r(p) \leq n - m + 1$.

- For every s, the map $q \mapsto n_s(q)$ is a lower semi-continuous function from M to $\mathbb{N}$. Therefore, the set of regular points is open and dense in M.

- For every $i = 1, \ldots, n$, the weight $w_i(\cdot)$ is an upper semi-continuous function. In particular, this is the case for the degree of nonholonomy $r(\cdot) = w_n(\cdot)$, that is, $r(q) \leq r(p)$ for q near p. As a consequence $r(\cdot)$ is bounded on any compact subset of M.

- The degree of nonholonomy may be unbounded on M (see example 3.6 above). Thus determining if a nonholonomic system is controllable is a non decidable problem: the computation of an infinite number of brackets may be needed to decide if Chow's Condition is satisfied.

  However, in the case of polynomial vector fields on $\mathbb{R}^n$ (relevant in practice), it can be shown that the degree of nonholonomy is bounded by a universal function of the degree k of the polynomials (see [Gab95, GJR98]):

  $$r(x) \leq 2^{3n^2} n^{2n} k^{2n}.$$

The meaning of the sequence of weights is best understood in terms of basis of T_pM. Choose first vector fields $Y_1, \dots, Y_{n_1}$ in Δ^1 whose values at p form a basis of $\Delta^1(p)$. Choose then vector fields $Y_{n_1+1}, \dots, Y_{n_2}$ in Δ^2 such that the values $Y_1(p), \dots, Y_{n_2}(p)$ form a basis of $\Delta^2(p)$. For each s, choose $Y_{n_{s-1}+1}, \dots, Y_{n_s}$ in Δ^s such that $Y_1(p), \dots, Y_{n_s}(p)$ form a basis of $\Delta^s(p)$. We obtain in this way a family of vector fields $Y_1, \dots, Y_n$ such that

$$\begin{cases} Y_1(p), \dots, Y_n(p) \text{ is a basis of } T_pM, \\ Y_i \in \Delta^{w_i}, \ i = 1, \dots, n. \end{cases} \tag{11}$$

A family of n vector fields satisfying (11) is called an *adapted frame at* p. The word "adapted" means here "adapted to the flag (10)", since the values at p of an adapted frame contain a basis $Y_1(p), \dots, Y_{n_s}(p)$ of each subspace $\Delta^s(p)$ of the flag. By continuity, at a point q close enough to p, the values of $Y_1, \dots, Y_n$ still form a basis of T_qM. However, if p is singular, this basis may not be adapted to the flag (10) at q.

Let us explain now the relation between weights and orders. We write first the tangent space as a direct sum,

$$T_pM = \Delta^1(p) \oplus \Delta^2(p)/\Delta^1(p) \oplus \cdots \oplus \Delta^r(p)/\Delta^{r-1}(p),$$

where $\Delta^s(p)/\Delta^{s-1}(p)$ denotes a supplementary of $\Delta^{s-1}(p)$ in $\Delta^s(p)$, and take a local system of coordinates $(y_1, \dots, y_n)$. The dimension of each space $\Delta^s(p)/\Delta^{s-1}(p)$ is equal to $n_s - n_{s-1}$, and we can assume that, up to a reordering, we have $dy_j(\Delta^s(p)/\Delta^{s-1}(p)) \neq 0$ for $n_{s-1} < j \leq n_s$.

Take an integer j such that $0 < j \leq n_1$. From the assumption above, there holds $dy_j(\Delta^1(p)) \neq 0$, and consequently there exists X_i such that $dy_j(X_i(p)) \neq 0$. Since $dy_j(X_i) = X_i y_j$ is a first-order nonholonomic derivative of y_j, we have $\mathrm{ord}_p(y_j) \leq 1 = w_j$.

Take now an integer j such that $n_{s-1} < j \leq n_s$ for $s > 1$, that is, $w_j = s$. Since $dy_j(\Delta^s(p)/\Delta^{s-1}(p)) \neq 0$, there exists a vector field Y in Δ^s such that $dy_j(Y(p)) = (Yy_j)(p) \neq 0$. By definition of Δ^s, the Lie derivative Yy_j is a linear combination of nonholonomic derivatives of y_j of order not greater than s. One of them must be nonzero, and so $\mathrm{ord}_p(y_j) \leq s = w_j$.

Finally, any system of local coordinates $(y_1, \dots, y_n)$ satisfies $\mathrm{ord}_p(y_j) \leq w_j$ up to a reordering (or $\sum_{i=1}^n \mathrm{ord}_p(y_i) \leq \sum_{i=1}^n w_i$ without reordering). The coordinates with the maximal possible order will play an important role.

Definition 3.5. A *system of privileged coordinates at* p is a system of local coordinates $(z_1, \dots, z_n)$ such that $\mathrm{ord}_p(z_j) = w_j$ for $j = 1, \dots, n$.

Notice that privileged coordinates $(z_1, \dots, z_n)$ satisfy

$$dz_i(\Delta^{w_i}(p)) \neq 0, \quad dz_i(\Delta^{w_i - 1}(p)) = 0, \quad i = 1, \dots, n, \tag{12}$$

or, equivalently, $\partial_{z_i}|_p$ belongs to $\Delta^{w_i}(p)$ but not to $\Delta^{w_i-1}(p)$. Local coordinates satisfying (12) are called *linearly adapted coordinates* ("adapted" because the differentials at p of the coordinates form a basis of T_p^*M dual to the values of an adapted frame). Thus privileged coordinates are always linearly adapted coordinates. The converse is false, as shown in the example below.

Example 3.7. Take $X_1 = \partial_x$, $X_2 = \partial_y + (x^2+y)\partial_z$ in $\mathbb{R}^3$. The weights at 0 are $(1,1,3)$ and (x,y,z) are adapted at 0. But they are not privileged: indeed, the coordinate z is of order 2 at 0 since $(X_2X_2z)(0) = 1$.

Remark 3.1. As it is suggested by Kupka [Kup96], one can define *privileged functions at p* to be the smooth functions f on U such that

$$\mathrm{ord}_p(f) = \min\{s \in \mathbb{N} \ : \ df(\Delta^s(p)) \neq 0\}.$$

It results from the discussion above that some local coordinates $(z_1, \ldots, z_n)$ are privileged at p if and only if each z_i is a privileged function at p.

Let us now show how to compute orders using privileged coordinates. We fix a system of privileged coordinates $(z_1, \ldots, z_n)$ at p. Given a sequence of integers $\alpha = (\alpha_1, \ldots, \alpha_n)$, we define the weighted degree of the monomial $z^\alpha = z_1^{\alpha_1} \cdots z_n^{\alpha_n}$ to be $w(\alpha) = w_1\alpha_1 + \cdots + w_n\alpha_n$ and the weighted degree of the monomial vector field $z^\alpha\partial_{z_j}$ as $w(\alpha) - w_j$. The weighted degrees allow to compute the orders of functions and vector fields in a purely algebraic way.

Proposition 3.2. *For a smooth function f with a Taylor expansion*

$$f(z) \sim \sum_\alpha c_\alpha z^\alpha,$$

the order of f is the least weighted degree of monomials having a nonzero coefficient in the Taylor series.

For a vector field X with a Taylor expansion

$$X(z) \sim \sum_{\alpha,j} a_{\alpha,j} z^\alpha \partial_{z_j},$$

the order of X is the least weighted degree of a monomial vector fields having a nonzero coefficient in the Taylor series.

In other words, when using privileged coordinates, the notion of nonholonomic order amounts to the usual notion of vanishing order at some point, only assigning weights to the variables.

Proof. For $i = 1, \dots, n$, we have $\partial_{z_i}|_p \in \Delta^{w_i}(p)$. Then there exist n vector fields $Y_1, \dots, Y_n$ which form an adapted frame at p and such that $Y_1(p) = \partial_{z_1}|_p$, $\dots$, $Y_n(p) = \partial_{z_n}|_p$. For every i, the vector field Y_i is of order $\geq -w_i$ at p since it belongs to Δ^{w_i}. Moreover we have $(Y_i z_i)(p) = 1$ and $\mathrm{ord}_p(z_i) = w_i$. Thus $\mathrm{ord}_p(Y_i) = -w_i$.

Take a sequence of integers $\alpha = (\alpha_1, \dots, \alpha_n)$. The monomial z^α is of order $\geq w(\alpha)$ at p and the differential operator $Y^\alpha = Y_1^{\alpha_1} \cdots Y_n^{\alpha_n}$ is of order $\geq -w(\alpha)$. Observing that $(Y_i z_j)(p) = 0$ if $j \neq i$, we easily see that $(Y^\alpha z^\alpha)(p) = \frac{1}{\alpha_1! \dots \alpha_n!} \neq 0$, whence $\mathrm{ord}_p(z^\alpha) = w(\alpha)$.

In the same way, we obtain that, if z^α, z^β are two different monomials and λ, μ two nonzero real numbers, then $\mathrm{ord}_p(\lambda z^\alpha + \mu z^\beta) = \min\big(w(\alpha), w(\beta)\big)$. Thus the order of a series is the least weighted degree of monomials actually appearing in the series itself. This shows the statement on order of functions.

As a consequence, for any smooth function f, the order at p of $\partial_{z_i} f$ is $\geq \mathrm{ord}_p(f) - w_i$. Since moreover $\partial_{z_i} z_i = 1$, we obtain that $\mathrm{ord}_p(\partial_{z_i})$ is equal to $-w_i$. The second part of the statement follows. □

A notion of homogeneity is also naturally associated with a system of privileged coordinates $(z_1, \dots, z_n)$ defined on U. We define first the one-parameter family of dilations

$$\delta_t : (z_1, \dots, z_n) \mapsto (t^{w_1} z_1, \dots, t^{w_n} z_n), \qquad t \geq 0.$$

Each dilation δ_t is a map from $\mathbb{R}^n$ to $\mathbb{R}^n$. By abuse of notations, for $q \in U$ and t small enough we write $\delta_t(q)$ instead of $\delta_t(z(q))$, where $z(q)$ are the coordinates of q. A dilation δ_t acts also on functions and vector fields by pull-back: $\delta_t^* f = f \circ \delta_t$ and $\delta_t^* X$ is the vector field such that $(\delta_t^* X)(\delta_t^* f) = \delta_t^*(Xf)$.

Definition 3.6. A function f is *homogeneous of degree s* if $\delta_t^* f = t^s f$. A vector field X is *homogeneous of degree σ* if $\delta_t^* X = t^\sigma X$.

For a smooth function (resp. a smooth vector field), this is the same as being a finite sum of monomials (resp. monomial vector fields) of weighted degree s. As a consequence, if a function f is homogeneous of degree s, then it is of order s at p.

A typical degree 1 homogeneous function is the so-called *pseudo-norm at p*, defined by:

$$z \mapsto \|z\|_p = |z_1|^{1/w_1} + \cdots + |z_n|^{1/w_n}. \tag{13}$$

When composed with the coordinates function, the pseudo-norm at p is a (non smooth) function of order 1, that is,

$$\|z(q)\|_p = O\big(d(p,q)\big).$$

Actually, it results from Proposition 3.2 that the order of a function $f \in C^\infty(p)$ is the least integer s such that $f(q) = O(\|z(q)\|_p^s)$.

Examples of privileged coordinates. Of course all the results above on algebraic computation of orders hold only if privileged coordinates do exist. Two types of privileged coordinates are commonly used in the literature.

a. Exponential coordinates. Choose an adapted frame $Y_1, \dots, Y_n$ at p. The inverse of the local diffeomorphism

$$(z_1, \dots, z_n) \mapsto \exp(z_1 Y_1 + \cdots + z_n Y_n)(p)$$

defines a system of local privileged coordinates at p, called *canonical coordinates of the first kind*. These coordinates are mainly used in the context of hypoelliptic operator and for nilpotent Lie groups with right (or left) invariant sub-Riemannian structure.

The inverse of the local diffeomorphism

$$(z_1, \dots, z_n) \mapsto \exp(z_n Y_n) \circ \cdots \circ \exp(z_1 Y_1)(p)$$

also defines privileged coordinates at p, called *canonical coordinates of the second kind*. They are easier to work with than the one of the first kind. For instance, in these coordinates, the vector field Y_n read as ∂_{z_n}. One can also exchange the order of the flows in the definition to obtain any of the Y_i as ∂_{z_i}. The fact that canonical coordinates of both first and second kind are privileged is proved in Section B.1.

We leave it to the reader to verify that the diffeomorphism

$$(z_1, \dots, z_n) \mapsto \exp(z_n Y_n + \cdots + z_{s+1} Y_{s+1}) \circ \exp(z_s Y_s) \cdots \circ \exp(z_1 Y_1)(p)$$

also induces privileged coordinates. As a matter of fact, any "mix" between first and second kind canonical coordinates defines privileged coordinates.

b. Algebraic coordinates. There exist also effective constructions of privileged coordinates (the construction of exponential coordinates is not effective in general since it requires to integrate flows of vector fields). We present here Bellaïche's algorithm, but other constructions exist (see [Ste86, AS87]).

1. Choose an adapted frame $Y_1, \dots, Y_n$ at p.
2. Choose coordinates $(y_1, \dots, y_n)$ centered at p such that $\partial_{y_i}|_p = Y_i(p)$.
3. For $j = 1, \dots, n$, set

$$z_j = y_j - \sum_{k=2}^{w_j - 1} h_k(y_1, \dots, y_{j-1}),$$

where, for $k = 2, \dots, w_j - 1$,

$$h_k(y_1, \dots, y_{j-1}) = \sum_{\substack{|\alpha|=k \\ w(\alpha)<w_j}} Y_1^{\alpha_1} \dots Y_{j-1}^{\alpha_{j-1}} \Big(y_j - \sum_{q=2}^{k-1} h_q(y) \Big)(p) \, \frac{y_1^{\alpha_1}}{\alpha_1!} \cdots \frac{y_{j-1}^{\alpha_{j-1}}}{\alpha_{j-1}!},$$

with $|\alpha| = \alpha_1 + \cdots + \alpha_n$.

The fact that coordinates $(z_1, \dots, z_n)$ are privileged at p will be proved in Section B.3.

Coordinates $(y_1, \dots, y_n)$ are linearly adapted coordinates. They can be obtained from any original system of coordinates by an affine change. The privileged coordinates $(z_1, \dots, z_n)$ are then obtained from $(y_1, \dots, y_n)$ by an expression of the form

$$\begin{aligned} z_1 &= y_1, \\ z_2 &= y_2 + \mathrm{pol}(y_1), \\ &\vdots \\ z_n &= y_n + \mathrm{pol}(y_1, \dots, y_{n-1}), \end{aligned}$$

where each pol is a polynomial function without constant nor linear terms. The inverse change of coordinates takes the same triangular form, which makes the use of these coordinates easy for computations.

3.3 Nilpotent approximation

Fix a system of privileged coordinates $(z_1, \dots, z_n)$ at p. Every vector field X_i is of order ≥ -1, hence it has, in z coordinates, a Taylor expansion

$$X_i(z) \sim \sum_{\alpha,j} a_{\alpha,j} z^\alpha \partial_{z_j},$$

where $w(\alpha) \geq w_j - 1$ if $a_{\alpha,j} \neq 0$. Grouping together the monomial vector fields of same weighted degree, we express X_i as a series

$$X_i = X_i^{(-1)} + X_i^{(0)} + X_i^{(1)} + \cdots$$

where $X_i^{(s)}$ is a homogeneous vector field of degree s.

Proposition 3.3. *Set $\widehat{X}_i = X_i^{(-1)}$, $i = 1, \dots, m$. The family of vector fields $\widehat{X}_1, \dots, \widehat{X}_m$ is a first-order approximation of $X_1, \dots, X_m$ at p and generate a nilpotent Lie algebra of step $r = w_n$.*

Proof. The fact that the vector fields $\widehat{X}_1, \dots, \widehat{X}_m$ form a first-order approximation of $X_1, \dots, X_m$ results from their construction.

Note further that any homogeneous vector field of degree smaller than $-w_n$ is zero, as it is easy to check in privileged coordinates. Moreover, if X and Y are homogeneous of degree respectively k and l, then the bracket $[X, Y]$ is homogeneous of degree $k + l$ because $\delta_t^*[X, Y] = [\delta_t^* X, \delta_t^* Y] = t^{k+l}[X, Y]$.

It follows that every iterated bracket of the vector fields $\widehat{X}_1, \dots, \widehat{X}_m$ of length k (i.e. containing k of these vector field) is homogeneous of degree $-k$ and is zero if $k > w_n$. □

Definition 3.7. The family $(\widehat{X}_1, \dots, \widehat{X}_m)$ is called the *(homogeneous) nilpotent approximation* of $(X_1, \dots, X_m)$ at p associated with the coordinates z.

Example 3.8 (unicycle). Consider the vector fields on $\mathbb{R}^2 \times S^1$ defining the kinematic model of a unicycle (see example 1.1), that is, $X_1 = \cos\theta\partial_x + \sin\theta\partial_y$, $X_2 = \partial_\theta$. We have $[X_1, X_2] = \sin\theta\partial_x - \cos\theta\partial_y$, so the weights are $(1, 1, 2)$ at every point. At $p = 0$, the coordinates (x, θ) have order 1 and y has order 2, consequently (x, θ, y) is a system of privileged coordinates at 0. Taking the Taylor expansion of X_1 and X_2 in the latter coordinates, we obtain the homogeneous components:

$$X_1^{(-1)} = \partial_x + \theta\partial_y, \quad X_1^{(0)} = 0, \quad X_1^{(1)} = -\frac{\theta^2}{2}\partial_x - \frac{\theta^3}{3!}\partial_y, \quad \dots$$

and $X_2^{(-1)} = X_2 = \partial_\theta$. Therefore the homogeneous nilpotent approximation of (X_1, X_2) at 0 in coordinates (x, θ, y) is

$$\widehat{X}_1 = \partial_x + \theta\partial_y, \quad \widehat{X}_2 = \partial_\theta.$$

We easily check that the Lie brackets of length 3 of these vectors are zero, that is, $[\widehat{X}_1, [\widehat{X}_1, \widehat{X}_2]] = [\widehat{X}_2, [\widehat{X}_1, \widehat{X}_2]] = 0$, and so the Lie algebra $Lie(\widehat{X}_1, \widehat{X}_2)$ is nilpotent of step 2.

The homogeneous nilpotent approximation is not intrinsic to the frame $(X_1, \dots, X_m)$, since it depends on the chosen system of privileged coordinates. However, if $\widehat{X}_1, \dots, \widehat{X}_m$ and $\widehat{X}'_1, \dots, \widehat{X}'_m$ are the nilpotent approximations associated with two different systems of coordinates, then their Lie algebras $Lie(\widehat{X}_1, \dots, \widehat{X}_m)$ and $Lie(\widehat{X}'_1, \dots, \widehat{X}'_m)$ are isomorphic. If moreover p is a regular point, then $Lie(\widehat{X}_1, \dots, \widehat{X}_m)$ is isomorphic to the graded nilpotent Lie algebra

$$\mathrm{Gr}(\Delta)_p = \Delta(p) \oplus (\Delta^2/\Delta^1)(p) \oplus \cdots \oplus (\Delta^{r-1}/\Delta^r)(p).$$

Remark 3.2. The nilpotent approximation denotes in fact two different objects. Each $\widehat{X}_i$ can be seen as a vector field on $\mathbb{R}^n$ or as the representation in z coordinates of the vector field $z^*\widehat{X}_i$ defined on a neighbourhood of p in M. This will cause no confusion since the nilpotent approximation is associated with a given system of privileged coordinates.

It is worth to notice the particular form of the nilpotent approximation in privileged coordinates. Write $\widehat{X}_i = \sum_{j=1}^n f_{ij}(z)\partial_{z_j}$. Since $\widehat{X}_i$ is homogeneous of degree -1 and ∂_{z_j} of degree $-w_j$, the function f_{ij} is a homogeneous polynomial of weighted degree $w_j - 1$. In particular it can not involve variables of weight greater than $w_j - 1$, that is,

$$\widehat{X}_i(z) = \sum_{j=1}^{n} f_{ij}(z_1, \ldots, z_{n_{w_j}-1})\partial_{z_j}.$$

The nonholonomic control system $\dot{z} = \sum_{i=1}^m u_i \widehat{X}_i(z)$ associated with the nilpotent approximation is then polynomial and in a triangular form,

$$\dot{z}_j = \sum_{i=1}^{m} u_i f_{ij}(z_1, \ldots, z_{n_{w_j}-1}).$$

Computing the trajectories of a system in such a form is rather easy: given the input function $(u_1(t), \ldots, u_m(t))$, it is possible to compute the coordinates z_j one after the other, only by integration.

As vector fields on $\mathbb{R}^n$, $\widehat{X}_1, \ldots, \widehat{X}_m$ generate a sub-Riemannian distance on $\mathbb{R}^n$ which is homogeneous with respect to the dilation δ_t.

Lemma 3.4.

(i) The family $(\widehat{X}_1, \ldots, \widehat{X}_m)$ satisfies Chow's Condition on $\mathbb{R}^n$.

(ii) The growth vector at 0 of $(\widehat{X}_1, \ldots, \widehat{X}_m)$ is equal to the one at p of $(X_1, \ldots, X_m)$.

Let $\widehat{d}$ be the sub-Riemannian distance on $\mathbb{R}^n$ associated with $(\widehat{X}_1, \ldots, \widehat{X}_m)$.

(iii) The distance $\widehat{d}$ is homogeneous of degree 1,

$$\widehat{d}(\delta_t x, \delta_t y) = t\widehat{d}(x, y).$$

(iv) There exists a constant $C > 0$ such that, for all $z \in \mathbb{R}^n$,

$$\frac{1}{C}\|z\|_p \leq \widehat{d}(0, z) \leq C\|z\|_p,$$

where $\|\cdot\|_p$ denotes the pseudo-norm at p (see (13)).

Proof. Through the coordinates z we identify the neighbourhood U of p in M with a neighbourhood of 0 in $\mathbb{R}^n$.

For every iterated bracket $X_I = [X_{i_k}, \dots, [X_{i_2}, X_{i_1}]]$ of the vector fields $X_1, \dots, X_m$, we set$\widehat{X}_I = [\widehat{X}_{i_k}, \dots, [\widehat{X}_{i_2}, \widehat{X}_{i_1}]]$, and for $k \geq 1$ we set $\widehat{\Delta}^k = \operatorname{span}\{\widehat{X}_I \,:\, |I| \leq k\}$. As noticed in the proof of Proposition 3.3, a bracket $\widehat{X}_I$ of length $|I| = k$ is homogeneous of weighted degree $-k$, and by construction of the nilpotent approximation, there holds $X_I = \widehat{X}_I +$ terms of order $> -k$. Therefore,

$$\widehat{X}_I(0) = X_I(p) \bmod \operatorname{span}\{\partial_{z_j}\big|_p \,:\, w_j < k\} = X_I(p) \bmod \Delta^{k-1}(p).$$

As a consequence, for any integer $k \geq 1$, we have

$$\dim \widehat{\Delta}^k(0) = \dim \Delta^k(p), \tag{14}$$

and property (ii) follows. Moreover, if $X_{I_1}, \dots, X_{I_n}$ form an adapted frame at p, then the family $(\widehat{X}_{I_1}(0), \dots, \widehat{X}_{I_n}(0))$ is of rank n, which implies that its determinant is nonzero. Since the determinant of $X_{I_1}, \dots, X_{I_n}$ is an homogeneous polynomial of weighted degree 0, it is nonzero everywhere, which implies (i).

As for the property (iii), consider the nonholonomic system defined by the nilpotent approximation, that is, $\dot{z} = \sum_{i=1}^m u_i \widehat{X}_i(z)$. Observe that, if $\widehat{\gamma}$ is a trajectory of this system, that is, if

$$\dot{\widehat{\gamma}}(t) = \sum_{i=1}^m u_i \widehat{X}_i\big(\widehat{\gamma}(t)\big), \quad t \in [0, T],$$

then the dilated curve $\delta_\lambda \widehat{\gamma}$ satisfies

$$\frac{d}{dt}\delta_\lambda \widehat{\gamma}(t) = \sum_{i=1}^m \lambda u_i \widehat{X}_i\big(\delta_\lambda \widehat{\gamma}(t)\big), \quad t \in [0, T].$$

Thus $\delta_\lambda \widehat{\gamma}$ is a trajectory of the same system, with extremities $(\delta_\lambda \widehat{\gamma})(0) = \delta_\lambda(\widehat{\gamma}(0))$ and $(\delta_\lambda \widehat{\gamma})(T) = \delta_\lambda(\widehat{\gamma}(T))$, and its length equals $\lambda \operatorname{length}(\widehat{\gamma})$. This proves the homogeneity of $\widehat{d}$.

Finally, since $(\widehat{X}_1, \dots, \widehat{X}_m)$ satisfies Chow's Condition, the distance $\widehat{d}(0, \cdot)$ is continuous on $\mathbb{R}^n$ (see Corollary 2.5). We can then choose a real number $C > 0$ such that, on the compact set $\{\|z\| = 1\}$, we have $1/C \leq \widehat{d}(0, z) \leq C$. Both functions $\widehat{d}(0, z)$ and $\|z\|$ being homogeneous of degree 1, the inequality of Property (iv) follows. □

3.4 Distance estimates

As it is the case for Riemannian distances, in general it is impossible to compute analytically a sub-Riemannian distance (it would require to determine all minimizing curves). This is very important to obtain estimates of the distance, at least locally. In a Riemannian manifold (M, g), the situation is rather simple: in local coordinates x centered at a point p, the Riemannian distance d_R satisfies:

$$d_R(q, q') = \|x(q) - x(q')\|_{g_p} + o(\|x(q)\|_{g_p} + \|x(q')\|_{g_p}),$$

where $\|\cdot\|_{g_p}$ is the Euclidean norm induced by the value g_p of the metric g at p. This formula has two consequences: first, it shows that the Riemannian distance behaves at the first-order as the Euclidean distance associated with $\|\cdot\|_{g_p}$; secondly, the norm $\|\cdot\|_{g_p}$ gives explicit estimates of d_R near p, such as

$$\frac{1}{C}\|x(q)\|_{g_p} \leq d_R(p, q) \leq C\|x(q)\|_{g_p}.$$

In sub-Riemannian geometry, the two properties above hold, but do not depend on the same function: the first-order behaviour near p is characterized by the distance $\widehat{d}_p$ defined by a nilpotent approximation at p, whereas explicit local estimates of $d(p, \cdot)$ are given by the pseudo-norm at p $\|\cdot\|_p$ defined in (13). We first present the latter estimates, often referred to as the "Ball-Box Theorem", and then the first-order expansion of d in Theorem 3.8.

Theorem 3.5. *The following statement holds if and only if $z_1, \ldots, z_n$ are privileged coordinates at p:*
there exist constants C_p and $\varepsilon_p > 0$ such that, if $d(p, q^z) < \varepsilon_p$, then

$$\frac{1}{C_p}\|z\|_p \leq d(p, q^z) \leq C_p\|z\|_p \tag{15}$$

(as previously, q^z denotes the point near p with coordinates z and $\|\cdot\|_p$ the pseudo-norm at p).

Corollary 3.6 (Ball-Box Theorem)**.** *Expressed in a given system of privileged coordinates, the sub-Riemannian balls $B(p, \varepsilon)$ satisfy, for $\varepsilon < \varepsilon_p$,*

$$\mathrm{Box}\big(\frac{1}{C_p}\varepsilon\big) \subset B(p, \varepsilon) \subset \mathrm{Box}\big(C_p\varepsilon\big),$$

where $\mathrm{Box}(\varepsilon) = [-\varepsilon^{w_1}, \varepsilon^{w_1}] \times \cdots \times [-\varepsilon^{w_n}, \varepsilon^{w_n}]$.

Remark 3.3. The constants C_p and ε_p depend on the base point p. Around a regular point p_0, it is possible to construct systems of privileged coordinates depending continuously on the base point p. In this case, the corresponding constants C_p and ε_p depend continuously on p. This is no longer true at a singular point. In particular, if p_0 is singular, the estimate (15) does not hold uniformly near p_0: we can not choose the constants C_p and ε_p independently on p near p_0. We will see in section 4.2 uniform versions of the Ball-Box Theorem.

The Ball-Box Theorem is stated in different papers, often under the hypothesis that the point p is regular. To our knowledge, two valid proofs exist, the ones in [NSW85] and in [Bel96]. The result also appears without proof in [Gro96] and in [Ger84], and with erroneous proofs in [Mit85] and in [Mon02].

We present here a proof adapted from the one of Bellaïche (our is much simpler because Bellaïche actually proves a more general result, namely (20)). Basically, the idea is to compare the distances d and $\widehat{d}$. The main step is Lemma 3.7 below, which is essential in other respects to explain the role of nilpotent approximations in control theory.

Fix a point $p \in M$, and a system of privileged coordinates at p. Through these coordinates we identify a neighbourhood of p in M with a neighbourhood of 0 in $\mathbb{R}^n$. As in the preceding subsection, we denote by $\widehat{X}_1, \ldots, \widehat{X}_m$ the homogeneous nilpotent approximation of $X_1, \ldots, X_m$ at p (associated with the given system of privileged coordinates) and by $\widehat{d}$ the induced sub-Riemannian distance on $\mathbb{R}^n$. Recall also that $r = w_n$ denotes the degree of nonholonomy at p.

Lemma 3.7. *There exist constants C and $\varepsilon > 0$ such that, for any $x_0 \in \mathbb{R}^n$ and any $t \in \mathbb{R}^+$ with $\tau = \max(\|x_0\|_p, t) < \varepsilon$, we have*

$$\|x(t) - \widehat{x}(t)\|_p \leq C\tau t^{1/r},$$

where $x(\cdot)$ and $\widehat{x}(\cdot)$ are trajectories of the nonholonomic systems defined respectively by $X_1, \ldots, X_m$ and $\widehat{X}_1, \ldots, \widehat{X}_m$, starting at the same point x_0, associated with the same control function $u(\cdot)$, and satisfying $\|u(t)\| = 1$ a.e.

Proof. The first step is to prove that $\|x(t)\|_p$ and $\|\widehat{x}(t)\|_p \leq Cst\,\tau$ for small enough τ, where *Cst* is a constant. Let us do it for $x(t)$, the proof being exactly the same for $\widehat{x}(t)$.

The equation of the control system associated with $X_1, \ldots, X_m$ is

$$\dot{x}_j = \sum_{i=1}^{m} u_i\big(f_{ij}(x) + r_{ij}(x)\big), \quad j = 1, \ldots, n,$$

where $f_{ij}(x) + r_{ij}(x)$ is of order $\leq w_j - 1$ at 0. Thus, for $j = 1, \ldots, n$ and $i = 1, \ldots, m$, $|f_{ij}(x)+r_{ij}(x)| \leq Cst\, \|x\|_p^{w_j-1}$ when $\|x\|_p$ is small enough. Note that, along the trajectory $x(t)$, $\|x\|_p$ is small when τ is. Since $\|u(t)\| = 1$ a.e., we get:

$$|\dot{x}_j| \leq Cst\, \|x\|_p^{w_j-1}. \tag{16}$$

To integrate this inequality, choose an integer N such that all N/w_j are even integers and set $\|x\|_N = \big(\sum_{i=1}^n |x_i|^{N/w_i}\big)^{1/N}$. The function $\|x\|_N$ is equivalent to $\|x\|_p$ in the norm sense, and it is differentiable except at the origin. Inequality (16) implies $\frac{d}{dt}\|x\|_N \leq Cst$, and then, by integration,

$$\|x(t)\|_N \leq Cst\, t + \|x(0)\|_N \leq Cst\, \tau.$$

The functions $\|x\|_N$ and $\|x\|_p$ being equivalent, we obtain, for a trajectory starting at x_0, $\|x(t)\|_p \leq Cst\, \tau$ when τ is small enough.

The second step is to prove $|x_j(t)-\widehat{x}_j(t)| \leq Cst\, \tau^{w_j} t$. The function $x_j-\widehat{x}_j$ satisfies the differential equation

$$\begin{aligned}
\dot{x}_j - \dot{\widehat{x}}_j &= \sum_{i=1}^m u_i\big(f_{ij}(x) - f_{ij}(\widehat{x}) + r_{ij}(x)\big), \\
&= \sum_{i=1}^m u_i\Big(\sum_{\{k\,:\,w_k<w_j\}} (x_k - \widehat{x}_k) Q_{ijk}(x, \widehat{x}) + r_{ij}(x)\Big),
\end{aligned}$$

where $Q_{ijk}(x, \widehat{x})$ is a homogeneous polynomial of weighted degree $w_j - w_k - 1$. For $\|x\|_p$ and $\|\widehat{x}\|_p$ small enough, we have

$$|r_{ij}(x)| \leq Cst\, \|x\|_p^{w_j} \quad \text{and} \quad |Q_{ijk}(x, \widehat{x})| \leq Cst\, (\|x\|_p + \|\widehat{x}\|_p)^{w_j-w_k-1}.$$

Using the inequalities of the first step, we obtain finally, for τ small enough,

$$|\dot{x}_j(t) - \dot{\widehat{x}}_j(t)| \leq Cst \sum_{\{k\,:\,w_k<w_j\}} |x_k(t) - \widehat{x}_k(t)| \tau^{w_j-w_k-1} + Cst\, \tau^{w_j}. \tag{17}$$

This system of inequalities has a triangular form, hence it can be integrated iteratively. For $w_j = 1$, the inequality is $|\dot{x}_j(t) - \dot{\widehat{x}}_j(t)| \leq Cst\, \tau$, and so $|x_j(t) - \widehat{x}_j(t)| \leq Cst\, \tau t$. By induction, let $j > n_1$ and assume $|x_k(t) - \widehat{x}_k(t)| \leq Cst\, \tau^{w_k} t$ for $k < j$. Inequality (17) implies

$$|\dot{x}_j(t) - \dot{\widehat{x}}_j(t)| \leq Cst\, \tau^{w_j-1} t + Cst\, \tau^{w_j} \leq Cst\, \tau^{w_j},$$

and so $|x_j(t) - \widehat{x}_j(t)| \leq Cst\, \tau^{w_j} t$.

Finally,

$$\|x(t) - \widehat{x}(t)\|_p \leq Cst\, \tau (t^{1/w_1} + \cdots + t^{1/w_n}) \leq Cst\, \tau t^{1/r},$$

which completes the proof of the lemma. □

Proof of Theorem 3.5. Observe first that, by definition of order, a system of coordinates z is privileged if and only if $d(p, q^z) \geq Cst\, \|z\|_p$. What remains to prove is that, if z are privileged coordinates, then $d(p, q^z) \leq Cst\, \|z\|_p$.

We will show that, for $\|x^0\|_p$ small enough,

$$d(0, x^0) \leq 2\widehat{d}(0, x^0),$$

and so $d(0, x^0) \leq Cst\, \|x^0\|_p$ by Lemma 3.4. This will prove Theorem 3.5.

Fix $x^0 \in \mathbb{R}^n$, $\|x^0\|_p < \varepsilon$. Let $\widehat{x}_0(t)$, $t \in [0, T_0]$, be a minimizing curve for $\widehat{d}$, having velocity one, and joining x^0 to 0. According to Remark 1.2, such a curve exists, and there exists a control $u_0(\cdot)$ associated with $\widehat{x}_0$ such that $\|u_0(t)\| = 1$ a.e. Moreover, $T_0 = \widehat{d}(0, x^0)$.

Let $x_0(t)$, $t \in [0, T_0]$, be the trajectory of the control system associated with $X_1, \ldots, X_m$ starting at x^0 and defined by $u_0(\cdot)$. We have $\text{length}\big(x_0(\cdot)\big) \leq T_0$. Set $x^1 = x_0(T_0)$. By Lemma 3.7,

$$\|x^1\|_p = \|x_0(T_0) - \widehat{x}_0(T_0)\|_p \leq C\tau T_0^{1/r},$$

where $\tau = \max(\|x^0\|_p, T_0)$. By Lemma 3.4, $T_0 = \widehat{d}(0, x^0)$ satisfies $T_0 \geq \|x^0\|_p / C'$, so $\tau \leq C' T_0$, and

$$\widehat{d}(0, x^1) \leq C' \|x^1\|_p \leq C'' \widehat{d}(0, x^0)^{1+1/r},$$

with $C'' = C'^2 C$.

Choose now $\widehat{x}_1(t)$, $t \in [0, T_1]$, a minimizing curve for $\widehat{d}$ of velocity one joining x^1 to 0. There exists a control $u_1(\cdot)$ associated with $\widehat{x}_1$ such that $\|u_1(t)\| = 1$ a.e. Let $x_1(t)$, $t \in [0, T_1]$, be the trajectory of the control system associated with $X_1, \ldots, X_m$ starting at x^1 and defined by $u_1(\cdot)$. Set $x^2 = x_1(T_1)$. As previously, we have $\text{length}\big(x_1(\cdot)\big) = \widehat{d}(0, x^1)$ and $\widehat{d}(0, x^2) \leq C'' \widehat{d}(0, x^1)^{1+1/r}$.

Repeating this construction, we obtain a sequence of points $x^0, x^1, x^2, \ldots$ such that $\widehat{d}(0, x^{k+1}) \leq C'' \widehat{d}(0, x^k)^{1+1/r}$, and a sequence of trajectories $x_k(\cdot)$ joining x^k to x^{k+1} of length equal to $\widehat{d}(0, x^k)$.

Taking $\|x^0\|_p$ small enough, we can assume $C'' \widehat{d}(0, x^0)^{1/r} \leq 1/2$. We have then $\widehat{d}(0, x^1) \leq \widehat{d}(0, x^0)/2$, $\ldots$, $\widehat{d}(0, x^k) \leq \widehat{d}(0, x^0)/2^k, \ldots$ Consequently, x^k tends to 0 as $k \to +\infty$. Putting end to end the curves $x_k(\cdot)$ gives a trajectory joining x^0 to 0 of length $\widehat{d}(0, x^0) + \widehat{d}(0, x^1) + \cdots \leq 2\widehat{d}(0, x^0)$. This implies $d(0, x^0) \leq 2\widehat{d}(0, x^0)$, and the proof is complete. □

Now, the distance $\widehat{d}$ on $\mathbb{R}^n$ induces a distance $\widehat{d}_p$ on a neighbourhood of p in M by setting $\widehat{d}_p(q, q') = \widehat{d}(z(q), z(q'))$. This distance gives the first-order term in the expansion of $d(p, \cdot)$.

Theorem 3.8. *On a neighbourhood of p in M there holds*

$$d(p,q) = \widehat{d}_p(p,q)\left(1 + O\left(\widehat{d}_p(p,q)\right)\right).$$

Remark 3.4. By Theorem 3.5 and Lemma 3.4, when $d(p,q)$ is small enough we get the estimate

$$\frac{1}{C}\widehat{d}_p(p,q) \leq d(p,q) \leq C\widehat{d}_p(p,q), \tag{18}$$

where C is some positive constant. Theorem 3.8 essentially states that this constant can be chosen arbitrarily close to 1.

Proof. Fix $\delta > 0$. We have to prove that there exists $\varepsilon > 0$ such that, if $d(p,q) < \varepsilon$, then

$$(1-\delta)\widehat{d}_p(p,q) \leq d(p,q) \leq (1+\delta)\widehat{d}_p(p,q). \tag{19}$$

Let q be a point in M. Setting $x^0 = q$ in the proof of Theorem 3.5 furnishes a trajectory joining q to 0 which length is equal to $\sum_{k=0}^{\infty}\widehat{d}(0,x^k)$, the points x^k being such that $\widehat{d}(0,x^{k+1}) \leq C''\widehat{d}(0,x^k)^{1+1/r}$.

From (18), there exists $\varepsilon > 0$ such that $d(p,q) < \varepsilon$ implies $C''\widehat{d}(0,x^0)^{1/r} \leq \delta/(1+\delta)$. In this case the trajectory from q to 0 is of length not greater than $(1+\delta)\widehat{d}(0,x^0)$ and we have

$$d(p,q) \leq (1+\delta)\widehat{d}_p(p,q).$$

To prove the other inequality in (19), we use the same argument but reverse the role of d and $\widehat{d}$. Let $x_0(t)$, $t \in [0,T_0]$, be a minimizing curve for d of velocity one joining x^0 to 0, and let $u_0(\cdot)$ be a control associated with $x_0(\cdot)$ such that $\|u_0(t)\| = 1$ a.e. We have $T_0 = d(0,x^0)$. Let $\widehat{x}_0(t)$, $t \in [0,T_0]$, be the trajectory of the control system associated with $\widehat{X}_1,\ldots,\widehat{X}_m$ starting at x^0 and defined by the control $u_0(\cdot)$. In particular, $\mathrm{length}\big(x_0(\cdot)\big) \leq T_0$.

Set $x^1 = x_0(T_0)$. By Lemma 3.7,

$$\|x^1\|_p = \|x_0(T_0) - \widehat{x}_0(T_0)\|_p \leq C\tau T_0^{1/r},$$

where $\tau = \max(\|x^0\|_p, T_0)$. Theorem 3.5 implies $\tau \leq C_p T_0$, and

$$d(0,x^1) \leq C_p\|x^1\|_p \leq C''d(0,x^0)^{1+1/r},$$

with $C'' = C_p^2 C$.

Repeating this construction gives a trajectory of $\widehat{X}_1,\ldots,\widehat{X}_m$ joining q to p whose length is equal to $\sum_{k=0}^{\infty} d(0,x^k)$, where $d(0,x^{k+1}) \leq C''d(0,x^k)^{1+1/r}$.

For $d(p,q)$ small enough, we have $C''d(0,x^0)^{1/r} \leq \delta/(1-\delta)$ and the trajectory from q to 0 is of length $\leq 1/(1-\delta)d(0,x^0)$, which leads to

$$\widehat{d}_p(p,q) \leq \frac{1}{(1-\delta)} d(p,q).$$

This completes the proof. □

3.5 Approximate motion planning

Given a control system (Σ), the *motion planning problem* is to steer (Σ) from an initial point to a final point. For nonholonomic systems, the exact problem is in general unsolvable. However methods exist for a particular class of systems, namely for nilpotent (or nilpotentizable) systems. It is then of interest to devise approximate motion planning techniques based on nilpotent approximations. These techniques are Newton type methods, the nilpotent approximation playing the role of the usual linearization.

Precisely, consider a nonholonomic control system

$$(\Sigma) \ : \qquad \dot{x} = \sum_{i=1}^{m} u_i X_i(x), \quad x \in \mathbb{R}^n,$$

and initial and final points a and b in $\mathbb{R}^n$. Denote by $\widehat{X}_1, \ldots, \widehat{X}_m$ a nilpotent approximation of $X_1, \ldots, X_m$ at b. The k-step of an approximate motion planning algorithm take the following form (x^k denotes the state of the system, x^0 being the initial point a):

1. compute a control $u(t)$, $t \in [0,T]$, steering the control system associated with $\widehat{X}_1, \ldots, \widehat{X}_m$ from x^k to b;
2. compute the trajectory $x(\cdot)$ of (Σ) with control $u(\cdot)$ starting from x^k;
3. set $x^{k+1} = x(T)$.

Is this algorithm convergent or, at least, locally convergent? The answer to the latter question is positive under an extra hypothesis on the control given in point 2 of the algorithm, namely,

(H) there exists a constant K such that, if x^k and b are close enough, then

$$\int_0^T \|u(t)\| dt \leq K \widehat{d}(b, x_k).$$

Note that a control corresponding to a minimizing curve for $\widehat{d}$ satisfies this condition. Other standards methods using Lie groups (like the one in [LS91]) or based on the triangular form of the homogeneous nilpotent approximation also satisfy this hypothesis.

The local convergence is then proved exactly in the same way than Theorem 3.5. We normalize first the control, so that $\|u(t)\| = 1$ a.e. Then from Lemmas 3.7 and 3.4, we have $\widehat{d}(b, x^{k+1}) \leq C'' T^{1+1/r}$, and using hypothesis (H), we obtain

$$\widehat{d}(b, x^{k+1}) \leq C'' K^{1+1/r} \widehat{d}(b, x_k)^{1+1/r}.$$

If a is close enough to b, we have, at each step of the algorithm, $\widehat{d}(b, x^{k+1}) \leq \widehat{d}(b, x_k)/2$, which proves the local convergence of the algorithm. In other words, *for each point $b \in M$, there exists a constant $\varepsilon_b > 0$ such that, if $d(a, b) < \varepsilon_b$, then the approximate motion planning algorithm steering the system from a to b converges.*

To obtain a globally convergent algorithm, a natural idea is to iterate the locally convergent one. This requires the construction of a finite sequence of intermediate goals $b_0 = a, b_1, \ldots, b_N = b$ such that $d(b_{i-1}, b_i) < \varepsilon_{b_i}$. However the constant ε_b depends on b and, as already noticed for Theorem 3.5, it is not possible to have a uniform nonzero constant near singular points. Thus this method may provide a globally convergent algorithm only when every point is regular.

4 Tangent structure to Carnot-Carathéodory spaces

Consider a manifold M endowed with a sub-Riemannian distance d on M. The so-defined metric space (M, d) is called a *Carnot-Carathéodory space*. The notion of first-order approximation introduced in the previous section has a metric interpretation and will allow us to describe the local structure of a Carnot-Carathéodory space.

4.1 Metric tangent space

In describing the tangent space to a manifold, we essentially look at smaller and smaller neighbourhoods of a given point, the manifold being fixed. Equivalently, we can look at a fixed neighbourhood, but expanding the manifold. As noticed by Gromov, this idea can be used to define a notion of tangent space for a general metric space.

If X is a metric space with distance d, we define $\lambda\mathsf{X}$, for $\lambda > 0$, to be the metric space with same underlying set as X and distance λd. A *pointed metric space* (X, x) is a metric space with a distinguished point x.

Loosely speaking, a metric tangent space to the metric space X at x is a pointed metric space $(C_x\mathsf{X}, y)$ such that

$$(C_x\mathsf{X}, y) = \lim_{\lambda\to+\infty} (\lambda\mathsf{X}, x).$$

Of course, for this definition to make sense, we have to define the limit of pointed metric spaces.

Let us first define the Gromov-Hausdorff distance between metric spaces. Recall that, in a metric space X, the Hausdorff distance H-dist(A, B) between two subsets A and B of X is the infimum of $\rho > 0$ such that any point of A is within a distance ρ of B and any point of B is within a distance ρ of A. The Gromov-Hausdorff distance GH-dist(X, Y) between two metric spaces X and Y is the infimum of Hausdorff distances H-dist$(i(\mathsf{X}), j(\mathsf{Y}))$ over all metric spaces Z and all isometric embeddings $i : \mathsf{X} \to \mathsf{Z}$, $j : \mathsf{Y} \to \mathsf{Z}$.

Thanks to Gromov-Hausdorff distance, one can define the notion of limit of a sequence of pointed metric spaces: (X_n, x_n) converge to (X, x) if, for any positive r,

$$\text{GH-dist}\big(B^{\mathsf{X}_n}(x_n, r), B^{\mathsf{X}}(x, r)\big) \to 0 \quad \text{as } n \to +\infty$$

where $B^{\mathsf{Y}}(y, r)$ is considered as a metric space, endowed with the distance of Y. Note that all pointed metric spaces isometric to (X, x) are also limit of (X_n, x_n). However the limit is unique up to an isometry provided the closed balls around the distinguished point are compact [BBI01, Sect. 7.4].

Finally, one says that $(\mathsf{X}_\lambda, x_\lambda)$ converge to (X, x) when $\lambda \to \infty$ if, for every sequence λ_n, $(\mathsf{X}_{\lambda_n}, x_{\lambda_n})$ converge to (X, x).

Definition 4.1. A pointed metric space $(C_x\mathsf{X}, y)$ is a *metric tangent space* to the metric space X at x if $(\lambda\mathsf{X}, x)$ converge to $(C_x\mathsf{X}, y)$ as $\lambda \to +\infty$. If it exists, it is unique up to an isometry provided the closed balls around x in $(\lambda\mathsf{X}, x)$ are compact.

For a Riemannian metric space (M, d_R) induced by a Riemannian metric g on a manifold M, metric tangent spaces at a point p exist and are isometric to the Euclidean space (T_pM, g_p), that is, the standard tangent space endowed with the scalar product defined by the quadratic form g_p.

For a Carnot-Carathéodory space (M, d), the metric tangent space is given by the nilpotent approximation.

Theorem 4.1. *A Carnot-Carathéodory space (M, d) admits metric tangent spaces (C_pM, y) at every point $p \in M$. The space C_pM is itself a Carnot-Carathéodory space isometric to $(\mathbb{R}^n, \widehat{d})$, where $\widehat{d}$ is the sub-Riemannian distance associated with a homogeneous nilpotent approximation at p.*

This theorem, due to Bellaïche, is a consequence of a strong version of Theorem 3.5: for q and q' in a neighbourhood of p,

$$|d(q, q') - \widehat{d}(q, q')| \leq Cst\, \widehat{d}(p, q) d(q, q')^{1/r}. \tag{20}$$

In these notes, we present neither the proof of this result, nor the one of Theorem 4.1, and we refer the reader to [Bel96].

Remark 4.1. Recall that $\widehat{d}$ is not intrinsic to the frame $(X_1, \ldots, X_m)$. Thus Theorem 4.1 does not provide an intrinsic characterization of the metric tangent space. Such characterizations exist for sub-Riemannian manifolds (M, D, g_R) in [MM00] and [FJ03], and the latter could easily be adapted to the case of a sub-Riemannian geometry associated with a nonholonomic system. However these characterizations are intrinsic to the differentiable manifold M equipped with the sub-Riemannian structure (D, g_R), or to M equipped with the frame $(X_1, \ldots, X_m)$, not to the metric space (M, d). To our knowledge, the problem of finding a definition of the metric tangent space C_pM depending only on the Carnot-Carathéodory space (M, d) is still open.

The question we want to address now is: what is the algebraic structure of C_pM? Of course C_pM is not a linear space in general: for instance, $\widehat{d}$ is homogeneous of degree 1 but with respect to dilations δ_t but not with respect to the usual Euclidean dilations. We will see that C_pM has a natural structure of group, or at least of quotient of groups.

Denote by G_p the group generated by the diffeomorphisms $\exp(t\widehat{X}_i)$ acting on the left on $\mathbb{R}^n$. Since $\mathfrak{g}_p = Lie(\widehat{X}_1, \ldots, \widehat{X}_m)$ is a nilpotent Lie algebra, $G_p = \exp(\mathfrak{g}_p)$ is a simply connected group, having $\mathfrak{g}_p$ as its Lie algebra.

This Lie algebra $\mathfrak{g}_p$ splits into homogeneous components

$$\mathfrak{g}_p = \mathfrak{g}^{-1} \oplus \cdots \oplus \mathfrak{g}^{-r},$$

where $\mathfrak{g}^{-s}$ is the set of homogeneous vector fields of degree $-s$, and so $\mathfrak{g}_p$ is a graded Lie algebra. The first component $\mathfrak{g}^{-1} = \mathrm{span}\langle \widehat{X}_1, \ldots, \widehat{X}_m \rangle$ generates $\mathfrak{g}_p$ as a Lie algebra. All these properties imply that G_p is what we call a Carnot group.

Definition 4.2. A *Carnot group* is a simply connected Lie group, such that the associated Lie algebra is graded, nilpotent, and generated by its first component.

Note that the dilations δ_t act on $\mathfrak{g}_p$ as a multiplication by t^{-s} on $\mathfrak{g}^{-s}$. This action extends to G_p by the exponential mapping.

Example 4.1 (Heisenberg group). The simplest non Abelian Carnot group is the *Heisenberg group* $\mathbb{H}^3$ which is the connected and simply connected Lie group whose Lie algebra satisfies

$$\mathfrak{g} = \mathfrak{g}^{-1} \oplus \mathfrak{g}^{-2}, \qquad \text{with } \dim \mathfrak{g}^{-1} = 2.$$

As a consequence, $\dim \mathbb{H}^3 = 3$. Choosing a basis X, Y, $Z = [X, Y]$ of $\mathfrak{g}$, we define coordinates on $\mathbb{H}^3$ by the exponential mapping

$$(x, y, z) \mapsto \exp(xX + yY + zZ).$$

By the Campbell-Hausdorff formula (see Section A.1), the law group on $\mathbb{H}^3$ in these coordinates is

$$(x, y, z)\dot{(}x', y', z') = (x + x', y + y', z + z' + \frac{1}{2}(xy' - x'y)),$$

which is homogeneous with respect to the dilation $\delta_t(x, y, z) = (tx, ty, t^2 z)$.

Finally, denote by X_1, X_2 the left-invariant vector fields on $\mathbb{H}^3$ whose values at the identity are respectively X and Y. In coordinates (x, y, z), these vector fields write as

$$X_1 = \partial_x - \frac{y}{2}\partial_z \quad \text{and} \quad X_2 = \partial_y + \frac{x}{2}\partial_z,$$

which are the vector fields of what we have called the Heisenberg case in examples 3.3 and 3.4.

Let $\widehat{\xi}_1, \ldots, \widehat{\xi}_m$ be the right-invariant vector fields on G_p such that $\widehat{\xi}_i(\mathrm{id}) = \widehat{X}_i$, where id is the identity of G_p. Equivalently,

$$\widehat{\xi}_i(g) = \frac{d}{dt}\big[\exp(t\widehat{X}_i)g\big]\big|_{t=0}.$$

With $(\widehat{\xi}_1, \ldots, \widehat{\xi}_m)$ is associated a right-invariant sub-Riemannian metric and a sub-Riemannian distance d_{G_p} on G_p.

The action of G_p on $\mathbb{R}^n$ is smooth and transitive. Indeed, for every $x \in \mathbb{R}^n$, the orbit of x under the action of G_p is the set

$$\left\{\exp(t_{i_1}\widehat{X}_{i_1}) \circ \cdots \circ \exp(t_{i_k}\widehat{X}_{i_k})(x) \ : \ k \in \mathbb{N},\ t_{i_j} \in \mathbb{R},\ i_j \in \{1, \ldots, m\}\right\}.$$

By Chow-Rashevsky's theorem (or more precisely from Remark 2.3), this set is the whole $\mathbb{R}^n$ since $(\widehat{X}_1, \ldots, \widehat{X}_m)$ satisfies Chow's Condition on $\mathbb{R}^n$ (Lemma 3.4).

To understand the algebraic structure of $C_p M$ we will use the following standard result on transitive action of Lie groups (see for instance [Lee03, Th. 9.24]).

Theorem 4.2. *Let G be a Lie group acting on the left smoothly and transitively on a manifold M. Let $q \in M$ and H be the isotropy subgroup of q which is defined by $H = \{g \in G \ : \ g \cdot q = q\}$. Then H is a closed subgroup of G, the left coset space G/H is a manifold of dimension $\dim G - \dim H$, and the map $F : G/H \to M$ defined by $F(gH) = g \cdot q$ is an equivariant diffeomorphism.*

Let H_p be the isotropy subgroup of $0 \in \mathbb{R}^n$ under the action of G_p. According to Theorem 4.2, the map $\phi_p : G_p \to \mathbb{R}^n$, $\phi_p(g) = g(0)$, induces a diffeomorphism

$$\psi_p : G_p/H_p \to \mathbb{R}^n, \qquad \psi_p(gH_p) = g(0).$$

Observe that H_p is invariant under dilations, since $\delta_t g(\delta_t x) = \delta_t(g(x))$. Hence H_p is connected and simply connected, and so $H_p = \exp(\mathfrak{h}_p)$, where $\mathfrak{h}_p$ is the Lie sub-algebra of $\mathfrak{g}_p$ containing the vector fields vanishing at 0,

$$\mathfrak{h}_p = \{Z \in \mathfrak{g}_p \ : \ Z(0) = 0\}.$$

As $\mathfrak{g}_p$, $\mathfrak{h}_p$ is invariant under dilations and splits into homogeneous components.

Now, the elements $\widehat{X}_1, \dots, \widehat{X}_m$ of $\mathfrak{g}_p$ act on the left on G_p/H_p with the notation $\overline{\xi}_1, \dots, \overline{\xi}_m$,

$$\overline{\xi}_i(gH_p) = \frac{d}{dt}\big[\exp(t\widehat{X}_i)gH_p\big]\big|_{t=0}.$$

These vector fields define a sub-Riemannian metric and a sub-Riemannian distance $\overline{d}$ on G_p/H_p. We clearly have $\psi_{p*}\overline{\xi}_i = \widehat{X}_i$, so ψ_p maps the sub-Riemannian metric on G_p/H_p associated with $(\overline{\xi}_1, \dots, \overline{\xi}_m)$ to the one on $\mathbb{R}^n$ associated with $(\widehat{X}_1, \dots, \widehat{X}_m)$.

Theorem 4.3. *The metric tangent space C_pM and $(\mathbb{R}^n, \widehat{d})$ are isometric to the coset space G_p/H_p endowed with the sub-Riemannian distance $\overline{d}$.*

Example 4.2 (Grušin plane). Consider the vector fields $X_1 = \partial_x$ and $X_2 = x\partial_y$ on $\mathbb{R}^2$. The Carnot-Carathéodory space defined by these vector fields is called the *Grušin plane*.

The only nonzero bracket is $X_{12} = \partial_y$. Thus, at $p = 0$, the weights are $(1, 2)$, and (x, y) are privileged coordinates. Since X_1 and X_2 are homogeneous with respect to this system of coordinates, we have $\widehat{X}_1 = X_1$ and $\widehat{X}_2 = X_2$. The Lie algebra they generate is

$$\mathfrak{g}_0 = \operatorname{span}(X_1, X_2, X_{12})$$

which is of dimension 3, and the group $\exp(\mathfrak{g}_0)$ is actually the *Heisenberg group* $\mathbb{H}^3$ (see example 4.1). The Lie sub-algebra $\mathfrak{h}_0$ of $\mathfrak{g}_0$ containing the vector fields vanishing at 0 is

$$\mathfrak{h}_0 = \operatorname{span}(X_2),$$

which is one-dimensional. Thus the Grušin plane is isometric to $\mathbb{H}^3/\exp(\mathfrak{h}_0)$ endowed with the distance $\overline{d}$.

Example 4.3 (Martinet case). Consider the Martinet case, defined on $\mathbb{R}^3$ by

$$X_1 = \partial_x \quad \text{and} \quad X_2 = \partial_y + \frac{x^2}{2}\partial_z.$$

As noticed in example 3.5, at $p = 0$, the coordinates (x, y, z) are privileged and, by homogeneity, $\widehat{X}_1 = X_1$ and $\widehat{X}_2 = X_2$. Moreover the only nonzero bracket are $X_{12} = x\partial_z$ and $X_{112} = \partial_z$. Thus,

$$\mathfrak{g}_0 = \operatorname{span}(X_1, X_2, X_{12}, X_{112})$$

which is of dimension 4. The group $\exp(\mathfrak{g}_0)$ is called the *Engel group*, and is denoted by $\mathbb{E}^4$. The Lie sub-algebra $\mathfrak{h}_0$ of $\mathfrak{g}_0$ containing the vector fields vanishing at 0 is

$$\mathfrak{h}_0 = \operatorname{span}(X_{12}).$$

When the point p is regular, this theorem can be refined thanks to the following result.

Lemma 4.4. *If p is a regular point, then* $\dim G_p = n$.

Proof. Let $X_{I_1}, \ldots, X_{I_n}$ be an adapted frame at p. Due to the regularity of p, $X_{I_1}, \ldots, X_{I_n}$ is also an adapted frame near p, so any bracket X_J can be written as

$$X_J(z) = \sum_{\{i\,:\,|I_i|\leq|J|\}} a_i(z) X_{I_i}(z),$$

where each a_i is a function of order $\geq |I_i| - |J|$. Taking the homogeneous terms of degree $-|J|$ in this expression, we obtain

$$\widehat{X}_J(z) = \sum_{\{i\,:\,|I_i|=|J|\}} a_i(0) \widehat{X}_{I_i}(z),$$

and so $\widehat{X}_J \in \operatorname{span}\langle \widehat{X}_{I_1}, \ldots, \widehat{X}_{I_n} \rangle$. Thus $\widehat{X}_{I_1}, \ldots, \widehat{X}_{I_n}$ is a basis of $\mathfrak{g}_p$, and so $\dim G_p = n$. □

As a consequence H_p is of dimension zero. Since H_p is invariant under dilations, $H_p = \{\mathrm{id}\}$, and hence the mapping $\phi_p : G_p \to \mathbb{R}^n$, $\phi_p(g) = g(0)$, is a diffeomorphism. Moreover $\phi_{p*}\widehat{\xi}_i = \widehat{X}_i$, which implies that ϕ_p maps the sub-Riemannian metric on G_p associated with $(\widehat{\xi}_1, \ldots, \widehat{\xi}_m)$ to the one on $\mathbb{R}^n$ associated with $(\widehat{X}_1, \ldots, \widehat{X}_m)$. This gives the following result*.

Proposition 4.5. *When p is a regular point, the metric tangent space C_pM and the Carnot-Carathéodory space $(\mathbb{R}^n, \widehat{d})$ are isometric to the Carnot group G_p endowed with the right-invariant sub-Riemannian distance d_{G_p}.*

Thus Carnot groups have the same role in sub-Riemannian geometry as Euclidean spaces have in Riemannian geometry. For this reason they are sometimes referred to as "non Abelian linear spaces": the internal operation – addition – is replaced by the law group and the external operation – multiplication by a real number – by the dilations. Note that, when G_p is Abelian (i.e. commutative) then G_p has a linear structure and the sub-Riemannian metric on G_p is a Euclidean metric.

Example 4.4 (unicycle). In the case of the distance d associated with the unicycle (examples 1.1 and 3.8), the growth vector is $(2, 3)$ at every point. Hence every point $p \in \mathbb{R}^2 \times S^1$ is regular, and the Lie algebra generated by the nilpotent approximation satisfies

$$\mathfrak{g}_p = \mathfrak{g}^{-1} \oplus \mathfrak{g}^{-2}, \qquad \text{with } \dim \mathfrak{g}^{-1} = 2.$$

As a consequence, $G_p = \mathbb{H}^3$ (see example 4.1), and so the metric tangent space to $(\mathbb{R}^2 \times S^1, d)$ at every point p has the structure of the Heisenberg group.

4.2 Desingularization and uniform distance estimate

We have already highlighted in Remark 3.3 that singular points may cause difficulties, in particular because of the loss of uniformity of distance estimates. Therefore it is necessary to study carefully the behaviour of the distance at such points. We proceed as it is usual for singularities, that is, we consider a singularity as the projection of a regular object. To this aim we exploit the algebraic structure of the metric tangent space, which provides a good way of lifting and projecting Carnot-Carathéodory spaces.

Let us begin with nilpotent approximations. We keep the notations and definitions of the preceding subsection. At a singular point p, we have the

*This result appeared first in [Mit85], but with an erroneous proof. The presentation given here is inspired from the one of [Bel96].

following diagram,

$$\begin{array}{ccc} & (G_p, d_{G_p}) & \\ & \pi\downarrow \qquad \phi_p\searrow & \\ (G_p/H_p, \overline{d}) & \overset{\psi_p}{\longrightarrow} & (\mathbb{R}^n, \widehat{d}) \end{array}$$

Since the sub-Riemannian metric on G_p is a right-invariant, every point in the space (G_p, d_{G_p}) is regular.

Definition 4.3. A Carnot-Carathéodory space (M, d) is said to be *equiregular* if every point in M is regular.

Thus $(\mathbb{R}^n, \widehat{d})$ is the projection of the equiregular space (G_p, d_{G_p}). Recall now that $\widehat{\xi}_1, \dots, \widehat{\xi}_m$ (resp. $\overline{\xi}_1, \dots, \overline{\xi}_m$) are mapped to $\widehat{X}_1, \dots, \widehat{X}_m$ by ϕ_p (resp. ψ_p). Working in a system of coordinates, we identify G_p/H_p with $\mathbb{R}^n$ and $\overline{\xi}_i$ with $\widehat{X}_i$. These coordinates on $\mathbb{R}^n \simeq G_p/H_p$, denoted by x, induce coordinates $(x, z) \in \mathbb{R}^N$ on G_p for which we have

$$\widehat{\xi}_i(x, z) = \widehat{X}_i(x) + \sum_{j=n+1}^{N} b_{ij}(x, z)\partial_{z_j}. \tag{21}$$

Let $(x(\cdot), z(\cdot))$ be a trajectory of the nonholonomic system in G_p defined by $\widehat{\xi}_1, \dots, \widehat{\xi}_m$. Then, for every control $u(\cdot)$ associated with the trajectory,

$$(\dot{x}(t), \dot{z}(t)) = \sum_{i=1}^{m} u_i(t)\widehat{\xi}_i(x, z),$$

It follows from (21) that $x(\cdot)$ is a trajectory in $\mathbb{R}^n$ of the system defined by $\widehat{\xi}_1, \dots, \widehat{\xi}_m$, which is associated with the same controls $u(\cdot)$, so that

$$\operatorname{length}\big(x(\cdot)\big) = \operatorname{length}\big((x, z)(\cdot)\big).$$

Thus $\widehat{d}$ can be obtained from the sub-Riemannian distance d_{G_p} in G_p by

$$\widehat{d}(q_1, q_2) = \inf_{\widetilde{q}_2 \in q_2 H_p} d_{G_p}(\widetilde{q}_1, \widetilde{q}_2), \qquad \text{for any } \widetilde{q}_1 \in q_1 H_p,$$

or, equivalently, $B^{\widehat{d}}(q_1, \varepsilon) = \phi_p\big(B^{d_{G_p}}(\widetilde{q}_1, \varepsilon)\big)$.

We will use this idea to desingularize the original space (M, d). Choose for x privileged coordinates at p, so that

$$X_i(x) = \widehat{X}_i(x) + R_i(x) \qquad \text{with } \operatorname{ord}_p R_i \geq 0.$$

Set $\widetilde{M} = M \times \mathbb{R}^{N-n}$, and in local coordinates (x, z) on $\widetilde{M}$, define vector fields on a neighbourhood of $(p, 0)$ by

$$\xi_i(x, z) = X_i(x) + \sum_{j=n+1}^{N} b_{ij}(x, z)\partial_{z_j},$$

with the same functions b_{ij} as in (21). Such vector fields are called a *lifting* of the vector fields $X_1, \ldots, X_m$: denoting by $\pi : \widetilde{M} \to M$ the canonical projection, we have $X_i = \pi_* \xi_i$ for $i = 1, \ldots, m$, that is, X_i is the projection of ξ_i.

We define in this way a nonholonomic system on an open set $\widetilde{U} \subset \widetilde{M}$ whose nilpotent approximation at $(p, 0)$ is $(\widehat{\xi_1}, \ldots, \widehat{\xi_m})$, by construction. Unfortunately, $(p, 0)$ can be itself a singular point. Indeed, a point can be singular for a system and regular for the nilpotent approximation taken at this point.

Example 4.5. Take the vector fields $X_1 = \partial_{x_1}$, $X_2 = \partial_{x_2} + x_1\partial_{x_3} + x_1^2\partial_{x_4}$ and $X_3 = \partial_{x_5} + x_1^{100}\partial_{x_4}$ on $\mathbb{R}^5$. The origin 0 is a singular point. However the nilpotent approximation at 0 is $\widehat{X}_1 = X_1$, $\widehat{X}_2 = X_2$, $\widehat{X}_3 = \partial_{x_5}$, for which 0 is not singular.

To avoid this difficulty, we take a group bigger than G_p, namely the free nilpotent group N_r of step r with m generators. N_r is a Carnot group and its Lie algebra $\mathfrak{n}_r$ is the free nilpotent Lie algebra of step r with m generators. The given of m generators $\alpha_1, \ldots, \alpha_m$ of $\mathfrak{n}_r$ define on N_r a right-invariant sub-Riemannian distance d_N.

The group N_r can be thought as a group of diffeomorphisms, and so it defines a left action on $\mathbb{R}^n$. Denoting by J the isotropy subgroup of 0 for this action, we obtain that $(\mathbb{R}^n, \widehat{d})$ is isometric to N_r/J endowed with the restriction of the distance d_N.

Reasoning as above, we are able to lift locally the vector fields $X_1, \ldots, X_m$ on M to vector fields on $M \times \mathbb{R}^{\widetilde{n}-n}$, $\widetilde{n} = \dim N_r$, having $\alpha_1, \ldots, \alpha_m$ for nilpotent approximation at $(p, 0)$. Moreover $(p, 0)$ is a regular point for the associated nonholonomic system in $M \times \mathbb{R}^{\widetilde{n}-n}$ since N_r is free up to step r. We obtain in this way a result of desingularization.

Lemma 4.6. *Let p be a point in M, r the degree of nonholonomy at p, $\widetilde{n} = \widetilde{n}(m, r)$ the dimension of the free Lie algebra of step r with m generators, and $\widetilde{M} = M \times \mathbb{R}^{\widetilde{n}-n}$. Then there exist a neighbourhood $\widetilde{U} \subset \widetilde{M}$ of $(p, 0)$, a neighbourhood $U \subset M$ of p with $U \times \{0\} \subset \widetilde{U}$, local coordinates (x, z) on $\widetilde{U}$,*

and smooth vector fields on $\widetilde{U}$*,*

$$\xi(x,z) = X_i(x) + \sum_{j=n+1}^{N} b_{ij}(x,z)\partial_{z_j}, \tag{22}$$

such that:

- $\xi_1, \dots, \xi_m$ *satisfy Chow's Condition and have* r *for degree of nonholonomy everywhere (so the Lie algebra they generate is free up to step* r*);*
- *every* $\widetilde{q}$ *in* $\widetilde{U}$ *is regular;*
- *denoting by* $\pi : \widetilde{M} \to M$ *the canonical projection, and by* $\widetilde{d}$ *the sub-Riemannian distance defined by* $\xi_1, \dots, \xi_m$ *on* $\widetilde{U}$*, we have* $\pi_* \xi_i = X_i$*, and for* $q \in U$ *and* $\varepsilon > 0$ *small enough,*

  $$B(q,\varepsilon) = \pi\Big(B^{\widetilde{d}}\big((q,0),\varepsilon\big)\Big),$$

  or, equivalently,

  $$d(q_1,q_2) = \inf_{\widetilde{q}_2 \in \pi^{-1}(q_2)} \widetilde{d}\big((q_1,0),\widetilde{q}_2\big).$$

Remark 4.2. The lemma still holds if we replace r by any integer greater than the degree of nonholonomy at p.

Thus any Carnot-Carathéodory space (M,d) is locally the projection of an equiregular Carnot-Carathéodory space $(\widetilde{M},\widetilde{d})$. This projection preserves the trajectories, the minimizers, and the distance.

Example 4.6 (Martinet case). Consider the vector fields of the Martinet case (see example 3.5), defined on $\mathbb{R}^3$ by:

$$X_1 = \partial_x \quad \text{and} \quad X_2 = \partial_y + \frac{x^2}{2}\partial_z.$$

Let $\pi : \mathbb{R}^4 \to \mathbb{R}^3$ be the projection with respect to the last coordinates, $\pi(x,y,z,w) = (x,y,z)$. Then X_1 and X_2 are the projections of the vector fields defining the Engel group $\mathbb{E}^4$ (see example 4.3),

$$\xi_1 = \partial_x \quad \text{and} \quad \xi_2 = \partial_y + \frac{x^2}{2}\partial_z + x\partial_w,$$

that is $\pi_* \xi_i = X_i$. Thus, for every pair of points $q_1, q_2 \in \mathbb{R}^3$,

$$d_{\mathrm{Mart}}(q_1,q_2) = \inf_{w \in \mathbb{R}} d_{\mathbb{E}^4}\big((q_1,0),(q_2,w)\big),$$

where d_{Mart} and $d_{\mathbb{E}^4}$ are the sub-Riemannian distance in respectively the Martinet space and the Engel group.

Example 4.7 (Grušin plane). Consider the vector fields

$$X_1 = \partial_x, \qquad X_2 = x\partial_y,$$

on $\mathbb{R}^2$, which define the Grušin plane (see example 4.2). Let $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ be the projection with respect to the last coordinates, $\pi(x, y, z) = (x, y)$. Then $X_1 = \pi_* \xi_1$ and $X_2 = \pi_* \xi_2$, where

$$\xi_1 = \partial_x \quad \text{and} \quad \xi_2 = \partial_z + x\partial_z,$$

are, up to a change of coordinates, the vector fields defining the Heisenberg case (see example 3.3).

Application: uniform Ball-Box theorem

The key feature of regular points is uniformity:

- uniformity of the flag (10);
- uniformity w.r.t. p of the convergence $(\lambda(M, d), p) \to C_p M$ (as explained by Bellaïche [Bel96, Sect. 8], this uniformity is responsible for the group structure of the metric tangent space);
- uniformity of distance estimates (see Remark 3.3).

In particular the last property is essential to compute Hausdorff dimensions (see Section 4.3) or to prove the global convergence of approximate motion planning algorithms. Recall what we mean by uniformity in this context: in a neighbourhood of a regular point p_0, we can construct privileged coordinates depending continuously on the base point p and such that the distance estimate (15) holds with C_p and ε_p independent of p.

As already noticed, all these uniformity properties are lost at singular points. However, using the desingularization of a sub-Riemannian manifold, we are able to give a uniform version of distance estimates.

Let $\Omega \subset M$ be a compact set. We denote by $r_{\max}$ the maximum of degrees of nonholonomy at points in Ω. As noticed in Section 3.2, $r_{\max}$ is finite. We assume that M is an oriented manifold, so that the determinant det is well-defined (see [AM78, Def. 2.5.18]).

Let $\mathfrak{X}$ be the set of n-tuples $\mathbf{X} = (X_{I_1}, \ldots, X_{I_n})$ of brackets of length $|I_i| \leq r_{\max}$. Since $r_{\max}$ is finite, $\mathfrak{X}$ is a finite subset of $Lie(X_1, \ldots, X_m)^n$. Given $q \in \Omega$ and $\varepsilon > 0$, we define a function $f_{q,\varepsilon} : \mathfrak{X} \to \mathbb{R}$ by

$$f_{q,\varepsilon}(\mathbf{X}) = \left|\det\left(X_{I_1}(q)\varepsilon^{|I_1|}, \ldots, X_{I_n}(q)\varepsilon^{|I_n|}\right)\right|.$$

We say that $\mathbf{X} \in \mathfrak{X}$ is an *adapted frame at* (q, ε) if it achieves the maximum of $f_{q,\varepsilon}$ on $\mathfrak{X}$.

The values at q of an adapted frame at (q, ε) clearly form a basis of $T_q M$. Moreover, q being fixed, the adapted frames at (q, ε) are adapted frames at q for ε small enough.

Theorem 4.7 (Uniform Ball-Box theorem)**.** *There exist positive constants K and ε_0 such that, for $q \in \Omega$, $\varepsilon < \varepsilon_0$, and any adapted frame $\mathbf{X}$ at (q, ε), there holds*

$$\mathrm{Box}_{\mathbf{X}}(q, \frac{1}{K}\varepsilon) \subset B(q, \varepsilon) \subset \mathrm{Box}_{\mathbf{X}}(q, K\varepsilon),$$

where $\mathrm{Box}_{\mathbf{X}}(q, \varepsilon) = \{\exp(x_1 X_{I_1}) \circ \cdots \circ \exp(x_n X_{I_n})(q) : |x_i| \leq \varepsilon^{|I_i|}, 1 \leq i \leq n\}$.

Of course, q being fixed, this estimate is equivalent to the one of the Ball-Box theorem for ε smaller than some $\varepsilon_1(q) > 0$. However, the main difference is that here ε_0 does not depend on q, whereas in the Ball-Box theorem $\varepsilon_1 = \varepsilon_1(q)$ which can be infinitely close to 0 as q varies.

Let vol be any Riemannian volume on the manifold M. As a direct consequence of the Uniform Ball-Box theorem, we have an estimate of the volume of a small sub-Riemannian ball.

Corollary 4.8. *There exist positive constants K and ε_0 such that, for all $q \in \Omega$ and $\varepsilon < \varepsilon_0$,*

$$\frac{1}{K} \max_{\mathbf{X}} f_{q,\varepsilon}(\mathbf{X}) \leq \mathrm{vol}(B(q, \varepsilon)) \leq K \max_{\mathbf{X}} f_{q,\varepsilon}(\mathbf{X}),$$

the maximum of $f_{q,\varepsilon}(\mathbf{X}) = \left|\det\left(X_{I_1}(q)\varepsilon^{|I_1|}, \ldots, X_{I_n}(q)\varepsilon^{|I_n|}\right)\right|$ being taken over all families $\mathbf{X} = (X_{I_1}, \ldots, X_{I_n})$ of brackets of length $|I_i| \leq r_{\max}$.

If moreover all points in Ω are regular, then for all $q \in \Omega$ and $\varepsilon < \varepsilon_0$,

$$\frac{1}{K}\varepsilon^Q \leq \mathrm{vol}(B(q, \varepsilon)) \leq K\varepsilon^Q, \tag{23}$$

where $Q = \sum_{i=1}^n w_i(q)$ does not depend on q.

4.3 Hausdorff dimension

Consider a metric space (M, d) and denote by $\operatorname{diam} S$ the diameter of a set $S \subset M$. Let $k \geq 0$ be a real number. For every subset $A \subset M$, we define the *k-dimensional Hausdorff measure* $\mathcal{H}^k$ of A as $\mathcal{H}^k(A) = \lim_{\varepsilon \to 0^+} \mathcal{H}^k_\varepsilon(A)$, where

$$\mathcal{H}^k_\varepsilon(A) = \inf\left\{\sum_{i=1}^{\infty} (\operatorname{diam} S_i)^k \,:\, A \subset \bigcup_{i=1}^{\infty} S_i,\ S_i \text{ closed set},\ \operatorname{diam} S_i \leq \varepsilon\right\},$$

and the *k-dimensional spherical Hausdorff measure* $\mathcal{S}^k$ of A as $\mathcal{S}^k(A) = \lim_{\varepsilon\to 0^+} \mathcal{S}^k_\varepsilon(A)$, where

$$\mathcal{S}^k_\varepsilon(A) = \inf\left\{\sum_{i=1}^{\infty} (\operatorname{diam} S_i)^k \,:\, A \subset \bigcup_{i=1}^{\infty} S_i,\ S_i \text{ is a ball, } \operatorname{diam} S_i \le \varepsilon\right\}.$$

In the Euclidean space $\mathbb{R}^n$, k-dimensional Hausdorff measures are often defined as $2^{-k}\alpha(k)\mathcal{H}^k$ and $2^{-k}\alpha(k)\mathcal{S}^k$, where $\alpha(k)$ is defined from the usual gamma function as $\alpha(k) = \Gamma(\frac{1}{2})^k/\Gamma(\frac{k}{2}+1)$. This normalization factor is necessary for the n-dimensional Hausdorff measure and the Lebesgue measure to coincide on $\mathbb{R}^n$.

For a given set $A \subset M$, $\mathcal{H}^k(A)$ is a decreasing function of k, infinite when k is smaller than a certain value, and zero when k is greater than this value. We call *Hausdorff dimension* of A the real number

$$\dim_{\mathcal{H}} A = \sup\{k \,:\, \mathcal{H}^k(A) = \infty\} = \inf\{k \,:\, \mathcal{H}^k(A) = 0\}.$$

Note that $\mathcal{H}^k \le \mathcal{S}^k \le 2^k\mathcal{H}^k$, so the Hausdorff dimension can be defined equivalently from Hausdorff or spherical Hausdorff measures.

There exist only few results on Hausdorff measures in sub-Riemannian geometry, except for specific cases [ABB11, GJ12]. The most general result is the following one.

Theorem 4.9. *Let (M, d) be an equiregular Carnot-Carathéodory space and p a point in M. Then the Hausdorff dimension of a small enough ball $B(p, r)$ is $\dim_{\mathcal{H}} B(p, r) = Q$, where*

$$Q = \sum_{i=1}^{n} w_i(p) = \sum_{i\ge 1} i\left(\dim \Delta^i(p) - \dim \Delta^{i-1}(p)\right)$$

does not depend on p. Moreover $\mathcal{H}^Q(B(p, r))$ is finite.

Proof. Fix a Riemannian volume vol on $B(p, r)$ (it is possible for a small enough r). It results from Corollary 4.8 that, for $q \in B(p, r)$ and ε small enough,

$$\frac{1}{K}\varepsilon^Q \le \operatorname{vol}(B(q, \varepsilon)) \le K\varepsilon^Q. \tag{24}$$

Define N_ε to be the maximal number of disjoints balls of radius ε included in $B(p, r)$, and consider such a family $B(q_i, \varepsilon)$, $i =, \dots, N_\varepsilon$, of disjoints balls. By (24),

$$\frac{1}{K}\varepsilon^Q N_\varepsilon \le \operatorname{vol}(B(p, r)) \quad \Rightarrow \quad N_\varepsilon \le K\varepsilon^{-Q}\operatorname{vol}(B(p, r)).$$

On the other hand the union $\bigcup_i B(q_i, 2\varepsilon)$ covers $B(p,r)$, and by Theorem 4.7 every ball $B(q_i, 2\varepsilon)$ is of diameter $\geq \frac{4}{K}\varepsilon$ if ε is small enough. This implies

$$\mathcal{S}^Q(B(p,r)) \leq \liminf_{\varepsilon \to 0} N_\varepsilon \left(\frac{4\varepsilon}{K}\right)^Q < \infty.$$

Therefore $\dim_{\mathcal{H}} B(p,r) \leq Q$.

Conversely, let $\bigcup_i B(q_i, r_i)$ be a covering of $B(p,r)$ with balls of diameter not greater than ε. If ε is small enough, every r_i is smaller than ε_0 and there holds

$$\mathrm{vol}(B(p,r)) \leq \sum_i \mathrm{vol}(B(q_i, r_i)) \leq K \sum_i r_i^Q.$$

As a consequence, we have $\mathcal{S}^Q(B(p,r)) \geq \mathrm{vol}(B(p,r))/K$, which in turn implies $\dim_{\mathcal{H}} B(p,r) \geq Q$. This ends the proof. □

When (M,d) is not equiregular, the Hausdorff dimension of balls centered at singular points behaves in a different way. Let us show it on an example.

Consider the Martinet space (see Example 3.5), that is, $\mathbb{R}^3$ endowed with the sub-Riemannian distance associated with the vector fields

$$X_1 = \partial_x \quad \text{and} \quad X_2 = \partial_y + \frac{x^2}{2}\partial_z.$$

A point $q = (x,y,z)$ is regular if $x \neq 0$ and in this case $\sum_i w_i(q) = 4$, otherwise it is singular and $\sum_i w_i(q) = 5$.

Lemma 4.10. *Let p a point in the Martinet space.*

- *If p is regular, then $\dim_{\mathcal{H}} B(p,r) = 4$, and $\mathcal{H}^4(B(p,r))$ is finite.*
- *If p is singular, then $\dim_{\mathcal{H}} B(p,r) = 4$, but $\mathcal{H}^4(B(p,r))$ is not finite.*

Proof. When p is regular, the result is a direct consequence of Theorem 4.9. Let us consider a singular point p and a radius $r > 0$. Since regular points form an open set, $B(p,r)$ contains small balls centered at regular points, and thus $\dim_{\mathcal{H}} B(p,r) \geq 4$. Moreover, it results from Corollary 4.8 that, for $q = (x,y,z)$ close enough from p and for $\varepsilon > 0$ small enough,

$$\frac{1}{K}\varepsilon^4 \max(|x|, \varepsilon) \leq \mathrm{vol}(B(q,\varepsilon)) \leq K\varepsilon^4 \max(|x|,\varepsilon). \tag{25}$$

We proceed as in the proof of Theorem 4.9. Define N_ε to be the maximal number of disjoints balls of radius ε included in $B(p,r)$, and consider such a family $B(q_i,\varepsilon)$, $i = , \ldots, N_\varepsilon$, of disjoints balls, with $q_i = (x_i, y_i, z_i)$. Notice

that the first coordinate x is of nonholonomic order ≤ 1 at any point; this implies that there exists a constant $K' > 0$ such that

$$B(q_i, \varepsilon) \subset B(p, r) \cap \{q = (x, y, z) \ : \ |x - x_i| \leq K'\varepsilon\}.$$

As a consequence, for an integer k, every ball $B(q_i, \varepsilon)$ such that $(k-1)\varepsilon \leq |x_i| < k\varepsilon$ is included in the set $B(p, r) \cap \{q = (x, y, z) \ : \ |x| \in ((k - 1 - K')\varepsilon, (k + K')\varepsilon]\}$. The volume of the latter set is smaller than $K''\varepsilon$, where K'' is a constant (depending neither on k nor ε). Then it results from (25) that

$$KN_\varepsilon(k)k\varepsilon^5 \leq K''\varepsilon,$$

where $N_\varepsilon(k)$ is the number of points q_i such that $(k-1)\varepsilon \leq |x_i| < k\varepsilon$. The Ball-Box Theorem implies that $N_\varepsilon(k) = 0$ when $k > K'r/\varepsilon$, and hence

$$N_\varepsilon = \sum_{k=1}^{\lceil K'r/\varepsilon \rceil} N_\varepsilon(k) \leq \frac{\text{const}}{\varepsilon^4} \sum_{k=1}^{\lceil K'r/\varepsilon \rceil} \frac{1}{k} \leq \frac{\text{const}}{\varepsilon^4} \log\left(\frac{1}{\varepsilon}\right),$$

where $\lceil t \rceil$ denotes the integer part of a number t. Now the union $\bigcup_i B(q_i, 2\varepsilon)$ covers $B(p, r)$ and every ball $B(q_i, 2\varepsilon)$ is of diameter $\geq \frac{4}{K}\varepsilon$ if ε is small enough. This implies that, for any real number $s > 4$,

$$\mathcal{S}^s(B(p, r)) \leq \lim_{\varepsilon \to 0} \left(\frac{4\varepsilon}{K}\right)^s N_\varepsilon \leq \lim_{\varepsilon \to 0} \text{const}\, \varepsilon^{s-4} \log\left(\frac{1}{\varepsilon}\right) = 0.$$

Consequently $\dim_{\mathcal{H}} B(p, r) \leq 4$, and hence $\dim_{\mathcal{H}} B(p, r) = 4$, since the converse inequality holds.

We are left to show that $\mathcal{H}^4(B(p, r))$, or equivalently $\mathcal{S}^4(B(p, r))$, is not finite. Let $\bigcup_i B(q_i, r_i)$ be a covering of $B(p, r)$ with balls of diameter not greater than ε. For an integer $k \geq 1$, denote by I_k the set of indices such that $\bigcup_{i \in I_k} B(q_i, r_i)$ is a covering of the set $B(p, r) \cap \{q = (x, y, z) \ : \ |x| \in ((k-1)\varepsilon, k\varepsilon]\}$. Thus

$$\sum_{i \in I_k} \text{vol}(B(q_i, r_i)) \geq \text{const}\, \varepsilon.$$

On the other hand $i \in I_k$ implies $\text{vol}(B(q_i, r_i)) \leq \text{const}\, r_i^4 k\varepsilon$, and so

$$\sum_{i \in I_k} r_i^4 \geq \frac{\text{const}}{k}.$$

Summing up over k, we obtain, for a small enough ε,

$$\sum_i r_i^4 \geq \text{const}\, \log\left(\frac{1}{\varepsilon}\right).$$

As a consequence, $\mathcal{S}^4_\varepsilon(B(p,r)) \geq \text{const } \log(\frac{1}{\varepsilon})$, and so $\mathcal{S}^4_\varepsilon(B(p,r)) = \infty$. This ends the proof. $\square$

Appendix

A Flows of vector fields

This section is dedicated to the proof of Campbell-Hausdorff type formulas for flows of vector fields. The result in Section A.1 has been used in Section 2.1, the one in Section A.2 will be necessary for the next section.

Let U be an open subset of $\mathbb{R}^n$ and $VF(U)$ the set of smooth vector fields on U. Given a vector field $X \in VF(U)$, we denote its flow by $\exp(tX)$.

A.1 Campbell-Hausdorff formula for flows

We will need in this section the Campbell-Hausdorff formula which we recall briefly here (for a more detailed presentation see for instance [Bou72, Ch. II]). Let x and y be two non commutative indeterminates, and $[x, y] = xy - yx$ their commutator, also denoted by $[x, y] = (\mathrm{ad}x)y$. The length of an iterated commutator $(\mathrm{ad}x_1) \cdots (\mathrm{ad}x_{k-1})x_k$, where each $x_1, \ldots, x_k$ equals x or y, is defined to be the number of occurrences k of x and y. Define also e^x and e^y to be the series $\sum_{k\geq 0} \frac{x^k}{k!}$ and $\sum_{k\geq 0} \frac{y^k}{k!}$. Then we have $e^x e^y = e^{H(x,y)}$ in the sense of formal power series, where

$$H(x, y) = x + y + \frac{1}{2}[x, y] + R(x, y), \tag{26}$$

and $R(x, y)$ is a series whose terms are linear combination of iterated commutators of x and y of length greater than 2. For an integer N we denote by $H_N(x, y)$ the partial sum of $H(x, y)$ containing only iterated commutators of length not greater than N. In particular, $H_1 = x+y$ and $H_2 = x+y+\frac{1}{2}[x, y]$.

Consider now two vector fields $X, Y \in VF(U)$. Given $t \in \mathbb{R}$ and an integer N, $H_N(tY, tX)$ is a smooth vector field on U which writes as $\sum_{i=1}^{N} t^i Y_i$, where the vector fields $Y_1, \ldots, Y_N$ belong to the Lie algebra generated by X and Y.

Lemma A.1. *Let $p \in M$. There exist positive constants δ and C such that $|t| < \delta$ implies*

$$\|\exp(tX) \circ \exp(tY)(p) - \exp(H_N(tY, tX))(p)\| \leq C|t|^{N+1}.$$

Proof. Set $\psi(t) = \exp(tX) \circ \exp(tY)(p)$, which is a function defined and C^∞ in a neighbourhood of $0 \in \mathbb{R}$, and let $(x_1, \ldots, x_n)$ be a system of local coordinates on a neighbourhood of p in U. We will compute the Taylor

expansion of every component $x_i(\psi(t))$, for $i = 1, \ldots, n$. To do this, we introduce the function $\phi(t,s) = \exp(tX) \circ \exp(sY)(p)$, so that $\psi(t) = \phi(t,t)$, and we compute the partial derivatives of $x_i \circ \phi$ at $0 \in \mathbb{R}^2$. We have:

$$\frac{\partial x_i \circ \phi}{\partial t}(t,s) = \frac{d}{dt}\left[x_i \circ \exp(tX)\right](\exp(sY)(p)) = Xx_i(\phi(t,s)),$$

where Xx_i denotes the Lie derivative of x_i along X. Repeating this computation, we obtain for any integer k,

$$\frac{\partial^k x_i \circ \phi}{\partial t^k}(t,s) = X^k x_i(\phi(t,s)).$$

In the same way, we have:

$$\begin{aligned}\frac{\partial^{k+l} x_i \circ \phi}{\partial s^l \partial t^k}(0,0) &= \frac{\partial^l}{\partial s^l}\frac{\partial^k x_i \circ \phi}{\partial t^k}(0,s)\big|_{s=0} \\ &= \frac{\partial^l}{\partial s^l}\left[X^k x_i(\exp(sY)(p))\right]\big|_{s=0} = Y^l X^k x_i(p).\end{aligned}$$

We then deduce that the formal Taylor series of $x_i(\psi(t)) = x_i(\phi(t,t))$ at 0 is

$$\sum_{k,l\geq 0} \frac{t^{k+l}}{k!l!} Y^l X^k x_i(p) = \left[\sum_{l\geq 0} \frac{t^l}{l!} Y^l\right]\left[\sum_{k\geq 0} \frac{t^k}{k!} X^k\right] x_i(p),$$

where X and Y are considered as derivation operators. From the Campbell-Hausdorff formula, the product of the formal series $e^{tY} = \sum_{l\geq 0} \frac{t^l}{l!} Y^l$ with $e^{tX} = \sum_{k\geq 0} \frac{t^k}{k!} X^k$ is equal to the series $e^{H(tY,tX)}$. As a consequence, the Taylor expansion of $x_i(\psi(t))$ up to degree N is given by the terms of degree $\leq N$ in the series $e^{H(tY,tX)} x_i(p)$, which coincide with the terms of degree $\leq N$ in the series $e^{H_N(tY,tX)} x_i(p)$.

On the other hand, it results from Lemma A.2 below that $e^{H_N(tY,tX)} x_i(p)$ is the Taylor series at 0 of the function $t \mapsto x_i \circ \exp(H_N(tY,tX))(p)$. Thus

$$x_i \circ \psi(t) - x_i \circ \exp(H_N(tY,tX))(p) = O(|t|^{N+1})$$

for every coordinate x_i, and the lemma follows. □

Lemma A.2. *Let $Y_1, \ldots, Y_\ell$ be vector fields on U, $f : U \to \mathbb{R}$ a smooth function, and $p \in U$. The formal Taylor series at $0 \in \mathbb{R}^\ell$ of the function $(z_1, \ldots, z_\ell) \mapsto f(\exp(\sum_i z_i Y_i)(p))$ is given by*

$$\sum_{k\geq 0} \frac{1}{k!}\left(\sum_i z_i Y_i\right)^k f(p) = e^{\sum_i z_i Y_i} f(p).$$

The formal Taylor series at $0 \in \mathbb{R}$ *of the function* $t \mapsto f(\exp(Y(t))(p))$, *where* $Y(t) = \sum_{i=1}^{\ell} t^i Y_i$, *is given by*

$$\sum_{k\geq 0} \frac{1}{k!} Y(t)^k f(p) = e^{Y(t)} f(p). \tag{27}$$

Proof. The second statement is obviously a consequence of the first one, so it is sufficient to prove the latter. We introduce the functions $g(z) = f(\exp(\sum_i z_i Y_i)(p))$ and $G(t,z) = g(tz)$ which are well-defined and smooth on a neighbourhood of 0 in $\mathbb{R}^\ell$, respectively $\mathbb{R} \times \mathbb{R}^\ell$. We are looking for the Taylor series of g at 0.

Since $G(t,z) = f(\exp(t \sum_i z_i Y_i)(p))$, we have, for any integer $k \geq 0$,

$$\frac{\partial^k G}{\partial t^k}(0,z) = (\sum_i z_i Y_i)^k f(p).$$

On the other hand $G(t,z) = g(tz)$, and hence the previous derivative can also be computed as

$$\frac{\partial^k G}{\partial t^k}(0,z) = \sum_{\alpha_1+\cdots+\alpha_\ell=k} \frac{k!}{\alpha_1! \cdots \alpha_\ell!} z_1^{\alpha_1} \cdots z_\ell^{\alpha_\ell} \frac{\partial^k g}{\partial z_1^{\alpha_1} \cdots \partial z_\ell^{\alpha_\ell}}(0).$$

Combining both expressions, we obtain

$$\sum_{\alpha_1+\cdots+\alpha_\ell=k} \frac{z_1^{\alpha_1} \cdots z_\ell^{\alpha_\ell}}{\alpha_1! \cdots \alpha_\ell!} \frac{\partial^k g}{\partial z_1^{\alpha_1} \cdots \partial z_\ell^{\alpha_\ell}}(0) = \frac{1}{k!} \left(\sum_i z_i Y_i \right)^k f(p),$$

and the lemma follows. □

Lemma A.1 can be extended in two ways. First, since the vector fields and their flows are smooth on U, the estimate holds uniformly with respect to p. Second, by Lemma A.2, the vector fields tX and tY may be replaced by the one-parameter families of vector fields $X(t) = tX_1 + \cdots + t^k X_k$ and $Y(t) = tY_1 + \cdots + t^\ell Y_\ell$, where $X_1, \ldots, X_k$ and $Y_1, \ldots, Y_\ell$ are vector fields on U. As an example, $H_N(tY, tX)$ is of this form. To summarize, a slight change in the proof of Lemma A.1 actually shows the following result.

Corollary A.3. *Let* $K \subset U$ *be a compact. There exist two positive constants* δ, C *such that, if* $p \in K$ *and* $|t| < \delta$, *then:*

$$\|\exp(X(t)) \circ \exp(Y(t))(p) - \exp(H_N(Y(t), Y(t)))(p)\| \leq C|t|^{N+1}.$$

We are now in a position to prove formula (6), that we used in the proof of Lemma 2.1. Let $X_1, \dots, X_m$ be m elements of $VF(U)$. For every multi-index $I \in \{1, \dots, m\}^k$, $k \in \mathbb{N}$, we define the local diffeomorphisms ϕ_t^I on U by induction on the length $|I|$ of I. Let $\phi_t^i = \exp(tX_i)$ and set, if $I = iJ$,

$$\phi_t^{iJ} = \phi_{-t}^J \circ \phi_{-t}^i \circ \phi_t^J \circ \phi_t^i.$$

Proposition A.4. *Let $K \subset U$ be a compact and I a multi-index. There exist two positive constants δ, C such that, if $p \in K$ and $|t| < \delta$, then*

$$\left\|\phi_t^I(p) - p - t^{|I|}X_I(p)\right\| \;\leq\; C|t|^{|I|+1}. \tag{28}$$

Proof. For $\delta > 0$ small enough, the mapping $(p,t) \mapsto \phi_t^I(p)$ is defined and C^∞ on $K \times (-\delta, \delta)$. As a consequence, we are reduced to prove (28) for a fixed $p \in K$.

When $|I| = 1$, $\phi_t^I(p) = \exp(tX_i)(p)$ for some $i \in \{1, \dots, m\}$, which is equal to $p + tX_i(p) + O(|t|^2)$.

Now, let I be a multi-index and $N > |I|$ an integer. Corollary A.3 implies that $\phi_t^I(p) = \exp(H_N^I(t))(p) + O(t^{N+1})$ where the series $H_N^I(t)$ is defined by induction: if $I = iJ$, then

$$H_N^{iJ}(t) = H_N(H_N(tX_i, H_N^J(t)), H_N(-tX_i, -H_N^J(t))).$$

Applying (26) iteratively, we can write $H_N^I(t) = t^{|I|}X_I + t^{|I|+1}R_I(t)$, the latter term being a one-parameter vector field. As a consequence,

$$\phi_t^I(p) = p + t^{|I|}X_I(p) + \text{terms of degree greater than } |I|+1,$$

which completes the proof. □

A.2 Push-forward formula

Given two vector fields $X, Y \in VF(U)$, we write $(\mathrm{ad}X)Y$ for $[X,Y]$, $(\mathrm{ad}X)^2Y$ for $(\mathrm{ad}X)((\mathrm{ad}X)Y)$, etc.

Proposition A.5. *Let $K \subset U$ be a compact, N a positive integer, and X, Y, $Y_1, \dots, Y_\ell$ vector fields in $VF(U)$. There exist two positive constants δ, C such that, if $p \in K$, $t \in \mathbb{R}$ and $z \in \mathbb{R}^\ell$ satisfy $|t| < \delta$ and $\|z\| < \delta$, then*

$$\left\|\exp(tY)_*X(p) - \textstyle\sum_{k=0}^N \frac{t^k}{k!}(\mathrm{ad}Y)^k X(p)\right\| \leq C|t|^{N+1},$$

$$\left\|\exp(\textstyle\sum_{i=1}^\ell z_iY_i)_*X(p) - \sum_{k=0}^N \frac{1}{k!}\left(\mathrm{ad}\sum_{i=1}^\ell z_iY_i\right)^k X(p)\right\| \leq C\|z\|^{N+1},$$

where $\exp(tY)_*X = d(\exp(tY)) \circ X \circ \exp(-tY)$ denotes the push-forward of the vector field X by the diffeomorphism $\exp(tY)$.

Proof. Let us begin with the first inequality. Set $\phi_p(t) = \exp(tY)_* X(p)$. For $\delta > 0$ small enough, the mapping $(p,t) \mapsto \phi_p(t)$ is defined and C^∞ on $K \times (-\delta, \delta)$. As a consequence, there exists a constant $C > 0$ such that, for every $p \in K$ and $|t| < \delta$, we have

$$\left\| \phi_p(t) - \sum_{k=0}^{N} \frac{t^k}{k!} \frac{d^k \phi_p}{dt^k}(0) \right\| \leq C|t|^{N+1}.$$

It remains to prove that $\frac{d^k \phi_p}{dt^k}(0) = (\mathrm{ad}Y)^k X(p)$ for any integer k. Note first that $\phi_p(0) = X(p)$, and that

$$\frac{d\phi_p}{dt}(0) = \frac{d}{dt} \left[\exp(tY)_* X \right] \big|_{t=0}(p)$$

is by definition equal to $-L_Y X(p)$, where $L_Y X$ is the Lie derivative of X along Y (see for instance [Boo86]). Since $L_Y X(p) = (\mathrm{ad}Y)X(p)$, the cases $k = 0$ and $k = 1$ are done.

We need now to compute $\frac{d\phi_p}{dt}(t)$ at $t \neq 0$. Let us write $\phi_p(t+s)$ as $\exp(tY)_* \exp(sY)_* X(p)$. We have

$$\begin{aligned} \frac{d\phi_p}{dt}(t) = \frac{d\phi_p(t+s)}{ds} \mid_{s=0} &= \exp(tY)_* \frac{d}{ds} \left[\exp(sY)_* X \right] \big|_{s=0}(p) \\ &= \exp(tY)_* ((\mathrm{ad}Y)X)(p). \end{aligned}$$

This derivative has the same form as $\phi_p(t)$, X being replaced by $(\mathrm{ad}Y)X$. Iterating the argument above, we obtain by induction $\frac{d^k \phi_p}{dt^k}(0) = (\mathrm{ad}Y)^k X(p)$, and the first inequality of the proposition is proved.

As for the second inequality, the same reasoning applies and we only need to compute the partial derivatives at $0 \in \mathbb{R}^\ell$ of the function $\widetilde{\phi}(z) = \exp(\sum_i z_i Y_i)_* X(p)$. This can be done as in the proof of Lemma A.2. The proposition follows. □

B Different systems of privileged coordinates

This appendix is devoted to the proof that the examples of coordinates introduced in Section 3.2 are actually privileged coordinates.

B.1 Canonical coordinates of the second kind

Let $p \in M$ and $Y_1, \ldots, Y_n$ an adapted frame at p (see (11), page 21). The map

$$\phi : (z_1, \ldots, z_n) \mapsto \exp(z_n Y_n) \circ \cdots \circ \exp(z_1 Y_1)(p)$$

is a local diffeomorphism near $0 \in \mathbb{R}^n$ and its inverse defines some coordinates called *canonical coordinates of the second kind near p*.

The following result is due to Hermes [Her91].

Lemma B.1. *Canonical coordinates of the second kind are privileged at p.*

For sake of simplicity, we will write the compositions of maps as products; for instance, we write

$$\phi(z) = \exp(z_1 Y_1) \cdots \exp(z_n Y_n)(p).$$

Proof. First, let us recall that ϕ is a local diffeomorphism at $z = 0$ because its differential at 0 is an isomorphism. This results from

$$\frac{\partial \phi}{\partial z_i}(0) = \frac{d}{dt}\left(\phi(0, \ldots, t, \ldots, 0)\right)\big|_{t=0} = \frac{d}{dt}\left(\exp(tY_i)(p)\right)\big|_{t=0} = Y_i(p),$$

for $i = 1, \ldots, n$. This computation also reads as $\phi_* \frac{\partial}{\partial z_i}(p) = Y_i(p)$, which implies $Y_i z_i(p) = 1$ (as in Section 3.1, $Y_i z_i$ denotes the Lie derivative of the function z_i along the vector field Y_i). Hence the order of z_i at p is not greater than w_i.

It remains to show that the order of z_i at p is at least w_i for each $i = 1, \ldots, n$. This is a direct consequence of the following assertion.

Claim. *Let X be one of the vector fields $X_1, \ldots, X_m$. Then, for $i = 1, \ldots, n$, the Taylor expansion at $z = 0$ of the function $a_i(z) = X z_i\left(\phi(z)\right)$ is a sum of homogeneous polynomials in the coordinates z of weighted degree $\geq w_i - 1$.*

From the very definition of $a_i(z)$, we have

$$X(\phi(z)) \;=\; \sum_{i=1}^{n} a_i(z) \frac{\partial \phi}{\partial z_i}(z). \tag{29}$$

Given z, let φ be the diffeomorphism defined on a neighbourhood of p by $\varphi(q) = \exp(z_1 Y_1) \cdots \exp(z_n Y_n)(q)$. In particular, $\varphi(p) = \phi(z)$. In order to obtain an equality in T_pM, we apply the isomorphism $(d\varphi_p)^{-1}$ to both sides of (29), and we get

$$(\varphi^{-1})_* X(p) \;=\; \sum_{i=1}^{n} a_i(z) \left(\varphi^{-1}\right)_* \frac{\partial}{\partial z_i}(p).$$

This equality is of the form $W = \sum_{i=1}^{n} a_i V_i$, where the vectors $W = W(z)$ and $V_i = V_i(z)$, $i = 1, \ldots, n$, belong to T_pM. If we denote by $b = b(z) \in \mathbb{R}^n$ the coordinates of W in the basis $(Y_1(p), \ldots, Y_n(p))$ of T_pM, and by $P = P(z)$

the $(n \times n)$-matrix of the coordinates of $V_1, \ldots, V_n$ in the same basis, then the vector $a(z) = (a_1(z), \ldots, a_n(z))$ appears as the solution of $Pa = b$.

Note first that $P(0)$ equals the identity matrix $\mathbb{I}$. Therefore both matrices $P(z)$ and $P(z)^{-1}$ are equal to $\mathbb{I}$+homogeneous terms of positive degree. Hence the Taylor expansion of $a_i(z)$ and the one of $b_i(z)$ have the same homogeneous terms of lower degree.

On the other hand, since $\varphi^{-1} = \exp(-z_n Y_n) \cdots \exp(-z_1 Y_1)$, we have

$$W(z) \;=\; \exp(-z_n Y_n)_* \cdots \exp(-z_1 Y_1)_* X(p).$$

Let us choose an integer N bigger than all the weights w_i, and apply Proposition A.5 to $\exp(-z_1 Y_1)_* X$, then to $\exp(-z_2 Y_2)_* (\mathrm{ad}Y_1)^{l_1} X$, and so on,

$$\begin{aligned} W(z) \;&=\; \exp(-z_n Y_n)_* \cdots \exp(-z_2 Y_2)_* \sum_{l_1=0}^{N} \frac{(-z_1)^{l_1}}{l_1!} (\mathrm{ad}Y_1)^{l_1} X(p) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad +O(|z_1|^{N+1}) \\ &\;\;\vdots \\ &= \sum_{l_1,\ldots,l_n=0}^{N} \frac{(-z_1)^{l_1}}{l_1!} \cdots \frac{(-z_n)^{l_n}}{l_n!} \, (\mathrm{ad}Y_n)^{l_n} \cdots (\mathrm{ad}Y_1)^{l_1} X(p) + O(|z|^{N+1}). \end{aligned}$$

Hence every coordinate $b_i(z)$ of $W(z)$ satisfies

$$b_i(z) \;=\; \sum_{l_1,\ldots,l_n=0}^{N} \frac{(-z_1)^{l_1}}{l_1!} \cdots \frac{(-z_n)^{l_n}}{l_n!} \beta_i^l + O(|z|^{N+1}), \tag{30}$$

β_i^l being the ith coordinate in the basis $(Y_1(p), \ldots, Y_n(p))$ of the vector $(\mathrm{ad}Y_n)^{l_n} \cdots (\mathrm{ad}Y_1)^{l_1} X(p)$. The latter vector belongs to $\Delta^w(p)$, where $w = 1 + l_1 w_1 + \cdots + l_n w_n$ (recall that $X \in \Delta^1$ and $Y_i \in \Delta^{w_i}$). Since $(Y_1, \ldots, Y_n)$ is an adapted frame at p, β_i^l is zero when $1 + l_1 w_1 + \cdots + l_n w_n < w_i$. It follows that $b_i(z)$ – and then $a_i(z)$ – contains only homogeneous terms of weighted degree greater than or equal to $w_i - 1$. This ends the proofs of both the claim and the lemma. □

B.2 Canonical coordinates of the first kind

Let $p \in M$ and $Y_1, \ldots, Y_n$ an adapted frame at p. The map

$$\widetilde{\phi} : (z_1, \ldots, z_n) \mapsto \exp(z_1 Y_1 + \cdots + z_n Y_n)(p)$$

is a local diffeomorphism near $0 \in \mathbb{R}^n$ since its differential at 0 is an isomorphism. This results from

$$\frac{\partial \widetilde{\phi}}{\partial z_i}(0) = \frac{d}{dt}\left(\widetilde{\phi}(0,\dots,t,\dots,0)\right)\big|_{t=0} = \frac{d}{dt}\left(\exp(tY_i)(p)\right)\big|_{t=0} = Y_i(p),$$

for $i = 1,\dots,n$. The inverse of $\widetilde{\phi}$ defines some local coordinates near p called *canonical coordinates of the first kind*.

Lemma B.2. *Canonical coordinates of the first kind are privileged at p.*

The first proof of this lemma appeared in [RS76], with a different formulation. The proof we present here is rather different.

Proof. The proof follows exactly the same lines as the one of Lemma B.1, replacing ϕ by $\widetilde{\phi}$, and φ by $\widetilde{\varphi} = \exp(\sum_j z_j Y_j)$. We are left to compute the coordinates $\widetilde{b}_i(z)$, $i = 1,\dots,n$, of the vector $\widetilde{W}(z) = (\widetilde{\varphi}^{-1})_* X(p)$ in the basis $(Y_1(p),\dots,Y_n(p))$ of T_pM. It results directly from Proposition A.5 that

$$\begin{aligned}\widetilde{W}(z) &= \sum_{k=0}^{N} \frac{1}{k!}\left(\operatorname{ad}\sum_{i=1}^{\ell} z_j Y_j\right)^k X(p) + O(|z|^{N+1}),\\ &= \sum_{k=0}^{N}\sum_{l_1+\cdots+l_n=k} a_l z_1^{l_1}\cdots z_n^{l_n} Z_l(p) + O(|z|^{N+1}),\end{aligned}$$

where Z_l belongs to $\Delta^w(p)$, with $w = 1 + l_1 w_1 + \cdots + l_n w_n$. Thus every coordinate $\widetilde{b}_i(z)$ of $\widetilde{W}(z)$ satisfies

$$\widetilde{b}_i(z) = \sum_{k=0}^{N}\sum_{l_1+\cdots+l_n=k} a_l z_1^{l_1}\cdots z_n^{l_n}\widetilde{\beta}_i^l + O(|z|^{N+1}),$$

$\widetilde{\beta}_i^l$ being the ith coordinate of Z_l in the basis $(Y_1(p),\dots,Y_n(p))$. This expression is similar to (30), and the same conclusion follows. □

B.3 Algebraic coordinates

Let us recall the construction of the algebraic coordinates $(z_1,\dots,z_n)$ given in page 24. Let $Y_1,\dots,Y_n$ be an adapted frame at p, and $(y_1,\dots,y_n)$ be local coordinates centered at p such that $\partial_{y_i}|_p = Y_i(p)$. For $j = 1,\dots,n$, we set

$$z_j = y_j - \sum_{k=2}^{w_j-1} h_k(y_1,\dots,y_{j-1}), \tag{31}$$

where, for $k = 2, \dots, w_j - 1$,

$$h_k(y_1, \dots, y_{j-1}) = \sum_{\substack{|\alpha|=k \\ w(\alpha)<w_j}} Y_1^{\alpha_1} \dots Y_{j-1}^{\alpha_{j-1}} \Big(y_j - \sum_{q=2}^{k-1} h_q(y) \Big)(p) \ \frac{y_1^{\alpha_1}}{\alpha_1!} \cdots \frac{y_{j-1}^{\alpha_{j-1}}}{\alpha_{j-1}!},$$

with $|\alpha| = \alpha_1 + \cdots + \alpha_n$.

Lemma B.3. *The algebraic coordinates $(z_1, \dots, z_n)$ are privileged at p.*

The proof of the lemma is based on the following result.

Lemma B.4. *A function f is of order $\geq s$ at p if and only if*

$$(Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} f)(p) = 0$$

for all α such that $w(\alpha) < s$.

Proof. Let f be a function of order $\geq s$ at p. Using the rules (9), we have $\mathrm{ord}_p(Y_i) \geq -w_i$ for $i = 1, \dots, n$, and hence $\mathrm{ord}_p(Y_1^{\alpha_1} \cdots Y_n^{\alpha_n}) > -s$ for every $\alpha = (\alpha_1, \dots, \alpha_n)$ such that $w(\alpha) < s$. Consequently, for such an α the function $Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} f$ is of positive order, and so vanishes at p.

Conversely, let f be a function of order $< s$ at p. We introduce the canonical coordinates of the second kind $(x_1, \dots, x_n)$ defined by means of the adapted basis $Y_1, \dots, Y_n$. Proposition 3.2 implies that there exists α such that $w(\alpha) = \mathrm{ord}_p(f) < s$ and $(\partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n} f)(p) \neq 0$. Moreover, every vector field Y_i, $i = 1, \dots, n$, writes in coordinates x as

$$\sum_{j=1}^{n} Y_i^j(x) \partial_{x_j}, \quad \text{where } \mathrm{ord}_p(Y_i^j) \geq w_j - w_i.$$

There also holds $Y_i^j(0) = \delta_{ij}$ since $Y_i(p) = \partial_{x_i}$. As a consequence,

$$Y_1^{\alpha_1} \cdots Y_n^{\alpha_n}(p) = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n}(p) + \sum_{w(\beta)<w(\alpha)} a_\beta \partial_{x_1}^{\beta_1} \cdots \partial_{x_n}^{\beta_n}(p),$$

and thus $(Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} f)(p) = (\partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n} f)(p) \neq 0$ since $w(\alpha) = \mathrm{ord}_p(f)$. This ends the proof. □

Proof of Lemma B.3. Let $i \in \{1, \dots, n\}$. Note first that $Y_i z_i(p) = Y_i y_i(p) = 1$, which implies $\mathrm{ord}_p(z_i) \leq w_i$. It remains to show that $\mathrm{ord}_p(z_i) \geq w_i$. For this we will use the criterion of Lemma B.4.

Let α such that $w(\alpha) < w_i$ (and so $|\alpha| < w_i$). Using the expression (31) of z_i, we obtain

$$\begin{aligned} Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} z_i &= Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} \left(y_i - \sum_{k=2}^{w_i - 1} h_k(y) \right) \\ &= Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} \left(y_i - \sum_{k=2}^{|\alpha|-1} h_k(y) \right) - Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} \left(\sum_{k=|\alpha|}^{w_i - 1} h_k(y) \right). \quad (32) \end{aligned}$$

The functions h_k are given by

$$h_k(y) = \sum_{\substack{|\beta|=k \\ w(\beta)<w_i}} Y_1^{\beta_1} \dots Y_{i-1}^{\beta_{i-1}} \Big(y_i - \sum_{q=2}^{k-1} h_q(y) \Big)(p) \ \frac{y_1^{\beta_1}}{\beta_1!} \cdots \frac{y_{i-1}^{\beta_{i-1}}}{\beta_{i-1}!}.$$

Therefore, we clearly have $\left(Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} h_k\right)(p) = 0$ if $k > |\alpha|$, and

$$\left(Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} h_{|\alpha|}\right)(p) = Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} \left(y_i - \sum_{k=2}^{|\alpha|-1} h_k(y) \right)(p).$$

Plugging this expression into (32), we obtain $(Y_1^{\alpha_1} \cdots Y_n^{\alpha_n} z_i)(p) = 0$, which ends the proof. □

References

[ABB11] A. Agrachev, D. Barilari, and U. Boscain. On the Hausdorff volume in sub-Riemannian geometry. *Calculus of Variations and Partial Differential Equations*, 43(3-4):355–388, 2011.

[ABB12] A. Agrachev, D. Barilari, and U. Boscain. Introduction to Riemannian and sub-Riemannian geometry (from Hamiltonian viewpoint). Preprint SISSA 09/2012/M, 2012.

[AM78] R. Abraham and J. E. Marsden. *Foundations of Mechanics*. Addison-Wesley, 1978.

[AS87] A. A. Agrachev and A. V. Sarychev. Filtrations of a Lie algebra of vector fields and nilpotent approximations of control systems. *Dokl. Akad. Nauk SSSR*, 285:777–781, 1987. English transl.: *Soviet Math. Dokl.*, 36:104–108, 1988.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001.

[Bel96] A. Bellaïche. The tangent space in sub-Riemannian geometry. In A. Bellaïche and J.-J. Risler, editors, *Sub-Riemannian Geometry*, Progress in Mathematics. Birkhäuser, 1996.

[Boo86] W. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.

[Bou72] N. Bourbaki. *Groupes et Algèbres de Lie*. Hermann, Paris, 1972.

[DLPR12] Lance Drager, Jeffrey Lee, Efton Park, and Ken Richardson. Smooth distributions are finitely generated. *Annals of Global Analysis and Geometry*, 41:357–369, 2012. 10.1007/s10455-011-9287-8.

[FJ03] E. Falbel and F. Jean. Measures of transverse paths in sub-Riemannian geometry. *Journal d'Analyse Mathematique*, 91:231–246, 2003.

[Gab95] A. Gabrielov. Multiplicities of zeroes of polynomials on trajectories of polynomial vector fields and bounds on degree of non-holonomy. *Mathematical Research Letters*, 2:437–451, 1995.

[Ger84] V. Ya. Gershkovich. Two-sided estimates of metrics generated by absolutely non-holonomic distributions on Riemannian manifolds. *Sov. Math. Dokl.*, 30:506–510, 1984.

[GJ12] R. Ghezzi and F. Jean. A new class of $(\mathcal{H}^k, 1)$-rectifiable subsets of metric spaces. *Com. on Pure and Applied Analysis*, 2012. To appear.

[GJR98] A. Gabrielov, F. Jean, and J.-J. Risler. Multiplicity of polynomials on trajectories of polynomials vector fields in C^3. In W. Pawłucki B. Jakubczyk and J. Stasica, editors, *Singularities Symposium – Łojasiewicz 70*, volume 44, pages 109–121. Banach Center Publications, Warszawa, 1998.

[Gro96] M. Gromov. Carnot-Carathéodory spaces seen from within. In A. Bellaïche and J.-J. Risler, editors, *Sub-Riemannian Geometry*, Progress in Mathematics. Birkhäuser, 1996.

[Her91] H. Hermes. Nilpotent and high-order approximations of vector field systems. *SIAM Review*, 33(2):238–264, 1991.

[Kha01] H. Khalil. *Nonlinear Systems*. Prentice Hall, 3nd edition, 2001.

[Kup96] I. Kupka. Géométrie sous-riemannienne. In *Séminaire N. Bourbaki*, volume 817, June 1996.

[Lee03] John. M. Lee. *Introduction To Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, 2003.

[LM67] E.B. Lee and L. Markus. *Foundations of Optimal Control Theory*. Wiley, New York, 1967.

[LS91] G. Lafferriere and H. Sussmann. Motion planning for controllable systems without drift. In *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, Sacramento, California, 1991.

[Mit85] J. Mitchell. On Carnot-Carathéodory metrics. *Journal of Differential Geom.*, 21:35–45, 1985.

[MM00] G. A. Margulis and G. D. Mostow. Some remarks on the definition of tangent cones in a Carnot-Carathéodory space. *Journal d'Analyse Mathématique*, 80:299–317, 2000.

[Mon02] R. Montgomery. *A tour of sub-Riemannian geometries, their geodesics and applications*. Math. Surveys and Monographs. American Math. Soc., 2002.

[Nag66] T. Nagano. Linear differential systems with singularities and an application to transitive Lie algebras. *J. Math. Soc. Japan*, 18:398–404, 1966.

[NSW85] A. Nagel, E. M. Stein, and S. Wainger. Metrics defined by vector fields. *Acta Math.*, 155:103–147, 1985.

[Rif] L. Rifford. Nonholonomic variations: An introduction to sub-Riemannian geometry. In preparation.

[RS76] L. P. Rothschild and E. M. Stein. Hypoelliptic differential operators and nilpotent groups. *Acta Math.*, 137:247–320, 1976.

[Rud70] W. Rudin. *Real and complex analysis*. McGraw-Hill, 1970.

[Ste74] P. Stefan. Accessible sets, orbits, and foliations with singularities. *Proc. London Math. Soc.*, 29(3):699–713, 1974.

[Ste86] G. Stefani. On local controllability of a scalar-input system. In Lindquist Byrnes, editor, *Theory and Appl. of Nonlinear Control Syst.*, pages 167–179. North Holland, Amsterdam, 1986.

[Sus73] H. J. Sussmann. Orbits of families of vector fields and integrability of distributions. *Trans. Amer. Math. Soc.*, 180:171–188, 1973.

[Sus74] H. J. Sussmann. An extension of theorem of Nagano on transitive Lie algebras. *Proc. Amer. Math. Soc.*, 45:349–356, 1974.