Universiteit Utrecht

Graduate School of Natural Sciences

Bracket-generating distributions and filtered structures

Master Thesis

C.J.B. Bruijnen

Mathematical Sciences

Supervisors:

Dr. Á. del Pino Gómez
Universiteit Utrecht

Prof. Dr. E.P. van den Ban
Universiteit Utrecht

11th August 2020

Contents

1 Introduction

Distributions are subbundles of the tangent bundle of a manifold. Many problems in control theory can be modelled using distributions. An easy example of this is a coin rolling on a table (adapted from [3]).

The position of the coin can be described as $(x, y, \theta) \in \mathbb{R}^2 \times \mathbb{S}^1$. Here, the first two coordinates $(x, y) \in \mathbb{R}^2$ indicate the position of the coin on the table, and the last coordinate $\theta \in \mathbb{S}^1$ describes the direction of the coin. We use $\mathbb{S}^1$, because we remember heads and tails. Intuitively, it is clear that the motion of the coin is restricted. The change in position has to be proportional to the direction, since the coin can only move in the direction it points to.

Concretely, if $\gamma(t) = (x(t), y(t), \theta(t))$ is a path describing the movement of the coin, then the change in position $(x'(t), y'(t))$ has to be proportional to the direction $(\cos(\theta(t)), \sin(\theta(t)))$. This means that γ has to satisfy any of the following equivalent conditions:

- $\det \begin{pmatrix} x'(t) & \cos(\theta(t)) \\ y'(t) & \sin(\theta(t)) \end{pmatrix} = \sin(\theta(t))x'(t) - \cos(\theta(t))y'(t) = 0.$
- $\gamma^*\alpha = 0$ for the 1-form $\alpha = \sin(\theta)dx - \cos(\theta)dy$.
- γ is tangent to $\xi = \ker(\alpha) = \langle \cos(\theta)\partial_x + \sin(\theta)\partial_y, \partial_\theta \rangle$.

As we can see, ξ is a subbundle of the tangent bundle $T(\mathbb{R}^2 \times \mathbb{S}^1)$, so ξ is a distribution on $\mathbb{R}^2 \times \mathbb{S}^1$. Moreover,

$$[\cos(\theta)\partial_x + \sin(\theta)\partial_y, \partial_\theta] = \sin(\theta)\partial_x - \cos(\theta)\partial_y.$$

Especially,

$$\langle \cos(\theta)\partial_x + \sin(\theta)\partial_y, \partial_\theta, [\cos(\theta)\partial_x + \sin(\theta)\partial_y, \partial_\theta] \rangle = \mathfrak{X}(\mathbb{R}^2 \times \mathbb{S}^1).$$

The distribution ξ on $\mathbb{R}^2 \times \mathbb{S}^1$ is an example of a bracket-generating distribution, meaning that the set of vector fields on the manifold can be spanned by bracket expressions of vector fields in the framing of the distribution. Chow's Theorem 3.13 tells us that if a manifold is endowed with a bracket-generating distribution, we can connect any two points on the manifold by a path tangent to the distribution ([8], [3]). In the case of the rolling coin, this means that the coin can roll to any point on the table, as is intuitively clear.

When studying distributions, it is often useful to look at the vector fields tangent to the distribution, instead of looking at the distribution directly. Vector fields provide more useful operations, the Lie bracket chief among them. Bracket-generating is one of the properties of distributions which is defined through Lie brackets.

In this thesis, we will also study filtered structures, which is a filtration of the tangent bundle by distributions which behaves well with respect to the Lie bracket ([9]). A filtered structure is of the form $\xi_1 \subset \cdots \subset \xi_m = TM$, where the ξ_i are distributions. Under some additional regularity conditions (weak regularity), a bracket-generating distribution can generate a special case of a filtered structure. In general, the lowest-order distribution ξ_1 does not generate the other distributions in the filtered structures. An example of a filtered structure on $\mathbb{R}^2 \times \mathbb{S}^1$ is given by $\xi_1 \subset \xi_2 \subset \xi_3$ with

$$\begin{aligned} \xi_1 &= \langle \cos(\theta)\partial_x + \sin(\theta)\partial_y + \partial_\theta \rangle \\ \xi_2 &= \langle \cos(\theta)\partial_x + \sin(\theta)\partial_y, \partial_\theta \rangle \\ \xi_3 &= T(\mathbb{R}^2 \times \mathbb{S}^1). \end{aligned}$$

In particular, ξ_1 does not generate ξ_2 or ξ_3.

For a filtered structure, the growth of the filtration and the behaviour with respect to the Lie bracket can be packed into one algebraic object, called the nilpotentisation. It is given by $\xi_1 \oplus \xi_2/\xi_1 \oplus \cdots \oplus \xi_m/\xi_{m-1}$. When restricted to a point, it is a Lie algebra.

Given a filtered structure, we can induce a local distance function by privileged coordinates, which is a weighted system of coordinates centred at a point, where the weighting is induced by the filtered structure ([9]). Through an open cover of the manifold, we can combine local distances on the sets in the cover to create a global distance function on the manifold. The creation of this global distance function is one of the original contributions of this thesis.

In Chapters 5 and 6, we will study the tangent cone of a manifold with a filtered structure. The tangent cone appears when we 'stretch' the distances to infinity, while one point remains fixed ([8], [6]). One would expect that by this zooming in process, we would lose all structure, but Mitchell's Theorem tells us otherwise. The main contribution of this thesis is a version of Mitchell's Theorem for filtered structures, Theorem 5.19. (The usual version of Mitchell's Theorem, Theorem 5.18, is stated in terms of weakly regular bracket-generating distributions. [8]) It tells us that the tangent cone exists for a manifold with a filtered structure, and that it is equal to the simply-connected Lie group with Lie algebra the nilpotentisation at the point.

2 Distributions

In this chapter, we will study distributions and their properties, laying the foundation for the rest of the thesis. Instead of looking at a distribution directly, we will often study the set of vector fields tangent the distribution, which have more useful operations. The most important operation will be the Lie bracket of vector fields.

For a distribution ξ, we can study the growth of the set $\Gamma(\xi)$ of vector fields tangent to the distribution with respect to the Lie bracket, giving us the associated Lie flag. The two extremes in behaviour with respect to the Lie bracket are involutive distributions and bracket-generating distributions. For the former, $\Gamma(\xi)$ forms a Lie algebra subalgebra of the Lie algebra of vector fields on the manifold. In the latter case, $\Gamma(\xi)$ generates the set of all vector fields on the manifold.

Under certain regularity conditions, the associated Lie flag induces a notion of curvature, an analogue of curvature for the Riemannian setting. As with Riemannian curvature, it measures how far removed one is from the flat case. In the case of distributions, the flat case is an integrable (or involutive) distribution. In the case of Riemannian metrics, it is the Euclidean case.

In Section 2.7 we will introduce the notion of filtered structures, which is a filtration of the tangent bundle through distributions. This will be useful in later chapters, since it removes the assumption of the lowest-order distribution generating the rest of the filtration.

This chapter is based on [3] and chapters 2 and 4 of [8] (for distributions), [9] (for filtered structures), and [10] and chapter 3 of [4] (for the Campbell-Baker-Hausdorff formula). For more on Pfaffian systems, see [12]. For more on the Campbell-Baker-Hausdorff Formula, see [4] and [10].

2.1 Distributions, an introduction

Let us restate the definition of a distribution.

Definition 2.1. Let M be a smooth manifold. A **distribution** is a subbundle $\xi \subset TM$. The **rank** of ξ is the dimension of its fibres.

From the introduction we know that the model of the rolling coin on the table is an example of a distribution. Another example is the following:

Example 2.2 (Martinet distribution)**.** Consider the smooth manifold $\mathbb{R}^3(x, y, z)$. Then

$$\xi = \ker(dy - z^2 dx) = \langle \partial_x + z^2 \partial_y, \partial_z \rangle$$

is a distribution, called the **Martinet distribution**. It has rank 2. △

It is useful to look at the vector fields tangent to a given distribution, which have more useful operations. These tangent vector fields can tell us more about the properties of a distribution.

Definition 2.3. Let (M, ξ) a manifold endowed with a distribution. Then $\Gamma(\xi) \subset \mathfrak{X}(M)$ the set of all smooth vector fields on M tangent to ξ.

As we will show shortly, the set $\Gamma(\xi)$ is a $C^\infty(M)$-module, which is defined as follows:

Definition 2.4. A $C^\infty(M)$**-module** is a set V endowed with multiplication $V \times V \to V$, $(v, w) \mapsto v + w$ and scalar multiplication $C^\infty(M) \times V \to V$, $(f, v) \mapsto fv$ such that $(V, +)$ is an abelian group, and such that for all $v, w \in V$, $f, g \in C^\infty(M)$ we have

- $f(gv) = (fg)v$
- $(f+g)v = fv + gv$
- $f(v+w) = fv + fw$.

Proposition 2.5. *Let ξ a distribution on M. Then $\Gamma(\xi)$ is a C^∞-module of vector fields on M.*

Proof. Let $X, Y \in \Gamma(\xi)$, $f \in C^\infty(M)$. Using the usual addition of vector fields and multiplication with smooth functions, we have $(X+Y)_p = X_p + Y_p$ for all $p \in M$, and $(fX)_p = f(p)X(p)$. By the properties of these operations, it is enough to prove that $X+Y$, $fX \in \Gamma(\xi)$. Note that for all $(p,v) \in \xi$, we have $(X+Y)_p(v) = X_p(v) + Y_p(v) = 0$, and $(fX)_p(v) = f(p)X_p(v) = 0$, which proves our assertion. □

From this proposition, it follows that there is a natural map which sends distributions on M to C^∞-modules of vector fields on M given by $\xi \mapsto \Gamma(\xi)$.

Definition 2.6. Let Γ a C^∞-module of vector fields on M. Let

$$\Gamma(p) = \{X_p : X \in \Gamma\}$$

the evaluation at $p \in M$. Then Γ is of **pointwise constant rank** if there is $k \in \mathbb{N}$ such that $\dim(\Gamma(p)) = k$ for all $p \in M$.

Because ξ is a subbundle of TM, the set $\Gamma(\xi)$ is of pointwise constant rank. Moreover, if M is a closed manifold, we can recover a distribution from a $C^\infty(M)$-module of vector fields of pointwise constant rank by evaluating at every point.

2.2 The associated flag

One of the most useful operations for studying distributions is the Lie bracket of vector fields. Since $\mathfrak{X}(M)$ is a Lie algebra, we may look at the behaviour of $\Gamma(\xi)$ with respect to the Lie bracket.

Bracket expressions will allow us to write iterated Lie brackets in a more convenient manner.

Definition 2.7. A **bracket expression of length** 1 is a string of the form "a_1", which has the formal variable a_1 as input. A **bracket expression of length** $i+j$ is a string of the form "$[A(a_1, \dots, a_i), B(a_{i+1}, \dots, a_{i+j})]$", where $A(-)$ and $B(-)$ are bracket expressions of length i, j respectively.

The following lemma tells us that we can reduce any bracket expression into a linear combination of bracket expressions of a nicer form, namely $[, B(-)]$. This result will be very useful later.

Lemma 2.8. *Let $A(-)$ be a bracket expression of length i whose entries are vector fields. Then we can write it as linear combination of bracket expressions of the form $[, B(-)]$, with $B(-)$ a bracket-expression of length at most $i-1$.*

Proof. If $i = 0, 1, 2$ the statement is trivial.
For $i = 3$, we have two options. Either $A(-)$ is of the form $[, B(-)]$ for $B(-)$ a bracket expression of length 2, or $A(-)$ is of the form $[[,],[,]]$. Using the Jacobi-identity, we have:

$$[[X,Y],[Z,W]] = [Z,[W,[X,Y]]] + [W,[Z,[X,Y]]].$$

For $i \geq 4$, the result follows from similar computations, namely through repeatedly applying the Jacobi identity until we have the desired result. □

By keeping track of the lengths of bracket expressions, the Lie bracket gives rise to the associated Lie flag. This will be a useful object for describing infinitesimal properties of the distribution.

Definition 2.9. The **associated Lie flag** of a distribution ξ is defined as a sequence of $C^\infty(M)$-modules

$$\Gamma^{(1)}(\xi) \subset \Gamma^{(2)}(\xi) \subset \Gamma^{(3)}(\xi) \subset \cdots \subset \mathfrak{X}(M)$$

where

$$\Gamma^{(i)}(\xi) = \langle A_i(X_1, \ldots, X_i) : X_1, \ldots, X_i \in \Gamma(\xi) \rangle_{C^\infty(M)}$$

with $A_i(-)$ ranging over all bracket expressions of length at most i. Here, the brackets indicate the span over $C^\infty(M)$.

Note that $\Gamma^{(1)}(\xi) = \Gamma(\xi)$.

We return to our previous example of the Martinet distribution. For this distribution, it is easy to calculate the associated Lie flag.

Example 2.10 (Martinet distribution)**.** Consider again the Martinet distribution on $\mathbb{R}^3(x, y, z)$, which is given by

$$\xi = \ker(dy - z^2 dx) = \left\langle \partial_x + z^2\partial_y, \partial_z \right\rangle.$$

We have:

$$\begin{aligned} [\partial_z, \partial_x + z^2\partial_y] &= 2z\partial_y \\ [\partial_z, [\partial_z, \partial_x + z^2\partial_y]] &= 2\partial_y. \end{aligned}$$

It follows that $\Gamma^{(2)}(\xi)$ has rank 3 everywhere, except at the hypersurface $\{z = 0\}$ where it has rank 2, so $\Gamma^{(2)}(\xi)$ is **not** a distribution. Moreover, $\Gamma^{(3)}(\xi) = \mathfrak{X}(M)$. $\triangle$

As we see in the example above, $\Gamma^{(i)}(\xi)$ may not correspond to a distribution. This motivates us to define the following:

Definition 2.11. A distribution ξ is **weakly regular** if for all i, $\Gamma^{(i)}(\xi)$ is the $C^\infty(M)$-submodule of vector fields which is tangent to a distribution, which we will call $\xi^{(i)}$.

For a general distribution ξ, we have by definition that $\Gamma^{(1)}(\xi) \subset \Gamma^{(2)}(\xi) \subset \cdots \subset \mathfrak{X}(M)$. It is possible that the subsets remain strict, and that the sequence $\Gamma^{(1)}(\xi) \subset \Gamma^{(2)}(\xi) \subset \ldots$ keeps growing. On the other hand, it is also possible that the sequence 'stops' at some point, i.e. we have

$$\Gamma^{(1)}(\xi) \subset \Gamma^{(2)}(\xi) \subset \cdots \subset \Gamma^{(m)}(\xi) = \Gamma^{(m+1)}(\xi) = \cdots \subset \mathfrak{X}(M).$$

This is the case when ξ stabilises.

Definition 2.12. Let (M, ξ) a manifold endowed with a distribution. Then ξ **stabilises** if there exists $m \in \mathbb{N}$ such that $\Gamma^{(i)}(\xi) = \Gamma^{(m)}(\xi)$ for all $i \geq m$.

There is a link between a distribution being stabilising and being weakly regular, as we see in the following lemma.

Lemma 2.13. *Let (M, ξ) a manifold endowed with a distribution. If ξ is weakly regular, then it stabilises.*

Proof. Suppose that the Lie flag does not stabilise, i.e. for every m there is $i > m$ such that $\Gamma^{(i)}(\xi) \neq \Gamma^{(m)}(\xi)$. By definition, $\Gamma^{(m)}(\xi) \subset \Gamma^{(i)}(\xi)$, so it follows that $\text{rank}(\Gamma^{(m)}(\xi)) < \text{rank}(\Gamma^{(i)}(\xi))$. This means that for every $k \in \mathbb{N}$ there is i such that $\text{rank}(\Gamma^{(i)}(\xi)) \geq k$. However, $\Gamma^{(i)}(\xi)$ is the $C^\infty(M)$-module tangent to the distribution $\xi^{(i)}$, and therefore $\xi^{(i)} \subset TM$, so $\text{rank}(\Gamma^{(i)}(\xi)) = \text{rank}(\xi^{(i)}) \leq \dim(M)$. This leads to a contradiction. So, there is $m \in \mathbb{N}$ such that $\Gamma^{(i)}(\xi) = \Gamma^{(m)}(\xi)$ for all $i \geq m$. □

Having established this link, we can introduce the notion of a growth vector. As the name suggests, it is a vector which measures how a weakly regular distribution grows with respect to its associated Lie flag.

Definition 2.14. The **growth vector** of a weakly regular distribution ξ is the vector

$$(\text{rank}(\xi^{(1)}), \text{rank}(\xi^{(2)}), \ldots, \text{rank}(\xi^{(m)}))$$

where m is the step in which the associated rank stabilises.

2.3 Differential systems

Even when a distribution ξ is not weakly regular, not all is lost. The modules $\Gamma^{(i)}(\xi)$ in the associated Lie flag are still reasonably well-behaved. For example:

Lemma 2.15. *The $\Gamma^{(i)}(\xi)$ are locally finitely generated as $C^\infty(M)$-modules.*

Proof. We will prove the statement by induction on i.

As a bundle, ξ is trivial over any ball $U \subset M$. So there exists a local framing $\{X_1, \ldots, X_k\} \subset \mathfrak{X}(U)$. Any vector field tangent to ξ can then be expressed as $X = \sum_{j=1}^k f_j X_j$ for some smooth functions $f_i : U \to \mathbb{R}$. Hence, $\Gamma^{(1)}(\xi)$ is locally finitely generated.

Next, suppose that $\Gamma^{(i-1)}(\xi)$ is finitely generated over U, and let $\{Y_1, \ldots, Y_m\} \subset \mathfrak{X}(U)$ a local framing for $\Gamma^{(i-1)}(\xi)$. By lemma 2.8, any bracket expression of length i can be written as a linear combination of bracket expressions of the form $[, B(-)]$ with $B(-)$ a bracket expression of length $i-1$. So, $X \in \Gamma^{(i)}(\xi)$ can be written as

$$X = \sum_{j=1}^{m} \sum_{l=1}^{k} f_j Y_j + g_{jl}[X_l, Y_j]$$

for some smooth functions $f_j, g_{jl} : U \to \mathbb{R}$. Hence, $\Gamma^{(i)}(\xi)$ is locally finitely generated. □

This motivates the following definition:

Definition 2.16. A **differential system** is a locally finitely generated $C^\infty(M)$-module of vector fields.

2.4 For completeness: Pfaffian systems and annihilators

So far, we have looked at distributions and the tangent bundle. But we can define a very similar construction for the cotangent bundle, namely Pfaffian systems. Apart from being analogous to distributions, there is a direct correspondence between distributions and Pfaffian systems. But first, let us give a definition of a Pfaffian system.

Definition 2.17. A **Pfaffian system** I on a manifold M is a subbundle of T^*M.

As we have alluded to before, Pfaffian systems arise naturally from distributions through the annihilator of a distribution.

Example 2.18. Let ξ be a distribution on a smooth manifold M. The annihilator of ξ is the set

$$\xi^\perp = \{\omega \in T^*M : \omega(v) = 0, \, \forall v \in \xi\}.$$

Especially, it is a Pfaffian system. △

Let I be a Pfaffian system. Then we denote by $\Gamma(I)$ its space of sections, which is a $C^\infty(M)$-submodule of $\Omega^1(M)$. On this space we can define similar (and dual) notions to $\Gamma(\xi)$, the space of sections of a distribution ξ.

Definition 2.19. Given a Pfaffian system I, its **dual flag** is given by

$$\Gamma^{(1)}(I) \supset \Gamma^{(2)}(I) \supset \dots$$

where $\Gamma^{(1)}(I) = \Gamma(I)$, and we define inductively

$$\Gamma^{(i)}(I) = \{\alpha \in \Gamma^{(i-1)}(I) : d\alpha(\omega, -) \in \Gamma^{(i-1)}(I), \, \forall \omega \in I^\perp\}.$$

Analogous to the growth vector of a weakly regular distribution, we can define the dual growth vector.

Definition 2.20. Given a Pfaffian system I, its **dual growth vector** is the vector

$$(\mathrm{corank}(\Gamma^{(1)}(I)), \mathrm{corank}(\Gamma^{(2)}(I)), \dots, \mathrm{corank}(\Gamma^{(m)}(I)))$$

where m is the step in which the dual flag stabilises.

Note that the dual growth vector may change with the point.

The following lemma shows that the annihilator provides a duality between distributions and Pfaffian systems.

Lemma 2.21. *If ξ is a weakly regular distribution, then* $(\Gamma^{(i)}(\xi))^\perp = \Gamma^{(i)}(\xi^\perp)$.

Proof. We prove the statement by induction on i. For $i = 1$, we have

$$(\Gamma^{(1)}(\xi))^\perp = \Gamma(\xi)^\perp = \Gamma(\xi^\perp) = \Gamma^{(1)}(\xi^\perp).$$

Now suppose the statement is true for $i - 1$. We then have $\alpha \in \Gamma^{(i)}(\xi^\perp)$ if and only if $d\alpha(v, -) \in \Gamma^{(i-1)}(\xi^\perp)$ for all $v \in \Gamma(\xi)$ by definition. So, if and only if $d\alpha(v, w) = 0$ for all $w \in \Gamma^{(i-1)}(\xi)$, because $(\Gamma^{(i-1)}(\xi))^\perp = \Gamma^{(i-1)}(\xi^\perp)$ by the induction hypothesis. By Cartan's magic formula, we therefore have $0 = d\alpha(v, w) = \alpha([v, w])$. Because $\Gamma^{(i)}(\xi)$ is spanned by all elements of the form $[v, w]$ for $v \in \Gamma(\xi) = \Gamma^{(1)}(\xi)$ and $w \in \Gamma^{(i-1)}(\xi)$, it follows that $\alpha \in \Gamma^{(i)}(\xi^\perp)$ if and only if $\alpha \in (\Gamma^{(i)}(\xi))^\perp$. □

If ξ is weakly regular, the dual growth vector of $\xi^\perp$ is equal to the growth vector of ξ. Generally, the two might differ. This happens for example in the Martinet distribution.

Example 2.22 (The Martinet distribution)**.** As we have seen before, the Martinet distribution is the distribution

$$\xi = \ker(dy - z^2dx) = \langle \partial_x + z^2\partial_y, \partial_z \rangle$$

on the manifold $\mathbb{R}^3(x, y, z)$. Its annihilator is

$$\xi^\perp = \langle \alpha = dy - z^2dx \rangle .$$

We will show that if $d(g\alpha)(v, w) = 0$ for all $v, w \in \Gamma(\xi)$, then $g = 0$.
Because ξ is bracket-generating, we have that g must be zero in $\{z \neq 0\}$. By continuity, we then have that g must also be zero in $\{z = 0\}$. Hence, $\Gamma^{(1)}(\xi^\perp) = 0$. On the other hand, we have

$$[\partial_z, \partial_x + z^2\partial_y] = 2z\partial_y.$$

Therefore, $\Gamma^{(1)}(\xi^\perp) \neq (\Gamma^{(1)}(\xi))^\perp$. △

2.5 Some relevant cases

Both the associated Lie flag and the growth vector are differential invariants of a distribution. The two extreme cases are given by involutive distributions and bracket-generating distributions. For the former, the set of all smooth vector fields tangent to the distribution forms a Lie algebra. In the latter, the smooth vector fields tangent to the distribution generate the set of all vector fields on the manifold.

Definition 2.23. A distribution ξ is **involutive** if $\Gamma^{(i)}(\xi) = \Gamma^{(1)}(\xi)$ for all i.

Equivalently, a distribution is involutive if $\Gamma(\xi)$ is a Lie subalgebra of $\mathfrak{X}(M)$.

Definition 2.24. Let ξ be a distribution, with $\mathrm{rank}(\xi) = k$. Then ξ is **integrable** if there exist local coordinates $(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n)$, where ξ is given as the kernel of $dx_{k+1}, \ldots, dx_n$.

Integrable distributions are foliations.

Definition 2.25. A **foliation** of dimension k on a manifold M is a partition $\mathcal{F}$ of M into disjoint, connected, immersed submanifolds of dimension k such that for every point $p \in M$ there is a chart (U, ϕ) with $p \in U$ and $\phi(U)$ a k-dimensional cube in $\mathbb{R}^n$. For every $N \in \mathcal{F}$, we have either $N \cap U = \emptyset$, or $N \cap U$ is a countable union of k-dimensional slices of the form $x_{k+1} = c_{k+1}, \ldots, x_n = c_n$. Lastly, a submanifold $N \in \mathcal{F}$ is called a **leaf**.

Example 2.26. The collection of tangent spaces to the leaves of a foliation forms an involutive distribution on M. △

The following theorem is an important and useful result (which we will not prove here).

Theorem 2.27 (Frobenius)**.** *A distribution is involutive if and only if it is integrable.*

For a proof of Frobenius's Theorem, see [5].
It is immediate that an integrable distribution is involutive. The other direction is harder to prove.
On the other end of the spectrum of distributions are the bracket-generating distributions. In that case, $\mathfrak{X}(M)$ is completely generated by bracket-expressions of vector fields in $\Gamma(\xi)$.

Definition 2.28. A distribution ξ is **bracket-generating** if there is an integer m such that $\Gamma^{(m)}(\xi) = \mathfrak{X}(M)$.

Thus, if ξ is bracket-generating, $\Gamma(\xi)$ generates $\mathfrak{X}(M)$ as an algebra.

2.6 Curvatures

As we have seen before, the associated Lie flag is comprised of bracket-expressions of elements in $\Gamma(\xi)$, and thus measures the non-integrability and weak regularity of a distribution ξ. If ξ is weakly regular, the Lie bracket induces a well-defined morphism

$$\Gamma^{(i)}(\xi) \times \Gamma^{(j)}(\xi) \to \Gamma^{(i+j)}(\xi).$$

Before we can define the morphism, we have to check that the Lie bracket is well-behaved with respect to the grading.

Lemma 2.29. *Let $X \in \Gamma^{(i)}(\xi)$, $Y \in \Gamma^{(j)}(\xi)$. Then $[X,Y] \in \Gamma^{(i+j)}(\xi)$.*

Proof. The Lie bracket is bilinear, so we may assume without loss of generality that $X = A(X_1,\dots,X_i)$ and $Y = B(Y_1,\dots,Y_j)$. Here, $A(-)$, $B(-)$ are bracket expressions of length at most i, j respectively, and $X_1,\dots,X_i,Y_1,\dots,Y_j \in \Gamma(\xi)$. Then $[A(-),B(-)]$ is a bracket expression of length at most $i+j$, and therefore

$$[X,Y] = [A(X_1,\dots,X_i), B(Y_1,\dots,Y_j)] \in \Gamma^{(i+j)}(\xi).$$

$\square$

Since the Lie bracket is a derivation, we get the following result:

Lemma 2.30. *The Lie bracket yields a well-defined bundle morphism*

$$\Omega_{(i,j)}(\xi) : \left(\xi^{(i)}/\xi^{(i-1)}\right) \times \left(\xi^{(j)}/\xi^{(j-1)}\right) \to \xi^{(i+j)}/\xi^{(i+j-1)}.$$

*It is called the (i,j)-**curvature**.*

Proof. It suffices to show that the map

$$\Gamma^{(i)}(\xi) \times \Gamma^{(j)}(\xi) \to \Gamma^{(i+j)}(\xi) \to \Gamma^{(i+j)}(\xi)/\Gamma^{(i+j-1)}(\xi)$$

is $C^\infty(M)$-linear. Here, the second map is the quotient.

Let $X \in \Gamma^{(i)}(\xi)$, $Y \in \Gamma^{(j)}(\xi)$, and f a smooth map. We have

$$[X, fY] = df(X)Y + f[X,Y].$$

Since $df(X)Y \in \Gamma^{(j)}(\xi) \subset \Gamma^{(i+j)}(\xi)$, we have

$$[X, fY] \equiv f[X,Y] \mod \Gamma^{(i+j-1)}(\xi).$$

By a similar computation, $[fX,Y] \equiv f[X,Y] \mod \Gamma^{(i+j-1)}(\xi)$. Hence, the map above is linear. $\square$

We know from Lemma 2.8 that any bracket expression $A(-)$ of length i can be reduced to a linear combination of bracket expressions of the form $[,B(-)]$ where $B(-)$ is a bracket expression of length $i-1$. This behaviour of bracket expressions translates to curvatures in the following way:

Lemma 2.31. *All curvatures are determined by the ones of the form*

$$\Omega_{(1,i)}(\xi) : \xi^{(1)} \times \left(\xi^{(i)}/\xi^{(i-1)}\right) \to \xi^{(i+1)}/\xi^{(i)}.$$

Proof. By Lemma 2.8, we can write any bracket expression $A(-)$ of length i as a linear combination of bracket expressions of the form $[, B(-)]$, where $B(-)$ is a bracket expression of length $i-1$. $\square$

The assumption of weak regularity is necessary for the curvatures to be well-defined. However, the rank of the curvature may vary with the point.

Example 2.32. Consider the manifold $\mathbb{R}^5(x_1, x_2, y_1, y_2, z)$ endowed with a distribution ξ given by

$$\xi = \ker(dz - y_1 dx_1 - f(y_2)dx_2) = \langle \partial_{y_1}, \partial_{y_2}, \partial_{x_1} + y_1\partial_z, \partial_{x_2} + f(y_2)\partial_z \rangle,$$

with $f : \mathbb{R} \to \mathbb{R}$ such that f is identically zero on the region $\{y_2 \geq 0\}$, and strictly increasing otherwise. The only non-trivial Lie brackets are:

$$\begin{aligned}
[\partial_{y_1}, \partial_{x_1} + y_1\partial_z] &= \partial_z \\
[\partial_{y_2}, \partial_{x_2} + f(y_2)\partial_z] &= f'(y_2)\partial_z.
\end{aligned}$$

We see that ξ is bracket-generating and weakly regular. In particular, $\Omega_{(1,1)}$ has maximal rank (namely, 4) if $y_2 < 0$, but rank 2 if $y_2 \geq 0$. $\triangle$

2.7 Filtered structures

As we have seen, a weakly regular, bracket-generating distribution induces a filtration of the tangent bundle through a sequence of distributions.

In Chapter 5 of the thesis, we will use filtered structures for weighted analysis. In Chapter 6, we will state and prove a version of Mitchell's Theorem for filtered structures, the main contribution of this thesis.

Definition 2.33. Let M be a manifold. A **filtered structure** on M is a sequence of distributions ξ_i on M with $\xi_i \subsetneq \xi_{i+1}$ for $i = 1, \dots, m-1$ such that

$$\xi_1 \subset \xi_2 \subset \dots \xi_m = TM$$

and $[\Gamma(\xi_i), \Gamma(\xi_j)] \subset \Gamma(\xi_{i+j})$.

Weakly regular bracket-generating distributions generate a special case of a filtered structure, in which the lowest-order distribution generates the other distributions in the filtered structure. Generally, this is not the case.

Example 2.34. Let (M, ξ) be a manifold endowed with a weakly regular, bracket-generating distribution. Let ξ_i be the distribution corresponding to $\Gamma^{(i)}(\xi)$, and let m be the step for which the Lie flag stabilises. Then

$$\xi = \xi_1 \subset \dots \subset \xi_m = TM$$

is a filtered structure on M. $\triangle$

We can think of vector fields as differential operators of order 1 acting on functions. In that manner, we can compose them to yield higher-order differential operators. When we use filtered structures, however, we want to regard certain vector fields as higher order operators. In some analytic settings, this will allow us to define calculi of differential operators in which the notion of ellipticity is different from the usual one.

Definition 2.35. Let $X \in \mathfrak{X}(M)$. Then X has **order** $-i$ if $X \in \Gamma(\xi_i)$, but $X \notin \Gamma(\xi_{i+1})$. We write: $\operatorname{ord}(X) = -i$.

Let $\operatorname{rank}(\xi_i) = n_i$. In particular $n_m = n$, because $\xi_m = TM$. Moreover, let $X_1, \dots, X_n$ be a framing of TM such that $X_1, \dots, X_{n_i}$ is a framing of ξ_i. Then by definition, we have

$$\begin{aligned} \operatorname{ord}(X_1) = \cdots = \operatorname{ord}(X_{n_1}) &= -1 \\ \operatorname{ord}(X_{n_1+1}) = \cdots = \operatorname{ord}(X_{n_2}) &= -2 \\ &\vdots \\ \operatorname{ord}(X_{n_{m-1}+1}) = \cdots = \operatorname{ord}(X_n) &= -m. \end{aligned}$$

We write $-a_i = \operatorname{ord}(X_i)$.

Remark 2.36. If $\xi_1 \subset \cdots \subset \xi_m$ is generated by a weakly regular bracket-generating distribution ξ, then the order of $X = A(X_1, \dots, X_{n_1})$ is related to the length of A. Indeed, $\operatorname{ord}(X) = -a$ if and only if the bracket-expression A has length a and cannot be reduced to a linear combination of bracket-expressions of length strictly smaller than a.

Normally, a Lie bracket is a differential operator of order 1, because it is a vector field. Even though it is a sum of operators of order 2, cancellations take place. By definition, a bracket expression is an iterated Lie bracket. △

Similar to weakly regular distributions, we can also define the curvature for filtered structures.

Lemma 2.37. *The Lie bracket yields a well-defined bundle morphism*

$$\Omega_{(i,j)} : (\xi_i/\xi_{i-1}) \times (\xi_j/\xi_{j-1}) \to \xi_{i+j}/\xi_{i+j-1}$$

called the (i,j)-curvature.

Proof. As in the proof of Lemma 2.30, it suffices to prove that the map

$$\Gamma(\xi_i) \times \Gamma(\xi_j) \to \Gamma(\xi_{i+j}) \to \Gamma(\xi_{i+j})/\Gamma(\xi_{i+j-1})$$

is C^∞-linear, where the second map is just the quotient map. So, let $X \in \Gamma(\xi_i)$, $Y \in \Gamma(\xi_j)$ and let $f \in C^\infty(M)$. We have

$$[X, fY] = df(X)Y + f[X,Y]$$

and $df(X)Y \in \Gamma(\xi_i) \subset \Gamma(\xi_{i+j})$, and therefore

$$[X, fY] \equiv f[X,Y] \mod \Gamma(\xi_{i+j-1}).$$

By a similar argument, we have $[fX, Y] \equiv f[X,Y] \mod \Gamma(\xi_{i+j-1})$. Hence, the map is $C^\infty(M)$-linear. □

For a distribution ξ we have that $\Gamma(\xi)$ generates all $\Gamma^{(i)}(\xi)$, and therefore all curvatures are determined by ones of the form $\Omega_{(1,i)}(\xi)$, as we have seen in Lemma 2.31. However, in the case of a filtered structure, $\Gamma(\xi_1)$ in general does not generate all $\Gamma(\xi_i)$, and therefore $\Omega_{(1,i)}$ do not necessarily generate all curvatures.

2.8 Graded Lie algebras

We want to put curvatures together into an algebraic object which is easier to deal with. This will be the nilpotentisation of a distribution. First, we need to introduce the notion of a graded Lie algebra.

Definition 2.38. A **graded Lie algebra** is a Lie algebra $(V, [,])$ such that

- V is a graded vector space with decomposition

$$V = \oplus_{i=1}^m V_i \oplus V_\infty$$

  for some non-negative integer m.

- The Lie bracket is compatible with the grading: any bracket expression involving V_∞ is zero, and $[V_i, V_j] \subset V_{i+j}$.

A Lie algebra is nilpotent if any sufficiently long bracket expression is zero. A graded Lie algebra is a special case of a nilpotent Lie algebra. More explicitly, we have:

Definition 2.39. A Lie algebra $(V, [,])$ is **nilpotent** if there is $m \in \mathbb{N}$ such that

$$A(v_1, \ldots, v_i) = 0$$

for all $v_1, \ldots, v_i \in V$, and $A(-)$ a bracket expression of length $i > m$.

Lemma 2.40. *A graded Lie algebra* $(V, [,])$ *is nilpotent.*

Proof. We prove by induction. By definition, we have $[V_i, V_j] \subset V_{i+j}$. Now, suppose that for $B(-)$ a bracket expression of length $i-1$, we have

$$B(V_{j_1}, \ldots, V_{j_{i-1}}) \subset V_{j_1+\cdots+j_{i-1}}.$$

Let $A(-)$ a bracket expression of length i. Without loss of generality, we may assume that $A(-) = [, B(-)]$, where $B(-)$ is a bracket expression of length $i-1$. Indeed, by a similar argument as Lemma 2.8, any bracket expression $A(-)$ of length i is a linear combination of bracket expressions of the form $[, B(-)]$, where $B(-)$ is a bracket expression of length $i-1$.

We then have

$$\begin{aligned} A(V_{j_1}, \ldots, V_{j_i}) &= [V_{j_1}, B(V_{j_2}, \ldots, V_{j_i})] \\ &\subset [V_{j_1}, V_{j_2 \ldots j_i}] \\ &\subset V_{j_1+\cdots+j_i}. \end{aligned}$$

Especially, if $A(-)$ is a bracket-expression of length $i > m$, we have

$$A(V_{j_1}, \ldots, V_{j_i}) = 0.$$

Hence, V is nilpotent. ☐

Remark 2.41. We can also build a graded Lie algebra with negative grading. In that case, it is a Lie algebra $(V, [,])$ such that

$$V = \oplus_{i=-m_1}^{m_2} V_i \oplus V_\infty$$

for some non-negative integers m_1, m_2. The Lie bracket then is compatible as before, i.e. any bracket expression involving V_∞ is zero, and $[V_i, V_j] \subset V_{i+j}$. However, it is not necessarily nilpotent. Since $v \in V$ can have negative or positive grading, we can keep increasing or decreasing the grade of a bracket expression so that we never land in V_∞. △

We will classify all Lie algebras of dimensions 2 and 3.

Example 2.42 (Heisenberg algebra)**.** The first non-trivial example of a graded Lie algebra appears in $\dim(V) = 3$, which is the **Heisenberg algebra**. It is defined by the relation $[e_1, e_2] = e_3$. We can choose many gradings, all of which are of the form:

$$\mathrm{gr}(e_1) = i,\ \mathrm{gr}(e_2) = j,\ \mathrm{gr}(e_3) = i + j.$$

△

Proposition 2.43. *Let* $(V, [,])$ *a graded Lie algebra, with* $\dim(V) = 3$*. Then* V *is either trivial or isomorphic to the Heisenberg algebra.*

Proof. If $(V, [,])$ has no non-trivial bracket expressions, then it is trivial.

Suppose $(V, [,])$ has one non-trivial bracket expression.

Claim 1. By possibly changing the basis, we have $[e_1, e_2] = e_3$.

Proof. Suppose $[e_1, e_2]$ is not linearly independent from e_1, e_2. Then $[e_1, e_2] = \lambda e_1 + \mu e_2$ for some $\lambda, \mu \in \mathbb{K}$ (where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ depending on whether V is a real or complex vector space). We have

$$[e_1, \lambda e_1 + \mu e_2] = \mu[e_1, e_2]$$

and similarly

$$[e_2, e_3] = -\lambda[e_1, e_2].$$

However, we have by the Jacobi identity that

$$\begin{aligned} [e_1, [e_1, e_2]] &= [e_1, [e_2, e_1]] - [e_2, [e_1, e_1]] \\ &= [-e_1, [e_1, e_2]] \\ &= -[e_1, e_3] \end{aligned}$$

and therefore $[e_1, [e_1, e_2]] = 0$. By a similar calculation, it follows that $[e_2, [e_1, e_2]] = 0$. It follows that $\mu = \lambda = 0$. However, this contradicts the assumption that $[e_1, e_2]$ is non-trivial. Hence, by possibly changing the basis we have $[e_1, e_2] = e_3$. □

So, V is isomorphic to the Heisenberg algebra.

Now, suppose that $(V, [,])$ has two or more non-trivial bracket expressions. By possibly changing the basis, we have $[e_1, e_2] = e_3$ and $[e_2, e_3] \neq 0$. If $\mathrm{gr}(e_1) = i$ and $\mathrm{gr}(e_2) = j$, then $\mathrm{gr}(e_3) = i + j$ by the first bracket expression. However, then

$$\mathrm{gr}([e_2, e_3]) = \mathrm{gr}(e_2) + \mathrm{gr}(e_3) = i + 2j > i + j.$$

This contradicts the grading. So, $(V, [,])$ is either trivial or isomorphic to the Heisenberg algebra. □

Example 2.44 (Engel algebra)**.** In $\dim(V) = 4$ there is a non-trivial graded Lie algebra defined by $[e_1, e_2] = e_3$, $[e_1, e_3] = e_4$, called the **Engel algebra**. Again, there are many possible gradings, all of which are of the form:

$$\mathrm{gr}(e_1) = i,\ \mathrm{gr}(e_2) = j,\ \mathrm{gr}(e_3) = i + j,\ \mathrm{gr}(e_4) = 2i + j.$$

△

Proposition 2.45. *Let* $(V, [,])$ *a graded Lie algebra with* $\dim(V) = 4$*. Then* V *is either trivial, isomorphic to the Engel algebra, or can be decomposed into the Heisenberg algebra and a trivial part.*

Proof. If $(V, [,])$ has no non-trivial bracket expressions, then it is trivial.

Suppose $(V, [,])$ has one non-trivial bracket expression. Then by possibly changing the basis, we have $[e_1, e_2] = e_3$. So, V can be decomposed into the Heisenberg algebra and a trivial part.

Suppose $(V, [,])$ has two non-trivial bracket expressions. By possibly changing the basis, we have $[e_1, e_2] = e_3$ and $[e_1, e_3] = e_4$. So V is isomorphic to the Engel algebra.

Now suppose that $(V, [,])$ has three or more non-trivial bracket expressions. Then by possibly changing the basis, we have $[e_1, e_2] = e_3$ and $[e_1, e_3] = e_4$, and some other non-trivial bracket expression. By the linearity of the Lie bracket, the last non-trivial bracket expression is either $[e_2, e_3]$ or $[e_k, e_4]$ for $k \in \{1, 2, 3\}$. If $\mathrm{gr}(e_1) = i$ and $\mathrm{gr}(e_2) = j$, then $\mathrm{gr}(e_3) = i + j$ and $\mathrm{gr}(e_4) = 2i + j$. If $[e_k, e_4] \neq 0$ for $k \in \{1, 2, 3\}$, then

$$\mathrm{gr}([e_k, e_4]) = \mathrm{gr}(e_k) + \mathrm{gr}(e_4) > 2i + j$$

which contradicts the grading. If $[e_2, e_3] \neq 0$, then $\mathrm{gr}([e_2, e_3]) = i + 2j$. It follows that $i = j = 1$, and $[e_2, e_3] = \lambda e_4$ for some $\lambda \in \mathbb{R} \setminus \{0\}$. Then we have

$$0 = \lambda[e_1, e_3] - [e_2, e_3] = [\lambda e_1 - e_2, e_3].$$

In that case, we should change the basis to $e_1, \lambda e_1 - e_2, e_3, e_4$. In that case, $(V, [,])$ is isomorphic to the Heisenberg algebra.

So, $(V, [,])$ is either trivial, isomorphic to the Engel algebra, or it can be decomposed into the Heisenberg algebra and a trivial part. □

We will now take the notion of a graded Lie algebra to a bundle-theoretical framework.

Definition 2.46. A **weak bundle of graded Lie algebras** is a pair $(L \to M, [,])$ satisfying

- $L \to M$ is a graded vector bundle, i.e. $L = \oplus_{i=1}^m L_i \oplus L_\infty$.
- The bracket $[,]$ is an antisymmetric tensor from L to L such that it turns the fibres of L into graded Lie algebras. Moreover, the pairing $[,]$ depends smoothly on the base point in M.

If $(L \to M, [,])$ is locally trivial (i.e. modelled on a single Lie algebra), it is called a **bundle of graded Lie algebras**.

2.9 Nilpotentisation

Given a distribution on a manifold, we can pack the information given by the growth vector and the curvatures into a single algebraic object, as we alluded to before. This is the nilpotentisation. We can do the same for filtered structures.

Definition 2.47. Given (M, ξ) a manifold with a weakly regular distribution. Let $m \in \mathbb{N}$ the minimal integer such that $\Gamma^{(m)}(\xi) = \Gamma^{(i)}(\xi)$ for all $i \geq m$. The **nilpotentisation** associated to ξ is the weak bundle of graded Lie algebras given by

$$L(\xi) = \oplus_{i=1}^{m} L_i(\xi) \oplus L_\infty(\xi)$$

where $L_1(\xi) = \xi^{(1)}$, $L_i(\xi) = \xi^{(i)}/\xi^{(i-1)}$, and $L_\infty(\xi) = TM/\xi^{(m)}$. Here, the Lie bracket is the curvature.

Similarly, given M a manifold endowed with a filtered structure $\xi_1 \subset \cdots \subset \xi_m$. The **nilpotentisation** associated to the filtered structure is the weak bundle of graded Lie algebras given by

$$L(\xi_1 \subset \cdots \subset \xi_m) = \oplus_{i=1}^{m} L_i(\xi_1 \cdots \subset \xi_m),$$

where $L_1(\xi_1 \subset \cdots \subset \xi_m) = \xi_1$ and $L_i(\xi_1 \subset \cdots \subset \xi_m) = \xi_i/\xi_{i-1}$.

In general, the nilpotentisation is just a weak bundle of graded Lie algebras.

Definition 2.48. A distribution ξ is **regular** if the nilpotentisation $L(\xi)$ is a bundle of graded Lie algebras.

Similarly, a filtered structure $\xi_1 \subset \cdots \subset \xi_m$ is **regular** if $L(\xi_1 \subset \cdots \subset \xi_m)$ is a bundle of graded Lie algebras.

2.10 Lie group multiplication: the Campbell-Baker-Hausdorff formula

As we have seen, the nilotentisation is a weak bundle of graded Lie algebras. In particular, this means that the restriction of the nilpotentisation to a point is a graded Lie algebra. Lie's Third Theorem (see [2]) there exists a corresponding simply-connected Lie group whose Lie algebra is the nilpotentisation at a point. More generally, Lie's Third Theorem states that for any finite-dimensional real Lie algebra, there exists a corresponding simply-connected Lie group.

More concretely, let $\mathfrak{g}$ a Lie algebra. Then the elements of the corresponding Lie group G are given by $\exp(X)$ for $X \in \mathfrak{g}$. Given X, $Y \in \mathfrak{g}$ we want to find $Z \in \mathfrak{g}$ such that

$$\exp(X)\exp(Y) = \exp(Z)$$

in order to define the group multiplication. The Campbell-Baker-Hausdoff Formula then gives an explicit solution for $Z = \log(\exp(X)\exp(Y))$.

Theorem 2.49 (Campbell-Baker-Hausdorff formula)**.** *The solution Z for $\exp(X)\exp(Y) = \exp(Z)$ can be written as a formal series of bracket expressions of X and Y. The first few terms of the series are*

$$\log(\exp(X)\exp(Y)) = X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}[X, [X, Y]] - \frac{1}{12}[Y, [X, Y]] + \ldots$$

See [4] and [10] for an explicit formula of the series, and a proof.

In the case of a nilpotent Lie algebra $\mathfrak{g}$, this series is always finite. Therefore, the series converges over the whole Lie algebra. Hence, the space underlying the Lie group is Euclidean. This means that, as a space, the Lie group is equal to the Lie algebra, which is a Euclidean space.

For ξ a weakly regular bracket-generating distribution, the Lie group with Lie algebra $L(\xi)(p)$ is a Carnot group.

Definition 2.50. A **Carnot group** is a Lie group with a graded Lie algebra of the form $\mathfrak{g} = \oplus_{i=1}^{m} V_i$, where every V_i is spanned by bracket expressions of elements of V_1.

3 Some control theory: the endpoint map and Chow's Theorem

In this chapter we prove Chow's Theorem 3.13, which tells us that for a manifold with a bracket-generating distribution, we can connect any two points on the manifold by a path tangent to the distribution. We have seen this before in the rolling coin example in the Introduction (Chapter 1): the coin could reach any point on the table because the distribution modelling the coin's movement was bracket-generating.

We will prove Chow's Theorem using an endpoint map, which is defined in Section 3.2. An endpoint map sends an L^2-function to a point in M. The endpoint map is a control problem, in which the L^2-function dictates the path. It returns the endpoint of the path. For Chow's Theorem, we will use a specific version of the endpoint map using the flows of vector fields in the framing of the distribution and commutators of those flows. These paths are all tangent to the distribution, and will approximate the flows of bracket expressions of those vector fields. So, to prove Chow we will create an explicit path.

In section 3.4, we will define the Carnot-Carathéodory distance. Similar to a Riemannian distance, it measures the distance between two points on a manifold by shortest paths. In the Carnot-Carathéodory distance, however, we restrict to paths tangent to the distribution. Hence, it measures the distance between two points through geodesics which are tangent to the distribution. It is not a Riemannian distance, although it does arise from one. If two points cannot be connected by a path tangent to the distribution, their Carnot-Carathéodory distance is infinite. By Chow's Theorem, we can connect any two points by a path tangent to the distribution, as long as the distribution is bracket-generating. Therefore, the Carnot-Carathéodory distance between any two points is always finite for a bracket-generating distribution.

This chapter is based on [8] (chapter 2), [1] and [3]. For more on Carnot-Carathéodory metrics, see [7].

3.1 Flows

We will first discuss the relationship between the flow of a bracket expression of vector fields and the commutator of the corresponding flows.

Definition 3.1. Let X be a vector field. Then we will write ϕ_X^t for the **flow** of X.

Let us start with the easiest case of a bracket expression, namely one of length 2. For two vector fields X, Y we can approximate the flow of $[X, Y]$ by the commutator of flows $\left[\phi_X^{\sqrt{t}}, \phi_Y^{\sqrt{t}}\right]$.

Proposition 3.2. *Let X, $Y \in \mathfrak{X}(M)$ be two vector fields. Then $\phi_{[X,Y]}^t = [\phi_X^{\sqrt{t}}, \phi_Y^{\sqrt{t}}] + o(t)$.*

Proof. We will show that $\phi_{[X,Y]}^{t^2} = [\phi_X^t, \phi_Y^t] + o(t^2)$, using local coordinates. Let $p \in M$. W first consider the Taylor expansion of $\phi_{[X,Y]}^{t^2}(p)$ around $t = 0$, which is

$$\phi_{[X,Y]}^{t^2}(p) = \phi_{[X,Y]}^{t=0}(p) + t^2[X, Y]_p + o(t^2).$$

On the other hand, we have by partial differentiation that

$$\left.\frac{d}{dt}\right|_{t=0}[\phi_X^t,\phi_Y^s](p) = \left.\frac{d}{dt}\right|_{t=0}\phi_X^t\circ\phi_Y^s\circ\phi_X^{-t}\circ\phi_Y^{-s}(p)$$
$$= X_p - \phi_Y^s\circ X_p\circ\phi_Y^{-s}(p) + o(t^2).$$

Differentiating again, we get

$$\left.\frac{d}{ds}\right|_{s=0} X_p - \phi_Y^s\circ X_p\circ\phi_Y^{-s}(p) + o(t^2) = [X,Y]_p + o(s^2).$$

Hence, we have

$$[\phi_X^t,\phi_Y^t](p) = \left[\phi_X^{t=0},\phi_Y^{t=0}\right](p) + t^2[X,Y]_p + o(t^2)$$
$$= \phi_{[X,Y]}^{t=0}(p) + t^2[X,Y]_p + o(t^2).$$

Hence, $\phi_{[X,Y]}^{t^2} = [\phi_X^t,\phi_Y^t] + o(t^2)$, and therefore $\phi_{[X,Y]}^t = \left[\phi_X^{\sqrt{t}},\phi_Y^{\sqrt{t}}\right] + o(t)$. ∎

We can generalise this result for bracket expressions of arbitrary length.

Proposition 3.3. *Let $A(-)$ a bracket-expression of length i, and let $X_1,\ldots,X_i\in\mathfrak{X}(M)$ be vector fields on M. Then we have*

$$\phi_{A(X_1,\ldots,X_i)}^t = A\left(\phi_{X_1}^{t^{1/i}},\ldots,\phi_{X_i}^{t^{1/i}}\right) + o(t).$$

Proof. We will prove the statement by induction. For $i=1$, the result is trivial. In Proposition 3.2, we proved the case for $i=2$. Now, suppose that the result holds $i-1$. By Lemma 2.8, we can write $A(-)$ as a linear combination of bracket-expressions of the form $[,B(-)]$ for $B(-)$ a bracket-expression of length $i-1$. Without loss of generality, assume that $A(X_1,\ldots,X_i) = [X_1, B(X_2,\ldots,X_i)]$ for some bracket-expression $B(-)$ of length $i-1$. By the induction hypothesis, we have

$$\phi_{B(X_2,\ldots,X_i)}^t = B\left(\phi_{X_2}^{t^{1/(i-1)}},\ldots,\phi_{X_i}^{t^{1/(i-1)}}\right) + o(t).$$

Therefore, we have

$$A(\phi_{X_1}^t,\ldots,\phi_{X_i}^t) = \left[\phi_{X_1}^t, B\left(\phi_{X_2}^t,\ldots,\phi_{X_i}^t\right)\right]$$
$$= \left[\phi_{X_1}^t, \phi_{B(X_2,\ldots,X_i)}^{t^i} + o(t^{i-1})\right]$$
$$= \left[\phi_{X_1}^t, \phi_{B(X_2,\ldots,X_i)}^{t^i}\right] + o(t^i).$$

Let $p\in M$. As in the previous proposition, we will use the Taylor expansion at $t=0$. We have

$$\left[\phi_{X_1}^t,\phi_{B(X_2,\ldots,X_i)}^{s^{i-1}}\right](p) = \left.\frac{d}{dt}\right|_{t=0}\phi_{X_1}^t\circ\phi_{B(X_2,\ldots,X_i)}^{s^{i-1}}\circ\phi_{X_1}^{-t}\circ\phi_{B(X_2,\ldots,X_i)}^{-s^{i-1}}(p)$$
$$= (X_1)_p - \phi_{B(X_2,\ldots,X_i)}^{s^{i-1}}\circ(X_1)_p\circ\phi_{B(X_2,\ldots,X_i)}^{-s^{i-1}}(p) + o(t^2).$$

Differentiating again, we get

$$\begin{aligned}
&\left.\frac{d}{ds}\right|_{s=0} (X_1)_p - \phi^{s^{i-1}}_{B(X_2,\ldots,X_i)} \circ (X_1)_p \circ \phi^{-s^{i-1}}_{B(X_2,\ldots,X_i)}(p) + o(t^2)\\
&= [X_1, B(X_2,\ldots,X_i)]_p + o(s^i)\\
&= A(X_1,\ldots,X_i)_p + o(s^i).
\end{aligned}$$

So, we have $\phi^{t^i}_{A(X_1,\ldots,X_i)} = A(\phi^t_{X_1},\ldots,\phi^t_{X_i}) + o(t^i)$, and therefore $\phi^t_{A(X_1,\ldots,X_i)} = A(\phi^{t^{1/i}}_{X_1},\ldots,\phi^{t/i}_{X_i}) + o(t)$. $\square$

With this, we can now define the endpoint map.

3.2 The endpoint map

Given (M,ξ) a manifold with a distribution, and $p \in M$. We can ask which points on the manifold we can reach through a path tangent to the distristribution which starts at p. Chow's Theorem 3.13 tells us that if ξ is bracket-generating, then we can reach any point in the manifold through such a path. In order to prove Chow's Theorem, and to solve the question which points we can reach by a path tangent to the distribution, we will first look at which points near p we can reach by a path tangent to ξ. For this, we will use a special version of the endpoint map.

In this section, we let $\{X_1,\ldots,X_k\}$ a framing of ξ centred at p. We write ϕ^t_i for the flow of X_i at time t.

Definition 3.4. The **endpoint map** based at p is the map

$$\begin{aligned}
L^2([0,1],\mathbb{R}^k) &\to M\\
u &\mapsto x(1)
\end{aligned}$$

where $x(t)$ is the unique solution of the local control problem given by

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^k u_i(t) X_i(x(t)) \\ x(0) = p. \end{cases}$$

As we said, we are specifically interested in paths which are tangent to the distribution. By choosing u carefully, we can create such paths through solving the control problem. We know from the framing that we can reach the points of the form $q = \phi^{t_i}_i(p)$. We can express these explicitly using the endpoint map as follows:

Definition 3.5. Let $\gamma^{t_i}_i : [0,1] \to \mathbb{R}^k$ with

$$\begin{aligned}
\gamma^{t_1}_1(t) &= (t_1, 0, \ldots, 0)\\
\gamma^{t_2}_2(t) &= (0, t_2, 0, \ldots, 0)\\
&\vdots\\
\gamma^{t_k}_k(t) &= (0, \ldots, 0, t_k).
\end{aligned}$$

Then $\phi^{t_i}_i(p) = x(1)$, where $x(t)$ is the solution of

$$\begin{cases} \dot{x}(t) = \sum_{j=1}^k \left(\gamma^{t_i}_i\right)_j (t) X_j(x(t)) = t_i X_i(x(t)) \\ x(0) = p. \end{cases}$$

We can even reach the points of the form $q = \left[\phi_i^t, \phi_j^t\right](p)$. In order to do this, we first define the commutator of paths.

Definition 3.6. Let $u, v : [0,1] \to \mathbb{R}^k$ be paths, given by $u(t) = (u_1(t), \ldots, u_k(t)$ and $v(t) = (v_1(t), \ldots, v_k(t))$. We define the **concatenation of paths** as follows:

$$u \centerdot v(t) = \begin{cases} 2u(t) & \text{if } t \in [0, \frac{1}{2}] \\ 2v(t) & \text{if } t \in [\frac{1}{2}, 1]. \end{cases}$$

We define the **commutator of paths** as

$$[u, v] = u \centerdot v \centerdot (-u) \centerdot (-v).$$

Remark 3.7. The concatenation of paths is not the same as the usual concatenation in algebraic topology for the product in the fundamental group or the fundamental groupoid. In the definition of the concatenation of paths we are adding a dilation, because these paths are the derivatives of the paths which are of interest to us. The derivatives do get dilated when we run over the path faster. $\triangle$

In the next proposition, we see how the commutator of flows can be described as a solution of a differential equation involving the commutator of paths. We will later use a generalisation of this to define the endpoint map.

Proposition 3.8. *The commutator of flows $[\phi_i^t, \phi_j^t](p)$ equals $x(1)$, where $x(s)$ is the solution to the differential equation*

$$\begin{cases} \dot{x}(s) = \sum_{l=1}^k [\gamma_i^t, \gamma_j^t]_l(s) X_l(x(s)) \\ x(0) = p. \end{cases}$$

Proof. We have

$$\gamma_i^t \centerdot \gamma_j^t \centerdot \gamma_i^{-t} \centerdot \gamma_j^{-t} = ((\gamma_i^t \centerdot \gamma_j^t) \centerdot \gamma_i^{-t}) \centerdot \gamma_j^{-t},$$

and therefore

$$\gamma_i^t \centerdot \gamma_j^t \centerdot \gamma_i^{-t} \centerdot \gamma_j^{-t}(s) = \begin{cases} \gamma_i^{8t}(8s) & \text{if } s \in [0, \frac{1}{8}] \\ \gamma_j^{8t}(8s-1) & \text{if } s \in [\frac{1}{8}, \frac{1}{4}] \\ \gamma_i^{-4t}(4s-1) & \text{if } s \in [\frac{1}{4}, \frac{1}{2}] \\ \gamma_j^{-2t}(2s-1) & \text{if } s \in [\frac{1}{2}, 1]. \end{cases}$$

In order to get a reparametrisation of the integrated path, we reparametrise and rescale this path as follows:

$$\gamma(s) = \begin{cases} \gamma_i^{4t}(4s), & \text{if } s \in [0, \frac{1}{4}] \\ \gamma_j^{4t}(4s-1) & \text{if } s \in [\frac{1}{4}, \frac{1}{2}] \\ \gamma_i^{-4t}(4s-2) & \text{if } s \in [\frac{1}{2}, \frac{3}{4}] \\ \gamma_j^{-4t}(4s-3) & \text{if } s \in [\frac{3}{4}, 1]. \end{cases}$$

Hence, $[\phi_i^t, \phi_j^t](p)$ is indeed a solution of

$$\dot{x}(s) = \sum_{l=1}^k [\gamma_i^t, \gamma_j^t]_l(s) X_l(x(s)),$$

where $x(0) = p$. $\square$

We can generalise this result for bracket expressions of any length:

Proposition 3.9. *Let $A(-)$ a bracket expression of length i, then $A(\phi^t_{j_1}, \dots, \phi^t_{j_i})(p) = x(1)$, where $x(s)$ is the solution of the differential equation*

$$\begin{cases} \dot{x}(s) = \sum_{j=1}^k A(\gamma^t_{j_1}, \dots, \gamma^t_{j_i})_j(s) X_j(x(s)) \\ x(0) = p. \end{cases}$$

Proof. We prove by induction. It follows by Proposition 3.8 that the statement is true for bracket expressions of length 2.

Now suppose the statement is true for bracket expressions of length $i-1$. By a similar argument as Lemma 2.8 we can write $A(-)$ as a linear combination of bracket expressions of the form $[, B(-)]$, with $B(-)$ a bracket expression of length $i-1$. Without loss of generality, assume that $A(-) = [, B(-)]$. We have

$$[\gamma^t_{j_1}, B(\gamma^t_{j_2}, \dots, \gamma^t_{j_i})](s) = \begin{cases} \gamma^{8t}_{j_1}(8s) & \text{if } s \in [0, \frac{1}{8}] \\ B(\gamma^{8t}_{j_2}, \dots, \gamma^{8t}_{j_i})(8s-1) & \text{if } s \in [\frac{1}{8}, \frac{1}{4}] \\ \gamma^{-4t}_{j_1}(4s-1) & \text{if } s \in [\frac{1}{4}, \frac{1}{2}] \\ B(\gamma^{-2t}_{j_2}, \dots, \gamma^{-2t}_{j_i})(2s-1) & \text{if } s \in [\frac{1}{2}, 1]. \end{cases}$$

To obtain a reparametrisation of the integrated path, we reparametrise and rescale as follows:

$$\gamma(s) = \begin{cases} \gamma^{4t}_{j_i}(4s) & \text{if } s \in [0, \frac{1}{4}] \\ B(\gamma^{4t}_{j_2}, \dots, \gamma^{4t}_{j_i})(4s-1) & \text{if } s \in [\frac{1}{4}, \frac{1}{2}] \\ \gamma^{-4t}_{j_1}(4s)(4s-2) & \text{if } s \in [\frac{1}{2}, \frac{3}{4}] \\ B(\gamma^{-4t}_{j_2}, \dots, \gamma^{-4t}_{j_i})(4s-3) & \text{if } s \in [\frac{3}{4}, 1]. \end{cases}$$

So then $A(\phi^t_{j_1}, \dots, \phi^t_{j_i})(p) = [\phi^t_{j_1}, B(\phi^t_{j_2}, \dots, \phi^t_{j_i})](p)$ is the solution of the differential equation

$$\dot{x}(s) = \sum_{j=1}^k A(\gamma^t_{j_1}, \dots, \gamma^t_{j_i})_j(s) X_j(x(s))$$

with $x(0) = p$. □

3.3 Chow's theorem

We now come to the main result of this chapter, which is Chow's Theorem. It tells us that for M a smooth manifold, endowed with a bracket-generating distribution ξ, we can connect any two points on the manifold by a path tangent to the distribution.

The following remark tells us that the bracket-generating condition is indeed necessary.

Remark 3.10. In general, we cannot connect any two points on a manifold by a path tangent to the distribution. For example, the result does not hold for an involutive distribution. If ξ is involutive, then Frobenius's Theorem 2.27 tells us it is also integrable. This means that the paths tangent to ξ through a fixed point $p \in M$ form a smooth immersed manifold of dimension $k = \text{rank}(\xi)$, which is called the **leaf** through p. Then, if $q \in M$ does not lie on the leaf through p, we cannot connect them via a path tangent to ξ. △

Before we state Chow's Theorem, we first define a special case of the endpoint map. We restrict the input of paths u so that we can only use paths which are concatenations of paths using the framing of the distribution. This allows us to only move in directions which are tangent to the distributions.

Definition 3.11. Let (M, ξ) a manifold of dimension n with a bracket-generating distribution of rank k, and let $p \in M$. Let $X_1, \ldots, X_k$ be a framing of ξ around p. For $k+1 \leq i \leq n$, let $X_i = A_i(X_1, \ldots, X_k)$ such that $X_1, \ldots, X_n$ forms a framing of TM around p. Here A_i are bracket expressions of length a_i. Write

$$\phi_i^t = \begin{cases} \phi_{X_i}^t & \text{if } 1 \leq i \leq k \\ A_i\left(\phi_{X_1}^{t^{1/a_i}}, \ldots, \phi_{X_k}^{t^{1/a_i}}\right) & \text{if } k+1 \leq i \leq n. \end{cases}$$

Then the **endpoint map adapted to the framing** $X_1, \ldots, X_n$ is defined as:

$$\begin{aligned} \psi : \mathbb{R}^n &\to M \\ (t_1, \ldots, t_n) &\mapsto \phi_n^{t_n} \circ \cdots \circ \phi_2^{t_2} \circ \phi_1^{t_1}(p). \end{aligned}$$

Remark 3.12. This is a special case of the original endpoint map. Indeed, we can write it as a control problem, where $\psi(t_1, \ldots, t_n) = x(1)$ is the solution to

$$\begin{cases} \dot{x}(t) = \sum_{i=1}^n u(t) X_i(x(t)) \\ x(0) = p \end{cases}$$

with

$$u(t) = \sum_i^k \gamma_i^{t_i}(t) + \sum_{j=k+1}^n A_j\left(\gamma_1^{t_j^{1/a_j}}, \ldots, \gamma_n^{t_j^{1/a_j}}\right)(t).$$

Through this, we effectively reduce the domain of the original endpoint map to a finite-dimensional subspace, which we identify with $\mathbb{R}^k$. The domain of the original endpoint map was $L^2([0,1], \mathbb{R}^k)$, which is an infinite-dimensional space. △

Theorem 3.13 (Chow)**.** *Let (M, ξ) be a manifold of dimension n with a bracket-generating distribution of rank k. Then any two points $p, q \in M$ can be connected by a path tangent to ξ.*

Proof of Theorem 3.13. Let $X_1, \ldots, X_n$ and $\phi_1^t, \ldots, \phi_n^t$ be as in Definition 3.11. By Proposition 3.3, $\phi_j^t = \phi_{X_j}^t + o(t)$.

We claim that for any j, the points p and $\phi_j^t(p)$ can be connected by a path tangent to ξ. Indeed, for $j \leq k$, the flow lines of ϕ_j^t are tangent to ξ. If $j > k$, then $\phi_j^t(p)$ can be reached from p by flowing iteratively along each element in the bracket expression that defines ϕ_j^t.

We use the endpoint map ψ at p. For any point $p' \in M$, the flows $\phi_{X_j}^t(p')$ at p' are C^∞-functions in the variable t. Therefore, for $j = 1, \ldots, n$ we have that $\phi_j^t(p')$ are also C^∞-functions, since they are either a vector field flow at p', or a bracket-expression of flows at p'. It follows that ψ is also C^∞, since it is the composition of C^∞-functions. Moreover, $\psi_*(\partial/\partial t_j)|_{(t_1,\ldots,t_n)=(0,\ldots 0)} = X_j$ for $j = 1, \ldots, n$. By the inverse function theorem, ψ is a C^∞-diffeomorphism near the origin. By construction, p can be connected to any point in the image of ψ with a path tangent to ξ. We can construct an endpoint map for any point $p \in M$, so any two points in M can be connected by a piecewise linear map tangent to ξ. □

For an alternative proof of Chow's Theorem, see [8].

A direct consequence of Chow's Theorem is the following:

Corollary 3.14. *Let (M, ξ) be a manifold with a bracket-generating distribution. Let $p, q \in M$, and let γ a path connecting p and q. Then for every $\epsilon > 0$, there is a path γ_ϵ tangent to ξ which connects p and q, and $||\gamma - \gamma_\epsilon||_{\sup} < \epsilon$.*

Proof. Let $\epsilon > 0$. Let $\gamma : [0, 1] \to M$, $\gamma(0) = p$, $\gamma(1) = q$. Divide $[0, 1]$ into intervals of length $1/N_\epsilon$ for some integer $N_\epsilon > 1/\epsilon$.
For every $i \in \{0, \ldots, N_\epsilon - 1\}$, let $U_i \subset M$ be a neighbourhood of $\gamma([i/N_\epsilon, (i + 1)/N_\epsilon])$ such that for all $p_1 \in \gamma([i/N_\epsilon, (i + 1)/N_\epsilon])$, $p_2 \in U_2$ we have $||p_1 - p_2|| < \epsilon$. Then $\cup_{i=0}^{N_\epsilon - 1}$ covers $\gamma([0, 1])$, and every U_i is a manifold with bracket-generating distribution $\xi|_{U_i}$.
By Chow's Theorem, we can connect the points $\gamma(i/N_\epsilon)$, $\gamma((i+1)/N_\epsilon) \in U_i$ with a path $\tilde{\gamma}_i$ tangent to $\xi|_{U_i}$. Concatenating these paths, we get a path $\gamma_\epsilon = \tilde{\gamma}_0 * \cdots * \tilde{\gamma}_{N_\epsilon - 1}$ which connects $\gamma_\epsilon(0) = p$ and $\gamma_\epsilon(1) = q$, which is tangent to ξ, and such that $||\gamma - \gamma_\epsilon||_{\sup} < \epsilon$. □

The following example shows that the result of Chow's Theorem can hold without the assumption that the distribution is bracket-generating.

Example 3.15. Consider $\mathbb{R}^3(x, y, z)$ with distribution

$$\xi = \ker(dz + f(y)dx) = \langle \partial_x - f(y)\partial_z, \partial_y \rangle ,$$

where $f : \mathbb{R} \to \mathbb{R}$ such that $f(y) = 0$ if $y \leq 0$, and $f'(y) > 0$ if $y > 0$. We have

$$[\partial_x - f(y)\partial_z, \partial_y] = f'(y)\partial_z.$$

Hence, ξ is a foliation on $\{y \leq 0\}$, and is bracket-generating on $\{y > 0\}$.

Consider two points (x, y, z) and (x', y', z') with $y, y' \leq 0$. We will show that we can connect these points by a path tangent to ξ.

Start at (x, y, z) and then move to $(x, 0, z)$ via a straight path (which is contained in the leaf). Then move from $(x, 0, z)$ to $(x', 0, z')$ on $\{y \geq 0\}$ via Chow's theorem. Lastly, move from $(x', 0, z')$ to (x', y', z') via a straight path, which is contained in the leaf. △

3.4 A note on Carnot-Carathéodory metrics

Given a manifold with a distribution, we can ask which pairs of points we can connect through paths tangent to the distribution. As we have seen in Chow's Theorem, we can connect any two points by a path tangent to the distribution if the distribution is bracket-generating. However, this is not true in general. The Carnot-Carathéodory is defined as the infimum over all paths tangent to the distribution which connect a given pair of points. If this distance is infinite, no such paths exist.

Definition 3.16. Let (M, ξ) be a manifold endowed with a distribution, and let g be a Riemannian metric on M. The **Carnot-Carathéodory distance** between two points p, $q \in M$ is defined as

$$d_c(p, q) = \inf_{\gamma \in C_{p,q}} \{\text{length}(\gamma)\},$$

where $C_{p,q}$ is the set of all curves tangent to the distribution ξ which join p, q. Here, length(γ) indicates the length with respect to g.

In the case of a bracket-generating distribution, then $d_c(p, q)$ is always finite. This follows directly from Chow's Theorem 3.13. As we mentioned in Section 2.10, there is a Carnot group whose Lie algebra is the nilpotentisation at the point. Given a Riemannian metric g on the manifold, we can endow the first level ξ of the nilpotentisation with the metric g, and thus build a bundle of Carnot groups.

Control theory deals with understanding the paths that minimise the Carnot-Carathéodory distance. For more one these minimising paths, see [8] Chapters 1 and 3 and Appendix E.

4 Weighted analysis

Given a filtered structure on a manifold, we can introduce weights to the vector fields on M to partly 'remember' the structure of the nilpotentisation. The filtration at the bundle level given by a filtered structure defines a filtration at the level of sections.

As we mentioned before in Section 2.7 in the discussion before Definition 2.35, the order of vector fields, that we can think of vector fields as differential operators of order 1. However, in the context of filtered structures, we regard certain vector fields as higher-order differential operators.

Concretely, if $\xi_1 \subset \cdots \subset \xi_m$ is a filtered structure on a smooth manifold M, we say that $X \in \mathfrak{X}(M)$ has weight i if $X \in \Gamma(\xi_i)$ and $X \notin \Gamma(\xi_{i+1})$. Using the notion of weight, we can define the order of smooth functions on M, by keeping track of the weights of the vector fields for which the directional derivative of the function vanishes.

Furthermore, we will introduce the notion of privileged coordinates, which are coordinate functions on the manifold which 'remember' the filtered structure. This will give another (equivalent) way to calculate the order of a function.

Lastly, we will define dilations with respect to the privileged coordinates, and introduce the notion of homogeneous vector fields. This will allow us to find a basis for the nilpotentisation.

So, let M a smooth manifold of dimension n, and let $\xi_1 \subset \cdots \subset \xi_m = TM$ a filtered structure. Let n_i the rank of the distribution ξ_i. In particular, $n_m = n$. We can choose a framing $X_1, \ldots, X_n$ of TM such that for all i we have that $X_1, \ldots, X_{n_i}$ is a framing of ξ_i.

This chapter is based on [1], [6], and chapters 2 and 8 of [8] (distributions) and [9] (filtered structures).

4.1 Nonholonomic derivatives, orders

Using the vector fields in the framing, we can introduce nonholonomic derivatives. It is a kind of directional derivative, but one which remembers the order of the vector fields which define the direction.

For all i, let the integer a_i be such that $X_i \in \Gamma(\xi_{a_i})$ and $X_i \notin \Gamma(\xi_{a_i+1})$.

Definition 4.1. Let $p \in M$ and let f be a smooth function defined on a neighbourhood of p. A **nonholonomic derivative of f of order** a is defined as

$$X_{i_1} \ldots X_{i_j} f$$

where $a_{i_1} + \cdots + a_{i_j} = a$.

Using the notion of nonholonomic derivatives, we can define the order of smooth functions at a point.

Definition 4.2. The **order** of f at p is greater than or equal to s if all nonholonomic derivatives of f of order smaller than or equal to $s-1$ vanish at p. We write $\mathrm{ord}_p(f) \geq s$. Moreover, $\mathrm{ord}_p(f) = s$ whenever $\mathrm{ord}_p(f) \geq s$ but not $\mathrm{ord}_p(f) \geq s+1$.

We can also define the order of differential operators and vector fields. For differential operators, we can express it in the vector fields of the framing.

Proposition 4.3. *If D is a linear differential operator, we can write D as*

$$D = \sum_{\alpha} c_\alpha X_1^{\alpha_1} \ldots X_n^{\alpha_n}$$

where the sum is taken over all multi-indices $\alpha = (\alpha_1, \dots, \alpha_n)$, and $c_\alpha \in C^\infty(M, \mathbb{R})$ is non-zero for finitely many terms.

Definition 4.4. A vector field X has **order** greater than or equal to $-s$ whenever $X \in \Gamma(\xi_s)$. We write $\mathrm{ord}(X) \geq -s$. Moreover, $\mathrm{ord}(X) = -s$ whenever $\mathrm{ord}(X) \geq -s$ but not $\mathrm{ord}(X) \geq -s+1$.

A differential operator D has **order** greater than or equal to $-s$ if for all α such that $c_\alpha \neq 0$ we have $\sum_{i=1}^n \alpha_i a_i \leq s$. We write $\mathrm{ord}(D) \geq -s$. Moreover, $\mathrm{ord}(D) = -s$ whenever $\mathrm{ord}(D) \geq -s$ but not $\mathrm{ord}(D) \geq -s+1$.

Remark 4.5. A vector field X has weight i if and only if $\mathrm{ord}(X) = -i$, meaning there is a direct connection between weight and order. △

4.2 Privileged coordinates

In this section, we will define privileged coordinates. It is a coordinate system, centred at a point p on the manifold, which 'remembers' the filtered structure. More specifically, it remembers weight. Using the privileged coordinates, we will show how they relate to the framing of the filtered structure, and use them to define a local distance function. This distance function will allow us to measure the distance between the base point p and points near p. If a metric is fixed, this will function as an approximation of the Carnot-Carathéodory metric for filtered structures.

Definition 4.6. Let $x_1, \dots, x_n$ be a system of coordinates centred at p. Then $x_1, \dots, x_n$ are **privileged** if $\mathrm{ord}_p(x_i) = a_i$ and $dx_1, \dots, dx_n$ form a basis of T_p^*M adapted to the filtered structure, meaning that $dx_1, \dots, dx_{n_i}$ form a basis of $\Gamma(\xi_i)(p)$ for all $1 \leq i \leq n$.

Let $x_1, \dots, x_n$ be a system of privileged coordinates. By possibly changing the order of the indices and rescaling, we may assume that $X_i x_i(p) = 1$. Fixing a metric g on the manifold, we can choose X_i to be orthonormal for g. In that case, the x_i are roughly orthogonal.

Proposition 4.7. *We can express X_i in these privileged coordinates as follows:*

$$X_i = \frac{\partial}{\partial x_i} + \sum_{\substack{j \ s.t. \\ a_j \geq a_i}} B_i^j \frac{\partial}{\partial x_j}$$

where B_i^j is a smooth function with $\mathrm{ord}_p\left(B_i^j\right) \geq a_j - a_i$.

Proof. First, let us start with a naive expression, namely

$$X_i = \frac{\partial}{\partial x_i} + \sum_{j \neq i} B_i^j \frac{\partial}{\partial x_j}.$$

We will prove that $B_i^j \equiv 0$ whenever $a_j < a_i$.

Claim 2. The order of B_i^j at p is greater than or equal to $a_j - a_i$.

Proof. By definition, $\mathrm{ord}_p(x_j) = a_j$, and therefore $X_{i_1} \dots X_{i_k} X_i x_j(p) = 0$ whenever $a_{i_1} + \dots + a_{i_k} + a_i \leq a_j - 1$. Therefore,

$$X_{i_1} \dots X_{i_k} B_i^j(p) = \left((X_{i_1} \dots X_{i_k})(X_i x_j)\right)(p) = 0$$

whenever $a_{i_1} + \dots + a_{i_k} \leq a_j - a_i - 1$. Hence, $\mathrm{ord}_p\left(B_i^j\right) \geq a_j - a_i$. □

If $a_j \geq a_i$, we are done.

Now, let $a_j < a_i$. Since B_i^j is smooth, we always have $\mathrm{ord}_p\left(B_i^j\right) \geq 0$. Because $x_1, \ldots, x_n$ are privileged coordinates, we have $\mathrm{ord}\left(\frac{\partial}{\partial x_j}\right) = -a_j$. We have for

$$\begin{aligned}\mathrm{ord}\left(B_i^j \frac{\partial}{\partial x_j}\right) &\geq \mathrm{ord}_p\left(B_i^j\right) + \mathrm{ord}\left(\frac{\partial}{\partial x_j}\right)\\ &\geq 0 - a_j\\ &> -a_i.\end{aligned}$$

If $B_i^j \not\equiv 0$, we have $B_i^j \frac{\partial}{\partial x_j} \in \Gamma(\xi_{a_i+1}) \setminus \Gamma(\xi_{a_i})$. But then $X_i \in \Gamma(\xi_{a_i+1}) \setminus \Gamma(\xi_{a_i})$, so $\mathrm{ord}(X_i) > a_i$ which is a contradiction. Hence, $B_i^j \equiv 0$ whenever $a_j < a_i$. So indeed,

$$X_i = \frac{\partial}{\partial x_i} + \sum_{\substack{j \text{ s.t.} \\ a_j \geq a_i}} B_i^j \frac{\partial}{\partial x_j}.$$

□

Definition 4.8. The **local distance function at p induced by the privileged coordinates** $x_1, \ldots, x_n$ is

$$d_p(p, q) = d(0, (x_1, \ldots, x_n)) = |x_1| + \cdots + |x_i|^{1/a_i} + \cdots + |x_n|^{1/m}$$

where q lies in the domain of the privileged coordinates $x_1, \ldots, x_n$.

Note that the local distance function is not equivalent to the usual distance. It measures the shortest distance with respect to (the weighting induced by) the filtered structure, which means that growth in certain directions is not linear. Moreover, it is truly local. Since it is defined by a system of privileged coordinates, it can only measure the distance from p to a point q in the domain of the system of privileged coordinates. We cannot measure from p to *any* point on the manifold.

The local distance function satisfies the triangle inequality (for norms).

Proposition 4.9. *For any q_1, $q_2 \in M$ we have*

$$d_p(p, q_1 + q_2) \leq d_p(p, q_1) + d_p(p, q_2).$$

Proof. By definition, we have

$$\begin{aligned}d_p(p, q_1 + q_2) &= d(0, (x_1 + y_1, \ldots, x_n + y_n))\\ &= |x_1 + y_1| + \cdots + |x_i + y_i|^{1/a_i} + \cdots + |x_n + y_n|^{1/m}.\end{aligned}$$

Claim 3. For all $1 \leq i \leq n$, we have $|x_i + y_i|^{1/a_i} \leq |x_i|^{1/a_i} + |y_i|^{1/a_i}$.

Proof. We have $|x_i + y_i|^{1/a_i} \leq |x_i|^{1/a_i} + |y_i|^{1/a_i}$ if and only if

$$\begin{aligned}|x_i + y_i| &\leq \left(|x_i|^{1/a_i} + |y_i|^{1/a_i}\right)^{a_i} \\ &= \sum_{j=0}^{a_i} \binom{a_i}{j} |x_i|^{(a_i-j)/a_i} |y_i|^{j/a_i}.\end{aligned}$$

Note that we always have

$$\begin{aligned}|x_i + y_i| &\leq |x_i| + |y_i| \\ &\leq \sum_{j=0}^{a_i} \binom{a_i}{j} |x_i|^{(a_i-j)/a_i} |y_i|^{j/a_i}.\end{aligned}$$

Therefore, we always have $|x_i + y_i|^{1/a_i} \leq |x_i|^{1/a_i} + |y_i|^{1/a_i}$. □

From the claim, it follows that

$$\begin{aligned}d_p(p, q_1 + q_2) &= \sum_{i}^{n} |x_i + y_i|^{1/a_i} \\ &\leq \sum_{i=1}^{n} |x_i|^{1/a_i} + |y_i|^{1/a_i} \\ &= d_p(p, q_1) + d_p(p, q_2).\end{aligned}$$

□

Lastly, we will define a weighted norm, which is similar to the usual notion of a norm, but it is required to 'behave well' with respect to dilations, which are defined as follows for a general graded metric space:

Definition 4.10. Let (V, d) be a graded metric space. Then a **dilation** on V are functions function $\delta : V \to V$ for $r > 0$ such that for every $x, y \in V$ we have $d(\delta_r x, \delta_r y) = rd(x, y)$.

Definition 4.11. A **weighted norm** on $\mathbb{R}^n$ (as a graded vector space) is a continuous function $|| \cdot ||$ such that $||x|| > 0$ for all $x \neq 0$, for all x, y we have $||x + y|| \leq ||x|| + ||y||$ and such that $||\delta_t x|| = t||x||$ for all $t > 0$.

Example 4.12. Two examples of weighted norms are

$$||x|| = \sum_{i=1}^{n} |x_i|^{1/a_i}$$

and

$$||x|| = \sup_i |x_i|^{1/a_i}.$$

△

Remark 4.13. Any two weighted norms are equivalent, since they are a finite distortion of each other. △

4.3 Privileged distances, and the order of functions

We will now take a step back from the local distance function, to consider the privileged distances, of which the local distance function is an example. These are distance functions which behave well with respect to the order. Using privileged distances, we can give another definition of the order of a smooth function. We will show that the local distance function is a privileged distance, providing a new order for smooth functions. We will prove that this is equivalent to our earlier definition of order.

Definition 4.14. A distance function d is **privileged** if for all smooth functions f around p and for all q in the domain of f, we have

$$\mathrm{ord}_p(f) \geq s \Leftrightarrow |f(q)| = O(d(p,q)^s).$$

An alternative way of defining the order of a smooth function is to define it through privileged coordinates.

Definition 4.15. Let f a smooth function defined in a neighbourhood of p. Then f has **order** s with respect to d whenever $|f(q)| = O(d(p,q)^s)$ but $|f(q)| \neq O(d(p,q)^{s+1})$. Here, d is a privileged distance function.

The next example shows that two privileged distance functions do not necessarily have to be equivalent.

Example 4.16. Consider $\mathbb{R}$ as a weighted vector space, of weight 1. Now, consider two sequences $(p_i)_{i\in\mathbb{N}}$ and $(q_i)_{i\in\mathbb{N}}$ with

$$p_i = \frac{1}{2^{2^i}}, \qquad q_i = \frac{1}{2^{2^i} + \frac{1}{2}}.$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that

$$f(p_i) = \frac{1}{2^{2^i}}, \qquad f(q_i) = \frac{1}{2^{2^{i+1}} - \frac{1}{2}}.$$

Especially, we have

$$p_i > q_i > p_{i+1} = p_i^2,$$

and we have

$$p_i = f(p_i) > f(q_i) > f(p_{i+1}) = p_{i+1}.$$

Now, let d be the Euclidean metric, which is a privileged distance function at 0. Then f^*d is also a privileged distrance function, with

$$f^*d(0,p_i) = |p_i|, \qquad f^*d(0,q_i) = \frac{1}{2^{2^{i+1}} - \frac{1}{2}}.$$

To check that f^*d is privileged, it is enough to check that $x = O(f^*d(0,x))$ for all $x \in \mathbb{R}$, since any smooth function can be described by its Taylor series at 0. We have $p_i = O(f^*(0,p_i))$, because $f(p_i) = p_i$. Moreover, we have $q_i = O(f^*d(0,q_i))$ because $f(q_i) < q_i$. Therefore, we have by continuity of f that $x = O(f^*d(0,x))$ for all x close to zero.

Although both d and f^*d are privileged distances, they are not equivalent. We have $\lim_{i\to\infty} f(q_i) - q_i^2 \to 0$, which implies that $f^*d(0,q_i)$ is more or less $\sqrt{q_i}$, which is clearly very different from $|q_i| = d(0,q_i)$. △

We will show that the definition of order with respect to the local distance function is equivalent to our earlier definition of order (which was defined through nonholonomic derivatives). In order to prove this, we will first show that the local distance function, which arises from a system of privileged coordinates, is a privileged distance function.

Proposition 4.17. *The local distance function induced by privileged coordinates is privileged, i.e.* $ord_p(f) \geq s$ *if and only if*

$$|f(x_1, \dots, x_n)| = O\left(\left(|x_1| + \cdots + |x_i|^{1/a_i} + \cdots + |x_n|^{1/m}\right)^s\right).$$

Proof. Without loss of generality, assume that $X_i x_j(p) = \delta_{ij}$. Let f a smooth function defined in a neighbourhood of p. Let $\sum_\alpha c_\alpha x^\alpha$ the Taylor expansion of f in the privileged coordinates, with $\alpha = (\alpha_1, \dots, \alpha_n)$ varying over all multi-indices.

Claim 4. We have

$$X^\alpha x^\beta(p) = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{otherwise.} \end{cases}$$

Especially, we have $\mathrm{ord}_p(x^\alpha) = \sum_{i=1}^n \alpha_i a_i =: w(\alpha)$.

Proof. We can calculate explicitly. We have

$$\begin{aligned} X^\alpha x^\beta(p) &= \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} x_1^{\beta_1} \dots x_n^{\beta_n}(p) \\ &= \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

From this, it follows immediately that $\mathrm{ord}_p(x^\alpha) = w(\alpha)$. □

It follows from the claim and the Taylor expansion of f that $X^\alpha f(p) = c_\alpha$. Therefore, $\mathrm{ord}_p(f) \geq s$ if and only if $c_\alpha = 0$ whenever $w(\alpha) \leq s-1$. Especially, $\mathrm{ord}_p(f) \geq s$ if and only if $c_\alpha = 0$ whenever $\alpha = (\beta_1, \dots, \beta_i/a_i, \dots, \beta_n/m)$ with β another multi-index and $\sum_{i=1}^n \beta_i = w(\alpha) \leq s-1$.

Claim 5. We have $\mathrm{ord}_p\left(\left(x_1 + \cdots + x_i^{1/a_i} + \cdots + x_n^{1/m}\right)^s\right) = s$.

Proof. We have

$$\left(x_1 + \cdots + x_i^{1/a_i} + \cdots + x_n^{1/m}\right)^s = \sum_{\substack{\alpha \text{ s.t.} \\ \alpha_1 + \cdots + \alpha_n = s}} x_1^{\alpha_1} \dots x_i^{\alpha_i/a_i} \dots x_n^{\alpha_n/m}.$$

Moreover, we have

$$\mathrm{ord}_p\left(x_1^{\alpha_1} \dots x_i^{\alpha_i/a_i} \dots x_n^{\alpha_n/m}\right) = \sum_{i=1}^n (\alpha_i/a_i) \cdot a_i = \sum_{i=1}^n \alpha_i.$$

Hence, we have

$$\mathrm{ord}_p\left(\sum_{\substack{\alpha \text{ s.t.} \\ \alpha_1 + \cdots + \alpha_n = s}} x_1^{\alpha_1} \dots x_i^{\alpha_i/a_i} \dots x_n^{\alpha_n/m}\right) = s.$$

□

From the claim, it follows that $\mathrm{ord}_p(f) \geq s$ if and only if

$$|f(x_1, \ldots, x_n)| = O\left(\left(|x_1| + \cdots + |x_i|^{1/a_i} + \cdots + |x_n|^{1/m}\right)^s\right).$$

□

Now, we can prove that the order of a smooth function at a point is the same under the old order of smooth functions, and with respect to the local distance function.

Lemma 4.18. *The two definitions of order are equivalent.*

Proof. Let f a smooth function defined on a neighbourhood of p. By Proposition 4.17 we have that all nonholonomic derivatives of of f of order $\leq s-1$ vanish at p if and only if

$$|f(x_1, \ldots, x_n)| = O\left(\left(|x_1| + \cdots + |x_i|^{1/a_i} + \cdots + |x_n|^{1/m}\right)^s\right).$$

Moreover, there is a nonholonomic derivative of f of order s which does not vanish at p if and only if

$$|f(x_1, \ldots, x_n)| \neq O\left(\left(|x_1| + \cdots + |x_i|^{1/a_i} + \cdots + |x_n|^{1/m}\right)^{s+1}\right).$$

Hence, the two definitions of order are equivalent. □

4.4 Dilations and homogeneity

In this section, we will use privileged coordinates to define dilation at a point. Dilations are somewhat similar to scalar multiplications, but they remember the grading. Using dilations, we will define homogeneity of vector fields and smooth functions, and will construct a homogeneous approximation for vector fields. This will ultimately allow us to use the framing of a filtered structure to create a basis for the nilpotentisation.

Once again, let M be a manifold endowed with a filtered structure $\xi_1 \subset \cdots \subset \xi_m = TM$.

Definition 4.19. A **dilation induced by a system of privileged coordinates** $x_1, \ldots, x_n$ at the point p is a function δ_r^p with $r > 0$, which satisfies

$$(\delta_r^p)^* (x_1, \ldots, \xi_i, \ldots, x_n) = (rx_1, \ldots, r^{a_i} x_i, \ldots, r^m x_n).$$

For a smooth function f, the dilation δ_r^p satisfies

$$(\delta_r^p)^* f(x_1, \ldots, x_i, \ldots, x_n) = f(rx_1, \ldots, r^{a_i} x_i, \ldots, r^m x_n).$$

Definition 4.20. Let $X \in \mathfrak{X}(M)$ be a vector field on M of order $-s$. Then X is **homogeneous** at p if $(\delta_r^p)^* X = r^{-s} X$. Similarly, a smooth function f on M of order a is **homogeneous** at p if $(\delta_r^p)^* f = r^a f$.

Note that the privileged coordinates are homogeneous, in particular.

Proposition 4.21. *Let X, $Y \in \mathfrak{X}(M)$ homogeneous of order $-s_1$, $-s_2$ respectively. Then $[X, Y]$ is homogeneous as well.*

Proof. We calculate:

$$\begin{aligned}(\delta_r^p)^*[X,Y] &= [(\delta_r^p)^*X, (\delta_r^p)^*Y] \\ &= [r^{-s_1}X, r^{-s_2}Y] \\ &= r^{-(s_1+s_2)}[X,Y].\end{aligned}$$

□

We can decompose a vector field into homogeneous parts.

Proposition 4.22. *Let X a vector field of order $-s$ at p. Then we can write*

$$X = X^{(-s)} + X^{(-s+1)} + \cdots + X^{(-1)} + X^{(0)} + X^{(1)} + \dots$$

where $X^{(i)}$ is a homogeneous vector field with $\mathrm{ord}_p\left(X^{(i)}\right) = i$.

Proof. We calculate the decomposition inductively. First, we define

$$X^{(-s)} = \lim_{r\to 0} r^s \delta_r^* X.$$

Because $\mathrm{ord}_p(X) = -s$, we have $X^{(-s)} \neq 0$. Moreover, $\mathrm{ord}_p\left(X - X^{(-s)}\right) \geq -s+1$ whenever $X \neq X^{(-s)}$. So, we can calculate

$$X^{(-s+1)} = \lim_{r\to 0} r^{s-1}\delta_r^*\left(X - X^{(-s)}\right).$$

We have $X^{(-s+1)}$ is either homogeneous at p of order $-s+1$ or zero. Continuing in this manner, we define for general $i \geq -s$ that

$$X^{(i)} = \lim_{r\to 0} r^{-i}\delta_r^*\left(X - X^{(-s)} - X^{(-s+1)} - \cdots - X^{(i-1)}\right).$$

By induction, we have $X^{(i)}$ either zero or homogeneous at p of order i. Hence, we have

$$X = X^{(-s)} + X^{(-s+1)} + \cdots + X^{(-1)} + X^{(0)} + X^{(1)} + \dots$$

□

Remark 4.23. In the case of the trivial grading, this is just the usual Taylor expansion. △

We can define $\hat{X} = X^{(-s)}$, with $\mathrm{ord}_p(X) = s$. By the following lemma, we can equivalently define $\hat{X} = \lim_{r\to 0} r^s(\delta_r^p)^*X$.

Lemma 4.24. *We have for all $i \geq -s$ that*

$$\lim_{r\to 0} r^s(\delta_r^p)^*X^{(i)} = \begin{cases} X^{(-s)} & \text{if } i = -s \\ 0 & \text{otherwise.}\end{cases}$$

Proof. Because $X^{(i)}$ is homogeneous of degree i, we can calculate explicitly

$$\lim_{r\to 0} r^s(\delta_r^p)^* X^{(i)} = r^{s+i}X^{(i)} = \begin{cases} X^{(-s)} & \text{if } i = -s \\ 0 & \text{otherwise.} \end{cases}$$

□

The following proposition shows that the map $X \mapsto \hat{X}$ is well-behaved with respect to the Lie bracket. When we will later prove that $\hat{X}_1, \ldots, \hat{X}_n$ generates a finite-dimensional nilpotent Lie algebra (which is isomorphic to the nilpotentisation), this ensures that bracket expressions still respect the grading (which is especially important in the case of a filtered structure arising from a bracket-generating distribution). It is remarkable that these homogeneous vector fields produce such a Lie algebra, since most pairs of vector fields generate infinite-dimensional Lie algebras.

Proposition 4.25. *Let $X, Y \in \mathfrak{X}(M)$ two vector fields, with order $-s_1$, $-s_2$ respectively. If $[\hat{X}, \hat{Y}] \neq 0$, then*

$$\widehat{[X,Y]} = [\hat{X}, \hat{Y}]$$

and $\widehat{[X,Y]}$ is homogeneous of order $-(s_1+s_2)$.

Proof. We can expand the vector fields in their homogeneous parts, i.e.

$$X = X^{(-s_1)} + X^{(-s_1+1)} + \ldots$$

and

$$Y = Y^{(-s_2)} + Y^{(-s_2+1)} + \ldots.$$

By definition, $\hat{X} = X^{(-s_1)}$ and $\hat{Y} = Y^{(-s_2)}$. Therefore,

$$[\hat{X}, \hat{Y}] = \left[X^{(-s_1)}, Y^{(-s_2)}\right].$$

Moreover, we have

$$\begin{aligned} [X,Y] &= \left[\sum_{i=-s_1}^{\infty} X^{(i)}, \sum_{j=-s_2}^{\infty} Y^{(j)}\right] \\ &= \sum_{i=-s_1}^{\infty}\sum_{j=-s_2}^{\infty}\left[X^{(i)}, Y^{(j)}\right]. \end{aligned}$$

Particularly, we have

$$[X,Y]^{(l)} = \sum_{i+j=l}\left[X^{(i)}, Y^{(j)}\right].$$

Since $[\hat{X}, \hat{Y}] \neq 0$, we have

$$[X,Y] = [X,Y]^{(-(s_1+s_2))} + [X,Y]^{(-(s_1+s_2)+1)} + \ldots$$

and $[X,Y]^{(-(s_1+s_2))} = \left[X^{(-s_1)}, Y^{(-s_2)}\right]$. Especially, we have

$$\widehat{[X,Y]} = \left[X^{(-s_1)}, Y^{(-s_2)}\right] = \left[\hat{X}, \hat{Y}\right].$$

□

Remark 4.26. Let us consider (M, ξ) a manifold endowed with a bracket-generating distribution. Let $X_1, \dots, X_k$ a framing of ξ. By definition, there are $X_{k+1}, \dots, X_n$ with $X_i = A_i(X_1, \dots, X_k)$ such that $X_1, \dots, X_n$ form a framing of TM. By the previous proposition, we have that $\hat{X}_i = A_i\left(\hat{X}_1, \dots, \hat{X}_k\right)$. △

The next lemma shows that applying the map $X \mapsto \hat{X}$ to the framing of the filtered structure, gives a basis for the nilpotentisation.

Proposition 4.27. *The Lie algebra spanned by $\hat{X}_1, \dots, \hat{X}_n$ is the nilpotentisation of $\xi_1 \subset \dots \subset \xi_m$.*

Proof. For every i, we have that $X_1, \dots, X_{n_i}$ is a framing of ξ_i. Therefore,

$$X_1 \mod (\Gamma(\xi_{i-1})), \dots, X_{n_i} \mod (\Gamma(\xi_{i-1}))$$

spans $\Gamma(\xi_i)/\Gamma(\xi_{i-1})$. Since $X_1, \dots, X_{n_{i-1}} \in \Gamma(\xi_{i-1})$, it follows that for all $1 \leq j \leq n_{i-1}$

$$X_j \equiv 0 \mod (\Gamma(\xi_{i-1}))$$

and therefore

$$X_{n_{i-1}+1} \mod (\Gamma(\xi_{i-1})), \dots, X_{n_i} \mod (\Gamma(\xi_{i-1}))$$

span $\Gamma(\xi_i)/\Gamma(\xi_{i-1})$. We have $\dim(\Gamma(\xi_i)/\Gamma(\xi_{i-1})) = n_i - n_{i-1}$, so the above is indeed a basis for $\Gamma(\xi_i)/\Gamma(\xi_{i-1})$.
Similarly, $\hat{X}_{n_{i-1}+1}, \dots, \hat{X}_{n_i}$ are linearly independent, and are homogeneous of weight i. Let $\hat{\Gamma}_i$ be spanned by $\hat{X}_{n_{i-1}+1}, \dots, \hat{X}_{n_i}$ as a vector space over $\mathbb{R}$. If $[\hat{X}_i, \hat{X}_j] \neq 0$, then $[\hat{X}_i, \hat{X}_j]$ is homogeneous of order $-(a_i + a_j)$. Therefore, we have $\left[\hat{\Gamma}_i, \hat{\Gamma}_j\right] \subset \hat{\Gamma}_{i+j}$.
As a vector space, $\Gamma(\xi_i)/\Gamma(\xi_{i-1})$ is isomorphic to $\hat{\Gamma}_i$, by

$$X_j \mod (\Gamma_{i-1}) \mapsto \hat{X}_j.$$

Moreover, $\hat{\Gamma}_1 \oplus \dots \oplus \hat{\Gamma}_m$ is isomorphic as a graded Lie algebra to the nilpotentisation of $\xi_1 \subset \dots \subset \xi_m$. □

Therefore, this construction of the Lie algebra of homogeneous vector fields does not depend on the choice of a framing of the filtered structure up to isomorphism.

4.5 Global distance

We have seen before how we can construct a local distance function on a manifold endowed with a filtered structure. We can also construct a global distance function. We do this by taking the infimums over the lengths of concatenations of the integral curves of frames, where the distance on each integral curve is given by the degree of a vector field. This gives a Manhattan-style distance, since we are only allowed to move along curves tangent to the framing.

More specifically, let $(M, \xi_1 \subset \xi_m)$ be a manifold endowed with a filtered structure. We can cover M by open balls U_i with $i \in I$ for some indexing set I, i.e. $M \subset \bigcup_{i \in I} U_i$. Choose the balls small enough so that for every $i \in I$ there is a framing $X_1^i, \dots, X_n^i$ of the filtered structure on U_i, with $X_1^i, \dots, X_{n_j}^i$ a framing of ξ_j for $j = 1, \dots, m$. Similar to Section 3.2, we want to reach any point on the manifold through concatenations of integral curves of the framing.

Let us first define the length of a flow curve.

Definition 4.28. Let $X \in \mathfrak{X}(M)$ of order $-s$, and let ϕ_X^t be the time t flow of X. Then the **length** of ϕ_X^t is given by

$$\text{length}\left(\phi_X^t\right) = |t|^{1/s}.$$

We can extend this definition to concatenations of vector fields.

Definition 4.29. Let $X_1, \dots, X_k \in \mathfrak{X}(M)$ of orders $-s_1, \dots, -s_k$ respectively. Then the **length** of $\phi_{X_1}^{t_1} \circ \dots \circ \phi_{X_k}^{t_k}$ is given by

$$\begin{aligned}\text{length}\left(\phi_{X_1}^{t_1} \circ \dots \circ \phi_{X_k}^{t_k}\right) &= \text{length}\left(\phi_{X_1}^{t_1}\right) + \dots + \text{length}\left(\phi_{X_k}^{t_k}\right)\\ &= |t_1|^{1/s_1} + \dots + |t_k|^{1/s_k}.\end{aligned}$$

Remark 4.30. For the X_j^i we have in particular that length $\left(\phi_{X_j^i}^t\right) = t^{1/a_j}$. △

Using this, we can define a global distance as the infimum over all paths which are concatenations of the flows of the vector fields in the framing over U_i.

Definition 4.31. The **global distance** d associated to the framing $X_1, \dots, X_n$ on M is defined as

$$d(p,q) = \inf_{\phi \in C_{p,q}} \{\text{length}(\phi)\}$$

where $C_{p,q}$ is the set of paths which are concatenations of $\phi_{X_j^i}^t$ for $i \in I$ and $1 \leq j \leq n_i$.

The global distance depends on the choice of framing, since that is what dictates which directions we are allowed to take.

Lemma 4.32. *The global distance is a metric, i.e. for any $p, q, r \in M$ the global metric satisfies*

1. $d(p,q) = 0 \iff p = q$
2. $d(p,q) = d(q,p)$
3. $d(p,q) \leq d(p,r) + d(r,q)$.

Proof. If $p = q$ then it is trivially true that $d(p,q) = 0$. Now, suppose that $d(p,q) = 0$. Let $C'_{p,q}$ be the set of all paths connecting p and q. Then we have

$$0 = d(p,q) = \inf_{\phi \in C_{p,q}} \{\text{length}(\phi)\} \geq \inf_{\phi \in C'_{p,q}} \{\text{length}(\phi)\} \geq 0.$$

Therefore, $p = q$. So, $d(p,q) = 0 \iff p = q$.

Moreover, we have

$$\begin{aligned}d(p,q) &= \inf_{\phi \in C_{p,q}} \{\text{length}(\phi)\}\\ &= \inf_{\phi \in C_{q,p}} \{\text{length}(-\phi)\}\\ &= \inf_{\phi \in C_{q,p}} \{\text{length}(\phi)\}\\ &= d(q,p).\end{aligned}$$

Lastly, let $C_{p,r,q} \subset C_{p,q}$ be the subset of paths which flow through r. We then have

$$\begin{aligned} d(p,q) &= \inf_{\phi \in C_{p,q}} \{\text{length}(\phi)\} \\ &\leq \inf_{\phi \in C_{p,r,q}} \{\text{length}(\phi)\} \\ &\leq \inf_{\phi \in C_{p,r},\, \psi \in C_{r,q}} \{\text{length}(\phi) + \text{length}(\psi)\} \\ &= \inf_{\phi \in C_{p,r}} \{\text{length}(\phi)\} + \inf_{\psi \in C_{q,r}} \{\text{length}(\psi)\} \\ &= d(p,r) + d(r,q). \end{aligned}$$

□

We will leave it as an open question to further define the properties of this global distance. For example, if a filtered structure arises from a bracket-generating distribution, it is unclear how it relates to the Carnot-Carathéodory distance.

5 The tangent cone

In this section, we will define the tangent cone to a metric space. For a metric space (X, d) and a point $x_0 \in X$, its tangent cone at x_0 appears when we stretch the distances to infinity, while the position of the point x_0 remains the same.

In order to properly define the tangent cone, we first need to define the Gromov-Hausdorff distance, which measures the distance between two metric spaces. This in turn allows us to define the limit of a sequence of metric spaces, which is what the tangent cone is, if it exists.

Although one would expect that the tangent cone forgets all structure, Mitchell's theorem will show us that this is not the case. Indeed, for a manifold endowed with a filtered structure, the tangent cone exists, and is equal to the Lie group whose Lie algebra is the nilpotentisation of the filtered structure at a point.

Lastly, we will construct an explicit group product on the tangent cone, using dilations and privileged coordinates. Being a Lie group, the tangent cone has a group structure. It is not clear at the start, however, how the group structure relates to the filtered structure.

This chapter is based on [1], [7] and chapter 8 of [8].

5.1 Dilations

Before we do anything else, we will consider the following two definitions. These notions will be used throughout this chapter, and it is useful to know what exactly these mean.

Definition 5.1. A **pointed metric space** is a pair (X, x_0), where X is a metric space and $x_0 \in X$. The point x_0 is called the **base point**.

The next definition is a restatement, and expansion, of Definition 4.10.

Definition 5.2. Let (X, d) a metric space, and $r > 0$. An **r-dilation** is a map $\delta_r : X \to X$ such that $d(\delta_r(x), \delta_r(y)) = rd(x, y)$ for all x, $y \in X$.
If (X, x_0) is a pointed metric space, then we require that the r-dilation fixes the base point, i.e. $\delta_r(x_0) = x_0$.

5.2 Convergence of metric spaces

In this section, we will define the Gromov-Hausdorff distance, which measures the distance between two metric spaces. First we define the Hausdorff distance, which measures the distance between two subsets of a metric space. This allows us to define the Gromov-Hausdorff distance by measuring the distances between embeddings of two metric spaces into another (larger) metric space. Moreover, we will define what it means for two metric spaces to converge. We will use this to check whether the tangent cone exists.

In order to define the Hausdorff distance, we will need the following definition.

Definition 5.3. Let (X, d) a metric space, and $A \subset X$. Then the **δ-neighbourhood** of A is the set

$$N_\delta(A) = \{x \in X : \exists a \in A \text{ such that } d(x, a) < \delta\}.$$

Definition 5.4. Let (X, d) a metric space, A, $B \subset X$. Then the **Hausdorff distance** between A and B is

$$d_H(A, B) = \inf\{\delta : A \subset N_\delta(B),\, B \subset N_\delta(A)\}.$$

We can generalise the concept of Hausdorff distance to measure the distance between two 'abstract' metric spaces, by embedding them into a larger metric space while remembering their sizes through preserving the distances on the images.

Definition 5.5. Let (X, d_X), (Y, d_Y) be two metric spaces. An **isometric embedding** is an embedding $f : X \to Y$ such that $f^* d_Y = d_X$, i.e. f preserves distances.

Definition 5.6. Let (X, d_X), (Y, d_Y) two metric spaces. Then the **Gromov-Hausdorff distance** between X and Y is defined as

$$d_{GH}(X, Y) = \inf d_H(i(X), j(Y)),$$

where the infimum is taken over all metric spaces (Z, d_Z) and isometric embeddings $i : X \to Z$ and $j : Y \to Z$ (i.e. all spaces Z for which such isometric embeddings exist).

The above definition of a Gromov-Hausdorff space is rather impractical when doing computations. Lemma 5.8 will provide an easier alternative through the notion of compatible metrics.

Definition 5.7. Let (X, d_X), (Y, d_Y) be metric spaces. Then metric d on the disjoint union $X \coprod Y$ is **compatible** if $d|_X = d_X$ and $d|_Y = d_Y$.

Lemma 5.8. *Let X, Y be metric spaces. We have*

$$d_{GH}(X, Y) = \inf\{d(X, Y) : d \textit{ compatible on } X \coprod Y\}.$$

Proof. Define $\delta = \inf\{d(X, Y) : d \text{ compatible on } X \coprod Y\}$. Let d be a compatible metric, and let $i : X \to X \coprod Y$ and $j : Y \to X \coprod Y$ be isometric embeddings. By definition of the Gromov-Hausdorff metric, we have $d_{GH}(X, Y) \leq d(i(X), j(Y))$. Hence, $d_{GH}(X, Y) \leq \delta$.

On the other hand, let $i : X \to Z$ and $j : Y \to Z$ be isometric embeddings into a metric space (Z, d_Z). If $i(X) \cap j(Y) = \emptyset$, then $X \coprod Y \cong i(X) \cup j(Y)$, so the restriction of d_Z to $i(X) \cup j(Y)$ gives a metric d on $X \coprod Y$ such that $d(X, Y) = d_Z(i(X), j(Y)) \geq \delta$.

Now consider the case that $i(X) \cap j(Y) \neq \emptyset$. For $\epsilon > 0$, define $I_\epsilon = [0, \epsilon]$. On $Z \times I_\epsilon$, we define the metric

$$d((x, t), (y, s)) = \sqrt{d_Z}(x, y)^2 + |t - s|^2.$$

Embed X into $Z \times I_\epsilon$ by $i_0(x) = (i(x), 0)$, and embed Y into $Z \times I_\epsilon$ by $j_\epsilon(y) = (j(y), \epsilon)$. Then $i_0(X) \cap j_\epsilon(Y) = \emptyset$, and

$$d(i_0(X), j_\epsilon(Y)) = \sqrt{d_Z(i(X), j(Y))^2 + \epsilon^2} \geq \delta.$$

Letting $\epsilon \to 0$, we have $d_Z(i(X), j(Y)) \geq \delta$.

Thus, $d_{GH}(X, Y) \geq \delta$. Hence, $d_{GH}(X, Y) = \delta$. □

When two compact metric spaces are isometric, we already know the Gromov-Hausdorff distance.

Proposition 5.9. *Let X, Y be compact metric spaces. If X and Y are isometric, then $d_{GH}(X, Y) = 0$.*

Proof. If X and Y are isometric, there is a bijective isometry $f : X \to Y$. So, $f(X) = Y$. By definition, for every $A \subset Y$, we have $N_0(A) = A$. Especially, we have

$$d_{GH}(X, Y) \leq d_H(f(X), Y) = 0.$$

□

Now that we can measure the distance between two metric spaces, we can define convergence for metric spaces.

Definition 5.10. Let $\{X_i\}$ a sequence of metric spaces. They **converge** to a metric space (Y, d_Y) if $d_{GH}(X_i, Y) \to 0$ as $i \to \infty$.

Although the definition above is natural, it is not as useful when a sequence of metric spaces converges to an unbounded metric space. Intuitively, the sequence of intervals $\{(-i,i)\}_{i\in\mathbb{N}}$ should converge to $\mathbb{R}$. However, since $\mathbb{R}$ is unbounded, we have $\lim_{i\to\infty} d_{GH}((-i,i),\mathbb{R}) = \infty$. In the case of pointed metric spaces, it therefore is useful to introduce a new notion of convergence.

Definition 5.11. Let (X_i, x_i) a sequence of pointed metric spaces. It **converges** to a pointed metric space (Y, y) if for every $r > 0$, the r-balls around x_i in X_i converge to the r-balls around y in Y.

The notion of an approximate isometry will be useful when proving that a sequence of metric spaces converges. An approximate isometry will allow us to find an upper bound for the Gromov-Hausdorff distance in some cases.

Definition 5.12. A map $f : A \to Y$ satisfying the assumptions of Lemma 5.13 is called an **approximate isometry**, or a $(\delta, \epsilon_X, \epsilon_Y)$**-isometry**, between X and Y.

Lemma 5.13 (Approximate isometry criterion)**.** *Let X, Y be metric spaces, $A \subset X$, and $f : A \to Y$ not necessarily continuous, $\delta > 0$ such that for all a_1, $a_2 \in A$*

$$|d_Y(f(a_1), f(a_2)) - d_X(a_1, a_2)| \leq \delta.$$

Suppose that there are ϵ_X, $\epsilon_Y > 0$ such that for the metric d_X we have $x \in N_{\epsilon_X}(A)$ for all $x \in X$, and for the metric d_Y we have $y \in N_{\epsilon_Y}(f(A))$ for all $y \in Y$. Then

$$d_{GH}(X, Y) \leq \max\{\epsilon_X, \epsilon_Y\} + \delta/2.$$

Proof. We will define a compatible metric d on $X \coprod Y$. For $x \in X$, $y \in Y$ define

$$d(x, y) = \inf_{a\in A}\{d_X(x, a) + d_Y(f(a), y)\} + \delta/2.$$

We first check that the triangle inequality holds, so that d is indeed a metric. Let $x_1, x_2 \in X$, $y \in Y$. We check that $d(x_1, y) + d(y, x_2) \geq d(x_1, x_2)$. We have

$$\begin{aligned} d(x_1, y) + d(y, x_2) = \inf_{a_1\in A}&\{d_X(x_1, a) + d_Y(f(a_1), y)\} \\ &+ \inf_{a_2\in A}\{d_X(x_2, a_2) + d_Y(f(a_2), y)\} + \delta. \end{aligned}$$

On the other hand, we have

$$d(x_1, x_2) = d_X(x_1, x_2) \leq d_X(x_1, a_1) + d_X(a_1, a_2) + d_X(a_2, x_2).$$

By assumption, we have

$$\begin{aligned} d_X(a_1, a_2) &\leq d_Y(f(a_1), f(a_2)) + \delta \\ &\leq d_Y(f(a_1), y) + d_Y(y, f(a_2)) + \delta. \end{aligned}$$

Hence,
$$d(x_1,x_2) \leq d_X(x_1,a_1) + d_Y(f(a_1),y) + d_Y(y,f(a_2)) + d_X(a_2,x_2) + \delta.$$
Taking the infimum over all a_1, a_2, we get that $d(x_1,x_2) \leq d(x_1,y) + d(y,x_2)$.
Now, let $x \in X$, $y_1,y_2 \in Y$. We check that $d(y_1,x) + d(x,y_2) \geq d(y_1,y_2)$. We have
$$d(y_1,x) + d(x,y_2) = \inf_{a_1 \in A}\{d_X(x,a_1) + d_Y(f(a_1),y_1)\} + \inf_{a_2 \in A}\{d_X(x,a_2) + d_Y(f(a_2),y_2)\} + \delta.$$
On the other hand, we have
$$\begin{aligned} d(y_1,y_2) &\leq d_Y(y_1,y_2) \\ &\leq d_Y(y_1,f(a_1)) + d_Y(f(a_1),f(a_2)) + d_Y(f(a_2),y_2). \end{aligned}$$
By assumption, we have
$$\begin{aligned} d_Y(f(a_1),f(a_2)) &\leq d_X(a_1,a_2) + \delta \\ &\leq d_X(a_1,x) + d_X(x,a_2) + \delta. \end{aligned}$$
Hence,
$$d(y_1,y_2) \leq d_Y(y_1,f(a_1)) + d_X(a_1,x) + d_X(x,a_2) + d_Y(f(a_2),y_2) + \delta.$$
Taking the infimum over all a_1, a_2, we get that $d(y_1,y_2) \leq d(y_1,x) + d(x,y_2)$.
So, d is indeed a metric.
By assumption, for all $x \in X$, there is $a \in A$ such that $d_X(a,x) < \epsilon_X$. We have
$$d(x,a) \leq d_X(a,x) + \delta/2 < \epsilon_X + \delta/2.$$
So $X \subset N_{\delta_1}(Y)$ for the metric d, where $\delta_1 = \epsilon_X + \delta/2$. Similarly, for every $y \in Y$, there is $a \in A$ such that $d_Y(f(a),y) < \epsilon_Y$. We have
$$d(a,y) \leq d_Y(f(a),y) + \delta/2 < \epsilon_Y + \delta/2.$$
So $Y \subset N_{\delta_2}(X)$ for d, where $\delta_2 = \epsilon_Y + \delta/2$. So by Lemma 5.8, we have
$$d_{GH}(X,Y) \leq \max\{\delta_1,\delta_2\} = \max\{\epsilon_X,\epsilon_Y\} + \delta/2.$$
□

5.3 The tangent cone and Mitchell's theorem

Using the tools from the previous section, we can now properly define the tangent cone. Moreover, we will formulate Mitchell's Theorem, which shows how the tangent cone of a manifold endowed with a filtered structure and the nilpotentisation are related.

Definition 5.14. Let (X,x_0) a pointed metric space, with metric d. Then its **tangent cone** is given by $\lim_{\lambda\to\infty}(X,x_0,\lambda d)$, if the limit exists.

Our goal for the next few sections is to prove the following theorem:

Theorem 5.15 (Mitchell's theorem for filtered structures)**.** *Let $(M,\xi_1 \subset \cdots \subset \xi_m)$ a manifold endowed with a filtered structure, and $p \in M$. Let $X_1,\ldots,X_n$ be a framing of TM such that $X_1,\ldots,X_{n_i}$ is a framing of ξ_i for $i=1,\ldots,m$. Let d be the global metric associated to $X_1,\ldots,X_n$. Then the tangent cone of $(M,\xi_1 \subset \cdots \subset \xi_m)$ at p exists for d, and is equal tot he simply-connected Lie group with Lie algebra $L(\xi_1 \subset \cdots \subset \xi_m)(p)$, the nilpotentisation at p.*

5.4 Continuous expansion property

To prove Mitchell's theorem, we will need the notion of the continuous expansion property. If a metric space has this property, it means that the δ-neighbourhood of a ball of radius ϵ looks like we would expect, namely a ball of radius $\delta + \epsilon$. In Lemma 5.17 we will show how we can use this property to prove that the tangent cone exists.

Definition 5.16. A metric space X has the **continuous expansion property** at a point $x \in X$ if $N_\delta(B(x,\epsilon)) = B(x, \delta+\epsilon)$ for all δ, $\epsilon > 0$.

The continuous expansion property holds for Riemannian metrics, Carnot-Carathéodory metrics and global metrics for a manifold with a filtered structure (because they are all defined by lengths of curves). It may fail, however, for discrete metric spaces.

Lemma 5.17. *Let (X, x_0) a pointed metric space with metric d_0 that admits dilations δ_r. Suppose that d is a metric defined in a d_0-neighbourhood U of x_0, and $|\delta_r^*(d-d_0)| = o(r)$ uniformly on U as $r \to 0$. If both d and d_0 have the continuous expansion property at x_0, then the tangent cone of (X, x_0) exists and is equal to (X, d_0, x_0).*

Proof. Let $B_r(\epsilon)$, $B_0(\epsilon)$ be the ball of radius ϵ around x_0 with respect to the metric $(1/r)d$, d_0 respectively. By assumption, for all $x_1, x_2 \in U$ we have

$$|d(\delta_r x_1, \delta_r x_2) - d_0(\delta_r x_1, \delta_r x_2)| \leq f(r)$$

for some function f such that $f(r)/r \to 0$ as $r \to 0$. Since the δ_r are d_0-dilations, we can rewrite the estimate above as

$$|(1/r)d(\delta_r x_1, \delta_r x_2) - d_0(x_1, x_2)| \leq f(r)/r.$$

Let $h(r) = f(r)/r$. By the estimate above, δ_r is an approximate isometry between $B_0(\epsilon)$ and $B_r(\epsilon)$. Setting $x_1 = x_0$ gives

$$|(1/r)d(x_0, \delta_r x_2) - d_0(x_0, x_2)| \leq h(r).$$

So, $\delta_r(B_0(\epsilon - h(r)) \subset B_r(\epsilon)$ and $B_r(\epsilon - h(r)) \subset \delta_r(B_0(\epsilon))$. Take $A = \delta_r^{-1}(B_r(\epsilon)) \cap B_0(\epsilon)$ as the domain of δ_r. We then have that $B_0(\epsilon - h(r)) \subset A$ and $B_r(\epsilon - h(r)) \subset \delta_r(A)$. From the continuous expansion property, it follows that δ_r is a $(h(r), h(r), h(r))$-approximate isometry, and therefore $d_{GH}(B_r(\epsilon), B_0(\epsilon)) \leq 3h(r)/2$. Since $h(r) \to 0$ as $r \to 0$, the result follows. □

5.5 Metrics on the Lie group with Lie algebra the nilpotentisation

First, consider (M, ξ) a manifold with a weakly regular, bracket-generating distribution, and let $p \in M$. Then $L(\xi)(p)$ is a Lie algebra associated to a Carnot group G. We can apply left-multiplication by elements of the Carnot group on $\xi^{(1)}(p) \subset L(\xi)(p)$ to extend it to all of G. By taking the Riemannian metric g on $\xi^{(1)}(p)$, we can extend it by left-invariance to all of G, which gives us a Carnot-Carathéodory metric.

Now, let $(M, \xi_1 \subset \cdots \subset \xi_m)$ a manifold with a filtered structure, and $p \in M$. Then $L(\xi_1 \subset \cdots \subset \xi_m)(p)$ is a Lie algebra for a simply-connected Lie group G. Applying left-multiplication by elements of the Lie group on $L(\xi_1 \subset \cdots \subset \xi_m)(p)$ lets us extend it to all of G, providing a left-invariant filtered structure $\hat{\xi}_1 \subset \cdots \subset \hat{\xi}_m$ on G. With this, we can create a global metric on G by taking a framing of the filtered structure $\hat{\xi}_1 \subset \cdots \subset \hat{\xi}_m$.

5.6 Mitchell: the proof

Finally, we can prove Mitchell's theorem. First, we will state and prove the regular version of Mitchell's Theorem, which is the original version. After that, we will prove Mitchell's Theorem for filtered structures, which is an original contribution of this thesis.

Theorem 5.18 (Mitchell, regular version)**.** *Let (M, ξ) a manifold endowed with a bracket-generating, weakly regular distribution, and $p \in M$. The tangent cone at p exists for the Carnot-Carathéodory metric, and is equal to the Carnot group with a left-invariant Carnot-Carathéodory metric and with graded Lie algebra $L(\xi)(p)$, the nilpotentisation of ξ at p.*

Proof. Let $\{X_1, \dots, X_k\}$ be a framing of ξ around p. Because ξ is bracket-generating, there exist $X_{k+1}, \dots, X_n$ given by $X_j = A_j(X_{i_1^j}, \dots, X_{i_{a_j}^j})$, such that $\{X_1, \dots, X_n\}$ span $\mathfrak{X}(M)$. Here $A_j(-)$ is a bracket expression of length a_j. Let m the step for which the Lie flag stabilises. By possibly rearranging the order, assume that $a_{k+1} \leq a_{k+2} \leq \dots \leq a_n = m$. For $j \leq k$, let $a_j = 1$.
For $j \leq k$, let $\phi_j^t = \phi_{X_j}^t$, the flow of X_j. For $j > k$, let $\phi_j^t = A_j\left(\phi_{X_{i_1^j}}^{t/a_j}, \dots, \phi_{X_{i_{a_j}^j}}^{t/a_j}\right)$. As in Chow's Theorem, we define the endpoint map ψ based at p:

$$\begin{aligned}\psi : \mathbb{R}^n &\to M\\ (t_1, \dots, t_n) &\mapsto \phi_n^{t_n} \circ \dots \circ \phi_2^{t_2} \circ \phi_1^{t_1}(p).\end{aligned}$$

The t_i are privileged coordinates. For $r > 0$ we define the dilation

$$\delta_r(t_1, \dots, t_j, \dots, t_n) = (rt_1, \dots, r^{a_j}t_j, \dots, r^m t_n).$$

Let $u \in L^2([0,1], \mathbb{R}^k)$. Let $x(t)$ be a solution to the control problem

$$\dot{x}(t) = \sum_{j=1}^k u_j(t) X_j(x(t)), \quad x(0) = 0,$$

and let $\hat{x}(t)$ be a solution to

$$\dot{\hat{x}}(t) = \sum_{j=1}^k u_j(t) \hat{X}_j(\hat{x}(t)), \quad \hat{x}(0) = 0.$$

We have $X_j = X_j^{(-a_j)} + X_j^{(-a_j+1)} + \dots$, and $X_j^{(-a_j)} = \hat{X}_j$. Hence, we can write $R_j = X_j - \hat{X}_j$, and $\mathrm{ord}_p(R_j) > -a_j$. Therefore we have

$$\sum_{j=1}^k u_j(t) X_j = \sum_{j=1}^k u_j(\hat{X}_j + R_j),$$

and $\mathrm{ord}_p(R_j) > -1$. Let d be the Carnot-Carathéodory distance associated with ξ, and let d_0 be the Carnot-Carathéodory distance associated with $\hat{\xi}$, which is the distribution with framing $\hat{X}_1, \dots, \hat{X}_k$. We have $|x_j(t) - \hat{x}_j(t)| \leq Ct^2$ for some constant C, so $d_0(x(t), \hat{x}(t)) \leq Ct^{1+1/m}$, so

$$|\delta_t^*(d - d_0)| = O(t^{1+1/m}) = o(t).$$

So by Lemma 5.17 we have that the tangent cone exists and is equal to the Lie group with Lie algebra the nilpotentisation at p. $\square$

Theorem 5.19 (Mitchell for filtered structures)**.** *Let $(M, \xi_1 \subset \cdots \subset \xi_m)$ be a manifold endowed with a filtered structure, and let $p \in M$. Then the tangent cone at p exists for the global metric associated with the filtered structure, and is equal to the simply-connected graded Lie group with left-invariant global metric and with Lie algebra $L(\xi_1 \subset \cdots \subset \xi_m)(p)$, the nilpotentisation of the filtered structure at p.*

Proof. Let $X_1, \ldots, X_n$ be a framing of TM around p adapted to the filtered structure, i.e. $X_1, \ldots, X_{n_i}$ is a framing of ξ_i for $1 \leq i \leq m$. We define the endpoint map ψ based at p as:

$$\psi : \mathbb{R}^n \to M$$
$$(t_1, \ldots, t_n) \mapsto \phi_{X_n}^{t_n} \circ \cdots \circ \phi_{X_1}^{t_1}(p).$$

Then the t_i are privileged coordinates. For $r > 0$ we define the dilation

$$\delta_r(t_1, \ldots, t_i, \ldots, t_n) = (rt_1, \ldots, r^{a_i} t_i, \ldots, r^m t_n),$$

where a_i is the weight of X_i. Let $u \in L^2([0,1], \mathbb{R}^n)$. Let $x(t)$ be a solution to the control problem

$$\dot{x}(t) = \sum_{i=1}^{n} u_i(t) X_i(x(t)), \qquad x(0) = 0,$$

and let $\hat{x}(t)$ be a solution to

$$\dot{\hat{x}}(t) = \sum_{i=1}^{n} u_i(t) \hat{X}_i(\hat{x}(t)), \qquad \hat{x}(0) = 0.$$

We have $X_i = X_i^{(-a_i)} + X_i^{(-a_i+1)} + \ldots$, and $X_i^{(-a_i)} = \hat{X}_i$. Hence, we may define $R_i = X_i - \hat{X}_i$, and $\operatorname{ord}_p(R_i) > -a_i$. Therefore, we have

$$\sum_{i=1}^{n} u_i(t) X_i = \sum_{i=1}^{n} u_i(t)(\hat{X}_i + R_i).$$

Let d be the global distance associated with ξ, and let d_0 be the global distance associated with $\hat{\xi}$, the filtered structure arising from $\hat{X}_1, \ldots, \hat{X}_n$. We have $|x_i(t) - \hat{x}_i(t)| \leq Ct^{a_i+1}$ for some constant C, so $d_0(x(t), \hat{x}(t)) \leq Ct^{1+1/m}$, so

$$|\delta_t^*(d - d_0)| = O(t^{1+1/m}) = o(t).$$

So by Lemma 5.17, the tangent cone exists and is equal to the simply-connected Lie group with Lie algebra $L(\xi_1 \subset \cdots \subset \xi_m)(p)$. $\square$

6 The tangent cone as a group

In this section we will offer a geometric interpretation interpretation of the group structure of the tangent cone, and see how it arises from the filtered structure on the manifold. As we have seen in Mitchell's Theorem, the tangent cone is the Lie group whose Lie algebra is the nilpotentisation of the filtered structure at the base point. The nilpotentisation at the base point induces a group multiplication on the tangent cone. In this section we will show that the group multiplication arises from the dilations associated to the filtered structure.

We construct a group product by mimicking the standard construction for constructing a group product on a smooth manifold. Here, we think of vectors as equivalence classes of paths, and of vector fields as equivalence classes of (time-dependent) flows. The group operation then is the composition of flows, which at equivalence class level is addition.

We will explicitly define a group multiplication on the tangent cone by defining a group multiplication on equivalence classes of paths on the manifold which are rectifiable at the origin and which start at the base point. A path is rectifiable at $t = 0$ when its growth is at most linear near the origin, measured with respect to the local distance function induced by privileged coordinates. (It therefore must behave well with respect to the dilations.) Since every element of the tangent cone can be embedded in one of the equivalence classes of rectifiable paths, the group multiplication for the rectifiable paths is also a group multiplication on the tangent cone.

This chapter is based on [6] and [7], which state the results in terms of distributions. We adapted these results to suit filtered structures. For a global perspective on the tangent cone, see [11].

6.1 Rectifiable paths and rectifiable families

Let us start by giving an explicit definition for paths rectifiable at $t = 0$, and what it means for them to be equivalent. As we have said before, our main goal for this section is to find a group multiplication on the equivalence classes of these rectifiable paths.

As a reminder, the local distance function is defined as follows:

Definition 6.1. The **local distance function at p induced by the privileged coordinates** $x_1, \dots, x_n$ is

$$d_p(p, q) = d(0, (x_1, \dots, x_n)) = |x_1| + \dots + |x_i|^{1/a_i} + \dots + |x_n|^{1/m}$$

where q lies in the domain of the privileged coordinates $x_1, \dots, x_n$.

Instead of using the local distance function, one could do the same construction with the global distance function.

Definition 6.2. A smooth path $\gamma(t)$ on M with $\gamma(0) = p$ is **rectifiable at** $t = 0$ if

$$d_p(p, \gamma(t)) \leq Ct$$

as $t \to 0$ for some positive constant C, where d_p is the local distance function.

Definition 6.3. Let $\gamma_1(t)$, $\gamma_2(t)$ two smooth paths on M with $\gamma_1(0) = \gamma_2(0) = p$. Then $\gamma_1(t)$ and $\gamma_2(t)$ are **equivalent at** p if

$$t^{-1} d_p(p, \gamma_1(t) - \gamma_2(t)) \to 0$$

as $t \to 0$.

Lemma 6.4. *The tangent cone of M at p is the set of equivalence classes of all smooth paths γ with $\gamma(0) = p$ which are rectifiable at $t = 0$.*

In order to define a group product on the equivalence classes of smooth paths rectifiable at $t = 0$, we will need the notion of a family of segments rectifiable at $t = 0$ for which we can define a product. Families of segments rectifiable at $t = 0$ are a collection of paths rectifiable at $t = 0$ under some additional conditions. Similar to equivalence of rectifiable paths, we can also define equivalence of families of segments rectifiable at $t = 0$. Later, we will show that the equivalence classes of families of rectifiable segments only depend on the equivalence classes of the rectifiable paths which represent them, and therefore defines a product on equivalence classes of rectifiable paths, and therefore the tangent cone.

Definition 6.5. A **family of segments rectifiable at** $t = 0$ is a smooth map

$$\mathcal{F} : M \times I \to M$$

where I is an open neighbourhood of $0 \in \mathbb{R}$, such that

1. $\mathcal{F}(p, 0) = p$ for all $p \in M$
2. $t \mapsto \mathcal{F}(p, t)$ is uniformly rectifiable at $t = 0$ for all $p \in M$, i.e. for every compact neighbourhood $K \subset M$ of p there is a constant $C_K > 0$ and a compact neighbourhood $I_K \subset I$ of 0 such that for all $(q, t) \in K \times I_K$
$$d_q(q, \mathcal{F}(q, t)) \leq C_K t.$$

Definition 6.6. The **product** of two families $\mathcal{F}_1$, $\mathcal{F}_2$ of segments rectifiable at $t = 0$ is defined as

$$\mathcal{F}_1 \circ \mathcal{F}_2(p, t) = \mathcal{F}_1(\mathcal{F}_2(p, t), t).$$

If $\mathcal{F}_1 : M \times I_1 \to M$ and $\mathcal{F}_2 : M \times I_2 \to M$, then $\mathcal{F}_1 \circ \mathcal{F}_2 : M \times (I_1 \cap I_2) \to M$.

By Lemma 6.4, we can embed any representative of an element of the tangent cone in a family of segments rectifiable at $t = 0$. This means that if the product respects equivalence, then it defines a product on the tangent cone. First, let us check that the product of two families of segments rectifiable at $t = 0$ is a family of segments rectifiable at $t = 0$.

Lemma 6.7. *Let $\mathcal{F}_1$, $\mathcal{F}_2$ two families of segments rectifiable at $t = 0$. Then $\mathcal{F}_1 \circ \mathcal{F}_2$ is a family of segments rectifiable at $t = 0$.*

Proof. We check that $\mathcal{F}_1 \circ \mathcal{F}_2$ satisfies all conditions in the definition of a family of segments rectifiable at $t = 0$.

Because $\mathcal{F}_1$ and $\mathcal{F}_2$ are smooth, it follows that $\mathcal{F}_1 \circ \mathcal{F}_2$ is also smooth.

We have for all $p \in M$ that

$$\mathcal{F}_1 \circ \mathcal{F}_2(p, 0) = \mathcal{F}_1(\mathcal{F}_2(p, 0), 0) = \mathcal{F}_1(p, 0) = p.$$

Lastly, we check that $t \mapsto \mathcal{F}_1 \circ \mathcal{F}_2(p, t)$ is uniformly rectifiable for all $p \in M$. By definition, for all compact neighbourhoods $K \subset M$ of p there is a constant $C_K > 0$ and a compact neighbourhood $I_K \subset I$ of 0 such that for all $(q, t) \in K \times I_K$ we have

$$d_q(q, \mathcal{F}_2(q, t)) \leq C_K t.$$

Now, let $L = \mathcal{F}_2(K \times I_K)$. Because $\mathcal{F}_2$ is continuous and $\mathcal{F}_2(K \times \{0\}) = \mathrm{id}_K$, we have that L is a compact neighbourhood of p. By definition, there is a constant $C_L > 0$ and a compact neighbourhood $I_L \subset I$ of 0 such that $d_q(q, \mathcal{F}_1(q,t)) \leq C_L t$ for all $(q,t) \in L \times I_L$. In particular,

$$d_{\mathcal{F}_2(q,t)}(\mathcal{F}_2(q,t), \mathcal{F}_1(\mathcal{F}_2(q,t))) \leq C_L t$$

as $t \to 0$. Because d_q depends smoothly on the point q, there is a constant C'_L such that

$$d_q(q, \mathcal{F}_1(\mathcal{F}_2(q,t),t) - \mathcal{F}_2(q,t)) \leq C'_L t$$

as $t \to 0$. So, we have for all $(q,t) \in K \times (I_K \cap I_L)$ that

$$\begin{aligned}
d_q(q, \mathcal{F}_1 \circ \mathcal{F}_2(q,t)) &= d_q(q, \mathcal{F}_1(\mathcal{F}_2(q,t),t)) \\
&= d_q(q, \mathcal{F}_2(q,t) - \mathcal{F}_2(q,t) + \mathcal{F}_1(\mathcal{F}_2(q,t),t)) \\
&\leq d_q(q, \mathcal{F}_2(q,t)) + d_q(q, \mathcal{F}_1(\mathcal{F}_2(q,t),t) - \mathcal{F}_2(q,t)) \\
&\leq C_K t + C'_L t \\
&= (C_K + C'_L)t
\end{aligned}$$

as $t \to 0$. □

We also have the notion of equivalence, which is a similar notion to the equivalence of paths rectifiable at $t = 0$.

Definition 6.8. Two families $\mathcal{F}_1$, $\mathcal{F}_2$ of segments rectifiable at $t = 0$ are **equivalent** if

$$t^{-1} d_{\mathcal{F}_1(p,t)}(\mathcal{F}_1(p,t), \mathcal{F}_2(p,t)) \to 0 \text{ as } t \to 0$$

uniformly on compact sets on the domain of $\mathcal{F}_1 \circ \mathcal{F}_2$. We denote the equivalence by $\mathcal{F}_1 \sim \mathcal{F}_2$.

6.2 Weighted norms

We will now return more explicitly to the filtered structure, using dilations and privileged coordinates. Ultimately, this will provide a more concrete way of checking whether a smooth path is rectifiable at $t = 0$, through checking its behaviour with respect to privileged coordinates, and giving a more coordinate-based approach.

From now on, we will once again assume that M is a manifold of dimension n, endowed with a filtered structure $\xi_1 \subset \cdots \subset \xi_m = TM$. Let $n_i = \mathrm{rank}(\xi_i)$, and let $X_1, \ldots, X_n$ a framing of TM such that $X_1, \ldots, X_{n_i}$ is a framing of ξ_i. Lastly, let $-a_i$ be the order of X_i.

Given $K \subset M$ a compact neighbourhood of p, we can choose a neighbourhood $U \subset \mathbb{R}^n$ such that the map

$$\exp_p : x \mapsto \exp\left(\sum_{i=1}^{n} x_i X_i\right)(p)$$

is a diffeomorphism from U to its image $\exp_p(U)$ for all $p \in K$. Furthermore, if we fix the point $p \in K$, we can define

$$\delta_t^p \exp_p(x) = \exp_p(\delta_t(x))$$

for all t for which the right-hand side is defined. Now, let

$$\xi_i^p = (\exp_p)\xi_i.$$

Moreover, let

$$\xi^p_{i,t} = \delta_t \xi^p_i$$

and

$$\xi^p_{i,\infty} = \lim_{t\to\infty} \delta_t \xi^p_i = \lim_{t\to\infty} \xi^p_{i,t}.$$

Note that $\xi^p_1 \subset \cdots \subset \xi^p_m$ is filtered structure on $\mathbb{R}^n$ which corresponds to the grading introduced by the original filtered structure.

For the rest of this section, we assume that $||x|| = \sup_i |x_i|^{1/a_i}$. Moreover, we will write $||\cdot||_t$ when we want to indicate that the norm-distance is with respect to the framing $\delta_t^* X_1, \ldots, \delta_t^* X_n$. Note that $||\cdot||_t = ||\cdot||$. Moreover, we write $||\cdot||_\infty$ for the weighted norm with respect to the framing of the nilpotentisation. To indicate the base point, we write $||\cdot||^p_t$ or $||\cdot||^p_\infty$ or $||\cdot||^p$.

Lemma 6.9. *We have $lim_{t\to\infty}||x||_t = ||x||_\infty$.*

Proof. This follows from the fact that $\lim_{r\to\infty} r^{-a_i}\delta_r^* = \hat{X}_i$ for all $i = 1, \ldots, n$. □

In the following proposition, we show that the weighted norms are well-behaved when the base point p varies over a compact subset.

Proposition 6.10. *Given $\epsilon > 0$ and $K \subset M$ compact, there is $\delta > 0$ such that for every $p \in K$ and $0 < t < \delta$ we have*

$$(1-\epsilon)||x||^p_\infty \leq t^{-1}||\delta^p_t x||^p \leq (1+\epsilon)||x||^p_\infty.$$

Proof. Fix $p \in K$. We have

$$\frac{||x||_t}{t||\delta_{t^{-1}}x||} = 1.$$

Because $\lim_{t\to\infty} ||x||_t = ||x||_\infty$ we have

$$\frac{t^{-1}||\delta_t x||}{||x||_\infty} = 1.$$

□

6.3 The weighted norm and rectifiability

Because $\exp_p : U \to \exp_p(U)$ distorts distances by a bounded factor, we can use the norm-distance to check whether a curve is rectifiable at the origin through its behaviour with respect to $\exp_p$ and the dilations.

Corollary 6.11. *Let γ a curve in $\exp_p(U) \subset M$ with $\gamma(0) = p$. Then $\gamma(t)$ is rectifiable at $t = 0$ if and only if $(\exp_p)^{-1}\delta^p_{t^{-1}}\gamma(t)$ has a limit in $\mathbb{R}^n$ as $t \to 0$. Let this limit be x^0. Then x^0 is the unique element in $\mathbb{R}^n$ such that $t \mapsto \exp_p \delta_t x$ is equivalent to $t \mapsto \gamma(t)$.*

Proof. We have

$$\begin{aligned} ||(\exp_p)^{-1}\delta^p_{t^{-1}}\gamma(t)|| &= ||\delta_{t^{-1}}(\exp_p)^{-1}\gamma(t)|| \\ &\sim t^{-1} d_p(p, \gamma(t)) \end{aligned}$$

as $t \to 0$, where $\sim$ means distortion by a bounded factor. If $\gamma(t)$ is rectifiable at $t = 0$, then $d_p(p, \gamma(t))$ is bounded as $t \to 0$. Therefore, $\sup_i |t^{-a_i}(\exp_p)^{-1}\gamma(t)|^{1/a_i}$ is bounded as $t \to 0$. Write $x_i(t) = ((\exp_p)^{-1}\gamma(t))_i$. Then we have $x_i(t) = t^{a_i} y_i(t)$ for all i, where y_i is a smooth function. So $\lim_{t\to 0} y_i(t) = y_i(0)$.

Lastly, assume that x, x^0 are elements of $\mathbb{R}^a$ such that

$$t^{-1} d_p(p, \exp_p(\delta_t x) - \exp_p(\delta_t x^0)) \to 0$$

as $t \to 0$. Then $||x - x_0||_\infty = 0$, so $x = x_0$. □

The following lemma shows that there are explicit correspondences between paths and families of segments rectifiable at $t = 0$ and smooth paths in U. Especially, we can express paths rectifiable at $t = 0$ and families of segments rectifiable at $t = 0$ as smooth paths or functions on the manifold.

Lemma 6.12. *Let $x(t)$ be a smooth path in U with $x(0) = 0$, and let $\gamma(t) = \exp_p(x(t))$. Then the following holds:*

1. *$\gamma(t)$ is smooth, and $\gamma(0) = p$*
2. *$\gamma(t)$ is rectifiable at $t = 0$ if and only if $\lim_{t\to 0} \delta^p_{t^{-1}}\gamma(t) = \lim_{t\to 0} \exp(\delta_{t^{-1}} x(t))(p)$ exists.*
3. *Let $\mathcal{F}$ be a family of segments rectifiable at $t = 0$. Then $\lim_{t\to 0}(\exp_p)^{-1}\delta_{t^{-1}}(\mathcal{F}(p,t)) = x^0(p)$ exists, and we have*
$$(\exp_p)^{-1}\mathcal{F}(p,t) = \delta_t(x^0(p) + t x^1(p,t)),$$
where $x^0(p)$, $x^1(p,t)$ are smooth functions. Conversely, any family $\mathcal{F}(p,t)$ satisfying this condition is a family of segments rectifiable at $t = 0$.
4. *Set $\mathcal{F}^0(p,t) = \exp_p \delta_t x_0$. Then $\mathcal{F} \sim \mathcal{F}^0$.*
5. *The map $\mathcal{F} \mapsto \mathcal{F}^0$ is constant on equivalence classes.*

Proof. The first three assertions follow from Corollary 6.11.

We have

$$\begin{aligned} t^{-1} d_p(p, \mathcal{F}(p,t) - \mathcal{F}^0(p,t)) &= t^{-1} d_p(p, \exp_p(\delta_t(x^0 + t x^1)) - \exp_p(\delta_t x^0)) \\ &\sim ||x_0 + t x_1 - x_0||_\infty \to 0 \end{aligned}$$

as $t \to 0$.

Next, suppose $\mathcal{F}_1$, $\mathcal{F}_2$ are two families of segments rectifiable at $t = 0$ such that $\mathcal{F}_1 \sim \mathcal{F}_2$. For $i = 1, 2$ we have $\mathcal{F}_i \sim \mathcal{F}_i^0$, and therefore $\mathcal{F}_1^0 \sim \mathcal{F}_2^0$. So, we have

$$t^{-1}||\delta_t(x_1^0(p) - x_2^0(p))||^p_\infty = ||x_1^0(p) - x_2^0(p)||^p_\infty \to 0$$

as $t \to 0$. So for all p we have $x_1^0(p) = x_2^0(p)$. □

The following two lemmas will allow us to prove that the product on families of segments rectifiable at $t = 0$ provides a product on the equivalence classes of paths rectifiable at $t = 0$, and therefore a product on the tangent cone.

Lemma 6.13. *The product of families of segments rectifiable at $t = 0$ respects equivalence.*

Proof. Let $\mathcal{F}_1$, $\mathcal{F}_2$ be two families of segments rectifiable at $t = 0$. By the previous lemma, we have $\mathcal{F}_i \sim \mathcal{F}_i^0$ with

$$\mathcal{F}_i^0(p,t) = \exp_p(\delta_t x_i^0) \text{ and } \mathcal{F}_i(p,t) = \exp_p(\delta_t(x_i^0 + x_i^1)).$$

It suffices to prove that

$$t^{-1} d_p(p, \mathcal{F}_1 \circ \mathcal{F}_2(p,t) - \mathcal{F}_1^0 \circ \mathcal{F}_2^0(p,t)) \to 0$$

as $t \to 0$ uniformly on compact sets in M. We have

$$\mathcal{F}_1 \circ \mathcal{F}_2(p,t) = \exp_p(\delta_t(x_1^0 + tx_1^1 + x_2^0 + tx_2^1))$$

and

$$\mathcal{F}_1^0 \circ \mathcal{F}_2^0(p,t) = \exp_p(\delta_t(x_1^0 + x_2^0)).$$

So, we have

$$\begin{aligned}
&t^{-1} d_p(p, \mathcal{F}_1 \circ \mathcal{F}_2(p,t) - \mathcal{F}_1^0 \circ \mathcal{F}_2^0(p,t)) \\
&\sim ||\delta_{t^{-1}}(\delta_t(x_1^0 + tx_1^1 + x_2^0 + tx_2^1) - \delta_t(x_1^0 + x_2^0))||_\infty \\
&= ||x_1^0 + tx_1^1 + x_2^0 + tx_2^1 - (x_1^0 + x_2^0)||_\infty \\
&= ||tx_1^1 + tx_2^1||_\infty \to 0
\end{aligned}$$

as $t \to 0$ on compact sets on M. Here, $\sim$ signifies distortion with a bounded factor. Hence, $\mathcal{F}_1 \circ \mathcal{F}_2 \sim \mathcal{F}_1^0 \circ \mathcal{F}_2^0$. Because equivalence is transitive, the result follows. □

Lemma 6.14. *Let $x_1(t)$, $x_2(t)$, $y_1(t)$, $y_2(t)$ be paths rectifiable at $t = 0$ with $x_i(0) = y_i(0) = p_0$. Assume $x_1 \sim x_2$ and $y_1 \sim y_2$. Let $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{G}_1$, $\mathcal{G}_2$ be families of segments rectifiable at $t = 0$ with $\mathcal{F}_i(p_0,t) = x_i(t)$ and $\mathcal{G}_i(p_0,t) = y_i(t)$. Then*

$$t^{-1} d_{p_0}(p_0, \mathcal{F}_1 \circ \mathcal{G}_1(p_0,t) - \mathcal{F}_2 \circ \mathcal{G}_2(p_0,t)) \to 0$$

as $t \to 0$.

Proof. Set

$$\zeta_{i,p} = \lim_{t \to 0} \delta_{t^{-1}} \exp_p^{-1} \mathcal{F}_i(p,t),$$

so $\mathcal{F}_i^0 = \exp_p(\delta_t \zeta_{i,p})$. Similarly, set

$$\eta_{i,p} = \lim_{t \to 0} \delta_{t^{-1}} \exp_p^{-1} \mathcal{G}_i(p,t)$$

so that $\mathcal{G}_i^0 = \exp_p(\delta_t \eta_{i,p})$. It suffices to prove that

$$t^{-1} d_{p_0}(p_0, \mathcal{F}_1^0 \circ \mathcal{G}_1^0(p_0,t) - \mathcal{F}_2^0 \circ \mathcal{G}_2^0(p_0,t)) \to 0 \tag{1}$$

as $t \to 0$.

Because $x_1(t) = \mathcal{F}_1(p_0,t)$ is equivalent to $x_2(t) = \mathcal{F}_2(p_0,t)$, we have $\zeta_{1,p_0} = \zeta_{2,p_0}$, i.e. $\mathcal{F}_1^0(p_0,t) = \mathcal{F}_2^0(p_0,t)$. Similarly, we have $\eta_{1,p_0} = \eta_{2,p_0}$, so $\mathcal{G}_1^0(p_0,t) = \mathcal{G}_2^0(p_0,t)$. Let $p = p(t) = \exp_{p_0}(\delta_t \eta_{1,p_0}) = \exp_{p_0}(\delta_t \eta_{2,p_0})$. Then equation 1 is equal to

$$d_{p_0}\left(p_0, \delta_{t^{-1}}^{p_0}\left(\exp\left(\delta_t \sum_{j=1}^n (\zeta_{1,p})_j X_j\right) \circ \exp_{p_0}(\delta_t \eta_{1,p_0}) - \exp\left(\delta_t \sum_{j=1}^n (\zeta_{2,p})_j X_j\right) \circ \exp_{p_0}(\delta_t \eta_{2,p_0})\right)\right).$$

We have

$$\begin{aligned}
&\delta^{p_0}_{t^{-1}} \exp\left(\delta_t \sum_{j=1}^{n} (\zeta_{i,p})_j X_j\right) \circ \exp_{p_0}(\delta_t \eta_{i,p_0}) \\
&= \delta^{p_0}_{t^{-1}} \exp\left(\delta_t \sum_{j=1}^{n} (\zeta_{i,p})_j X_j\right) \circ \exp\left(\delta_t \sum_{j=1}^{n} (\eta_{i,p_0})_j X_j\right)(p_0) \\
&= \exp\left(\sum_{j=1}^{n} t^{a_j} (\zeta_{i,p})_j \delta_{t^{-1}} X_j\right) \circ \exp\left(\sum_{j=1}^{n} t^{a_j} (\eta_{i,p_0})_j \delta_{t^{-1}} X_j\right)(p_0).
\end{aligned}$$

As $t \to 0$, we have $t^{a_j} \delta_{t^{-1}} X_j \to \hat{X}_j$ on some neighbourhood of p_0. Therefore, as the expression above approaches

$$\exp\left(\sum_{j=1}^{n} (\zeta_{i,p_0})_j \hat{X}_j\right) \circ \exp\left(\sum_{j=1}^{n} (\eta_{i,p_0})_j \hat{X}_j\right)(p_0)$$

as $t \to 0$. Hence,

$$t^{-1} d_{p_0}(p_0, \mathcal{F}^0_1 \circ \mathcal{G}^0_1(p_0,t) - \mathcal{F}^0_2 \circ \mathcal{G}^0_2(p_0,t)) \to 0$$

as $t \to 0$. □

By the previous two lemmas, it follows that:

Theorem 6.15. *Let $x_1(t)$, $x_2(t)$ be equivalent smooth paths in M rectifiable at $t = 0$. Let $\mathcal{F}_1$, $\mathcal{F}_2$ be families of segments rectifiable at $t = 0$ with $\mathcal{F}_i(p,t) = x_i(t)$. Then the equivalence class of $\mathcal{F}_1 \circ \mathcal{F}_2(p,t)$ depends only on the equivalence classes of $x_1(t)$ and $x_2(t)$, and hence defines a product of elements of the tangent cone at p.*

References

[1] A. Bellaïche. The tangent space in sub-riemannian geometry. In A. Bellaïche and J.-J. Risler, editors, *Sub-Riemannian Geometry*, volume 144 of *Progress in Mathematics*. Birkhäuser Verlag, 1996.

[2] N. Bourbaki. *Lie Groups and Lie Algebras*. Springer-Verlag, 1975.

[3] Á. del Pino. Lecture notes: topological aspects in the study of tangent distributions, 2018.

[4] B. C. Hall. *Lie Groups, Lie Algebras, and Representations*. Springer-Verlag, 2003.

[5] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2 edition, 2013.

[6] 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.

[7] J. Mitchell. On carnot-carathéodory metrics. *Journal of Differential Geometry*, 21:35–45, 1985.

[8] R. Montgomery. *A Tour of Subriemannian Geometries, Their Geodesics and Apllications*. Mathematical Surveys and Monographs. American Mathematical Society, 2006.

[9] M. K. Neusser. *Weighted Jet Bundles and Differential Operators for Parabolic Geometries*. PhD thesis, Universität Wien, 2010.

[10] S. Sternberg. Lie algebras, 2004.

[11] E. van Erp and R. Yuncken. On the tangent groupoid of a filtered manifold. *arXiv e-prints*, page arXiv:1611.01081, nov 2016.

[12] A. M. Vershik and V. Y. Gershkovich. Nonholonomic problems and the theory of distributions. *Acta Applicandae Mathematicae*, 12:181–209, 1988.